Oracle® Secure Enterprise Search

Administrator's Guide

11*g* Release 2 (11.2.1)

E17332-04

May 2011

ORACLE®

Oracle Secure Enterprise Search Administrator's Guide, 11*g* Release 2 (11.2.1)

E17332-04

Primary Author: Donna Carver, Vishwanath Sreeraman, Kathy Rich, Michele Cyran

Contributors: Shashi Anand, Sachin Bhatkar, Greg Brunet, Stefan Buchta, Yujie Cao, Thomas Chang, Mohammad Faisal, Roger Ford, Cindy Hsin, Marvin Huang, Diego Iglesias, Rahul Joshi, Sana Karam, Hiroshi Koide, Belinda Leung, Valarie Moore, Huyen Nguyen, Yiming Qi, Birinder Tiwana, Luke Wang, Steve Yang, Yan Zhao

Contents

Preface xiii

Audience xiii
Documentation Accessibility xiii
Related Documents xiii
Conventions xiv

What's New xv

Features in Release 11.2.1 xv

Part I Learning the Basics

1 Introduction to Oracle Secure Enterprise Search

Overview of Oracle Secure Enterprise Search 1-1
Source Types 1-2
Oracle Secure Enterprise Search Components 1-3
Oracle Secure Enterprise Search Administration GUI 1-3
Oracle Secure Enterprise Search Crawler 1-4
Oracle Secure Enterprise Search APIs 1-5
Secure Search in Oracle Fusion Applications 1-6
Scheduling Crawls 1-6
Administrative Tools 1-7
Oracle Secure Enterprise Search Features 1-7
Secure Search 1-8
Federated Search 1-8

2 Getting Started with the Oracle SES Administration GUI

Getting Started Basics for the Administration GUI 2-1
Understanding the Oracle SES Administration GUI 2-2
Home Tab 2-2
Search Tab 2-3
Global Settings Tab 2-4
Starting and Stopping Oracle SES 2-5

3 Understanding Crawling

Overview of the Oracle Secure Enterprise Search Crawler 3-1
Crawler URL Queue 3-2
Understanding Access URLs and Display URLs 3-2
Modifying the Crawler Parameters 3-2
Overview of Crawler Settings 3-3
Crawling Mode 3-3
URL Boundary Rules 3-4
Inclusion Rules 3-4
Exclusion Rules 3-4
Examples of Inclusion and Exclusion Rules 3-5
Document Types 3-5
Crawling Depth 3-6
Robots Exclusion 3-6
Index Dynamic Pages 3-6
Title Fallback 3-7
Character Set Detection 3-7
Special Considerations with Automatic Character Set Detection 3-8
Language Detection 3-8
Deleting the Secure Cache 3-8
Overview of XML Connector Framework 3-9
Example Using the XML Connector 3-10
XML Configuration File 3-10
Configuring Support for Image Metadata 3-12
Identifying the Search Attributes for Image Metadata 3-12
Supporting XMP Metadata 3-14
Supporting DICOM Metatags 3-15
Example: Adding an Attribute to the Default attr-config.xml File 3-18
Creating an Image Document Service Connector 3-18
Using the Image Document Service Connector 3-19
Searching Image Metadata 3-19
Troubleshooting the Image Document Service Connector 3-20
Overview of Attributes 3-20
Attributes For Different Source Types 3-21
Using Lists of Values for Search Attributes 3-21
System-Defined Search Attributes 3-22
Understanding the Crawling Process 3-22
The Initial Crawl 3-23
Queuing and Caching Documents 3-23
Indexing Documents 3-23
Oracle SES Stoplist 3-23
Maintenance Crawls 3-24
Automatic Forced Recrawls 3-24
Monitoring the Crawling Process 3-25
Crawler Statistics 3-25
Crawler Log Files 3-26
Crawler Configuration 3-27

Parallel Query Indexing 3-27
Document Partition Model and Storage Areas 3-27

4 Customizing the Search Results

Adding Suggested Content in Search Results 4-1
Suggested Content Providers 4-1
Security Options 4-2
Example Configuring Google OneBox for Suggested Content 4-3
Customizing the Relevancy of Search Attributes 4-3
Providing Faceted Navigation 4-4

Part II Creating Data Sources

5 Configuring Access to Built-in Sources

Setting Up Web Sources 5-1
Boundary Rules for Web Sources 5-3
Web Document Attributes 5-3
Setting Up Table Sources 5-4
Table Search Attributes 5-5
Setting Up File Sources 5-5
File Document Attributes 5-6
Tips for Using File Sources 5-7
Crawling File Sources with Non-ASCII Character Sets 5-7
Crawling File Sources with Symbolic Links 5-7
Crawling File URLs 5-7
Crawling File Sources from a Network Drive 5-7
Setting Up E-Mail Sources 5-8
Setting Up Mailing List Sources 5-8
Setting Up OracleAS Portal Sources 5-10
Crawling a Folder or Page 5-10
OracleAS Portal Search Attributes 5-10
Tips for Using OracleAS Portal Sources 5-11
Setting Up Federated Sources 5-12
Federation Trusted Entities 5-13
Example Creating a Federated Source 5-15
Customizing Federated Sources 5-15
Route Queries to the Federated Source 5-15
Set Search Restrictions 5-16
Retrieve Attributes 5-17
Map Attributes 5-17
Tips for Using Federated Sources 5-18
Looping Among Federated Sources 5-18
Federated Search Characteristics 5-18
Federated Search Limitations 5-19

6 Configuring Access to Content Management Sources

Setting Up EMC Documentum Content Server Sources 6-1
Important Notes for EMC Documentum Content Server Sources 6-1
Required Software 6-1
Required Tasks 6-2
Known Issues 6-3
Configuration for Documentum Content Server 6.5 6-3
Setting Up Identity Management for EMC Documentum Content Server 6-4
Activating the Documentum Identity Plug-in 6-4
Activating the Oracle Internet Directory Identity Plug-In 6-5
Activating the AD Identity Plug-In 6-6
Activating SunOne Identity Plug-In 6-8
Creating an EMC Documentum Content Server Source 6-10
Setting Up Microsoft SharePoint Sources 6-11
Important Notes About SharePoint 2007 Sources 6-12
Known Limitations of the SharePoint 2007 Connector 6-12
Known Issues for SharePoint 2007 Connector 6-13
Supported Platforms 6-14
Creating a SharePoint 2007 Source 6-14
Deploying the Web Service on MOSS 2007 6-18
Setting Up Oracle Content Database Sources 6-18
Important Notes for Oracle Content Database Sources 6-18
Setting Up Identity Management for Oracle Content Database Sources 6-19
Creating an Oracle Content Database JDBC Source 6-19
Creating an Oracle Content Database Source 6-21
Required Tasks for Oracle Content Database Release 10.1.3 6-22
Oracle Content Database Source Attributes 6-24
Setting Up Oracle Content Server Sources 6-25
Oracle Content Server Security Model 6-26
Roles and Groups 6-27
Accounts 6-27
Setting Up Identity Management for Oracle Content Server 6-28
Creating an Oracle Content Server Source 6-28

7 Configuring Access to Collaboration Sources

Setting Up EMC Documentum eRoom Sources 7-1
Documentum eRoom Web Services 7-2
Important Notes for Documentum eRoom Sources 7-2
Supported Platforms 7-2
Required Software 7-2
Required Tasks 7-2
Known Issues 7-3
Creating a Documentum eRoom Source 7-3
Setting Up Lotus Notes Sources 7-4
Important Notes for Lotus Notes Sources 7-5
Required Software 7-5
Required Tasks 7-5

Known Issues ... 7-6
Setting Up Identity Management for Lotus Notes ... 7-6
Creating a Lotus Notes Source ... 7-6
Displaying the Parent URL in the Search Results ... 7-8
Setting Up Microsoft Exchange Sources ... 7-8
Important Notes for Microsoft Exchange Sources ... 7-9
Required Software ... 7-9
Required Tasks ... 7-9
Known Issues ... 7-10
Setting Up Identity Management for Microsoft Exchange ... 7-10
Creating a Microsoft Exchange Source ... 7-11
Microsoft Exchange Source Attributes ... 7-11
Setting Up NTFS Sources for Windows ... 7-11
Important Notes for NTFS Sources ... 7-12
Required Software ... 7-12
Required Tasks ... 7-12
Setting Up Identity Management for NTFS Sources ... 7-13
Creating an NTFS Source ... 7-13
NTFS Source Attributes ... 7-14
Setting Up NTFS Sources for UNIX ... 7-14
Important Notes for NTFS Sources ... 7-14
Required Software ... 7-15
Setting Up Identity Management with NTFS Sources ... 7-15
Creating an NTFS Source ... 7-15
Installing and Configuring Windows Services ... 7-16
Required Software ... 7-16
Required Tasks ... 7-16
Installing Oracle Search File Change Detector ... 7-17
Modifying the File Change Detector Configuration File ... 7-17
Installing the NTFS Web Service ... 7-19
Configuring the NTFS Connector ... 7-21
Known Issues ... 7-21
Setting Up Oracle Calendar Sources ... 7-22
Setting Up Identity Management for Oracle Calendar ... 7-22
Creating an Oracle Calendar Source ... 7-22
Oracle Calendar Attributes ... 7-23
Setting Up Oracle Collaboration Suite E-Mail Sources ... 7-23
Important Notes for Oracle Collaboration Suite E-Mail Sources ... 7-23
Required Tasks ... 7-24
Setting Up Identity Management for Oracle Collaboration Suite E-Mail Sources ... 7-24
Creating an Oracle Collaboration Suite E-Mail Source ... 7-24

8 Configuring Access to Applications Sources

Setting Up Oracle Fusion Sources ... 8-1
Setting Up Identity Management System ... 8-1
Defining a Fusion Source ... 8-2
Setting Up Oracle WebCenter Sources ... 8-3

Defining a WebCenter Source ... 8-3
Setting Up Oracle E-Business Suite Sources ... 8-4
Setting Up Database Sources ... 8-6
Required Columns in Database Sources ... 8-6
Optional Columns in Database Sources ... 8-7
Configuring the JDBC Driver ... 8-8
Query File XML Schema Definition ... 8-8
Creating Public Database Sources ... 8-9
Defining User-Defined Security for Database Sources ... 8-11
Database Search Attributes ... 8-12
Example of Creating a Database Source With User-Defined Security ... 8-12
Setting Up Siebel 7.8 Sources ... 8-14
Requirements for Siebel 7.8 Sources ... 8-14
Installing the JDBC Driver for Microsoft SQL Server ... 8-15
Starting the Decompression Server ... 8-15
Setting Up Identity Management for Siebel 7.8 ... 8-16
Creating a Secured Siebel 7.8 Source ... 8-16
Creating a Public Siebel 7.8 Source ... 8-19
Queries to Crawl Siebel 7.8 Business Components ... 8-19
Service Request Attachments ... 8-20
Accounts ... 8-21
Products ... 8-22
Literature ... 8-23
Solution ... 8-24
Service Request ... 8-25
Contacts ... 8-26
Activity ... 8-27
Activity Attachment ... 8-29
Setting Up Siebel 8 Sources ... 8-32

Part III Advanced Topics

9 Security in Oracle Secure Enterprise Search

Overview of Oracle Secure Enterprise Search Security ... 9-1
Oracle Secure Enterprise Search Security Model ... 9-1
Changing the Administration Password ... 9-2
Temporary Passwords ... 9-4
Authentication and Authorization ... 9-4
About Oracle SES Authentication ... 9-4
About Oracle SES User Authorization ... 9-4
Restrictions on Changing the ACL Policy ... 9-6
Activating an Identity Plug-in ... 9-7
Re-registering Pre-Installed Identity Plug-ins ... 9-9
Restrictions on Changing the Identity Plug-in ... 9-10
Authentication Methods ... 9-10
Oracle Secure Enterprise Search User Repository ... 9-11
Oracle SES Authentication Interface ... 9-11

Enabling Secure Search 9-12
User Authorization Cache 9-12
Federated User Authorization Cache 9-14
Modifying the Remote Cache Configuration File 9-15
XML Schema Definition for Remote Cache Configuration Files 9-16
Administrator-Based Authorization 9-17
Identity-Based Secure Search 9-17
Query-time Authorization 9-18
Self Service Authorization 9-19
Configuring Secure Search with OracleAS Single Sign-On 9-20
Configuring Oracle HTTP Server 9-21
Configuring OracleAS and Oracle SES for Single Sign-on Security 9-25
Adding OSSO Identity Asserter 9-26
Adding Oracle Internet Directory Authenticator 9-26
Configuring Secure Search with Oracle Access Manager Single Sign-On 9-28
Configuring Oracle Identity Management 9-28
Configuring Oracle HTTP Server 9-29
Installing and Configuring WebGate 9-30
Creating a WebGate Instance 9-30
Installing WebGate 9-30
Updating the WebGate Web Server Configuration 9-30
Integrating Oracle Access Manager with Oracle SES 9-30
Configuring QueryPlan.xml in Oracle SES 9-33
SSL and HTTPS Support in Oracle Secure Enterprise Search 9-34
Understanding SSL 9-34
Managing the Keystore 9-35
Importing SSL Certificates into the Java Virtual Machine 9-36
Maintaining a Keystore 9-36
Oracle SES Acting as an SSL Client 9-36
Oracle SES Acting as an SSL Server 9-37
Configuring Oracle Secure Enterprise Search to Require SSL 9-37
Configuring Oracle HTTP Server to Require SSL 9-39
Changing the Master Encryption Key 9-43

10 Administering Oracle SES Instances

Increasing Data Storage Capacity 10-1
Tuning Crawl Performance 10-4
Crawler Schedule 10-5
Crawler Schedules in Oracle Fusion Applications 10-5
Stuck Scheduling Requests 10-6
Proxy Servers 10-6
Boundary Rules 10-7
Notes for File Sources 10-7
Dynamic Pages 10-8
Crawler Depth 10-8
Robots Rule 10-8
Duplicate Documents 10-8

Redirected Pages 10-9
URL Looping 10-9
Oracle Redo Log 10-10
What to Do Next 10-11
Tuning Search Performance and Scalability 10-12
Suggested Links 10-12
Authentication and Authorization 10-13
Parallel Query and Index Partitioning 10-14
Storage Areas 10-14
Configuring a Partition 10-14
Index Fragmentation 10-14
Indexing Parameters 10-15
Indexing Batch Size 10-15
Indexing Memory Size 10-16
Parallel Indexing Degree 10-16
Search Statistics 10-16
Relevancy Boosting 10-16
Load Balancing on Oracle RAC 10-17
Configuring a Small Index 10-17
WebLogic Search Server Configuration 10-17
Database Initialization Parameters 10-18
Oracle UNDO Tablespace 10-19
Buffer Cache 10-19
Turning On Debug Mode 10-20
Supporting Failover in Oracle RAC 10-20
Monitoring Oracle Secure Enterprise Search 10-21
Integrating with Google Desktop 10-21
Accessing the Oracle WebLogic Server Administration Console 10-21

11 Oracle Secure Enterprise Search APIs

Overview of Oracle Secure Enterprise Search APIs 11-1
Oracle Secure Enterprise Search Web Services APIs 11-1
Web Services APIs Installation 11-2
Query Web Services Location 11-2
Administration Web Services Location 11-3
Web Services Concepts 11-3
Web Services 11-3
Simple Object Access Protocol 11-4
Web Services Description Language 11-4
Web Services Architecture 11-4
Development Platforms 11-5
Query Web Services Common Data Types 11-5
Base Data Types 11-5
XML-to-Java Data Type Mappings 11-6
Complex Types 11-6
Array Types 11-10
Query Web Services Operations 11-10

Overview of Query Web Services Operations 11-10
Authentication Operations 11-11
Search Operations 11-12
Browse Operations 11-19
Metadata Operations 11-20
Search Hit Operations 11-22
User Feedback Operations 11-24
Query Web Services Query Syntax 11-24
Search Term 11-24
Phrase 11-24
Operators 11-25
Default Search: Implicit AND 11-25
Word Separator 11-25
Filter Conditions (Advanced Conditions) 11-25
Special Search Terms 11-25
Query Web Services Example: Basic Search 11-26
Default-Factor Element 11-28
Query Web Services Example: Customizing Relevancy 11-29
Filter Element 11-29
Ranking Element 11-30
Global-Settings Element 11-30
Custom-Factor Element 11-31
Applying Ranking Factors 11-32
Client-Side Query Java Proxy Library 11-33

A XML Connector Examples and Schemas

Configuration File XML Schema Definition A-1
Control Feed Example A-2
Control Feed XML Schema Definition A-3
Data Feed Example A-5
Data Feed XML Schema Definition A-7

B URL Crawler Status Codes

C Third Party Licenses

Apache Software C-1
Eclipse Software C-4
Egothor Software C-8
Javascript Bubbling Library C-9
Plug-in Software C-9
Snowball Software C-10
Visigoth Software C-10
Yahoo! Inc. C-11

D Error Messages

Index

Preface

The *Oracle Secure Enterprise Search Administrator's Guide* explains how to administer Oracle Secure Enterprise Search instances. You will learn how to set up a variety of information sources, crawl and index those sources, and customize the search results.

This Preface contains these topics:

- Audience
- Documentation Accessibility
- Related Documents
- Conventions

Audience

Oracle Secure Enterprise Search Administrator's Guide is intended for administrators and application developers who perform the following tasks:

- Install and configure Oracle Secure Enterprise Search
- Administer Oracle Secure Enterprise Search
- Develop Oracle Secure Enterprise Search applications

Documentation Accessibility

For information about Oracle's commitment to accessibility, visit the Oracle Accessibility Program website at
http://www.oracle.com/pls/topic/lookup?ctx=acc&id=docacc.

Access to Oracle Support

Oracle customers have access to electronic support through My Oracle Support. For information, visit
http://www.oracle.com/pls/topic/lookup?ctx=acc&id=info or visit
http://www.oracle.com/pls/topic/lookup?ctx=acc&id=trs if you are hearing impaired.

Related Documents

For more information about Oracle Secure Enterprise Search, refer to the following resources:

- *Oracle Fusion Applications Installation Guide*

Provides instructions for provisioning an Oracle Fusion Applications environment and installing applications.

- *Oracle Fusion Applications Enterprise Deployment Guide*

  Provides information for system administrators who are responsible for implementing and configuring Oracle Fusion Applications enterprise deployments.

- *Oracle Fusion Applications Administrator's Guide*

  Provides information for administrators of enterprise applications.

- *Oracle Secure Enterprise Search Release Notes*

  Provides version information and identifies known issues.

- *Oracle Secure Enterprise Search Administration API Guide*

  Provides a guide to the various interfaces to the Administration API.

Up-to-date Release Notes are posted on Oracle Technology Network (OTN). You must register online before using OTN. Registration is free and can be done at this location:

`http://www.oracle.com/technetwork/community/join/overview/index.htm`

If you have a user name and password for OTN, then you can go directly to the documentation section of OTN at this location:

`http://www.oracle.com/technetwork/indexes/documentation/index.html`

Conventions

The following text conventions are used in this document:

Convention	Meaning
boldface	Boldface type indicates graphical user interface elements associated with an action, or terms defined in text or the glossary.
italic	Italic type indicates book titles, emphasis, or placeholder variables for which you supply particular values.
`monospace`	Monospace type indicates commands within a paragraph, URLs, code in examples, text that appears on the screen, or text that you enter.
`MW_HOME`	*`MW_HOME`* is the top-level directory for Oracle Fusion Middleware products. This directory is called Middleware home. The Middleware home for Fusion Applications is named `fusionapps`. Each application is installed in its own Oracle home under *MW_HOME*.
`ORACLE_HOME`	*`ORACLE_HOME`* for Oracle Secure Enterprise Search represents the path *`MW_HOME`*`/ses` by default; a different name can be specified during installation. It is the directory where Oracle SES is installed.
`/`	A slash separates levels of a directory path. On Windows, use a backslash (`\`) instead of a slash (`/`).

What's New

This chapter provides a brief description about the new features available in Oracle Secure Enterprise Search (Oracle SES) 11*g* Release 2 (11.2.1). It also provides pointers to additional information.

Features in Release 11.2.1

Oracle SES is now integrated with Oracle Fusion Applications to provide secure search functionality for all of its enterprise applications, including WebCenter, Universal Content Management (UCM), and Business Intelligence. Oracle Fusion Applications is Oracle's next-generation applications suite built on a service-oriented platform. It brings together next-generation enterprise technologies, applications, and services, including Oracle Fusion Middleware, which supports complex, distributed software applications.

To provide a search solution for Oracle Fusion Applications, Oracle SES has transitioned to an open architecture from the bundled architecture of earlier releases. This open architecture enables you to install Oracle SES in an existing Oracle Database -- either a single instance or Oracle RAC -- and in an existing WebLogic domain -- either a single managed server or a WebLogic cluster.

See Also: "Secure Search in Oracle Fusion Applications" on page 1-6

New features for administrators:

- Oracle SES is installed with Oracle Fusion Applications. You can install Oracle Database and WebLogic Server at the same time, or you can install Oracle SES into an existing Database 11.2.0.2 or WebLogic Server 11.1.1. Refer to the *Oracle Fusion Applications Installation Guide*.

  Installation topology changes:

  - The Oracle SES middle-tier software bits are installed in Fusion Middleware Home (MW_HOME). *ORACLE_HOME* for is named `ses` by default.
  - The Oracle SES middle-tier components are deployed in the SES cluster in the Fusion Middleware Common Domain. This change improves scalability.
  - The crawler runs on the Oracle Enterprise Scheduler cluster in the WebLogic Common Domain. Be sure to provide sufficient resources in this cluster for the crawler. See "Secure Search in Oracle Fusion Applications" on page 1-6.
  - The Oracle SES schema is installed in the Fusion database (either a single instance of Oracle Database or Oracle RAC). Oracle RAC enables a single database to run across a cluster of servers, making it highly fault tolerant,

scalable, and available. See "Supporting Failover in Oracle RAC" on page 10-20.

  - Oracle SES uses these default tablespaces: SEARCH_DATA, SEARCH_INDEX, and SEARCH_TEMP. You or your Oracle Database administrator can create additional tablespaces before initiating the first crawl, and you can add data files and temp files to the existing tablespaces any time after Oracle Database is installed. See "Increasing Data Storage Capacity" on page 10-1.

- Upgrading from previous releases of Oracle SES is not supported.
- The administrative user is named SEARCHSYS.
- The following configuration files have been removed: `search.properties`, `search.config`, `searchctl.conf`, `crawler.dat`, and `ranking.xml`. The configuration parameters from these files were moved to Oracle Database. Some parameters can be modified using the Administration API, while others are not configurable.
- You can change the master encryption key with the `searchctl rollover_key` command. See "Changing the Master Encryption Key" on page 9-43.
- Schedules are managed by Oracle Enterprise Scheduler. See "Crawler Schedule" on page 10-5.
- The RSS crawler is recoverable when it is used in control feed mode. On restart, the crawler resumes processing from where it was interrupted.
- XML data feeds do not have a default content type. If the content type is not specified explicitly, then Oracle SES automatically detects the data type. Some filtering overhead is incurred in this process.
- Debug mode is turned on using either Oracle Enterprise Manager or the WebLogic Server Administration Scripting Tool. See "Turning On Debug Mode" on page 10-20.
- Faceted navigation can be configured through the Administration API. See "Providing Faceted Navigation" on page 4-4.
- Wildcard queries are not enabled by default. You can turn on wildcards by modifying the `queryConfig` object using the Administration API. See *Oracle Secure Enterprise Search Administration API Guide*.

New features for end users:

- Oracle SES supports crawling of the following additional document formats:
  - Microsoft Office 2008 for Mac. You can search for Word, Excel, and PowerPoint files.
  - PDF version 1.7
- In addition, there are some enhancements to how Oracle SES searches the following document formats:
  - Microsoft PowerPoint 97- 2003: You can now search files that are set to Read Only.
  - Microsoft Office 2007: Support for SmartArt.

Disabled features in this release:

- Oracle SES Search Application; use the Oracle Fusion Applications Search User Interface instead
- `searchctl` command operations except `rollover_key`

- These source types:
  - FileNet Content Engine
  - FileNet Image Services
  - Hummingbird
  - IBM DB2
  - Open Text Livelink
- Creating, deleting, or modifying user-defined source types
- User-defined document service managers
- Custom authentication plug-ins
- Custom authorization plug-ins
- Custom URL rewriter plug-ins
- Windows native authentication
- Backup and recovery; see the *Oracle Fusion Middleware Administrator's Guide* for alternative backup and recovery procedures
- Space management
- Configuration of these crawler parameters:

  ```
  archiveFileRecurseDepth
  CACHEFULLACTION
  CACHE_QUEUE_HIGHWATER_MARK
  CACHE_QUEUE_LOWWATER_MARK
  COLLECT
  FILE_WRITE_BUF_SIZE
  FILTERPATH NO_FILTER
  FILTER_OPTIONS NO_PDF_ROTATE (obsolete)
  IDM_USER_CACHE_SIZE
  IDM_GROUP_CACHE_SIZE
  INDEX_PARSED_ATTRIBUTES
  MAPPING
  MAX_DOC_PLAINTEXT_LENGTH
  portalNoIndexContainerPage
  SECURE_JDBC_CC
  SECURE_JDBC_CT
  SECURE_JDBC_EC
  SECURE_JDBC_ET
  SQL_CALLBACK
  SQL_COMMAND_HOOK
  SQL_RESPONSE_HOOK
  SYSTEM_PROPERTIES
  zipFileJavaPackage
  ```

- Configuration of these query parameters:

  ```
  ses.qapp.allowed_redirect_hosts
  ses.qapp.convert_timezone
  ses.qapp.group_tab_order
  ses.qapp.groupable_attrs
  ses.qapp.email.max_thread_docs
  ses.qapp.email.thread_sort_by
  ses.qapp.sortable_attrs
  multiSeqThreshold
  userStatsTypes
  ```

Part I

Learning the Basics

This part provides the information you need to administer Oracle SES instances. It contains the following chapters:

- Chapter 1, "Introduction to Oracle Secure Enterprise Search"
- Chapter 2, "Getting Started with the Oracle SES Administration GUI"
- Chapter 3, "Understanding Crawling"
- Chapter 4, "Customizing the Search Results"

1

Introduction to Oracle Secure Enterprise Search

This chapter describes the basic components of Oracle Secure Enterprise Search: the sources, crawler, and user interfaces. It contains the following topics:

- Overview of Oracle Secure Enterprise Search
- Source Types
- Oracle Secure Enterprise Search Components
- Secure Search in Oracle Fusion Applications
- Oracle Secure Enterprise Search Features

Overview of Oracle Secure Enterprise Search

Oracle Secure Enterprise Search enables a secure, high quality, easy-to-use search across all enterprise information assets. Key features include:

- The ability to search and locate public, private and shared content across Intranet Web servers, databases, IMAP e-mail, document management systems, applications, and portals
- Highly secure crawling, indexing, and searching
- A simple, intuitive search interface leading to an excellent user experience
- Excellent search quality, with the most relevant items for a query shown first, even when the query spans diverse public and private data sources
- Analytics on search results and usage patterns
- Sub-second query performance
- Ease of administration and maintenance, leveraging existing IT expertise

See Also:

- *Oracle Secure Enterprise Search Installation Guide* for requirements, tips, and information on getting started using Oracle SES
- Oracle Technology Network for updated information on known issues, code samples, and best practices:

  `http://www.oracle.com/technetwork/search/oses/overview/index.html`
- The *Oracle Secure Enterprise Search Release Notes* for version information and known issues

Source Types

A collection of information is called a source. Each source has a type that identifies where the information is stored, such as on a Web site or in a database table. Oracle SES provides several built-in source types and an architecture for adding new types.

Additionally, Oracle SES provides access to more third-party data repositories than any other enterprise search engine, without requiring you to generate any additional coding. While these data sources are classified as user-defined source types, they are available as the built-in source types. This guide organizes these user-defined source types into content management sources, collaboration sources, and applications sources.

Oracle SES also provides authorization cache sources for facilitating access to secure data.

Built-in Sources

- **Web**: Represents the content on a specific Web site. Web sources facilitate maintenance crawling of specific Web sites.
- **Table**: Represents content in a table or view in Oracle Database.
- **File**: The set of documents that can be accessed through the file system protocol.
- **E-mail**: Derives content from e-mails sent to a specific e-mail address. When Oracle SES crawls an e-mail source, it collects e-mail from all folders set up in the e-mail account, including Drafts, Sent Items, and Trash e-mails.
- **Mailing list**: Derives its content from e-mails sent to a specific mailing list.
- **OracleAS Portal**: Lets you search across multiple OracleAS Portal repositories, such as Web pages, files on disk, and pages in other OracleAS Portal instances.
- **Federated**: Enables you to share content across multiple Oracle SES instances.

Content Management Sources

- EMC Documentum Content Server
- Microsoft SharePoint 2007
- Oracle Content Database
- Oracle Content Database (JDBC)
- Oracle Content Server (formerly Stellent Content Server)

You may need to install client libraries and obtain a license from the vendor for some content sources to work. For example, EMC Documentum requires installation of a

compatible version of Documentum Foundation Classes (DFC), which is a Java library, on the computer running Oracle SES. Oracle SES does not ship with DFC.

Collaboration Sources

- EMC Documentum eRoom
- IBM Lotus Notes
- Microsoft Exchange
- Microsoft NT File Systems (NTFS)
- Oracle Calendar
- Oracle Collaboration Suite E-Mail

Oracle Applications Sources

- Database
- Oracle E-Business Suite
- Siebel 7.8
- Siebel 8 (Public)
- Oracle Fusion
- Oracle WebCenter

Authorization Sources

- User Authorization Cache
- Federated User Authorization Cache

> **See Also:** *Oracle Secure Enterprise Search Release Notes* for a list of supported platforms

Oracle Secure Enterprise Search Components

Oracle SES includes the following components:

- Oracle Secure Enterprise Search Administration GUI
- Oracle Secure Enterprise Search Crawler
- Oracle Secure Enterprise Search APIs

Oracle Secure Enterprise Search Administration GUI

The Oracle Secure Enterprise Search Administration GUI enables you to manage and monitor Oracle SES components using a browser-based interface. These are among the tasks that you perform:

- Define sources and crawling scope
- Configure the search application
- Monitor crawl progress and search quality
- Customize search results

See Also:

- "Understanding the Oracle SES Administration GUI" on page 2-2
- Oracle SES administration tutorial for help understanding common administrator tasks:

  `http://st-curriculum.oracle.com/tutorial/SESAdmin Tutorial/index.htm`
- Oracle SES Administration GUI Help

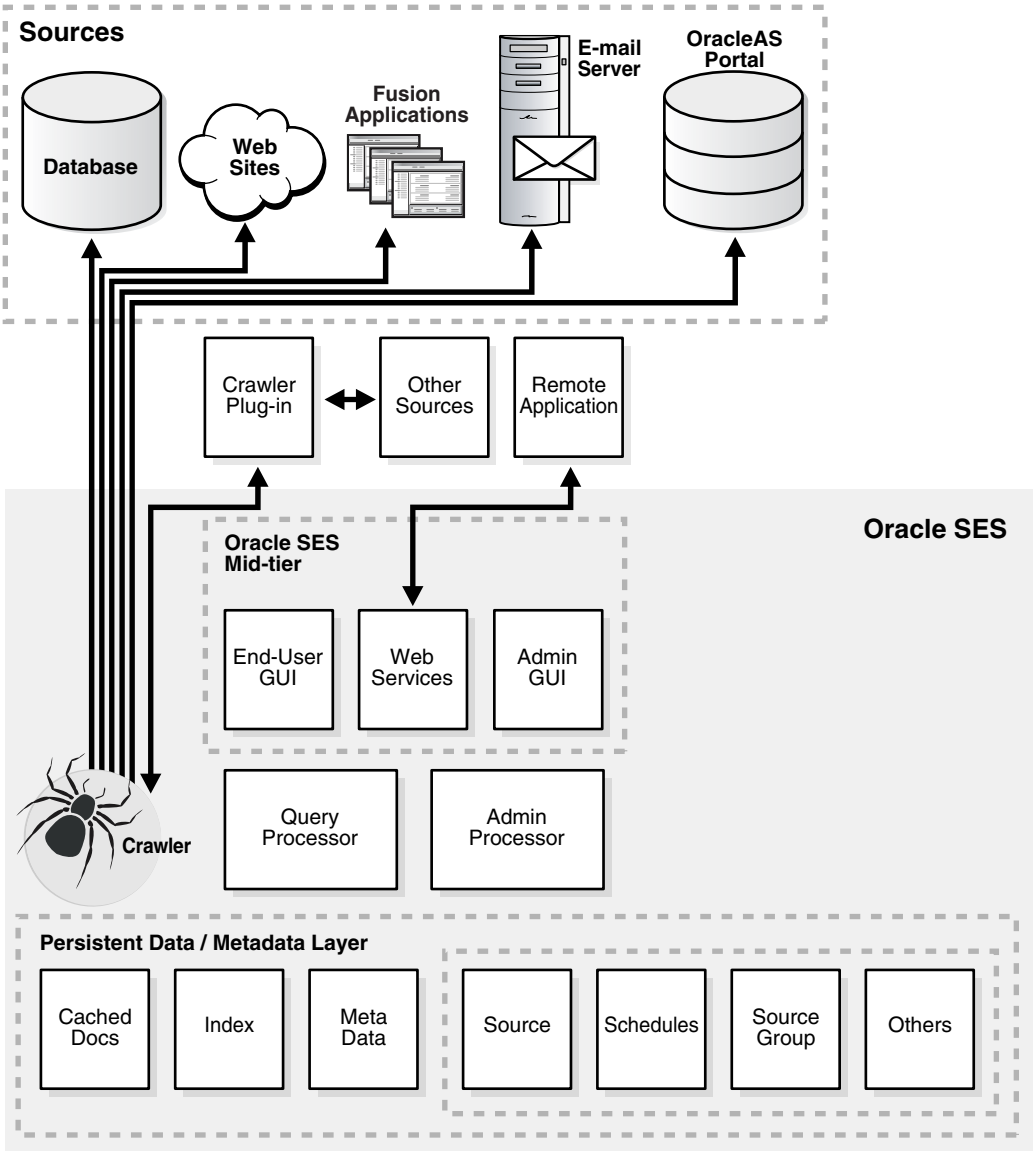
Oracle Secure Enterprise Search Crawler

Oracle SES uses a crawler to collect data from the sources. The Oracle SES crawler is a Java process activated by a schedule. When activated, the crawler spawns a configurable number of processor threads that fetch information from various sources and index the documents. This **index** is used for searching **sources**.

The crawler maps links and analyzes relationships. Whenever the crawler encounters embedded non-HTML, or non-textual documents during the crawling, it automatically detects the document type, and filters and indexes the document.

Figure 1–1 shows the crawler in relation to other Oracle SES components and a variety of data sources.

Figure 1–1 Crawler Collecting Information for Oracle SES



See Also: Chapter 3, "Understanding Crawling"

Oracle Secure Enterprise Search APIs

Oracle Secure Enterprise Search provides several APIs. For example, with the Web Services API, you can integrate Oracle SES search capabilities into your search application. Using the Administration API, you can manage multiple installations of Oracle SES more easily than using a graphical interface.

See Also:

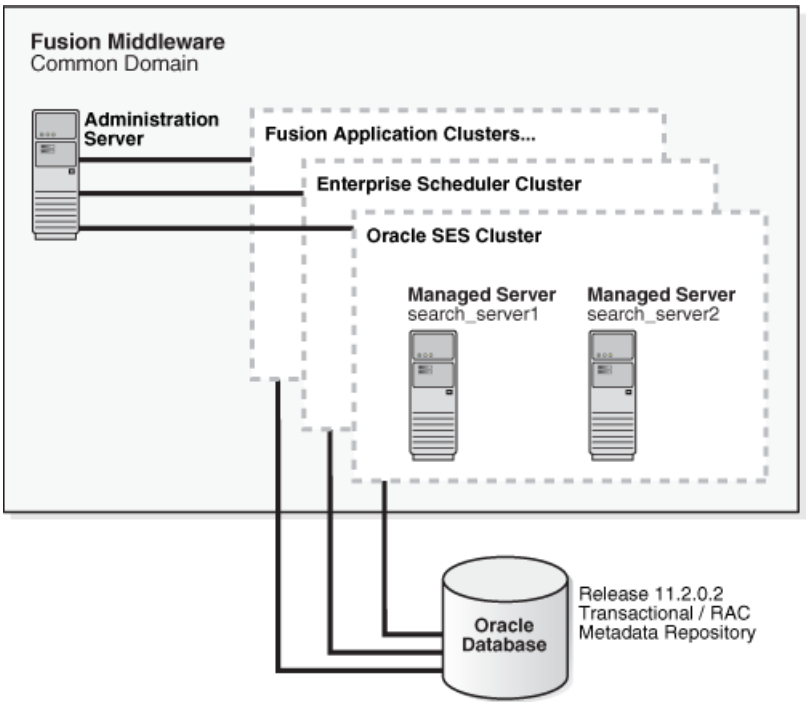
- Chapter 11, "Oracle Secure Enterprise Search APIs"
- *Oracle Secure Enterprise Search Administration API Guide*
- *Oracle Secure Enterprise Search Java API Reference*

Secure Search in Oracle Fusion Applications

Oracle SES is integrated with Oracle Fusion Applications to provide secure search functionality for all of its enterprise applications. In Fusion Applications, Oracle SES has an open architecture that enables you to install Oracle SES in an existing Oracle Database (either a single instance or Oracle RAC) and in an existing WebLogic domain (either a single managed server or a WebLogic cluster).

Figure 1–2 shows the relationships among the components in this configuration. Oracle SES can be configured as a single managed server or as a cluster in the Fusion Middleware Functional Setup domain.

Figure 1–2 Oracle SES in Oracle Fusion Applications

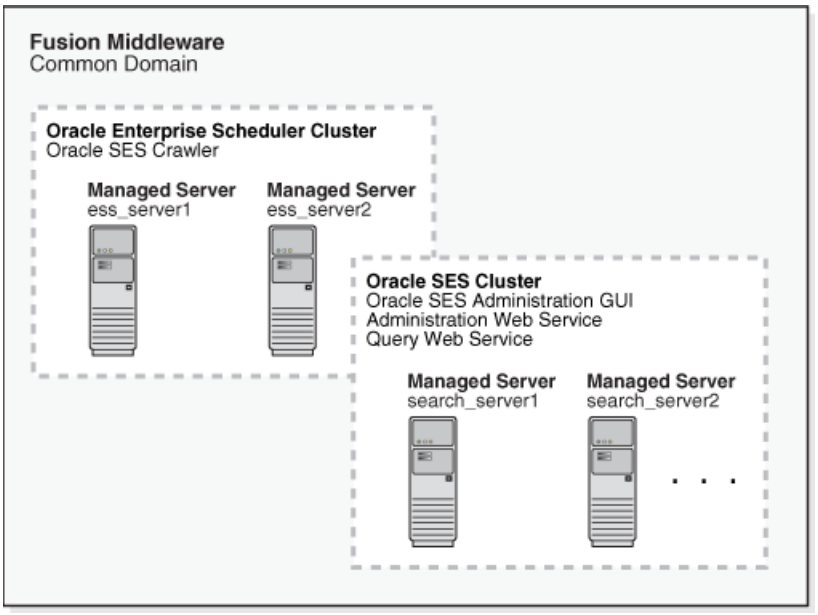


For the best performance under a production search load, Oracle recommends that you provide Oracle SES with a separate Oracle RAC cluster for the database back end.

Scheduling Crawls

In Oracle Fusion Applications, Oracle SES uses Oracle Enterprise Scheduler to schedule crawls of data sources. Figure 1–3 shows the distribution of the work load between the Oracle SES cluster and the Oracle ESS cluster: The crawler runs on the Oracle ESS cluster, and the Web applications run on the Oracle SES cluster.

Figure 1–3 Distribution of the Work Load Among Clusters



See Also: *Oracle Fusion Applications Administrator's Guide*

Administrative Tools

Oracle Fusion Applications provides additional tools that you can use to manage Oracle SES:

- Oracle Enterprise Manager Fusion Applications Control: You can control and monitor scheduled jobs in both the Oracle SES Administration GUI and in Fusion Applications Control. See "Monitoring the Crawling Process" on page 3-25.
- Oracle Enterprise Manager Database Control: You can use Database Control to monitor database performance metrics, alerts, warnings and errors associated with the SEARCHSYS administrative identity and the FUSION_APPS_SEARCH_APPID application identity.

Oracle Secure Enterprise Search Features

Information in an enterprise can be spread across Web pages, databases, mail servers or other collaboration software, document repositories, file servers, and desktops. Oracle SES searches all your data through the same interface. Oracle SES is fully globalized and works with many languages including Chinese, Japanese, Korean, Arabic, and Hebrew.

This section introduces a few of the features in Oracle SES. It includes the following topics:

- Secure Search
- Federated Search

See Also: Chapter 3, "Understanding Crawling" for more features relating to the crawler

Secure Search

Much of the information within an organization is publicly accessible. Anyone is allowed to view it. Therefore, it is relatively easy for a **crawler** to find and index that information.

However, there are other sources that are protected. These protected sources might be viewable only by certain users or groups of users. For example, while users can search in their own e-mail folders, they should not be able to search anyone else's e-mail.

For protected sources, the Oracle SES crawler indexes any document with the proper access control list. When end users perform a search, only documents that they have privileges to view are returned.

See Also: "Enabling Secure Search" on page 9-12

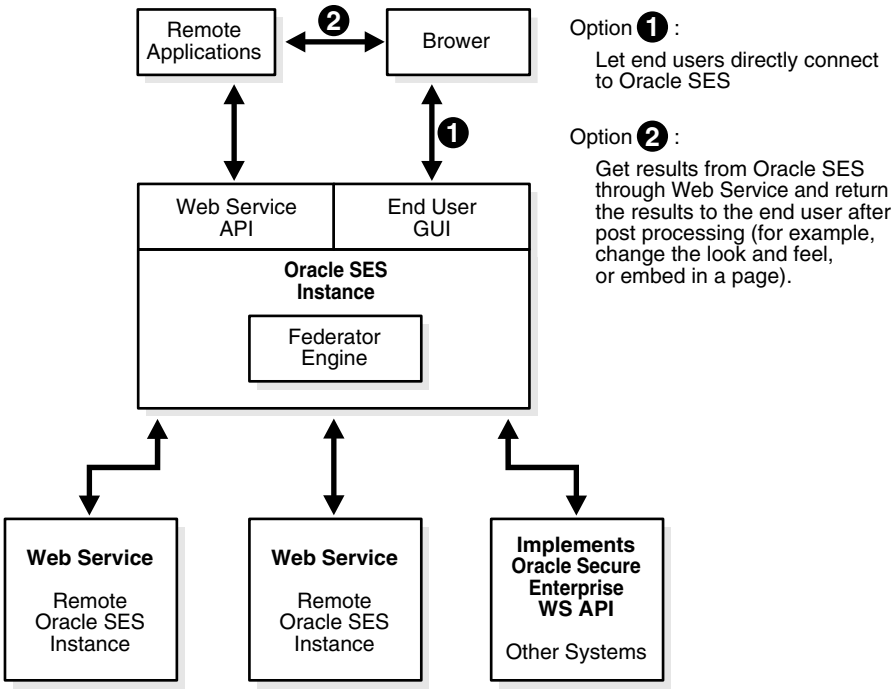
Federated Search

Oracle SES can search multiple Oracle SES instances with their own document repositories and indexes. It provides a unified framework to search the different repositories that are crawled, indexed, and maintained separately.

Federated search allows a single query to run across all Oracle SES instances. It aggregates the search results to show one unified result list to the user. User credentials are passed along with the query so that each federation endpoint can authenticate the user against its own document repository.

Figure 1–4 illustrates the federation architecture and two options for an end user to connect through a browser to Oracle SES. Option 1 allows users to connect their browsers directly to Oracle SES using the end-user graphical interface. Option 2 retrieves results from Oracle SES through Web Services after arbitrary post-processing, such as changing the look-and-feel or embedding the results in a page. For this option, the browser connects to remote applications, which connect to the Web Services API.

Figure 1–4 Federation Architecture



2

Getting Started with the Oracle SES Administration GUI

This chapter provides a brief introduction to using Oracle Secure Enterprise Search (Oracle SES). More information about the Oracle SES Administration GUI is provided later in this book and in the online Help.

This chapter contains the following topics:

- Getting Started Basics for the Administration GUI
- Understanding the Oracle SES Administration GUI
- Starting and Stopping Oracle SES

Getting Started Basics for the Administration GUI

After you have successfully installed Oracle SES, you can define your data sources and start crawling them. The following are general procedures.

To open the Oracle SES Administration GUI:

1. Open a browser and enter the URL for the Oracle SES Administration GUI. This URL is provided after the installation and has the form `http://host:port/search/admin/index.jsp`
2. Log on with the user name `SEARCHSYS` and the password specified during installation.

To administer Oracle SES:

1. Define one or more sources for the data you want to search.

   For example, if your data is stored in Web pages, then select the Web source.
2. Check the crawler progress and status on the Home - Schedules page. Click **Refresh Status**. From the status page, you can view the statistics of the crawl.
3. Monitor the search statistics on the Home - General page and the Home - Statistics page.

The following procedures expand these steps.

> **Note:** For specific information about a source, see Part II, "Creating Data Sources".

To create a source:

1. On the Home page, select the **Sources** secondary tab to display the Sources page.
2. Select a source type.
3. Click **Create** to display the Create Source page.
4. Complete the page. Click **Help** for a description of each item.
5. Click **Create** or **Create & Customize**.

   A crawl schedule is automatically created along with the source. If you select **Start Crawling Immediately**, then the crawler starts crawling when you click **Create**.

To customize a source:

1. When creating a source, click **Create & Customize** on the Create Source page to display the Customize Source page.

   or

   After creating a source, click the **Edit** icon for the source on the Home - Sources page.
2. Click the subtabs and make the desired changes.
3. Click **Apply**.

To crawl and index an existing source:

1. On the Home page, select the **Schedules** secondary tab to display the Crawler Schedules page.
2. Select the schedule.
3. To set up the frequency and other crawling options for one or more sources, click **Edit**. (Optional)
4. Click **Start** to start crawling.

Understanding the Oracle SES Administration GUI

The Oracle SES Administration GUI provides many options for managing and customizing Oracle SES to suit your enterprise. This section describes some tasks you can accomplish using the Oracle SES Administration GUI.

Home Tab

The **Home** tab consists of the **General**, **Sources**, **Schedules**, and **Statistics** secondary tabs.

- **Home - General**

  This is the home page for Oracle SES. The Summary section shows an overview of the system search performance, both quality and speed, over the past seven days. The Failed Schedules section lists all schedules that have failed. A failed schedule is one in which the crawler encountered an irrecoverable error, such as an indexing error or a source-specific login error, and cannot proceed. A failed schedule could be because of a partial collection and indexing of documents.

- **Home - Sources**

  A collection of information is called a source. Each source has a type, such as a Web site or a database table. User-defined source types are listed on the Global Settings - Source Types page. You can create as many sources as you want.

- **Home - Schedules**

  This page lets you view, edit, create, delete, stop, or start a schedule. Schedules define the frequency at which the index is updated with information about each source.

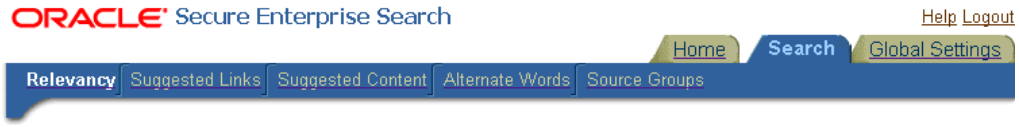
- **Home - Statistics**

  This page provides numerous search and crawler statistics, such as the most popular queries and crawler progress.

  Some statistics constantly show up-to-date information, while others are cached hourly to improve performance. The Last Refreshed time shows the actual time of the statistics displayed. Check the online help for each statistics page to confirm if the statistics are up-to-date or cached hourly.

Search Tab

The **Search** tab consists of the **Relevancy**, **Suggested Links**, **Suggested Content**, **Alternate Words**, and **Source Groups** secondary tabs. These pages help you improve search quality.



- **Search - Relevancy**

  Make important documents easier to find with relevancy boosting. Oracle SES lets you influence the order of documents in the result list for a particular search. For example, your company Web site could have a home page for documentation that should appear high in the results of any search for "documentation".

- **Search - Suggested Links**

  Direct users to a particular Web site for a search string. For example, when users search for "Oracle SES documentation" or "Enterprise Search documentation" or "Search documentation", you could suggest `http://www.oracle.com/technetwork`. In the default search page, suggested links are displayed at the top of the search result list. This is especially useful to provide links to important Web pages that are not crawled by Oracle SES.

- **Search - Suggested Content**

  Suggest actual content (as opposed to links) to be displayed along with the result list. For example, when an end user searches for contact information on a coworker, Oracle SES fetches the content from the suggested content provider and returns the contact information (e-mail address, phone number, and so on) for that person with the result list. Suggested content results appear in tabbed panes above the query results.

- **Search - Alternate Words**

  Use alternate words to suggest alternative search queries to users. This is useful for fixing common errors that users make in searching (for example, entering

Oracel instead of Oracle). Also, synonyms can provide more relevant results; for example, cellular phones for cell phones or wireless phones. Additional uses for alternate keywords are for product code names and abbreviations.

- **Search - Source Groups**

  Source groups are groups of sources that can be searched together. A source group consists of one or more sources, and a source can be assigned to multiple source groups. On the basic Search page, users can browse source groups that the administrator created. If no source group is selected, then all documents are searched.

Global Settings Tab

The **Global Settings** tab includes links to configure settings for your Oracle SES environment.

This section describes some global configuration pages.

- **Crawler Configuration**

  This page configures global crawler settings, such as crawling depth, language, and maximum document size.

  After a source has been created, you can define crawling parameters, such as URL boundary rules and crawling depth, for that source by editing that source on the Home - Sources page.

  See Also: "Overview of Crawler Settings" on page 3-3

- **Query Configuration**

  This page includes several options, including the following: maximum number of results returned, display URL, spell checking, statistics collection, URL submission, federated search, and secure search.

- **Identity Management Setup**

  This page lets you set up connections between Oracle Secure Enterprise Search and any identity management system to validate and authenticate users. This is necessary for secure searches. Oracle SES uses an **identity plug-in** as an interface to an identity management system.

- **Configure Search Results List**

  This page lets you customize the look and feel of the search result list using XSLT and CSS style sheets.

- **Configure Clustering in Search Results**

  Real-time clustering organizes search results into various groups to provide end users with different views on the top results. Clustered documents within one group (called a cluster node) share the same common topics or property values. A cluster node with a large document set can be categorized into child cluster nodes, and a hierarchy cluster result tree is built. Users can navigate directly to a specific cluster node or refine their query by combining the original query and cluster results.

  See Also:

  - Oracle SES Administration Tutorial for help with common administrator tasks:

    `http://st-curriculum.oracle.com/tutorial/SESAdminTutorial/index.htm`

  - Oracle SES Administration GUI context sensitive online Help
  - Oracle SES home page for updated information on known issues, code samples, and best practices:

    `http://www.oracle.com/technetwork/search/oses/overview/index.html`

Starting and Stopping Oracle SES

You can start, suspend, and stop Oracle SES from the WebLogic administration console from `http://wls_host:wls_port/console`.

Startup and shutdown of the database and listener services are managed independently of the middle tier.

To start Oracle SES:

1. Verify that Oracle Database is up and running. If it is down, start it using the tool of your choice, such as Oracle Enterprise Manager or SQL*Plus.
2. Start WebLogic Server, and enter your WebLogic administrative user name and password when prompted.
   - For Linux:

     ```
     startWebLogic.sh
     ```

   - For Windows:

     ```
     startWebLogic.bat
     ```

3. In a separate window, start Oracle SES.
   - For Linux:

     ```
     startManagedWebLogic.sh search_server1
     ```

   - For Windows:

     ```
     StartManagedWebLogic.bat search_server1
     ```

4. Log in again as the `weblogic` administrator.

To stop Oracle SES:

1. Oracle Database must be up and running to follow these procedures.
2. Start WebLogic Server.
   - For Linux:

     ```
     startWebLogic.sh
     ```

   - For Windows:

     ```
     startWebLogic.bat
     ```

3. Log in as the weblogic user with the password set under "Configuring WebLogic Server for Oracle SES 11.2.1".
4. In a separate window, stop Oracle SES.
   - For Linux:

     ```
     stopManagedWebLogic.sh search_server1
     ```

   - For Windows:

     ```
     StopManagedWebLogic.bat search_server1
     ```

   If needed, log in again as the WebLogic administrator.

3

Understanding Crawling

This chapter discusses the Oracle SES crawler. It contains the following topics:

- Overview of the Oracle Secure Enterprise Search Crawler
- Overview of Crawler Settings
- Overview of XML Connector Framework
- Configuring Support for Image Metadata
- Overview of Attributes
- Understanding the Crawling Process
- Monitoring the Crawling Process
- Parallel Query Indexing

See Also:

- "Tuning Crawl Performance" on page 10-4 and "Tuning Search Performance and Scalability" on page 10-12
- The Oracle Secure Enterprise Search tutorials at

  `http://www.oracle.com/technetwork/search/oses/overview/index.html`

Overview of the Oracle Secure Enterprise Search Crawler

The Oracle Secure Enterprise Search (Oracle SES) crawler is a J2SE process activated by a set schedule that runs on the middle tier. When activated, the crawler spawns processor threads that fetch documents from **sources**. The crawler caches the documents, and when the cache reaches the maximum batch size of 250 MB, the crawler indexes the cached files. This index is used for searching.

The document cache, called Secure Cache, is stored in Oracle Database in a compressed SecureFile LOB. Oracle Database provides excellent security and compact storage.

In the Oracle SES Administration GUI, you can create schedules with one or more sources attached to them. Schedules define the frequency at which the Oracle SES index is kept up-to-date with existing information in the associated sources.

See "Understanding the Crawling Process" on page 3-22 for more detailed information about the crawling process.

Crawler URL Queue

In the process of crawling, the crawler maintains a list of URLs of the discovered documents that are fetched and indexed in an internal URL queue. The queue is persistently stored, so that crawls can be resumed after the Oracle SES instance is restarted.

Understanding Access URLs and Display URLs

A *display* URL is a URL string used for search result display. This is the URL used when users click the search result link. An *access* URL is an optional URL string used by the crawler for crawling and indexing. If it does not exist, then the crawler uses the display URL for crawling and indexing. If it does exist, then it is used by the crawler instead of the display URL for crawling. For regular Web crawling, only display URLs are available. But in some situations, the crawler needs an access URL for crawling the internal site while keeping a display URL for the external use. For every internal URL, there is an external mirrored URL.

For example, for file sources with display URLs, end users can access the original document with the HTTP or HTTPS protocols. They provide the appropriate authentication and personalization and result in a better user experience.

Display URLs can be provided using the URL Rewriter API. Or, they can be generated by specifying the mapping between the prefix of the original file URL and the prefix of the display URL. Oracle SES replaces the prefix of the file URL with the prefix of the display URL.

For example, if the file URL is

```
file://localhost/home/operation/doc/file.doc
```

and the display URL is

```
https://webhost/client/doc/file.doc
```

then specify the file URL prefix as

```
file://localhost/home/operation
```

and the display URL prefix as

```
https://webhost/client
```

Modifying the Crawler Parameters

You can alter the crawler's operating parameters at two levels:

- At the global level for all sources
- At the source level for a particular defined source

Global parameters include the default values for language, crawling depth, and other crawling parameters, and the settings that control the crawler log and cache.

To configure the crawler:

1. Click the **Global Settings** tab.
2. Under Sources, click **Crawler Configuration**.
3. Make the desired changes on the Crawler Configuration page. Click **Help** for more information about the configuration settings.
4. Click **Apply**.

To configure the crawling parameters for a specific source:

1. From the Home page, click the **Sources** secondary tab to see a list of sources you have created.
2. Click the edit icon for the source whose crawler you want to configure, to display the Edit Source page.
3. Click the **Crawling Parameters** subtab.
4. Make the desired changes. Click **Help** for more information about the crawling parameters.
5. Click **Apply**.

The parameter values for a particular source can override the default values set at the global level. For example, for Web sources, Oracle SES sets a default crawling depth of 2, irrespective of the crawling depth you set at the global level.

Also note that some parameters are specific to a particular source type. For example, Web sources include parameters for HTTP cookies.

Overview of Crawler Settings

This section describes crawler settings and other mechanisms to control the scope of Web crawling:

- Crawling Mode
- URL Boundary Rules
- Document Types
- Crawling Depth
- Robots Exclusion
- Index Dynamic Pages
- Title Fallback
- Character Set Detection

> **See Also:** "Tuning Crawl Performance" on page 10-4 for more detailed information on these settings and other issues affecting crawl performance

Crawling Mode

For initial planning purposes, you might want the crawler to collect URLs without indexing. After crawling is finished, examine the document URLs and status, remove unwanted documents, and start indexing. The crawling mode is set on the Home - Schedules - Edit Schedules page.

> **See Also:** Appendix B, "URL Crawler Status Codes"

These are the crawling mode options:

- **Automatically Accept All URLs for Indexing**: This crawls and indexes all URLs in the source. For Web sources, it also extracts and indexes any links found in those URLs. If the URL has been crawled before, then it is reindexed only if it has changed.

- **Examine URLs Before Indexing**: This crawls but does not index any URLs in the source. It also crawls any links found in those URLs.
- **Index Only**: This crawls and indexes all URLs in the source. It does not extract any links from those URLs. In general, select this option for a source that has been crawled previously under "Examine URLs Before Indexing".

URL Boundary Rules

URL boundary rules limit the crawling space. When boundary rules are added, the crawler is restricted to URLs that match the indicated rules. The order in which rules are specified has no impact, but exclusion rules always override inclusion rules.

Boundary rules are set on the Home - Sources - Boundary Rules page.

Inclusion Rules

Specify an inclusion rule that a URL contain, start with, or end with a term. Use an asterisk (*) to represents a wildcard. For example, `www.*.example.com`. Simple inclusion rules are case-insensitive. For case-sensitivity, use regular expression rules.

An inclusion rule ending with example.com limits the search to URLs ending with the string `example.com`. Anything ending with `example.com` is crawled, but `http://www.example.com.tw` is not crawled.

If the URL Submission functionality is enabled on the Global Settings - Query Configuration page, then URLs that are submitted by end users are added to the inclusion rules list. You can delete URLs that you do not want to index.

Oracle SES supports the regular expression syntax used in Java JDK 1.4.2 Pattern class (`java.util.regex.Pattern`). Regular expression rules use special characters. The following is a summary of some basic regular expression constructs.

- A caret (^) denotes the beginning of a URL and a dollar sign ($) denotes the end of a URL.
- A period (.) matches any one character.
- A question mark (?) matches zero or one occurrence of the character that it follows.
- An asterisk (*) matches zero or more occurrences of the pattern that it follows. You can use an asterisk in the starts with, ends with, and contains rules.
- A backslash (\) escapes any special characters, such as periods (\.), question marks (\?), or asterisks (*).

See Also:

`http://www.oracle.com/technetwork/java/index.html` for a complete description in the Java documentation

Exclusion Rules

You can specify an exclusion rule that a URL contains, starts with, or ends with a term. An exclusion of `uk.example.com` prevents the crawling of Example hosts in the United Kingdom.

Default Exclusion Rules

The crawler contains a default exclusion rule to exclude non-textual files. The following file extensions are included in the default exclusion rule.

- **Image**: bmp, png, tif

- **Audio**: wav, wma, mp3
- **Video**: avi, wmv, mpeg, mpg
- **Binary**: bin, cab, dll, dmp, ear, exe, iso, jar, scm, so, tar, war, wmv

To crawl a file with these extensions, update the `globalBoundaryRules` object object using the Administration API. See the *Oracle Secure Enterprise Search Administration API Guide*

> **Note:** Only the file name is indexed when crawling multimedia files, unless the file is crawled using a crawler plug-in that provides a richer set of attributes, such as the Image Document Service plug-in.

Examples of Inclusion and Exclusion Rules

The following example uses several regular expression constructs that are not described earlier, including range quantifiers, non-grouping parentheses, and mode switches. For a complete description, see the Java documentation.

To crawl only HTTPS URLs in the `example.com` and `examplecorp.com` domains, and to exclude files ending in .doc and .ppt:

- Inclusion: URL regular expression

  ```
  ^https://.*\.example(?:corp){0,1}\.com
  ```

- Exclusion: URL regular expression `(?i:\.doc|\.ppt)$`

Document Types

You can customize which document types are processed for each source. By default, PDF, Microsoft Excel, Microsoft PowerPoint, Microsoft Word, HTML and plain text are always processed.

To add or remove document types:

1. On the Home page, click the **Sources** secondary tab.
2. Choose a source from the list and select Edit to display the Customize Source page.
3. Select the **Document Types** subtab.

   The listed document types are supported for the source type.
4. Move the types to process to the Processed list and the others to the Not Processed list.
5. Click **Apply**.

Keep the following in mind about graphics file formats:

- For graphics format files (JPEG, JPEG 2000, GIF, TIFF, **DICOM**), only the file name is searchable. The crawler does not extract any metadata from graphics files or make any attempt to convert graphical text into indexable text, unless you enable a document service plug-in. See "Configuring Support for Image Metadata" on page 3-12.

Oracle SES allows up to 1000 files in zip files and LHA files. If there are more than 1000 files, then an error is raised and the file is ignored.

See Also: *Oracle Text Reference* Appendix B for supported document types

Crawling Depth

Crawling depth is the number of levels to crawl Web and file sources. A Web document can contain links to other Web documents, which can contain more links. Specify the maximum number of nested links for the crawler to follow. Crawling depth starts at 0; that is, if you specify 1, then the crawler gathers the starting (seed) URL plus any document that is linked directly from the starting URL. For file crawling, this is the number of directory levels from the starting URL.

Set the crawling depth on the Home - Sources - Crawling Parameters page.

Robots Exclusion

You can control which parts of your sites can be visited by robots. If robots exclusion is enabled (default), then the Web crawler traverses the pages based on the access policy specified in the Web server robots.txt file. The crawler also respects the page-level robot exclusion specified in HTML metatags.

For example, when a robot visits `http://www.example.com/`, it checks for `http://www.example.com/robots.txt`. If it finds it, then the crawler checks to see if it is allowed to retrieve the document. If you own the Web sites, then you can disable robots exclusions. However, when crawling other Web sites, always defer to robots.txt by enabling robots exclusion.

Set the robots parameter on the Home - Sources - Crawling Parameters page.

Index Dynamic Pages

By default, Oracle SES processes dynamic pages. Dynamic pages are generally served from a database application and have a URL that contains a question mark (?). Oracle SES identifies URLs with question marks as dynamic pages.

Some dynamic pages appear as multiple search results for the same page, and you might not want them all indexed. Other dynamic pages are each different and must be indexed. You must distinguish between these two kinds of dynamic pages. In general, dynamic pages that only change in menu expansion without affecting its contents should not be indexed.

Consider the following three URLs:

```
http://example.com/aboutit/network/npe/standards/naming_convention.html

http://example.com/aboutit/network/npe/standards/naming_convention.html?nsdnv=14z1

http://example.com/aboutit/network/npe/standards/naming_convention.html?nsdnv=14
```

The question marks (?) in two URLs indicate that the rest of the strings are input parameters. The three results are essentially the same page with different side menu expansion. Ideally, the search yields only one result:

```
http://example.com/aboutit/network/npe/standards/naming_convention.html
```

Note: The crawler cannot crawl and index dynamic Web pages written in Javascript.

Set the dynamic pages parameter on the Home - Sources - Crawling Parameters page.

Title Fallback

You can override a default document title with a meaningful title if the default title is irrelevant. For example, suppose that the result list shows numerous documents with the title "Daily Memo". The documents had been created with the same template file, but the document properties had not been changed. Overriding this title in Oracle SES can help users better understand their search results.

Title fallback can be used for any source type. Oracle SES uses different logic for each document type to determine which fallback title to use. For example, for HTML documents, Oracle SES looks for the first heading, such as `<h1>`. For Microsoft Word documents, Oracle SES looks for text with the largest font.

If the default title was collected in the initial crawl, then the fallback title is only used after the document is reindexed during a re-crawl. Thus, if there is no change to the document, then you must force the change by setting the re-crawl policy to **Process All Documents** on the Home - Schedules - Edit Schedule page.

To implement title fallback, modify the `crawlerSettings` object using the Administration API. Set the `<search:indexNullTitleFallback>` element to `indexForAll`, and list the bad titles in the `<search:badTitles>` elements. See the *Oracle Secure Enterprise Search Administration API Guide*.

Title fallback is not currently supported in the Oracle SES Administration GUI, and by default, it is turned off.

Special considerations with title fallback

- With Microsoft Office documents:
  - Font sizes 14 and 16 in Microsoft Word correspond to normalized font sizes 4 and 5 (respectively) in converted HTML. The Oracle SES crawler only picks up strings with normalized font size greater than 4 as the fallback title.
  - Titles must contain more than five characters.
- When a title is null, Oracle SES automatically indexes the fallback title for all binary documents (for example, .doc, .ppt, .pdf).

  For HTML and text documents, Oracle SES does *not* automatically index the fallback title. Thus, the replaced title on HTML or text documents cannot be searched with the title attribute on the Advanced Search page. To turn on indexing for HTML and text documents, modify the `crawlerSettings` object using the Administration API. Set the `<search:indexNullTitleFallback>` parameter to `indexForAll`.

Character Set Detection

This feature enables the crawler to automatically detect character set information for HTML, plain text, and XML files. Character set detection allows the crawler to properly cache files during crawls, index text, and display files for queries. This is important when crawling multibyte files (such as files in Japanese or Chinese).

To enable character set detection, update the `crawlerSettings` object using the Administration API. Set the `<search:charsetDetection>` parameter to `true`. See the *Oracle Secure Enterprise Search Administration API Guide* for more information about changing crawler settings.

Special Considerations with Automatic Character Set Detection

- To crawl XML files for a source, be sure to add XML to the list of processed document types on the Home - Source - Document Types page. XML files are currently treated as HTML format, and detection for XML files may not be as accurate as for other file formats.

Language Detection

With multibyte files, besides turning on character set detection, be sure to set the **Default Language** parameter. For example, if the files are all in Japanese, select Japanese as the default language for that source. If automatic language detection is disabled, or if the crawler cannot determine the document language, then the crawler assumes that the document is written in the default language. This default language is used only if the crawler cannot determine the document language during crawling.

If your files are in multiple languages, then turn on the **Enable Language Detection** parameter. Not all documents retrieved by the crawler specify the language. For documents with no language specification, the crawler attempts to automatically detect language. The language recognizer is trained statistically using trigram data from documents in various languages (for instance, Danish, Dutch, English, French, German, Italian, Portuguese, and Spanish). It starts with the hypothesis that the given document does not belong to any language and ultimately refutes this hypothesis for a particular language where possible. It operates on Latin-1 alphabet and any language with a deterministic Unicode range of characters (like Chinese, Japanese, Korean, and so on).

The crawler determines the language code by checking the HTTP header content-language or the `LANGUAGE` column, if it is a table source. If it cannot determine the language, then it takes the following steps:

- If the language recognizer is not available or if it cannot determine a language code, then the default language code is used.
- If the language recognizer is available, then the output from the recognizer is used.
- Oracle SES uses different lexers for space-delimited languages (such as English), Chinese, Japanese, and Korean. See the `lexer` object description in the *Oracle Secure Enterprise Search Administration API Guide*.

The **Default Language** and the **Enable Language Detection** parameters are on the Global Settings - Crawler Configuration page (globally) and also the Home - Sources - Crawling Parameters page (for each source).

> **Note:** For file sources, the individual source setting for **Enable Language Detection** remains false regardless of the global setting. In most cases, the language for a file source should be the same, and set from, the **Default Language** setting.

Deleting the Secure Cache

You can manage the Secure Cache either on the global level or at the data source level. The data source configuration supersedes the global configuration.

The cache is preserved by default and supports the Cached link feature in the search result page. If you do not use the **Cache** link, then you can delete the cache, either for specific sources or globally for all of them. Without a cache, the **Cached** link in a search result page returns a `File not found` error.

To delete the cache for all sources:

1. Select the **Global Settings** tab in the Oracle SES Administration GUI.
2. Choose **Crawler Configuration**.
3. Set **Preserve Document Cache** to **No**.
4. Click **Delete Cache Now** to remove the cache from all sources, except any that are currently active under an executing schedule. The cache is deleted in the background, and you do not have to wait for it to complete.
5. Click **Apply**.

To delete the cache for an individual source:

1. Select the **Sources** secondary tab on the Home page.
2. Click **Edit** for the source.
3. Click the **Crawling Parameters** subtab.
4. Set **Preserve Document Cache** to **No**.
5. Click **Apply**.

Overview of XML Connector Framework

Oracle SES provides an XML connector framework to crawl any repository that provides an XML interface to its contents. The connectors for Oracle Content Server, Oracle E-Business Suite 12, and Siebel 8 use this framework.

Every document in a repository is known as an **item**. An item contains information about the document, such as author, access URL, last modified date, security information, status, and contents.

A set of items is known as a **feed** or **channel**. To crawl a repository, an XML document must be generated for each feed. Each feed is associated with information such as feed name, type of the feed, and number of items.

To crawl a repository with the XML connector, place data feeds in a location accessible to Oracle SES over one of these protocols: HTTP, FTP, or FILE. Then generate an XML Configuration File that contains information such as feed location and feed type. Create a source with a source type that is based on this XML connector and trigger the crawl from Oracle SES to crawl the feeds.

There are two types of feeds:

- **Control feed**: Individual feeds can be located anywhere, and a single control file is generated with links to the feeds. This control file is input to the connector through the configuration file. A link in control feed can point to another control feed. Control feed is useful when data feeds are distributed over many locations or when the data feeds are accessed over diverse protocols such as FTP and file.
- **Directory feed**: All feeds are placed in a directory, and this directory is input to the connector through the configuration file. Directory feed is useful when the data feeds are available in a single directory.

Guidelines for the target repository generating the XML feeds:

- XML feeds are generated by the target repository, and each file system has a limit on how many files it can hold. For directory feeds, the number of documents in each directory should be less than 10,000. There are two considerations:
  - **Feed files**: The number of items in each feed file should be set such that the total number of feed files in the feed directory is kept under 10,000.

  - **Content files**: If the feed files specify content through attachment links and the targets of these links are stored in the file system, then ensure that the targets are distributed in multiple directories so that the total number of files in each directory is kept under 10,000.

- When feeds are generated real-time over HTTP, ensure that the component generating the feeds is sensitive to time-out issues of feed requests. The feed served as the response for every request should be made available within this time out interval; otherwise, the request from Oracle SES times out. The request is retried as many times as specified while setting up the source in Oracle SES. If all these attempts fail, then the crawler ignores this feed and proceeds with the next feed.

  See Also:

  - Appendix A, "XML Connector Examples and Schemas"
  - "Setting Up Oracle Content Server Sources" on page 6-25
  - "Setting Up Oracle E-Business Suite Sources" on page 8-4
  - "Setting Up Siebel 8 Sources" on page 8-32

Example Using the XML Connector

The courses in the Oracle E-Business Suite Learning Management application can be crawled and indexed to readily search the courses offered, location and other details pertaining to the courses.

To crawl and index courses in Oracle E-Business Suite Learning Management:

1. Generate an XML feed containing the courses. Each course can be an item in the feed. The properties of the course such as location and instructor can be set as attributes of the item.
2. Move the feed to a location accessible to Oracle SES through HTTP, FTP, or file protocol.
3. Generate a control file that points to that feed.
4. Generate a configuration file to point to this feed. Specify the feed type as control, the URL of the control feed, and the source name in the configuration file.
5. Create an **Oracle E-Business Suite 12** source in Oracle SES, specifying in the parameters the location of the configuration file, the user ID and the password to access the feed.

XML Configuration File

The configuration file is an XML file conforming to a set schema.

The following is an example of a configuration file to set up an XML-based source:

```
<rsscrawler xmlns="http://xmlns.oracle.com/search/rsscrawlerconfig">
     <feedLocation>ftp://my.host.com/rss_feeds</feedLocation>
     <feedType>directoryFeed</feedType>
     <errorFileLocation>/tmp/errors</errorFileLocation>
     <securityType>attributeBased</securityType>
     <sourceName>Contacts</sourceName>
     <securityAttribute name="EMPLOYEE_ID" grant="true"/>
</rsscrawler>
```

Where

- `feedLocation` is one of the following:
  - URL of the directory, if the data feed is a directory feed

    This URL should be the FTP URL or the file URL of the directory where the data feeds are located. For example:

    ```
    ftp://example.domain.com/relativePathOfDirectory
    file://example.domain.com/c:\dir1\dir2\dir3
    file://example.domain.com//private/home/dir1/dir2/dir3
    ```

    File URL if the data feeds are available on the same computer as Oracle SES. The path specified in the URL should be the absolute path of the directory.

    FTP URL to access data feeds on any other computer. The path of the directory in the URL can be absolute or relative. The absolute path should be specified following the slash (/) after the host name in the URL. The relative path should be specified relative to the home directory of the user used to access FTP feeds.

    The user ID used to crawl the source should have write permissions on the directory, so that the data feeds can be deleted after crawl.
  - URL of the control file, if the data feed is a control feed

    This URL can be HTTP, HTTPS, file, or FTP URL. For example:

    ```
    http://example.com:7777/context/control.xml
    ```

    The path in FTP and file protocols can be absolute or relative.
- `feedType` indicates the type of feed. Valid values are `directoryFeed`, `controlFeed`, and `dataFeed`.
- `errorFileLocation` (optional) specifies the directory where status feeds should be uploaded.

  A status feed is generated to indicate the status of the processing feed. This status feed is named *data_feed_file_name*.suc or *data_feed_file_name*.err depending on whether the processing was successful. Any errors encountered are listed in the error status feed. If a value is specified for this parameter, then the status feed is uploaded to this location. Otherwise, the status feed is uploaded to the same location as the data feed.

  The user ID used to access the data feed should have write permission on the directory.

  If `feedLocation` is an HTTP URL, then `errorFileLocation` also should be an HTTP URL, to which the status feeds are posted. If no value is specified for `errorFileLocation`, then the status feeds are posted to the URL given in `feedLocation`.

  If an error occurs while processing a feed available over file or FTP protocol, then the erroneous feed is renamed *filename*.prcsdErr in the same directory.
- `sourceName` (optional) specifies the name of the source.
- `securityType` (optional) specifies the security type. Valid values are the following:
  - `noSecurity`: There is no security information associated with this source at the document level. This is the default value.
  - `identityBased`: Identity-based security is used for documents in the feed.

  - `attributeBased`: Attribute-based security is used for documents in the feed. With this security model, security attributes should be specified in the `securityAttribute` tag, and the values for these attributes should be specified for each document.

- `securityAttribute` specifies attribute-based security. One or more tags of this type should be specified, and each tag should contain the following attributes:
  - `name`: Name of the security attribute.
  - `grant`: Boolean parameter indicating whether this is a grant/deny attribute. The security attribute is considered a grant attribute if the value is true and a deny attribute if the value is false.

See Also: "Configuration File XML Schema Definition" on page A-1

Configuring Support for Image Metadata

The Oracle SES crawler initially is set to search only text files. You can change this behavior by configuring an image document service connector to search the metadata associated with image files. Image files can contain rich metadata that provide additional information about the image itself.

The Image Document Service connector integrates Oracle Multimedia (formerly Oracle *inter*Media) images with Oracle SES. This connector is separate from any specific data source.

The following table identifies the metadata formats (EXIF, IPTC, XMP, DICOM) that can be extracted from each supported image format (JPEG, TIFF, GIF, JPEG 2000, DICOM).

	JPEG	TIFF	GIF	JPEG2000	DICOM
EXIF	Yes	Yes	No	No	No
IPTC	Yes	Yes	No	No	No
XMP	Yes	Yes	Yes	Yes	No
DICOM	No	No	No	No	Yes

See Also: *Oracle Multimedia User's Guide* and *Oracle Multimedia Reference* for more information about image metadata

Identifying the Search Attributes for Image Metadata

Image files can contain metadata in multiple formats, but not all of it is useful when performing searches. A configuration file in Oracle SES enables you to control the metadata that is searched and published to an Oracle SES Web application.

If you upgraded from a previous release, then the default configuration file remains `ordesima-sample.xml`.

The default configuration file is named `attr-config.xml`. You can modify this file, which is located at *ORACLE_HOME*`/search/lib/plugins/doc/ordim/config/`. Oracle recommends that you create a copy of the default configuration file before editing it. Note that the configuration file must conform to the XML schema *ORACLE_HOME*`/search/lib/plugins/doc/ordim/xsd/ordesima.xsd`.

Oracle SES indexes and searches only those image metadata tags that are defined within the `metadata` element (between `<metadata>...</metadata>`) in the

configuration file. By default, the configuration file contains a set of the most commonly searched metadata tags for each of the file formats. You can add other metatags to the file based on your specific requirements.

Image files can contain metadata in multiple formats. For example, an image can contain metadata in the EXIF, XMP, and IPTC formats. An exception to this are DICOM images, which contain only DICOM metadata. Note that for IPTC and EXIF formats, Oracle Multimedia defines its own image metadata schemas. The metadata defined in the configuration file must conform to the Oracle Multimedia defined schemas.

Because different metadata formats use different tags to refer to the same attribute, it is necessary to map metatags and the search attributes they define. Table 3–1 lists some commonly used metatags and how they are mapped in Oracle SES.

Table 3–1 Metatag Mapping

Oracle SES Attribute Name	Oracle SES Predefined Name	EXIF Metatag	IPTC Metatag	XMP Metatag
Author	Author	Artist	Author	photoshop:Creator
AuthorTitle	X	X	AuthorTitle	photoshop:AuthorsPosition
Description	Description	ImageDescription	Caption	dc:Description
Title	Title	X	ObjectName	dc:Title
Description Writer	X	X	captionWriter	photoshop:CaptionWriter
Headline1	Headline1	X	Headline	photoshop:Headline
Category	X	X	Category	photoshop:Category
Scene	X	X	X	Iptc4xmpCore:Scene
Publisher	X	X	X	dc:Publisher
Source	X	X	Source	photoshop:Source
Copyright	X	Copyright	Copyright	dc:rights
Keywords	Keywords	X	Keyword	dc:subject
Provider	X	X	Credit	photoshop:Credit
City	X	X	City	photoshop:City
State	X	X	provinceState	photoshop:State
Country	X	X	Country	photoshop:Country
Location	X	X	location	Iptc4xmpCore:Location
EquipmentMake	X	Make	X	tiff:Make
EquipmentModel	X	Model	X	tiff:Model

Oracle SES provides this mapping in the configuration file `attr-config.xml`. You can edit the file to add other metatags. Oracle recommends that you make a copy of the original configuration file before editing the settings. The configuration file defines the display name of a metatag and how it is mapped to the corresponding metadata in each of the supported formats.

This is done using the <searchAttribute> tag, as shown in the example below:

```
<searchAttribute>
 <displayName>Author</displayName>
 <metadata>
   <value format="iptc">byline/author</value>
   <value format="exif">TiffIfd/Artist</value>
   <value format="xmp">dc:creator</value>
   <value format="xmp">tiff:Artist</value>
 </metadata>
</searchAttribute>
```

For each search attribute, the value of <displayName> is an Oracle SES attribute name that is displayed in the Oracle SES web application when an Advanced Search - Attribute Selection is performed. If any of the listed attributes are detected during a crawl, then Oracle SES automatically publishes the attributes to the SES web application.

For the <value> element, the format attribute must take the value of a supported format, such as iptc, exif, xmp, or dicom.

The value defined within the element, for example, byline/author, is the XML path when the image format is IPTC, EXIF, or XMP. For DICOM, this value must be the standard tag number or value locator.

For IPTC and EXIF formats, the XML path must conform to the metadata schemas defined by Oracle Multimedia. These schemas are defined in the files ordexif.xsd and ordiptc.xsd located at *ORACLE_HOME*/search/lib/plugins/doc/ordim/xsd/.

You do not need to specify the root elements defined in these schemas (iptcMetadata, exifMetadata) in the configuration file. For example, you can specify byline/author as the xmlPath value of the author attribute in IPTC format. Oracle Multimedia does not define XML schemas for XMP metadata, so refer to the Adobe XMP specification for the xmlPath value.

Within the <searchAttribute> tag, you can also specify an optional <dataType> tag if the attribute carries a date or numeric value. For example,

```
<searchAttribute>
     <displayName>AnDateAttribute</displayName>
     <dataType>date</dataType>
     <metadata>
        ...
     </metadata>
</searchAttribute>
```

The default data type is string, so you do not have to explicitly specify a string.

Supporting XMP Metadata

Oracle SES supports both standard and custom XMP metadata searches. Because all XMP properties share the same parent elements <rdf:rdf><rdf:description>, you must specify only the real property schema and property name in the configuration file. For example, specify photoshop:category instead of rdf:rdf/rdf:description/photoshop:category. The same rule applies to XMP custom metadata also. However, for XMP structure data, you must specify the structure element in the format parent/child 1/child 2/…child N, where child N is a leaf node. For example, Iptc4xmpCore:CreatorContactInfo/Iptc4xmpCore:CiPerson. Note that the image plug-in does not validate the metadata value for XMP metadata.

XMP metatags consist of 2 components separated by a colon(:). For example, `photoshop:Creator`, which corresponds to the `Author` attribute (see Table 3–1). In this example, `photoshop` refers to the XMP schema namespace. The other common namespaces include `dc`, `tiff`, and `Iptc4xmpCore`.

Before defining any XMP metadata in the configuration file, you must ensure that the namespace is defined. For example, before defining the metadata `photoshop:Creator`, you must include the namespace `photoshop` in the configuration file. This rule applies to both the standard and custom XMP metadata namespaces. As a best practice, Oracle recommends that you define all the namespaces at the beginning of the configuration file. If the namespace defined in the configuration file is different from the one in the image, then Oracle SES cannot find the attributes associated with this namespace. You can define namespaces as shown:

```
<xmpNamespaces>
<namespace
prefix="Iptc4xmpCore">http://iptc.org/std/Iptc4xmpCore/1.0/xmlns/</namespace>
<namespace prefix="dc">http://purl.org/dc/elements/1.1/</namespace>
<namespace prefix="photoshop">http://ns.adobe.com/photoshop/1.0/</namespace>
<namespace prefix="xmpRights">http://ns.adobe.com/xap/1.0/rights/</namespace>
<namespace prefix="tiff">http://ns.adobe.com/tiff/1.0/</namespace>
</xmpNamespaces>
```

The Adobe XMP Specification requires that XMP namespaces end with a slash (/) or hash (#) character.

> **See Also:** *Adobe Extensible Metadata Platform (XMP) Specification* for the XMP metadata schema and a list of standard XMP namespace values.
>
> http://partners.adobe.com/public/developer/en/xmp/sdk/XMPspecification.pdf

Custom XMP metadata must be explicitly added to `attr-config.xml`. An example of a custom metadata is:

```
<xmpNamespaces>
  <namespace prefix="hm">http://www.oracle.com/ordim/hm/</namespace>
</xmpNamespaces>
<searchattribute>
  <displayname>CardTitle</displayname>
  <metadata>
    <value format="xmp">hm:cardtitle</value>
  </metadata>
</searchattribute>
```

Supporting DICOM Metatags

Oracle SES 11*g* supports DICOM metatags, and these metatags are available in the default configuration file `attr-config.xml`.

DICOM metatags are either DICOM standard tags or DICOM value locators.

DICOM Standard Tags DICOM standard tags are 8-digit hexadecimal numbers, represented in the format `ggggeeee` where `gggg` specifies the group number and `eeee` specifies the element number. For example, the DICOM standard tag for the attribute `performing physician's name` is represented using the hexadecimal value 00081050.

The group number gggg must take an even value, excepting 0000, 0002, 0004, and 0006, which are reserved group numbers.

The DICOM standard defines over 2000 standard tags.

The file attr-config.xml contains a list of predefined DICOM standard metatags. You can add new metatags to the file as shown in the following example:

```
<searchAttribute>
      <displayName>PerformingPhysicianName</displayName>
      <metadata>
        <value format="dicom">00081050</value>
      </metadata>
</searchAttribute>
```

Note: The image connector does not support SQ, UN, OW, OB, and OF data type tags. Therefore, do not define such tags in the configuration file.

See Also: http://medical.nema.org for more information about the standard tags defined in DICOM images, and the rules for defining metatags

DICOM Value Locators Value locators identify an attribute in the DICOM content, either at the root level or from the root level down.

A value locator contains one or more sublocators and a tag field (optional). A typical value locator is of the format:

```
sublocator#tag_field
```

Or of the format:

```
sublocator
```

Each sublocator represents a level in the tree hierarchy. DICOM value locators can include multiple sublocators, depending on the level of the attribute in the DICOM hierarchy. Multiple sublocators are separated by the dot character (.). For example, value locators can be of the format:

```
sublocator1.sublocator2.sublocator3#tag_field
```

Or of the format:

```
sublocator1.sublocator2.sublocator3
```

A tag_field is an optional string that identifies a derived value within an attribute. A tag that contains this string must be the last tag of a DICOM value locator. The default is NONE.

A sublocator consists of a tag element and can contain other optional elements. These optional elements include definer and item_num. Thus, a sublocator can be of the format:

```
tag
```

Or it can be of the format

```
tag(definer)[item_num)
```

Table 3–2 Sub Components of a Sublocator

Component	Description
tag	A DICOM standard tag represented as an 8-digit hexadecimal number.
definer	A string that identifies the organization creating the tag. For tags that are defined by the DICOM standard, the default value (which can be omitted) is `DICOM`. Oracle SES supports DICOM standard tags alone. It does not support private tags.
item_num	An integer that identifies a data element within an attribute, or a wildcard character ("*") that identifies all data elements within an attribute. It takes a default value of 1, the first data element of an attribute. This parameter is optional.

The following example shows how to add a value locator to the `attr-config.xml` file:

```
<searchAttribute>
  <displayName>PatientFamilyName</displayName>
  <metadata>
  <value format="dicom">00100010#UnibyteFamily</value>
  </metadata>
</searchAttribute>
```

where `UnibyteFamily` is a tag_field of person name.

The following example shows how to define a value locator from the root level.

```
<searchAttribute>
      <displayName>AdmittingDiagnosisCode</displayName>
      <metadata>
        <value format="dicom">00081084.00080100</value>
      </metadata>
</searchAttribute>
<searchAttribute>
      <displayName>AdmittingDiagnosis</displayName>
      <metadata>
        <value format="dicom">00081084.00080104</value>
      </metadata>
</searchAttribute>
```

In the above example, the tag 00081084 represents the root tag `Admitting Diagnoses Code Sequence`. This tag includes four child tags: `code value` (0008, 0100), `coding scheme designator` (0008, 0102), `coding scheme version` (0008, 0103) and `code meaning` (0008, 0104). In this example, the value locators are `code value`: 00081084.00080100 and `code meaning`: 00081084.00080104.

Note: The image connector does not support SQ, UN, OW, OB, and OF data type value locators. Therefore, ensure that the last sublocator of a value locator does not specify such data types.

See Also: *Oracle Multimedia DICOM Developer's Guide* for more information about DICOM value locators

Example: Adding an Attribute to the Default attr-config.xml File

To search for information about image caption writer:

1. Open Oracle SES Administration GUI and create the DescriptionWriter attribute:

   Specify **DescriptionWriter** as an Oracle SES attribute name (shown on the Advanced Search - Attribute Selection page).

2. Examine the following sources for information relevant to modifying the default `attr-config.xml` file:

   - Oracle Multimedia IPTC schema at *ORACLE_HOME*`/search/lib/plugins/doc/ordim/xsd/ordiptc.xsd`. The IPTC metadata for image caption writer is shown as `captionWriter`.
   - Adobe XMP Specification for XMP Metadata. The XMP path for this property is defined as `photoshop:CaptionWriter`.
   - Oracle Multimedia EXIF schema. There is no caption writer metadata in EXIF.

3. Add the following section to `attr-config.xml`:

```
<searchAttribute>
   <displayName>DescriptionWriter</displayName>
   <metadata>
       <xmlPath format="iptc">captionWriter</xmlPath>
       <xmlPath format="xmp">photoshop:CaptionWriter</xmlPath>
   </metadata>
</searchAttribute>
```

4. If the `photoshop` XMP namespace is not registered in the configuration file, then add the `namespace` element to `xmpNamespaces` as shown here:

```
<xmpNamespaces>
   <namespace prefix="photoshop">http://ns.adobe.com/photoshop/1.0/</namespace>
      .
      .   existing namespaces
      .
</xmpNamespaces>
```

Creating an Image Document Service Connector

A default Image Document Service connector instance is created during the installation of Oracle SES. You can configure the default connector or create a new one.

To create an Image Document Service instance:

1. In the Oracle SES Administration GUI, click **Global Settings**.
2. Under Sources, click **Document Services** to display the Global Settings - Document Services page.
3. To configure the default image service instance:
   a. Click **Expand All**
   b. Click **Edit** for the default image service instance.

   or

   To create a new image service instance:

   a. Click **Create** to display the Create Document Service page.

b. For **Select From Available Managers**, choose **Secure Enterprise Search Image Document Service** and click **Next**.

c. Provide a name for the instance.

4. Provide a value for the **attributes configuration file** parameter.

   The default value of **attributes configuration file** is `attr-config.xml`. The file is located at *ORACLE_HOME*`/search/lib/plugins/doc/ordim/config/`.

5. Click **Apply**.

6. Click **Document Services** in the locator links to return to the Document Services page.

7. Add the Image Document Service plug-in to either the default pipeline or a new pipeline.

To add the default Image Document Service plug-in to the default pipeline:

1. Under Document Service Pipelines, click **Edit** for the default pipeline.
2. Move the Image Document Service instance from **Available Services** to **Used in Pipeline**.
3. Click **Apply**.

To create a new pipeline for the default Image Document Service plug-in:

1. Under Document Service Pipelines, click **Create** to display the Create Document Service Pipeline page.
2. Enter a name and description for the pipeline.
3. Move the Image Document Service instance from **Available Services** to **Used in Pipeline**.
4. Click **Create**.

Using the Image Document Service Connector

You must either create a source to use the connector or enable the connector for an existing source.

To enable the connector for an existing source:

1. Click **Sources** on the Home page.
2. Click the **Edit** icon for the desired source.
3. Click **Crawling Parameters**.
4. Select the pipeline that uses the Image Document Service and enable the pipeline for this source.
5. Click **Document Types**. From the **Not Processed** column, select the image types to search and move them to the **Processed** column. The following sources are supported: JPEG, JPEG2000, GIF, TIFF, DICOM.

Searching Image Metadata

You can search image metadata from either the Oracle SES Basic Search page or the Advanced Search - Attribute Selection page.

For Basic Search, Oracle SES searches all the metadata defined in the configuration file for each supported image document (JPEG, TIFF, GIF, JPEG 2000, and DICOM). It returns the image document if any matching metadata is found.

Advanced Search enables you to search one or more specified attributes. It also supports basic operations for date and number attributes. Oracle SES returns only those image documents that contain the specified metadata.

Oracle SES does not display the Cache link for image search results.

Troubleshooting the Image Document Service Connector

If the Image Document Service Connector fails, then check the following:

- Is the pipeline with an Image Document Service connector instance enabled for the source?
- Are the image types added to the source?
- For a web source, are the correct MIME types included in the HTTP server configuration file?

  For example, if you use Oracle Application Server, then check the Apache `mime.types` file. If the following media types are missing, then add them:

MIME Type	Extensions
image/jp2	jp2
application/dicom	dcm

- If a connection is established, and all the image files are not crawled, then check whether the recrawl policy is set to `Process Documents That Have Changed`. If so, change it to `Process All Documents`.

  To do this step, go to Home - Schedules, and under Crawler Schedules, click **Edit** for the specific source. This opens the Edit Schedule page. Under Update Crawler Recrawl Policy, select **Process All Documents**.

  You can change the recrawl policy back to **Process Documents That Have Changed**, after the crawler has finished crawling all the documents in the new source.

Overview of Attributes

Each source has its own set of document attributes. Document attributes, like metadata, describe the properties of a document. The crawler retrieves values and maps them to a search attributes. This mapping lets users search documents based on their attributes. Document attributes in different sources can be mapped to the same search attribute. Therefore, users can search documents from multiple sources based on the same search attribute.

After you crawl a source, you can see the attributes for that source. Document attribute information is obtained differently depending on the source type.

Document attributes can be used in tasks such as document management, access control, or version control. Different sources can have different attribute names that are used for the same idea; for example, `version` and `revision`. It can also have the same attribute name for different ideas; for example, "language" as in natural language in one source but as programming language in another. Document attribute information is obtained differently depending on the source type.

Oracle SES has several default search attributes. They can be incorporated in search applications for a more detailed search and richer presentation.

Search attributes are defined in the following ways:

- System-defined search attributes, such as title, author, description, subject, and mimetype.
- Search attributes created by the Oracle SES administrator.
- Search attributes created by the crawler. During crawling, the crawler plug-in maps the document attribute to a search attribute with the same name and data type. If not found, then the crawler creates a new search attribute with the same name and type as the document attribute defined in the crawler plug-in.

> **Note:** Search attribute names must be unique; two attributes cannot have the same name. For example, if a search attribute exists with a String data type, and another search attribute is discovered by the crawler with the same name but a different data type, then the crawler ignores the second attribute.
>
> To prevent this conflict and allow Oracle SES to index both attributes, check the list of Oracle SES attribute names and types in *Oracle SES Attributes* before creating new attributes.

Attributes For Different Source Types

Table and database sources have no predefined attributes. The crawler collects attributes from columns defined during source creation. You must map the columns to the search attributes.

For Siebel 7.8 sources, specify the attributes in the query while creating the source. For Oracle E-Business Suite and Siebel 8 sources, specify the attributes in the XML data feed.

For many source types, such as OracleAS Portal, e-mail, NTFS, and Microsoft Exchange sources, the crawler picks up key attributes offered by the target systems. For other sources, such as Documentum eRoom or Lotus Notes, an **Attribute list** parameter is in the Home - Sources - Customize User-Defined Source page. Any attributes that you define are collected by the crawler and available for search.

Using Lists of Values for Search Attributes

The list of values (LOV) for a search attribute can help you specify a search. Global search attributes can be specified on the Global Settings - Search Attributes page.

For user-defined sources where LOV information is supplied through a crawler plug-in, the crawler registers the LOV definition. Use the Oracle SES Administration GUI or the crawler plug-in to specify attribute LOVs, attribute value, attribute value display name, and its translation.

When multiple sources define the LOV for a common attribute, such as title, the user sees all the possible values for the attribute. When the user restricts search within a particular source group, only LOVs provided by the corresponding sources in the source group are shown.

LOVs can be collected automatically. The following example shows Oracle SES collecting LOV values to crawl a fictitious URL.

1. Create a Web source with `http://www.example.com` as the starting URL. Do not start crawling yet.

2. From the Global Settings - Search Attributes page, select the Attribute for Oracle SES to collect LOVs and click **Manage Lov**. (For example, click **Manage Lov** for **Author**.)
3. Select **Source-Specific** for the created source, and click **Apply**.
4. Click **Update Policy**.
5. Choose **Document Inspection** and click **Update**, then click **Finish**.
6. From the Home - Schedules page, start crawling the Web source. After crawling, the **LOV** button in the Advanced Search page shows the collected LOVs.

System-Defined Search Attributes

There are also two system-defined search attributes, `Urldepth` and `Infosource Path`.

`Urldepth` measures the number of levels down from the root directory. It is derived from the URL string. In general, the depth is the number of slashes, not counting the slash immediately following the host name or a trailing slash. An adjustment of -2 is made to home pages. An adjustment of +1 is made to dynamic pages, such as the example in Table 3–3 with the question mark in the URL.

`Urldepth` is used internally for calculating relevance ranking, because a URL with a smaller URL depth is typically more important.

Table 3–3 lists the `Urldepth` of some example URLs.

Table 3–3 Depth of Example URLs

URL	Urldepth
http://example.com/portal/page/myo/Employee_Portal/MyCompany	4
http://example.com/portal/page/myo/Employee_Portal/MyCompany/	4
http://example.com/portal/page/myo/Employee_Portal/MyCompany.htm	4
http://example.com/finance/finhome/topstories/wall_street.html?.v=46	4
http://example.com/portal/page/myo/Employee_Portal/home.htm	2

`Infosource Path` is a path representing the source of the document. This internal attribute is used in situations where documents can be browsed by their source. The `Infosource Path` is derived from the URL string.

For example, for this URL:

```
http://example.com/portal/page/myo/Employee_Portal/home.htm
```

The `Infosource Path` is:

```
portal/page/myo/Employee_Portal
```

If the document is submitted through a connector, this value can be set explicitly by using the `DocumentMetadata.setSourceHierarchy` API.

Understanding the Crawling Process

The first time the crawler runs, it must fetch data (Web pages, table rows, files, and so on) based on the source. It then adds the document to the Oracle SES index.

The Initial Crawl

This section describes a Web source crawling process for a schedule. It is divided into these phases:

- Queuing and Caching Documents
- Indexing Documents

Queuing and Caching Documents

The crawling cycle involves the following steps:

1. Oracle spawns the crawler according to the schedule you specify with the Oracle SES Administration GUI. When crawling is initiated for the first time, the URL queue is populated with the seed URLs.
2. The crawler initiates multiple crawling threads.
3. The crawler thread removes the next URL in the queue.
4. The crawler thread fetches the document from the Web. The document is usually an HTML file containing text and hypertext links. When the document is not in HTML format, the crawler converts the document into HTML before caching.
5. The crawler thread scans the HTML file for hypertext links and inserts new links into the URL queue. Duplicate links in the document table are discarded.
6. The crawler caches the HTML file.
7. The crawler registers the URL in the URL table.
8. The crawler thread starts over by repeating Step 3.

Fetching a document, as described in Step 4, can be time-consuming because of network traffic or slow Web sites. For maximum throughput, multiple threads fetch pages at any given time.

Indexing Documents

When the cache is full (default maximum size is 250 MB), the indexing process begins. At this point, the document content and any searchable attributes are pushed into the index.

When the **Preserve Document Cache** parameter is set to false, the crawler automatically deletes the cache after indexing the documents.

Oracle SES Stoplist

Oracle SES maintains a stoplist. A stoplist is a list of words that are ignored during the indexing process. These words are known as stopwords. Stopwords are not indexed because they are deemed not useful, or even disruptive, to the performance and accuracy of indexing. The Oracle SES stoplist contains only English words, and cannot be modified.

When you run a phrase search with a stopword in the middle, the stopword is not used as a match word, but it is used as a placeholder. For example, the word "on" is a stopword. If you search for the phrase "oracle on demand", then Oracle SES matches a document titled "oracle on demand" but not a document titled "oracle demand". If you search for the phrase "oracle on on demand", then Oracle SES matches a document entitled "oracle technology on demand" but not a document titled "oracle demand" or "oracle on demand".

Maintenance Crawls

After the initial crawl, a URL page is only crawled and indexed if it changed since the last crawl. The crawler determines whether it has changed with the HTTP If-Modified-Since header field or with the checksum of the page. URLs that no longer exist are marked and removed from the index.

To update changed documents, the crawler uses an internal checksum to compare new Web pages with cached Web pages. Changed Web pages are cached and marked for re-indexing.

Data synchronization involves the following steps:

1. Oracle spawns the crawler according to the schedule specified in the Oracle SES Administration GUI. The URL queue is populated with the seed URLs of the source assigned to the schedule.
2. The crawler initiates multiple crawling threads.
3. Each crawler thread removes the next URL in the queue.
4. Each crawler thread fetches a document from the Web. The page is usually an HTML file containing text and hypertext links. When the document is not in HTML format, the crawler converts the document into HTML before caching.
5. Each crawler thread calculates a checksum for the newly retrieved page and compares it with the checksum of the cached page. If the checksum is the same, then the page is discarded and the crawler goes to Step 3. Otherwise, the crawler continues to the next step.
6. The crawler thread scans the document for hypertext links and inserts new links into the URL queue. Links that are in the document table are discarded. Oracle SES does not follow links from filtered binary documents.
7. The crawler marks the URL as accepted. The URL is crawled in future maintenance crawls.
8. The crawler registers the URL in the document table.
9. If the cache is full or if the URL queue is empty, then caching stops. Otherwise, the crawler thread starts over at Step 3.

A maintenance or a forced recrawl does not move a cache from the file system to the database, or the reverse. The cache location for a source remains the same until it is migrated to a different location.

Automatic Forced Recrawls

When you configure a data source, certain operations trigger an automatic forced recrawl of the data source. These operations include the following:

- Deleting a document attribute from the data source
- Remapping a document attribute to a different search attribute
- Changing the crawler configuration "Index Dynamic Page" from `No` to `Yes` for a Web source.

These operations set the Force Recall flag, but no notice is given of this change in mode.

Monitoring the Crawling Process

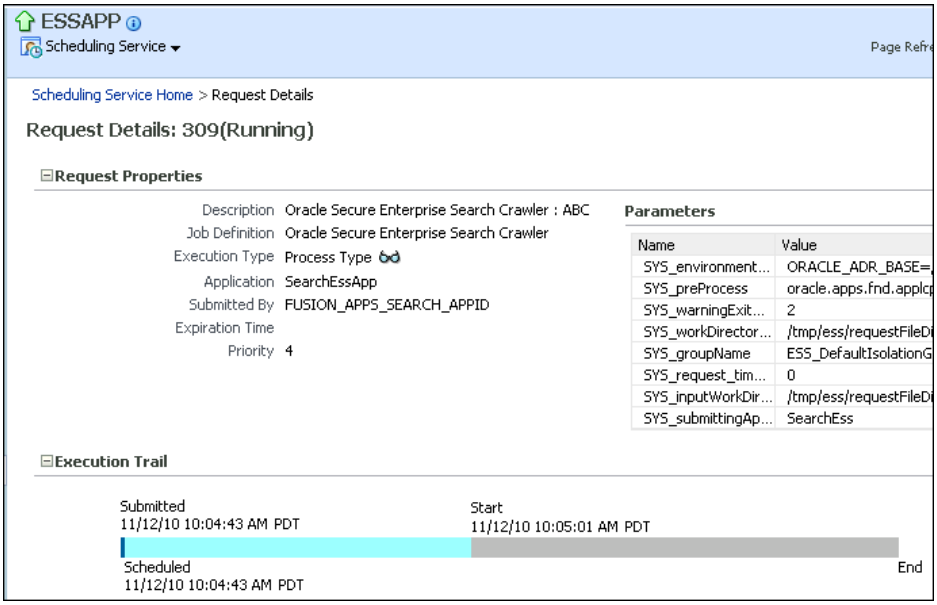
Monitor the crawling process in the Oracle SES Administration GUI by using a combination of the following:

- Check the crawl progress and crawl status on the Home - Schedules page. (Click **Refresh Status**.)
- Monitor your crawler statistics on the Home - Schedules - Crawler Progress Summary page and the Home - Statistics page.
- Monitor the log file for the current schedule.

See Also: "Tuning Crawl Performance" on page 10-4

In Oracle Fusion Applications, you can also monitor crawler jobs in Enterprise Manager Fusion Applications Control. Figure 3–1 shows a crawler schedule named ABC, which appears in the Scheduling Services with a prefix of Oracle Secure Enterprise Search Crawler. The FUSION_APPS_SEARCH_APPID application identity submits all crawler jobs. All Oracle SES connectors use this identity to crawl searchable repositories within Fusion Applications.

Figure 3–1 Oracle SES Crawls Reported in Fusion Applications Control



Crawler Statistics

The following crawler statistics are shown on the Home - Schedules - Crawler Progress Summary page. Some statistics are also shown in the log file under "Crawling results".

- **Documents to Fetch**: Number of URLs in the queue waiting to be crawled. The log file uses the phrase "Documents to Process".
- **Documents Fetched**: Number of documents retrieved by the crawler.
- **Document Fetch Failures**: Number of documents whose contents cannot be retrieved by the crawler. This could be due to an inability to connect to the Web site, slow server response time causing time-outs, or authorization requirements. Problems encountered after successfully fetching the document are not considered

here; for example, documents that are too big or duplicate documents that were ignored.

- **Documents Rejected**: Number of URL links encountered but not considered for crawling. The rejection could be due to boundary rules, the robots exclusion rule, the mime type inclusion rule, the crawling depth limit, or the URL rewriter discard directive.
- **Documents Discovered**: Total number of documents discovered so far. This is roughly equal to (documents to fetch) + (documents fetched) + (document fetch failures) + (documents rejected).
- **Documents Indexed**: Number of documents that have been indexed or are pending indexing.
- **Documents Non-Indexable**: Number of documents that cannot be indexed; for example, a file source directory or a document with robots NOINDEX metatag.
- **Document Conversion Failures**: Number of document filtering errors. This is counted whenever a document cannot be converted to HTML format.

Crawler Log Files

The log file records all crawler activity, warnings, and error messages for a particular schedule. It includes messages logged at startup, run time, and shutdown. Logging everything can create very large log files when crawling a large number of documents. However, in certain situations, it can be beneficial to configure the crawler to print detailed activity to each schedule log file.

On the Global Settings - Crawler Configuration page, you can select either to log everything or to log only summary information. You can also select the language the crawler uses to generate the log file.

A new log file is created when you restart the crawler. The location of the crawler log file can be found on the Home - Schedules - Crawler Progress Summary page. The crawler maintains the past seven versions of its log file. The most recent log file is shown in the Oracle SES Administration GUI. You can view the other log files in the file system.

The format of the log file name is:

`search.crawler.i`*`SES_Instance_ID`*`ds`*`Data_Source_ID`*`.`*`timestamp`*`.log`

Where:

- *SES_Instance_ID* is the SID of the SES database.
- *Data_Source_ID* is the identifier of the data source being crawled.
- *timestamp* is the starting time in Greenwich Mean Time (GMT) 24-hour MMDDHHmm format (month, day, hour, minute).

Each logging message in the log file is one line, containing the following six tab delimited columns, in order:

1. Timestamp
2. Message level
3. Crawler thread name
4. Component name. It is typically the name of the executing Java class.
5. Module name. It can be internal Java class method name

6. Message

Crawler Configuration

Most crawler configuration tasks are controlled in the Oracle SES Administration GUI, but certain features (like title fallback, character set detection, and indexing the title of multimedia files) are controlled only by the Administration API. Configuration of the crawler is described by the `crawlerSettings` object.

The crawler uses a set of codes to indicate the crawling result of the crawled URL. Besides the standard HTTP status codes, it uses its own codes for non-HTTP related situations.

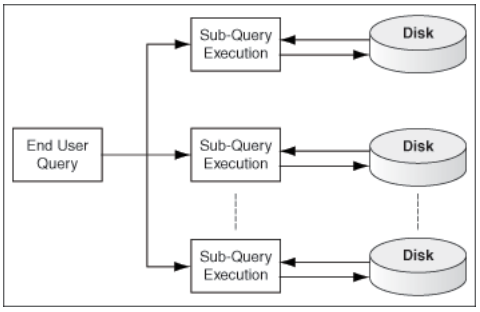
See Also: Appendix B, "URL Crawler Status Codes"

Parallel Query Indexing

In order to scale up the indexed data size while maintaining satisfactory query response time, the indexed data can be stored in independent disks to perform disk I/O operations in parallel. The major features of this architecture are:

- Oracle SES index is partitioned, so that the sub-queries are executed in parallel.
- Disks perform I/O operations independent of one another. As a result, the I/O bus contention does not create a significant bottleneck on the collective I/O throughput.
- Partition rules are used to control the document distribution among the partitions.

Figure 3–2 End User Query Partitioning



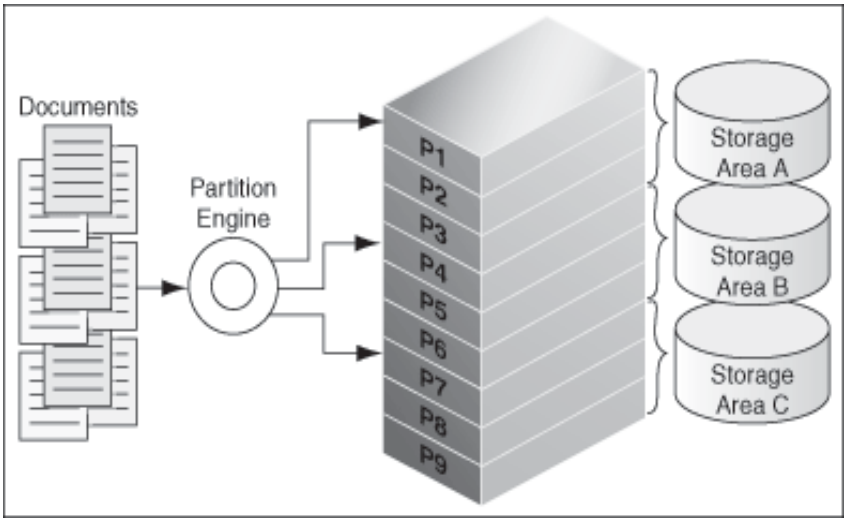
Document Partition Model and Storage Areas

Storage areas are used to store the partitions when the partitioning option is enabled. There are two kinds of partition mechanisms for improving query performance, attribute-based partitioning and hash-based partitioning. Currently, Oracle SES supports only hash-based partitioning.

Hash-based partitioning uses a hash function to distribute a large set of documents into multiple partitions. A partition engine controls the partition logic at both crawl time and query time. When a large data set must be searched without pruning the conditions, the end user request is broken into multiple parallel sub-queries so that the I/O and CPU resources can be used in parallel. After the result sets of the sub-queries are returned by the independent query processors, a merged result set is returned to the end user.

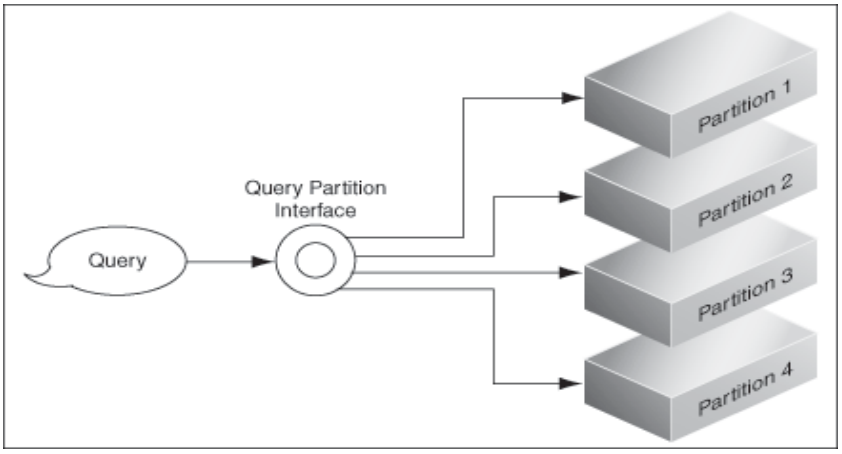
Figure 3–3 shows how the mechanism works during crawl time. The documents are partitioned and stored in different storage areas. Note that the storage areas are created on separate physical disks, so that I/O operations can be performed in parallel to improve the search turn around time.

Figure 3–3 Document Partitioning at Crawl Time



At query time, the query partition engine generates sub-queries and submits them to the storage areas, as shown in Figure 3–4.

Figure 3–4 Generation of Sub Queries at Query Time



See "Parallel Query and Index Partitioning" on page 10-14 for more information.

4

Customizing the Search Results

This chapter explains the various ways available for customizing the search results. It contains the following topics:

- Adding Suggested Content in Search Results
- Customizing the Relevancy of Search Attributes
- Providing Faceted Navigation

Adding Suggested Content in Search Results

Suggested content lets you display real-time data content along with the result list in the default query application. Oracle SES retrieves data from content providers and applies a style sheet to the data to generate an HTML fragment. The HTML fragment is displayed in the result list and is available through the Web Services API. For example, when an end user searches for contact information on a coworker, Oracle SES can fetch the content from the suggested content provider and return the contact information (e-mail address, phone number, and so on) for that person with the result list. Suggested content results appear in tabbed panes above the query results. When the query returns no results, suggested content is not displayed.

Configure suggested content on the Search - Suggested Content page in the Oracle SES Administration GUI. Enter the maximum number of suggested content results (up to 20) to be included with the Oracle SES result list. The results are rendered on a first-come, first-served basis.

Suggested Content Providers

Regular expressions (as supported in the Java regular expression API `java.util.regex`) are used to define query patterns for suggested content providers. The regular expression-based pattern matching is case-sensitive. For example, a provider with the pattern `dir\s(\S+)` is triggered on the query `dir james` but not on the query `Dir James`. To trigger on the query `Dir James`, the pattern could be defined either as `[Dd][Ii][Rr]\s+(\S+)` or as `(?i)dir\s+(\S+)`. A provider with a blank query pattern is triggered on all queries.

The URL you enter for the suggested content provider can contain the following variables: `$ora:q`, `$ora:lang`, `$ora:q1...$ora:q`*n* and `$ora:username`.

- `$ora:q` is the end user full query.
- `$ora:lang` is the two-letter code for the browser language.

- `$ora:qn` is the *n*th regular expression match group from the end user query. *n* starts from 1. If no *n*th group is matched, then the empty string replaces the variable.
- `$ora:username` is the end user name.

Enter an XSLT style sheet to define rules (for example, the size and style) for transforming XML content from a provider into an HTML fragment. This HTML fragment is displayed in the result list or returned over the Web Services API. If you do not enter an XSLT style sheet, then Oracle SES assumes that the suggested content provider returns HTML. If you do not enter an XSLT style sheet and the provider returns XML, then the result list displays the plain XML.

Note: As an administrator, you are responsible for verifying that the suggested content providers return valid and safe content. Corrupted or incomplete content returned by a suggested content provider can affect the formatting of the default query application results page.

Security Options

There are three security options for how Oracle SES passes the end user's authentication information to the suggested content provider:

- **None**: No security policy is used. (Default)
- **Cookie**: The end user first must be authenticated by the suggested content provider. A cookie is set for the user to maintain a session. Oracle SES must know the cookie used by the provider for authentication, and it is made available during registration of the suggested content provider. When the user enters a query, Oracle SES grabs the cookies from the user's request header and passes them to the provider. The cookie scope must be set to the common domain of the provider site and the Oracle SES site by the provider.

  For example, suppose the provider site is `http://provider.example.com` and the Oracle SES site is `http://ses.example.com`. After the end user logs in to the provider site, the site could set the value of the security cookie loginCookie with domain scope `.example.com`. When the end user searches in Oracle SES, Oracle SES gets the loginCookie value from the end user browser and forwards it to the provider site to get the suggested content (without login to the provider site again). However, if the provider site is accessed as `http://provider` or if the Oracle SES site is accessed as `http://SES`, then no domain cookie is available for sharing between the two sites and this security mechanism does not work.

  You can decide what happens when suggested content is available but the user is not logged in to the suggested content provider or the cookie for the provider is not available. For **Unauthenticated User Action**, if you select **Ignore content**, then content from that provider is not displayed in the result list. If you select **Display login message**, then Oracle SES returns a message that there is content available from this provider but the user is not logged in. The message also provides a link to log in to that provider. Enter the link for the suggested content provider login in the **Login URL** field.
- **Service-to-Service**: A one-way trusted relationship is established between Oracle SES and the suggested content provider. Any user logged in to Oracle SES does not need to be authenticated by the provider again. The provider only authenticates the Oracle SES application and trusts the Oracle SES application to act as the end user.

The end user identity is sent from Oracle SES to the provider site in the HTTP header `ORA_S2S_PROXY_USER`. The trusted entity could be a proxy user configured in the identity management system used by the provider, or it could be a name-value pair.

If the secured content provider must authenticate the end user, and it sets the domain level security cookie to maintain login information after the end user login, then use the cookie method for form authentication. The Oracle SES end user must login manually to the provider site, and the security cookie is stored in the browser. Oracle SES searches on the provider for the end user without additional login.

However, if the domain security cookie is not allowed for the provider, then the provider must support service-to-service security. The provider must allow an Oracle SES application account to search after passing HTTP basic or digest authentication. Also, if the provider has different secured content for different Oracle SES end users, then it must respect the end user security (in the HTTP header `ORA_S2S_PROXY_USER`) for the Oracle SES search request.

To register a provider that requires either HTTP basic or HTTP digest authentication, specify the authentication user name in the **Entity Name** field and specify the authentication password in the **Entity Password** field.

Example Configuring Google OneBox for Suggested Content

Existing OneBox providers can be configured as Oracle SES suggested content providers. For example, for a Google OneBox provider, the provider URL might be `http://host.company.com/apps/directory.jsp` and the trigger might be `dir\s(\S+)`. When the user query is "dir james", the provider receives the request with a query string similar to the following: `apiMaj=10&apiMin=1&oneboxName=app&query=james`.

With a suggested content provider, set the URL template as `http://host.company.com/apps/directory.jsp?apiMaj=10&apiMin=1&oneboxName=app&query=$ora:q1`. The provider pattern is the same: `dir\s(\S+)`. The XSLT used for Google OneBox can be re-used with a minor change. Look for the line:

```
<xsl:template name="apps">
```

and change that line in your template to

```
<xsl:template match="/OneBoxResults">
```

Customizing the Relevancy of Search Attributes

You can customize the default Oracle SES ranking to create a more relevant search result list for your enterprise. Ranking is determined by default and custom attributes. Default attributes include title, keywords, description, and others. Different weights indicate the importance of each attribute for document relevancy. For example, Oracle SES gives more weight to titles than to keywords.

To customize the relevancy of search results, you can use the Administration API or the Query Web Services API.

See Also:

- The `relevanceRanking` object in the *Oracle Secure Enterprise Search Administration API Guide*
- "Query Web Services Example: Customizing Relevancy" on page 11-29

Providing Faceted Navigation

Facets are a way of labeling data so that it can be navigated in different ways. It enables users to browse information based on a particular characteristic of the data by narrowing the possible search results. Many commercial sites on the Internet use facets to guide their customers to a purchase.

Faceted navigation provides these advantages:

- Searches return more relevant visits. Most queries are short and ambiguous, but a faceted navigation system provides an easy-to-use interface that enables users to clarify their intentions.
- Navigation becomes a form of discovery, in which users may find related items that they did not realize were what they are looking for.
- Users can pursue multiple paths to the same information. Each path is dynamically defined by the user through the navigation effort.

Different types of data have different facets. These are a few examples:

- Wine: Color, variety, country, vintage, price
- Recipes: Main ingredient, course, ethnicity, dietary restriction, holiday
- Shoes: Gender, size, color, style

For a description of faceted navigation in Oracle Fusion Applications, see the *Oracle Fusion Applications Administrator's Guide*.

To create a faceted navigation system, create a facet tree using the Administration API.

See Also: The `facetTree` object in the *Oracle Secure Enterprise Search Administration API Guide*.

Part II

Creating Data Sources

This part describes the data sources that you can set up for Oracle SES to crawl and index. It contains the following chapters:

- Chapter 5, "Configuring Access to Built-in Sources"
- Chapter 6, "Configuring Access to Content Management Sources"
- Chapter 7, "Configuring Access to Collaboration Sources"
- Chapter 8, "Configuring Access to Applications Sources"

5

Configuring Access to Built-in Sources

Among the built-in sources are the data repositories familiar to everyone, such as Web sites and e-mail. Most of them can be set up very quickly. This chapter contains the following topics:

- Setting Up Web Sources
- Setting Up Table Sources
- Setting Up File Sources
- Setting Up E-Mail Sources
- Setting Up Mailing List Sources
- Setting Up OracleAS Portal Sources
- Setting Up Federated Sources

Setting Up Web Sources

A Web source enables users to search a Web site. The following procedures identify the basic steps for setting up a Web source using the Oracle SES Administration GUI. For more information on each page, click **Help**.

Oracle SES is configured to crawl Web sites on the intranet within the corporate fire wall. To crawl Web sites on the Internet (external Web sites), Oracle SES requires the HTTP proxy server information. See the Global Settings - Proxy Settings page.

You should review the default crawling parameters before you start crawling Internet sources.

To create a Web source:

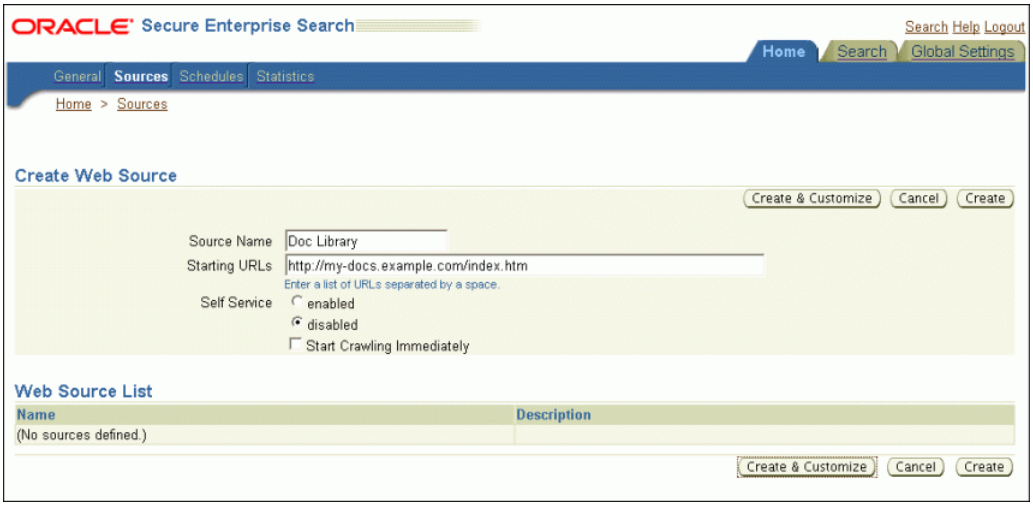
1. On the Home page, select the **Sources** secondary tab to display the Sources page.
2. For Source Type, select **Web**.
3. Click **Create** to display the Create Web Source page.
4. Complete the following fields:
   - **Source Name**: Name that you assign to this Web source.
   - **Starting URLs**: The HTTP or HTTPS address of the Web site, starting at the top page to be searched.
   - **Self Service**: **Disabled** to use an identity management system or **Enabled** to prompt users for their credentials.

- **Start Crawling Immediately**: Select this option to accept the default parameters and begin crawling, or deselect it to defer crawling.

5. Click **Create** or **Create & Customize**.
6. Follow the steps for crawling and indexing a source in "Getting Started Basics for the Administration GUI" on page 2-1.

Figure 5–1 shows the Create Web Source page.

Figure 5–1 Creating a Web Source



To customize a Web source:

1. When creating a Web source, click **Create & Customize** on the Create Web Source page to display the Customize Web Source page.

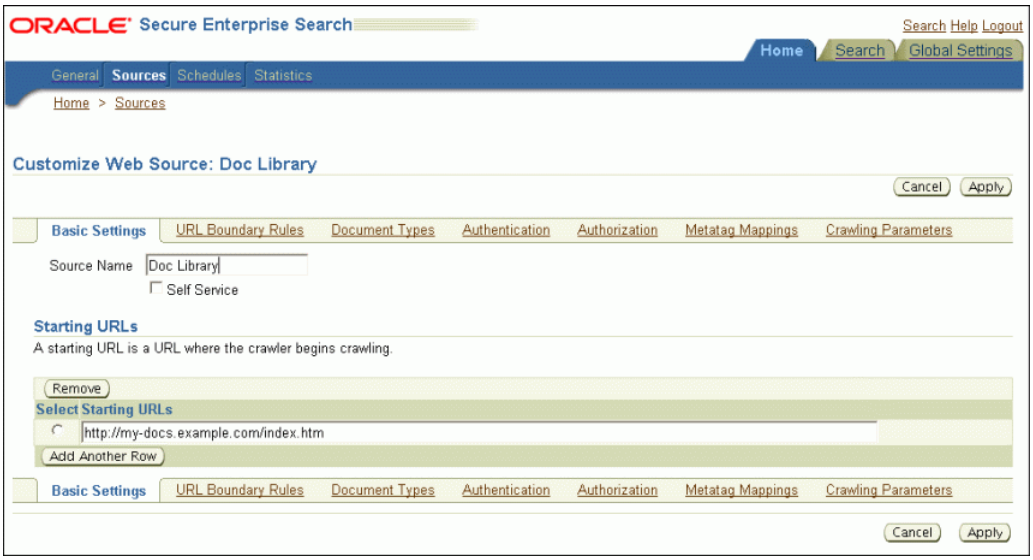
   or

   After creating a source, click the **Edit** icon on the Home - Sources page.

2. Click the following subtabs and make the desired changes.
   - **Basic Settings**: The choices entered on the Create Web Source page.
   - **Boundary Rules**: Contents of a URL that include or exclude a page from crawling.
   - **Document Types**: Common document and image types that you can include or exclude from crawling. By default, Oracle SES crawls HTML, Excel, PowerPoint, Word, PDF and plain text.
   - **Authentication**: Configuration of HTTP, HTML forms, or Oracle Single-Sign-On methods of authentication. By default, no authentication is required.
   - **Authorization**: Configuration of an Access Control List or an authorization manager plug-in.
   - **Metatag Mappings**: Maps document attributes to Oracle SES search attributes. See "Web Document Attributes" on page 5-3.
   - **Crawling Parameters**: Sets a variety of crawling conditions, such as depth, language, HTTP cookies.
3. Click **Apply**.

Figure 5–2 shows the Customize Web Source page.

Figure 5–2 Customizing a Web Source



Boundary Rules for Web Sources

When creating a Web source, the host name of the seed (top level URL) is automatically added to the boundary rule. However, subsequent changes to the seed URL are not reflected automatically to the rule. Remember to synchronize the boundary rule if there is any change to the seed URL. Currently, Oracle SES does not remove crawled URLs even if the original seed is removed: everything is controlled by the boundary rules.

Web Document Attributes

Oracle SES crawls and indexes these Web document attributes:

- Title
- Author
- Description
- Host
- Keywords
- Language
- LastModifiedDate
- Mimetype
- Subject: Mapped to "Description". If there is no description metatag in the HTML file, then it is ignored.
- Headline1: The highest H tag text; for example, "Annual Report" from `<H2>Annual Report</H2>` when there is no H1 tag in the page.
- Headline2: The second highest H tag text
- Reference Text: The anchor text from another Web page that points to this page.

You can define additional HTML metatags to map to a String attribute on the Home - Sources - Metatag Mapping page.

Setting Up Table Sources

A table source represents content in a database table or view. Table sources and database sources are similar, in that they both crawl database tables. The important difference between them is that the table source type was designed for ease of use, while the database source type was designed for high performance. If you are creating a source for a production environment, then see "Setting Up Database Sources" on page 8-6.

Database object names may be represented with a quoted identifier. A quoted identifier is case-sensitive and begins and ends with double quotation marks ("). If the database object is represented with a quoted identifier, then you must use the double quotation marks and the same case whenever you refer to that object.

When creating a table source in Oracle SES, if the table name is a quoted identifier, such as "1 (Table)", then in the **Table Name** field enter "1 (Table)", with the same case and double quotation marks. Similarly, if a primary key column or content column is named using a quoted identifier, then enter that name exactly as it appears in the database with double quotation marks.

> **See Also:** *Oracle Database SQL Reference* for more information about schema object names and qualifiers

The following procedures identify the basic steps for setting up a table source using the Oracle SES Administration GUI. For more information on each page, click **Help**.

To create a table source:

1. On the Home page, select the **Sources** secondary tab to display the Sources page.
2. For Source Type, select **Table**.
3. Click **Create** to display the Create Table Source page.
4. Complete the following fields. Click **Help** for additional information.

   Database Information

   - **Source Name**: Name that you assign to this table source.
   - **Database Host Name**: Name or IP address of the host computer for the database.
   - **TCP Port Number**: Port number of Oracle Net Listener. The default port number for Oracle Database is 1521.
   - **SID**: System identifier or instance name of the database.
   - **User Name**: Database user name with SELECT access rights to the table or view.
   - **Password**: Password for **User Name**.
   - **Delete Passwords After Crawl**: Select to ensure that Oracle SES does not store the database schema password for this source longer than is needed to crawl the source.

   Table Information

   - **Schema**: Name of the database schema that owns the table or view.

- **Table Name**: Name of the table or view.
- **Primary Key Column**: Column or set of columns that uniquely identify each row in the table or view. For performance, the primary key must be indexed. Click **Locate Table** for a list of columns in the table.
- **Content Column**: Column with searchable content for Oracle SES to crawl and index. The content column cannot be the same as the primary key column.
- **Content Type**: Binary, plain text, or HTML.

5. Click **Create** or **Create & Customize**.
6. Follow the steps for crawling and indexing a source in "Getting Started Basics for the Administration GUI" on page 2-1.

To customize a table source:

1. When creating a table source, click **Create & Customize** on the Create Table Source page to display the Customize Table Source page.

   or

   After creating a source, click the **Edit** icon on the Home - Sources page.
2. Click the following subtabs and make the desired changes.
   - **Basic Settings**: Identifies the source name and schema password. Any other changes to the original settings require a new source definition.
   - **Table Column Mappings**: Maps columns to Oracle SES search attributes. See "Table Search Attributes" on page 5-5.
   - **Language**: Identifies the default language and different column languages.
   - **Display URL**: Specifies the URL that users see for security reasons instead of the actual URL.
   - **Authorization**: Configuration of an Access Control List or an authorization manager plug-in.
3. Click **Apply**.

Table Search Attributes

Table sources have no predefined attributes. The crawler collects attributes from columns defined during source creation. You must map the columns to the search attributes.

Setting Up File Sources

A file source enables users to search files that are accessible from the Oracle SES query nodes through a commonly shared file system.

The following procedures identify the basic steps for setting up a file source using the Oracle SES Administration GUI. For more information on each page, click **Help**.

To create a file source:

1. On the Home page, select the **Sources** secondary tab to display the Sources page.
2. For Source Type, select **File**.
3. Click **Create** to display the Create File Source page.
4. Complete the following fields. Click **Help** for additional information.

- **Source Name**: Name that you assign to this table source.
- **Starting URL**: The URL of the top directory where the crawler begins. See "Tips for Using File Sources" on page 5-7.

5. Click **Create** or **Create & Customize**.
6. Follow the steps for crawling and indexing a source in "Getting Started Basics for the Administration GUI" on page 2-1.

To customize a file source:

1. When creating a file source, click **Create & Customize** on the Create File Source page to display the Customize File Source page.

   or

   After creating a source, click the **Edit** icon on the Home - Sources page.
2. Click the following subtabs and make the desired changes.
   - **Basic Settings**: Source name, language, and starting URL.
   - **URL Boundary Rules**: Contents of a URL that include or exclude a page from crawling.
   - **Document Types**: Common document and image types that you can include or exclude from crawling. By default, Oracle SES crawls HTML, Excel, PowerPoint, Word, PDF, and plain text.
   - **Display URL**: URL that users see for security reasons instead of the actual URL.
   - **Authorization**: Configuration of an Access Control List.

     or an authorization manager plug-in.
   - **Attribute Mapping**: Maps document attributes to Oracle SES search attributes. See "File Document Attributes" on page 5-6.
   - **Crawling Parameters**: Crawling conditions, such as depth, language, HTTP cookies.
3. Click **Apply**.

File Document Attributes

Oracle SES crawls and searches various attributes. By default, Oracle SES maps these search attributes to common document attributes, such as AUTHOR, CREATOR, KEYWORD, and SUBJECT. You can enter and map additional document attributes.

Oracle SES crawls and indexes these document attributes:

- Title
- Author
- Description
- Host
- Keywords
- Language
- LastModifiedDate
- Mimetype

- Subject

Tips for Using File Sources

This section contains the following topics:

- Crawling File Sources with Non-ASCII Character Sets
- Crawling File Sources with Symbolic Links
- Crawling File URLs
- Crawling File Sources from a Network Drive

Crawling File Sources with Non-ASCII Character Sets

For file sources to successfully crawl and display multibyte environments, the locale of the computer that starts the Oracle SES server must be the same as the target file system. This way, the Oracle SES crawler can detect the multibyte files and paths.

If the locale is different in the installation environment, then Oracle SES must be reinstalled from the environment with the correct locale. For example, for a Korean environment, either set LC_ALL to ko_KR **or** set both LC_LANG **and** LANG to ko_KR.KSC5601. Then restart Oracle SES as described in "Starting and Stopping Oracle SES" on page 2-5.

Crawling File Sources with Symbolic Links

When crawling file sources on UNIX, the crawler resolves any symbolic link to its true directory path and enforce the boundary rule on it. For example, suppose directory /tmp/A has two children, B and C, where C is a link to /tmp2/beta. The crawl has the following URLs:

- /tmp/A
- /tmp/A/B
- /tmp2/beta
- /tmp/A/C

If the inclusion rule is `/tmp/A`, then /tmp2/beta is excluded. The seed URL is treated as is.

Crawling File URLs

For a plug-in to return file URLs to the crawler, the file URLs must be fully qualified. For example, `file://localhost/`.

If a file URL is to be used "as is", without going through Oracle SES to retrieve the file, then "file" in the Display URL Prefix should be upper case "FILE". For example, `FILE://localhost/...`. The starting URL is not case sensitive.

"As is" means that when a user clicks the search link of the document, the browser tries to use the specified file URL on the client computer to retrieve the file. Without that, Oracle SES uses this file URL on the server computer and sends the document through HTTP to the client computer.

Crawling File Sources from a Network Drive

If the files are crawled from a network drive, then the Oracle process should be started as a user who has access to the drive.

See Also: "Required Tasks" on page 7-12 for instructions on how to change the user running the Oracle process.

Setting Up E-Mail Sources

An e-mail source enables users to search e-mail messages on an IMAP server. The following procedures identify the basic steps for setting up an e-mail source using the Oracle SES Administration GUI. For more information on each page, click **Help**.

For Oracle Collaboration Suite and Oracle Beehive mail sources, use the "Setting Up Oracle Collaboration Suite E-Mail Sources" on page 7-23.

To create an e-mail source:

1. On the Home page, select the **Sources** secondary tab to display the Sources page.
2. For Source Type, select **E-mail**.
3. Click **Create** to display the Create E-Mail Source page.
4. Complete the following fields. Click **Help** for additional information.
   - **Source Name**: Name that you assign to this table source.
   - **IMAP Server**: Address of the IMAP server, such as `mail.example.com`.
   - **Archive**: Directory where crawled messages are stored. On Oracle RAC, this directory must be in the shared crawler log directory, which is set on the Global Settings - Crawler Configuration page.
   - **Self Service**: **Enabled** to prompt users for their credentials, or **Disabled** to provide credentials for a single user on this page.
5. Click **Create**.
6. Follow the steps for crawling and indexing a source in "Getting Started Basics for the Administration GUI" on page 2-1.

E-Mail Attributes

Oracle SES crawls and searches these search attributes.

- Author
- Title
- Subject
- Language
- LastModifiedDate

Setting Up Mailing List Sources

A mailing list source enables users to search messages that were sent to a mailing list on an IMAP server.

The Oracle SES crawler is IMAP4 compliant. To crawl mailing list sources, you need an IMAP e-mail account. Oracle recommends that you create an e-mail account that is used solely for Oracle SES to crawl mailing list messages. The crawler is configured to crawl one IMAP account for all mailing list sources. Therefore, all mailing list messages to be crawled must be found in the Inbox of the e-mail account specified on this page. This e-mail account should be subscribed to all the mailing lists. New

postings for all the mailing lists are sent to this single account and subsequently crawled.

Messages deleted from the global mailing list e-mail account are not removed from the Oracle SES index. The mailing list crawler deletes messages from the IMAP e-mail account as it crawls. The next time the IMAP account for mailing lists is crawled, the previous messages no longer exist. Any new messages in the account are added to the index and consequently deleted from the account. This keeps the global mailing list IMAP account clean. The Oracle SES index serves as a complete archive of all the mailing list messages.

The following procedures identify the basic steps for setting up a mailing list source using the Oracle SES Administration GUI. For more information on each page, click **Help**.

To create a mailing list source:

1. Enter the global mailing list settings:
   a. On the Global Settings page, choose **Mailing List Settings** under Sources to display the Global Mailing List Settings page.
   b. Complete the following fields. Click **Help** for additional information.

      User Name: IMAP e-mail account that is used to crawl the messages. This user must be on all of the mailing lists identified as a mailing list source.

      Password: Password for User Name.

      IMAP Server: Address of the IMAP server, such as `mail.example.com`.
   c. Click **Apply**.
2. On the Home page, select the **Sources** secondary tab to display the Sources page.
3. For Source Type, select **Mailing List**.
4. Click **Create** to display the Create Mailing List Source page.
5. Complete the following fields. Click **Help** for additional information.
   - **Source Name**: Name that you assign to this table source.
   - **Mailing List**: Name of the mailing list to be searched, such as `news@example.com`.
6. Click **Create**.
7. Follow the steps for crawling and indexing in "Getting Started Basics for the Administration GUI" on page 2-1 for the mailing list schedule.

Mailing List Attributes

Oracle SES crawls and indexes these search attributes.

- Author
- Title
- Subject
- Language
- LastModifiedDate

Setting Up OracleAS Portal Sources

An OracleAS Portal source enables users to search across multiple portal installations and repositories, such as Web pages, disk files, and pages on other OracleAS Portal instances. Oracle Secure Enterprise Search can securely crawl both public and private OracleAS Portal content.

To create an OracleAS Portal source:

1. On the Home page, select the **Sources** secondary tab to display the Sources page.
2. For Source Type, select **OracleAS Portal**.
3. Click **Create** to display the Create OracleAS Portal Source page.
4. Complete the following fields. Click **Help** for additional information.
   - **Source Name**: Name that you assign to this OracleAS Portal source.
   - **URL Base**: Base URL for OracleAS Portal.
   - **Page Groups**: List of page groups in OracleAS Portal retrieved when you click **Retrieve Page Groups**. Select the ones to crawl.
5. Click **Create & Customize**.
6. Select the **Authentication** tab.
7. Select **Enable OracleAS Single Sign-On Authentication** and enter your credentials.
8. Click **Apply**.
9. Follow the steps for crawling and indexing in "Getting Started Basics for the Administration GUI" on page 2-1 for the mailing list schedule.

Crawling a Folder or Page

The portal crawler can crawl a subtree under a specific folder or page instead of under an entire portal tree.

To set the boundary rule to crawl a specific folder or page:

1. On the Home page, click the **Sources** secondary tab to display the Sources page.
2. Select a source and click **Edit** to display the Edit User-Defined Source page.
3. Click the **URL Boundary Rules** subtab.
4. Under **Inclusion Rules** for the URL, select the **starts with** rule and enter the value of the `PORTAL_PATH` for the folder or page.

   For example, to crawl only the P2 subtree of a portal tree, enter the path from the root to P2, such as `/Proot/P1/P2`.

OracleAS Portal Search Attributes

The crawler picks up key attributes offered by OracleAS Portal, as described in Table 5–1.

Table 5–1 OracleAS Portal Source Attributes

Attribute	Description
createdate	Date the document was created
creator	User name of the person who created the document

Table 5–1 (Cont.) OracleAS Portal Source Attributes

Attribute	Description
author	User-editable field so that they can specify a full name or whatever they want
page_path	Hierarchy path of the portal page/item in the portal tree (contains page titles)
portal_path	Hierarchy path of the portal page/item in the portal tree, used for browsing and boundary rules (contains page names) When searching OracleAS Portal 10.1.2, portal_path appears as upper case in the browse. When searching OracleAS Portal 10.1.4, portal_path appears in lowercase.
title	Title of the document
description	Brief description of the document
keywords	Keywords of the document
expiredate	Expiration date of the document
host	Portal host
infosource	Path of the Portal page in the browse hierarchy
language	Language of the portal page or item
lastmodifieddate	Last modified date of the document
mimetype	Usually 'text/html' for portal
perspectives	User-created markers that can be applied to pages or items, such as 'INTERNAL ONLY', 'REVIEWED', or 'DESIGN SPEC'. For example, a Portal containing recipes could have items representing recipes with perspectives such as 'Breakfast', 'Tea', 'Contains Nuts', 'Healthy' and one particular item could have several perspectives assigned to it.
wwsbr_name_	Internal name of the portal page or item
wwsbr_charset_	Character set of the portal page or item
wwsbr_category_	Category of the portal page or item
wwsbr_updatedate_	Date the last time the portal page or item was updated
wwsbr_updator_	Person who last updated the page or item
wwsbr_subtype_	Subtype of the portal page/item (for example, container)
wwsbr_itemtype_	Portal item type
wwsbr_mime_type_	Mimetype of the portal page or item
wwsbr_publishdate_	Date the portal page or item was published
wwsbr_version_number_	Version number of the portal item

Tips for Using OracleAS Portal Sources

- An OracleAS Portal source name cannot exceed 35 characters.
- URL boundary rules are not enforced for URL items. A URL item is the metadata that resides on the OracleAS Portal server. Oracle SES does not touch the display URL or the boundary rules for URL items.
- The `portal_path` attribute is used to compare boundary rules. Portal pages and items are organized in a tree structure. When a page is included or excluded, its entire subtree starting with that node is included or excluded.

- If OracleAS Portal user privileges change, the content the crawler collects might not be properly authorized. For example, in a Portal crawl, the user specified in the Home - Sources - Authentication page does not have privileges to see certain Portal pages. However, after privileges are granted to the user, on subsequent incremental crawls, the content still is not picked up by the crawler. Similarly, if privileges are revoked from the user, the content might still be picked up by the crawler.

  To be certain that Oracle SES has the correct set of documents, whenever a user's privileges change, update the crawler re-crawl policy to **Process All Documents** on the Home - Schedules - Edit Schedules page, and restart the crawl.

Setting Up Federated Sources

Secure federated search enables searching secure content across distributed Oracle SES instances. An end user is authenticated to the Oracle SES federation broker. Along with querying the secure content in its own index, the federation broker federates the query to each federation endpoint on behalf of the authenticated end user. This mechanism necessitates propagation of user identity between the Oracle SES instances. In building a secure federated search environment, an important consideration is the secure propagation of user identities between the Oracle SES instances. This section explains how Oracle SES performs secure federation.

To create a federated source:

1. On the Home page, select the **Sources** secondary tab to display the Sources page.
2. For Source Type, select **Federated**.
3. Click **Create** to display the Create Federated Source page.
4. Complete the following fields. See "Federation Trusted Entities" on page 5-13 and click **Help** for additional information.
   - **Source Name**: Name that you assign to this federated source.
   - **Web Service URL**: The URL for the Web service.
   - **Remote Entity Name**: Name of the federation trusted entity on the federation endpoint.
   - **Remote Entity Password**: Password for Remote Entity Name.
   - **Search User Attribute**: Attribute used to authenticate users on the federation endpoint instance.
   - **Filter Rule**: Conditions for routing queries to this federated source. Filter rules can improve scalability. If no rule is defined, then the federation agent sends all queries to the federated source to perform the search.
5. Click **Create** or **Create & Customize**.
6. Follow the steps for crawling and indexing a source in "Getting Started Basics for the Administration GUI" on page 2-1.

To customize a federated source:

1. When creating a federated source, click **Create & Customize** on the Create File Source page to display the Customize File Source page.

   or

   After creating a source, click the **Edit** icon on the Home - Sources page.

2. Click the following subtabs and make the desired changes. See "Customizing Federated Sources" on page 5-15.
   - **Basic Settings**: Source name, Web Service URL, and so forth.
   - **Search Restrictions**: Controls whether the search is restricted, and if so, which source groups are searched.
   - **Attribute Retrieval**: Lists search attributes to retrieve at query time.
   - **Attribute Mapping**: Maps local and remote search attributes.
3. Click **Apply**.

See Also: "Configuring Secure Search with OracleAS Single Sign-On" on page 9-20

Federation Trusted Entities

When performing a secure search on a federation endpoint, the federation broker must pass the identity of the logged-in user to the federation endpoint. If the endpoint instance trusts the broker instance, then the broker instance can proxy as the end user. To establish this trust relationship, Oracle SES instances should exchange some secret. This secret is exchanged in the form of a *trusted entity*.

A trusted entity consists of two values: entity name and entity password. Each Oracle SES instance can have one or more trusted entities that it can use to participate in secure federated search. (A trusted entity is also referred to as a proxy user.)

An Oracle SES instance can connect to an identity management (IDM) system for managing users and groups. An IDM system can be an LDAP-compliant directory, such as Oracle Internet Directory or Active Directory.

Each trusted entity can be authenticated by either an IDM system or by the Oracle SES instance directly, independent of an IDM system. For authentication by an IDM system, check the box **Use Identity Plug-in for authentication** when creating a trusted entity. In this case, the entity password is not required. This is useful when there is a user configured in the IDM system that can be used for proxy authentication. Ensure that the entity name is the name of the user that exists in the IDM system and is going to be used as the proxy user.

For authentication of the proxy user by Oracle SES, deselect **Use Identity Plug-in for authentication** when creating a trusted entity. Then use any name and password pair to create a trusted entity.

Use **Authentication Attribute** to specify the format of the user credential that the Oracle SES federation endpoint expects for this particular trusted entity in proxy authentication. The identity plug-in registered on the federation endpoint should be able to map this user identity to the default authentication format used on the federation endpoint. This is useful when a federation broker cannot send user identity in the default authentication format used on the federation endpoint for proxy authentication, but the identity plug-in registered on the federation endpoint can map the value from the attribute in which it receives the user identity during proxy authentication to the default authentication format used on the federation endpoint.

To use a proxy entity, use the Web services API `proxyLogin` user name and password for the entity name and entity password. The identity plug-in can validate the password instead of storing it. When a request is sent for `proxyLogin`, Oracle SES calls the identity plug-in (which returns the call) to authenticate the entity. The `proxyLogin` must supply a valid trusted entity registered in the federation trusted entities.

User names are not case sensitive.

To perform secure federated search, both the broker and the endpoint instances involved in the federation must have identity plug-ins registered. The identity plug-ins may or may not talk to the same IDM system.

> **Note:** All user names should be unique across all Oracle SES instances. If not, then there should be a clear mapping for the users to make them unique across all IDMs involved in the secure federation.

Carefully specify the following parameters under the section **Secure Federated Search** when creating a federated source on the broker instance:

- **Remote Entity Name**: This is the name of the federation trusted entity on the federation endpoint. It is provided by the administrator of the endpoint instance.
- **Remote Entity Password**: This is the password of the federation trusted entity on the federation endpoint. It is provided by the administrator of the endpoint instance.
- **Search User Attribute**: This attribute identifies, and is used to authenticate, a user on the federation endpoint instance. This parameter is optional parameter, unless the broker and endpoint use different authentication attributes to identify end users. For example, on the broker instance, an end user can be identified by user name; on the endpoint instance, the end user can be identified by e-mail address.

  The identity plug-in registered on the broker instance should be able to map the user identity to this attribute based on the authentication attribute used during the registration of the identity plug-in. If this attribute is not specified during creation of the federation source, then the user identity on the broker instance is used to search on the endpoint instance.

  > **Note:** If these parameters are not specified during the creation of the federated source, then the federated source is treated as a public source (that is, only public content is available to the search users).

- **Secure Oracle HTTP Server-Oracle SES channel**: Because any Oracle HTTP Server can potentially connect to the AJP13 port on the Oracle SES instances and masquerade as a specific person, either the channel between the Oracle HTTP Server and the Oracle SES instance must be SSL-enabled or the entire Oracle HTTP Server and Oracle SES instance computers must be protected by a fire wall.

  > **Notes:**
  >
  > - In a secure federated search environment, the broker or the endpoint instance might or might not be using OracleAS Single Sign-On. However, the Web service URL of the endpoint should not be behind OracleAS Single Sign-On.
  > - Oracle strongly recommends that you SSL-protect the channel between Oracle HTTP Server and Oracle SES for secure content. The endpoint instance should be SSL-enabled, or you should be able to access the Web service using HTTPS.

See Also: "Tips for Using Federated Sources" on page 5-18

Example Creating a Federated Source

This section describes the steps for setting up a federated source that connects to Active Directory.

1. Activate the Active Directory identity plug-in on both the endpoint and broker instances. For example, on the Global Settings - Identity Management Setup page, enter the following:
   - **Parameter Name**: value
   - **Directory URL**: `ldap://ad.oracle.com:389`
   - **Directory account name**: administrator@ad.oracle.com
   - **Directory account password**: Password for **Directory account name**.
   - **Directory subscriber**: dc=ad,dc=oracle,dc=com
   - **Directory security protocol**: none
2. Create federation trusted entities on the endpoint instance. For example, login to Oracle SES on the endpoint instance, navigate to the Global Settings - Federation Trusted Entities page, and enter the following:
   - **Entity Name**: Entity name
   - **Entity Password**: Password for **Entity Name**
3. Create a federated source on the broker side. For example, login to Oracle SES on the broker instance, navigate to the Home - Sources page, select the source type as Federated, and enter the following:
   - **Source Name**: Sourcename1
   - **Web Service URL**: `http://endpoint.cn.oracle.com:7777/search/query/OracleSearch`
   - **Remote Entity Name**: Entity name
   - **Remote Entity Password**: Password
4. To browse the federated source on broker side, create a source group and then add the federated source to the group.

Customizing Federated Sources

On the Home - Sources - Customize Federated Source page, you can change the source name, Web Service URL, remote entity name and password, and search user attribute.

This section describes the other ways you can customize a federated source:

- Route Queries to the Federated Source
- Set Search Restrictions
- Retrieve Attributes
- Map Attributes

Route Queries to the Federated Source

Enter a filter rule, which sets conditions for routing queries to the federated source, on the Home - Sources - Customize Federated Source page. Filter rules can improve scalability. If no rule is defined, then the federation agent sends all queries to the

federated source to perform the search. The rules are applied only against the search query filter. They are not applied when an end user enters the attribute shortcut query.

Each rule has an attribute, a colon (:) and an expression. Rules can be based on end user properties, such as name or e-mail address, and on query information, such as document language, author, or document modified date. For example, an identity attribute could be `mail` or `dn`. A query attribute could be `author` or `lastmodifieddate`.

Multiple rules for the source are joined with the AND and OR operators. The attribute name and the operators are not case-sensitive. For example, the following rule defines that the federated source is for English documents and for users having an e-mail address starting with A in the identity management system:

```
(language:en ) AND (idm::mail:a.*)
```

The attribute can be Date, String, or Number type. For String attributes, the rule expression is regular expression. Oracle SES supports the regular expression syntax used in Java JDK 1.4.2 Pattern class (java.util.regex.Pattern). For Date and Number attributes, the expression contains the operator and value. The operators are =, >, >=, <, <=.

Filter Rule Examples The following rule defines that the federated source is for documents larger than 1 M:

```
content-length:>1000000
```

The following rule defines that the federated source is for documents published after 12/31/2006:

```
lastmodifieddate:> 12/31/2006
```

The following example defines that the federated source has only documents for the last week:

```
lastmodifieddate:> sysdate - 7
```

The following rule defines that the federated source is for the login name, which could be an attribute of the identity management repository:

```
username:test00.*
```

Set Search Restrictions

Restrict search to a specific list of source groups on the Home - Sources - Customize Federated Source - Search Restrictions page.

Available source groups from the federated source are retrieved when the page is loaded. When Source Group Restricted Search is selected, you can move the source groups between the **Not Searched** and **Searched** lists. When **Unrestricted Search** is selected, all source groups on the remote instance are searched.

The **Refresh Source Groups** button refreshes the available source groups from the remote instance. If a source group is no longer available, then it is marked **Not Available**. All newly available source groups after a refresh appear in the **Not Searched** list by default, and all existing source groups remain in the list they are presently in. If a remote source group is renamed, the old name is marked **Not Available** and the new name appears in the **Not Searched** list. Unavailable source groups persist while they remain in the Searched list.

If the federated source is unavailable, then the available source groups are loaded from local storage. A warning message then states that Oracle SES cannot retrieve the available source groups from the remote instance, indicating that the available source groups may be out of date.

> **Note:** A federated source can be restricted to only explicitly-created source groups on the remote Oracle SES instance. For example, a federated source cannot be restricted to the Miscellaneous group on the remote Oracle SES instance.

Retrieve Attributes

Identify which attributes to retrieve from the federated source on the Home - Sources - Customize Federated Source - Attribute Retrieval page.

Available attributes from the federated source are retrieved when the page is loaded. Move search attributes to retrieve between the **Not Retrieved** column and the **Retrieved** column. Attributes that are always retrieved by Oracle SES by default are in the **Retrieved** list and marked **Mandatory**. These attributes cannot be saved in the **Not Retrieved** list.

The **Refresh Attributes** button refreshes the available attributes from the remote instance. If an attribute is no longer available, then it is marked **(Not Available)**. All newly available attributes after a refresh appear in the **Not Retrieved** list by default, and all existing attributes remain in the list they are presently in. If a remote attribute is renamed, then the old attribute name is marked **Not Available**, and the new name appears in the **Not Retrieved** list. Unavailable attributes persist while they remain on the Retrieved list or are used in an explicit attribute mapping.

If the federated source is unavailable, then the available attributes are loaded from local storage. A warning message then states that Oracle SES cannot retrieve the available attributes from the remote instance, so the available attributes may be out of date.

Map Attributes

Map local search attributes with federated search attributes on the Home - Sources - Customize Federated Source - Attribute Mapping page. For example, a local search attribute named Creator can be mapped to a remote attribute named Author. This is an explicit attribute mapping. Only one-to-one mappings between attributes of the same data type are supported.

> **Note:** For default Oracle SES search attributes, Oracle SES implicitly maps local attributes to remote attributes. For example, a remote attribute named Author is always mapped to local search attribute name Author. For all other attributes, explicit mappings must be created.

Local search attributes are the available attributes on the local instance, as defined on the Global Settings - Search Attributes page. Local search attributes that are used in a mapping cannot be deleted on the Global Settings - Search Attributes page. Initially, there are no mappings.

Remote search attributes are the available attributes on the federated source. This list is retrieved when the page is loaded. If a remote attribute is mapped to a local attribute but the remote attribute is no longer available, then the remote attribute is marked

(Not available). Only attribute mappings involving available remote attributes are used during queries.

Tips for Using Federated Sources

- The Oracle SES federator caches the federator configuration (that is, all federation-related parameters including federated sources). As a result, any change in the configuration takes effect within five minutes.
- If you entered proxy settings on the Global Settings - Proxy Settings page, then add the Web Services URL for the federated source as a proxy exception.
- If the federation endpoint instance is set to secure mode 3 (require login to search secure and public content), then all documents (ACL stamped or not) are secure. For secure federated search, create a trusted entity in the federation endpoint instance, then edit the federated source with the trusted entity user name and password.
- There can be consistency issues if you have configured a BIG-IP system as follows:
  - You have two Oracle SES instances configured identically (same crawls, same sources, and so on) behind a BIG-IP load balancer to act as a single logical Oracle SES instance.
  - You have two other Oracle SES instances configured identically along with Oracle HTTP Server and OracleAS Web Cache fronting each one and both servers behind BIG-IP. Each of these two instances federate to the logical Oracle SES instance. Web Cache is clustered between these two nodes to act as a single logical Oracle SES instance called broker instance.

  When a user performs a search on the broker Oracle SES instance and tries to access the documents in the result, document access may not be consistent each time. As a work-around, ensure that the load balancer sends all the requests in one user session to the exact same node each time.

Looping Among Federated Sources

A federation loop or cycle refers to a deployment in which multiple SES instances federate to each other. For example, if SES Instance A federates to SES Instance B, and SES Instance B federates back to instance A, then a federation cycle is in the deployment. Federation cycles can cause a flood of queries and high CPU load on the participating SES instances.

SES does not detect federation cycles, thus the Oracle SES administrator is responsible for avoiding them. You can explicitly remove them from the deployments or use source-group-restricted federation. The previous example can be fixed with a source-group restriction: the source groups on Instance B selected for federation on Instance A do not have any federated sources for Instance A, and the converse. See "Set Search Restrictions" on page 5-16.

Federated Search Characteristics

- Federated search can improve performance by distributing query processing on multiple computers. It can be an efficient way to scale up search service by adding a cluster of Oracle SES instances.
- The federated search quality depends on the network topology and the throughput of the entire federated Oracle SES environment.

Federated Search Limitations

- There is a size limit of 200KB for the cached documents existing on the federation endpoint to be displayed on the Oracle SES federation broker instance.
- For infosource browse, if the source hierarchies for both local and federated sources under one source group start with the same top level folder, then a sequence number is added to the folder name belonging to the federated source to distinguish the two hierarchies on the Browse page.
- For federated infosource browse, a federated source should be put under an explicitly created source group.
- On the Oracle SES federation broker, there is no direct access to documents on the federation endpoint through the display URL in the search result list for the following source types:
  - File (local files, not UNC)
  - Table
  - E-mail
  - Mailing list

  For these source types, only the cached version of documents is accessible.

6

Configuring Access to Content Management Sources

This chapter contains the following topics:

- Setting Up EMC Documentum Content Server Sources
- Setting Up Microsoft SharePoint Sources
- Setting Up Oracle Content Database Sources
- Setting Up Oracle Content Server Sources

Setting Up EMC Documentum Content Server Sources

Documentum data is stored in DocBases, which can contain cabinets and folders. A Documentum Content Server instance can have one or more DocBases crawled with an EMC Documentum Content Server source. The Documentum Content Server source navigates through the DocBases and the inline cabinets to crawl all the documents in Documentum Content Server. Oracle SES creates an index, stores the metadata, and accesses information in Oracle SES to provide search capabilities according to the end user permissions.

Oracle SES supports incremental crawling; that is, it crawls and indexes only those documents that have changed after the most recent crawling was scheduled. A document is re-crawled if either the content or metadata or the direct security access information of the document has changed. A document is also re-crawled if it is moved within Documentum Content Server and the end user has to access the same document with a different URL. Documents deleted from a DocBase are removed from the index during incremental crawling.

Important Notes for EMC Documentum Content Server Sources

The Documentum source in Oracle SES must use the administrator account of a DocBase for crawling and indexing documents of that DocBase.

Required Software

- Documentum Content Server DA (Documentum Administrator) *or* Documentum Content Server WebTop application must be installed and configured.
- Documentum Foundation Classes (DFC) must be installed on the server running Oracle SES.
- Currently supported Documentum version is 6.5.

Required Tasks

- Because EMC Documentum Content Server software is not included with Oracle SES, certain files must be copied manually into Oracle SES.

  The DFC installation asks for destination directory and user directory. For Windows, the default destination directory is `C:\Program Files\Documentum` and default user directory is `C:\Documentum`.

  For UNIX, you must create a DFC program root and a DFC user root. For example, DFC program root might be *user_home*/documentum_shared and DFC user root might be *user_home*/documentum.

- Copy files from EMC Documentum Content Server. The files may be stored in the `shared`, `dfc`, or `config` subdirectories.

  – Copy these files to *ORACLE_HOME*`/search/lib/plugins/dcs/`:

  ```
  dctm.jar
  dfc.jar
  dfcbase.jar
  aspectjrt.jar
  certjFIPS.jar
  jsafeFIPS.jar
  configservice-api.jar
  dfc.properties
  ```

  – Create a subdirectory with a name such as `dcsothers` in *ORACLE_HOME*`/search/lib/plugins/dcs/`, and make a second copy of `dfc.properties` in it.

- Add the following to `DMCL.ini`:

  ```
  max_session_count = 20
  max_connection_per_session = 20
  ```

  In Windows, `DMCL.ini` is located in the `WINNT` folder. In Linux, `DMCL.ini` is available in the Documentum folder (DFC user root).

- In Windows 2003 server, copy `dmcl40.dll` from *DFC_destination_directory*`/shared/` to *ORACLE_HOME*`/product/`*version*`/`*SES Instance Name*`/BIN`. For UNIX platforms, copy the file according to Table 6–1.

- The environment variables $DOCUMENTUM_SHARED (DFC Program root) and $DOCUMENTUM (DFC user directory) must be created before installing DFC on Linux. Also note that these variables must to be exported again, and Oracle SES must be restarted when the system restarts. These variables can also be exported permanently in Linux.

  Use the following commands to export environmental variables in Linux:

  For DOCUMENTUM:

  ```
  export DOCUMENTUM=/home/sesuser/DOCUMENTUM
  ```

  For DOCUMENTUM_SHARED:

  ```
  export DOCUMENTUM_SHARED=/home/sesuser/DOCUMENTUM_SHARED
  ```

- Restart the middle tier.

  On Windows, restart the computer after installing DFC.

Table 6–1 DFC Files to Copy for UNIX Platforms

Platform	Copy File	From	To
Linux x86	libdmcl40.so	*DFC_destination_directory*/dfc	$ORACLE_HOME/lib
Linux x86-64	libdmcl40.so	*DFC_destination_directory*/dfc	$ORACLE_HOME/lib32
Solaris SPARC (64-bit)	libdmcl40.so	*DFC_destination_directory*/dfc	$ORACLE_HOME/lib32
HP-UX PA-RISC (64-bit)	libdmcl40.sl	*DFC_destination_directory*/dfc	$ORACLE_HOME/lib32
AIX 5L Based Systems (64-bit)	libdmcl40.so	*DFC_destination_directory*/dfc	$ORACLE_HOME/lib32
HP-UX Itanium	libdmcl40.so	*DFC_destination_directory*/dfc	$ORACLE_HOME/lib32

Known Issues

- In this release, search results cannot be viewed in Documentum desktop. The documents and folders can be viewed only using Documentum Administrator (DA) or Webtop applications.
- For the **Container name** parameter, a value of repository name alone might not work. Enter the value of *RepositoryName/CabinetName*. For example, *DocBaseName/CabinetName/FolderName/SubFolderName*.
- Incremental crawls do not recognize an ACL modification of access permissions from None to Browse and Browse to None. The DCSCHECKSUM attribute value is same for both settings.

Configuration for Documentum Content Server 6.5

For Windows, the JAR files can be taken from the application server directory where DA is deployed. For DFC installation on Linux, it is a prerequisite to create DFC program root and DFC user root. For example, the DFC program root can be *USER HOME*/DOCUMENTUM_SHARED and the DFC user root can be *USER HOME*/DOCUMENTUM. Table 6–2 lists the location of the JAR files in Windows and Linux.

Table 6–2 Location of the JAR Files

JAR File Name	Windows Location	Linux Location
dfc.jar	*Application server home directory/da deployment directory*/WEB-INF/lib	*DFC_destination_directory*
aspectjrt.jar	*Application server home directory/da deployment directory*/WEB-INF/lib	*DFC_destination_directory*/dfc
certjFIPS.jar	*Application server home directory/da deployment directory*/WEB-INF/lib	*DFC_destination_directory*/dfc
jsafeFIPS	*Application server home directory/da deployment directory*/WEB-INF/lib	*DFC_destination_directory*/dfc
dfc.properties	*Application server home directory/da deployment directory*/WEB-INF/classes	*DFC_user_directory*/config/
configservice-api.jar	*Application server home directory/da deployment directory*/WEB-INF/lib	*DFC_destination_directory*/dfc

To configure the crawler plug-in:

1. Create a new directory under *ORACLE_HOME*/search/lib/plugin/dcs/. For example, dcsothers.
2. Copy dfc.properties to the folder created in the previous step (dcsothers) and to the main folder (dcs).
3. Copy dfc.jar, aspectjrt.jar, certjFIPS.jar, jsafeFIPS.jar, configservice-api.jar to the dcs folder in the following path *ORACLE_HOME*/search/lib/plugin/dcs.
4. The environment variables $DOCUMENTUM_SHARED (DFC Program root) and $DOCUMENTUM (DFC user directory) must be created before installing DFC on Linux. Also note that the environment variables $DOCUMENTUM_SHARED, $DOCUMENTUM, and $CLASSPATH must be exported again, and Oracle SES must be restarted when the computer restarts. These variables can also be exported permanently in Linux.

   Export environmental variables in Linux using commands like these:

   For DOCUMENTUM:

   ```
   export DOCUMENTUM=/home/sesuser/DOCUMENTUM
   ```

   For DOCUMENTUM_SHARED:

   ```
   export DOCUMENTUM_SHARED=/home/sesuser/DOCUMENTUM_SHARED
   ```

   For CLASSPATH:

   ```
   export CLASSPATH=$DOCUMENT_SHARED/dctm.jar:$DOCUMENTUM_SHARED/config
   ```

Setting Up Identity Management for EMC Documentum Content Server

Setting up identity management requires administration steps in both Oracle SES and EMC Documentum. It includes the following steps:

- Activating the Documentum Identity Plug-in
- Activating the Oracle Internet Directory Identity Plug-In
- Activating the AD Identity Plug-In
- Activating SunOne Identity Plug-In

Activating the Documentum Identity Plug-in

To activate the Documentum identity plug-in, perform the following steps:

1. Select **Documentum Identity Plug-in**.
2. Click **Activate**.
3. Enter a valid DocBase name.
4. Enter a valid user name and password.
5. Ensure that the environment variable DOCUMENTUM and DOCUMENTUM_SHARED are set correctly.
6. Click **Finish**.

Activating the Oracle Internet Directory Identity Plug-In

Before activating the Oracle Internet Directory Identity plug-in, Documentum Content Server should be synchronized with Oracle Internet Directory as an LDAP server. For synchronization, you must import the users and groups from Oracle Internet Directory to Documentum.

To synchronize users and groups in Oracle Internet Directory and Documentum Content Server:

1. Create an LDAP Configuration Object in Documentum Administrator (DA):
   a. Login to DA.
   b. Navigate to **Administration**, **User Management**, **LDAP**.
   c. In the File Menu, select **File**, **New**, **LDAP Configuration Object**.
   d. In the Name field, enter a name for LDAP Configuration Object.
   e. Select **dm_user** as the user subtype.
   f. Under Communication Mode, select **Regular**.
   g. Under Import, select **Users** and **Groups**.
   h. Select **Default Configuration Object** to use this configuration object in the server field.
   i. Click **Next**.
   j. In the Directory Type field, select **Oracle Internet Directory Server**.
   k. In the Bind Type field, select **Bind by Searching for Distinguished Name**.
   l. In the Binding Name field, provide the administrative user name of Oracle Internet Directory. This is usually `cn=orcladmin`.
   m. In the Binding Password field, provide the administrative user password.
   n. In the Host Name field, provide the Oracle Internet Directory host name.
   o. Retain the default port number of Oracle Internet Directory (`389`).
   p. In the Person Object Class field, provide the information of Base Person Object, typically the value is `inetOrgPerson`.
   q. In the Person Search Base field, provide the person search base defined in Oracle Internet Directory. For example, `cn=Users`, `dc=us`, `dc=oracle`, `dc=com`.
   r. In the Person Search Filter field, specify `cn=*`.
   s. In the Group Object Class field, provide the Group Object. Typically the value is `groupOfUniqueNames`.
   t. In the Group Search Filter field, specify `cn=*`.
   u. Click **Next**.
   v. The Attribute Map information is displayed. Click **Finish**.
2. Run the LDAP_Synchronization job:
   a. Login to DA.
   b. Navigate to **Administration**, **Job Management**, **Jobs**.
   c. Open the job **dm_LDAPsynchronization**.

d. In the state field, select **Active**.

e. Select **Deactivate On Failure**.

f. In Designated Server, select the host name of Documentum Server.

g. Select **Run After Update**.

h. Click the **Schedule** tab.

i. In the **Start Date And Time** field, set the current date and time.

j. Select **Repeat time** from the Repeat list.

k. Set the Frequency field to any numeric value.

l. Select **End Date And Time** and specify how long the Synchronization job should run.

m. Click the `Method` tab.

n. Select **Pass Standard Argument**.

o. Click the **SysObject info** tab.

p. Click **OK**.

After synchronizing the Documentum Content Server with Oracle Internet Directory, you must activate the Oracle Internet Directory activity plug-in in Oracle SES.

To activate the Oracle Internet Directory Activity Plug-in:

1. Log in to Oracle SES as the admin user.
2. Click **Global Settings**.
3. Select **System**, **Identity Management Setup**.
4. Select **Oracle Internet Directory identity plug-in manager** and click **Activate**.
5. Select `nickname` from the **Authentication Attribute** list.
6. Provide the following values:
   - **Host name**: The host name of the computer where Oracle Internet Directory is running.
   - **Port**: The default LDAP port number, `389`.
   - **Use SSL**: `true` or `false` based on your preference.
   - **Realm**: The Oracle Internet Directory realm, for example, **dc=us.dc=oracle.dc=com**
   - **User name**: The Oracle Internet Directory administrative user name, for example, **cn=orcladmin**.
   - **Password**: Administrative password

Activating the AD Identity Plug-In

Before activating the AD Identity plug-in for validating the users in AD, Documentum Content Server must be synchronized with AD as an LDAP server. For synchronization, you must import users and groups from AD to Documentum.

To configure Documentum Content Server as an LDAP server:

1. Create an LDAP Configuration Object in DA:

   a. Log in to DA.

b. Navigate to **Administration**, **User Management**, **LDAP**.

c. Select **File**, **New**, **LDAP Configuration Object**.

d. Enter a name for ldap configuration object.

e. Select **dm_user** as User Subtype.

f. In the Communication Mode field, select **Regular**.

g. In the Import field, select **Users** and **Groups**.

h. Select **Default Configuration Object** in the server field, and click **Next**.

i. Provide the following values:

Directory Type: Select **Active Directory Server**.

Bind Type: Select **Bind by Searching for Distinguished Name**

Binding Name: Provide the admin user name of AD. It is normally **domainName/Administrator**.

Binding Password: The password of the AD admin user.

Host Name: AD host name.

Port: Default port number of AD, `389`.

Person Object Class: The Base Person Object, typically the value is `user`.

Person Search Base: The person search base defined in AD, for example `cn=Users,dc=us, dc=oracle,dc=com`.

Person Search Filter: Enter `cn=*`.

Group Object Class: The group object. Typically the value is `group`.

Group Search Base: The group search base defined in AD. For example, `dc=us,dc=oracle,dc=com`.

Group Search Filter: Enter `cn=*`.

j. Click **Next**.

k. The Attribute Map information is displayed. Click **Finish**.

2. Run the LDAP_Synchronization job:

a. Login to DA.

b. Navigate to **Administration**, **Job Management**, **Jobs**.

c. Open the job **dm_LDAPsynchronization**.

d. In the state field, select **Active**.

e. Select **Deactivate On Failure**.

f. In Designated Server, select the host name of Documentum Server.

g. Select **Run After Update**.

h. Click the **Schedule** tab.

i. In the **Start Date And Time** field, set the current date and time.

j. Select **Repeat time** from the Repeat list.

k. Set the Frequency field to any numeric value.

l. Select **End Date And Time** and specify how long the Synchronization job should run.

m. Click the `Method` tab.

n. Select **Pass Standard Argument**.

o. Click the **SysObject info** tab.

p. Click **OK**.

After synchronizing the Documentum Content Server with the AD, you must activate the identity for AD Identity plug-in.

To activate the identity plug-in:

1. Log in to Oracle SES as admin user.
2. Click **Global Settings**, and then select **System**, **Identity Management Setup**.
3. Select **Activity Directory Identity Plug-in Manager**, and click **Activate**.
4. Provide the following values:
   - **Authentication Attribute**: Select `USER_NAME`.
   - **Directory URL**: Provide the host name and the port number. For example, `ldap://ldapserverhost:port`.
   - **Directory account name**: Provide the AD user name, for example `Administrator`.
   - **Directory account password**: AD user password.
   - **Directory subscriber**: Provide the directory subscriber (ldap base). For example, `dc=us,dc=oracle,dc=com`.
   - **Directory security protocol**: Specify either `none` or `portnumber`.
5. Click **Finish**.

Activating SunOne Identity Plug-In

Before activating the SunOne Identity plug-in for validating the users in SunOne, you must synchronize Documentum Content Server with SunOne as an LDAP server. For synchronization, you must import the users and groups from Oracle Internet Directory to Documentum Content Server.

To import users and groups from Oracle Internet Directory:

1. Create an LDAP Configuration Object in DA:
   a. Log in to DA.
   b. Navigate to **Administration**, **User Management**, **LDAP**.
   c. Select **File**, **New**, **LDAP Configuration Object**.
   d. Enter a name for ldap configuration object.
   e. Select **dm_user** as User Subtype.
   f. In the Communication Mode field, select **Regular**.
   g. In the Import field, select **Users** and **Groups**.
   h. Select **Default Configuration Object** in the server field, and click **Next**.
   i. Provide the following values:

Directory Type: Select **Netscape/iPlanet Directory Server**

Bind Type: Select **Bind by Searching for Distinguished Name**

Binding Name: Provide the admin user name of SunOne. It is normally **cn=Administrator**.

Binding Password: The password of the SunOne admin user.

Host Name: SunOne host name.

Port: Enter the port number used for SunOne. The default port number of SunOne is `389`.

Person Object Class: The Base Person Object, typically the value is `person`.

Person Search Base: The person search base defined in SunOne, for example `cn=Users,dc=us, dc=oracle,dc=com`.

Person Search Filter: Enter `cn=*`.

Group Object Class: The group object. Typically the value is `groupOfUniqueNames`.

Group Search Base: The group search base defined in AD. For example, `dc=us,dc=oracle,dc=com`.

Group Search Filter: Enter `cn=*`.

j. Click **Next**.

k. The Attribute Map information is displayed. Click **Finish**.

2. Run the LDAP_Synchronization job:

a. Login to DA.

b. Navigate to **Administration**, **Job Management**, **Jobs**.

c. Open the job **dm_LDAPsynchronization**.

d. In the state field, select **Active**.

e. Select **Deactivate On Failure**.

f. In Designated Server, select the host name of Documentum Server.

g. Select **Run After Update**.

h. Click the **Schedule** tab.

i. In the **Start Date And Time** field, set the current date and time.

j. Select **Repeat time** from the Repeat list.

k. Set the Frequency field to any numeric value.

l. Select **End Date And Time** and specify how long the Synchronization job should run.

m. Click the `Method` tab.

n. Select **Pass Standard Argument**.

o. Click the **SysObject info** tab.

p. Click **OK**.

After the Documentum Content Server is synchronized with SunOne, the identity is activated for SunOne Identity plug-in.

To activate the identity for the SunOne plug-in:

1. Log in to Oracle SES as the administrative user.
2. Click **Global Settings**, and then select **System**, **Identity Management Setup**.
3. Select **Sun Java System Directory Server Manager**, and click **Activate**.
4. Provide the following values:
   - **Authentication Attribute**: Select `USER_NAME`.
   - **Directory URL**: Provide the host name and the port number. For example, `ldap://ldapserverhost:port`.
   - **Directory account name**: Provide the Directory Server user name, for example `Administrator`.
   - **Directory account password**: Directory Server user password.
   - **Directory subscriber**: Provide the directory subscriber (ldap base). For example, `dc=us,dc=oracle,dc=com`.
   - **Directory security protocol**: Specify either `none` or `portnumber`.
5. Click **Finish**.

Creating an EMC Documentum Content Server Source

Create an EMC Documentum Content Server source on the Home - Sources page. Select EMC Documentum Content Server from the Source Type list, and click **Create**. Enter values for the following parameters:

- **Container name**: The names of the containers to be crawled by Oracle SES. You can crawl an entire Documentum DocBase or a specific *`repository/cabinet/folder`*. The format is *`DocBaseName/CabinetName/FolderName/SubFolderName`*. Multiple comma-delimited container names can be entered. This parameter is case-sensitive; hence, enter the exact same cabinet name as in the Documentum repository. Required

  These are examples of container names:

  - `DocBase1`: The entire DocBase1 is crawled.
  - `DocBase2/Cabinet21`: Cabinet21 and its sub-folders within DocBase2 are crawled.
  - `DocBase2/Cabinet21/Folder11`: Folder11 and its sub-folders are crawled.
  - `DocBase1, DocBase2/Cabinet21/Folder11`: The entire DocBase1 and Folder 11 in DocBase2/Cabinet21 are crawled.
- **Attribute list**: The comma-delimited list of Documentum attributes along with their data types to be searchable. The format is `AttributeName:AttributeType, AttributeName:AttributeType`. Valid values are String, Number, and Date. See Table 6–3, " Documentum Data Type Mapping".

  While crawling a DocBase, an attribute is indexed only if both name and type match the configured name and type; otherwise, it is ignored. This is an optional parameter.

  For example, assume that you have the following Documentum attributes with the indicated data types

  - account name: String
  - account ID: Integer
  - creation date: Date

  To make these attributes searchable, enter this value for **Attribute list**:

  ```
  Account Name:String, Account ID:Number, Creation Date:Date
  ```

  The default searchable attributes for Documentum Content Server are Modified Date, Title, and Author.

  Multiple attributes with same name are not allowed, such as `Emp_ID:String` and `Emp_ID:Number`.

- **User name:** Enter the user name of a valid Documentum Content Server user. The user should be an administrator user or a user who has access to all cabinets, folders, and documents of the DocBases configured in the **Container name** parameter. The user should be able to retrieve content, metadata, and ACL from cabinets, folders, documents and other custom sub classes of all DocBases configured in **Container name** parameter. Required.
- **Password**: Password of the Documentum user. Required.
- **Crawl versions**: Indicate whether multiple versions of documents should be crawled, either `true` or `false`. The default value is `false`. Any other value is `false` and only the latest versions of a document are crawled. Optional.
- **Crawl folder attributes**: Indicate whether folder attributes must be crawled, either `true` or `false`. This is an optional parameter. The default value is `false`. Any other value is interpreted as `false`.
- **URL for viewing the documents**: A valid URL for Documentum WebTop or DA application used for viewing the Oracle SES search results. For example:

  `http://`*`IP_address:port`*`/da`

  or

  `http://`*`IP_address:port`*`/webtop`

- **Authentication Attribute**: This parameter is used to set ACLs. This parameter lets you set multiple LDAP servers. If Oracle SES and Documentum Content Server are synchronized with Active Directory, then enter the value `USER_NAME`. If Oracle Internet Directory is used, then enter `nickname`.

Table 6–3 Documentum Data Type Mapping

Sr. No	Documentum Data Type	Oracle SES Data Type
1	Boolean	Number
2	Integer	Number
3	String	String
4	ID	String
5	Time or Date	Date
6	Double	Number

Setting Up Microsoft SharePoint Sources

The SharePoint Crawler connector enables Oracle SES to provide secure search over SharePoint Portal Server and Microsoft Office SharePoint Server 2007 (MOSS). The

connector extends the searching capabilities of Oracle SES and enables it to search into an external repository. Oracle SES can crawl through the documents, items, and related metadata in SharePoint repositories and provide secure, full-text search. The connector also provides metadata search and browse functionality, which allows a search to be done against a specific subfolder in the hierarchy.

In SharePoint, data is stored in different libraries such as the Document Library, Picture Library, Lists, Discussion Boards, and so on. A SharePoint instance can have one or more sites and sub-sites that the SharePoint Crawler connector can crawl after you set up the appropriate configuration parameters in the Oracle SES Administration GUI. The SharePoint Crawler connector navigates through the Libraries and Lists to crawl all the documents and items from a SharePoint repository. It creates an index, stores the metadata, and accesses information in Oracle SES to provide search capabilities according to the end user permissions.

The SharePoint Crawler connector supports incremental crawling, which means that it crawls and indexes only those documents that have changed after the most recent crawl. A document is re-crawled if the content, metadata, or direct security access information of the document has changed since the previous crawl. Documents deleted from a Library are removed from the index during incremental crawling.

Important Notes About SharePoint 2007 Sources

- When the **Crawl Security Settings** parameter is set to either `NORMAL` or `STRICT`, the SharePoint Crawler for the Container must use the SharePoint administrator account for crawling and indexing documents.
- When the **Crawl Security Settings** parameter is set to `RELAX`, any user that has at least Visitor (Read) permissions can be identified in the SharePoint source for crawling and indexing documents.
- The supported versions of SharePoint Server are:
  - 2003 or 2.0 for SharePoint Portal Server
  - 2007 or 3.0 for MOSS 2007
- SharePoint Container names in Oracle SES should not contain any special characters. Enter a backslash (\) before a slash or a comma. Otherwise, the crawler does not recognize the Container.

Known Limitations of the SharePoint 2007 Connector

- Passwords entered through the Oracle SES Administration GUI are case insensitive.
- Storing more than 200 files in a single folder may result in degraded performance and increased crawling time.
- An administrator must own the SharePoint Server site with the documents to be crawled. The crawler does not have sufficient access rights to crawl the documents if it uses the identity of a non-administrative user.

  To grant administrative rights to a SharePoint user:

  1. Open the SharePoint Site UI and select **Site Settings**.
  2. Select **Users and Permissions**, then **Site Collection Administrators** to display the Site Collection Administrators page.
  3. Enter the user name of the SharePoint Server site in the Site Collection Administrators field.

4. Click **OK**.

- If the Crawler Security Settings parameter is set to `RELAX`, then the user ID specified in the User Name parameter does not require administrative privileges. Visitor (Read) permissions on the site are sufficient. However, Read must have Browse Directories permissions to access any sub-sites. Otherwise, the sub-sites are not crawled.

  To add Browse Directories permissions for SharePoint 2007:

  1. Open People and Groups - Site Permissions.
  2. Under Settings - Permission Levels, select **READ**.
  3. Under Site Permissions, select **Browse Directories**.
  4. Click **Submit**.

  To add Browse Directories permissions for SharePoint 2003:

  1. Open the Created subarea and select **Manage Security**.
  2. Select the user and edit permissions.
  3. Select READ.
  4. Click **Advanced Permissions**.
  5. Under Advanced Permissions, select **Browse Directories**.
  6. Click **OK**.

- SharePoint does not allow users without administrative privileges to browse user profiles.

  If the user ID specified in the User Name parameter does not have administrative privileges, then this user needs permission to manage profiles.

  To grant permission to manage profiles:

  1. Open SharePoint Central Administration 3.02.
  2. Click **Shared Services Administration - SharedServices1**.
  3. Under **User Profile** and **My Sites**, select **Personalization Service Permissions**.
  4. Add user *user1* and select permissions **Manage user profiles**.
  5. Save and submit the user.

  User profiles are crawled if the user has specified the root site in the Site/Sub-Site URL parameter of the source configuration.

Known Issues for SharePoint 2007 Connector

- Versions of list items whose object type is folder are not getting crawled and indexed.
- Site Collection Administrator users are not able to see documents if they are not listed among the document permission users.
- `Unable to type cast null` message is not error. This information is provided when the crawler tries to crawl attachments that are not supported for a particular entity.
- `Principal` *`user_name`* `cannot be validated` error is returned when the crawler obtains a user name from the SharePoint repository that is not present in the Active Directory.

- Performance of the SharePoint connector can be impacted when the Crawl Versions attribute is set to true.

Supported Platforms

The following platforms are supported by the SharePoint Crawler connector:

- Red Hat Linux 4
- Windows 2003 Server Standard Edition and above with the latest Service Pack

Creating a SharePoint 2007 Source

Create a source for the newly-created user-defined source type on the Home - Sources page. Enter a source name. Provide values for the configuration parameters described in the following list. Also see Table 6–4, " Supported Values for SharePoint Source Parameters".

- **SharePoint Version**: Version of the SharePoint server (SharePoint Portal Server/MOSS 2007) to crawl. (Required)
- **Container name**: Contains the names of the containers to be crawled by Oracle SES. You can specify multiple container names as a comma-delimited list. (Required)

  You can crawl an entire area or site or a specific folder. The format for specifying a container folder is *AreaName/LibraryName/FolderName/SubFolderName*.

  To crawl all documents in the Area or Library, the format is *AreaName* or *AreaName/LibraryName*.

  To index the entire SharePoint portal, enter a slash (/).

  To crawl all sites, enter `sites`.

  Examples for SharePoint Portal Server:

  – Container name: *AreaName*

    The entire Area is crawled.

  – Container name: *AreaName/LibraryName/Folder21*

    Folder21 and its subfolders within *LibraryName* are crawled.

  – Container name: *LibraryName*

    All documents inside the Library and its subfolders are crawled.

  Examples for MOSS 2007:

  – Container name: *LibraryName/Folder21*

    Folder21 and its sub-folders within *LibraryName* are crawled.

  – Container name: *LibraryName*

    All documents inside the Library and its subfolders are crawled.

    The path for the container cannot contain any special characters. Enter a backslash (\) before a slash or a comma.

- **Attribute list**: A comma-delimited list of attributes, as described in Table 6–5. The format for an attribute list is *AttributeName*, *AttributeName*. Multiple attributes with same name are not allowed, such as `Emp_ID, Emp_ID`.

In MOSS 2007, all attributes viewable from the UI are indexed by default. List all custom attributes to index, using the names displayed in the user interface.

In SPPS (SP 2003), the Title, LastModifiedDate, and Author attributes are indexed by default. List any other attributes to index, using the names displayed in the UI.

If you update the attribute list from the administrator parameters, then perform a forced recrawl to delete the indexes of the old attribute list and to create indexes for the new attribute list.

- **Domain name**: The domain name of the user that is used to crawl the SharePoint site. For example, if you intend to use the `OracleDomain\Administrator` user for crawling, then enter `OracleDomain` for this parameter. Do not include `.com` or `.in` or any other suffix in the name. (Required)
- **User name**: Specifies the user name of a valid SharePoint Portal Server/MOSS 2007 user. Do not include the domain name for this user. For example, for `OracleDomain\Administrator`, enter `Administrator`. (Required)
- **Password**: Specifies the password of the SharePoint user specified in User name. (Required)
- **Authentication attribute**: Format of the user and group identity stored in the ACL of SharePoint objects. This format must be an authentication attribute of the Oracle SES active identity plug-in, such as `USER_NAME` for an Active Directory identity plug-in. Otherwise, the ACL validation fails during indexing. (Required and case sensitive)

  For example, this value is `USER_NAME` for the Microsoft Active Directory identity plug-in.
- **SPS Site/Sub-Site URL**: The URL of the Site or Sub-site of the SharePoint Portal, which is used for viewing the search results. (Required)

  This URL has the form `http://`*`HostName:PortNumber`* or `http://`*`HostName:PortNumber/SubSiteName`*.
- **Crawl Security Settings**: Sets security on documents for indexing. (Required)

  This setting can be one of the following:

  – `NORMAL`: The regular crawl uses site-level access control lists (ACLs) but not document-level ACLs.

  – `RELAX`: When the SharePoint Site Administrator user information is not available and the SharePoint user has visitor (or read) permissions on the site, this user is not able to crawl subsites under the main site. This mode is intended for exposing public documents temporarily and quickly to search. The SES administrator must be careful not to expose documents to other users inadvertently. See the work-around for this in "Known Limitations of the SharePoint 2007 Connector" on page 6-12.

  – `STRICT`: Captures even document-level security. This mode requires that an additional Web Service agent, Oracle MOSS Web Service, be installed on the SharePoint 2007 server. See "Deploying the Web Service on MOSS 2007" on page 6-18.
- **Simple Include**: Only include URLs having at least one word mentioned in this parameter. Separate the words with commas.
- **Simple Exclude**: Exclude all URLs having one or more word(s) mentioned in this parameter. Separate the words with commas.

- **Regular Expression Include**: Include all URLs that match the expression provided in this parameter.
- **Regular Expression Exclude**: Exclude all URLs that match the expression provided in this parameter.
- **Crawl versions**: Controls whether multiple versions of documents are crawled. Valid values are `true` and `false`. Any other value is interpreted as `false`. The default value is `false`, so only the latest version is crawled. (Optional)
- **Crawl folder attributes**: Controls whether folder attributes are crawled. The default value is `false`. Valid values are `true` or `false`, and any other value is interpreted as `false`. (Optional)
- **Crawl attachments**: This parameter indicates whether attachments should be crawled. The default value is `false`. Valid values are `true` or `false`, and any other value is interpreted as `false`. (Optional)
- **LDAP URL**: URL of the LDAP server, such as `ldap://IP:port`, where the default port number is 389.
- **LDAP Search Base**: LDAP Search Base, such as, `DC=abc`, `DC=com`. When the value of **Authentication Attribute** is DN, specify the LDAP URL and the LDAP search base of the LDAP server configured in the identity plug-in. Otherwise, leave these parameters blank.

Table 6–4 summarizes the supported values for the configuration parameters of the SharePoint Crawler connector.

Table 6–4 Supported Values for SharePoint Source Parameters

Parameter Name	SharePoint Portal Server	MOSS 2007
SharePoint Version	2003, 2.0	2007, 3.0
Container name	(/) for full site, Library Name, List Name, Area Name	(/) for full site, Library Name, List Name
Attribute list	*AttributeName1*, *AttributeName2*	*AttributeName1*, *AttributeName2*
Domain Name	Domain name of the user	Domain name of the user
User name	Valid administrator user for SharePoint Portal server	Valid administrator user for MOSS 2007
Password	Password for the user	Password for the user
Authentication attributes	`USER_NAME`	`USER_NAME`
SPC Site/Sub-Site URL	IP address or host name with port on which SharePoint Portal Server is installed	IP address or host name with port on which MOSS 2007 is installed
Crawl Security Settings	`NORMAL`, `RELAX`	`NORMAL`, `RELAX`, `STRICT`
Simple Include	Part of URL	Part of URL
Simple Exclude	Part of URL	Part of URL
Regular Expression Include	All URLs that match the expression	All URLs that match the expression
Regular Expression Exclude	All URLs that match the expression	All URLs that match the expression
Crawl versions	`true` or `false`	`true` or `false`

Table 6–4 (Cont.) Supported Values for SharePoint Source Parameters

Parameter Name	SharePoint Portal Server	MOSS 2007
Crawl folder attachments	`true` or `false`	`true` or `false`
Crawl attachments	`true` or `false`	`true` or `false`
LDAP URL	URL of the LDAP server	URL of the LDAP server
LDAP Search Base	LDAP Search Base	LDAP Search Base

Table 6–5 Attributes for List Items and Versions Crawled for SharePoint 2007

List Item Type	Attributes
Document Library	Title, Author, Created, Modified
Picture Library	Title, ImageSize, ImageCreateDate, Description, Keywords
Form Library	Title, Author, Created, Modified
Translation Library	Title, Name, Language, Base Document Version, Translation Status, Created
Data Connection Library	Connection Type, Description, Keywords, Title, UDC Purpose, Created
Slide Library	Name, Presentation, Description, Created
Report Library	Name, Title, Author, Created, Report Category, Report Status
Dash Board	Name, Title, Author, Created
Wiki Page Library	Title, Author, Created, Modified
Announcements	Title, Body, Editor, Modified, Author, Created
Contacts	Company, WorkCity, Created, Email, Comments, Title, Editor, HomePhone, JobTitle, Modified, WorkZip, WorkPhone, WorkState, FirstName, Author, FullName, WorkCountry, CellPhone, WorkFax, WorkAddress
Links	Comments, Editor, Modified, Author, URL, Created
Discussion Reply	Body, Created, DiscussionTitle, Editor, Modified, Author
Calendar	EventType, Title, EventDate, Duration, Editor, WorkspaceLink, Modified, EndDate, Description, fRecurrence, Author, fAllDayEvent, Created
Task	Title, StartDate, Body, Status, Editor, Priority, AssignedTo, DueDate,Modified, Author, PercentComplete, Created
Project Task	Title, StartDate, Body, Status, Editor, Priority, AssignedTo, DueDate,Modified, Author, PercentComplete, Created
Issue Tracking	Category, LinkIssueIDNoMenu, RelatedIssues, IssueID, Priority, DueData, Comment, V3Comments, IsCurrent, Created, Title, Status, Editor, AssignedTo, Modified, Author
Custom List	Title, Editor, Modified, Author, Created
Languages and Translators	Language_x0020_From,Language_x0020_To,Modified,Author,Translator,Created, Editor
KPI List	Title, PercentExpression, Editor, ViewGuid, Modified, Value, AutoUpdate, KpiComments, Author, Goal, ValueExpression, Warning, KpiDescription, DataSource, LowerValuesAreBetter, Created

Deploying the Web Service on MOSS 2007

For MOSS 2007, if the **Crawl Security Settings** parameter is set to `STRICT`, then you must install an extra web service, Oracle MOSS Web Service. The following installation and deinstallation files are provided by the OracleMOSSService installer at *ORACLE_HOME*`/search/lib/plugins/sps/WebService.zip`:

- OracleMossService.wsp
- install.cmd
- de-install.cmd

To install or deinstall the Oracle MOSS Web Service:

1. Click install.cmd to install, or click de-install.cmd to deinstall.
2. Verify that the STSADM.exe file is in the following location: *Drive*:\Program Files\Common Files\Microsoft Shared\web server extensions\12\BIN.

   If STSADM.exe is not in that folder, specify the correct path when the installer prompts for it.
3. Press any key to continue.

Setting Up Oracle Content Database Sources

Documents in Oracle Content Database are organized into **folders**. Oracle SES navigates the folder hierarchy to crawl all documents in Oracle Content Database. It creates an index, stores the metadata, and accesses information in Oracle SES to provide search according to the end users' permissions.

The metadata crawled includes *`folder_url`* (URL of the folder containing the document) and *`folder_path`* (path of the folder containing the document). These let you show the direct folder path and direct folder URL for each document hit.

Oracle SES supports incremental crawling; that is, it only crawls and indexes documents that have changed since the last crawling. A document is re-crawled if either the content or the direct security access information of the document changes. A document is also re-crawled if it is moved within Oracle Content Database and the end user has to access the same document with a different URL. Deleted documents are removed from the index during incremental crawling.

Important Notes for Oracle Content Database Sources

This book uses the product name Oracle Content Database to mean *both* Oracle Content Database *and* Oracle Content Services. Oracle Content Database sources are certified with Oracle Content Database release 10.2 and release 10.1.3 and Oracle Content Services release 10.1.2.3.

Known Issues:

- The administrator account used by the Oracle Content Database source must have the **ContentAdministrator** role on the site that is being crawled and indexed. Also, end users searching documents in Oracle Content Database must have the **GetContent** and **GetMetadata** permissions.
- By default, Oracle Content Database has a limit of three concurrent requests (simultaneous operations) for each user. However, Oracle SES has a default of five concurrent crawler threads. When crawling Oracle Content Database, only three of the five threads can successfully crawl, which causes the crawl to fail.

Workaround: For an Oracle Content Database source, change the **Number of Crawler Threads** on the Home - Sources - Crawling Parameters page to a value of 3 or fewer.

Or, modify the Oracle Collaboration Suite configuration in Oracle Enterprise Manager to allow more than three concurrent requests. For example:

1. Access the Enterprise Manager page for the Collaboration Suite Midtier. For example: `http://example.domain:1156/`.
2. Click the Oracle Collaboration Suite midtier standalone instance name. For example: `ocsapps.example.domain`.
3. In the **System Components** table, click **Content**.
4. From **Administration**, click **Node Configurations**.
5. In the **Node Configurations** table, click **HTTP_Node**. For example: ocsapps.computer.domain_HTTP_Node.
6. On **Properties**, change the value for **Maximum Concurrent Requests Per User**. Enter a value larger than or equal to the number of crawling threads used by Oracle SES. This value is listed on the Global Settings - Crawler Configuration page.

Setting Up Identity Management for Oracle Content Database Sources

The Oracle SES instance and the Oracle Content Database instance must be connected to the same or mirrored Oracle Internet Directory system or other LDAP server.

To set up a secure Oracle Content Database source:

1. Read "Known Issues:" on page 6-18 and confirm that the number of crawler threads does not exceed the available concurrent connection settings for each user in Oracle Content Database.
2. Activate the Oracle Internet Directory identity plug-in for the Oracle Content Database instance on the Global Settings - Identity Management Setup page in Oracle SES.
3. For Oracle Content Database 10.1.2.3 and 10.2.0.4, use the following LDIF file to create an *application entity* for the plug-in. (An application entity is a data structure within LDAP used to represent and keep track of software applications accessing the directory with an LDAP client.)

   ```
   ORACLE_HOME/bin/ldapmodify -h oidHost -p OIDPortNumber -D "cn=orcladmin" -w
   password -f  ORACLE_HOME/search/config/ldif/csPlugin.ldif
   ```

   This defines the entity that is used for the connector: `orclApplicationCommonName=ocsCsPlugin,cn=ifs,cn=products,cn=oraclecontext`. The entity has the password `welcome1`.

Creating an Oracle Content Database JDBC Source

The Content Database JDBC connector is an alternative to the Content Database connector provided in Oracle SES Release 10.1. The JDBC connector greatly improves the performance of incremental crawls. If the elapsed time of an incremental crawl is an important consideration in your deployment of Oracle SES, then use the JDBC connector.

Oracle SES crawler supports crawling from Oracle Content Database 10.1.2.0.4 or later. See the readme file for Oracle Content Database 10.2.1.0.4 patchset for details on configuring high volume full and incremental crawls in Oracle Content Database.

You may need to grant the SES user access to an Oracle Content Database object. Use this command:

```
GRANT SELECT ON ODMC_ALERT_SEQ TO sesuser
```

where `sesuser` is the SES user.

For example,

```
GRANT SELECT ON ODMC_ALERT_SEQ TO SEARCHSYS
```

> **Note:** The JDBC connector requires installation of a patch to Oracle Content Database. If the patch is not available for your version of Content Database, then use the older connector as described in "Creating an Oracle Content Database Source" on page 6-21.

To create an Oracle Content Database JDBC source:

1. Open the Oracle SES Administration GUI to the Home page.
2. Select the Sources secondary tab.
3. For Source Type, select **Oracle Content Database (JDBC)**, then click **Create** to display Step 1 Parameters.
4. Enter a source name and the values for the parameters described in Table 6–6.
5. Click **Next** to display Step 2 Authorization.
6. Enter the settings described in Table 6–7.
7. Click **Create** or **Create and Customize** to create the source.

Table 6–6 Oracle Content Database JDBC Source Parameters (Step 1)

Parameter	Value
Database Connection String	JDBC connection string to Oracle Content Database in the form `jdbc:oracle:thin@server:port:sid`. For example, `jdbc:oracle:thin@example.com:1521:rel11g`
Content DB System User	SYSTEM user for Content Database.
Alert Table Name	Name of the Alert table for Content Database, which typically has the form `ODMC_ALERT_name`.
Database User ID for Crawl	Valid user ID for the Content DB database.
Database Password for Crawl	Password associated with the user ID for crawling.
Document Count	Maximum number of documents to be crawled.
URL Prefix	URL to Oracle Content Database in the form `HTTP://hostname:port/CONTENT`. For example, `HTTP://example.com:7778/CONTENT`.
Document Access (DAV) User ID	Valid Content Database user ID for using WebDAV to access documents.
Document Access (DAV) Password	Password associated with the DAV user ID.

Table 6–6 (Cont.) Oracle Content Database JDBC Source Parameters (Step 1)

Parameter	Value
Starting Path for Crawl	Full path where the crawl starts. Enter / to crawl the entire Content Database hierarchy.

Table 6–7 Oracle Content Database JDBC Authorization Parameters (Step 2)

Parameter	Value
Authorization Database JDBC Connection String	JDBC connection string to Oracle Content Database in the form `jdbc:oracle:thin@server:port:sid`. For example, `jdbc:oracle:thin@example.com:1521:rel11g`
Content DB System User	System user for Content Database, such as `CONTENT` or `IFS_SYS`.
Database User ID	User ID to connect to the database.
Database Password	Password associated with the database user ID.
Use the Run-Time Result Filter	Controls use of a final security check: `TRUE`: Performs a final security check on each row in the result set. `FALSE`: Does not do a final check. (Default)
Authorization User ID Format	Format of user ID in the authorization query. Enter a supported authentication attributes of the active ID plugin, such as `nickname`.

Creating an Oracle Content Database Source

If Oracle Content Database release 10.2 or Oracle Content Services release 10.1.2 is used, then the **Entity name** and **Entity password** parameters are required, the last six parameters related with keystore are not required, and the crawler plug-in uses service to service (S2S) authentication to connect to Oracle Content Database.

If Oracle Content Database release 10.1.3 is used, then the last six parameters in the following table are required, the **Entity name** and **Entity password** are not required, and Oracle SES uses Web services authentication to connect to Oracle Content Database. See "Required Tasks for Oracle Content Database Release 10.1.3" on page 6-22.

Create an Oracle Content Database source on the Home - Sources page. Select **Oracle Content Database** from the Source Type list, and click **Create**.

Enter values for the parameters listed in Table 6–8.

Table 6–8 Oracle Content Database Source Parameters

Parameter	Value
Oracle Content Database URL	`http://host name:port/content`
Starting paths	/
Depth	-1
Oracle Content Database admin user	`orcladmin`
Entity name	`orclApplicationCommonName=ocsCsPlugin,cn=ifs,cn=products,cn=oraclecontext`
Entity password	welcome1

Table 6–8 (Cont.) Oracle Content Database Source Parameters

Parameter	Value
Crawl only	`false`
Use e-mail for authorization	`false`
Oracle Content Database Version	For example, 10.1.3.2.0
SES keystore location	For example, /scratch/ocs/cdb/cdb-ses/keystore/sesClientKeystore.jks
SES keystore type	jks
SES keystore password	*******
SES private key alias	client
SES private key password	*******
CDB Server public key alias	server

Table 6–9 Oracle Content Database Authorization Manager Plug-in Parameters

Parameter	Value
Oracle Content Database URL	http://*host name*:*port*/content
Oracle Content Database admin user	orcladmin
Entity name	`orclApplicationCommonName=ocsCsPlugin,cn=ifs,cn=products,cn=oraclecontext`
Entity password	welcome1
Use e-mail for authorization	`false`
Use result filter for authorization	`false` You can use a real-time result filter (query-time authorization) to ensure that the user has access to each result document. Set this parameter to `true` to remove documents that the user has lost access to since the last crawl.
Oracle Content Database Version	For example, 10.1.3.2.0
SES keystore location	For example, /scratch/ocs/cdb/cdb-ses/keystore/sesClientKeystore.jks
SES keystore type	jks
SES keystore password	********
SES private key alias	client
SES private key password	*******
CDB Server public key alias	server

Required Tasks for Oracle Content Database Release 10.1.3

This section describes the required steps for Web services authentication when using Oracle Content Database release 10.1.3. This procedure uses the JDK keytool to create the keys.

See Also: "Setting Up a Server Keystore for WS-Security" in the *Oracle Fusion Middleware Administrator's Guide for Oracle Universal Online Archive* at http://download.oracle.com/docs/cd/B32110_01/content.1013/b32191/security.htm#CHDGCJEH

1. Configure a server keystore at the Oracle Content Database middle tier if the keystore is not set up yet.

   The file *ORACLE_HOME*/j2ee/OC4J_Content/config/oc4j.properties defines the keystore type and the keystore properties file location. If you use a different file name for the keystore, then edit the file on the following entry:

   ```
   oracle.ifs.security.KeyStoreLocation=/home/oracle/product/10.1.3.2.0/OracleAS_1/content/settings/server-keystore.jks
   ```

   a. Change to the settings directory:

      ```
      cd ORACLE_HOME/content/settings
      ```

   b. Create the Oracle Content Database server keystore with the following keytool command, substituting a secure password for *password*.

      ```
      ORACLE_HOME/jdk/bin/keytool -genkey -keyalg RSA -validity 5000
      -alias server -keystore server-keystore.jks -dname "cn=server" -keypass
      password -storepass password
      ```

      To list the keys in the store:

      ```
      ORACLE_HOME/jdk/bin/keytool -list -keystore server-keystore.jks
      -keypass password -storepass password
      ```

   c. Sign the key before using it:

      ```
      ORACLE_HOME/jdk/bin/keytool -selfcert -validity 5000 -alias server
      -keystore server-keystore.jks -keypass password -storepass password
      ```

   d. Export the server public key from the server keystore to a file:

      ```
      ORACLE_HOME/jdk/bin/keytool -export -alias server -keystore
      server-keystore.jks -file cdbServer.pubkey -keypass password -storepass
      password
      ```

   e. Store both the keystore password and the private server key password in a secure location so Oracle Content Database can access the keystore and the private key.

      ```
      ORACLE_HOME/content/bin/changepassword -k
      ```

      When prompted for the old password, press [Enter] if it is the first time to set the password; otherwise, enter the previous password. Then, enter and confirm the keystore password (`-storepass password`) that you provided in step 1.b.

      See *ORACLE_HOME*/content/log/changepassword.log.

2. Configure a client keystore at the Oracle SES installation.

   a. Create the SES client keystore with the following keytool command, substituting a secure password for *password*:

      ```
      ORACLE_HOME/jdk/bin/keytool -genkey -keyalg RSA -validity 5000
      -alias client -keystore sesClientKeystore.jks -dname "cn=client"
      ```

```
-keypass password -storepass password
```

To list the keys in store:

```
ORACLE_HOME/jdk/bin/keytool -list -keystore sesClientKeystore.jks
-keypass password -storepass password
```

b. Sign the key before using the key:

```
ORACLE_HOME/jdk/bin/keytool -selfcert -validity 5000 -alias client
-keystore sesClientKeystore.jks -keypass password -storepass password
```

Restart the WebCenter middle tier from the Oracle Enterprise Manager console.

c. Export the server public key from the server keystore to a file:

```
ORACLE_HOME/jdk/bin/keytool -export -alias client -keystore
sesClientKeystore.jks -file sesClient.pubkey -keypass password
-storepass password
```

3. Import Oracle SES client public keys into the Oracle Content Database server keystore (`sesClient.pubkey` must be copied to Oracle Content Database):

```
cd ORACLE_HOME/content/settings

ORACLE_HOME/jdk/bin/keytool -import -alias client -file
sesClient.pubkey -keystore server-keystore.jks -keypass password
-storepass password
```

4. Import Oracle Content Database server public keys into the Oracle SES keystore. (`cdbServer.pubkey` must be copied to Oracle SES):

```
ORACLE_HOME/jdk/bin/keytool -import -alias server -file
cdbServer.pubkey -keystore sesClientKeystore.jks -keypass password
-storepass password
```

Note: Check the server logs at *ORACLE_HOME*`/content/logs` for keystore issues with the crawler plug-in.

Oracle Content Database Source Attributes

Oracle SES crawls the following attributes for Oracle Content Database Sources:

- `AUTHOR`
- `CREATE_DATE`
- `DESCRIPTION`
- `FILE_NAME`
- `LASTMODIFIEDDATE`
- `LAST_MODIFIED_BY`
- `TITLE`
- `MIMETYPE`
- `ACL_CHECKSUM`: The check sum calculated over the ACL submitted for the document.

- DOCUMENT_LANGUAGE: Oracle SES language code taken from Oracle Content Database language string. For example, if Oracle Content Database uses "American", then Oracle SES submits it as "en-us".
- DOCUMENT_CHARACTER_SET: The character set for the Oracle Content Database document.

Oracle SES also can search categories or customized attributes created by the user in Oracle Content Database.

You can apply categories to files and links, and divide categories into subcategories having one or more attributes. When a document in Oracle Content Database is attached to a category, you can search on the attribute of category. (The attributes appear in the list of search attributes.)

For example, suppose you create a category named testCategory with testAttr1 and testAttr2. Document X is created and assigned to testCategory. You must assign the value to the testCategory attributes. After crawling, testAttr1 and testAttr2 appears in the search attribute list.

Customized attribute values can be the following types: String, Integer, Long, Double, Boolean, Date, User, Enumerated String, Enumerated Integer, and Enumerated Long:

- Index Long, Double, Integer, Enumerated Integer, and Enumerated Long type customized attributes are type Number attributes in Oracle SES. The display name has an _N suffix.
- Index Date customized attributes are type Date attributes in Oracle SES. The display name has a _D suffix).
- Index String, Enumerated String, and User customized attributes are type String attributes in Oracle SES.

Limitations on Custom Attributes for Oracle Content Database

- The Oracle Content Database SDK has more features than the Oracle Content Database Web GUI. The Web GUI does not support String arrays, but the SDK does. If you use the SDK to build customized administration and user GUIs that support the String array type, then a customized attribute can have multiple values.
- If a document in Oracle Content Database is attached to a category and the attributes in that category are left blank, then the attribute is not available in the attribute list for an Advanced Search. The crawler skips attributes with null values. However, if another document has the same attribute with a real value, then the attribute is indexed.

Setting Up Oracle Content Server Sources

The Oracle Content Server connector enables Oracle SES to search Oracle Content Server (formerly Stellent Server), which is the foundation of the Oracle Universal Content Management solution. Users throughout the organization can contribute content from native desktop applications, manage content through rich library services, publish content to Web sites or business applications, and access the content with a browser.

The Content Server connector supports Oracle Content Server 7.5.2 or 10gR3 with XMLCrawlerExport (the Oracle Content Server RSS component).

Oracle Content Server includes an RSS feed generator component (XMLCrawlerExport) on top of the content server. This component generates RSS

feeds as XML files from its internal indexer, based on indexer activity. It has access to the original content (for example, a Microsoft Word document), the Web viewable rendition, and all the metadata associated with each document. The component also has a template that contains a Idoc script that applies the metadata values from the indexer to generate the XML document. (Idoc is an Oracle Content Server proprietary scripting language.) Oracle Content Server generates feeds for all documents for the initial crawl, and feeds for updated and deleted documents for the incremental crawl. Each document can be an item in the feed, with the operation on the item (such as insert, delete, update), its metadata (such as author, summary), URL links, and so on.

The Oracle Content Server connector reads the feeds provided by Oracle Content Server according to a crawling schedule. Oracle SES parses and extracts the metadata information, and fetches the document content, using its generic RSS crawler framework.

Oracle SES supports the control feed method, in which individual feeds can be located anywhere and a control feed file is generated containing the links to other feeds. This control file is input to the connector through the configuration file. Control feed must be used when two computers are on different domains or on different platforms, or if they use remote access protocol, such as HTTP or FTP, for communication between the two servers.

See Also:

- "Overview of XML Connector Framework" on page 3-9
- Oracle Content Management page at `http://www.oracle.com/technetwork/middleware/content-management/overview/index.html`

Oracle Content Server Security Model

The Oracle Content Server security model is based on the concept of permissions, which defines the privileges a user has on a document. The following table shows the set of permissions supported by Oracle Content Server. Each permission is a superset of the previous ones. For example, Write permission includes Read permission. Admin permission is a superset of all the permissions.

Table 6–10 Oracle Content Server Permissions

Permission	Description
Read	View documents
Write	View, Check In, Check Out, and Get Copy of documents
Delete	View, Check In, Check Out, Get Copy, and Delete documents
Admin	View, Check In, Check Out, Get Copy, and Delete documents An Administration user with Workflow rights can start or edit a workflow for the document. An Administration user can also check in documents with another user specified as the Author.

Oracle Content Server provides multiple security models, including an out-of-the-box security system and integration with centralized security models such as LDAP and Active Directory.

Oracle Universal Content Management security can work in these modes:

- Universal Content Management native identity plugin where Universal Content Management is not connected to a directory

- Oracle Internet Directory
- Active Directory only where Universal Content Management is connected to Active Directory using LDAP. A connection to Active Directory using Microsoft Security is not supported.

The Oracle SES Oracle Content Server connector supports the two most popular security models among current Oracle Content Server customers: Roles and Groups, and Accounts.

Roles and Groups

A security group is a set of files grouped under a unique name. Every file in the library belongs to a security group. Access to security groups is controlled by the permissions, which are assigned to roles, which are assigned to users. For example, the EngAdmin role has Read, Write, Delete, and Admin permission to all content in the EngDocs security group. User Joe is assigned to role EngAdmin; therefore, Joe has all permissions to the documents in EngDocs group.

Accounts

Accounts provide greater flexibility and granularity than groups. An account is a group of content. It introduces another metadata field that is filled out upon content check-in. When accounts are enabled, content items also can be assigned to an account in addition to the security group. A user must have access to the account to read, write, delete or administer content in that account. When accounts are used, the account becomes the primary permission to satisfy before security group permissions are applied.

A user's access to a document is like the intersection between their account permissions and security group permissions. For example, a user is assigned the EngAdmin role, which has all permissions to the documents in EngDocs security group. At the same time, the user is also assigned Read and Write permission to the EngProjA account. Therefore, the user has only Read and Write permission to a content item that is in the EngDocs security group and the EngProjA account.

Accounts can also be set up in a hierarchical structure. A user has permission to the entire subtree starting from the account node. For instance, a user assigned to the Eng account has access to Eng/AbcProj and Eng/XyzProj, or any accounts beginning with Eng. In other words, users that have permission to a particular account prefix also have access to all accounts with that prefix.

Note: Oracle Content Server uses a prefix test for account filtering, so a slash (/) has no special meaning. A user granted permission to account A has access to any documents in account A*, such as A, AB, or A/B. The hierarchical structure takes advantage of the prefix semantics, but it is enforced with the account model. Hence, there is no special character as the level divider when testing for account permissions.

See Also: Oracle Universal Content Management documentation at

http://www.oracle.com/technetwork/middleware/content-management/index-094708.html

Setting Up Identity Management for Oracle Content Server

To activate the Oracle Content Server identity plug-in:

1. On the Global Settings page, select **Identity Management Setup** under the System heading.

   The Global Settings - Identity Management Setup page is displayed.

2. Select **Oracle Content Server** and click **Activate**.
3. Enter values for the parameters described in Table 6–11, then click **Finish**.

Table 6–11 Oracle Content Server Connector Setup Parameters

Parameter	Value
HTTP endpoint for authentication	HTTP endpoint for Oracle Content Server authentication. For example, `http://my.host.com:`*`port`*`/idc/idcplg`
Admin User	Administrative user who accesses the Oracle Content Server Identity Service API
Password	Administrative user password

Creating an Oracle Content Server Source

To create an Oracle Content Server source using the Oracle SES Administration GUI:

1. On the Home page, click the **Sources** secondary tab to display the Sources page.
2. Select **Oracle Content Server** from the **Source Type** list, then click **Create** to display Step 1 Parameters.
3. Enter values for the parameters described in Table 6–12.
4. Click **Next** to display Step 2 Authorization, then set values for the parameters described in Table 6–12.
5. Scroll down to Security Attributes to verify that `ACCOUNT` and `DOCSECURITYGROUP` are listed. If they are not, then the source was not created correctly. Verify that the Configuration URL in Step 1 is correct.
6. Click **Create** to create the Oracle Content Server source.

   After processing each data feed, a status feed is uploaded to the location specified in the configuration file. This status feed is named one of the following:

   - *`data_feed_file_name`*`.suc` indicates the data feed was processed successfully.
   - *`data_feed_file_name`*`.err` indicates that an error was encountered while processing the feed. The errors are listed in this status feed.

     Tip: To index multibyte character sets, set the default character set of the crawler to UTF-8 regardless of the character set of Oracle Content Server. See "Modifying the Crawler Parameters" on page 3-2.

Table 6–12 Oracle Content Server Source Parameters (Step 1)

Parameter	Value
Configuration URL	URL of the XML configuration file providing details of the source, such as the data feed type, location, security attributes, and so on. Obtain the location of the file from the Oracle Content Server administrator. Use the following format to enter the configuration URL: `http://`*`host_name/instance_name`*`/idcplg?IdcService=SES_CRAWLER_DOWNLOAD_CONFIG&source=`*`source_name`*
Authentication Type	Java authentication type. Set this parameter when the data feeds are accessed over HTTP. Enter one of the following values: ■ `NATIVE`: Proprietary XML over HTTP authentication ■ `ORASSO`: Oracle Single Sign-on.
User ID	User ID to access the data feeds. The access details of the data feed are specified in the configuration file. Obtain a user ID from the Oracle Content Server administrator.
Password	Password for **User ID**. Obtain the password from the Oracle Content Server administrator.
Realm	Realm of the Oracle Content Server instance.
Oracle SSO Login URL	URL that protects all OracleAS Single Sign-on applications. Set this parameter when the **Authentication Type** is ORASSO.
Oracle SSO Action URL	URL that authenticates OracleAS Single Sign-on user credentials. The login form is submitted to this URL. Set this parameter when **Authentication Type** is ORASSO.
Scratch Directory	Directory where Oracle SES can write temporary status logs. The directory must be on the same system where Oracle SES is installed. Optional.
Maximum number of connection attempts	Maximum number of attempts to connect to the target server for access to the data feed.

Table 6–13 Oracle Content Server Connector Authorization Parameters (Step 2)

Parameter	Value
HTTP Endpoint for Authorization	HTTP endpoint for Oracle Content Server authorization, such as `http://example.com:7777/idc/idcplg`.
Display URL Prefix	HTTP host information to prefix the partial URL specified in the access URL of the documents in RSS feeds to form the complete URL. This complete URL is displayed as the URL when a user clicks the document link in the Oracle SES search results page. For example, you might display `http://example.com:7777/idc` (*not* `http://example.com/`, as shown on the user interface page).
Administrator User	Administrative user to access the Authorization Service API of Oracle Content Server.
Administrator Password	Administrative user password.
Display Crawled Version	Controls access to the crawled documents: ■ `true`: Search results point to the crawled version of the document. ■ `false`: Search results point to the content information page.

Table 6–13 (Cont.) Oracle Content Server Connector Authorization Parameters (Step 2)

Parameter	Value
Authorization User ID Format	Format of the user ID used by the Oracle Content Server authorization API, such as `username`, `email`, `nickname`, `user_name`. When no value is specified, the canonical form of the user identity in the active identity plug-in is submitted to the authorization API.
Use Cached User and Role Information to Authorize Results	Controls user authorization: ■ `true`: Uses the cached user query filter. This setting removes the query time dependency on Oracle Content Server. ■ `false`: Queries Oracle Content Server for authorization.
User Role Data Source to Cache the Filter	The name of the Oracle Content Server Users source that has crawled the user's SecurityGroup and Account information.
Authentication Type	Java authentication type. Enter `NATIVE` for proprietary XML over HTTP authentication, or `ORASSO` for Oracle Single Sign-on. Set this parameter when the data feeds are accessed over HTTP.
Realm	Realm of the Oracle Content Server instance.
Oracle SSO Login URL	URL that protects all OracleAS Single Sign-on applications. Set this parameter when the **Authentication Type** is ORASSO.
Oracle SSO Action URL	URL that authenticates OracleAS Single Sign-on user credentials. The login form is submitted to this URL. Set this parameter when **Authentication Type** is ORASSO.

7

Configuring Access to Collaboration Sources

This chapter contains the following topics:

- Setting Up EMC Documentum eRoom Sources
- Setting Up Lotus Notes Sources
- Setting Up Microsoft Exchange Sources
- Setting Up NTFS Sources for Windows
- Setting Up NTFS Sources for UNIX
- Setting Up Oracle Calendar Sources
- Setting Up Oracle Collaboration Suite E-Mail Sources

Setting Up EMC Documentum eRoom Sources

The EMC Documentum eRoom Server plug-in extends the searching capabilities of Oracle SES and enables it to search Documentum eRoom Server repositories. Oracle SES can crawl through the documents and related metadata in the Documentum eRoom and provide secure, full-text search. It also provides metadata search and browse functionality.

Documentum eRoom data is stored in an eRoom, which in turn can contain other containers and content. A Documentum eRoom Server instance can have one or more items that can be crawled using the Documentum eRoom Server plug-in by configuring parameters in Oracle SES. The Documentum eRoom Server plug-in navigates through all the containers and the inline contents to crawl all the documents/items in Documentum eRoom Sever. It creates an index, stores the metadata, and accesses information in Oracle SES to provide search according to the end user permissions.

The Documentum eRoom Server plug-in supports incremental crawling; that is, it crawls and indexes only those documents which have changed after the most recent crawling was performed. A document is re-crawled if either the content or metadata or the direct security access information of the document has changed. A document is also re-crawled if it is moved within Documentum eRoom Server and the end user has to access the same document with a different URL. Documents deleted from items are removed from the index during incremental crawling.

Documentum eRoom Web Services

The Documentum eRoom application is a COM-based application. To interact with the crawler plug-in, a Web service has been created to fetch the data from eRoom (through eRoom APIs) and provide it to the crawler plug-in.

Important Notes for Documentum eRoom Sources

- The eRoom crawler plug-in should use the administrator account for crawling and indexing eRoom items.
- The Documentum eRoom Server version must be 7.3.

Supported Platforms

The following platforms are supported by this release of Documentum eRoom Web Services:

- Windows 2000/2003 Server
- Microsoft Internet Information Server (IIS) 5.0 or higher

Required Software

- Documentum eRoom Server version 7.3 must be installed and configured
- Oracle SES must be installed
- Documentum eRoom Server Administrator
- The server hosting eRoom must contain Windows .NET Framework 1.1

Required Tasks

The following tasks must be performed before installing the Documentum eRoom Server plug-in:

- ***Microsoft Active Directory Identity Plug-in:*** Configure Oracle SES to Active Directory Identity Plug-in:

  This task must be performed if the identity plug-in for Active Directory is being used for authentication.

  In the Oracle SES Administration GUI, navigate to the Global Settings - Identity Management Setup page. Select **The Active Directory Identity Plug-in Manager implemented based on Oracle User & Role API,** and click **Activate**.

  – For **Authentication Attribute**, select 'USER_NAME'.

  – For **Directory URL**, enter the host name and port number, for example 'ldap://ldapserverhost:port'.

  – For **Directory account name**, enter Active Directory User, for example 'Administrator'.

  – For **Directory account password**, enter the password for **Directory account name**.

  – For **Directory subscriber**, enter the Active Directory information (ldap base); for example, 'dc=us,dc=oracle,dc=com'.

  – For **Directory security protocol**, enter the appropriate value: 'none' or 'port number'.

  Click **Finish**.

- ***Microsoft Active Directory Identity Plug-in:*** Synchronize users and groups from Active Directory to eRoom:
  1. Login to eRoom Server and navigate to Community Setting.
  2. On the right side, click **Directories - Select add a Directory connection**. For **Name**, enter a name for the LDAP Directory Connection. Select the **LDAP Directory** option. Click **Next**.
  3. Enter the URLs for the LDAP directory you want to connect to. Provide the user name and password of the LDAP server. Click **Next**. For **Search Root**, specify dc=us,dc=oracle,dc=com.
  4. For **Search Filter**, specify cn=*. Click **Next**.
  5. Display the test query of connection information. Click **Next**.
  6. Attribute Map information is displayed. Click **Next**.
  7. Display the test Mapping. If these are correct, click **OK**.
  8. Run the LDAP_Synchronization job: To synchronize a connection, click synchronize all connection. Click **OK**.
- Set up the eRoom Web Service:
  1. Check the pre-installation requisites before proceeding.
  2. Navigate to the *ORACLE_HOME*`/search/lib/plugins/eroom` folder. Unzip `EroomServices.zip` to any temporary folder on the computer where the IIS instance for eRoom is installed.
  3. Run `Setup.Exe` to install the Web service on the server that is hosting eRoom. Provide a name for the virtual directory to be created. This name is required when entering the **URL for Web Service** parameter in Oracle SES.
  4. Verify that the Web service is installed by checking the following URL:

     `http://`*`iisServerIP/host/VirtualDirectoryName`*

Known Issues

- The number of votes cast does not get crawled.
- To validate and authenticate users, an eRoom source can use either the Oracle Internet Directory or the Microsoft Active Directory identity plug-in. This connector does *not* support the native eRoom identity management system.

Creating a Documentum eRoom Source

Create a source for the user-defined eRoom source type on the Home - Sources page. Enter a source name. Provide values for the following parameters.

- **Container name**: The names of the containers to be crawled by Oracle SES. You can crawl the entire Site, Community, Facility, or eRoom item. Required.

  The format for specifying container is as follows:

  ```
  <siteName>   OR
  <siteName>/<communityName>   OR
  <siteName>/<communityName>/<FacilityName>   OR
  <siteName>/<communityName>/<FacilityName>/<eRoomName>
  ```

  For example:

  ```
  Container name:OracleSite/OracleCommunity/OracleFacility/OracleRoom
  ```

OracleRoom is crawled.

- **Attribute list**: The comma-delimited list of eRoom custom attributes along with their data types to be searchable. The format is *attributeName:attributeType, attributeName:attributeType*. Valid values are String, Number, and Date.

  While crawling eRoom, an attribute is indexed only if both name and type match the configured name and type; otherwise, it is ignored. This is an optional field. For example, to make the following eRoom attributes searchable:

  – Attribute Name: Account Name Attribute Type: String

  – Attribute Name: Account ID Attribute Type: Integer

  – Attribute Name: Creation Date Attribute Type: Date

  The value should be:

  ```
  Account Name: String, Account ID: Number, Creation Date: Date
  ```

  The default searchable attributes for Documentum eRoom Server are Modified Date, Title, Author, CreateDate, and MimeType.

- **User name**: User name of a valid Documentum eRoom Server user. The user should be an administrator or a user who has access to all content, metadata, and ACL from all folders and documents of items configured in **Container name**. Required.
- **Password**: Password of the Documentum user configured previously. Required.
- **Crawl versions**: Controls whether multiple versions of documents are crawled. Valid values are `true` or `false`. The default value is `false`. Any other value is interpreted as `false` and only the latest version of a file is crawled. Optional
- **URL for Web Services**: A valid URL where eRoom Web service has been installed. (`http://`*`server/virtualName`*) For example, `http://10.113.10.82/EroomServices`.
- **URL for viewing the documents**: A valid IP address or host name with port number (*`IP_ address:port`*) of the server hosting Documentum eRoom. It is used for viewing the Oracle SES search results; for example, `http://10.113.10.82/eRoom` or `http://10.113.10.82:7512/eRoom`.
- **Authentication Attribute**: Attribute used by the LDAP to validate the user. This varies based on the identity plug-in used for authentication. For Active Directory, set it to `USER_NAME`.

Setting Up Lotus Notes Sources

Lotus Notes data is stored in notes-databases, which can be further contained inside directories on a server. A Lotus Domino Server instance can have one or more databases that can be crawled using the Lotus Notes source. The Lotus Notes source navigates through the databases to crawl the documents (for example, e-mail, calendar, address book, and "to do") in the specified databases. It stores the metadata, and accesses information in Oracle SES to provide search according to the end user's credentials.

The Lotus Notes connector lets you enable or disable multiple attachment support with the `Attachment as Search Item` attribute. When this is disabled, the

additional attributes `Parent URL` and `Parent Title` are added for all attachment documents, to link it with the parent document.

The Lotus Notes source supports incremental crawling; that is, it crawls and indexes only those documents that have changed after recent most crawling was scheduled. A document is re-crawled if either the content, metadata, display URL or the direct security access information of the document has changed. Documents deleted from a database are removed from the index during incremental crawling.

To enable Oracle SES to launch Notes thick client, set the **Notes Thick Client** parameter to `true`.

Important Notes for Lotus Notes Sources

The user-account used to crawl Lotus Notes databases should preferably be an Administrator account, such that it has access on all databases and can retrieve and crawl all documents in the specified databases.

Required Software

- Lotus Domino Server R5.0.9/R6.5.4/R7.0
- Notes Clients R5.0.9/R6.5.4/R7.0

Required Tasks

The following tasks must be performed before installing the Lotus Notes source:

1. HTTP and DIIOP tasks must be running on Domino Server.
2. If the Active Directory identity plug-in is used, then the users and user-groups in the Domino Directory must be synchronized with Active Directory. While using the Active Directory identity plug-in, the short-name in the Lotus Notes person document is used for validating the user in Active Directory, so it should be a resolvable logon name in Active Directory.
3. Configure the server document:
   a. Open the server document on the Lotus Notes server that must be crawled.
   b. On the Configuration page, expand the **Server** section.
   c. On the Security page, in the **Programmability Restrictions** area, specify the appropriate security restrictions for your environment in the following fields:

      Run restricted Lotus Script/Java agents

      Run restricted Java/Javascript/COM

      Run unrestricted Java/Javascript/COM

      For example, you might specify an asterisk (*) to allow unrestricted access by Lotus Script/Java agents, and specify user names that are registered in the Domino Directory for the Java/Javascript/COM restrictions.

      Note: The crawler that you configure to crawl this server with the DIIOP protocol must be able to use the user names that you specify in these fields.

   d. Open the Internet Protocol page, then open the HTTP page, and set the **Allow HTTP Clients to Browse Database** option to **Yes**.

e. Configure the user document:

   Open the user document on the Lotus Notes server. This document is stored in the Domino directory.

   On the Basics page, for **Internet password**, specify a password.

f. Restart the DIIOP task on the server.

4. Before activating the Lotus Notes identity plug-in, copy these Lotus files

```
Note.jar
NCSO.jar
```

   to

```
ORACLE_HOME/search/lib/plugins/ln/
```

Known Issues

- A Lotus Notes source does not index encrypted fields, and the content of attachments with encrypted documents, for searching. With encrypted documents, the URL of the search result launches the Notes document instead of the attachment file, which is the case when non-encrypted documents are crawled.
- Deleted Notes documents and attachments in Notes documents are still searchable after an incremental crawl that was set by specifying 'Recrawl using last modified date' as true. To remove URLs from deleted documents or attachments from the Oracle SES index, either perform a force re-crawl (that is, change the re-crawl policy to **Process All Documents** on the Home - Schedules - Edit Schedule page) or mark the 'Recrawl using last modified date' source parameter as false.

Setting Up Identity Management for Lotus Notes

Activate an identity plug-in on the Global Settings - Identity Management Setup page.

The users/groups on Active Directory can be synchronized with Lotus Domino Directory such that all users/groups in Active Directory get registered in Domino as well. Thus, any ACL entry in a notes database or notes document can be validated in Active Directory also, and vice versa.

> **See Also:** "Activating the Active Directory Identity Plug-in" on page 9-8

Oracle SES also provides a Lotus Notes identity plug-in so the Lotus Domino Directory can authenticate and validate the notes native users and groups in Oracle SES.

Activate the Lotus Notes identity plug-in with the following parameters:

- **Server name**: The Domino server fully qualified host name/IP address. If the HTTP port on the Domino server is not 80, then the host name should be *server_name:HTTP_port_number*.
- **User name**: User name of a valid Lotus Domino Server user. Required.
- **Password**: Internet password of the Lotus Notes user. Required.

Creating a Lotus Notes Source

Create a Lotus Notes source on the Home - Sources page. Select **Lotus Notes** from the Source Type list, and click **Create**. Enter values for the following parameters:

- **Server Name**: The Domino server fully qualified host name or IP address. For example, if the Lotus Notes database name is `ses.nsf`, then enter `ses.nsf` for this parameter. If the HTTP port on the Domino server is not 80, then the host name should be *`server_name:HTTP_port_number`*. Required.
- **Attribute list**: The comma-delimited list of Lotus Notes attributes along with their data types to search. The format is *`AttributeName:AttributeType`*, *`AttributeName:AttributeType`*. The valid values are String, Number, and Date. For example: `Subject:String`

Table 7–1 Lotus Notes Data Type Mapping

Sr. No	Lotus Notes Data Type	Oracle SES Data Type
1	Boolean	String
2	Integer	Number (Big Decimal)
3	String	String
4	Date	Date

While crawling a database, an attribute is indexed only if both name and type match the configured name and type; otherwise, it is ignored. This is an optional parameter.

The default searchable attributes for Lotus Domino Server are LASTMODIFIEDDATE, Title, and Author. Multiple attributes with same name are not allowed.

- **User name**: The user name of a valid Lotus Domino Server user. The user should be an Administrator user or a user who has access to all folders and documents of the databases configured in the **Container name** parameter. The user should be able to retrieve content, metadata, and ACL from documents of all databases configured in **Container name** parameter. Required.
- **Password**: Internet password of the Lotus Notes user. Required.
- **Container Name**: Names of the containers to be crawled. Multiple container names must be comma delimited. The container name can include folders, databases, views and folders within databases. For example, `database-abc.nsf`, `folders-folder1`, `views-abc.nsf:By Author`, and `db-abc.nsf:folder\subfolder`. Note that Lotus Notes database file name must be specified with the extension.
- **Crawl Public Documents**: Indicate whether the public documents on notes databases must be crawled such that they are available to anonymous users in Oracle SES, either true or false. Required.
- **Authentication Attribute**: The attribute used to validate the ACL. With the Active Directory identity plug-in, set the value to `USER_NAME`. With the Lotus Notes identity plug-in, set the value to `NATIVE`. Required.
- **Mail Template Name**: This parameter is specific to the mail-databases and the mail template's name should be specified here if any/all of the databases being crawled are mail databases. This is a mandatory parameter if either the **Past Days** or **Future Days** parameter is specified.
- **Past Days**: If the user is crawling calendar entries, then this parameter specifies the number of days earlier for which the calendar entries are picked. The date of reference here is the start date of the event. This accounts for the number of days earlier, and it does not filter the search by time.

- **Future Days**: If the user is crawling calendar entries, then this parameter specifies the number of days in the future for which the calendar entries are picked. The date of reference here is the end date of the event. This accounts for the number of days in the future, and it does not filter the search by time.
- **Notes Title Field**: Because in Lotus Notes custom applications it is not mandatory to maintain a Title field, this parameter has been provided to specify those text fields that should be parsed to retrieve the title field. For example, you could enter `Subject`. With multiple field names, the first field available on the document is selected for the title. Required.
- **Notes Thick Client**: Enter `true` to use Lotus Notes (thick client). Enter `false` to use Lotus Notes Web access.
- **Recrawl using last modified date**: Enter `true` to enqueue only modified documents. Required.
- **Attachment As Search Item**: Enter `true` to have each document in the attachment be submitted individually as an independent document with the same set of attributes and ACLS as that of the parent document. Enter `false` to have attachments be added to the parent document and submitted as a unit.

Displaying the Parent URL in the Search Results

Take the following steps to display the Parent URL attribute in the search results for Lotus Notes connector:

1. On the Global Settings page, select **Use Advanced Configuration**. The Global Settings - Configure Search Result List is displayed.
2. Under Attribute Selection, move **ParentURL** to the **Included** list.
3. Under Style Sheets in the Enter an XSLT box, scroll to `<!-- Links link -->...</td>` and enter the following XSL code:

```
<tr>
   <td>
      <xsl:if test="parenturl [.!= '']">
         <xsl:text>ParentURL: </xsl:text>
         <a class="browseLink" href="{parenturl}">
            <xsl:value-of select="parenturl" />
         </a>
      </xsl:if>
   </td>
</tr>
```

4. Click **Apply**.

Setting Up Microsoft Exchange Sources

Oracle SES can crawl through and provide secure search for e-mail and calendar items, related metadata, attributes, ACLs, and attachments in Microsoft Exchange. It also provides attribute search and browse functionality, which allows search to be done against a specific subfolder in the hierarchy.

Oracle SES supports incremental crawling; that is, it crawls and indexes only those documents that have changed since the last crawl was scheduled. A document is re-crawled if either the content or metadata or the direct security access (permissions) information of the document has changed. A document is also re-crawled if it is

moved within Microsoft Exchange. Documents deleted from Exchange are removed from the index during incremental crawls.

A Microsoft Exchange source covers the following objects in Exchange:

- E-mail
- E-mail attachments
- Calendar events

Important Notes for Microsoft Exchange Sources

On the Exchange server, the super user must grant himself the `Send as` and `Receive as` privileges. You can enable privileges globally for all users in the system. No user-specific privilege grants are required.

See Also:

- *Microsoft Exchange 2003 Technical Reference Guide* and information about permissions in Microsoft Exchange: http://www.microsoft.com/technet/prodtechnol/exchange/default.mspx
- *Oracle Secure Enterprise Search Release Notes* on OTN for supported platforms

Required Software

- Microsoft Internet Information Server (IIS)

Note: The file ADODB.dll is usually included in the Windows .NET Framework SDK. However, if this file is not on your computer, then you must download the ADODB.dll appropriate for your system from Microsoft and install it using the following command:

```
gacutil /i adodb.dll
```

You can download the Windows .NET Framework from this site:

http://www.microsoft.com/downloads/details.aspx?FamilyID=FE6F2099-B7B4-4F47-A244-C96D69C35DEC&displaylang=en

---

Required Tasks

- Proper permissions on the Exchange server must be granted to the Exchange administrator. The Exchange server is crawled with the permission of a super user with the `Send as` and `Receive as` privileges. The easiest way to configure this is to use an administrator as super user or create a super user with the administrator privilege and the `Send as` and `Receive as` privileges targeting Exchange inbox store and public folders.
- To enable the Outlook Web Access logon page, you must enable forms-based authentication on the server. To enable forms-based authentication:
  1. On the Exchange server, log on with the Exchange administrator account, and then start Exchange System Manager.
  2. In the console tree, expand **Servers**.

3. Expand the server for which you want to enable forms-based authentication, and then expand **Protocols**.
4. Expand HTTP, right-click **Exchange Virtual Server**, and then click **Properties**.
5. In the **Exchange Virtual Server Properties** dialog box, on the **Settings** tab, in the **Outlook Web Access** pane, select the **Enable Forms Based Authentication** option.
6. Click **Apply**, and then click **OK**.
7. Restart the IIS server.

If you are using forms-based authentication with SSL off-loading, you must configure your Exchange Server front-end servers to handle this scenario.

See Also: *How to Enable Forms-Based Authentication* at

http://technet.microsoft.com/en-us/library/bb123832.aspx

Known Issues

E-mails with multibyte characters sent from a browser with a different language set than the characters in the mail are not indexed correctly in Oracle SES. The multibyte characters are converted to question marks (?).

This is a known e-mail content issue with Microsoft Exchange. To send future e-mails so that the Microsoft Exchange connector can crawl them properly, either of these workarounds can be applied:

- Change the browser language to the characters in the e-mail. For example, set it to "Japanese" to input Japanese characters.
- Change the value of the following registry key:

```
HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\MSExchangeWEB\OWA\UseRegio
nalCharset
     (Original) '1'
     (New)      Any number (except 1). For example, '0'
```

See Also:

- *How to Modify the Default Browser Language Settings for Outlook Web Access* at

  http://technet.microsoft.com/en-us/library/aa996640.aspx

- *Outlook Mobile Access and Exchange 2003* at

  http://technet.microsoft.com/en-us/library/dafc9621-7b71-42fa-b1cb-3ea63e15ad04.aspx

Setting Up Identity Management for Microsoft Exchange

The Microsoft Exchange connector uses WebDAV for best performance. Oracle recommends that Active Directory be used as identity management system for the Oracle SES instance. The Active Directory instance must be the same one that Microsoft Exchange is using to authenticate users on the file system.

For the Oracle SES instance to read the files during crawling, add permission to each folder and file to make them accessible by the operating system user that runs the

Oracle SES instance. Adding permissions to a folder automatically adds the same permissions to all the files and subfolders in the folder.

> **See Also:** "Activating the Active Directory Identity Plug-in" on page 9-8

Creating a Microsoft Exchange Source

Create a Microsoft Exchange source on the Home - Sources page. Select **Microsoft Exchange** from the Source Type list, and click **Create**.

Enter values for the following parameters:

- **User Name**: User name to authenticate between Oracle SES and Exchange
- **Password**: password to authenticate between Oracle SES and Exchange
- **Server**: Microsoft Exchange server IP
- **Domain**: Microsoft Exchange server domain
- **LDAP Port**: Microsoft Exchange LDAP port
- **Simple Include**: To limit crawling, specify up to 50 colon-delimited path inclusion boundary rules using simplified regular expressions. Specify an inclusion rule that a URL contain, start with, or end with a term. Only `*`, `^`, and `$` operators are permitted. An asterisk (`*`) is a wildcard. A caret (`^`) denotes the beginning of a URL, and a dollar sign (`$`) denotes the end. For example:
  `^https://*.oracle.com/.jpg$`
- **Simple Exclude**: To limit crawling, specify up to 50 colon-delimited path exclusion boundary rules using simplified regular expressions. Only `*`, `^`, and `$` operators are permitted.
- **Regular Expression Include**: To limit crawling, specify up to 50 colon-delimited path inclusion boundary rules using restricted (full java.util.regexp) regular expression rules. For example:

  ```
  ^https://.*\.oracle(?:corp){0,1}\.com
  ```

- **Regular Expression Exclude**: To limit crawling, specify up to 50 colon-delimited path exclusion boundary rules using restricted (full java.util.regexp) regular expression rules.

Microsoft Exchange Source Attributes

- ReceivedTime
- From
- To
- CC
- Subject
- Lastmodifieddate

Setting Up NTFS Sources for Windows

This section contains information for NTFS sources on Windows. For NTFS on UNIX, see "Setting Up NTFS Sources for UNIX" on page 7-14.

The NTFS connector enables Oracle SES to search file repositories in Microsoft NTFS. An Oracle SES NTFS source collects the content, metadata attributes and ACLs of files in NTFS. An NTFS source supports incremental crawl. After the initial crawl is performed, subsequent crawls only collect those documents that have changed since the last crawl. A document is re-crawled if the content, metadata, or the ACL information of the document has changed. A file is also re-crawled if it is moved between folders. Files deleted from NTFS are removed from the index during incremental crawls.

Important Notes for NTFS Sources

- The operating system user running the Oracle SES instance must have read permission on the NTFS file share being crawled. For example, if the remote file share `\\computer1\share1\directory1\` is crawled by the NTFS source, then the Oracle SES instance must be run as a domain user who has access to the file share.
- If you get the ACL in the form `<encrypted acl>@domain` for a folder on a remote computer, it probably means that the computer running the Oracle SES instance and the remote computer are on different domains and your computer cannot interpret the ACLs appropriately.
- Currently, the Oracle SES crawler considers the shared folder an empty document, but it is not indexed; therefore, the total number of unique documents indexed is less than the total number of documents fetched.
- An ACL error may appear when crawling an NTFS source as a built-in user or group, such as an Administrator user. As a workaround, set explicit access to the administrator user: Security - Administrator (user), All Permissions.
- *Everyone* is a special group that represents all current network users, including guests and users from other domains. When a user logs on to the network, the user is automatically added to the `Everyone` group. The NTFS connector supports the `Everyone` group. All documents for which the `Everyone` group has permission is crawled and accessed like public documents. There is no need to log in to the search application to access these public documents. However, if there is a "deny" to a user along with permissions to `Everyone` group to access the document, then all users except for the one for who "deny" has been granted can see the document, and these users must log in to the search application to see the document.
- When using Internet Explorer with files on a different domain, you must explicitly log on to Internet Explorer to open result links to those files.
- When you use the NTFS connector and search file types of `.txt`, `.zip`, or `.rtf`, only the Title and Author attributes are fetched and indexed. For these attributes, the crawler fetches the properties stored in the authoring program (typically accessed by selecting **Properties** from the File menu) and not the NTFS properties (accessed in Windows Explorer by right-clicking the file name and choosing **Properties**).

Required Software

Windows .NET Framework 2.0

Required Tasks

- If not previously installed, then download and install the Windows .NET 2.0 Framework from this site:

http://msdn.microsoft.com/netframework/downloads/updates/default.aspx

- Before crawling the first NTFS source, change the log on account of `OracleService`*`SID`* and `Oracle`*`SID`*`TNSListener` to the domain administrator and restart both services.

Setting Up Identity Management for NTFS Sources

When an NTFS source is used, Oracle recommends that Active Directory be used as identity management system for the Oracle SES instance. The Active Directory instance must be the same one that NTFS is using to authenticate users on the file system.

For the Oracle SES instance to read the files during crawling, add the permission to each folder and file to make it accessible by the operating system user that runs the Oracle SES instance. Adding permissions to a folder automatically adds the same permissions to all the files and sub-folders in the folder.

NTFS sources rely on Active Directory for security permissions. Because permissions at the server local group level are not defined in Active Directory, these permissions are not supported when crawling NTFS sources. Permissions for server local groups (not domain local groups) are ignored during crawling. Permissions for domain groups and users inherited from server local groups also are ignored.

> **See Also:** "Activating the Active Directory Identity Plug-in" on page 9-8

Creating an NTFS Source

Create an NTFS source on the Home - Sources page. Select NTFS from the Source Type list, and click **Create**. Enter values for the following parameters:

- **UNC Path**: UNC Paths, for example, `\\MyServer\Mysharedfolder`
- **Domain Name**: Domain name of the URL (**UNC Path**)
- **Simple Include**: To limit crawling, specify up to 50 colon-separated path boundary rules using simplified regular expressions. Only `*`, `^`, and `$` operators are permitted. For example: `^https://*.oracle.com/.jpg$`
- **Simple Exclude**: To limit crawling, specify up to 50 colon-separated path boundary rules using simplified regular expressions. Only `*`, `^`, and `$` operators are permitted.
- **Regular Expression Include**: To limit crawling, specify up to 50 colon-separated path boundary rules using restricted (full java.util.regexp) regular expression rules. For example: `^https://.*\.oracle(?:corp){0,1}\.com`
- **Regular Expression Exclude**: To limit crawling, specify up to 50 colon-separated path boundary rules using restricted (full java.util.regexp) regular expression rules.
- **Use Local Display URL**: Enter `true` to use the local display URL or `false` to use display the content in a web browser.
- **Authentication Attribute**: Authentication attribute used by the LDAP to validate the user. Use `USER_NAME` for Active Directory and `nickname` for Oracle Internet Directory.

After crawling an NTFS source, you may get a "No User Found Matching the Criteria" error message on the Home - Schedules - Data Synchronization page. This error is

signalled by the identity plug-in. The NTFS connector tries to validate the principal as user first. If that fails, then it tries to validate the principal as group. This error occurs if there are groups as ACL for a document, because the connector does not know if the given principal is a user or a group.

NTFS Source Attributes

- ACLS_
- FILEDATE
- Host
- Language
- LastModifiedDate
- Mimetype
- Title

Setting Up NTFS Sources for UNIX

This section contains information for NTFS sources on UNIX, which have additional setup steps not required on Windows. For NTFS sources on Windows, see "Setting Up NTFS Sources for Windows" on page 7-11.

An NTFS source collects the content, metadata attributes, and ACLs of files in NTFS. An NTFS source supports incremental crawl. After the initial crawl is performed, subsequent crawls only collect those documents that have changed since the last crawl. A document is re-crawled if the content, metadata or the ACL information of the document has changed. A file is also re-crawled if it is moved between folders. Files deleted from NTFS are removed from the index during incremental crawls.

Important Notes for NTFS Sources

- On the Windows server, the super user must have permission to read the NTFS file share.
- The super user must be the impersonate user in the IIS Server.
- The default behavior for NTFS for UNIX is to use local file display URL, so the client computer must have access to the file share.
- An ACL error may appear when crawling an NTFS source as a built-in user or group, such as an Administrator user. As a workaround, set explicit access to the administrator user: Security - Administrator (user), All Permissions.
- **Everyone** is a special group that represents all current network users, including guests and users from other domains. When a user logs on to the network, the user is automatically added to the Everyone group. The NTFS connector supports the Everyone group. All documents for which the Everyone group has permission is crawled and accessed like public documents. There is no need to log in to the search application to access these public documents. However, if a user is denied access to a document while the Everyone group has access, then all users except for the denied user can see the document, and these users must log in to the search application to see the document.
- When using Internet Explorer with files on a different domain, you must explicitly log on to Internet Explorer to open result links to those files.

Required Software

- Microsoft Internet Information Server (IIS)
- NET 2.0 Framework

Setting Up Identity Management with NTFS Sources

When an NTFS source is used, Oracle recommends that Active Directory be used as identity management system for the Oracle SES instance. The Active Directory instance must be the same one that NTFS is using to authenticate users on the file system.

For the Oracle SES instance to read the files during crawling, add permission to each folder and file to make them accessible by the operating system user that runs the Oracle SES instance. Adding permissions to a folder automatically adds the same permissions to all the files and sub-folders in the folder.

NTFS sources rely on Active Directory for security permissions. Because permissions at the server local group level are not defined in Active Directory, these permissions are not supported when crawling NTFS sources. Permissions for server local groups (not domain local groups) are ignored during crawling. Permissions for domain groups and users inherited from server local groups also are ignored.

> **See Also:** "Activating the Active Directory Identity Plug-in" on page 9-8

Creating an NTFS Source

To create an NTFS source on UNIX:

1. On the Home page, select the **Sources** secondary tab.
2. On the Sources page, select the **NTFS** source type and click **Create**.
3. Complete the Create User-Defined Source page. Table 7–2 describes the parameters.
4. Click **Create** or **Create & Customize**.

Table 7–2 NTFS Source Parameters for UNIX

Parameter	Description
UNC Path	UNC path for the NTFS system to crawl; for example, `\\MYSERVER\mysharedfolder`
WebService Endpoint	Target end point (HTTP or HTTPS); for example `https://mail.example.com/NTFSWebService/NTFSWebService.asmx`
WebService User Name	User name to authenticate the NTFS WebService for the Endpoint.
WebServicePassword	Password for **User Name**.
Simple Include	To limit crawling, specify up to 50 colon-delimited (`:`) path inclusion boundary rules using simplified regular expressions. Specify an inclusion rule that a URL contain, start with, or end with a term. Only `*`, `^`, and `$` operators are permitted. An asterisk (`*`) is a wildcard. A caret (`^`) denotes the beginning of a URL, and a dollar sign (`$`) denotes the end of a URL. For example: `^https://*.oracle.com/.jpg$`

Table 7–2 (Cont.) NTFS Source Parameters for UNIX

Parameter	Description
Simple Exclude	To limit crawling, specify up to 50 colon-delimited (:)path exclusion boundary rules using simplified regular expressions. Only `*`, `^`, and `$` operators are permitted.
Regular Expression Include	To limit crawling, specify up to 50 colon-delimited (:) path inclusion boundary rules using restricted (full java.util.regexp) regular expression rules. For example: `^https://.*\.oracle(?:corp){0,1}\.com`
Regular Expression Exclude	To limit crawling, specify up to 50 colon-delimited (:) path exclusion boundary rules using restricted (full java.util.regexp) regular expression rules.
ACL Validation Attribute	ACL attribute used to validate the user. Enter `USER_NAME` for Active Directory or `nickname` for Oracle Internet Directory.
Domain Name	Domain name of the URL (UNC Path).
Incremental Crawl With File Change Detector	Enter `true` to use the File Change Detector, or `false` to use scan-based incremental crawl. See "Installing Oracle Search File Change Detector" on page 7-17

After crawling an NTFS source, you may get a "No User Found Matching the Criteria" error message on the Home - Schedules - Data Synchronization page. If this error accompanies a crawl failure, then check that the principal is a valid user or group

Installing and Configuring Windows Services

NTFS sources on UNIX requires an NTFS agent to be installed and configured on the Windows domain where the NTFS files are to be crawled. The NTFS agent collects and sends content and metadata to the crawler plug-in on the Oracle SES computer in a crawl session. The communication protocol between Oracle SES and the NTFS agent is HTTP or HTTPS.

The NTFS agent must be installed on a Windows computer where IIS is present, and the computer must be in the same Windows domain where the NTFS file share to be crawled resides.

Typically, a remote file share is crawled with the permission of a domain administrator or a domain user with read privileges on the file share. The easiest way to configure this is to add the domain admin group to the administrators group of the target computer.

The Oracle SES instance must connect to the same Active Directory instance that the Microsoft NTFS domain connects to.

Required Software

Windows .NET Framework 2.0

Internet Information Services (IIS) Manager

Required Tasks

Verify that Windows .NET 2.0 Framework is installed. If it is not, then download and install it from this site:

`http://msdn.microsoft.com/netframework/downloads/updates/default.aspx`

Installing Oracle Search File Change Detector

By installing and configuring the Oracle Search File Change Detector service, you can realize significantly improved performance in incremental crawls. This service provides the crawler with a list of documents that are modified or deleted. This method is more efficient than scanning all files for changes.

The older, scan-based incremental crawl is still available. You can use it when you cannot deploy File Change Detector on your NTFS system or under the conditions listed in "Configuring the NTFS Connector" on page 7-21.

The following procedure installs File Change Detector in the Microsoft .NET Framework.

To install Oracle Search File Change Detector:

1. Copy `OracleSearchFileChangeDetector.zip` from *`ORACLE_HOME`*`/search/lib/plugins/ntfsLinWin` to the Windows server where Internet Information Services (IIS) is running.
2. Unzip the contents of OracleSearchFileChangeDetector.zip to a folder. It contains two files:
   - OracleSearchFileChangeDetector.exe
   - OracleSearchFileChangeDetector.exe.config
3. Open OracleSearchFileChangeDetector.exe.config in a text editor and modify the configuration settings as necessary. The settings are described in "Modifying the File Change Detector Configuration File" on page 7-17.
4. Open a command prompt window and navigate to the folder for .NET Framework Version 2.0. It has a name such as

   C:\WINDOWS\Microsoft.NET\Framework\v2.0.50727
5. Install the OracleSearchFileChangeDetector service by issuing a command like the following, where *path* is the folder containing the configuration file:

   ```
   InstallUtil path\OracleSearchFileChangeDetector.exe
   ```

   For example:

   ```
   installutil d:\OracleSearchFileChangeDetector\
   OracleSearchFileChangeDetector.exe
   ```

   The Set Service Login dialog box is displayed.
6. Enter the user credentials for the domain user identified in the ASP.NET Configuration Settings dialog box. For **Username**, use the format *`domain\username`*.
7. Open the Windows Services utility and start the OracleSearchFileChangeDetector service.
8. Install the NTFS Web service as described in "Installing the NTFS Web Service" on page 7-19.

Modifying the File Change Detector Configuration File

The `OracleSearchFileChangeDetector.exe.config` file is the XML configuration file for the File Change Detector. When you add new sources, this file is automatically updated with the UNC path of the sources. However, if you make changes to the path of an existing source, then you must restart File Change Detector for the new path to be watched.

Example 7–1 shows a sample configuration file.

Example 7–1 Oracle Search File Change Detector Configuration File

```
<?xml version="1.0" encoding="utf-8" ?>
<configuration>
   <configSections>
      <section name="StartupFolders"
         type="FileChangeDetector.StartupFoldersConfigSection,
         OracleSearchFileChangeDetector"/>
   </configSections>
   <StartupFolders>
      <DefaultInternalBufferSizeValue>
         <add internalBufferSize="32768" />
      </DefaultInternalBufferSizeValue>
      <Folders>
         <add sourceName="NTFS1" path="10.255.255.255\writeHere"/>
         <add sourceName="NTFS2" path="10.255.255.255\Work"
            internalBufferSize="40960"/>
      </Folders>
      <Results>
         <add directory="C:\NTFS\Data" />
      </Results>
      <SESBufferSizeValue>
         <add sesBufferSize="1" />
      </SESBufferSizeValue>
   </StartupFolders>
</configuration>
```

The XML elements are described in the following topics.

DefaultInternalBufferSizeValue

Oracle Search File Change Detector uses a Windows API to capture file update events. The API uses an internal buffer to cache events. The buffer size is specified in the internalBufferSize parameter of the nested add element:

```
<DefaultInternalBufferSizeValue>
   <add internalBufferSize="n" />
</DefaultInternalBufferSizeValue>
```

The internalBufferSize parameter specifies the default buffer size for all folders that the File Change Detector monitors, as specified in the Folders element.

The internal buffer is allocated from non-paged memory, which cannot be swapped to disk. Therefore, keep the value of internalBufferSize as small as possible. Increase the value for frequent, highly concurrent updates: More than 100 changes per second.

Folders

This element specifies the list of directories to be watched. Create one nested add element for each NTFS source:

```
<Folders>
   <add sourceName="name" path="path"/>
   <add sourceName="name" path="path" internalBufferSize="n"/>
</Folders>
```

The nested add element has these attributes:

- **SourceName**: A unique name within the configuration file to identify the NTFS source. (Required)
- **Path**: The UNC path specified in the NTFS source configuration. (Required)

  To specify multiple UNC paths, use a colon as the delimiter. For example:

  ```
  <add sourceName="ntfstest" path="\\server1\share1\Folder:\\server2\share1"/>
  ```

- **InternalBufferSize**: A value that overrides DefaultInternalBufferSizeValue for a source where extensive changes are expected. (Optional)

Results

Specifies the folder where the Oracle Search File Change Detector logs the changes. The value must be the same as the IncrementalCrawlData property in the Web service configuration.

```
<Results>
   <add directory="path" />
</Results>
```

SESBufferSizeValue

This element specifies the number of events cached in an internal buffer by the OracleSearchFileChangeDetector service before writing them to the log file. For example, a value of 1 indicates that every event is written immediately to the log file, while a value of 10 means that 10 events are cached before writing them to the log file.

Increase the value of the sesBufferSize parameter when capturing changes in folders where you expect extensive changes. However, the larger the buffer size is, the less up-to-date the changes in the log file are, because updates are less frequent. A reasonable value is the average number of concurrent updates to the crawled folders.

```
<SESBufferSizeValue>
   <add sesBufferSize="n" />
</SESBufferSizeValue>
```

Installing the NTFS Web Service

Install this service after you install Oracle Search File Change Detector, as described in "Installing Oracle Search File Change Detector" on page 7-17.

To install the NTFS Web service:

1. Copy `NTFSWebService.zip` from *ORACLE_HOME*`/search/lib/plugins/ntfsLinWin` to the Windows server where Internet Information Services (IIS) is running.
2. Unzip the files in `NTFSWebService.zip` into a permanent folder.
3. Create a virtual directory on the Internet Information Server with the path pointing to the folder created in the previous step.
   a. Select **Administrative Tools** from the Windows Start menu, then select **Internet Information Services (IIS) Manager**.
   b. Expand the navigator in IIS Manager and right-click a Web site.
   c. Select **New**, then **Virtual Directory**.
   d. Follow the steps of the Virtual Directory Creation wizard.

4. On the Virtual Directory Access Permissions page of the wizard, select **Read** and **Run Scripts (such as ASP)**.
5. Open NTFSWebService Properties.
6. On the ASP.NET tab, verify that ASP.NET is version 2.0.
7. On the General tab, enter the settings described in Table 7–3.
8. On the Application tab, select Local Impersonation and enter the user credentials in the form *domain**username*.

   The application user must have these permissions:

   - Read on the NTFS Web Service physical directory
   - Read on the file share to be crawled.
   - Write on the C:\WINDOWS\Microsoft.NET\Framework*version*\Temporary ASP.NET Files folder.

     If the application user does not have access to this directory, then the Web service cannot load the required DLLs and signals the following error when it tries to access the Web service:

     ```
     Server Error in '/NTFSWS683343' Application
     Could not load file or assembly 'WEBSESNTFS' or one of its dependencies.
     Access is denied.
     ```

Table 7–3 ASP.NET Configuration Settings

Parameter	Description
ServiceUsername	User name that authenticates Oracle SES to the NTFS Web service. You also enter this user name when creating the NTFS source. Oracle SES cannot access the Web service without the service username and password.
ServicePassword	Password for ServiceUsername. Ensure that this password is kept secure.
Batchsize	Determines the number of file URLs fetched for a Web service response. The NTFS connector processes a folder by fetching all the files in the folder.
FileChunkSize	Positive integer that specifies the chunk size. Large documents are sent in chunks to the NTFS connector. Enter a positive integer. For example, 1024000 divides the file into 1 MB chunks for sending over the Web. File chunk size should be the optimal data size that can transfer over the network.
IncrementalCrawlData	Path of the Results directory as specified in the Oracle Search File Change Detector configuration file. See "Modifying the File Change Detector Configuration File" on page 7-17. Choose the Application tab and impersonate as user that has read permission on the shared folder. In the example below, "OSES" is the domain and "NTFSCrawler" is a domain user that has read permissions on the shared folder.

To verify that the NTFS Web service is installed correctly:

1. Open Internet Information Services (IIS) Manager.
2. In the navigation tree, select **NTFSWebService** to display its contents in the right pane.

3. Right-click **NTFSWebService.asmx** and choose **Browse**.
4. Ensure that the Web service methods described in Table 7–4 are listed.

Table 7–4 NTFS Web Service Methods

Method	Description
`ClearFCDLog`	Clears the current Oracle Search File Change Detector log.
`ClearPreviousFCDLog`	Clears the previous Oracle Search File Change Detector log.
`GetDFList`	Gets all the files and subfolders in a specified folder.
`GetDocContainer`	Gets the file and the access URL, display URL, and actual content after encoding. It also gets the ACL for the files and attributes of the file.
`GetFileInParts`	Gets the file after breaking it into chunk. The FileChunkSize parameter controls the chunk size.
`GetMinimalMetadata`	Fetches the ACL for the document and the last modified date of the file to determine whether the file has changed.
`GetModifiedURLs`	Gets a list of modified files and folders from the Oracle Search File Change Detector.

Configuring the NTFS Connector

The NTFS connector must be configured to perform incremental crawls with the Oracle Search File Change Detector. The connector has an additional parameter.

To configure the NTFS connector:

1. Open the Oracle SES Administration GUI, and select the **Sources** secondary tab.
2. Create or edit the NTFS connector.
3. Set the **Incremental crawl with the File Change Detector** parameter to `true`.

When the **Incremental crawl with File Change Detector** parameter is set to `true`, the NTFS connector performs the incremental crawl using the detector change logs. It reverts automatically to a scan-based incremental crawl under these conditions:

- The Oracle Search File Change Detector service is stopped.
- The Oracle Search File Change Detector service is started after the previous crawl start time. Scan-based incremental crawl is performed because some changes in the NTFS system might not be captured by the File Change Detector.
- The internal buffer of the File Change Detector overflowed. When the buffer overflows, the file change detector might not capture some changes.

To revert manually to a scan-based incremental crawl, set the **Incremental crawl with the File Change Detector** parameter to `false`.

Known Issues

- The Oracle Search File Change Detector does not capture changes to top-level directories used in the crawler configuration (UNC Path). Note that other directories within the folder are detected correctly.
- Changes to the source configuration, such as boundary rules and maximum file size, do not affect incremental crawls. For these changes to take effect, run a scan-based incremental crawl by setting the **Incremental crawl with the File Change Detector** parameter to false.

- File Change Detector hangs after the Windows Server Active Directory is restarted. You must manually restart the File Change Detector service whenever Active Directory is restarted.

Setting Up Oracle Calendar Sources

Oracle recommends creating one source group for *archived* calendar data and another source group for *active* calendar data. One instance for the archived source can run less frequently, such as every week or month. This source should cover all history. A separate instance for the active source can run daily for only the most recent period.

Setting Up Identity Management for Oracle Calendar

The Oracle SES instance and the Oracle Calendar instance must be connected to the same Oracle Internet Directory system.

To set up a secure Oracle Calendar source:

1. On the Global Settings - Identity Management Setup page in the Oracle SES Administration GUI, select the **Oracle Internet Directory identity plug-in manager**, and click **Activate**.
2. Use the following LDIF file to create an **application entity** for the plug-in. An application entity is a data structure within LDAP used to represent and keep track of software applications accessing the directory with an LDAP client.

   ```
   Oracle_home/bin/ldapmodify -h oidHost -p OIDPortNumber -D "cn=orcladmin" -w
   password -f Oracle_home/search/config/ldif/calPlugin.ldif
   ```

   This string defines the entity that is used for the plug-in: `orclapplicationcommonname=ocscalplugin,cn=oses,cn=products,cn=oraclecontext`. The entity has the password `welcome1`.

Creating an Oracle Calendar Source

Create an Oracle Calendar source on the Home - Sources page. Select Oracle Calendar from the Source Type list, and click **Create**. Enter values for the parameters described in Table 7–5.

Table 7–5 Calendar Source Parameters

Parameter	Value
Calendar server	`http://host name:port`
Application entity name	`orclapplicationcommonname=ocscalplugin,cn=oses, cn=products,cn=oraclecontext`
Application entity password	welcome1
OID server hostname	Oracle Internet Directory *hostname*
OID server port	`389`
OID server SSL port	`636`
OID server ldapbase	`dc=us,dc=oracle,dc=com`
OID login attribute	`uid`
User query	`(objectclass=ctCalUser)`
Past days	`30`

Table 7–5 (Cont.) Calendar Source Parameters

Parameter	Value
Future days	`60`
Rollover	`true`
Calendar server for Display URL	Calendar endpoint URL to be used to formulate the display URL; for example, `http://calendarserver:7777`. If this parameter is blank, then the value provided for the **Calendar server** parameter is used to formulate the display URL.

Oracle Calendar Attributes

- Description
- Priority
- Status
- start date
- end date
- event Type
- Author
- Created Date
- Title
- Location
- Dial_info
- ConferenceID
- ConferenceKey
- Duration

Setting Up Oracle Collaboration Suite E-Mail Sources

Oracle Collaboration Suite 10*g* Mail (Oracle Mail) implements the IMAP protocol, which is used by Oracle SES to retrieve data. You must login to the mail server using the user name and password to retrieve information. Note that Oracle Collaboration Suite mail server has a flag that allows the administrator to crawl mails of all users. The IMAP connector uses this feature to crawl all the mails of all users using the mail server's administration login.

Important Notes for Oracle Collaboration Suite E-Mail Sources

Apart from the private folders, the Oracle Collaboration Suite E-Mail has shared folders. You can share any folder with another person by making it shared. Hence, while doing ACL stamping, the crawler must look if the mail is a part of a private folder or a shared folder and act accordingly.

The Oracle Collaboration Suite E-Mail has a Web interface to open mail. This same Web interface opens the searched mails from Oracle SES.

Required Tasks

For the e-mail administrator to crawl data, set this parameter:

Go to Farm - Midtier - Mail Application - IMAP Server - Default Settings, and set **Allow Admin to Access Any Account** to `true`.

Setting Up Identity Management for Oracle Collaboration Suite E-Mail Sources

Activate the identity plug-in on the Global Settings - Identity Management Setup page. Select **Oracle Internet Directory** identity plug-in and click **Activate**.

Enter values for the following parameters:

- **Authentication Attribute**: Select **nickname**.
- **Host name**: Enter the host name of the computer where Oracle Internet Directory is running.
- **Port**: Enter the value `389`, which is the default LDAP port number.
- **Use SSL**: Enter `true` or `false`.
- **Realm**: Enter the Oracle Internet Directory realm; for example, `dc=us,dc=oracle,dc=com`.
- **User name**: Enter the Oracle Internet Directory administrator user name; for example, `cn=orcladmin`.
- **Password**: Enter the password for the user name.

Creating an Oracle Collaboration Suite E-Mail Source

Create an Oracle Collaboration Suite E-Mail source on the Home - Sources page. Select **Oracle Collaboration Suite E-Mail** from the Source Type list, and click **Create**.

Enter values for the following parameters:

- **Email Server Address**: The IP address or DNS name of the IMAP e-mail server to be crawled, with the port number. This also specifies if the e-mail server follows IMAP or IMAPS protocol. Required.

  Use the format:

  `[imap | imaps]://`*`IPaddress:portNumber`*

  An exception is thrown if this parameter is null. If the server address is incorrect, then an exception is logged at the time of accessing the server.
- **Email Server Admin User**: The administration user name to access the e-mail server. Required.
- **Email Server Admin Password**: The password of the e-mail admin user. Required.
- **Authentication Attribute**: Attribute used to validate the user. This varies based on the identity plug-in used for authentication. Oracle Collaboration Suite E-Mail uses Oracle Internet Directory for authentication, so set this parameter to `mail`.
- **LDAP Server**: The LDAP server information (IP address or DNS name, and so on).
- **LDAP Server Port**: The LDAP server port number.
- **LDAP Base**: The domain to be searched; for example, `dc=oracle, dc=com`.
- **LDAP Query**: The query string defining the users whose e-mails must be crawled. This parameter is used for user-level partitioning.

For example, to crawl only users with names beginning with A and having an e-mail in the domain us.example.com, the query is `(|(cn=A*)(mail=*@us.example.com))`.

- **LDAP Admin User Name**: The administrator user name of the LDAP server. Required.
- **LDAP Admin Password**: The password of the admin user of the LDAP server.
- **Days to which the crawling needs to be done**: Specifies the number of days earlier to which the crawling must be done. The current date (time of crawl) is the base. For example, a value of 7 specifies crawling messages that are seven or more days old. `Today` is the default value.
- **Days from which crawling needs to be done**: The number of days earlier from which the crawling is done. The current date (time of crawl) is the base. For example, a value of 200 specifies crawling messages with dates that are 200 or fewer days old. All mail is the default value.
- **Folders to crawl**: The comma-delimited list of folders to be crawled. '*' means crawl all folders. Other valid values are INBOX, sent, and trash. This does not support regular expressions.
- **Folders not to crawl**: The comma-delimited list of folders not to be crawled. This list is considered only if the **Folders to crawl** parameter has the `*` wildcard as its value. Valid values are `INBOX`, `sent`, and `trash`. This parameter does not support regular expressions.
- **Remove Deleted messages from Index**: Indicates whether to keep the index for deleted mails in incremental recrawls. Valid values are `yes` and `no`. Any other value is considered to be `yes`.
- **Display URL template**: The display URL to be used for viewing the documents. This should have the placeholder for e-mail or user ID. For example, to see the full e-mail address in the display URL, enter the following:

  ```
  http://<>/um/templates/message_list.uix?state=message_list&cA
  ction=openmessage&message_wmuid=$EMAIL
  ```

  To see the user ID, enter the following:

  ```
  http://<>/um/templates/message_list.uix?state=message_list&cA
  ction=openmessage&message_wmuid=$UID
  ```

- **Email Server Version**: The email server to be crawled. Valid values are `ocs10g` or `beehive`.
- **Revisit Skipped Attachments**: Controls whether the crawler revisits attachments that were skipped in earlier crawls because they did not meet the document type inclusion rules. This setting provides an alternative to a force recrawl after changing the document type inclusion rules. Set to `TRUE` to revisit skipped attachments, or set to `FALSE` otherwise (default). The skipped attachments must have been crawled in Oracle SES 11.1.2.2 or later to be revisited.

8

Configuring Access to Applications Sources

This chapter explains how to set up sources for Oracle and third-party databases and for Oracle business applications. It contains the following topics:

- Setting Up Oracle Fusion Sources
- Setting Up Oracle WebCenter Sources
- Setting Up Oracle E-Business Suite Sources
- Setting Up Database Sources
- Setting Up Siebel 7.8 Sources
- Setting Up Siebel 8 Sources

Setting Up Oracle Fusion Sources

Using Oracle SES, you can search for documents within Oracle Fusion Applications. This is done by establishing a connection between Oracle SES and Oracle Fusion using a Fusion connector. To connect to and retrieve documents from Oracle Fusion, you must set up an Oracle SES Fusion identity management system using an identity plug-in, and an authorization management system using an authorization plug-in.

The identity plug-in enables Oracle SES to identify the set of users that can access the Fusion application. The authorization plug-in enables Oracle SES to determine the access rights that each user has for accessing different documents and data within WebCenter. Usually, all users may not have access to the entire data and document set within the application. Instead, each user may have access to a limited set of documents and data.

Setting Up Identity Management System

The identity management system enables Oracle SES to identify the set of users that can access the Fusion application. This is implemented using an identity plug-in.

To activate an identity plug-in for Fusion sources:

1. On the Global Settings page, click **Identity Management Setup** to open the Identity Management Setup page.
2. From the list of available sources, select **Oracle Fusion**, and click **Activate**.

   This opens the Activate Identity Plug-in page.
3. Enter values as described in Table 8–1. Obtain the values from the Fusion application administrator.
4. Click **Finish**.

Table 8–1 Identity Management Parameters for Oracle Fusion

Parameter	Description
HTTP end point for authentication	The HTTP endpoint to which user authentication/validation requests are sent.
User ID	Administration user ID to be used in the HTTP request for user authentication. This user ID is used to validate the authentication request in the Fusion repository. Obtain this ID from the Fusion application administrator.
Password	Administration password.

Defining a Fusion Source

A Fusion application source can be defined from the Source page. After you define the source, you can search for documents within the application.

To create a Fusion source:

1. On the Home page, click the **Sources** subtab.

   This opens the Sources page.

2. From Source Type list, select **Oracle Fusion** and click **Create**.

   This opens the Create Source page, which guides you through a multi-step procedure to enter source and authorization parameters.

3. On the Create Source page, enter the source parameter values listed in Table 8–2.
4. Click **Next** and specify values for the authorization parameters listed in Table 8–3.
5. Click **Create & Customize** to create the source.

Table 8–2 Fusion Connector Source Parameters

Parameter	Description
Configuration URL	URL of the XML configuration file providing details of the source, such as the data feed type, location, security attributes, and so on. The URL is a HTTP URL accessible over HTTP. Obtain this file from the Fusion application administrator.
Authentication Type	Enter the value `NATIVE`.
User ID	User ID to access the data feeds. The access details of the data feed are specified in the configuration file. The user id can be obtained from Fusion administrator.
Password	User password.
Realm	The realm of the application serving the feeds. The parameter is usually left blank.
Oracle SSO Login URL	Oracle Single Sign-On login URL that protects all Single Sign-On applications. Leave the parameter blank.
Oracle SSO Action URL	Oracle Single Sign-On action URL that authenticates Single Sign-On user credentials. Leave the parameter blank.
Scratch Directory	Local directory where status files can be temporarily written.
Maximum number of connection attempts	Maximum number of connection attempts to access data feed or upload status feed.

Table 8–3 Fusion Connector Authorization Parameters

Parameter	Description
HTTP endpoint for authorization	HTTP endpoint for Oracle Fusion authorization. For example, `http://my.host.com:port/AppSearch/SecurityService`
User ID	Administration user ID for Oracle Fusion authorization.
Password	Administration password.
Business component	Name of Oracle Fusion Business Component. For example, `oracle.apps.fnd.fwk.search.NavigationSVO`
Display URL Prefix	HTTP host to prefix the access URL to form the display URL. For example, `http://my.host.com:7777/`. This value must form a valid URL when concatenated with the access URL element of an item in the data feed. Be careful to avoid having either two slashes or none when the values are combined. Thus, enter a trailing slash (/) if the access URLs do not begin with a slash, or omit the trailing slash from the prefix if the access URLs begin with a slash.
Security attribute values for anonymous user	Comma-delimited list of authorized values of security attributes for anonymous user. When this parameter is left blank, the authorization service is contacted to retrieve the values of security attributes accessible for anonymous users.
User Identity Format	Format of user identity string posted to Oracle Fusion Authorization service. Default value is `FND`.

Setting Up Oracle WebCenter Sources

Use the WebCenter connector to connect to and search for documents within Oracle WebCenter 11*g*.

To set up the connector, you must define the source parameters for the connector and set up an authorization management system using an authorization plug-in.

The authorization plug-in enables Oracle SES to determine the access rights that each user has for different documents and data within WebCenter 11*g*. Usually, all users may not have access to the entire data and document set within the application. Instead, each user may have access to a limited set of documents and data.

> **Note:** A WebCenter source uses an Oracle Internet Directory identity plug-in by default. Hence, you need not explicitly set up an identity plug-in for a WebCenter source.

Defining a WebCenter Source

A WebCenter source can be defined from the Source page. After you define the source, you can search for documents within the application.

To create a WebCenter source:

1. On the Home page, click the **Sources** subtab.

   This opens the Sources page.

2. From Source Type list, select **Oracle WebCenter** and click **Create**.

   This opens the Create Source page, which guides you through a multi-step procedure to enter source and authorization parameters.

3. On the Create Source page, enter the source parameter values listed in Table 8–4.
4. Click **Next** and specify values for the authorization parameters listed in Table 8–5.
5. Click **Create & Customize** to create the source.

Table 8–4 WebCenter Connector Source Parameters

Parameter	Description
Configuration URL	URL of the XML configuration file providing details of the source, such as the data feed type, location, security attributes, and so on. The URL is a HTTP URL accessible over HTTP. Obtain this file from the WebCenter application administrator.
Authentication Type	Enter the value `NATIVE`.
User ID	User ID to access the data feeds. The access details of the data feed are specified in the configuration file. The user id can be obtained from Fusion administrator.
Password	User password.
Realm	The realm of the application serving the feeds. The parameter is usually left blank.
Oracle SSO Login URL	Oracle Single Sign-On login URL that protects all Single Sign-On applications. Leave the parameter blank.
Oracle SSO Action URL	Oracle Single Sign-On action URL that authenticates Single Sign-On user credentials. Leave the parameter blank.
Scratch Directory	Local directory where status files can be temporarily written.
Maximum number of connection attempts	Maximum number of connection attempts to access data feed or upload status feed.

Table 8–5 WebCenter Connector Authorization Parameters

Parameter	Description
Authorization Endpoint	URL servicing the lookup of authorization information.
Realm	Realm of the application serving the authorization information.
User ID	User ID to authenticate to the authorization URL.
Password	Password to authenticate to the authorization URL.
Authorization User ID Format	Format of the active identity plug-in user ID that is used by the WebCenter authorization endpoint. For example, this can be the username, email ID, or nickname.

Setting Up Oracle E-Business Suite Sources

The Oracle E-Business Suite connector uses the Oracle SES XML connector framework, where searching is based on Oracle E-Business Suite data available as XML feeds.

See Also: "Overview of XML Connector Framework" on page 3-9

To activate an identity plug-in for Oracle E-Business Suite sources:

1. On the Global Settings page, select **Identity Management Setup**.
2. Select **Oracle E-Business Suite** and click **Activate** to display the Activate Identity Plug-in page.

3. Enter values for the parameters as described in Table 8–6. Obtain the values for these parameters from the E-Business Suite administrator.
4. Click **Finish**.

Table 8–6 Oracle E-Business Suite Identity Management Parameters

Parameter	Value
HTTP endpoint for authentication	HTTP endpoint of Oracle E-Business Suite that provides the user authentication and validation service.
User ID	Administrator user ID for posting data to the endpoint specified in **HTTP endpoint for authentication**.
Password	Password for **User ID**.

To create an Oracle E-Business Suite source:

1. Activate an identity plug-in as described in the previous procedure.
2. On the Home page, select the **Sources** secondary tab.
3. Select **Oracle E-Business Suite** from the Source Type list, and click **Create**.
4. Enter the source parameters as described in Table 8–7.
5. Click **Next**.
6. Click **Get Parameters** to obtain a list of parameters for the authorization manager plug-in.
7. Enter the values for the authorization manager plug-in parameters as described in Table 8–8.
8. Click **Create**.

After processing each data feed, the crawler uploads a status feed to the location specified in the XML configuration file specified in the Configuration URL parameter. This status feed has a name in the following format:

- *`datafeedFilename`*`.suc` when the data feed was processed successfully.
- *`datafeedFilename`*`.err` when an error occurred during processing. The errors are listed in this file.

Table 8–7 Oracle E-Business Suite Source Parameters

Parameter	Value
Configuration URL	URL of the XML configuration file providing details of the source, such as the data feed type, location, security attributes, and so on. The URL is a HTTP URL accessible over HTTP. Obtain this file from the Oracle E-Business Suite administrator.
Authentication Type	Enter the value `Native`.
User ID	User ID to access the data feeds. The access details of the data feed are specified in the configuration file. The user id can be obtained from Oracle E-Business Suite administrator.
Password	Password for **User ID**.
Realm	The realm of the application serving the feeds. The parameter is usually left blank.
Oracle SSO Login URL	URL that protects all OracleAS Single Sign-on applications. Leave the parameter blank.

Table 8–7 (Cont.) Oracle E-Business Suite Source Parameters

Parameter	Value
Oracle SSO Action URL	URL that authenticates OracleAS Single Sign-on user credentials. Leave the parameter blank.
Scratch Directory	A directory on the same computer as Oracle SES, where the status logs are created temporarily.
Maximum number of connection attempts	Maximum number of attempts to connect to the target server to access the data feed.

Table 8–8 Oracle E-Business Suite Authorization Parameters

Parameter	Value
HTTP endpoint for authorization	HTTP endpoint of E-Business Suite that provides the user authorization service.
User ID	User ID.
Password	Password for **User ID**.
Business Component	Name of the Oracle E-Business Suite business component being crawled. The values of the security attributes for which the current user is authorized in the realm of this business component is retrieved to build the security filter for the user when the user logs into Oracle SES. For example, `oracle.apps.fnd.fwk.search.NavigationSVO`.
Security attribute values for anonymous user	Comma-delimited list of authorized values of security attributes for anonymous user. If the parameter is left blank, then the authorization service is contacted to retrieve the values of security attributes accessible for anonymous user.
Display URL Prefix	HTTP host information to prefix the partial URL specified in the access URL of the documents in XML feeds to form the complete URL. This complete URL is the display URL of the document when the document link in the Oracle SES search results page is clicked. This value must form a valid URL when concatenated with the access URL element of an item in the data feed. Be careful to avoid having either two slashes or none when the values are combined. Thus, enter a trailing slash (/) if the access URLs do not begin with a slash, or omit the trailing slash from the prefix if the access URLs begin with a slash.

Setting Up Database Sources

With a database source, you can crawl any JDBC-enabled database. A database source can crawl database content projected as a view or query. Each record in the view or query result set is interpreted as a document. You can create public database sources or secure database sources.

Required Columns in Database Sources

The view or query to be crawled must contain the columns described in Table 8–9. All column names must be in upper case.

Table 8–9 Database Source Required Columns

Column	Type	Description
CONTENT	VARCHAR2 or CLOB	Document content.
KEY	VARCHAR2 or RAW	Key to identify the record in the record set. You can use a custom name for this column by modifying drivers.properties. See "Configuring the JDBC Driver" on page 8-8.
LANG	VARCHAR2	Document language in ISO 639-1 language code; for example, en for English or ja for Japanese.
LASTMODIFIEDDATE	DATE	Last modified date of the document. If you do not have a column for the mandatory LastModifiedDate attribute, use a constant date value in the SQL query for the source. Use a format that the getTimestamp method of the corresponding JDBC driver accepts without errors. Incremental changes to records are not picked up by re-crawls, so always schedule a full crawl.
URL	VARCHAR2	Display URL for the document. The value for this column cannot be null. This connector requires that there is URL-based access to the records in the result set of the view or query.

Optional Columns in Database Sources

The view or query can contain the optional columns describe in Table 8–10. Any other column is considered an attribute of the document.

If the query or view contains both content and either an attachment or attachment link, then one column (in the following order) is considered document content:

1. ATTACHMENT_LINK
2. ATTACHMENT
3. CONTENT

Even if the ATTACHMENT_LINK or ATTACHMENT column is specified in the query, you should include the mandatory CONTENT column. However, the content of ATTACHMENT_LINK or ATTACHMENT is indexed as document content.

Table 8–10 Database Source Optional Columns

Column	Type	Description
ATTACHMENT	BLOB	Binary attachments for the document.
ATTACHMENT_LINK	VARCHAR2	A link to the attachment for the document. HTTP, HTTPS, FILE, and FTP are valid.)
CONTENTTYPE	VARCHAR2	Content type of the document; for example, "text/html" for HTML documents, "application/pdf" for PDF documents, or "application/msword" for Microsoft Word documents. Leave blank when the content type is unknown or varied so that is it not feasible to specify the content type for each document individually.

Table 8–10 (Cont.) Database Source Optional Columns

Column	Type	Description
PATH	VARCHAR2	Path to the document. It is used in the browse feature. It can represent the organizational hierarchy of the document. For example, level1#level2#level3.
TITLE	VARCHAR2	Title of the document to be displayed in the Oracle SES search result page.
LMD_TIMEZONE	VARCHAR2	Specifies the time zone for the date specified in LASTMODIFIEDDATE. For example CST. Oracle SES converts the last modified date from the specified time zone to Oracle SES time zone. If the time zone is not specified, then the date is considered to be in the Oracle SES time zone.

Configuring the JDBC Driver

Depending on your database source, you may need to configure the JDBC driver.

To crawl any third-party database:

1. Download the appropriate JDBC driver jar for JRE 1.6 into *ORACLE_HOME*/search/lib/plugins/oracleapplications.
2. Add the JRE 1.6 JDBC driver jar file name to the JDBC Driver Class parameter, as described in Table 8–11.
3. Add the JRE 1.6 JDBC driver jar file name to the classpath in MANIFEST.MF of appsjdbc.jar and DBCrawler.jar.
4. Restart the middle tier.

For a key attribute that is not named KEY:

1. When configuring the database connector, specify the column name in the Key Attribute Name parameter, as described in Table 8–11.
2. In the crawling query, use the key attribute name as the alias for the key value column name. In this example, ID was entered as the value of the Key Attribute Name parameter and is the alias for KEYVAL:

```
SELECT keyval id, content, url, lastmodifieddate, lang FROM sales_only
```

Query File XML Schema Definition

The following is the XSD that defines the format of the XML query file.

```
<!--[if !supportEmptyParas]-->XSD for the XML sub-queries file:<!--[endif]-->
<?xml version="1.0" encoding="windows-1252" ?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
xmlns="http://xmlns.oracle.com/ses/sqlconnector/detail-attribute-queries"
targetNamespace="http://xmlns.oracle.com/ses/sqlconnector/detail-attribute-queries
"  elementFormDefault="qualified">
  <xsd:complexType name="sqlQueriesType">
    <xsd:annotation>
      <xsd:documentation>
        Specify detail and attribute queries as a source parameter for
        each document fetched by the parent query.
      </xsd:documentation>
    </xsd:annotation>
    <xsd:sequence>
```

```
      <xsd:element name="attachmentQueries" maxOccurs="1" minOccurs="0">
        <xsd:annotation>
          <xsd:documentation>
            Specify detail queries to fetch detail records for each document
            represented by the parent record. The parent records, fetched by
            the parent query, are specified as a source parameter. Each record
            in the document (parent) query can be associated with several detail
            (child) records. Each of these child records has a single column
           specifying the content that will be indexed as attachment to the
           parent document. The child query should select a single column, and
           the WHERE clause should have bind variables of the form
           ##PARENT ATTR##, where the value of PARENT ATTR from the parent
           record is substituted while executing the detail query.
         </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
          <xsd:sequence>
            <xsd:element name="query" maxOccurs="unbounded" minOccurs="1">
             <xsd:complexType>
                <!--Attribute to specify whether the contents retrieved by the
                  query is inline attachment or link to an attachment. The value
                  "true" specifies that the content is a link to an attachment
                  and "false" indicates inline attachment. Default value is
                  false.-->
                <xsd:attribute name="link" default="false"/>
                <!--Content type of the attachment. If no value is specified,
                  SES will auto-detect the content type.-->
                <xsd:attribute name="contenttype" default="null"/>
              </xsd:complexType>
            </xsd:element>
          </xsd:sequence>
        </xsd:complexType>
      </xsd:element>
      <xsd:element name="attributeQueries" maxOccurs="1" minOccurs="0">
        <xsd:annotation>
          <xsd:documentation>
            Specify queries to retrieve values of attributes of the parent
            document. Use this feature if the attribute can contain multiple
            values for a document. If the attribute is a single-valued
            attribute, then it can be specified in the parent query. The WHERE
            clause should have bind variables of the form ##PARENT ATTR##,
            where the value of PARENT ATTR from the parent record is substituted
            while executing the query.
          </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
          <xsd:sequence>
            <xsd:element name="query" maxOccurs="unbounded" minOccurs="1"/>
          </xsd:sequence>
        </xsd:complexType>
      </xsd:element>
    </xsd:sequence>
  </xsd:complexType>
  <xsd:element name="sqlQueries" type="sqlQueriesType"/>
</xsd:schema>
```

Creating Public Database Sources

Public database sources have no security implemented in Oracle SES.

To create a public database source:

1. Create a database source on the Home - Sources page. Select **Database** from the Source Type list, and click **Create**.
2. Enter the database source parameters as described in Table 8–11.
3. Click **Next**.
4. Set authorization to **No Access Control List**, and clear the authorization manager class name and jar name.
5. Click **Create** to create the database source.

Table 8–11 Database Source Parameters

Parameter	Value
Database Connection String	JDBC connection string for the database with content to be crawled. The JDBC string is driver-specific. For example, `jdbc:oracle:thin:@server:port:SID`
User ID	User ID to log in to the database specified in **Database Connection String**. This user ID must have access to the schema owning the view specified in **View** or the query specified in **Query**.
Password	Password to log in to the database specified in **Database Connection String**.
View	Table or view to be crawled. Specify either **View** or **Query**, not both.
JDBC Driver Class	JDBC driver class to connect to the database. For example, oracle.jdbc.driver.OracleDriver. Leave blank to use the default driver: ■ Oracle Database: oracle.jdbc.driver.OracleDriver ■ SQL Server: com.microsoft.sqlserver.jdbc.SQLServerDriver
Key Attribute Name	Name of the KEY attribute in the crawling query/view. The default value is KEY.
Document Count	Maximum number of documents to be crawled before indexing. Enter `-1` to crawl all documents before indexing.
Query File	Path to the XML file specifying the subqueries to crawl attachments and attributes of documents corresponding to every record in the main query. See "Query File XML Schema Definition" on page 8-8.
Query	Query projecting the content to be crawled. Specify either **View** or **Query**, not both.
URL Prefix	String that precedes the content of the URL column and forms a display URL for the document.
Cache File	Prefix of a local file name in which the contents can be temporarily cached while crawling.
Path Separator	The character separating the tokens in the `PATH` of the document as returned by the query or view. It must be a single character, and it cannot be a space, a single or double quote, or a control character.

Table 8–11 (Cont.) Database Source Parameters

Parameter	Value

Parse Attributes	Enter `true` to extract the values of the attributes from the document content specified in the `SOLUTION` or `CONTENT` column. Enter `false` otherwise, or when the content is type text/html.
	In this example, attr1 and attr2 are extracted as attributes of the document with values 22 and 333 respectively:
	`<attr1>22</attr1> <attr2>333</attr2>`
	Content up to the first attribute is interpreted as the document content. The remaining portion is used to extract attributes only. In this example, only "page" is considered document content:
	`page<attr1>22</attr1> is <attr2>333</attr2> dispersed`

Remove Deleted Documents	Enter `true` to remove deleted documents from the index; otherwise, enter `false`.
Attachment Link Authentication Type	Standard Java authentication type used by the application serving the link in the `ATTACHMENT_LINK` column. Enter one of these values: ■ `PUBLIC`: No authentication. ■ `DIGEST`: Digest authentication ■ `BASIC`: Basic authentication ■ `NATIVE`: Native authentication in the source
Attachment Link User ID	User ID for accessing the links specified in the `ATTACHMENT_LINK` column. Required when the link targets are secure.
Attachment Link Password	Password for **Attachment Link User ID**.
Attachment Link Realm	Realm of the application serving the link in the `ATTACHMENT_LINK` column. Required when the link targets are secure.
Grant Security Attributes	Leave blank for public sources.
Deny Security Attributes	Leave blank for public sources.
JDBC Driver Class	JDBC driver class used to connect to the database. For example, oracle.jdbc.driver.OracleDriver.
Key Attribute Name	Name of the key column in the database source. The default value is KEY, as described in Table 8–9, " Database Source Required Columns".

Defining User-Defined Security for Database Sources

Some attributes in the view or query being crawled must be identified as security attributes. The values of these attributes determine if a user is authorized to view a document. These attributes can be either `GRANT` attributes or `DENY` attributes.

To create a database source with user-defined security:

1. On the Home - Sources page, select **Database** from the Source Type list and click **Create**.
2. Enter values for the parameters as described Table 8–11. Specify the `GRANT` and `DENY` attributes as values for parameters **Grant Security Attributes** and **Deny Security Attributes** respectively. If there are multiple `GRANT` or `DENY` security attributes, then separate attribute names with a space.

3. Click **Next**.
4. Enter values for the authorization plug-in parameters:
   - **Authorization Database Connection String**: JDBC connection string for the authorization database. The values of the security attributes to which a given user is authorized are retrieved from this database. The JDBC string is driver-specific.
   - **User ID**: User ID to login to the authorization database.
   - **Password**: Password to login to the authorization database.
   - **Authorization Query**: SQL query to retrieve the values of security attributes to which a given user is authorized. The `SELECT` clause of this query should have all the security attributes specified in Step 2 with identical names. This query can be of two types:
     - The query can return a single record for a given user. The value in each security attribute column should be a space-delimited list of values to which the user is authorized.
     - The query can return multiple records for a given user. The value in each security attribute column of every row of the result set of this query is interpreted as a single value.

       Specify a question mark (`?`) as the placeholder for the username in the query.
   - **Single Record Query**: Enter `true` if the authorization query returns a single record for a given user.
   - **Authorization User ID Format**: Format of the user ID to be used in the SQL query specified in **Authorization Query**. This format should be an authentication attribute of the active identity plug-in.

     For example, if Oracle SES is configured with the Oracle Internet Directory identity plug-in (which supports DN, nickname and e-mail address as authentication attributes), then this parameter can be specified as nickname. The nickname of the current user is then used in the SQL authorization query to build the security filter.

     If no value is specified for this parameter, then the user ID in the canonical form of the active identity plug-in is used in the authorization query to build the security filter.
5. Click **Create** to create the database source.

Database Search Attributes

Database sources have no predefined attributes. The crawler collects attributes from columns defined during source creation. You must map the columns to the search attributes.

Example of Creating a Database Source With User-Defined Security

The document set to be crawled is in tables `T1` and `T2` as specified by the following query:

```
SELECT
     T1.ID,
     T1.DESCRIPTION,
     T2.NAME,
```

```
     T1.LAST_UPDATE_DATE,
     T2.AUTH_ID, T1.HIERARCHY
FROM
     T1, T2
WHERE
     T1.ID = T2.DOC_ID
```

The document content is provided by the `T1.DESCRIPTION` column.

Each document has an HTTP access URL of the form `http://my.company.com/docserver?doc_id=`*`document_identifier`*.

The value of `T2.AUTH_ID` controls access to a document. For example, user SCOTT can access a document only if the value of `T2.AUTH_ID` for the document is in the list of AUTH_IDs for SCOTT as retrieved by the following query:

```
SELECT AUTH_ID FROM USER_AUTH A
     WHERE A.USER='SCOTT'
```

This source can be crawled as a database source type with the following source parameter values:

- **Database Connection String**: `jdbc:oracle:thin:@example:7777:ses`
- **User ID**: `apps_user`
- **Password**: ******
- **View**:
- **Document Count**: `-1`
- **Query**:

```
SELECT
     'docserver?doc_id='|| T1.ID URL,
     T1.ID "KEY",
     'en' LANG,
     T1.LAST_UPDATE_DATE LASTMODIFIEDDATE,
     T1.DESCRIPTION CONTENT,
     'text/plain' CONTENTTYPE,
     T2.NAME CUSTOMER_NAME,
     T2.AUTH_ID,
     T1.HIERARCHY PATH
FROM
     T1, T2
WHERE
     T1.ID=T2.DOC_ID
```

- **Query File**:

```
<?xml version="1.0" encoding="UTF-8" ?>
<sqlQueries xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://xmlns.oracle.com/ses/sqlconnector/detail-attribute-q
ueries detail-attribute-queries.xsd"
 xmlns="http://xmlns.oracle.com/ses/sqlconnector/detail-attribute-queries">
   <attachmentQueries>
     <query>
       <![CDATA[SELECT COMMENTS FROM COMMENTS_TBL WHERE DOCID=##KEY##]]>
     </query>
     <query link="true" contenttype='text/html'>
            <![CDATA[SELECT NOTES FROM NOTES_LINK_TBL WHERE DOCID=##KEY##]]>
     </query>
  </attachmentQueries>
```

```
  <attributeQueries>
       <query>
         <![CDATA[SELECT AUTHOR FROM AUTHORS_TBL WHERE DOCID=##KEY##]]>
       </query>
       <query>
                <![CDATA[SELECT KEYWORD FROM KEYWORD_TBL WHERE DOCID=##KEY##]]>
       </query>
  </attributeQueries>
</sqlQueries>
```

AUTHID and KEY are columns in the select list of the parent query.

Note: This must be the path to a local file containing the subqueries for attributes and attachments that are currently listed directly for query file.

- **URL Prefix**: `http://my.example.com/`
- **Cache File**: `/tmp/cacheFile`
- **Path Separator**: `#`
- **Parse Attributes**: `false`
- **Grant Security Attributes**: `AUTH_ID`
- **Deny Security Attributes**:

The following are sample parameter values for authorization.

- **Database Connection String**: `jdbc:oracle:thin:@example:7777:ses`
- **User ID**: `apps_user`
- **Password**: *****
- **Authorization Query**:

```
SELECT AUTH_ID
FROM USER_AUTH A
WHERE A.USER=UPPER(?)
```

- **Single Record Query**: `false`
- **Authorization User ID Format**: `username`

Setting Up Siebel 7.8 Sources

The Siebel 7.8 source crawler is based on crawling a view or query in a database. Each record in the view or query is considered a document.

The Siebel 7.8 connector supports Siebel installations on the following databases:

- Oracle
- SQL Server

Requirements for Siebel 7.8 Sources

Views and queries to be crawled must contain the columns described in Table 8–12.

Table 8–12 Siebel 7.8 Source Required Columns

Column Name	Datatype	Description
URL	VARCHAR2	Display URL for the document. The value for this column cannot be null. This connector requires that there is URL-based access to the records in the result set of the view or query.
SOLUTION or CONTENT	VARCHAR2 or CLOB	Document content.
LASTMODIFIEDDATE	DATE	Last modified date for crawl.
KEY	VARCHAR2	Primary key of the records.
LANG	VARCHAR2	Document language, such as en for English or ja for Japanese.

Any other columns in the views or queries is considered an attribute of the document.

Installing the JDBC Driver for Microsoft SQL Server

If Siebel 7.8 is installed over a Microsoft SQL Server database, then Oracle SES must have access to the JDBC driver for SQL Server.

To install the SQL Server JDBC driver for Oracle SES:

1. Download the SQL Server 2005 JDBC driver 1.1 from http://www.microsoft.com/downloads/details.aspx?FamilyId=6D483869-816A-44CB-9787-A866235EFC7C.
2. Follow the instructions at the same location to install the driver.
3. Copy sqljdbc4.jar for JRE 1.6 from the installed directory to *ORACLE_HOME*/search/lib/plugins/oracleapplications/.
4. Add sqljdbc4.jar to the classpath in MANIFEST.MF of appsjdbc.jar and DBCrawler.jar.
5. Restart the middle tier.

Starting the Decompression Server

If you are using Windows Native Access for the decompression server, then you can skip this procedure.

To start the decompression server on the Siebel host:

1. Install JDK if it is not already installed on the Siebel host.
2. Set the JAVA_HOME environment variable to the JDK folder.
3. Update the PATH environment variable to begin with JAVA_HOME\bin.
4. Start the RMI registry:
   - To use default port 1099, double-click JAVA_HOME\bin\rmiregistry.exe.
   - To use a different port number, execute this command from the Windows command prompt, replacing *port_number* as desired:

     %JAVA_HOME%\bin\rmiregistry *port_number*
5. Copy this file from the Oracle SES host to the Siebel host and unzip it into the SiebelDecompSvr folder:

ORACLE_HOME/search/lib/plugins/oracleapplications/DecompressionServer.zip

6. Update `SiebelDecompSvr\startDecompServer.bat` by replacing the placeholders in the `java` command with the following information:
   - Absolute path to the `sseunzip.exe` decompression utility.
   - Folder where Siebel attachment files can be temporarily decompressed.
   - Level for logging decompression server output: FINE, INFO, WARNING, SEVERE. Default is INFO.
   - For user credential-based security, the user name specified in the Oracle SES configuration. See Table 8–13, " Siebel 7.8 Identity Management Parameters".
   - For user credential-based security, the password specified in the Oracle SES configuration.
7. For RMI security, replace the placeholders in the `sec.policy` security policy file with the IP address or host name of the Oracle SES instances that crawls the Siebel attachments.
8. Double-click `SiebelDecompSvr\startDecompServer.bat` on the Siebel host to start the decompression server.

Setting Up Identity Management for Siebel 7.8

This procedure enables Oracle SES to validate users by querying the Siebel 7.8 identity management system.

To activate the Siebel 7.8 identity plug-in:

1. On the Global Settings page, select Identity Management Setup under the System heading.

   The **Global Settings - Identity Management Setup** page is displayed.
2. Select **Siebel 7.8** and click **Activate**.
3. Enter values for the parameters described in Table 8–13, then click **Finish**.

Table 8–13 Siebel 7.8 Identity Management Parameters

Parameter	Value
Authentication and Validation Database Connection String	JDBC connection string for the Siebel 7.8 database for authenticating and validating users.
User ID	Administrator ID for the Siebel 7.8 database (specified in **Authentication and Validation Database Connection String**) for validating users.
Password	Password for **User ID**.
User Validation Query	SQL query for validating users. The query must return 1 if the user is valid, and null otherwise. Use a question mark (?) as a placeholder for the user name. This query replaces this default query: `SELECT 1 FROM dbo.S_USER WHERE LOGIN=upper(?)`

Creating a Secured Siebel 7.8 Source

Oracle SES supports these Siebel 7.8 secured business components: Activity, Accounts, Contacts, Literature, Products, and Service Request.

To create a source for Siebel 7.8 secured business components:

1. On the Home page, click the **Sources** secondary tab to display the Sources page.
2. Select **Siebel 7.8** from the **Source Type** list, then click **Create** to display Step 1 Parameters.
3. Complete the form, entering values for the parameters described in Table 8–14.
4. Click **Next** to display Step 2 Authorization.
5. Provide values for the authorization parameters described in Table 8–15.
6. Click **Create**.

Table 8–14 Siebel 7.8 Source Parameters (Step 1)

Parameter	Description
Database Connection String	JDBC connection string for the Siebel 7.8 database from which the content has to be crawled.
User ID	User ID to login to the Siebel 7.8 database specified in **Database Connection String**. This user ID should have access to the schema owning the view specified in **View** or the query specified in **Query**.
Password	Password to login to the Siebel 7.8 database specified in **Database Connection String**.
View	Table or view with the columns needed for crawling. Leave this parameter blank.
Document Count	Maximum number of documents to be crawled before indexing. Enter –1 for this parameter.
Query	Query projecting the columns for crawling. Specify the query for the required business component given in "Queries to Crawl Siebel 7.8 Business Components" on page 8-19.
Query File	Path to the XML file specifying the subqueries to crawl attachments and attributes of documents corresponding to every record in the main query. Leave this parameter blank.
URL Prefix	String to prefix the content of URL column to form a complete display URL for the document
Cache File	Local file to which the contents can be temporarily cached while crawling.
Path Separator	Path separator character in the document path string. Leave this parameter blank.
Parse Attributes	Extracts attribute values from the document content specified in the SOLUTION or CONTENT column. Enter `false` for this parameter.
Grant Security Attributes	Space-separated list of grant security attributes. Enter `VISIBILITYID` for this parameter.
Deny Security Attributes	Space-separated list of deny security attributes. Leave this parameter blank.

Table 8–14 (Cont.) Siebel 7.8 Source Parameters (Step 1)

Parameter	Description
Remove Deleted Documents	Set this parameter to `false` to optimize incremental crawl. Only records that have changed since the previous crawl are crawled. The documents deleted in the database are not deleted from the Oracle SES index. Set this parameter to `true` if records that have been deleted from the database since the previous full crawl, or since a previous incremental crawl with this parameter set to `true`, should be deleted from the Oracle SES index as well. Crawling may take longer when this parameter is set to `true` than when it is set to `false`.
Attachment Link Authentication Type	Enter `NATIVE` for user credential-based security; otherwise, enter `PUBLIC` for this parameter.
Attachment Link User ID	User ID for accessing the link in the attachment link column. If the Attachment Link Authentication Type is `NATIVE` and attachments are accessed using RMI, then enter the user name provided in `SiebelDecompSvr\startDecompServer.bat`; otherwise, leave this parameter blank.
Attachment Link Password	Password for accessing the link in the attachment link column. If the Attachment Link Authentication Type is `NATIVE` and Attachments are accessed using RMI, then enter the password provided in `SiebelDecompSvr\startDecompServer.bat`; otherwise, leave this parameter blank.
Attachment Link Realm	Realm of the application serving the link in the attachment link column. Leave this parameter blank.

Table 8–15 Siebel 7.8 Authorization Parameters (Step 2)

Parameter	Value
Authorization Database Connection String	JDBC connection string for the authorization database. The values of the visibility IDs for users are retrieved from this database. Typically, this connection string matches the **Authentication and Validation Database Connection String** on Page 1.
User ID	Administrator user ID for the authorization database specified in **Authorization Database Connection String**
Password	Password for **User ID**. **Note**: Oracle SES does not allow a password to contain the characters `&`, `\`, `"`, `<` , or `>`.
Authorization Query	SQL query to retrieve the values of visibility IDs. Use a question mark (?) as a placeholder for the user name. See Example 8–1 for more information about authorization queries.

Example 8–1 Siebel 7.8 Authorization Queries

The following query is the default authorization query that can be used for the Service Request, Accounts, Products, Literature, Solution, and Contacts business components:

```
SELECT p.BU_ID visibilityid
FROM dbo.S_POSTN p
   INNER JOIN dbo.S_CONTACT c2 ON c2.PR_HELD_POSTN_ID = p.ROW_ID
   INNER JOIN dbo.S_USER    u  ON u.PAR_ROW_ID        = c2.PAR_ROW_ID
WHERE u.LOGIN = upper(?)
```

Use the following authorization query for Activity business components:

```
SELECT DISTINCT usr.LOGIN visibilityid
FROM
   S_PARTY_RPT_REL rpt,
   S_POSTN postn,
   S_USER usr,
   S_PARTY pty
WHERE
   rpt.PARTY_ID IN (
      SELECT T3.PR_HELD_POSTN_ID
      FROM dbo.S_PARTY T1
         INNER JOIN dbo.S_EMP_PER T2 ON T1.ROW_ID = T2.PAR_ROW_ID
         INNER JOIN dbo.S_CONTACT T3 ON T1.ROW_ID = T3.PAR_ROW_ID
         INNER JOIN dbo.S_USER    T4 ON T1.ROW_ID = T4.PAR_ROW_ID
      WHERE (T3.EMP_FLG = 'Y') AND T4.LOGIN=upper(?))
   AND rpt.SUB_PARTY_ID = postn.PAR_ROW_ID
   AND postn.PR_EMP_ID = usr.ROW_ID
   AND usr.PAR_ROW_ID = pty.ROW_ID
```

See "Queries to Crawl Siebel 7.8 Business Components" on page 8-19.

Creating a Public Siebel 7.8 Source

Oracle SES supports Solution as a public business component.

To create a source for Siebel 7.8 public business components:

1. On the Home page, click the **Sources** secondary tab to display the Sources page.
2. Select **Siebel 7.8 (Public)** from the **Source Type** list, then click **Create**.
3. Complete the form, entering values for the parameters described in Table 8–14. The two parameters to grant and deny security attributes are omitted from the configuration of a public source.
4. Click **Create**.

Queries to Crawl Siebel 7.8 Business Components

This section includes the queries to crawl the Siebel 7.8 business components supported by Oracle SES:

- Service Request Attachments
- Accounts
- Products
- Literature
- Solution
- Service Request
- Contacts
- Activity
- Activity Attachment

> **Note:** The Siebel 7.8 crawling queries provided in this section are supported only when used with default Siebel 7.8 database schema. Neither these queries nor any modifications to these queries are supported in any modified Siebel 7.8 database schema.

The queries appear in two forms: A multi-line query for readability and a single-line version to cut-and-paste into the Query parameter. Queries use SQL Server syntax unless otherwise noted and must be modified slightly for use with Oracle Database.

To use a sample query as the value of the Query parameter:

- Replace *HostName* with the name of the host where Siebel is installed.
- The values of the parameters SWEView and SWEApplet0 in the queries are the names of views and applets in a default Siebel installation. Change them as required if different names were used while installing Siebel 7.8.
- Add appropriate WHERE clauses to these queries depending on the search specification of views, applets and business components in the Siebel system. For example, if the Siebel system is configured to locate only internal service requests, then append the WHERE clause to the query for Service Request business component as follows: `WHERE c.SR_TYPE_CD = 'Internal'`.

To modify a query for use in Oracle Database instead of SQL Server:

- Replace the string concatenation operator '+' with '||'.
- Replace the table owner name dbo with the appropriate table owner name in Oracle Database.

Service Request Attachments

```
SELECT
   'callcenter_enu/start.swe?SWECmd=GotoView
      &SWEView=All+Service+Request+across+Organizations
      &SWERF=1
      &SWEHo=HostName
      &SWEBU=1
      &SWEApplet0=Service+Request+Detail+Applet
      &SWERowId0='+c.PAR_ROW_ID+'
      &SRAttId='+c.ROW_ID URL,
   'US' LANG,
   c.LAST_UPD LASTMODIFIEDDATE,
   'text/html' CONTENTTYPE,
   c.ROW_ID "KEY",
   coalesce('<b>Attachment Name:</b> '+c.FILE_NAME,'<null>')
      +coalesce(',<br><b>SR Number:</b> '+srv.SR_NUM,'<null>')
      +coalesce(',<br><b>SR Summary: </b>'+srv.SR_TITLE,'<null>') SOLUTION,
   c.ROW_ID sblrowid,
   c.CREATED created_on,
   c.CREATED_BY createdby,
   c.LAST_UPD_BY lastupdatedby,
   c.PAR_ROW_ID title,
   c.FILE_SRC_TYPE "type",
   c.FILE_EXT code01,
   c.COMMENTS "comment",
   c.FILE_SRC_PATH location,
   'Service Request Attachment' sblbctype,
   usr.LOGIN owner,
   srv.BU_ID visibilityid
```

```
FROM
   dbo.S_SR_ATT c
   INNER JOIN dbo.S_SRV_REQ srv      ON c.PAR_ROW_ID   = srv.ROW_ID
   LEFT OUTER JOIN dbo.S_USER usr    ON usr.PAR_ROW_ID = srv.OWNER_EMP_ID
   LEFT OUTER JOIN dbo.S_CONTACT con ON con.PAR_ROW_ID = c.LAST_UPD_BY
```

The following is the same query formatted as a single line that you can cut and paste into the Oracle SES Administration GUI:

```
SELECT
'callcenter_enu/start.swe?SWECmd=GotoView&SWEView=All+Service+Request+across+Organ
izations&SWERF=1&SWEHo=HostName&SWEBU=1&SWEApplet0=Service+Request+Detail+Applet&S
WERowId0='+c.PAR_ROW_ID+'&SRAttId='+c.ROW_ID URL, 'US' LANG, c.LAST_UPD
LASTMODIFIEDDATE, 'text/html' CONTENTTYPE, c.ROW_ID "KEY", coalesce('<b>Attachment
Name:</b> '+c.FILE_NAME,'<null>')+coalesce(',<br><b>SR Number:</b>
'+srv.SR_NUM,'<null>')+coalesce(',<br><b>SR Summary: </b>'+srv.SR_TITLE,'<null>')
SOLUTION, c.ROW_ID sblrowid, c.CREATED created_on, c.CREATED_BY createdby,
c.LAST_UPD_BY lastupdatedby, c.PAR_ROW_ID title, c.FILE_SRC_TYPE "type",c.FILE_EXT
code01, c.COMMENTS "comment", c.FILE_SRC_PATH location, 'Service Request
Attachment' sblbctype, usr.LOGIN owner, srv.BU_ID visibilityid FROM dbo.S_SR_ATT c
INNER JOIN dbo.S_SRV_REQ srv ON c.PAR_ROW_ID=srv.ROW_ID LEFT OUTER JOIN dbo.S_USER
usr ON usr.PAR_ROW_ID = srv.OWNER_EMP_ID LEFT OUTER JOIN dbo.S_CONTACT con ON
con.PAR_ROW_ID = c.LAST_UPD_BY
```

Accounts

```
SELECT
   'callcenter_enu/start.swe?SWECmd=GotoView
      &SWEView=All+Accounts+across+Organizations
      &SWERF=1
      &SWEHo=HostName
      &SWEBU=1
      &SWEApplet0=Account+List+Applet
      &SWERowId0='+T1.ROW_ID URL,
   'US' LANG,
   T2.LAST_UPD LASTMODIFIEDDATE,
   'text/html' CONTENTTYPE,
   T1.ROW_ID "KEY",
   coalesce('<b>Name:</b> '+T2.NAME,'<null>')
      +coalesce(',<br><b>Type:</b> '+T2.OU_TYPE_CD,'<null>')+',<br>
         <b>Address:</b>'
      +coalesce(T5.ADDR,'<null>')
      +coalesce(','+T5.CITY,'<null>')
      +coalesce(','+T5.STATE+' '+T5.ZIPCODE,'<null>')
      +coalesce(','+T5.COUNTRY,'<null>') SOLUTION,
   T1.ROW_ID sblrowid,
   T2.CREATED created_on,
   T2.CREATED_BY createdby,
   T2.LAST_UPD_BY lastupdatedby,
   T2.NAME title,
   T2.OU_NUM csn,
   T2.OU_TYPE_CD type,
   T2.LOC location,
   T10.LOGIN alias,
   T5.ADDR street,
   T5.CITY city,
   T5.STATE state,
   T5.COUNTRY country,
   T5.ZIPCODE zipcode,
   'Account' sblbctype,
   T2.BU_ID visibilityid
```

```
FROM
   dbo.S_PARTY T1
   INNER JOIN dbo.S_ORG_EXT T2       ON T1.ROW_ID      = T2.PAR_ROW_ID
   INNER JOIN dbo.S_ACCNT_POSTN T3   ON T2.PR_POSTN_ID = T3.POSITION_ID
      AND T2.ROW_ID = T3.OU_EXT_ID
   INNER JOIN dbo.S_PARTY T4         ON T3.POSITION_ID = T4.ROW_ID
   LEFT OUTER JOIN dbo.S_POSTN T9    ON T3.POSITION_ID = T9.PAR_ROW_ID
   LEFT OUTER JOIN dbo.S_ADDR_ORG T5 ON T2.PR_ADDR_ID  = T5.ROW_ID
   LEFT OUTER JOIN dbo.S_USER T10    ON T9.PR_EMP_ID   = T10.PAR_ROW_ID
   LEFT OUTER JOIN dbo.S_CONTACT T11 ON T11.PAR_ROW_ID = T2.LAST_UPD_BY
WHERE
   (T2.INT_ORG_FLG != 'Y' OR T2.PRTNR_FLG = 'Y')
```

The following is the same query formatted as a single line that you can cut and paste into the Oracle SES Administration GUI:

```
SELECT
'callcenter_enu/start.swe?SWECmd=GotoView&SWEView=All+Accounts+across+Organization
s&SWERF=1&SWEHo=HostName&SWEBU=1&SWEApplet0=Account+List+Applet&SWERowId0='+T1.ROW
_ID URL, 'US' LANG, T2.LAST_UPD LASTMODIFIEDDATE, 'text/html' CONTENTTYPE,
T1.ROW_ID "KEY", coalesce('<b>Name:</b>
'+T2.NAME,'<null>')+coalesce(',<br><b>Type:</b>
'+T2.OU_TYPE_CD,'<null>')+',<br><b>Address:</b>
'+coalesce(T5.ADDR,'<null>')+coalesce(','+T5.CITY,'<null>')+coalesce(','+T5.STATE+
' '+T5.ZIPCODE,'<null>')+coalesce(','+T5.COUNTRY,'<null>') SOLUTION,
T1.ROW_ID sblrowid, T2.CREATED created_on, T2.CREATED_BY createdby, T2.LAST_UPD_BY
lastupdatedby, T2.NAME title, T2.OU_NUM csn, T2.OU_TYPE_CD type, T2.LOC location,
T10.LOGIN alias, T5.ADDR street, T5.CITY city, T5.STATE state, T5.COUNTRY country,
T5.ZIPCODE zipcode, 'Account' sblbctype, T2.BU_ID visibilityid FROM dbo.S_PARTY T1
INNER JOIN dbo.S_ORG_EXT T2 ON T1.ROW_ID = T2.PAR_ROW_ID INNER JOIN
dbo.S_ACCNT_POSTN T3 ON T2.PR_POSTN_ID = T3.POSITION_ID AND T2.ROW_ID =
T3.OU_EXT_ID INNER JOIN dbo.S_PARTY T4 ON T3.POSITION_ID = T4.ROW_ID LEFT OUTER
JOIN dbo.S_POSTN T9 ON T3.POSITION_ID = T9.PAR_ROW_ID LEFT OUTER JOIN
dbo.S_ADDR_ORG T5 ON T2.PR_ADDR_ID=T5.ROW_ID LEFT OUTER JOIN dbo.S_USER T10 ON
T9.PR_EMP_ID = T10.PAR_ROW_ID LEFT OUTER JOIN dbo.S_CONTACT T11 ON
T11.PAR_ROW_ID=T2.LAST_UPD_BY WHERE (T2.INT_ORG_FLG != 'Y' OR T2.PRTNR_FLG = 'Y')
```

Products

```
SELECT
   'callcenter_enu/start.swe?SWECmd=GotoView
      &SWEView=All+Products+across+Organizations
      &SWERF=1
      &SWEHo=HostName
      &SWEBU=1
      &SWEApplet0=Product+List+Applet
      &SWERowId0='+c.ROW_ID URL,
   'US' LANG,
   c.LAST_UPD LASTMODIFIEDDATE,
   'text/html' CONTENTTYPE,
   c.ROW_ID "KEY",
   coalesce('<b>Name:</b> '+ c.NAME,'<null>')
      +coalesce(',<br><b>Part Number:</b> '+c.VENDR_PART_NUM,'<null>')
      +coalesce(',<br><b>Catalog/Category:</b> '+ c2.NAME,'<null>') SOLUTION,
   c.DESC_TEXT description,
   c.ROW_ID sblrowid,
   c.CREATED created_on,
   c.CREATED_BY createdby,
   c.NAME title,
   'Product Catalog' sblbctype,
   c.VENDR_PART_NUM name,
```

```
   c.VENDR_PART_NUM + ' ' + c3.PROD_ID + ' ' + c3.CTLG_CAT_ID summary,
   c.BU_ID visibilityid,
   c2.NAME sblvisibilityinfo,
   c.VERSION type
FROM
   dbo.S_PROD_INT c
   INNER JOIN      dbo.S_CTLG_CAT_PROD c3 ON c3.PROD_ID    = c.ROW_ID
   INNER JOIN      dbo.S_CTLG_CAT      c2 ON c2.ROW_ID     = c3.CTLG_CAT_ID
   LEFT OUTER JOIN dbo.S_CONTACT       c4 ON c4.PAR_ROW_ID = c.LAST_UPD_BY
```

The following is the same query formatted as a single line that you can cut and paste into the Oracle SES Administration GUI:

```
SELECT
'callcenter_enu/start.swe?SWECmd=GotoView&SWEView=All+Products+across+Organization
s&SWERF=1&SWEHo=HostName&SWEBU=1&SWEApplet0=Product+List+Applet&SWERowId0='+c.ROW_
ID URL, 'US' LANG, c.LAST_UPD LASTMODIFIEDDATE, 'text/html' CONTENTTYPE, c.ROW_ID
"KEY", coalesce('<b>Name:</b> '+ c.NAME,'<null>')+coalesce(',<br><b>Part
Number:</b> '+c.VENDR_PART_NUM,'<null>')+coalesce(',<br><b>Catalog/Category:</b>
'+ c2.NAME,'<null>') SOLUTION, c.DESC_TEXT description, c.ROW_ID sblrowid,
c.CREATED created_on, c.CREATED_BY createdby, c.NAME title, 'Product Catalog'
sblbctype, c.VENDR_PART_NUM name, c.VENDR_PART_NUM + ' ' + c3.PROD_ID + ' ' +
c3.CTLG_CAT_ID summary, c.BU_ID visibilityid, c2.NAME sblvisibilityinfo, c.VERSION
type FROM dbo.S_PROD_INT c INNER JOIN dbo.S_CTLG_CAT_PROD c3 ON
c3.PROD_ID=c.ROW_ID INNER JOIN dbo.S_CTLG_CAT c2 ON c2.ROW_ID=c3.CTLG_CAT_ID LEFT
OUTER JOIN dbo.S_CONTACT c4 ON c4.PAR_ROW_ID=c.LAST_UPD_BY
```

Literature

```
SELECT
   'callcenter_enu/start.swe?SWECmd=GotoView
      &SWEView=All+Sales+Tools+across+Organizations
      &SWERF=1
      &SWEHo=HostName
      &SWEBU=1
      &SWEApplet0=Sales+Tool+List+Applet
      &SWERowId0='+c.ROW_ID URL,
   'US' LANG,
   c.LAST_UPD LASTMODIFIEDDATE,
   'text/html' CONTENTTYPE,
   c.LAST_UPD created_on,
   c.LAST_UPD_BY lastupdatedby,
   c.ROW_ID "KEY",
   coalesce('<b>Name:</b> '+c.NAME,'<null>')
      +coalesce(',<br><b>Catalog/Category:</b> '+c4.NAME,'<null>') SOLUTION,
   c.DESC_TEXT description,
   c.NAME title,
   c.NAME name,
   c.FILE_REV_NUM +'' + c3.LIT_ID + ''+ c3.CTLG_CAT_ID + ''+ c4.ROW_ID + ''
      + c4.NAME summary,
   c.LIT_CD "type",
   c.BU_ID visibilityid,
   c4.NAME sblvisibilityinfo,
   'Sales Tool' sblbctype
FROM
   dbo.S_LIT c
   INNER JOIN      dbo.S_CTLG_CAT_LIT c3 ON c3.LIT_ID     = c.ROW_ID
   INNER JOIN      dbo.S_CTLG_CAT     c4 ON c4.ROW_ID     = c3.CTLG_CAT_ID
   LEFT OUTER JOIN dbo.S_CONTACT      c5 ON c5.PAR_ROW_ID = c.LAST_UPD_BY
```

The following is the same query formatted as a single line that you can cut and paste into the Oracle SES Administration GUI:

```
SELECT
'callcenter_enu/start.swe?SWECmd=GotoView&SWEView=All+Sales+Tools+across+Organizat
ions&SWERF=1&SWEHo=HostName&SWEBU=1&SWEApplet0=Sales+Tool+List+Applet&SWERowId0='+
c.ROW_ID URL, 'US' LANG, c.LAST_UPD LASTMODIFIEDDATE, 'text/html' CONTENTTYPE,
c.LAST_UPD created_on, c.LAST_UPD_BY lastupdatedby, c.ROW_ID "KEY",
coalesce('<b>Name:</b> '+c.NAME,'<null>')+coalesce(',<br><b>Catalog/Category:</b>
'+c4.NAME,'<null>') SOLUTION, c.DESC_TEXT description, c.NAME title, c.NAME name,
c.FILE_REV_NUM +'' + c3.LIT_ID + ''+ c3.CTLG_CAT_ID + ''+ c4.ROW_ID + '' + c4.NAME
summary, c.LIT_CD "type", c.BU_ID visibilityid, c4.NAME sblvisibilityinfo, 'Sales
Tool' sblbctype FROM dbo.S_LIT c INNER JOIN dbo.S_CTLG_CAT_LIT c3 ON
c3.LIT_ID=c.ROW_ID INNER JOIN dbo.S_CTLG_CAT c4 ON c4.ROW_ID=c3.CTLG_CAT_ID LEFT
OUTER JOIN dbo.S_CONTACT c5 ON c5.PAR_ROW_ID=c.LAST_UPD_BY
```

Solution

```
SELECT
   'callcenter_enu/start.swe?SWECmd=GotoView
      &SWEView=All+Solution+List+View
      &SWERF=1
      &SWEHo=HostName
      &SWEBU=1
      &SWEApplet0=Solution+List+Applet
      &SWERowId0='+c.ROW_ID URL,
   'US' LANG,
   c.LAST_UPD LASTMODIFIEDDATE,
   'text/html' CONTENTTYPE,
   c.ROW_ID "KEY",
   coalesce('<b>Name:</b> '+c.NAME,'<null>')
      +coalesce(',<br><b>Catalog/Category: </b>'+t.NAME,'<null>')
      +coalesce(',<br><b>Question: </b>'+ cast(c.FAQ_QUES_TEXT
         as nvarchar(4000)),'<null>')
      +coalesce(',<br><b>Resolution: </b>'+ cast(c.RESOLUTION_TEXT
         as nvarchar(4000)),'<null>') SOLUTION,
   c.ROW_ID sblrowid,
   c.CREATED created_on,
   c.CREATED_BY createdby,
   c.NAME title,
   c.FAQ_QUES_TEXT description,
   c.RESOLUTION_TEXT summary,
   c.TYPE_CD "type",
   c.STATUS_CD status,
   usr.LOGIN owner,
   usr.LOGIN alias,
   t.NAME location,
   'Solution' sblbctype
FROM
   dbo.S_RESITEM c
   INNER JOIN dbo.S_USER          usr ON c.CREATED_BY  = usr.PAR_ROW_ID
   INNER JOIN dbo.S_CTLGCT_RESITM cct ON c.ROW_ID      = cct.RES_ITEM_ID
   INNER JOIN dbo.S_CTLG_CAT      t   ON t.ROW_ID      = cct.CTLG_CAT_ID
   INNER JOIN dbo.S_CONTACT       c2  ON c2.PAR_ROW_ID = c.LAST_UPD_BY
```

The following is the same query formatted as a single line that you can cut and paste into the Oracle SES Administration GUI:

```
SELECT
'callcenter_enu/start.swe?SWECmd=GotoView&SWEView=All+Solution+List+View&SWERF=1&S
WEHo=HostName&SWEBU=1&SWEApplet0=Solution+List+Applet&SWERowId0='+c.ROW_ID URL,
```

```
'US' LANG, c.LAST_UPD LASTMODIFIEDDATE, 'text/html' CONTENTTYPE, c.ROW_ID "KEY",
coalesce('<b>Name:</b> '+c.NAME,'<null>')+coalesce(',<br><b>Catalog/Category:
</b>'+t.NAME,'<null>') + coalesce(',<br><b>Question: </b>'+ cast(c.FAQ_QUES_TEXT
as nvarchar(4000)),'<null>')+ coalesce(',<br><b>Resolution: </b>'+
cast(c.RESOLUTION_TEXT as nvarchar(4000)),'<null>') SOLUTION, c.ROW_ID sblrowid,
c.CREATED created_on, c.CREATED_BY createdby, c.NAME title, c.FAQ_QUES_TEXT
description, c.RESOLUTION_TEXT summary, c.TYPE_CD "type", c.STATUS_CD status,
usr.LOGIN owner, usr.LOGIN alias, t.NAME location, 'Solution' sblbctype FROM
dbo.S_RESITEM c INNER JOIN dbo.S_USER usr ON c.CREATED_BY = usr.PAR_ROW_ID INNER
JOIN dbo.S_CTLGCT_RESITM cct ON c.ROW_ID = cct.RES_ITEM_ID INNER JOIN
dbo.S_CTLG_CAT t ON t.ROW_ID = cct.CTLG_CAT_ID INNER JOIN dbo.S_CONTACT c2 ON
c2.PAR_ROW_ID=c.LAST_UPD_BY
```

Service Request

```
SELECT
   'callcenter_enu/start.swe?SWECmd=GotoView
      &SWEView=All+Service+Request+across+Organizations
      &SWERF=1
      &SWEHo=HostName
      &SWEBU=1
      &SWEApplet0=Service+Request+List+Applet
      &SWERowId0='+c.ROW_ID URL,
   'US' LANG,
   c.LAST_UPD LASTMODIFIEDDATE,
   'text/html' CONTENTTYPE,
   c.ROW_ID "KEY",
   coalesce('<b>SR Number:</b> '+c.SR_NUM,'<null>')
      +coalesce(',<br><b>Summary:</b> '+c.SR_TITLE,'<null>')
      +coalesce(',<br><b>Status:</b> '+c.SR_STAT_ID,'<null>')
      +coalesce(',<br><b>Area:</b> '+c.SR_AREA,'<null>')
      +coalesce(',<br><b>Subarea:</b> '+c.SR_SUB_AREA,'<null>')
      +coalesce(',<br><b>Resolution:</b> '+c.RESOLUTION_CD,'<null>') SOLUTION,
   c.DESC_TEXT description,
   c.BU_ID visibilityid,
   c.ROW_ID sblrowid,
   c.CREATED created_on,
   c.CREATED_BY createdby,
   c.SR_TITLE summary,
   a.NAME orgName,
   c.SR_AREA code01,
   a.OU_NUM csn,
   contact.FST_NAME firstName,
   contact.LAST_NAME lastName,
   c.SR_NUM title,
   c.SR_STAT_ID status,
   c.SR_SUB_AREA code02,
   usr.LOGIN owner,
   'Service Request' sblbctype
FROM
   dbo.S_ORG_EXT a
   INNER JOIN      dbo.S_SRV_REQ c        ON a.PAR_ROW_ID       = c.CST_OU_ID
   LEFT OUTER JOIN dbo.S_CONTACT contact  ON contact.PAR_ROW_ID = c.CST_CON_ID
   LEFT OUTER JOIN dbo.S_USER    usr      ON usr.PAR_ROW_ID     = c.OWNER_EMP_ID
   LEFT OUTER JOIN dbo.S_CONTACT c2       ON c2.PAR_ROW_ID      = c.LAST_UPD_BY
```

The following is the same query formatted as a single line that you can cut and paste into the Oracle SES Administration GUI:

```
SELECT
'callcenter_enu/start.swe?SWECmd=GotoView&SWEView=All+Service+Request+across+Organ
```

```
izations&SWERF=1&SWEHo=HostName&SWEBU=1&SWEApplet0=Service+Request+List+Applet&SWE
RowId0='+c.ROW_ID URL, 'US' LANG, c.LAST_UPD LASTMODIFIEDDATE, 'text/html'
CONTENTTYPE, c.ROW_ID "KEY", coalesce('<b>SR Number:</b>
'+c.SR_NUM,'<null>')+coalesce(',<br><b>Summary:</b>
'+c.SR_TITLE,'<null>')+coalesce(',<br><b>Status:</b>
'+c.SR_STAT_ID,'<null>')+coalesce(',<br><b>Area:</b>
'+c.SR_AREA,'<null>')+coalesce(',<br><b>Subarea:</b>
'+c.SR_SUB_AREA,'<null>')+coalesce(',<br><b>Resolution:</b>
'+c.RESOLUTION_CD,'<null>') SOLUTION, c.DESC_TEXT description, c.BU_ID
visibilityid, c.ROW_ID sblrowid, c.CREATED created_on, c.CREATED_BY createdby,
c.SR_TITLE summary, a.NAME orgName, c.SR_AREA code01, a.OU_NUM csn,
contact.FST_NAME firstName, contact.LAST_NAME lastName, c.SR_NUM title,
c.SR_STAT_ID status, c.SR_SUB_AREA code02, usr.LOGIN owner, 'Service Request'
sblbctype FROM dbo.S_ORG_EXT a INNER JOIN dbo.S_SRV_REQ c ON a.PAR_ROW_ID=
c.CST_OU_ID LEFT OUTER JOIN dbo.S_CONTACT contact ON contact.PAR_ROW_ID
=c.CST_CON_ID LEFT OUTER JOIN dbo.S_USER usr ON usr.PAR_ROW_ID = c.OWNER_EMP_ID
LEFT OUTER JOIN dbo.S_CONTACT c2 ON c2.PAR_ROW_ID=c.LAST_UPD_BY
```

Contacts

```
SELECT
   'callcenter_enu/start.swe?SWECmd=GotoView
      &SWEView=All+Contacts+across+Organizations
      &SWERF=1
      &SWEHo=HostName
      &SWEBU=1
      &SWEApplet0=Contact+List+Applet
      &SWERowId0='+c.PAR_ROW_ID URL,
   'US' LANG,
   c.LAST_UPD LASTMODIFIEDDATE,
   'text/html' CONTENTTYPE,
   c.PAR_ROW_ID "KEY",
   '<b>Name: </b>'
      +coalesce(c.LAST_NAME,'<null>')+' '
      +coalesce(c.FST_NAME,'<null>')
      +coalesce(',<br><b>Phone No.:</b> '+c.WORK_PH_NUM,'<null>')
      +coalesce(',<br><b>E-Mail ID:</b> '+ c.EMAIL_ADDR,'<null>') SOLUTION,
   t.PERS_AGENDA agenda,
   c.PAR_ROW_ID sblrowid,
   c.CREATED created_on,
   c.CREATED_BY createdby,
   a.NAME+'#'+c.JOB_TITLE PATH,
   c.LAST_NAME+' '+c.FST_NAME title,
   c.LAST_NAME lastName,
   c.FST_NAME firstName,
   c.EMP_ID owner,
   c.EMAIL_ADDR emailID,
   c.WORK_PH_NUM phoneNumber02,
   'Contacts' sblbctype,
   t.ACCOMPLISH summary,
   addr.ZIPCODE zipcode,
   addr.COUNTRY country,
   party.NAME name,
   addr.ADDR street,
   c.BU_ID visibilityid
FROM
   dbo.S_PARTY party
   INNER JOIN      dbo.S_CONTACT c     ON party.ROW_ID  = c.PAR_ROW_ID
   INNER JOIN      dbo.S_POSTN_CON T3  ON c.PR_POSTN_ID = T3.POSTN_ID
      AND c.ROW_ID = T3.CON_ID
```

```
   INNER JOIN      dbo.S_PARTY T4      ON T3.POSTN_ID   = T4.ROW_ID
   LEFT OUTER JOIN dbo.S_ORG_EXT a     ON a.PAR_ROW_ID  = c.PR_DEPT_OU_ID
   LEFT OUTER JOIN dbo.S_ADDR_ORG addr ON addr.ROW_ID   = c.PR_PER_ADDR_ID
   LEFT OUTER JOIN dbo.S_CONTACT_T t   ON c.ROW_ID      = t.PAR_ROW_ID
   LEFT OUTER join dbo.S_CONTACT c2    ON c2.ROW_ID     = c.LAST_UPD_BY
WHERE
   (c.PRIV_FLG = 'N')
```

The following is the same query formatted as a single line that you can cut and paste into the Oracle SES Administration GUI:

```
SELECT
'callcenter_enu/start.swe?SWECmd=GotoView&SWEView=All+Contacts+across+Organization
s&SWERF=1&SWEHo=HostName&SWEBU=1&SWEApplet0=Contact+List+Applet&SWERowId0='+c.PAR_
ROW_ID URL, 'US' LANG, c.LAST_UPD LASTMODIFIEDDATE, 'text/html' CONTENTTYPE,
c.PAR_ROW_ID "KEY", '<b>Name: </b>'+coalesce(c.LAST_NAME,'<null>')+'
'+coalesce(c.FST_NAME,'<null>')+coalesce(',<br><b>Phone No.:</b>
'+c.WORK_PH_NUM,'<null>')+coalesce(',<br><b>E-Mail ID:</b> '+
c.EMAIL_ADDR,'<null>') SOLUTION, t.PERS_AGENDA agenda, c.PAR_ROW_ID sblrowid,
c.CREATED created_on, c.CREATED_BY createdby, a.NAME+'#'+c.JOB_TITLE PATH,
c.LAST_NAME+' '+c.FST_NAME title, c.LAST_NAME lastName, c.FST_NAME firstName,
c.EMP_ID owner, c.EMAIL_ADDR emailID, c.WORK_PH_NUM phoneNumber02, 'Contacts'
sblbctype, t.ACCOMPLISH summary, addr.ZIPCODE zipcode, addr.COUNTRY country,
party.NAME name, addr.ADDR street, c.BU_ID visibilityid FROM dbo.S_PARTY party
INNER JOIN dbo.S_CONTACT c ON party.ROW_ID = c.PAR_ROW_ID INNER JOIN
dbo.S_POSTN_CON T3 ON c.PR_POSTN_ID = T3.POSTN_ID AND c.ROW_ID = T3.CON_ID INNER
JOIN dbo.S_PARTY T4 ON T3.POSTN_ID = T4.ROW_ID LEFT OUTER JOIN dbo.S_ORG_EXT a ON
a.PAR_ROW_ID = c.PR_DEPT_OU_ID LEFT OUTER JOIN dbo.S_ADDR_ORG addr ON addr.ROW_ID
= c.PR_PER_ADDR_ID LEFT OUTER JOIN dbo.S_CONTACT_T t ON c.ROW_ID=t.PAR_ROW_ID LEFT
OUTER join dbo.S_CONTACT c2 ON c2.ROW_ID=c.LAST_UPD_BY WHERE (c.PRIV_FLG = 'N')
```

Activity

For the queries shown in this section, before starting the crawl, you must replace *HostName* with the name or IP address of the Siebel host computer.

The following query can be used with Oracle Database:

```
SELECT
   'callcenter_enu/start.swe?SWECmd=GotoView
      &SWEView=All+Activity+List+View
      &SWERF=1
      &SWEHo=HostName
      &SWEBU=1
      &SWEApplet0=Activity+Form+Applet
      &SWERowId0='+T1.ROW_ID URL,
   'US' LANG,
   T1.ROW_ID "KEY",
   T1.LAST_UPD LASTMODIFIEDDATE,
   T1.NAME title,
   coalesce('Activity Name:'+T1.NAME,'<null>')
      +coalesce(', Activity Type:'+T1.TODO_CD,'<null>')
      +coalesce(', Creation Date:'+convert(varchar,T1.CREATED,103),'<null>')
      +coalesce(', Activity Status:'+T1.EVT_STAT_CD,'<null>')
      +coalesce(', ActivityPriority:'+T1.EVT_PRIORITY_CD,'<null>')
      +coalesce(', Activity Owner: '+T1.OWNER_LOGIN,'<null>') CONTENT,
   T1.NAME ActivityName,
   T1.CREATED CreatedOn,
   T1.CREATED_BY CreatedBy,
   T1.LAST_UPD_BY LastUpdatedBy,
   T1.TODO_CD ActivityType,
```

```
   T1.EVT_STAT_CD ActivityStatus,
   T1.EVT_PRIORITY_CD ActivityPriority,
   T1.TARGET_OU_ID Organization,
   T1.OWNER_LOGIN visibilityid,
   T1.OWNER_LOGIN ActivityOwner,
   T1.ROW_ID SBLROWID,
   T1.TODO_PLAN_START_DT StartDate,
   T1.TODO_PLAN_END_DT EndDate,
   T1.TODO_DUE_DT DueDate
FROM
   dbo.S_EVT_ACT T1
   INNER JOIN      dbo.S_ORG_EXT T2  ON T2.PAR_ROW_ID = T1.TARGET_OU_ID
   LEFT OUTER JOIN dbo.S_USER    usr ON usr.LOGIN     = T1.OWNER_LOGIN
WHERE
   ((T1.APPT_REPT_REPL_CD IS NULL)
   AND ((T1.TEMPLATE_FLG != 'Y')
   OR (T1.TEMPLATE_FLG = NULL)))
```

The following is the same query for use with Oracle Database, but formatted as a single line that you can cut and paste into the Oracle SES Administration GUI:

```
SELECT
'callcenter_enu/start.swe?SWECmd=GotoView&SWEView=All+Activity+List+View&SWERF=1&S
WEHo=HostName&SWEBU=1&SWEApplet0=Activity+Form+Applet&SWERowId0='+T1.ROW_ID
URL,'US' LANG,T1.ROW_ID "KEY", T1.LAST_UPD LASTMODIFIEDDATE, T1.NAME title ,
coalesce('Activity Name:'+T1.NAME,'<null>')
+coalesce(', Activity Type:'+T1.TODO_CD,'<null>')
+coalesce(', Creation Date:'+convert(varchar,T1.CREATED,103),'<null>')
+coalesce(', Activity Status:'+T1.EVT_STAT_CD,'<null>')
+coalesce(', ActivityPriority:'+T1.EVT_PRIORITY_CD,'<null>')
+coalesce(', Activity Owner: '+T1.OWNER_LOGIN,'<null>') CONTENT,T1.NAME
ActivityName,T1.CREATED CreatedOn, T1.CREATED_BY CreatedBy,T1.LAST_UPD_BY
LastUpdatedBy, T1.TODO_CD ActivityType,  T1.EVT_STAT_CD
ActivityStatus,T1.EVT_PRIORITY_CD ActivityPriority, T1.TARGET_OU_ID
Organization,T1.OWNER_LOGIN visibilityid, T1.OWNER_LOGIN ActivityOwner,T1.ROW_ID
SBLROWID,T1.TODO_PLAN_START_DT StartDate, T1.TODO_PLAN_END_DT EndDate,
T1.TODO_DUE_DT DueDate FROM dbo.S_EVT_ACT T1 INNER JOIN dbo.S_ORG_EXT T2 ON
T2.PAR_ROW_ID = T1.TARGET_OU_ID LEFT OUTER JOIN dbo.S_USER usr ON usr.LOGIN =
T1.OWNER_LOGIN WHERE ((T1.APPT_REPT_REPL_CD IS NULL) AND ((T1.TEMPLATE_FLG != 'Y')
OR (T1.TEMPLATE_FLG = NULL)))
```

The following query can be used with SQL Server:

```
SELECT
   'callcenter_enu/start.swe?SWECmd=GotoView
      &SWEView=All+Activity+List+View
      &SWERF=1
      &SWEHo=HostName
      &SWEBU=1
      &SWEApplet0=Activity+Form+Applet
      &SWERowId0='+T1.ROW_ID URL,
   'US' LANG,
   T1.ROW_ID "KEY",
   T1.LAST_UPD LASTMODIFIEDDATE,
   T1.NAME title,
   coalesce('Activity Name:'+T1.NAME,'<null>')
      +coalesce(', Activity Type:'+T1.TODO_CD,'<null>')
      +coalesce(', Creation Date:'+convert(varchar,T1.CREATED,103),'<null>')
      +coalesce(', Activity Status:'+T1.EVT_STAT_CD,'<null>')
      +coalesce(', ActivityPriority:'+T1.EVT_PRIORITY_CD,'<null>')
      +coalesce(', Activity Owner: '+T1.OWNER_LOGIN,'<null>') CONTENT,
```

```
   T1.NAME ActivityName,
   T1.CREATED CreatedOn,
   T1.CREATED_BY CreatedBy,
   T1.LAST_UPD_BY LastUpdatedBy,
   T1.TODO_CD ActivityType,
   T1.EVT_STAT_CD ActivityStatus,
   T1.EVT_PRIORITY_CD ActivityPriority,
   T1.TARGET_OU_ID Organization,
   T1.OWNER_LOGIN visibilityid,
   T1.OWNER_LOGIN ActivityOwner,
   T1.ROW_ID SBLROWID,
   T1.TODO_PLAN_START_DT StartDate,
   T1.TODO_PLAN_END_DT EndDate,
   T1.TODO_DUE_DT DueDate
FROM
   dbo.S_EVT_ACT T1
   INNER JOIN dbo.S_ORG_EXT   T2  ON T2.PAR_ROW_ID = T1.TARGET_OU_ID
   LEFT OUTER JOIN dbo.S_USER usr ON usr.LOGIN     = T1.OWNER_LOGIN
WHERE
   ((T1.APPT_REPT_REPL_CD IS NULL)
   AND ((T1.TEMPLATE_FLG != 'Y')
   OR (T1.TEMPLATE_FLG = NULL)))
```

The following is the same query for use with SQL Server, but formatted as a single line that you can cut and paste into the Oracle SES Administration GUI:

```
SELECT
'callcenter_enu/start.swe?SWECmd=GotoView&SWEView=All+Activity+List+View&SWERF=1&S
WEHo=HostName&SWEBU=1&SWEApplet0=Activity+Form+Applet&SWERowId0='+T1.ROW_ID
URL,'US' LANG,T1.ROW_ID "KEY", T1.LAST_UPD LASTMODIFIEDDATE, T1.NAME title,
coalesce('Activity Name:'+T1.NAME,'<null>')
+coalesce(', Activity Type:'+T1.TODO_CD,'<null>')
+coalesce(', Creation Date:'+convert(varchar,T1.CREATED,103),'<null>')
+coalesce(', Activity Status:'+T1.EVT_STAT_CD,'<null>')
+coalesce(', ActivityPriority:'+T1.EVT_PRIORITY_CD,'<null>')
+coalesce(', Activity Owner: '+T1.OWNER_LOGIN,'<null>') CONTENT, T1.NAME
ActivityName,T1.CREATED CreatedOn, T1.CREATED_BY CreatedBy,T1.LAST_UPD_BY
LastUpdatedBy, T1.TODO_CD ActivityType,  T1.EVT_STAT_CD
ActivityStatus,T1.EVT_PRIORITY_CD ActivityPriority, T1.TARGET_OU_ID
Organization,T1.OWNER_LOGIN visibilityid, T1.OWNER_LOGIN ActivityOwner,T1.ROW_ID
SBLROWID, T1.TODO_PLAN_START_DT StartDate, T1.TODO_PLAN_END_DT EndDate,
T1.TODO_DUE_DT DueDate  FROM dbo.S_EVT_ACT T1 INNER JOIN dbo.S_ORG_EXT T2 ON
T2.PAR_ROW_ID = T1.TARGET_OU_ID LEFT OUTER JOIN dbo.S_USER usr ON usr.LOGIN =
T1.OWNER_LOGIN WHERE ((T1.APPT_REPT_REPL_CD IS NULL) AND ((T1.TEMPLATE_FLG != 'Y')
OR (T1.TEMPLATE_FLG = NULL)))
```

Activity Attachment

For the queries shown in this section, before starting the crawl, substitute the following with the appropriate values:

- *Siebel_Host*: Siebel host name or IP address.
- *Folder_Path*: Absolute path to the folder containing attachment files on the Siebel host.
- *Attachment_Decompressor*: Choose the appropriate value:
  - When crawling from Linux or from different Windows domain, use RMI:

    rmi://*Siebel_Host*[:*RMI_Port*]/Decompressor

For example:

```
rmi://ses5-pc.example.com:2155/Decompressor
```

RMI_Port is the RMI port number. Specify it only when the RMI registry in the Siebel host is running on a port other than the default port 1099.

The RMI Decompression Server must be running.

– When crawling from Windows in the same domain, use Windows Native Access:

```
\\Siebel_Host\sseunzip_path
```

For example:

```
\\ses5-pc.example.com\BIN\sseunzip.exe
```

sseunzip_path is the path to the shared `sseunzip.exe` utility.

Use this query with Oracle Database:

```
SELECT
   'callcenter_enu/start.swe?SWECmd=GotoView
      &SWEView=Activity+Attachment+View
      &SWERF=1
      &SWEHo=Host_name
      &SWEBU=1
      &SWEApplet0=Activity+Form+Applet
      &SWERowId0='||T2.ROW_ID||'
      &SWEApplet1=Activity+Attachment+Applet
      &SWERowId1='||T1.ROW_ID URL,
   'US' LANG,
   T1.ROW_ID KEY,
   GREATEST(T1.LAST_UPD, T2.LAST_UPD) LASTMODIFIEDDATE,
   DECODE(T1.FILE_EXT, 'ppt', 'application/vnd.ms-powerpoint',
      'doc', 'application/msword', 'html', 'text/html',
      'txt', 'text/plain', 'pdf', 'application/pdf','xls',
      'application/vnd.ms-excel') CONTENTTYPE,
      Attachment_Decompressor
      Folder_Path\S_ACTIVITY_ATT_'+T1.ROW_ID+'_'
      +FILE_REV_NUM+'.SAF' ATTACHMENT_LINK,
   T1.FILE_NAME CONTENT,
   T1.FILE_NAME TITLE,
   T1.CREATED CreatedOn,
   T1.CREATED_BY CreatedBy,
   T1.LAST_UPD_BY LastUpdatedBy,
   T1.FILE_SRC_TYPE FileType,
   T1.FILE_DATE FileDate,
   replace(T1.FILE_NAME,' ','_') FileName,
   T1.FILE_SIZE FileSize,
   T1.FILE_EXT FileExtension,
   T1.COMMENTS AttachmentComment,
   'Activity Attachment' sblbctype,
   T2.OWNER_LOGIN ActivityOwner,
   T2.OWNER_LOGIN visibilityid,
   T2.TARGET_OU_ID Organization,
   T2.ROW_ID SBLROWID
FROM
   dbo.S_ACTIVITY_ATT T1
   INNER JOIN      dbo.S_EVT_ACT T2  ON T1.PAR_ROW_ID = T2.ROW_ID
   LEFT OUTER JOIN dbo.S_USER    usr ON usr.LOGIN     = T2.OWNER_LOGIN
```

The following is the same query formatted as a single line that you can cut and paste into the Oracle SES Administration GUI:

```
SELECT
'callcenter_enu/start.swe?SWECmd=GotoView&SWEView=Activity+Attachment+View&SWERF=1
&SWEHo=Host_name&SWEBU=1&SWEApplet0=Activity+Form+Applet&SWERowId0='||T2.ROW_ID||'
&SWEApplet1=Activity+Attachment+Applet&SWERowId1='||T1.ROW_ID URL,'US'
LANG,T1.ROW_ID KEY,GREATEST(T1.LAST_UPD, T2.LAST_UPD) LASTMODIFIEDDATE,
decode(T1.FILE_EXT, 'ppt', 'application/vnd.ms-powerpoint', 'doc',
'application/msword', 'html', 'text/html', 'txt', 'text/plain', 'pdf',
'application/pdf','xls', 'application/vnd.ms-excel') CONTENTTYPE,
'Attachment_Decompressor
Folder_Path\S_ACTIVITY_ATT_'+T1.ROW_ID+'_'+FILE_REV_NUM+'.SAF'
ATTACHMENT_LINK,T1.FILE_NAME CONTENT, T1.FILE_NAME TITLE, T1.CREATED
CreatedOn,T1.CREATED_BY CreatedBy,T1.LAST_UPD_BY LastUpdatedBy,T1.FILE_SRC_TYPE
FileType,T1.FILE_DATE FileDate,replace(T1.FILE_NAME,' ','_') FileName,T1.FILE_SIZE
FileSize, T1.FILE_EXT FileExtension, T1.COMMENTS AttachmentComment,'Activity
Attachment' sblbctype,T2.OWNER_LOGIN ActivityOwner,T2.OWNER_LOGIN
visibilityid,T2.TARGET_OU_ID Organization, T2.ROW_ID SBLROWID FROM
dbo.S_ACTIVITY_ATT T1 INNER JOIN dbo.S_EVT_ACT T2 ON T1.PAR_ROW_ID=T2.ROW_ID LEFT
OUTER JOIN dbo.S_USER usr ON usr.LOGIN = T2.OWNER_LOGIN
```

Use this query with SQL Server:

```
SELECT
   'callcenter_enu/start.swe?SWECmd=GotoView
      &SWEView=Activity+Attachment+View
      &SWERF=1
      &SWEHo=Host_name
      &SWEBU=1
      &SWEApplet0=Activity+Form+Applet
      &SWERowId0='+T2.ROW_ID+'
      &SWEApplet1=Activity+Attachment+Applet
      &SWERowId1='+T1.ROW_ID URL,
   'US' LANG,
   T1.ROW_ID "KEY",
   CASE
      WHEN (DATEDIFF(second, T1.LAST_UPD, T2.LAST_UPD)> 0) THEN T2.LAST_UPD
      ELSE T1.LAST_UPD
   END
LASTMODIFIEDDATE,
   Attachment_Decompressor Folder_Path
      \S_ACTIVITY_ATT_'+T1.ROW_ID+'_'+FILE_REV_NUM+'.SAF' ATTACHMENT_LINK,
   T1.FILE_NAME CONTENT,
   T1.FILE_NAME TITLE,
   T1.CREATED CreatedOn,
   T1.CREATED_BY CreatedBy,
   T1.LAST_UPD_BY LastUpdatedBy,
   T1.FILE_SRC_TYPE FileType,
   T1.FILE_DATE FileDate,
   REPLACE(T1.FILE_NAME,' ','_') FileName,
   T1.FILE_SIZE FileSize,
   T1.FILE_EXT FileExtension,
   T1.COMMENTS AttachmentComment,
   'Activity Attachment' sblbctype,
   T2.OWNER_LOGIN ActivityOwner,
   T2.OWNER_LOGIN visibilityid,
   T2.TARGET_OU_ID Organization,
   T2.ROW_ID SBLROWID,
   CASE T1.FILE_EXT
      WHEN 'ppt'  THEN 'application/vnd.ms-powerpoint'
```

```
      WHEN 'doc'  THEN 'application/msword'
      WHEN 'html' THEN 'text/html' when 'txt' then 'text/plain'
      WHEN 'pdf'  THEN 'application/pdf'
      WHEN 'xls'  THEN 'application/vnd.ms-excel'
   END
   CONTENTTYPE
FROM
   dbo.S_ACTIVITY_ATT T1
   INNER JOIN      dbo.S_EVT_ACT T2  ON T1.PAR_ROW_ID = T2.ROW_ID
   LEFT OUTER JOIN dbo.S_USER    usr ON usr.LOGIN     = T2.OWNER_LOGIN
```

The following is the same query formatted as a single line that you can cut and paste into the Oracle SES Administration GUI:

```
SELECT
'callcenter_enu/start.swe?SWECmd=GotoView&SWEView=Activity+Attachment+View&SWERF=1
&SWEHo=Host_name&SWEBU=1&SWEApplet0=Activity+Form+Applet&SWERowId0='+T2.ROW_ID+'&S
WEApplet1=Activity+Attachment+Applet&SWERowId1='+T1.ROW_ID URL,'US' LANG,T1.ROW_ID
"KEY",case when (datediff(second, T1.LAST_UPD, T2.LAST_UPD)> 0) then T2.LAST_UPD
else T1.LAST_UPD end LASTMODIFIEDDATE,'Attachment_Decompressor
Folder_Path\S_ACTIVITY_ATT_'+T1.ROW_ID+'_'+FILE_REV_NUM+'.SAF' ATTACHMENT_LINK,
T1.FILE_NAME CONTENT, T1.FILE_NAME TITLE, T1.CREATED CreatedOn,T1.CREATED_BY
CreatedBy, T1.LAST_UPD_BY LastUpdatedBy,T1.FILE_SRC_TYPE FileType,T1.FILE_DATE
FileDate,replace(T1.FILE_NAME,' ','_') FileName,T1.FILE_SIZE FileSize, T1.FILE_EXT
FileExtension, T1.COMMENTS AttachmentComment,'Activity Attachment'
sblbctype,T2.OWNER_LOGIN ActivityOwner,T2.OWNER_LOGIN visibilityid,T2.TARGET_OU_ID
Organization, T2.ROW_ID SBLROWID,  case T1.FILE_EXT when 'ppt' then
'application/vnd.ms-powerpoint' when 'doc' then 'application/msword' when 'html'
then 'text/html' when 'txt' then 'text/plain' when 'pdf' then 'application/pdf'
when 'xls' then 'application/vnd.ms-excel' end CONTENTTYPE FROM dbo.S_ACTIVITY_ATT
T1 INNER JOIN dbo.S_EVT_ACT T2 ON T1.PAR_ROW_ID=T2.ROW_ID LEFT OUTER JOIN
dbo.S_USER usr ON usr.LOGIN = T2.OWNER_LOGIN
```

Setting Up Siebel 8 Sources

The Siebel 8 connector uses the Oracle SES XML connector framework, where searching is based on Siebel data available as XML feeds.

See Also:

- "Overview of XML Connector Framework" on page 3-9
- Appendix A in the *Siebel Search Administration Guide* for searchable business components:

  http://download.oracle.com/docs/cd/B40099_01/80Siebel_HTML/books/Search/SearchTOC.html

- Siebel documentation on Oracle Technology Network (OTN) for information about supported Siebel modules:

  http://www.oracle.com/technetwork/indexes/documentation/index.html

To activate an identity plug-in for Siebel 8 sources:

1. On the Global Settings page, select **Identity Management Setup**.
2. Select **Siebel 8** and click **Activate** to display the Activate Identity Plug-in page.
3. Enter values for the parameters as described in Table 8–16. Obtain these values from the Siebel administrator.

4. Click **Finish**.

Table 8–16 Siebel 8 Identity Management Parameters

Parameter	Value
Siebel 8 Authentication Web Service Endpoint	HTTP endpoint of the Siebel Web service that provides the authentication service
Siebel 8 Validation Web Service Endpoint	HTTP endpoint of the Siebel Web service that provides the user validation service
User ID	Administrator ID for accessing the user validation service
Password	User password.

To create a Siebel 8 source:

1. Activate an identity plug-in as described in the previous procedure.
2. On the Home page, select the **Sources** secondary tab.
3. Select **Siebel 8** from the Source Types list, and click **Create**.
4. Enter the source parameters as described in Table 8–17.
5. Click **Next**.
6. Click **Get Parameters** to obtain a list of parameters for the authorization manager plug-in.
7. Enter the values for the authorization manager plug-in parameters as described in Table 8–18.
8. Click **Create**.

Table 8–17 Siebel 8 Source Parameters

Parameter	Value
Configuration URL	FILE protocol address of the XML configuration file providing details about the source, such as the data feed type, location, security attributes, and so on. Obtain this file from Siebel administrator and save it on the same computer as Oracle SES. Enter the configuration URL in the form: `file://localhost/`*`config_path`* where *`config_path`* is the absolute path to the configuration file. For example: `file://localhost/private/oracle/config.xml/`
Authentication Type	Standard Java authentication type used by the application serving the control and data feed. Leave this parameter blank as the feeds are accessed over file or FTP protocols.
User ID	User ID to login to the FTP server and access the data feeds. The access details of the data feed are specified in the configuration file. The user id can be obtained from Siebel administrator.
Password	Password for **User ID**.
Realm	The realm of the application serving the feeds. Leave this parameter blank since the feeds are accessed over file or FTP.
Oracle SSO Login URL	URL that protects all OracleAS Single Sign-on applications. Leave this parameter blank.
Oracle SSO Action URL	URL that authenticates OracleAS Single Sign-on user credentials. Leave the parameter blank.

Table 8–17 (Cont.) Siebel 8 Source Parameters

Parameter	Value
Scratch Directory	A directory on the same computer as Oracle SES, where the status logs are created temporarily.
Maximum number of connection attempts	Maximum number of attempts to connect to the target server to access the data feed.

Table 8–18 Siebel 8 Authorization Parameters

Parameter	Value
Siebel 8 authorization Web service endpoint	Webs service endpoint of the Siebel Web service that provides the authorization service.
User ID	Administrator ID for accessing the authorization service.
Password	Password for **User ID**.

Part III

Advanced Topics

This part provides information for experienced administrators. It contains the following chapters:

- Chapter 9, "Security in Oracle Secure Enterprise Search"
- Chapter 10, "Administering Oracle SES Instances"
- Chapter 11, "Oracle Secure Enterprise Search APIs"

9

Security in Oracle Secure Enterprise Search

This chapter describes the architecture and configuration for Oracle SES security model. It contains the following topics:

- Overview of Oracle Secure Enterprise Search Security
- Enabling Secure Search
- Configuring Secure Search with OracleAS Single Sign-On
- Configuring Secure Search with Oracle Access Manager Single Sign-On
- SSL and HTTPS Support in Oracle Secure Enterprise Search
- Changing the Master Encryption Key

Overview of Oracle Secure Enterprise Search Security

This section describes the Oracle SES security model. It contains the following topics:

- Oracle Secure Enterprise Search Security Model
- Changing the Administration Password
- Authentication and Authorization
- Authentication Methods

Oracle Secure Enterprise Search Security Model

Oracle SES provides access to a variety of content repositories through a single gateway. Each external repository has its own security model that determines whether a particular user can access a particular document. You must carefully consider all aspects of security in Oracle SES to respect the security of documents coming from multiple data repositories.

Oracle SES uses the following security services in its security model:

- **Identity plug-ins** can obtain user and group information directly from any identity management system. An identity plug-in is Java code between Oracle SES and an identity management system, allowing Oracle SES to read user and group information.
- **Secure Socket Layers** (SSL) is the industry standard protocol for managing the security of message transmission on the Internet. This is used for securing RMI connections, HTTPS crawling, and secure JDBC.

Connecting to the Oracle SES server using SQL*Plus, except as documented in this Guide, is not supported. Changing the Oracle SES server directly using SQL and

modifying initialization parameter files is not supported. User management, including password changes, should only be done using the Oracle SES Administration GUI or the Administration API.

As an additional security measure, Oracle SES is configured to reject connection requests using SQL*Plus from remote hosts. The only protocols supported for connection to Oracle SES from remote hosts are HTTP, HTTPS, and AJP13.

Changing the Administration Password

The administrator's user name for Oracle SES is SEARCHSYS. You can change the password specified during installation. A password must consist of at least eight ASCII characters, and contain at least one numeric and one alphabetic character. Note that the password length cannot exceed 30 characters or contain double-byte characters.

These are the basic steps for changing the SEARCHSYS password. More detailed instructions for these basic steps are provided afterward.

1. In the Oracle SES Administration GUI, change the password on the Global Settings - Change Password page.
2. In the WebLogic console, change the connection pool password.
3. Using the `wlst` script, change the password entry in the Credential Storage Framework.

Following are detailed instructions for steps 2 and 3.

To change the connection-pool password:

1. Log in to the WebLogic console, as described in "Accessing the Oracle WebLogic Server Administration Console" on page 10-21.
2. In the left panel under Change Center, click **Lock & Edit**.
3. Change the password for SearchAdminDS:
   a. In the left panel under Domain Structure, select **common domain** - **Services** - **Data Sources**. The Configuration tab of the Settings for SearchAdminDS appears in the main panel.
   b. In the Name column of the Data Sources table, click **SearchAdminDS**. The General tab of the Settings for SearchAdminDS is displayed.
   c. Select the **Connection Pool** tab.
   d. Enter the new password in the Password and Confirm Password fields.
   e. Click **Save**.
4. Change the password for SearchQueryDS:
   a. In the left panel under Domain Structure, select **common domain** - **Services** - **Data Sources**. The Configuration tab of the Settings for SearchQueryDS appears in the main panel.
   b. In the Name column of the Data Sources table, click **SearchQueryDS**. The General tab of the Settings for SearchQueryDS is displayed.
   c. Select the **Connection Pool** tab.
   d. Enter the new password in the Password and Confirm Password fields.
   e. Click **Save**.

5. In the left panel under Change Center, click **Activate Changes**.
6. Log out of the Weblogic console if desired.
7. Modify the password entry in the Credential Storage Framework (CSF), as described in the following procedure.

To change the password in the Credential Storage Framework:

1. Open a connection to the computer where Oracle Fusion Middleware is installed.
2. Go to *MW_HOME*/oracle_common/common/bin.
3. Start the WebLogic Server Administration Scripting Tool:
   - For Linux, enter `wlst.sh`.
   - For Windows, enter `wlst.cmd`.
4. Enter this command at the `wls/offline>` prompt:

   ```
   connect ()
   ```

5. Enter your WebLogic user name in response to the user-name prompt.
6. Enter your WebLogic password in response to the password prompt.
7. Enter the WebLogic server *hostname:port* in response to the server URL prompt.
8. Enter this command in response to the `wls:`*domain_name*`/serverConfig>` prompt, replacing *password* with the SEARCHSYS password. The Oracle SES credentials for connecting to Oracle Database are obtained from a CSF map named `oracle.apps.security` with a key named `FUSION_APPS_ECSF_SES_ADMIN-KEY`.

   ```
   updateCred(map="oracle.apps.security",key="FUSION_APPS_ECSF_SES_ADMIN-KEY",
   user="searchsys", password="password")
   ```

   See Also: *Oracle Fusion Middleware WebLogic Scripting Tool Command Reference*

Example 9–1 Modifying the CSF Entry

This interactive session shows the steps for modifying the password entry in the credential store. Be sure to enter your own user name, password, and URL.

```
> wlst.sh
     .
     .
     .
Initializing WebLogic Scripting Tool (WLST) ...

Welcome to WebLogic Server Administration Scripting Shell

Type help() for help on available commands

wls:/offline> connect()
Please enter your username : weblogic
Please enter your password : ********
Please enter your server URL [t3://localhost:7001]: example:9999
Connecting to t3:example:9999 with userid weblogic ...
Successfully connected to Admin Server 'AdminServer' that belongs to domain
'base_domain1'.
     .
```

```
      .
      .
wls:/base_domain1/serverConfig> updateCred(map ="oracle.apps.security",
key="FUSION_APPS_ECSF_SES_ADMIN-KEY", user="searchsys", password="password")
Location changed to domainRuntime tree...
```

Temporary Passwords

For added security, a temporary password feature is provided. You can enter login credentials for use by the crawler when creating table sources, e-mail, OracleAS Portal, or Web sources. For Web sources, authentication can be performed with HTTP authentication, HTML forms, and OracleAS Single Sign-On.

To use the temporary password feature:

- Select the **Delete Passwords After Crawl** option in the Oracle SES Administration GUI when creating or editing a source.

This option is not available when self service for Web sources is enabled.

If a source has the **Delete Passwords after Crawl** option enabled, then you are prompted for all required passwords whenever the schedule for that source starts. You must start these schedules manually, which enables you to respond to the prompts. The supplied passwords are removed immediately after the schedule completes.

Authentication and Authorization

Oracle SES security is implemented at two levels: user authentication and user authorization.

This section contains the following topics:

- About Oracle SES Authentication
- Restrictions on Changing the ACL Policy
- Activating an Identity Plug-in
- Re-registering Pre-Installed Identity Plug-ins
- Restrictions on Changing the Identity Plug-in

About Oracle SES Authentication

User authentication identifies a user through an identity management system. You can register an identity plug-in to any identity management system; Oracle SES provides registered identity plug-ins for many identity management systems. The plug-in that you activate is responsible for all authentication and validation activity in Oracle SES. Activation is performed on the Global Settings - Identity Management Setup page.

Security filter configuration for the identity plug-in is performed on the Global Settings - Query Configuration page. A login does not force a refresh of the user's security filter. For a query request, Oracle SES checks the timestamp of an existing cached security filter and refreshes it when the specified life span has expired. The default latency is 60 minutes.

About Oracle SES User Authorization

User authorization determines whether a user can access information about a particular item in the result list. It can be implemented in two layers.

The first layer uses access control lists (ACLs). An ACL lists the users or groups of users that have access to the document. The ACL can be assigned by the administrator

to an entire source through the Oracle SES Administration GUI (*source-level ACLs*), or it can be provided by the source itself for each document (*document-level ACLs*).

The second layer uses a Java class to dynamically filter documents at search time (*query-time authorization*).

Oracle SES can use the following types of ACL policies:

- **Source-level ACLs**: An individual source can be protected by a single ACL, which governs access to every document in that source. These ACLs are defined on the Home - Sources - Authorization page.
- **Document-level ACLs**: Oracle SES provides mapped security to repositories by retrieving the ACL for each document at the time of crawling and indexing. At crawl time, the ACL for each document is passed to the crawler along with the document content, and the ACL is stored in the index. Currently Oracle SES supports document-level ACLs for user-defined sources and OracleAS Portal sources. (The ACL policy is Documents Controlled by the Source.) With user-defined sources, ACLs are returned by the crawler plug-in implemented by the user. With OracleAS Portal sources, ACLs are returned by the OracleAS Portal server. At search time, Oracle SES does not need any connection with the repository to validate access privileges.

For both source-level ACLs and document-level ACLs, all users and roles defined in the ACLs must exist in the identity plug-in.

User names are not case sensitive.

Table 9–1 identifies when documents are visible with the document ACL types supported in Oracle SES:

Table 9–1 Document ACL Types in Oracle SES Security Model

Document ACL Type	Public User	Authenticated User	Authenticated User with Allow Permission to Document	Authenticated User with Deny Permission to Document
No ACL	document visible	document visible	N/A	N/A
Deny Permission Only	--	--	N/A	--
Allow Permission Only	--	--	document visible	N/A
Deny with Allow Permissions	--	--	document visible	--

Table 9–2 compares the document-level user authorization methods in Oracle SES.

Table 9–2 User Authorization Methods in Oracle Secure Enterprise Search

Method	How Authorization is Determined	Advantages	Disadvantages
ACLs	The ACL is supplied by a crawler plug-in or an OracleAS Portal server.	Faster secure search performance. No additional programming is required for ACL-based OracleAS Portal security. (If implementing a crawler plug-in, then some additional work is necessary to supply ACLs.)	ACLs are static: they are updated only when crawling the source repository or when the administrator changes Oracle SES ACLs in the Oracle SES Administration GUI
Query-time Authorization	`ResultFilterPlugin` Java class.	Dynamic authorization. Reflects real-time user access privilege on documents.	There is performance overhead in cases when the search is not selective, returning large number of rows before query-time authorization. Extra work is required to implement a `ResultFilterPlugin`.

For sources that do not fit the user/group model, an authorization plug-in provides a more flexible security model. With an authorization plug-in, a crawler plug-in can add security attributes similar to document attributes. The authorization plug-in is invoked at login time to build security filters onto the query string. The security filters are applied against the values of the security attributes for each document. Only documents whose security attribute values match the security filter are returned to the user.

See Also:

- "Administrator-Based Authorization" on page 9-17 for more information about ACLs
- "Query-time Authorization" on page 9-18 for more information on Java filter classes

Restrictions on Changing the ACL Policy

On the Home - Sources - Authorization page, you can set and change the ACL policy.

The following ACL policy options are available:

- **No ACL**: With this setting, all documents are considered searchable and visible
- **Oracle Secure Enterprise Search ACL**: With this setting (also known as **source-level ACLs**), you can protect the entire source with one ACL. The same ACL protects every document in that source.
- **ACLs Controlled by the Source**: This setting (also known as **document-level ACLs**) is available only for OracleAS Portal sources and user-defined sources. This preserves authorizations specified in OracleAS Portal. For user-defined sources, crawler plug-ins (or connectors) can supply ACL information with documents for indexing, which provides finer control document protection. (That is, each document in the source can have different access privileges.)

The following restrictions apply to changing the ACL policy:

- If the schedule associated with the source is not currently being crawled, and if the source has never been crawled, then all ACL policy changes are allowed.
- If the schedule associated with that source is currently being crawled (that is, the schedule status is Launching, Executing, or Stopping), then all ACL options are grayed out, and you cannot change the ACL policy.
- If the schedule associated with the source is not currently being crawled, but the source *has* been crawled earlier, then the only change allowed is between **No ACL** and **Oracle Secure Enterprise Search ACL** (in either direction). This is visible in the Oracle SES Administration GUI as follows:
  - If the ACL option selected before the crawl started was **No ACL** or **Oracle Secure Enterprise Search ACL**, then the **ACLs Controlled by the Source** option is grayed out.
  - If a secure ACL policy was selected but the identity plug-in is deactivated, then you can change the ACL policy to **No ACL** regardless of the crawl status.
- OracleAS Portal sources are subject to the same restrictions as other sources. That is, no changes are allowed while being crawled, and only changes between **No ACL** and **Oracle Secure Enterprise Search ACL** are allowed after crawling completes. However, the ACL policy for an OracleAS Portal source can also change if it is inheriting the ACL policy from its OracleAS Portal server parent; for example, when the OracleAS Portal server ACL policy is modified or when the OracleAS Portal source is changed from specifying the ACL policy locally to inheriting it from the server. Therefore, changes on an OracleAS Portal server are restricted so that no disallowed changes can occur on any children that inherit the ACL policy.

  If any child inheriting the ACL policy is being crawled, then no changes are allowed on the OracleAS Portal server. If any child inheriting the ACL policy has been crawled, then the only changes allowed are between **No ACL** and **Oracle Secure Enterprise Search ACL**. (If the OracleAS Portal server policy is **ACLs Controlled by the Source**, then no changes are allowed). Similarly, the OracleAS Portal source cannot be set to inherit its ACL policy from the OracleAS Portal server if the associated change in ACL policy would be disallowed.

> **Note:** A source that is being crawled is different from a source whose associated schedule is being crawled. Oracle SES restricts all ACL policy changes for a source when the schedule associated with that source is being crawled. A source might not be crawled, but the schedule associated with it could be crawled if another source in the same schedule is being crawled.

Activating an Identity Plug-in

You can register an identity plug-in to any identity management system. Oracle SES provides registered identity plug-ins for many identity management systems. The plug-in that you activate is responsible for all authentication and validation activity in Oracle SES. Activate an identity plug-in on the Global Settings - Identity Management Setup page.

When you activate a plug-in, the Global Settings - Activate Identity Plug-in page is displayed. When completing this page, ensure that the values for Additional User Base and Additional Group Base are not subsets of Directory Subscriber.

Caution: You must create the identity plug-in before creating an Oracle SES data source for any type of secure data. If the identity plug-in is not active, then the data source is crawled as a public data source.

The following table lists which identity plug-ins are available for each enterprise content source.

Table 9–3 Identity Plug-ins for Enterprise Content Sources

Source Type	Versions Supported	Identity Plug-in
Database	Any databases with a JDBC driver	Native
EMC Documentum Content Server	5.1, 5.2.5, 5.3 SP2	Active Directory, Oracle Internet Directory, Native
EMC Documentum eRoom	7.3	Active Directory, Oracle Internet Directory
Lotus Notes	5.0.9, 6.5.4,7.0	Active Directory, Oracle Internet Directory, Native
Microsoft Exchange	Windows 2000, Windows 2003	Active Directory
Microsoft SharePoint Portal Server	2003, 2007	Active Directory
NTFS	Windows 2000, Windows 2003	Active Directory, Oracle Internet Directory
Oracle Calendar	10.1.2 or later	Oracle Internet Directory
Oracle Content Database	Oracle Content Services 10.1.2 or later, Oracle Content Database 10.2 or 10.1.3	Native, Query-time authorization
Oracle E-Business Suite	11, 12	Native
Oracle Mail	10g	Oracle Internet Directory
Siebel 7.8	7.8	Native
Siebel 8	8	Native
Oracle Content Server	7.1.1, 7.5.2, 10gR3	Native

Activating the Active Directory Identity Plug-in When connecting to Active Directory, Oracle SES tries to resolve the Active Directory domain name to an IP address of the Active Directory Domain Controller. This is generally not possible, especially when Oracle SES is installed on a non-Windows system or on a Windows system in a different domain. You must add the IP address of the Active Directory domain to the host file.

For example, to connect to an Active Directory domain called `foobar.example.com`, you must add something similar to the hosts file: `10.123.1.2 foobar.example.com`. Search for the hosts file in `C:\Windows\System32\Drivers\etc\HOSTS` on Windows systems, and `/etc/hosts` on UNIX systems.

For the Active Directory identity plug-in enter values for the following parameters:

- **Directory URL**: `ldap://`*`ActiveDirectoryserver`*`:389`

- **Directory account name**: *UserLogonName* Confirm the user logon name on the Active Directory Users and Computers application. Under the **User** folder, right-click **username**. Select **Property** and go to the **Account** tab. For example, assume the user account `adtest` in domain `domain1.example.com`, which is associated with the target Active Directory. You may try `domain1\adtest` or `adtest@domain1.example.com` or `cn=adtest,cn=users,dc=domain1,dc=example,dc=com` if you are not sure the actual user logon name. The user account does not need to be an administrator account.
- **Directory account password**: *PasswordForDirectoryAccount*
- **Directory subscriber**: dc=domain1,dc=example,dc=com for the domain name `domain1.example.com`
- **Directory security protocol**: none

If you deactivate an identity plug-in, then you must restart the middle tier with `searchctl restart`.

Re-registering Pre-Installed Identity Plug-ins

If a pre-installed identity plug-in is accidentally removed, you can re-register it with the following steps:

1. On the Global Settings - Identity Management Setup page, click **Register New Identity Plug-in**.
2. Enter the class name and jar file name of the removed identity plug-in, as identified in Table 9–4.
3. Click **Finish**.

Table 9–4 Identity Plug-in Class Names and Jar File Names

Identity Plug-in	Plug-in Class Name	Jar File Name
Documentum Content Services	oracle.search.plugin.security.identity.dcs.DCSIdentityPluginManager	../dcs/DCSIdentityPlugin.jar
Database	oracle.search.plugin.security.identity.db.DBIdentityPluginManager	../oracleapplications/DBCrawler.jar
Oracle E-Business Suite	oracle.search.plugin.security.identity.ebs.EBSIdentityPluginMgr	.../oracleapplications/EBSCrawler.jar
Siebel 7.8	oracle.search.plugin.security.identity.siebel.Siebel78IdentityPluginMgr	../oracleapplications/Siebel78Crawler.jar
Siebel 8	oracle.search.plugin.security.identity.siebel.SiebelIdentityPluginMgr	../oracleapplications/Siebel8Crawler.jar
Oracle Content Server	oracle.search.plugin.security.identity.stellent.StellentIdentityPluginMgr	../oracleapplications/StellentCrawler.jar
Oracle Internet Directory	oracle.search.plugin.security.identity.oid.OIDPluginManager	OIDPlugins.jar
Active Directory	oracle.search.plugin.security.idm.IdentityPluginManagerADImpl	idm/idmPlugin.jar

Table 9–4 (Cont.) Identity Plug-in Class Names and Jar File Names

Identity Plug-in	Plug-in Class Name	Jar File Name
Sun Java System Directory Server	oracle.search.plugin.security.idm.IdentityPluginManagerIPlanetImpl	idm/idmPlugin.jar
OpenLDAP Directory	oracle.search.plugin.security.idm.IdentityPluginManagerOpenLdapImpl	idm/idmPlugin.jar
Lotus Notes	oracle.search.plugin.security.identity.ln.LNIdentityPluginManager	ln/LNIdentityPlugin.jar

Restrictions on Changing the Identity Plug-in

The information Oracle SES saves from the identity plug-in (that is, the correspondence between names and canonical attribute values) may not be valid on different identity plug-ins. If you keep the same identity plug-in server (for example, to change port numbers or to switch to SSL), or if you use a new directory server that has identical user information, then you can deactivate and re-activate the identity plug-in anytime without restriction. This section describes steps you must perform if you change identity plug-in servers with user information that is not identical.

If you have sources using the ACL policy Oracle Secure Enterprise Search ACL and you decide to use a different identity plug-in server, then you must clear the ACL data before deactivating the original identity plug-in. If the ACL data is not cleared, then the ACL policy configured for that source while connected to the old identity plug-in server is not correctly enforced after connecting to the new identity plug-in server.

The existing ACL data can be cleared using either of these methods:

- Before deactivating the identity plug-in, for each source using the ACL policy Oracle Secure Enterprise Search ACL, switch the ACL policy to No ACL. After connecting to the new identity plug-in server, restore the ACL policy to Oracle Secure Enterprise Search ACL and add the ACLs again. This temporarily makes the source public. If this is unacceptable, then use the next option.
- Before deactivating the identity plug-in, delete each source that uses the ACL policy Oracle Secure Enterprise Search ACL. After connecting to the new identity plug-in server, add the sources back and configure them again. The documents are never made public; but this may involve more work than the previous option.

If you have sources using the ACL policy ACLs Controlled by the Source and you decide to use a different identity plug-in server, then after activating the new identity plug-in server, each source that uses this ACL policy must be re-crawled with the **Process All Documents** option. This forces the reloading and indexing of all of ACL information for such sources using the new identity plug-in server. Select the **Process All Documents** option on the Home - Schedules - Edit Schedule page.

> **Note:** If the ACL data is not cleared before switching identity plug-in servers, then you see a message that some users and groups could not be found by the identity plug-in. Those users and groups are still displayed on the Home - Sources -Authorization page. They can be deleted manually.

Authentication Methods

The Oracle SES front-end interface collects user credentials, which are then validated against the active identity plug-in. In addition to authentication of search users, Oracle SES must also authenticate the crawler when accessing external data repositories.

Administrators supply credentials to crawl private content through the following authentication methods:

- HTTP authentication (both basic and digest authentication)
- HTML forms
- OracleAS Single Sign-On
- Oracle Access Manager Single Sign-On

It is the administrator's responsibility to check the authorization policy to ensure that crawled documents are properly protected.

Oracle Secure Enterprise Search User Repository

Oracle SES has two types of users:

- **Administrative User**: The administrative user is SEARCHSYS. This user is natively defined in Oracle SES. Only this user can use the Oracle SES Administration GUI.
- **Search Users**: Oracle SES lets you register an identity plug-in as an interface to any identity management system. (Oracle SES provides registered identity plug-ins for Oracle Internet Directory and other identity management systems.) The plug-in that you activate is responsible for all authentication and validation activity in Oracle SES. Use the Global Settings - Identity Management Setup page in the Oracle SES Administration GUI to associate Oracle SES with an identity management system.

Oracle Internet Directory is Oracle's native LDAP v3-compliant directory service. It is part of the Oracle Identity Management infrastructure. It is not included in Oracle SES. Use Oracle Internet Directory version 9.0.4 or 10.1.2 (with the latest patch release applied) for connection with Oracle SES. Oracle Internet Directory is not a part of Oracle SES, and therefore Oracle SES can be linked to any existing or new Oracle Internet Directory.

Oracle SES Authentication Interface

For the SEARCHSYS administrative user, a form login screen is available in the Oracle SES Administration GUI. This is the only way for an administrator to log in to Oracle SES.

For search users, there are three possible front-end authentication interfaces:

- HTML form login page. Oracle SES provides an authentication page, and it authenticates against the identity plug-in.
- Web Services API. The `login` and `logout` Web Services operations authenticate against the identity plug-in.
- Single sign-on login screen. This can be made available by front-ending Oracle SES with OracleAS Single Sign-On and Oracle HTTP Server. These are available as part of the Oracle Identity Management infrastructure in OracleAS.

Note:

- Only form login *or* single sign-on login can be used for search users at any point in time. Using single sign-on with the Web Services authentication interface is not supported.
- Oracle strongly recommends that you SSL-protect the channel between the Oracle HTTP Server and the Oracle WebLogic server instance for secure content.

Enabling Secure Search

Much of the information within an organization is publicly accessible. However, there are other sources that are protected. For example, while a user can search in their own e-mail folders, they should not be able to search anyone else's e-mail. A secure search returns only search results that the user is allowed to view based on access privileges.

Oracle SES can use the following two security modes: using OracleAS Single Sign-On or not. These options are set on the Global Settings - Query Configuration page:

- Require login for secure content only: anyone can search public content. This is the default. This is also known as secure mode 2.
- Require login for public and secure content. This is also known as secure mode 3.

The security mode is applied to both the default query application and Oracle SES Web services. In mode 3, if a user tries to perform any Web services operation (search or document service) without logging in first, then a SOAP exception is thrown indicating that this secure mode requires login for any operation.

This section describes the authorization methods that Oracle SES supports. The authorization methods prevent search users from accessing documents for which they do not have privileges.

Oracle Secure Enterprise Search offers several options for secure search:

- User Authorization Cache
- Federated User Authorization Cache
- Administrator-Based Authorization
- Identity-Based Secure Search
- Query-time Authorization
- Self Service Authorization

See Also: The Oracle SES administration tutorial at

`http://st-curriculum.oracle.com/tutorial/SESAdminTutorial/index.htm`

User Authorization Cache

The User Authorization Cache (UAC) source type can crawl and cache user authorization information such as groups and accessible values of user security attributes. This cached information is used at query time to build a security filter. Querying a local cache is much faster than retrieving the authorization information from external repository and identity systems, and thus it significantly reduces the

time to build the security filters for the current user. As a result, users can log in to Oracle SES much more quickly. Moreover, you can set up UAC sources to crawl the user authorization information off line, which reduces the load on target repositories at query time.

You can use UAC for sources that are based on either of the security models supported in Oracle SES:

- **Identity-based security**: UAC is enabled for Oracle Internet Directory, Active Directory, and Lotus Notes.
- **Attribute-based security**: UAC is enabled for Oracle Content Database and Oracle Content Server sources. This type of security is also called the user-defined security model.

The crawler stores the following information in the User Authorization Cache:

- User groups: The list of groups that a user belongs to.
- User attribute values: The values of a specified list of attributes for particular data sources. The values can be single values or arrays of values.

To create a UAC source for an identity plug-in:

1. Click the **Sources** tab.
2. For Source Type, select **User Authorization Cache**, then click **Create**.

   The Create User-Defined Source page is displayed.
3. Configure the UAC source with the parameters described in Table 9–5. Set **Retrieve user groups** to true.
4. Create and activate the identity plug-in. Configure the plug-in to use the cache.

   For example, set these parameters for an Oracle Internet Directory plug-in:

   - **Use User Cache**: `true`
   - **User Cache Source Name**: Name of the UAC source that caches the user group information.

To create a UAC source for attribute-based security:

1. Click the **Sources** tab.
2. For Source Type, select **User Authorization Cache**, then click **Create**.

   The Create User-Defined Source page is displayed.
3. Configure the UAC source with the parameters described in Table 9–5. Set **Source names for which security attributes should be crawled**.
4. Configure the data sources to use cached user authorization information for building security filter by setting appropriate UAC related parameters in the authorization plug-in.

   For example, set these parameters for Oracle Content Server source:

   - **Use cached user and role information to authorize results**: true
   - **User role data source to cache the filter**: Name of the UAC source that caches the user authorization information for this data source.

Table 9–5 User Authorization Cache Parameters

Parameter	Setting
User search query	Query expression defining the set of users to be crawled. For example, enter `a*` to crawl all users whose names begin with the letter a. Leave this parameter blank to crawl all users who have logged into Oracle SES. This expression is interpreted by the active identity plug-in. When the active identity plug-in is Oracle Internet Directory or Active Directory, specify the LDAP query to select the set of users to crawl.
User attributes to be synchronized from identity system	Comma-delimited list of user attributes to crawl, such as `name, groups`. Not used in this release: Leave blank.
Retrieve user groups	Enter `true` to cache groups for users; otherwise, enter `false`.
Source names for which security attributes should be crawled	Enter a comma-delimited list of source names with security attributes to crawl. For example, `source1, source2`. The security model in these sources must be attribute-based security.

Federated User Authorization Cache

The Federated User Authorization Cache maintains a single User Authorization Cache (UAC) for use by all Oracle SES instances in a federated environment. Any identity plug-in or authorization plug-in can use a Federated UAC.

Prerequisite

- Define one or more UAC sources for an Oracle SES instance in the federated environment, as described in "User Authorization Cache" on page 9-12. A local cache created on this system can be accessed as a federated UAC source by any other instance with an identically configured identity or authorization plug-in.

To create a federated UAC:

1. Click the **Sources** tab.
2. For Source Type, select **Federated UAC**, then click **Create**.

   The Create Federated UAC page is displayed.
3. Configure the federated UAC source with the parameters described in Table 9–6.
4. Configure an identity or authorization plug-in:
   a. Select the **Global Settings** secondary tab.
   b. Under System, choose **Identity Management Setup**.
   c. Activate an identity or authorization plug-in. Enter the name of the federated UAC as the value of the **User Cache Source Name** parameter.
5. Repeat these steps for each additional Oracle SES instance in the federated environment.

Table 9–6 Federated UAC Parameters

Parameter	Setting
Connection String	JDBC connection string to a remote Oracle SES instance with one or more UAC sources. (Required)
Remote Cache Config File	Full path name of the XML configuration file for the remote cache, as described in Example 9–2. (Required)

Table 9–6 (Cont.) Federated UAC Parameters

Parameter	Setting
Password	Password for the SEARCHSYS user on the Oracle SES instance identified by **Connection String**. (Required)

Modifying the Remote Cache Configuration File

You can create the remote cache configuration file anywhere on the computer where you defined a federated UAC. You identify the location of the file when defining a Federated UAC source. The file provides details about the remote UAC source. Example 9–2 shows a sample file.

Example 9–2 Remote Cache Configuration File

```
<?xml version="1.0" encoding="UTF-8"?>
<FederatedUAC>
   <UserCache>
      <Name>UAC1</Name>
   </UserCache>
   <UserCache>
      <Name>UAC2</Name>
      <UserRouting>u6-u10</UserRouting>
   </UserCache>
   <UserCache>
      <Name>UAC3</Name>
      <SourceMapping>
         <RemoteSourceName>S5</RemoteSourceName>
         <LocalSourceName>S1</LocalSourceName>
      </SourceMapping>
      <SourceMapping>
         <RemoteSourceName>S7</RemoteSourceName>
         <LocalSourceName>S3</LocalSourceName>
      </SourceMapping>
   </UserCache>
</FederatedUAC>
```

Name

The name of a UAC source in a remote instance of Oracle SES from which the federated UAC retrieves user and attribute information. (Required)

UserRouting

Not supported in this release.

SourceMapping

Maps the local source to a remote source with an identically configured authorization plug-in. (Required for attribute-based security)

Use these elements to provide the mapping information:

- **RemoteSourceName**: Name of a remote source defined with a UAC and an authorization plug-in that is configured identically to the plug-in for **LocalSourceName**. (Required)
- **LocalSourceName**: Name of the local source defined with a federated UAC and an authorization plug-in. (Required)

XML Schema Definition for Remote Cache Configuration Files

The following is the XML schema definition for remote cache configuration files:

```
<?xml version="1.0" encoding="windows-1252" ?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
            xmlns="http://www.example.org" attributeFormDefault="unqualified">
  <xsd:element name="FederatedUAC">
    <xsd:annotation>
      <xsd:documentation>Federated Source Configuration</xsd:documentation>
    </xsd:annotation>
    <xsd:complexType>
      <xsd:all>
        <xsd:element name="UserCache">
          <xsd:annotation>
            <xsd:documentation>
              Remote UAC Source Config Details
            </xsd:documentation>
          </xsd:annotation>
          <xsd:complexType>
            <xsd:all>
              <xsd:element name="Name" minOccurs="1">
                <xsd:annotation>
                  <xsd:documentation>
                    UAC source name in remote instance from which
                    user/attribute information must be retrieved
                  </xsd:documentation>
                </xsd:annotation>
              </xsd:element>
              <xsd:element name="UserRouting" minOccurs="0">
                <xsd:annotation>
                  <xsd:documentation>
                     Which users should be routed to this UAC cache. Used for
                     identity-based security model.
                  </xsd:documentation>
                </xsd:annotation>
              </xsd:element>
              <xsd:element name="SourceMapping" minOccurs="0" maxOccurs="1">
                <xsd:annotation>
                  <xsd:documentation>
                    To map data source prefix for user-defined security
                    attribute retrieval. This will be used in case of user
                    defined security model. The mapping information must be
                    mentioned in the form of remote and local data source names.
                  </xsd:documentation>
                </xsd:annotation>
                <xsd:complexType>
                  <xsd:all>
                    <xsd:element name="RemoteSourceName" maxOccurs="1"
minOccurs="1">
                      <xsd:annotation>
                        <xsd:documentation>
                          Remote Instance Data source name prefixed to the
                          attribute while caching security attribute information
                        </xsd:documentation>
                      </xsd:annotation>
                      <xsd:complexType/>
                    </xsd:element>
                    <xsd:element name="LocalSourceName" minOccurs="1">
                      <xsd:annotation>
                        <xsd:documentation>
```

```
                        Local Instance Data source name for which security
                        attribute is being retrieved
                      </xsd:documentation>
                    </xsd:annotation>
                  </xsd:element>
                </xsd:all>
              </xsd:complexType>
            </xsd:element>
          </xsd:all>
        </xsd:complexType>
      </xsd:element>
    </xsd:all>
  </xsd:complexType>
 </xsd:element>
</xsd:schema>
```

Administrator-Based Authorization

With administrator-based authorization, the administrator can specify an authorization policy when creating a source. This policy governs which users can view each document. Administrator-based authorization is based on ACLs. When a source is crawled, each document is stamped with an ACL. When a user enters a search, the result list only includes documents for which the user credentials match the document ACL.

> **See Also:** "Authentication and Authorization" on page 9-4 for more information about ACL policies

Oracle SES performs ACL duplicate detection. If a crawled document's ACL exists in the Oracle SES system, then that ACL is used to protect the document, instead of creating a new ACL within Oracle SES. This policy reduces storage space and increases performance.

Oracle SES supports only a single LDAP domain. The LDAP users and groups specified in the ACL must belong to the same LDAP domain.

Caution: If ACLs are crawled from sources, then ensure that the sources being crawled belong to the same LDAP domain. Otherwise, the end users might inadvertently be granted permission to documents that they should not be able to access.

When secure search is enabled, you may encounter up to a 15 minute delay viewing the private documents. This delay could be due to newly added secure sources or a user/group membership change in the identity management system.

Identity-Based Secure Search

You can do identity-based secure search with administrator-based authorization or custom crawler plug-ins.

Oracle SES provides identity plug-ins for OpenLDAP release 2.2 and 2.3 and Sun Java System Directory Server release 5.1 and 5.2. Activate either of these identity plug-ins on the Global Settings - Identity Management Setup page.

To use administrator-based authorization:

1. On the Home - Sources page, select a source to use administrator-based authorization.
2. On the Home - Sources - Customize Source page, click the **Authorization** tab.
3. Under Crawl-time ACL Stamping, select **Oracle Secure Enterprise Search ACL**.
4. For Type, select **User** or **Group**.
5. Click **Add Another Row**.
6. For **User**, select **USER_NAME** or something you want to use as **Format** and enter user name as **Name**. For **Group**, select **DN** as **Format** and input **cn=<Group>,<Group search bases>** as **Name**.
7. Click **Apply**.

Limitations with OpenLDAP and Sun Java System Directory Identity Plug-ins

The LDAP entry of users and groups on OpenLDAP or Sun Java System Directory Server requires the following conditions:

Users:

- Belong to the following objectClasses: `person`, `organizationalPerson`, and `inetOrgPerson`
- Have the following attributes: `dn`, `cn`, `sn`
- The entry's location: `uid=<User>,<User search bases>`

Groups:

- Belong to one objectClass: `groupOfUniqueNames`
- Have the following attributes: `dn, uniqueMember`
- The entry's location: `cn=<Group>,<Group search bases>`

Query-time Authorization

Query-time authorization provides another form of filtering. Query-time authorization can be enabled or disabled for Web, file, table, e-mail, mailing list, OracleAS Portal, and user-defined source types from the Home - Sources - Edit Source page. It is not available for federated or self-service sources. Query-time authorization can be used with or without ACLs. For example, a source could be stamped with a relatively broad ACL, while query-time authorization could be used to further filter the results.

In query-time authorization, the Oracle SES administrator associates a Java class that is called at run time. The Java class validates each document that is returned in a user query.

Query-time authorization requires these steps:

1. The Oracle SES administrator registers a Java class implementing the `ResultFilterPlugin` interface with a source that requires query-time authorization.
2. Oracle SES crawls, collects, and indexes all documents. If ACL stamping has been set up, then it ACL-stamps the documents.
3. At search time, the search result list initially contains all documents accessible under crawl-time ACL policies, unfiltered by query-time user privilege checking.
4. For the top-N results requested by the user, Oracle SES calls the registered Java class, passing in the search request and document information for any documents

belonging to the protected source. The Java class returns an integer value for each document indicating if the document should be removed from the result or not.

5. Only items the user is privileged to see are returned to the user in their result list.

Notes for Using Query-time Authorization:

- The Browse application is also filtered by the query-time authorization mechanism. The `ResultFilterPlugin` class controls which folders are visible to the user, and documents within folders are filtered by the same process as the standard search result list.
- Set the **Hit Count Method** to **Exact count (adjusted for query-time filtering)** on the Global Settings - Query Configuration page. If it is not set to **Exact Count**, then the hit count displayed could be larger than the actual number of documents the user is authorized to view. The page in the Oracle SES Administration GUI contains other query-time authorization configuration settings you might want to consider.

  Oracle SES reports an approximate count of search results. The number of documents Oracle SES fetches determines the accuracy of the estimation. When the hit count is high enough to go beyond one page of results, then the count changes to a more accurate count as you click **Next** pages. **Exact count** shows an accurate count, but this option impacts query response time.

  > **Note:** While crawling secure data sources, the crawler skips infosource count-hit. Changing the ACL policy for a data source from secure to public does not result in an automatic update of the infosource count-hit. If you wish to include the data source in the infosource count-hit, you must recrawl the data source or trigger infosource `update_doc_count` for the data source.

- If you modify the contents of the jar file containing the `ResultFilterPlugin` implementation classes, but do not change the location of the jar file, then you must restart WebLogic Server This ensures that the search application picks up your changes and that the Java Virtual Machine does not use a cached version of the class within the old jar file. Restart WebLogic Server as described in "Starting and Stopping Oracle SES" on page 2-5.
- If a `ResultFilterPlugin` class is enabled for an OracleAS Portal server, then all of its page group sources are automatically protected by that query-time filter.
- It may take up to five seconds for query-time authorization changes applied in the Oracle SES Administration GUI to take effect in the Oracle SES search engine. The relevant settings are the following:
  - Enabling a `ResultFilterPlugin` class for a source
  - The hit count method
  - The Query-time Authorization Configuration settings on the Global Settings - Query Configuration page.

Self Service Authorization

Self service authorization allows end users to enter their credentials needed to access an external content repository. Oracle Secure Enterprise Search crawls and indexes the repository using these credentials to authenticate as the end user. Only the self service user is authorized to see these documents in their search results. Self service

authorization works well out of the box, as the crawler appears to be a normally authenticated end user to the content repository.

To set up a self service source, create a template source, defining the target data repository but omitting the credentials needed to crawl. From the search application, an end user can view the Customize page and subscribe to a template source by entering their credentials in an input form. A new user-subscribed source is created, along with a copy of the template's schedule. Oracle SES creates an ACL for this user to be applied to the source.

User-subscribed sources are viewable in the Home - Sources - Manage Template Source page, and the associated schedules are administered in the Home - Schedules page. Any changes applied by the administrator to a template source are dynamically inherited by the associated user-subscribed sources for the next crawl.

The self service option is available for e-mail and Web sources. Self service e-mail sources require the administrator to specify the IMAP server address and the end user to specify the IMAP account user name and password. Self service Web sources are limited to content repositories that use OracleAS Single Sign-On authentication. The administrator specifies the seed URLs, boundary rules, document types, attribute mappings, and crawling parameters, and the end user specifies the single sign-on user name and password.

Crawling of user-subscribed sources is controlled by the administrator. End users do not see any search results for their subscribed source until that source is crawled by the administrator's schedule. Allowing a crawl to automatically launch immediately after an end user subscribes to a source might be useful. However, it makes it possible for users to load the system at inconvenient times.

Configuring Secure Search with OracleAS Single Sign-On

If you use OracleAS Single Sign-On, then you can configure Oracle SES to use your OracleAS server for authentication. This section describes the necessary configuration steps.

OracleAS supported version with the current Oracle SES installation is 10.1.4.0.1, with the latest patch sets applied. The supported Oracle HTTP Server versions are 10.1.3 and 11g.

Note: `ORACLEAS_HOME` refers to the Oracle home directory of the OracleAS middle tier installation, which is typically stored in a system variable. On UNIX systems, you can reference the path as `$AS_HOME`. On Windows, the equivalent is `%AS_HOME%`.

`ORACLEOHS_HOME` refers to the home directory of the Oracle HTTP server installation.

To enable Oracle SES for OracleAS Single Sign-on:

1. Front the Oracle SES instance with Oracle HTTP Server, as described in "Configuring Oracle HTTP Server".
2. Configure OracleAS and Oracle SES, as described in "Configuring OracleAS and Oracle SES for Single Sign-on Security" on page 9-25.

Configuring Oracle HTTP Server

The Oracle SES middle tier runs on Oracle WebLogic server, which is installed in *MW_HOME*/wlserver.

You can configure either Oracle HTTP Server 10.1.3 or Oracle HTTP Server 11g. Special configuration is necessary on both the Oracle SES side and the Oracle HTTP Server side.

> **Note:** When using Oracle HTTP Server fronting, Oracle SES allows the Oracle HTTP Server to assert the identity of the current user. You must limit this privilege to only trusted Oracle HTTP Server instances by SSL-protecting the communication between Oracle SES and Oracle HTTP Server.

To configure Oracle HTTP Server 10.1.3:

1. Copy the file mod_wl_20.so from *MW_HOME*/wlserver/server/plugin/linux/i686 to *ORACLEOHS_HOME*/ohs/modules.
2. Edit the file httpd.conf available at *ORACLEOHS_HOME*/ohs/conf/ to include the following lines:

```
LoadModule weblogic_module [Complete path to mod_wl_20.so]
<IfModule mod_weblogic.c>
   WebLogicHost [SES host name]
   WebLogicPort [SES HTTP port]
</IfModule>
<Location /search/query>
   SetHandler weblogic-handler
</Location>
<Location /search/admin>
   SetHandler weblogic-handler
</Location>
```

   For example, if the ORACLEOHS_HOME is /ohsHome, Oracle SES host is sesHost and the Oracle SES port is 8001, then the file contains the following content:

```
LoadModule weblogic_module  /ohsHome/ohs/modules/mod_wl_20.so
<IfModule mod_weblogic.c>
   WebLogicHost sesHost
   WebLogicPort 8001
</IfModule>
<Location /search/query>
   SetHandler weblogic-handler
</Location>
<Location /search/admin>
   SetHandler weblogic-handler
</Location>
```

3. Edit the httpd.conf to comment out or remove the following lines:

```
#LoadModule auth_module modules/mod_auth.so
#LoadModule auth_anon_module modules/mod_auth_anon.so
#LoadModule auth_db_module modules/mod_auth_dbm.so
```

4. Register Oracle HTTP Server mod_osso with Oracle Single Sign-On server 10.1.4. For registration, run the following command from *ORACLEAS_HOME*:

```
setenv ORACLE_HOME [full path to $OracleAS_Home]

$OracleAS_Home/sso/bin/ssoreg.sh -oracle_home_path  [Complete Path of
OracleAS_Home]  -site_name  [Hostname Of The Fronting OHS Server:Port]
-config_mod_osso TRUE  -mod_osso_url [http:// Hostname Of The Fronting OHS
Server:Port]  -update_mode  CREATE  -remote_midtier  -config_file [Path to
generate the file osso.conf]
```

For example, if you installed Oracle Identity Management 10g in `/asHome`, and the Oracle HTTP Server URL is `http://ohsserver:7779`, then you must run the following command:

```
setenv ORACLE_HOME /asHome

$OracleAS_Home/sso/bin./ssoreg.sh -oracle_home_path /asHome -site_name
ohsserver:7779 -config_mod_osso TRUE -mod_osso_url http://ohsserver:7779
-update_mode CREATE -remote_midtier -config_file /temp/osso.conf
```

5. Configure `mod_osso` to protect web resources with static directives:
   - Copy the file `osso.conf` generated in step 4 to: *ORACLEOHS_HOME*`/ohs/conf/osso`.
   - Edit the file `mod_osso.conf` available at *ORACLEOHS_HOME*`/ohs/conf/` to include the following:

```
LoadModule osso_module [path of OracleOHS_Home]/ohs/modules/mod_osso.so
<IfModule mod_osso.c>
  OssoIdleTimeout off
   OssoIpCheck on
   OssoConfigFile [path of OracleOHS_Home]/ohs/conf/osso/osso.conf
#OssoRedirectByForm off
#OssoSecureCookies on
#OssoProtectedOnly on
#OssoSecureCookies on
#OssoSendCacheHeaders on
#OssoHttpsFrontend on
#UseWebCacheIp on
<Location /search/query/formlogin.uix>
   require valid-user
   AuthType Basic
</Location>
</IfModule>
```

6. Edit the file `httpd.conf` to include the following:

```
include OracleOHS_Home/ohs/conf/mod_osso.conf
```

   For example, if your `ORACLEOHS_HOME` is `/ohsHome`, then the value would be:

```
include /ohsHome/ohs/conf/mod_osso.conf
```

7. Restart the HTTP server by issuing the following command:

```
OracleOHS_Home/opmn/bin/opmnctl restartproc process-type=HTTP_Server
```

To configure Oracle HTTP Server 11g:

1. Set *ORACLE_HOME* and *ORACLE_INSTANCE*.

   For example, if Oracle HTTP Server is installed at `/ohsHome`:

```
setenv ORACLE_HOME /ohsHome
```

If the instance is named myInstance

```
setenv ORACLE_INSTANCE /ohsHome/instances/myInstance
```

2. Edit the file mod_wl_ohs.conf available at *ORACLEOHS_HOME*/instances/instance1/config/OHS/ohs1/ to include the following:

```
<IfModule weblogic_module>
  WebLogicHost [SES host name]
  WebLogicPort [SES HTTP port]
  Debug ON
  WLLogFile Convenient Location of the log
</IfModule>
<Location /search/query>
  SetHandler weblogic-handler
</Location>
<Location /search/admin>
  SetHandler weblogic-handler
</Location>
# For monitor SES URL
<Location /monitor>
SetHandler weblogic-handler
</Location>
# For Help links in Admin side
<Location /search/ohw>
SetHandler weblogic-handler
</Location>
```

For example, if your Oracle SES host is sesHost and Oracle SES port is 8001,

```
<IfModule weblogic_module>
   WebLogicHost sesHost
   WebLogicPort 8001
   WLLogFile /scratch/exampleuser/weblogic.log
</IfModule>
<Location /search/query>
   SetHandler weblogic-handler
</Location>
<Location /search/admin>
   SetHandler weblogic-handler
</Location>
<Location /monitor>
   SetHandler weblogic-handler
</Location>
<Location /search/ohw>
   SetHandler weblogic-handler
</Location>
```

3. Register Oracle HTTP Server mod_osso with Oracle Single Sign-On Server 10.1.4. To register it, issue the following command from ORACLEAS_HOME:

```
setenv ORACLE_HOME [full path to $OracleAS_Home]

OracleAS_Home/sso/bin/ssoreg.sh -oracle_home_path  [Complete Path of
OracleAS_Home]  -site_name  [Hostname Of The Fronting OHS Server:Port]
-config_mod_osso TRUE  -mod_osso_url [http:// Hostname Of The Fronting OHS
Server:Port]  -update_mode  CREATE  -remote_midtier  -config_file [Path to
generate the file osso.conf]
```

For example, if you installed Oracle Identity Management 10g in /asHome, and the Oracle HTTP Server URL is http://ohsserver:7779, then you must run the following command:

```
setenv ORACLE_HOME /asHome

OracleAS_Home/sso/bin./ssoreg.sh -oracle_home_path /asHome -site_name
ohsserver:7779 -config_mod_osso TRUE -mod_osso_url http://ohsserver:7779
-update_mode CREATE -remote_midtier -config_file /temp/osso.conf
```

4. Configure mod_osso to protect Web resources with static directives:

   a. Copy osso.conf generated above to the location *OracleOHS_HOME*/instances/instance1/config/OHS/ohs1/conf

   b. Copy the file mod_osso.conf from *ORACLEOHS_HOME*/instances/instance1/config/OHS/ohs1/disabled/ to *OracleOHS_HOME*/instances/instance1/config/OHS/ohs1/moduleconf/.

   c. Edit the file mod_osso.conf located at *OracleOHS_HOME*/instances/instance1/config/OHS/ohs1/moduleconf to include the following lines:

```
LoadModule osso_module ${OracleOHS_HOME}/ohs/modules/mod_osso.so
<IfModule mod_osso.c>
  OssoIdleTimeout off
  OssoIpCheck on
  OssoSecureCookies Off
OssoConfigFile  <path_to_osso.conf_file>
#Location is the URI you want to protect
<Location />
  require valid-user
  AuthType Osso
</Location>
</IfModule>
```

   For example if your OracleOHS_Home is /ohsHome and you generated the osso.conf file at *ORACLEOHS_HOME*/instances/instance1/config/OHS/ohs1/conf, then you must include the following:

```
LoadModule osso_module "${OracleOHS_HOME}/ohs/modules/mod_osso.so"
<IfModule osso_module>
   OssoIpCheck off
   OssoIdleTimeout on
   OssoSecureCookies Off
   OssoConfigFile
/ohsHome/instances/instance1/config/OHS/ohs1/conf/osso.conf
<Location /search/query/formlogin.uix>
   require valid-user
   AuthType Osso
</Location>
</IfModule>
```

   d. Edit the file httpd.conf located at *ORACLEOHS_HOME*/instances/instance1/config/OHS/ohs1/ to include the following at the end of the file:

```
include
"${ORACLE_INSTANCE}/config/${COMPONENT_TYPE}/${COMPONENT_NAME}/moduleconf/m
```

```
od_osso.conf"

# Include the configuration files needed for mod_weblogic
include
"${ORACLE_INSTANCE}/config/${COMPONENT_TYPE}/${COMPONENT_NAME}/mod_wl_ohs.c
onf"

# Include the SSL definitions and Virtual Host container
include
"${ORACLE_INSTANCE}/config/${COMPONENT_TYPE}/${COMPONENT_NAME}/ssl.conf"
# Include the admin virtual host (Proxy Virtual Host) related configuration
include
"${ORACLE_INSTANCE}/config/${COMPONENT_TYPE}/${COMPONENT_NAME}/admin.conf"

include "moduleconf/*.conf"
```

5. Restart the HTTP server with the command

```
/$OracleOHS_Home/instances/instance1/bin/opmnctl startproc process-type=OHS
```

Configuring OracleAS and Oracle SES for Single Sign-on Security

To configure OracleAS to front Oracle SES:

1. Add providers for a WebLogic domain for OSSO. To add them, copy the file `ossoiap.jar` to the following location within Oracle WebLogic Server: *MW_HOME*`/wlserver/server/lib/mbeantypes/`. The file `ossoiap.jar` is located in *ORACLEOHS_HOME*`/modules/oracle.ossoiap_11.1.1/.`

   Note that `ossoiap.jar` is available only with Oracle HTTP Server 11g. If you have configured OracleAS Single Sign-On with Oracle HTTP Server 10g, then you must separately install Oracle HTTP Server 11g to obtain this file.

2. Restart WebLogic Server as described in "Starting and Stopping Oracle SES" on page 2-5.

3. Add OSSO Identity Asserter as described in "Adding OSSO Identity Asserter" on page 9-26.

4. Add Oracle Internet Directory Authenticator as described in "Adding Oracle Internet Directory Authenticator" on page 9-26.

5. Turn on the Single Sign-On flag in Oracle SES. This is done by updating the deployment plan for the query application:

   a. Edit the file `QueryPlan.xml` located in *ORACLE_HOME*`/search/tools/weblogic/deploy/plans/` to add the following:

```
<variable-definition>
 <variable>
  <name>sso_enabled</name>
   <value>true</value>
 </variable>

 <variable>
  <name>sso_vendor_name</name>
  <value>osso</value>
 </variable>

 <variable>
  <name>sso_user_guid_header</name>
```

```
  <value>Osso-User-Guid</value>
 </variable>

 <variable>
  <name>sso_username_header</name>
  <value>REMOTE_USER</value>
 </variable>
</variable-definition>
```

b. Redeploy the query application with the modified deployment plan. To redeploy, run the following command from *ORACLE_HOME*/search/tools/weblogic/deploy/:

```
./deployer.sh -serverURL t3://weblogic_url:port -user Weblogic Username
-password SES Admin Password -name application_name -plan plan_location
-process [redeploy|deploy]
```

For example, if you install Oracle SES on the host `myWlsServer` and port `7777`, and the Oracle SES admin password is `welcome1`, then you must issue the following command:

```
./deployer.sh -serverURL t3://myWlsServer:7777/ -user weblogic -password
welcome1 -name search_query -plan
$ORACLE_HOME/search/tools/weblogic/deploy/plans/QueryPlan.xml -process
redeploy
```

Adding OSSO Identity Asserter

To add an OSSO Identity Asserter to the domain, use the Oracle WebLogic Administration Console to perform the following steps:

1. Log in to the WebLogic Administration Console.
2. Click **Security Realms**, ***Default Realm Name***, **Providers**.
3. Click **New** under the Authentication Providers table.
4. Enter a name for the new provider, select its type, and then click **OK**. For example:

   Name: `OSSO Identity Asserter`

   Type: `OSSOIdentityAsserter`
5. Click the name of the newly added provider.
6. On the Common tab, set the appropriate values for common parameters and set the Control Flag to `SUFFICIENT` and then save the settings.
7. Save all configuration settings.
8. Stop and restart the Oracle WebLogic Server for the changes to take effect.

Adding Oracle Internet Directory Authenticator

To add an Oracle Internet Directory Authenticator to the domain, use the Oracle WebLogic Administration Console to perform the following steps:

1. Log in to the WebLogic Administration Console.
2. Click **Security Realms**, ***Default Realm Name***, **Providers**.
3. Click **New** under the Authentication Providers table.
4. Enter a name for the new provider, select its type, and then click **OK**. For example:

   Name: `OID Authenticator`

Type: `OracleInternetDirectoryAuthenticator`

5. Click **Save**.
6. Click the newly added authenticator to see the Settings page. Retain the default settings; do not change the Control Flag until you have verified that the Oracle Internet Directory configuration is valid.
7. On the Common tab, specify the following required settings and then save the settings.

   Propagate Cause For Login Exception: Check.

   Principal: LDAP administrative user. For example: cn=orcladmin

   Host: The Oracle Internet Directory host name

   Use Retrieved User Name as Principal: Check

   Credential: LDAP administrative user password. For example: password

   Confirm Credential: For example: password

   Group Base DN: Oracle Internet Directory group search base

   User Base DN: Oracle Internet Directory user search base.

   Port: Oracle Internet Directory port
8. Save all configuration settings.
9. Stop and restart the Oracle WebLogic Server for the changes to take effect.

After adding OracleAS Single Sign-On Identity Asserter and Oracle Internet Directory Authenticator as authentication providers, perform the following steps:

1. Log in to the WebLogic Administration Console.
2. Click **Security Realms**, **Default Realm Name**, **Providers**.
3. Select the **Users and Groups** tab to see a list of users and groups contained in the configured authentication providers.

   You should see user names from the Oracle Internet Directory configuration, which implicitly verifies that the configuration is working.
4. If the Oracle Internet Directory instance is configured successfully, you can change the control flag.

   If the Oracle Internet Directory authentication is sufficient for an application to identify the user, then choose the `SUFFICIENT` flag. `SUFFICIENT` indicates that if a user can be authenticated against Oracle Internet Directory, no further authentication is required. `REQUIRED` indicates that the authentication provider must authenticate the user even if another provider has already authenticated the user.
5. If your application requires the user name to be in the same case as in Oracle Internet Directory (uppercase, lowercase, initial capitals), select **Use Retrieved User Name as Principal**.
6. Save the changes.
7. Activate the changes and restart Oracle WebLogic Server.

Configuring Secure Search with Oracle Access Manager Single Sign-On

You can implement a single sign-on authentication mechanism for Oracle SES by using Oracle Access Manager.

Ensure that the following components are installed:

- Oracle Access Manager 10.1.4.3.0 or higher. See *Oracle Access Manager Installation Guide.*
- Oracle HTTP Server 11*g*
- Oracle Internet Directory 10.1.4.3.0 or higher. See *Oracle Fusion Middleware Administrator's Guide for Oracle Internet Directory*. Also see "Configuring Oracle Identity Management" on page 9-28 for information on configuring Oracle Internet Directory.
- Oracle HTTP Server WebGate.

  You must install Oracle Access Manager, and then add an entry for WebGate in Oracle Access Manager before installing WebGate. *Oracle Access Manager Installation Guide* provides detailed information about installing WebGate. Follow the steps as provided in this guide. However, for some steps, such as while creating a WebGate instance and while installing the WebGate, you must provide certain Oracle Access Manager-specific parameters, as listed in "Installing and Configuring WebGate" on page 9-30.

To implement the Oracle Access Manager single sign-on authentication on Oracle SES, you must configure Oracle HTTP Server, Oracle SES, Oracle Internet Directory, and Oracle Access Manager.

Configuring Oracle Identity Management

You must install Oracle Identity Management 10.1.4.3.0 or higher. This is required because the Oracle SES parameter `sso_user_guid_header` must be used to send the `ORCLGUID` attribute from Oracle Access Manager to SES, and this can be done only with Oracle Internet Directory 10.1.4.3.0 or higher.

To enable this on Oracle Internet Directory:

1. Add the following to the LDIF file:

   ```
   dn: cn=dsaconfig, cn=configsets,cn=oracle internet directory
   changetype: modify
   add: orclallattrstodn
   orclallattrstodn: cn=orcladmin
   ```

2. Import the LDIF file into Oracle Internet Directory:

   ```
   $LDAP_HOME/bin/ldapmodify -D cn=orcladmin -w password -h host -p port -c -v -f
   ldifFile
   ```

3. To verify that the changes you made to the LDIF file are reflected, use the following command:

   ```
   $LDAP_HOME/bin/ldapsearch -b "cn=dsaconfig,cn=configsets,cn=oracle internet
   directory" -s base  -h host -p port -w password -D "cn=orcladmin"
   "objectclass=*"
   ```

   You should see `orclallattrstodn` as an attribute of the `dsaconfig` entry.

4. Restart the Oracle Access Server and the Oracle Identity Server:

   ```
   $OAM_HOME/as/access/oblix/apps/common/bin/restart_access_server
   ```

```
$OAM_HOME/is/identity/oblix/apps/common/bin/restart_ois_server
```

Configuring Oracle HTTP Server

To configure Oracle HTTP Server, perform the following tasks:

1. Edit `mod_wl_ohs.conf` to include the following. The file is available at *ORACLEOHS_HOME*`/instances/`*instance1*`/config/OHS/ohs1/`, where *instance1* refers to the instance name of Oracle HTTP Server.

```
<IfModule weblogic_module>
     WebLogicHost [SES host name]
     WebLogicPort [SES HTTP port]
     WLLogFile Convenient Location of the log
</IfModule>

<Location /search/query>
     SetHandler weblogic-handler
</Location>

<Location /search/admin>
     SetHandler weblogic-handler
</Location>

# For monitor SES URL
<Location /monitor>
     SetHandler weblogic-handler
</Location>

# For Help links in Admin side
<Location /search/ohw>
     SetHandler weblogic-handler
</Location>
```

   For example, if your SES host is `sesHost` and the port is `8001`:

```
<IfModule weblogic_module>
     WebLogicHost sesHost
     WebLogicPort 8001
     WLLogFile /scratch/exampleuser/weblogic.log
</IfModule>

<Location /search/query>
     SetHandler weblogic-handler
</Location>

<Location /search/admin>
     SetHandler weblogic-handler
</Location>

<Location /monitor>
     SetHandler weblogic-handler
</Location>

<Location /search/ohw>
     SetHandler weblogic-handler
</Location>
```

2. Edit `httpd.conf` located at *ORACLEOHS_HOME*`/instances/instance1/config/OHS/ohs1/` to include the following at the end of the file:

```
# Include configuration for mod_weblogic
include
"${ORACLE_INSTANCE}/config/${COMPONENT_TYPE}/${COMPONENT_NAME}/mod_wl_ohs.conf"
```

Ensure that this line of code is on a single line.

3. Restart the HTTP server.

```
$ORACLEOHS_HOME/instances/instance1/bin/opmnctl restartproc process-type=OHS
```

Installing and Configuring WebGate

A WebGate is a Web server plug-in that is shipped out-of-the-box with Oracle Access Manager. The WebGate intercepts HTTP requests from users for Web resources and forwards them to the Access Server for authentication and authorization. See *Oracle Access Manager Installation Guide* for more information on installing a WebGate.

While installing WebGate, you must configure some parameters for the Oracle Access Manager single sign-on authentication.

Creating a WebGate Instance

Provide the following values while defining a WebGate instance in the Access System Console:

- **AccessGateName**: Set as `SESAccessGate`
- **Description**: Set as `Secure Enterprise Search Access Gate`
- **HostName**: This is the host name on which Oracle HTTP Server is installed.
- **AccessGate Password**: Set a password.
- **Port**: This is the port number set during Oracle HTTP Server installation.
- **Transport Security**: Set to `Open`.
- **Preferred HTTP Host**: The domain for Oracle HTTP Server. For example, if the Oracle HTTP Server hostname is `myhost.oracle.com`, then the domain is `oracle.com`.
- Ensure that **Access Management Service** is on.

Installing WebGate

Provide the following parameters while specifying the WebGate configuration details:

- **WebGate ID**: Enter `SESAccessGate`.
- **WebGate Password**: The same as **AccessGate** password.
- **Access Server ID**: Obtain this from Access System Console.
- **DNS hostname**: Obtain this from Access System Console.
- **Port number**: Obtain this from Access System Console.

Updating the WebGate Web Server Configuration

Use the option to automatically update the Web Server configuration.

Integrating Oracle Access Manager with Oracle SES

Perform the following tasks:

1. Create a login page for Oracle HTTP Server. For example, *ORACLEOHS_HOME*/ohs/htdocs/login/login.html:

```
<html>
<head>
<title>SES-OAM Test Login Page</title>
<body bgcolor="white">
<h1 align="center">SES-OAM SSO Login Page: Sign-In</h1>
<form method="POST" action="/myaction/test.html">
  <table border="0" cellspacing="5">
    <tr>
      <th align="right">Username:</th>
      <td align="left"><input type="text" name="usernamevar"></td>
    </tr>
    <tr>
      <th align="right">Password:</th>
      <td align="left"><input type="password" name="passwordvar"></td>
    </tr>
    <tr>
      <td align="right"><input type="submit" value="Log In"></td>
      <td align="left"><input type="reset"></td>
    </tr>
  </table>
</form>
</html>
```

2. Define a form-based authentication in Oracle Policy Manager:

   a. From `http://OAMHost:OAMPort/access/oblix`, select **Access System Console**, then **Access System Configuration**, and then **Authentication Management**.

   b. Create Form Login method with the following options:

      Name: OAMFormLogin

      Description: OAM Form-based login

      Level: 1

      Challenge Method: Form

      Challenge Parameter

      `form`: /login/login.html

      `creds`: usernamevar passwordvar

      `action`: /myaction/test.html

      `passthrough`: no

      SSL Required: No

      Enabled: Yes

   c. Set up the following plugins under the **Plugins** tab:

      credential_mapping:

```
obMappingBase="o=company,c=us",obMappingFilter="(&(&(objectclass=gensiteorg
person)(genuserid=%usernamevar%))(|(!(obuseraccountcontrol=*))(obuseraccoun
tcontrol=ACTIVATED)))"
```

      validate_password:

```
obCredentialPassword="passwordvar"
```

where `obMappingBase` is the base DN in the user search in the LDAP directory server, and `obMappingFilter` is the LDAP filter used to search for a user with a given userID. The directory login attribute is an attribute defined in the Identity System using a Semantic login type.

d. Ensure that a default step exists in the **Steps** tab to use the `credential_mapping` and `validate_password` plugins.

3. Create a policy in the Policy Manager to protect the query application login link using the form authentication created in the previous step:

   a. From `http://OAMHost:OAMPort/access/oblix`, select **Policy Manager**, and then **Create Policy Domain**.

   b. Protect an HTTP resource with `/search/query/formlogin.uix` as the URL prefix.

   c. In the **Authorization Rules** tab, add the role `myrole`. Also set the following:

      Enabled: Yes

      Allow takes precedence: Yes

   d. Under **Actions** tab for `myrole`, first add the following return action:

      Type: `HeaderVar`

      Name: `HTTP_USER_GUID`

      Return Attribute: `orclguid`

      Then add the following return action:

      Type: `HeaderVar`

      Name: `HTTP_USER_NAME`

      Return Attribute: `uid`

   e. Under **Allow Access** tab, ensure that anyone is allowed access.

   f. Enable the new policy under **My Policy Domains**.

   g. Click **Default Rules**, and under **Authentication Rule**, add a rule to use the form login scheme as the **Authentication Scheme**.

   h. Under **Authorization Expression**, ensure that `myrole` is selected for **Default Rules**.

4. Create a policy in Policy Manager to protect the HTTP resource `/search/query` with the **Anonymous Authentication** option. Note that the steps are identical to the previous step. However, for step 3g, the form login scheme must be **Anonymous Authentication** under **Authentication Rule**.

5. Configure Oracle SES to use `HTTP_USER_GUID` and `HTTP_USER_NAME` as the values of `sso_user_guid_header` and `sso_username_header` respectively. See "Configuring QueryPlan.xml in Oracle SES" on page 9-33.

6. Configure Oracle SES to use `OblixAnonymous` as the value for `sso_public_username`. See "Configuring QueryPlan.xml in Oracle SES" on page 9-33.

Configuring QueryPlan.xml in Oracle SES

To enable Oracle Access Manager single sign-on authentication:

1. Configure the parameters shown in Example 9–3in the `QueryPlan.xml` file, which is available at *ORACLE_HOME*`/search/tools/weblogic/deploy/plans/`.
2. Redeploy the query application with the modified deployment plan by running the following command from *ORACLE_HOME*`/search/tools/weblogic/deploy/`:

   ```
   sh ./deployer.sh -serverURL t3://host:port/ -user weblogic -password password
   -name search_query -plan ./plans/QueryPlan.xml -process redeploy
   ```

   If SES is deployed on a Windows system, then run the batch file `deployer.bat`, as shown:

   ```
   %ORACLE_HOME%\search\tools\weblogic\deploy\deployer.bat -serverURL
   t3://host:port/ -user weblogic -password password -name search_query -plan
   .\plans\QueryPlan.xml -process redeploy
   ```

   Where:

   host is the host name, and *port* is the WebLogic service port. This is the same port that you use to open the Administration GUI. *password* is the password for eqsys.

   For example, if you install Oracle SES on the host `myWlsServer` and port `7777`, and the Oracle SES administration password is `welcome1`, then issue the following command:

   ```
   ./deployer.sh -serverURL t3://myWlsServer:7777/ -user weblogic -password
   welcome1 -name search_query -plan ./plans/QueryPlan.xml -process redeploy
   ```

Example 9–3 QueryPlan.xml Parameters for Enabling Oracle Access Manager Single Sign-On Authentication

```
<variable>
   <name>sso_enabled</name>
   <value>true</value>
   <description>Whether SSO is enabled: true or false. The default is false.
</description>
</variable>

<variable>
   <name>sso_vendor_name</name>
   <value>oam</value>
   <description>The SSO vendor name. Supported values are osso or
oam.</description>
</variable>

<variable>
   <name>sso_user_guid_header</name>
   <value>HTTP_USER_GUID</value>
   <description>The HTTP header name that the SSO server uses to pass the user
GUID to SES. The value in the header should match the value of the users canonical
attribute for the active identity plugin.</description>
</variable>

<variable>
   <name>sso_username_header</name>
```

```
   <value>HTTP_USER_NAME</value>
   <description>The HTTP header name that the SSO server uses to pass the search
username to SES. The value in the header should match the value of the users
authentication attribute for the active identity plugin. Specify REMOTE_USER to
use getRemoteUser in the HTTP request to retrieve the username.</description>
</variable>

<variable>
   <name>sso_public_username</name>
   <value>OblixAnonymous</value>
   <description>(Optional) Specify the username of the public user if the SSO
server is configured to send a public user name in the sso_username_header for
unprotected or anonymously protected resources.</description>
</variable>
```

SSL and HTTPS Support in Oracle Secure Enterprise Search

For SSL support, Oracle SES uses JSSE, a highly-customizable SSL package included in Sun Microsystem's J2SE. Oracle SES uses SSL for many operations, some acting as the SSL client, and others acting as the SSL server.

Oracle SES can crawl HTTPS-based URLs, and the Oracle SES middle tier can be configured to support HTTPS-based access. HTTPS refers to HTTP running over a secure socket layer (SSL).

Understanding SSL

SSL is an encryption protocol for securely transmitting private content on the internet. Using SSL, two parties can establish a secure data channel. SSL uses a cryptographic system that uses two keys to encrypt data: a public key and a private key. Data encrypted with the public key can only be decrypted using the private key, and vice versa.

In SSL terms, the party that initiates the communication is considered the client. During the SSL handshake, authentication between the two parties occurs. The authentication can be one-way (server authentication only) or two-way (server and client authentication). The Oracle SES crawler supports one-way SSL. It does not support two-way SSL.

Server authentication is more common. It happens every time a Web browser accesses a URL that starts with HTTPS. Because of server authentication, the client can be certain of the server's identity and can trust that it is safe to submit secure data such as login username and password to the server.

The following list defines some common terms related to SSL:

- **Keystore**: A repository that includes the following:
  - Certificates identifying trusted entities. If a keystore contains certificates of only trusted entities, then it is referred to as a *truststore*.
  - Private-key and the matching certificate. This certificate is sent as a response to SSL authentication challenges.
- **Certificate**: A digital identification of an entity that contains the following:
  - SSL public key of the server
  - Information about the server
  - Expiration date

  - Digital signature by the issuer of the certificate used to verify the authenticity of the certificate

- **Certificate authority (CA)**: A well known and trusted entity (for example, VeriSign or Thawte). CAs are usually the issuers of other certificates.
- **Root certificate**: A self-signed certificate where the issuer is the entity that the certificate represents. CA certificates are typically root certificates.
- **Certificate chain**: This chain consists of the certificate, its issuer, the issuer of the issuer, and so on, all the way to the root certificate. If one certificate in the chain is trusted (that is, it is in the keystore), then the rest of the certificate can be verified for authenticity. This makes it possible for a keystore to contain only a few well-known and trusted root certificates from which most other certificates originate.

Every SSL connection starts with the SSL handshake. These are the basic steps:

1. The client contacts the server to establish a SSL connection.
2. The server looks in its keystore for its own SSL certificate and sends it back to the client.
3. The client checks its keystore to see if it trusts the server or any of the entities in the server's certificate chain. If not, then the handshake is aborted. Otherwise, the client positively identifies the server and deems it trusted. The expiration date of the certificate is also checked, and the name on the certificate is matched against the domain name of the server.
4. If the server is configured to require client authentication, then the server asks the client to identify itself, so the mirror image of steps 2 and 3 takes place.
5. Session keys are generated and used for encrypting the transmitted data.

Oracle strongly recommends that you use an SSL-protected channel to transmit password and other secure data over networks.

Typically, the following components transmit password and other secure data over a network:

- Federation
- Connectors
- Authorization plug-ins
- Identity plug-ins
- Suggested content
- Web Service APIs

Managing the Keystore

The keystore is populated with the root certificates representing well known certificate authorities. Most SSL-enabled Web sites use certificates that originate or chain from these main root certificates. See "Oracle SES Acting as an SSL Server" on page 37 for more information about managing the keystore.

> **See Also:** The Java Secure Socket Extension (JSSE) Reference Guide at
>
> `http://java.sun.com/j2se/1.4.2/docs/guide/security/jsse/JSSERefGuide.html`

Importing SSL Certificates into the Java Virtual Machine

For connectors that interact with external SSL-enabled repositories at crawl time and query time, you must import the SSL certificate into the keystore of the Oracle SES crawler Java Virtual Machine (JVM) and the mid-tier JVM. The keystore in the crawler's JVM is used at crawl time, and the keystore in the mid-tier JVM is used at query time to build the security filter.

Following are the paths to the two JREs into which to import the SSL certificate:

- *MW_HOME*/jrockit_160_20_D1.1.0-2119/jre

Maintaining a Keystore

Depending on requirements, the keystore might need maintenance. For example:

- If a main root certificate has expired, then it must be replaced by a new issue.
- If Oracle SES must trust another SSL-enabled peer whose certificate does not originate from a root certificate, then the peer's certificate, or one from its chain, must be added to the keystore.
- To enable SSL in the Oracle SES middle tier, Oracle SES must act as an SSL server, and that calls for the keystore to contain a private key and the corresponding certificate with the public key. (The same holds true for the SSL client role where the server requires client side SSL authentication.)

Maintenance of the keystore can be done using Sun Microsystem's `keytool` program, which ships with J2SE. You can find this utility under *MW_HOME*/jrockit_160_20_D1.1.0-2119/bin. Third-party keytool GUI wrapper programs are available.

> **See Also:** For detailed instructions on how to add, remove, or update certificates, generate keys, and create new keystores with a keytool:
>
> http://java.sun.com/j2se/1.4.2/docs/tooldocs/windows/keytool.html

Oracle SES Acting as an SSL Client

Oracle SES acts as the SSL client in the following situations:

- The crawler accesses a data repository that uses SSL (for example, HTTPS Web sites).
- The form registration wizard in the Oracle SES Administration GUI accesses HTTPS URLs.
- Oracle SES federates queries to other SSL-enabled search services (for example, an SSL-enabled Oracle SES instance).

  Note that for Oracle SES 11.1.2 instances, the broker and the endpoint do not have to exchange any certificates when the broker tries to create a federated source using the HTTPS Web service URL. This is because the Oracle SES instances share the same default certificates in the trusted store.

If you crawl an SSL-enabled Web site whose SSL key is not in the SSL keystore, the following error occurs:

```
@ javax.net.ssl.SSLHandshakeException:
sun.security.validator.ValidatorException: No trusted certificate found
```

To fix this error, you can add the key to the Oracle SES keystore.

To add an SSL certificate to the Oracle SES keystore:

1. Access the page in a browser, and accept the SSL certificate when prompted.
2. View the certificate through your browser options.
3. Import the certificate into the Oracle SES keystore.
4. Try the crawl again.

The following sections explain how to import certificates.

Oracle SES Acting as an SSL Server

Oracle SES acts as the SSL server when the middle tier, configured to use SSL, responds to HTTPS requests. The Oracle SES crawler connects to SSL-enabled sites using the JSSE package, which contains a keystore with a few default certificates from well known CAs.

This section contains the following topics:

- Configuring Oracle Secure Enterprise Search to Require SSL
- Configuring Oracle HTTP Server to Require SSL

Configuring Oracle Secure Enterprise Search to Require SSL

When Oracle SES is fronted by an Oracle HTTP Server, Oracle recommends that Oracle SES be configured to require SSL with client-side authentication for communication with the Oracle HTTP Server. Furthermore, it should use a keystore other than the default one. It is highly recommended that you create separate identity and trust keystores.

The communication channel between the client and Oracle SES is by default not SSL-enabled and not encrypted.

To configure Oracle SES to require SSL:

1. Create a new keystore. This step is optional but Oracle recommends it.
   a. Open the WebLogic console and log in, as described in "Accessing the Oracle WebLogic Server Administration Console" on page 10-21.
   b. Expand the **Environment** button and click **Servers**. This takes you to the configuration page for the servers.
   c. Click the name of the server for which you want to configure SSL.
   d. Click the **keystores** tab.
   e. From the keystores list, select **Custom Identity and Custom Trust**.
   f. In the **custom identity keystore** field add the complete path and name of the new keystore. The default keystore is located at *MW_HOME*`/wlserver/server/lib`. To create a new keystore `SESIdentity.jks`, add the path and name *MW_HOME*`/wlserver/server/lib/SESIdentity.jks` to the keystore field.
   g. Set the custom identity keystore type to be `jks`. Set a pass phrase for the store.
   h. In the custom trust field, add the complete path and name of the new keystore. The default keystore is located at *MW_HOME*`/wlserver/server/lib`. To

create a new keystore `SESTrust.jks`, add the path and name *MW_HOME*`/wlserver/server/lib/SESTrust.jks` to the keystore field.

i. Set the custom identity keystore type to be `jks`. Set a pass phrase for the store.

j. Click **Save** to create the new keystores.

2. Create new certificates for the identity and trust keystores, using the Java keytool utility, which is located in *MW_HOME*`/jdk160_21/jre/bin`. If this directory is not in your search path, then make it your current working directory for these steps.

   a. Generate the key for the identity keystore:

   ```
   keytool -genkey -alias [MyCertificateAlias] -keyalg RSA -keysize 1024
   -dname ["My DN"] -keypass [MyKeyPass ] -keystore [MyKeyStore ]  -storepass
   [PasswordOfTheKeystoreCreatedAbove] -storetype [StoreTypeCreatedAbove]
   ```

   For example:

   ```
   keytool -genkey -alias sescert -keyalg RSA -keysize 1024 -dname
   "CN=example0123.us.mycompany.com,OU=ses,O=oracle,C=us" -keypass welcome1
   -keystore $MW_HOME/wlserver/server/lib/SESIdentity.jks -storepass welcome1
   -storetype jks
   ```

   This example creates a certificate with the alias `sescert` and the given dn and keypass `welcome1`. It uses `SESIdentity.jks` as the keystore, which matches the one created in step 1. The storepass and the storetype are the same as supplied in step 1.

   b. Generate the key for the trust keystore:

   ```
   keytool -genkey -keyalg RSA -alias sescert -keysize 1024 -dname
   "CN=example0123.us.mycompany.com,OU=ses,O=oracle,C=us" -keypass welcome1
   -keystore $MW_HOME/wlserver/server/lib/SESTrust.jks -storepass welcome1
   -storetype jks
   ```

   c. Certify the generated keys:.

   ```
   keytool -selfcert -alias sescert -keyalg RSA -validity 2000 -keypass
   welcome1 -keystore $MW_HOME/wlserver/server/lib/SESIdentity.jks -storepass
   welcome1
   ```

   The command uses the alias, keypass, and the keystore location supplied in step 2.a. The store pass is the password of the store.

   Self-certify the keystore:

   ```
   keytool -selfcert -alias sescert -keyalg RSA -validity 2000 -keypass
   welcome1 -keystore $MW_HOME/wlserver/server/lib/SESTrust.jks -storepass
   welcome1
   ```

   Note: In addition to using the Java keytool utility to self-sign the generated key, you can use any of the options mentioned here: http://download.oracle.com/docs/cd/E12840_01/wls/docs103/secmanage/identity_trust.html

3. Configure Oracle SES to use the generated key:

   a. Log in to the admin console for WebLogic and select the server for which you want to configure SSL by expanding the **Environment** button and clicking on **Servers**. This takes you to the configuration page for the servers.

b. Click the **ssl** tab.

c. The private key location is set to **from Custom Identity Keystore**.

d. In the **Private Key Alias** field, provide the private key alias. This is the alias specified in step 2a.

e. Provide the private key pass phrase that you specified in step 2a.

f. Save the settings.

4. Enable SSL for Oracle SES:

a. Log in to the admin console for WebLogic and select the server for which you want to configure SSL by expanding the **Environment** button and clicking on **Servers**. This takes you to the configuration page for the servers.

b. Click the **General** tab.

c. Select **SSL Listen Port Enabled** and provide a port number. The default port is `7002`.

d. Save the settings.

e. Click the **Control** tab. You can access the control tab by expanding the **Environment** button and clicking on **Servers**.

f. From the control tab, restart SSL.

Configuring Oracle HTTP Server to Require SSL

Configuring Oracle HTTP Server to require SSL is a multistep process involving configuration of the server, modification of certain `.conf` files, and exchange of certificates.

To configure Oracle HTTP Server to require SSL:

1. Configure the Oracle HTTP server:

a. From *`ORACLEOHS_HOME`*`/bin`, run `owm`. This opens Oracle Wallet Manager, which is used to create the certificate for Oracle HTTP Server.

b. Click **Wallet** and then click **New**.

If you get a message indicating that the default directory is not set, click **Continue**.

c. Provide a password for the wallet. Click **No** for the option to configure user certificate request.

d. Click **Wallet** and then click **Save As**. Save the wallet to the directory *`ORACLE_HOME`*`/instances/`*`instanceName`*`/config/OHS/`*`componentName`*`/keystores/myWallet`. This creates a new wallet with the name `myWallet` for the Oracle HTTP server.

`instanceName` and `componentName` are specified during the installation of Oracle HTTP Server.

e. Create a key-cert pair (a user certificate) using the following command from *`ORACLEOHS_HOME`*`/bin`:

```
orapki wallet add -wallet [walletPath] -dn ["myDN"] -keysize 1024
-self_signed -validity 720
```

For example,

```
orapki wallet add -wallet
```

```
$ORACLEOHS_HOME/instances/instance1/config/OHS/ohs1/keystores/myWallet -dn
CN=example0123.us.mycompany.com,OU=ohs.ses,O=oracle,ST=ca,C=US -keysize
1024 -self_signed -validity 720
```

The command adds a user certificate with the given dn and the wallet located at *ORACLEOHS_HOME*/instances/instance1/config/OHS/ohs1/keystores/myWallet. Note that instance1 is the name of the instance provided during installation and ohs1 is the name of the component provided during installation.

f. Go back to the OWM utility and reopen the wallet: Close and open the wallet by selecting the correct directory. You should now see Certificate:[Ready] under the wallet.

g. Save the wallet.

h. Double-click **Certificate:[Ready]**, click the **Operations** tab, and select **export user certificate**. Export the user certificate file (/tmp/OHSIdentityCertificate.crt) to a suitable location.

2. Edit the file ssl.conf located at *ORACLEOHS_HOME*/instances/*instanceName*/config/OHS/*componentName/*.to include the following. Note that *instanceName* and *componentName* are specified during the installation of Oracle HTTP Server.

```
<VirtualHost*:dddd>
<IfModule mod_weblogic.c>
   WebLogicHost [SESHost]
   WebLogicPort [SESPort]
   Debug ALL
   WLLogFile [Location of the log file]
   SecureProxy On
   WlSSLWallet "MyWalletLocation"
<Location /weblogic>
   SetHandler weblogic-handler
   PathTrim /weblogic
</Location>
<Location /console>
   SetHandler weblogic-handler
</Location>
</IfModule>
</VirtualHost >
```

For example, if the host is sesHost, the port is 7002, and the wallet is located at *Oracle_Instance*/config/*Component_Type/Component_Name*/keystores/myWallet, then the following configuration file is helpful:

```
<IfModule mod_weblogic.c>
   WebLogicHost sesHost
   WebLogicPort 7002
   Debug ALL
   WLLogFile /scratch/exampleuser/Certificates/weblogic.log
   SecureProxy On
   WlSSLWallet
"${ORACLE_INSTANCE}/config/${COMPONENT_TYPE}/${COMPONENT_NAME}/keystores/myWall
et"
<Location /weblogic>
   SetHandler weblogic-handler
   PathTrim /weblogic
</Location>
```

```
<Location /console>
   SetHandler weblogic-handler
</Location>
</IfModule>
```

3. Edit the file `mod_wl_ohs.conf` located at *ORACLEOHS_HOME*`/instances/`*instanceName*`/config/OHS/`*componentName*`/` to include the following:

```
<IfModule weblogic_module>
WebLogicHost [SES host name]
  WebLogicPort [SES HTTP port]
  Debug ON
  WLLogFile [Location of the log]
</IfModule>
<Location /search/query>
  SetHandler weblogic-handler
</Location>
<Location /search/admin>
  SetHandler weblogic-handler
</Location>
# For monitor SES URL
<Location /monitor>
SetHandler weblogic-handler
</Location>
# For Help links in Admin side
<Location /search/ohw>
SetHandler weblogic-handler
</Location>
```

   For example if the Oracle SES host is `sesHost` and the port is `8001`, then the file would contain:

```
<IfModule weblogic_module>
  WebLogicHost sesHost
  WebLogicPort 8001
  Debug ON
  WLLogFile /scratch/exampleuser/weblogic.log
</IfModule>
<Location /search/query>
  SetHandler weblogic-handler
</Location>
<Location /search/admin>
  SetHandler weblogic-handler
</Location>
<Location /monitor>
  SetHandler weblogic-handler
</Location>
<Location /search/ohw>
  SetHandler weblogic-handler
</Location>
```

4. Exchange the certificates for Oracle HTTP Server and Oracle SES WebLogic servers. Use Oracle Wallet Manager to import and export certificates from and to the wallet, and use the Java keytool for the Oracle SES keystore. While importing a certificate, ensure that it is self-signed. If not, then you must import any of the certificates in the chain. See "Understanding SSL" on page 9-34 for more information about certificate chains.

   Perform the following steps to exchange certificates:

a. Export the SESIdentity key generated in step 2a of Configuring Oracle Secure Enterprise Search to Require SSL to a suitable location by running the following command:

```
keytool -export -alias sescert -keystore
$ORACLESES_HOME/wls/wlserver/server/lib/SESIdentity.jks -file
/tmp/SESIdentityCertificate.crt
```

The above command exports the certificate with the alias `sescert` and the keystore created in step 2a of Configuring Oracle Secure Enterprise Search to Require SSL to the file `/tmp/SESIdentityCertificate.crt`.

b. Import the exported Oracle HTTP Server certificate created in step 1h to Oracle SES. Issue this command from *MW_HOME*`/jdk160_21/jre/bin`:

```
keytool -file [LocationOfOHSIdentityCertificate]  -alias [MyOHSCerAlas]
-import -trustcacerts -keystore [LocationofSESTrustStore] -storepass
[MyPasswordForTheTrustStore] -storetype jks
```

For example, if the location of the exported Oracle HTTP Server identity certificate is `tmp/OHSIdentityCertificate.crt`, the Oracle SES trust store is at *MW_HOME*`/wlserver/server/lib/SESTrust.jks`, the store password is `welcome1`, and the alias is `ohsCert`, then run the following:

```
keytool -file tmp/OHSIdentityCertificate.crt  -alias ohsCert -import
-trustcacerts -keystore $MW_HOME/wlserver/server/lib/SESTrust.jks
-storepass welcome1 -storetype jks
```

c. Import the Oracle SES certificate into Oracle HTTP Server wallet. The Oracle SES certificate is the file `SESIdentityCertificate.crt` exported in step 4a. To import this certificate, from the Oracle Wallet Manager utility, click **Operations** and select **Import Trusted Certificate**. Navigate to the location of the exported Oracle SES certificate (`/tmp/SESIdentityCertificate.crt`), and import it as a trusted certificate.

d. Restart Oracle HTTP Server. Before restarting the server, ensure that the **Auto Login** option is enabled in Oracle Wallet Manager. The restart fails if the option is not enabled.

To restart the server, run the following command from *ORACLEOHS_HOME*`/instances/instance1/bin/`:

```
opmnctl restartproc process-type=OHS
```

e. Restart SSL for the Oracle WebLogic Server by using the control page of the server.

To access the control page, click **Environment**, then **Server**, and then **Control**.

See Also: *Oracle Database Advanced Security Administrator's Guide* for more information about Oracle Wallet Manager

Oracle Database Advanced Security Administrator's Guide for more information about the orapki utility

Oracle HTTP Server Administering a Standalone Deployment Based on Apache 2.0 for more information about enabling SSL for Oracle HTTP Server

Changing the Master Encryption Key

A master encryption key is used to encrypt secure fields in Oracle SES. You can change this key if its security is compromised or for any other reason.

To change the master encryption key:

1. Stop all crawler schedules.
2. Close all middle-tier applications, except for the Monitor application.
3. Open an interactive session on the Oracle SES middle-tier computer.
4. Issue a `searchctl rollover_key` command. See the following description.
5. Restart the crawler and the middle-tier applications.

searchctl rollover_key

This command has the following syntax:

`searchctl rollover_key` *`options`*

Options have the format *keyword=value*:

ses_db_conn_str

Local JDBC connection string for the Oracle SES database. For example, `localhost:5555:ses1`. Required.

ses_admin_passwd

Oracle SES administrative password, that is, for the SEARCHSYS user. If you omit this password from the command, then you are prompted for it.

wls_admin_server

URL to the WebLogic Server Administration Console. For example, `t3://wls_example:8000`. Required.

wls_admin_user

User name of the WebLogic administrative user. (Required)

wls_admin_passwd

Password of the WebLogic administrative user. If you omit this password from the command, then you are prompted for it.

master_key

New master key. If you omit this option, a random master key is set.

The following command changes the master key to "testing123";

```
searchctl rollover_key ses_db_conn_str=localhost:5555:ses1
ses_admin_passwd=password wls_admin_user=weblogic wls_admin_passwd=password
wls_admin_server=t3://asHost:8000 master_key=testing123
```

10

Administering Oracle SES Instances

This chapter provides information about tuning and general management of Oracle SES instances. It contains the following topics:

- Increasing Data Storage Capacity
- Tuning Crawl Performance
- Tuning Search Performance and Scalability
- Turning On Debug Mode
- Supporting Failover in Oracle RAC
- Monitoring Oracle Secure Enterprise Search
- Integrating with Google Desktop
- Accessing the Oracle WebLogic Server Administration Console

Increasing Data Storage Capacity

When crawling a large number of documents, the default Oracle SES tablespaces may not be big enough to complete a crawl. You or your DBA can add more data files to the Oracle SES tablespaces in Oracle Database. If a crawl fails because of insufficient space, then you can add data files and restart the crawl.

These tablespaces are used by Oracle SES:

- **`SEARCH_DATA`**: Stores the crawled document metadata and cached documents. If the Preserve Document Cache option is enabled, then this tablespace requires a large data file to host all the crawled documents.
- **`SEARCH_INDEX`**: Stores the SES index.
- **`SEARCH_TEMP`**: A temporary tablespace that the crawler uses while processing the data.

Fusion tablespaces such as `FUSION_TS_TX_IDX`, an index for transactional data, may also require additional space.

See Also:

- *Oracle Secure Enterprise Search Administration API Guide* for creating partitioned tablespaces.
- *Oracle Fusion Applications Installation Guide* for managing Fusion tablespaces.

The following procedures use SQL*Plus, but you can use Enterprise Manager or another tool of your choosing.

To check the current tablespace usage for auto-extensible data files:

1. Open SQL*Plus and log in to Oracle Database as a privileged user, such as `SYS` or `SYSTEM`.
2. Query the `DBA_DATA_FILES` data dictionary view, using a `SELECT` statement like the following:

   ```
   COLUMN tablespace_name FORMAT a20

   SELECT tablespace_name, SUM(bytes)/1024/1024 "Used Megabytes",
   SUM(maxbytes)/1024/1024 "Max Megabytes" FROM dba_data_files WHERE
   tablespace_name IN ('SEARCH_DATA', 'SEARCH_INDEX') GROUP BY tablespace_name;
   ```

   The query returns the number of megabytes currently used and the maximum number of megabytes available in autoextensible SES tablespaces, as shown in the following example. When the number of used megabytes approaches the maximum, add more data files. By adding more space before the next crawl, you can avoid having the crawl fail for lack of space.

   ```
   TABLESPACE_NAME      Used Megabytes Max Megabytes
   -------------------- -------------- -------------
   SEARCH_DATA                   420    32767.9844
   SEARCH_INDEX                  120    32767.9844
   ```

Auto-extensible data files are typically used for Oracle SES, but if you have fixed-size data files, then query the `DBA_FREE_SPACE` data dictionary view.

To add data files to the Oracle SES tablespaces:

1. Open SQL*Plus and log in to Oracle Database as a privileged user, such as `SYS` or `SYSTEM`.
2. Query the `DBA_DATA_FILES` data dictionary view for the full path to the `SEARCH_DATA` and `SEARCH_INDEX` tablespaces:

   ```
   SELECT tablespace_name, file_name FROM dba_data_files
      WHERE tablespace_name IN ('SEARCH_DATA','SEARCH_INDEX');
   ```

3. Issue `ALTER TABLESPACE` commands like the one shown here for each additional data file. Replace *`tablespace`* with `SEARCH_DATA` or `SEARCH_INDEX`, *`path/filename`* with the name of the new data file, and *`bytes`* with the desired initial size.

   ```
   ALTER TABLESPACE tablespace ADD DATAFILE
      'path/filename.dbf' SIZE bytes
      AUTOEXTEND ON MAXSIZE UNLIMITED;
   ```

4. Query `DBA_DATA_FILES` again to list the new data files. (Optional)

To add temp files to the Oracle SES temporary tablespace:

1. In SQL*Plus, query the `DBA_TEMP_FILES` data dictionary view for the full path to the `SEARCH_TEMP` tablespace:

   ```
   SELECT tablespace_name, file_name FROM dba_temp_files
      WHERE tablespace_name='SEARCH_TEMP';
   ```

2. Issue ALTER TABLESPACE commands like the one shown here for each additional temp file. Replace *tablespace* with SEARCH_TEMP, *path/filename* with the name of the new temp file, and *bytes* with the desired initial size.

```
ALTER TABLESPACE SEARCH_TEMP ADD TEMPFILE
   'path/filename.dbf' SIZE bytes
   AUTOEXTEND ON MAXSIZE UNLIMITED;
```

3. Query DBA_TEMP_FILES again to list the new temp files. (Optional)

To restart a failed crawl:

1. Open the Oracle SES Administration GUI and log in.
2. Click the **Schedules** subtab on the Home page.
3. Select the failed schedule, then click the **Start** button.

Example: Adding Data Files to Oracle SES Tablespaces

This example adds two data files to SEARCH_DATA, a data file to SEARCH_INDEX, and a temp file to SEARCH_TEMP.

1. Open SQL*Plus:

```
sqlplus / as sysdba
```

2. Obtain the full paths to the Oracle SES tablespaces:

```
SELECT tablespace_name, file_name FROM dba_data_files
   WHERE tablespace_name IN ('SEARCH_DATA','SEARCH_INDEX');

TABLESPACE_NAME FILE_NAME
--------------- --------------------------------------------------
SEARCH_DATA     /oracle/product/oradata/ses/SEARCH_DATA_1.dbf
SEARCH_INDEX    /oracle/product/oradata/ses/SEARCH_INDEX_1.dbf

SELECT tablespace_name, file_name FROM dba_temp_files
   WHERE tablespace_name='SEARCH_TEMP';

TABLESPACE_NAME FILE_NAME
--------------- --------------------------------------------------
SEARCH_TEMP     /oracle/product/oradata/ses/SEARCH_TEMP_1.dbf
```

3. Add two data files to the SEARCH_DATA tablespace:

```
ALTER TABLESPACE search_data ADD DATAFILE
   '/oracle/product/oradata/ses/SEARCH_DATA_2.dbf' SIZE 10M
   AUTOEXTEND ON MAXSIZE UNLIMITED;

Tablespace altered.

ALTER TABLESPACE search_data ADD DATAFILE
   '/oracle/product/oradata/ses/SEARCH_DATA_3.dbf' SIZE 10M
   AUTOEXTEND ON MAXSIZE UNLIMITED;

Tablespace altered.
```

4. Add a data file to the SEARCH_INDEX tablespace:

```
ALTER TABLESPACE search_index ADD DATAFILE
   '/oracle/product/oradata/ses/SEARCH_INDEX_2.dbf' SIZE 10M
   AUTOEXTEND ON MAXSIZE UNLIMITED;
```

```
Tablespace altered.
```

5. Add a temp file to the SEARCH_TEMP tablespace:

```
ALTER TABLESPACE search_index ADD TEMPFILE
   '/oracle/product/oradata/ses/SEARCH_TEMP_2.dbf' SIZE 10M
   AUTOEXTEND ON MAXSIZE UNLIMITED;

Tablespace altered.
```

6. Query the data dictionary for the new data files:

```
SELECT tablespace_name, file_name FROM dba_data_files
   WHERE tablespace_name IN ('SEARCH_DATA','SEARCH_INDEX')
   ORDER BY tablespace_name;

TABLESPACE_NAME FILE_NAME
--------------- ------------------------------------------------------------
SEARCH_DATA     /oracle/product/oradata/ses/SEARCH_DATA_1.dbf
SEARCH_DATA     /oracle/product/oradata/ses/SEARCH_DATA_2.dbf
SEARCH_DATA     /oracle/product/oradata/ses/SEARCH_DATA_3.dbf
SEARCH_INDEX    /oracle/product/oradata/ses/SEARCH_INDEX_1.dbf
SEARCH_INDEX    /oracle/product/oradata/ses/SEARCH_INDEX_2.dbf

SELECT tablespace_name, file_name FROM dba_temp_files
   WHERE tablespace_name='SEARCH_TEMP';

TABLESPACE_NAME FILE_NAME
--------------- ------------------------------------------------------------
SEARCH_TEMP     /oracle/product/oradata/ses/SEARCH_TEMP_1.dbf
SEARCH_TEMP     /oracle/product/oradata/ses/SEARCH_TEMP_2.dbf
```

Tuning Crawl Performance

Your Web crawling strategy can be as simple as identifying a few well-known sites that are likely to contain links to most of the other intranet sites in your organization. You could test this by crawling these sites without indexing them. After the initial crawl, you have a good idea of the hosts that exist in your intranet. You could then define separate Web sources to facilitate crawling and indexing on individual sites.

However, the process of discovering and crawling your organization's intranet, or the Internet, is generally an interactive one characterized by periodic analysis of crawling results and modification to crawling parameters. For example, if you observe that the crawler is spending days crawling one Web host, then you might want to exclude crawling at that host or limit the crawling depth.

This section contains the most common things to consider to improve crawl performance:

- Crawler Schedule
- Proxy Servers
- Boundary Rules
- Dynamic Pages
- Crawler Depth
- Robots Rule
- Duplicate Documents

- Redirected Pages
- URL Looping
- Oracle Redo Log
- What to Do Next

> **See Also:** "Monitoring the Crawling Process" on page 3-25 for more information on crawling parameters

Crawler Schedule

Schedules define the frequency at which the Oracle SES index is updated with information about each source. This section describes characteristics of the Oracle SES crawler schedule.

- The Failed Schedules section on the Home - General page lists all schedules that have failed. A failed schedule is one in which the crawler encountered an irrecoverable error, such as an indexing error or a source-specific login error, and cannot proceed. A failed schedule could be because of a partial collection and indexing of documents.
- The smallest granularity of the schedule interval is one hour. For example, you cannot start a schedule at 1:30 am.
- If a crawl takes longer to finish than the scheduled interval, then it starts again when the current crawl is done. There is no option to have the scheduled time automatically pushed back to the next scheduled time.
- When multiple sources are assigned to one schedule, the sources are crawled one by one following the order of their assignment in the schedule.
- The schedule starts crawling the assigned sources in the assigned order. Only one source is crawling under a schedule at any given time. If a source crawl fails, then the rest of the sources assigned after it are not crawled. The schedule does not restart. You must either resolve the cause of the failure and resume the schedule, or remove the failed source from the schedule.
- There is no automatic e-mail notification of schedule success or failure.

For more information about documents that the crawler does not index:

- Browse the crawler log in the Oracle SES Administration GUI. Select the Schedules subtab from the Home page, then click the Log File icon for the schedule.
- In the Oracle SES Administration GUI, select the **Statistics** subtab from the Home page. Under Crawler Statistics, choose **Problematic URLs**. This page lists errors encountered during the crawling process and the number of URLs that caused each error.

Crawler Schedules in Oracle Fusion Applications

In Oracle Fusion Applications, the schedule is executed by Oracle Enterprise Scheduler using the time zone of the middle tier. The schedule is not affected by failed crawls, so the next crawl still begins as scheduled. If you modify the schedule in Oracle SES, the revised schedule overwrites the previous one in Enterprise Scheduler. If you deactivate a schedule, then the schedule is canceled in Enterprise Scheduler.

In the Oracle Enterprise Scheduler log, indexed documents have a 200 status code, and documents that are not indexed have a different status code.

Stuck Scheduling Requests

If the status of a schedule is Error Manual Recovery, then you cannot perform operations such as start, stop, or delete. You must change the schedule to Failed, then start again.

To recover stuck scheduling requests:

1. Log in to the Enterprise Manager Fusion Middleware Control console for the Common Domain.
2. In the left panel, expand Scheduling Services and click **ESSAPP *(server name)***.
3. In the top panel, click **Scheduling Service** to open a cascading menu.
4. Click **Job Requests**, then **Search Job Requests** to display the Request Search page.
5. In the Search box for Application, select SearchEssApp.
6. For Status, select Error Manual Recovery.
7. Click **Search** to see a list of stuck scheduling requests.
8. Fix each stuck request by taking these steps:
   a. In the Request ID column, click a linked number.
   b. In the top right corner, click **Action**.
   c. Click **Recover Stuck Request**.

   The following message is displayed:

```
Request Details: 18201(Error)
#Request processing resulted in error with the below message.
ERROR_MANUAL_RECOVERY: Request could not be recovered
#The job request failed due to System error.
#The internal processing of the job request is still not complete.
To recover the request, click Action and select Recover Stuck Request.
```

After you fix all of the ESS requests and the Oracle SES schedule has a status of Failed, you can manage the schedules again.

Proxy Servers

By default, Oracle SES is configured to crawl Web sites in the intranet, so no additional configuration is required. However, to crawl Web sites on the Internet (also referred to as external Web sites), Oracle SES needs the HTTP proxy server information.

To register a proxy:

1. On the Global Settings page under Sources, select **Proxy Settings**.
2. Enter the proxy server name and port. Click **Set Proxy**.
3. Enter the internal host name suffix under Exceptions, so that internal Web sites do not go through the proxy server. Click **Set Domain Exceptions**.

   To exclude the entire domain, omit `http`, begin with `*.`, and use the suffix of the host name. For example, `*.us.example.com` or `*.example.com`. Entries without the `*.` prefix are treated as a single host. Use the IP address only when the URL crawled is also specified using the IP for the host name. They must be consistent.
4. If the proxy requires authentication, then enter the proxy authentication information on the Global Settings - Authentication page.

Boundary Rules

The seed URL you enter when you create a source is turned into an inclusion rule. For example, if `www.example.com` is the seed URL, then Oracle SES creates an inclusion rule that only URLs containing the string `www.example.com` are crawled.

However, suppose that the example Web site includes URLs starting with `www.exa-mple.com` or `example.com` (without the `www`). Many pages have a prefix on the site name. For example, the investor section of the site has URLs that start with `investor.example.com`.

Always check the inclusion rules before crawling, then check the log after crawling to see what patterns have been excluded.

In this case, you might add `www.example.com`, `www.exa-mple.com`, and `investor.example.com` to the inclusion rules. Or you might just add `example`.

To crawl outside the seed site (for example, if you are crawling `text.us.oracle.com`, but you want to follow links outside of `text.us.oracle.com` to oracle.com), consider removing the inclusion rules completely. Do so carefully. This action could lead the crawler into many, many sites.

Notes for File Sources

- If no boundary rule is specified, then crawling is limited to the underlying file system access privileges. Files accessible from the specified seed file URL are crawled, subject to the default crawling depth. The depth, which is 2 by default, is set on the Global Settings - Crawler Configuration page. For example, if the seed is `file://localhost/home/user_a/`, then the crawl picks up all files and directories under `user_a` with access privileges. It crawls any documents in the directory `/home/user_a/level1` due to the depth limit. The documents in the `/home/user_a/level1/level2` directory are at `level 3`.
- The file URL can be in UNC (universal naming convention) format. The UNC file URL has the following format for files located within the host computer:

  `file://localhost///`*`LocalComputerName/SharedFolderName`*

  For example, specify `\\stcisfcr\docs\spec.htm` as `file://localhost///stcisfcr/docs/spec.htm`

  where `stcisfcr` is the name of the host computer.

  The string `localhost` is optional. You can specify the URL path without the string `localhost` in the URL, in which case the URL format is:

  `file:///`*`LocalComputerName/SharedFolderName`*

  For example,

  `file:///stcisfcr/docs/spec.htm`

  Note that you cannot use the UNC format to access files on other computers.
- On some computers, the path or file name could contain non-ASCII and multibyte characters. URLs are always represented using the ASCII character set. Non-ASCII characters are represented using the hexadecimal representation of their UTF-8 encoding. For example, a space is encoded as `%20`, and a multibyte character can be encoded as `%E3%81%82`.

  You can enter spaces in simple (not regular expression) boundary rules. Oracle SES automatically encodes these URL boundary rules. For example, `Home Alone` is specified internally as `Home%20Alone`. Oracle SES does this encoding for the following:

- File source simple boundary rules
- URL string tests
- File source seed URLs

Oracle SES does not alter regular expression rules. You must ensure that the regular expression rule specified is against the encoded file URL. Spaces are not allowed in regular expression rules.

Dynamic Pages

Indexing dynamic pages can generate too many URLs. From the target Web site, manually navigate through a few pages to understand what boundary rules should be set to avoid crawling of duplicate pages.

Crawler Depth

Setting the crawler depth very high (or unlimited) could lead the crawler into many sites. Without boundary rules, a crawler depth of 20 probably crawls the entire World Wide Web from most locations.

Robots Rule

You can control which parts of your sites can be visited by robots. If robots exclusion is enabled (the default), then the Web crawler traverses the pages based on the access policy specified in the Web server `robots.txt` file.

The following sample `robots.txt` file specifies that no robots visit any URL starting with `/cyberworld/map/` or `/tmp/` or `/foo.html`:

```
# robots.txt for http://www.example.com/

User-agent: *
Disallow: /cyberworld/map/
Disallow: /tmp/
Disallow: /foo.html
```

If the Web site is under your control, then you can tailor a specific robots rule for the crawler by specifying Oracle Secure Enterprise Search as the user agent. For example:

```
User-agent: Oracle Secure Enterprise Search

Disallow: /tmp/
```

The robots `meta` tag can instruct the crawler either to index a Web page or to follow the links within it. For example:

```
<meta name="robots" content="noindex,nofollow">
```

Duplicate Documents

Oracle SES always removes duplicate (identical) documents. Oracle SES does not index a page that is identical to one it has already indexed. Oracle SES also does not index a page that it reached through a URL that it has already processed.

With the Web Services API, you can enable or disable *near* duplicate detection and removal from the result list. Near duplicate documents are similar to each other. They may or may not be identical to each other.

See Also: "Oracle Secure Enterprise Search Web Services APIs" on page 11-1

Redirected Pages

The crawler crawls only redirected pages. For example, a Web site might have Javascript that redirects users to another site with the same title. In such cases, only the redirected site is indexed.

Check for inclusion rules from redirects. The inclusion rules are based on the type of redirect. The `EQ_TEST.EQ$URL` table stores all of the URLs that have been crawled or are scheduled to be crawled. There are three kinds of redirects defined in it:

- **Temporary Redirect**: A redirected URL is always allowed if it is a temporary redirection (HTTP status code 302, 307). Temporary redirection is used for whatever reason that the original URL should still be used in the future. It's not possible to find out temporary redirect from EQ$URL table other than filtering out the rest from the log file.
- **Permanent Redirect**: For permanent redirection (HTTP status 301), the redirected URL is subject to boundary rules. Permanent redirection means the original URL is no longer valid and the user should start using the new (redirected) one. In EQ$URL, HTTP permanent redirect has the status code 954
- **Meta Redirect**: Metatag redirection is treated as a permanent redirect. Meta redirect has status code 954. This is always checked against boundary rules.

The `STATUS` column of `EQ_TEST.EQ$URL` lists the status codes. For descriptions of the codes, refer to Appendix B, "URL Crawler Status Codes."

Note: Some browsers, such as Mozilla and Firefox, do not allow redirecting a page to load a network file. Microsoft Internet Explorer does not have this limitation.

URL Looping

URL looping refers to the scenario where a large number of unique URLs all point to the same document. Looping sometimes occurs where a site contains a large number of pages, and each page contains links to every other page in the site. Ordinarily this is not a problem, because the crawler eventually analyzes all documents in the site. However, some Web servers attach parameters to generated URLs to track information across requests. Such Web servers might generate a large number of unique URLs that all point to the same document.

For example,

```
http://example.com/somedocument.html?p_origin_page=10
```

might refer to the same document as

```
http://example.com/somedocument.html?p_origin_page=13
```

but the `p_origin_page` parameter is different for each link, because the referring pages are different. If a large number of parameters are specified and if the number of referring links is large, then a single unique document could have thousands or tens of thousands of links referring to it. This is an example of how URL looping can occur.

Monitor the crawler statistics in the Oracle SES Administration GUI to determine which URLs and Web servers are being crawled the most. If you observe an

inordinately large number of URL accesses to a particular site or URL, then you might want to do one of the following:

- Exclude the Web server: This prevents the crawler from crawling any URLs at that host. (You cannot limit the exclusion to a specific port on a host.)
- Reduce the crawling depth: This limits the number of levels of referred links the crawler follows. If you are observing URL looping effects on a particular host, then you should take a visual survey of the site to find out an estimate of the depth of the leaf pages at that site. Leaf pages are pages that do not have any links to other pages. As a general guideline, add three to the leaf page depth, and set the crawling depth to this value.

Be sure to restart the crawler after altering any parameters. Your changes take effect only after restarting the crawler.

Oracle Redo Log

Oracle SES allocates 200M for the redo log during installation. 200M is sufficient to crawl a relatively large number of documents. However, if your disk has sufficient space to increase the redo log and if you are going to crawl a very large number of documents (for example, more than 300G of text), then increase the redo log file size for better crawl performance.

> **Note:** The biggest transaction during crawling is `SYNC INDEX` by Oracle Text. Check the AWR report or the V$SYSSTAT view to see the actual redo size during crawling. Roughly, 200M is sufficient to crawl up to 300G.

To increase the size of the redo log files:

1. Open SQL*Plus and connect as the SYSTEM user. It has the same password as SEARCHSYS.
2. Issue the following SQL statement to see the current redo log status:

```
SELECT vl.group#, member, bytes, vl.status
    FROM v$log vl, v$logfile vlf
    WHERE vl.group#=vlf.group#;

GROUP# MEMBER                                             BYTES STATUS
------ -------------------------------------------------- ---------- ----------
     3 /scratch/ses111/oradata/o11101/redo03.log         209715200 INACTIVE
     2 /scratch/ses111/oradata/o11101/redo02.log         209715200 CURRENT
     1 /scratch/ses111/oradata/o11101/redo01.log         209715200 INACTIVE
```

3. Drop the INACTIVE redo log file. For example, to drop group 3:

```
ALTER DATABASE DROP LOGFILE group 3;

Database altered.
```

4. Create a larger redo log file with a command like the following. If you want to change the file location, specify the new location.

```
ALTER DATABASE ADD LOGFILE '/scratch/ses111/oradata/o11101/redo03.log' 2
     size 400M reuse;
```

5. Check the status to ensure that the file was created.

```
SELECT vl.group#, member, bytes, vl.status
     FROM v$log vl, v$logfile vlf
     WHERE vl.group#=vlf.group#;

GROUP# MEMBER                                             BYTES STATUS
------ -------------------------------------------------- ---------- ----------
     3 /scratch/ses111/oradata/o11101/redo03.log          419430400 UNUSED
     2 /scratch/ses111/oradata/o11101/redo02.log          209715200 CURRENT
     1 /scratch/ses111/oradata/o11101/redo01.log          209715200 INACTIVE
```

6. To drop a log file with a CURRENT status, issue the following ALTER statement, then check the results.

```
ALTER SYSTEM SWITCH LOGFILE;

SELECT vl.group#, member, bytes, vl.status
     FROM v$log vl, v$logfile vlf
     WHERE vl.group#=vlf.group#;

GROUP# MEMBER                                             BYTES STATUS
------ -------------------------------------------------- ---------- ----------
     3 /scratch/ses111/oradata/o11101/redo03.log          419430400 CURRENT
     2 /scratch/ses111/oradata/o11101/redo02.log          209715200 ACTIVE
     1 /scratch/ses111/oradata/o11101/redo01.log          209715200 INACTIVE
```

7. Issue the following SQL statement to change the status of Group 2 from ACTIVE to INACTIVE:

```
ALTER SYSTEM CHECKPOINT;

SELECT vl.group#, member, bytes,  vl.status
     FROM v$log vl, v$logfile vlf
     WHERE vl.group#=vlf.group#;

GROUP# MEMBER                                             BYTES STATUS
------ -------------------------------------------------- ---------- ----------
     3 /scratch/ses111/oradata/o11101/redo03.log          419430400 CURRENT
     2 /scratch/ses111/oradata/o11101/redo02.log          209715200 INACTIVE
     1 /scratch/ses111/oradata/o11101/redo01.log          209715200 INACTIVE
```

8. Repeat steps 3, 4 and 5 for redo log groups 1 and 2.

What to Do Next

If you are still not crawling all the pages you think you should, then check which pages were crawled by doing one of the following:

- Check the crawler log file
- Create a search source group

To check the crawler log file:

1. On the Home page, click the **Schedules** secondary tab to display the Crawler Schedules page.
2. Click the Log File icon to display the log file for the source.
3. To obtain the location of the full log, click the **Status** link. The Crawler Progress Summary and Log Files by Source section displays the full path to the log file.

To create a search source group:

1. On the Search page, click the **Source Groups** subtab.
2. Click **New** to display Create New Source Group Step 1.
3. Enter a name, then click **Proceed to Step 2**.
4. Select a source type, then shuttle only one source from Available Sources to Assigned Sources.
5. Click **Finish**.

To search the source group:

1. On any page, click the **Search** link in the top right corner to open the Search application.
2. Select the group name, then issue a search term to list the matches within the source.
3. Select the group name, then click **Browse** to see a list of search groups:
   - The number after the group name identifies the number of browsed documents. Click the number to browse the search results.
   - Click the arrow before the group name to display a hierarchy of search results. The number of matches appears after each item in the hierarchy.

Tuning Search Performance and Scalability

Oracle SES contains features that you can tune to optimize search performance. This section contains suggestions on how to improve performance (such as response time and throughput) and scalability of Oracle SES. It identifies the most common ways to improve search quality.

- Suggested Links
- Authentication and Authorization
- Parallel Query and Index Partitioning
- Index Fragmentation
- Indexing Parameters
- Search Statistics
- Load Balancing on Oracle RAC
- WebLogic Search Server Configuration
- Database Initialization Parameters
- Oracle UNDO Tablespace
- Buffer Cache

Suggested Links

Suggested links enable you to direct users to a designated Web site for particular query keywords. For example, when users search for "Oracle Secure Enterprise Search documentation" or "Enterprise Search documentation" or "Search documentation", you could suggest `http://www.oracle.com/technetwork`.

Suggested link keywords are rules that determine which suggested links are returned (as suggestions) for a query. A rule can include query terms and logical operators. For example, "secure AND search". With this rule, the corresponding suggested link is

returned for the query "secure enterprise search", but it is not returned for the query "secure database".

The rule language used for the indexed queries supports the following operators:

Table 10–1 Suggested Link Keyword Operators

Operator	Example
ABOUT	about(dogs)
AND	dog and cat
NEAR	dog ; cat
OR	dog or cat
PHRASE	dog sled
STEM	$dog
THESAURUS	SYN(dog)

Note: Do not use special characters, such as #, $, =, and &, in keywords.

Suggested links appear at the top of the search result list. Oracle SES can display up to two suggested links for each query.

This feature is especially useful for providing links to important Web pages that are not crawled by Oracle Secure Enterprise Search. Add or edit suggested links on the Search - Suggested Links page in the Oracle SES Administration GUI.

Authentication and Authorization

By tuning the security filter settings, you can prevent time outs.

To change the configuration of the security filter:

1. Log in to the Oracle SES Administration GUI.
2. Click the Global Settings tab, then Query Configuration.
3. Scroll down to Security Filter Configuration and change these settings using the guidelines provided in the Help.
   - Security Filter Lifespan
   - Authentication Timeout
   - Authorization Timeout
   - Minimum Number of Threads
   - Maximum Number of Threads

You can further tune the security filter by using the Administration API to set the `<search:preserveStaleSecurityFilterOnError>` and `<search:securityFilterRefreshWaitTimeout>` parameters in the `queryConfig` object. For example, these settings allow an expired security filter to be used immediately when a fresh security filter is unavailable:

```
<search:preserveStaleSecurityFilterOnError>true
   </search:preserveStaleSecurityFilterOnError>
<search:securityFilterRefreshWaitTimeout>0
```

```
</search:securityFilterRefreshWaitTimeout>
```

The settings listed previously are also parameters of the `queryConfig` object and can be modified using the API. See the *Oracle Secure Enterprise Search Administration API Guide*.

Parallel Query and Index Partitioning

Parallel querying significantly improves search performance and facilitates searches of very large data sources. The query architecture is based on Oracle Database partitioning and enhancements in Oracle Text.

To make the best use of this feature, Oracle recommends that you run Oracle SES on a server with a 4-core CPU, with at least 8GB of RAM and multiple fast disk drives.

Parallel querying is automatically implemented on Oracle SES when the partitioning option is enabled. You can specify partitioning only during installation.

To enable partitioning:

1. Acquire a license for the Oracle Partitioning option.
2. During installation, answer Yes when the Repository Creation Utility (RCU) asks if you have a partitioning license. Then Oracle Database is installed with partitioning, and Oracle SES automatically supports parallel query.

Storage Areas

Database tablespaces are registered as storage areas in Oracle SES. To make optimum use of the parallel querying feature, you should distribute partitioned tablespaces across all physical disks and register them with Oracle SES.

Use the Administration API to manage `storageArea` objects.

Configuring a Partition

Configuring a partition includes listing the storage areas, identifying the partitioning attribute, and updating the partitioning rules. Use the Administration API to manage the `partitionConfig` object.

> **See Also:** *Oracle Secure Enterprise Search Administration API Guide* for information about managing storage areas and partitions

Index Fragmentation

Index fragmentation management allows the search engine index to be updated while Oracle SES is executing searches. This is achieved by temporarily saving index changes to an in-memory index and periodically merging them with the larger disk-based search engine index. This reduces fragmentation and leads to faster response times. Index fragmentation management is implemented automatically on Oracle SES, but it can be tuned by configuring Oracle Text, where you can turn index fragmentation management on and off, and specify the frequency of index merges.

Optimizing the index also reduces fragmentation, and it can significantly increase the speed of searches. Schedule index optimization on a regular basis. Also, optimize the index after the crawler has made substantial updates or if fragmentation is more than 50%. Verify that index optimization is scheduled during off-peak hours. Optimization of a very large index could take several hours.

You can see the fragmentation level and run index optimization on the Global Settings - Index Optimization page in the Oracle SES Administration GUI. Index optimization has these options:

Do Not Run Optimization Longer Than

Specify a maximum duration for the index optimization process. The actual time taken for optimization does not exceed this limit, but it can be shorter. A longer optimization time results in a more optimized index. In this mode, the optimization process does not require a large amount of free disk space.

Until the Optimization is Finished

Specifies that the optimization continues until it is finished. Allowing the optimization to complete creates a more compact index and supports better performance than a partial optimization.

In this mode, Oracle SES creates a temporary copy of the index. The required disk space almost equals the current index size. If sufficient free disk space is not available, then the optimization fails. Use the appropriate SQL query shown here to estimate the minimum disk requirement:

- **Oracle SES Without Partitioning**

```
SELECT SUM(bytes)/1048576 AS "MBytes"
   FROM dba_segments
   WHERE segment_name IN ('DR$EQ$DOC_PATH_IDX$I','DR$EQ$DOC_PATH_IDX$X');
```

- **Oracle SES With Partitioning**

```
SELECT SUM(sz) AS "MBytes"
   FROM
   (
      SELECT MAX(bytes)/1048576 sz FROM dba_segments
         WHERE segment_name LIKE 'DR#EQ$DOC_PATH_IDX$%I'
   UNION
      SELECT MAX(bytes)/1048576 sz FROM dba_segments
         WHERE segment_name LIKE 'DR#EQ$DOC_PATH_IDX$%X'
   ) ;
```

These queries return an estimate of the *minimum* disk space needed for optimization. Oracle SES may require more disk space than this estimate.

After the optimization is complete, Oracle SES releases the disk space consumed during the optimization. The space can be used by future crawls or any activity that consumes disk space.

Indexing Parameters

To improve indexing performance, adjust the following parameters on the Global Settings - Set Indexing Parameters page of the Oracle SES Administration GUI:

Indexing Batch Size

When the crawled data in the cache directory reaches Indexing Batch Size, Oracle SES starts indexing. The bigger the batch size, the longer it takes to start indexing each batch. Only indexed data can be searched: Data in the cache cannot be searched. The default size is 250M.

Document fetching and indexing run concurrently. While indexing is running, the Oracle SES crawler continues to fetch documents and store them in the cache directory.

Indexing Memory Size

This is the upper limit of memory used for indexing before flushing the index to disk.

A large amount of memory improves indexing performance because it reduces I/O. It also improves query performance because the created index is less fragmented from the beginning, while a fragmented index can be optimized later. Set this parameter as high as possible without causing memory paging.

A smaller amount of memory might be useful when indexing progress should be tracked or when run-time memory is scarce. The default size is 275M. In general, increasing the Indexing Memory Size parameter can reduce fragmentation.

Parallel Indexing Degree

The number of concurrent threads used for indexing. This parameter is disabled in the current version of Oracle SES; it is always set to 1.

Search Statistics

See the Home - Statistics page in the Oracle SES Administration GUI for lists of the most popular queries, failed queries, and ineffective queries. This information can lead to the following actions:

- Refer users to a particular Web site for failed queries on the Search - Suggested Links page.
- Fix common errors that users make in searching on the Search - Alternate Words page.
- Make important documents easier to find on the Search - Relevancy Boosting page.

Once daily, SES automatically summarizes logged queries. The summarizing task might use the server resource if there are a large number of logged queries, which may impact query performance. This issue is visible for stress tests where several queries are executed every second. The ideal solution in such instances is to disable the query statistics option.

To disable the query statistics option:

1. From the Administration GUI Home page, select the **Global Settings** tab, then click **Query Configuration**.
2. Under Query Statistics, select **No** for the **Enable Query Statistics** option.

Relevancy Boosting

Relevancy boosting lets administrators influence the order of documents in the result list for a particular search. You might want to override the default results for the following reasons:

- For a highly popular search, direct users to the best results
- For a search that returns no results, direct users to some results
- For a search that has no click-throughs, direct users to better results

In a search, each result is assigned a score that indicates how relevant the result is to the search; that is, how good a result it is. Sometimes you know the documents that are highly relevant to some search. For example, your company Web site could have a home page for XML (`http://example.com/XML-is-great.htm`), which you

want to appear high in the results of any search for `XML`. You would boost the score of the XML home page to 100 for an `XML` search.

The document also has a score computed for searches that are not among the boosted queries.

Two methods can help you locate URLs for relevancy boosting: *locate by search* and *manual URL entry*.

Relevancy boosting, like end user searching, is case-insensitve. For example, a document with a boosted score for `Oracle` is boosted for `oracle`.

See "Customizing the Relevancy of Search Attributes" on page 4-3.

Load Balancing on Oracle RAC

When Oracle SES is deployed in an Oracle Real Applications Cluster (Oracle RAC) environment, the usage profile is typically one of the following:

- Small index with a large query load
- Large index with a small-to-large query load

A third option, a small index and a small query load, typically operates on a single computer.

Configuring a Small Index

The load balancing solutions provided by Oracle RAC and the WebLogic Server are sufficient for this type of Oracle SES deployment. Most or all of the index can reside in memory or the buffer cache. You only need to set up the listeners appropriately for Oracle SES.

To set up the listeners:

- Provide a local listener on each Oracle RAC instance.
- Do not configure remote listeners.
- Oracle recommends dedicated processes over shared processes.

WebLogic Search Server Configuration

Oracle SES is installed in a WebLogic domain as described in "Secure Search in Oracle Fusion Applications" on page 1-6. The default settings for stuck threads can result in slow query performance even under a moderate load.

To change the search server configuration

1. Log in to the WebLogic console, as described in "Accessing the Oracle WebLogic Server Administration Console" on page 10-21.
2. In the left panel under Change Center, click **Lock & Edit**.
3. In the left panel under Domain Structure, expand Environment and click **Servers**. The Summary of Services page is displayed in the main panel.
4. In the Name column, click **search_server1**. The Settings for search_server1 page is displayed.
5. Select the **Configuration** tab.
6. Configure these settings:
   - Stuck Thread Max Time: 3600

- Stuck Thread Timer Interval: 1800

7. Click **Save**.
8. Repeat these steps for any other search server instances, such as search_server2.
9. In the left panel under Change Center, click **Activate Changes**.

Database Initialization Parameters

To support a large number of simultaneous users, you may need to increase the values of these database initialization parameters:

- `PROCESSES`
- `SESSIONS`
- `OPEN_CURSORS`

In Fusion Applications, the Oracle SES middle tier uses connection pooling to communicate with the backend database. The database connection uses dedicated server mode, so that when 10 users run concurrent searches, the database requires 10 user processes.

The crawler also uses several threads, and each thread uses several database connections. You can alter the number of crawler threads on the Home - Sources - Crawling Parameters page of the Oracle SES Administration GUI.

Use the combined estimate of concurrent user processes and crawler threads for the value of `PROCESSES`. Then modify `SESSIONS` to a compatible value, typically calculated as `1.1 * PROCESSES`.

You can monitor the number of open cursors using the statistics stored in the `V$SESSTAT` dynamic performance view. If the number of open cursors for user sessions frequently approaches the maximum, then you can increase that number.

See Also:

- *Oracle Database Reference* for initialization parameters and dynamic performance views
- *Oracle Database Performance Tuning Guide* for database tuning

To change the database initialization parameters:

1. Open SQL*Plus and log in to Oracle Database as a privileged user, such as `SYSTEM`.
2. For a list of all initialization parameters and their current settings, issue this SQL*Plus command:

   ```
   show parameters
   ```

3. Issue `ALTER SYSTEM` commands, using values appropriate for your system, to change the value of the parameters. For example, this command sets PROCESSES to 800:

   ```
   ALTER SYSTEM SET processes=800 SCOPE=spfile;
   ```

4. Restart Oracle Database for the new settings to take effect.

Oracle UNDO Tablespace

Heavy query load should not coincide with heavy crawl activity, especially when there are large-scale changes on the target site. If it does, such as when a crawl is scheduled around the clock, then increase the size of the Oracle UNDO tablespace with the `UNDO_RETENTION` parameter.

> **See Also:** *Oracle Database SQL Language Reference* and *Oracle Database Administrator's Guide* for more information about increasing the Oracle undo space

Buffer Cache

An Oracle SES search operation looks up the Oracle Text index and some internal tables to generate a hit list. To maintain the best search performance, reduce disk I/O as much as possible by keeping these objects in the buffer cache. If you have plenty of physical memory, you can enlarge the buffer cache so it can retain these objects.

The search operation accesses these database objects the most frequently:

Object Name	Partitioned Object Name	Object Type
`DR$EQ$DOC_PATH_IDX$X`	`DR#EQ$DOC_PATH_IDX`*`4-digit-ID`*`$X`	B-tree index
`DR$EQ$DOC_PATH_IDX$R`	`DR#EQ$DOC_PATH_IDX`*`4-digit-ID`*`$R`	Table
`DR$EQ$DOC_PATH_IDX$I`	`DR#EQ$DOC_PATH_IDX`*`4-digit-ID`*`$I`	Table

$X and $R are the most important and are typically smaller than $I. If the database has large KEEP pool or can support one, consider putting the $X and $R tables in it to maintain good performance when accessing them. While the $I table is also important for search, it can become too large to cache in its entirety.

Check the cache hit ratio for these objects regularly in Enterprise Manager or an Automatic Workload Repository (AWR) report. Crawling and optimization can change the size of these objects.

To put a table in the KEEP pool:

1. Open SQL*Plus or another SQL interface and connect as a privileged user.
2. Issue an `ALTER INDEX` command using this syntax, where *`table_name`* is the $R or $X table.

   ```
   ALTER INDEX table_name STORAGE(BUFFER_POOL KEEP)
   ```

3. Verify the new location of the table:

   ```
   SELECT buffer_pool FROM dba_indexes WHERE index_name = table_name;
   ```

Example 10–1 shows the SQL commands that put the $X file in the KEEP pool.

Example 10–1 Putting DREQDOC_PATH_IDX$X in the KEEP Pool

```
SQL> SELECT buffer_pool FROM dba_indexes WHERE index_name='DR$EQ$DOC_PATH_IDX$X';

BUFFER_POOL
----------------------
DEFAULT

SQL> ALTER INDEX dr$eq$doc_path_idx$x STORAGE (BUFFER_POOL KEEP);
```

```
Index altered.

SQL> SELECT buffer_pool FROM dba_indexes WHERE index_name='DR$EQ$DOC_PATH_IDX$X';

BUFFER_POOL
----------------------
KEEP
```

Turning On Debug Mode

The logging level for Oracle SES in Fusion Applications is controlled by either Enterprise Manager or the WebLogic Scripting Tool (WLST). The Oracle SES logger, named oracle.search, is set by default to the INFO level. For debugging, change the level of oracle.search to FINEST.

Messages for the Oracle SES server are logged in *MW_HOME*/servers/search_server1/logs/search_server1-diagnostic.log.

To change the logging level using WLST:

1. Go to *MW_HOME*/oracle_common/common/bin.
2. Start the WebLogic Server Administration Scripting Tool:
   - For Linux, enter `wlst.sh`.
   - For Windows, enter `wlst.cmd`.
3. Enter this command at the `wls/offline>` prompt:

   ```
   connect ()
   ```

4. Enter your WebLogic user name in response to the user-name prompt.
5. Enter your WebLogic password in response to the password prompt.
6. Enter the WebLogic server *hostname:port* in response to the server URL prompt.
7. Set the log level to `FINEST` for oracle.search:

   ```
   setLogLevel(target="search_server1",level="FINEST", logger="oracle.search")
   ```

> **See Also:** *Oracle Fusion Middleware WebLogic Scripting Tool Command Reference*

Supporting Failover in Oracle RAC

To support failover in Oracle RAC, you must change the database connection string in the credential store from a physical database node to a service representing multiple physical database nodes.

> **See Also:** "Introduction to Automatic Workload Management" in the *Oracle Real Application Clusters Administration and Deployment Guide*

To change the database connection string in the Credential Storage Framework:

1. Open a connection to the computer where Oracle Fusion Middleware is installed.
2. Go to *MW_HOME*/oracle_common/common/bin.
3. Start the WebLogic Server Administration Scripting Shell.
   - For Linux, enter `wlst.sh`.

- For Windows, enter `wlst.cmd`.

4. Enter this command at the `wls/offline>` prompt:

   ```
   connect ()
   ```

5. Enter your WebLogic user name in response to the user-name prompt.
6. Enter your WebLogic password in response to the password prompt.
7. Enter the WebLogic server URL in response to the server URL prompt. For example, `t3://localhost:7234`.
8. Enter this command in response to the `wls:`*`domain_name`*`/serverConfig>` prompt:

   ```
   updateCred("oracle.search","search_database",connect_string,"search")
   ```

   Where *`connect_string`* is the new database connection string, which uses the easy connect naming method in this format:

   ```
   jdbc:oracle:thin@hostname:port:sid
   ```

   For example:

   ```
   updateCred("oracle.search","search_database","jdbc:oracle:thin:@example.us.orac
   le.com:7890:fusion", "search")
   ```

   See Also: *Oracle Fusion Middleware WebLogic Scripting Tool Command Reference*

Monitoring Oracle Secure Enterprise Search

In a production environment, where a load balancer or other monitoring tools are used to ensure system availability, Oracle Secure Enterprise Search (Oracle SES) can be monitored easily at the following URL:

`http://`*`host:port`*`/monitor/check.jsp`.

The page should display the following message: **Oracle Secure Enterprise Search instance is up.**

This message is not translated to other languages because system monitoring tools might need to byte-compare this string.

If Oracle SES is not available, then the page displays either a connection error or the HTTP status code 503.

Integrating with Google Desktop

Oracle Secure Enterprise Search provides a GDfE plug-in to integrate with Google Desktop Enterprise Edition. You can include Google Desktop results in your Oracle SES hit list. You can also link to Oracle SES from the GDfE interface.

See Also: Google Desktop for Enterprise Plug-in Readme at `http://`*`host:port`*`/search/query/gdfe/gdfe_readme.html`

Accessing the Oracle WebLogic Server Administration Console

The Oracle WebLogic Server Administration Console is a Web browser-based user interface that displays the current status of the middle tier. For example, the Home

page shows a graph of the Response and Load, and the Performance page shows a graph of the Heap Usage.

To access the Oracle WebLogic Server Administration Console:

1. Enter the following URL in a Web browser, replacing *host:port* with the host name and port for the WebLogic Administration Console:

   `http://wls_host:wls_port/console`

2. Log in with your WebLogic administrative user name and password.

See Also: *Oracle Fusion Middleware Administrator's Guide*

11

Oracle Secure Enterprise Search APIs

This chapter explains the Oracle Secure Enterprise Search (Oracle SES) APIs and related information. This chapter contains the following topics:

- Overview of Oracle Secure Enterprise Search APIs
- Oracle Secure Enterprise Search Web Services APIs

See Also: *Oracle Secure Enterprise Search Java API Reference*

Overview of Oracle Secure Enterprise Search APIs

Oracle Secure Enterprise Search provides the following APIs:

Web Services APIs

The Web Services APIs are used to integrate Oracle SES search capabilities into your search application. Oracle SES provides Java proxy libraries. You either can use the Java libraries or create proxies, based on the published Web Services Description Language (WSDL) files, to access Oracle SES Web Services. Oracle SES provides two Web Services APIs:

- Query Web Services API
- Administration Web Services API

Oracle Secure Enterprise Search Web Services APIs

Oracle SES includes the following Web Services APIs:

- **Query Web Services API**: Enables you to perform search queries; for example, search for "oracle benefits" and return all the documents. You can also customize the default Oracle SES ranking to create a more relevant search result list for your enterprise or configure clustering for customized applications.
- **Administration Web Services API**: Enables you to perform various administrative tasks; for example, start or stop a crawl schedule, check schedule status, get the estimated index fragmentation level, and perform index optimization.

See Also:

- *Oracle Secure Enterprise Search Administration API Guide* for more information about Web services interface
- *Oracle Secure Enterprise Search Java API Reference*
- "Web Services Interface" section of the Oracle SES administration tutorial:

  `http://st-curriculum.oracle.com/tutorial/SESAdminTutorial/index.htm`

Oracle Secure Enterprise Search Web Services APIs let you write your own application to search and administer Oracle SES over the network. The APIs provide the following benefits:

- Applications can be deployed into any computer that connects to Oracle SES server through a standard Internet protocol.
- Web Services protocol is XML-based, which makes for easy application integration.

Oracle SES also provides the client-side Java proxies for marshalling and parsing Web Services SOAP messages. Client applications can use the library instead of creating SOAP requests and parsing SOAP responses by themselves to access Oracle SES Web Services.

This section contains the following topics:

- Web Services APIs Installation
- Web Services Concepts
- Web Services Architecture
- Query Web Services Common Data Types
- Query Web Services Operations
- Query Web Services Query Syntax
- Query Web Services Example: Basic Search
- Client-Side Query Java Proxy Library

Web Services APIs Installation

Oracle SES Web Services runs on Oracle WebLogic Server. They are installed and configured as part of the default installation. You can use Oracle SES Web Services immediately. Follow the same middle tier administration steps to start and stop Oracle SES Web Services. Note that the Query Web Service client API should be run in version JDK 1.6 or higher.

WebLogic provides a default Oracle SES Web Services administrator console. The administrator console URL matches the Oracle SES Web Services URL.

Query Web Services Location

The Query Web service is located at the following address for an Oracle SES installation:

`http://host:port/search/query/OracleSearch`

For example, if your Oracle SES middle tier is running on host `myhost` and the port number is `8888`, then the Query Web Services URL is the following:

```
http://myhost:8888/search/query/OracleSearch
```

You can obtain the following information from the administrator console:

- Oracle SES Query WSDL description
- List of Web Services messages and operations
- Client-side Java proxies and source codes

Administration Web Services Location

The Administration Web service is located at the following address for an Oracle SES installation:

`http://`*`host:port`*`/search/ws/admin/SearchAdmin`.

You can obtain the following information from the administrator console:

- Oracle SES Administration WSDL description
- List of Web Services messages and operations
- Client-side JavaScript stub

Web Services Concepts

Oracle SES Web Services consists of a remote procedure call (RPC) interface to Oracle SES that enables the client application to invoke operations on Oracle SES over the network. The client application uses WSDL specification published by Oracle SES Web Services URL to send a request message using Simple Object Access Protocol (SOAP). The server then responds to the client application with a SOAP response message.

This section explains the following concepts:

- Web Services
- Simple Object Access Protocol
- Web Services Description Language

Web Services

A Web Service is a software application identified by a URI whose interfaces and binding are capable of being defined, described, and discovered by XML artifacts. A Web Service supports direct interactions with other software applications using XML-based messages and internet-based products.

A Web Service does the following:

- Exposes and describes itself: A Web Service defines its functionality and attributes so that other applications can understand it. By providing a WSDL file, a Web Service makes its functionality available to other applications.
- Allows other services to locate it on the Web: A Web Service can be registered in a UDDI registry so that applications can locate it.
- Can be invoked: After a Web Service has been located and examined, the remote application can invoke the service using an Internet standard protocol.
- Web Services are of either request and response or one-way style, and they can use either synchronous or asynchronous communication. However, the fundamental

unit of exchange between Web Services clients and Web Services, of either style or type of communication, is a message.

Simple Object Access Protocol

The Simple Object Access Protocol (SOAP) is a lightweight XML-based protocol for exchanging information in a decentralized distributed environment. SOAP supports different styles of information exchange, including RPC-oriented and message-oriented exchange. RPC style information exchange allows for request-response processing, where an endpoint receives a procedure-oriented message and replies with a correlated response message. Message-oriented information exchange supports organizations and applications that must exchange messages or other types of documents where a message is sent, but the sender might not expect or wait for an immediate response. Message-oriented information exchange is also called document style exchange.

SOAP has the following features:

- Protocol independence
- Language independence
- Platform and operating system independence
- Support for SOAP XML messages incorporating attachments (using the multipart MIME structure)

Web Services Description Language

The Web Services Description Language (WSDL) is an XML format for describing network services containing RPC-oriented and message-oriented information. Programmers or automated development tools can create WSDL files to describe a service and can make the description available over the Internet. Client-side programmers and development tools can use published WSDL specifications to obtain information about available Web Services and to build and create proxies or program templates that access available services.

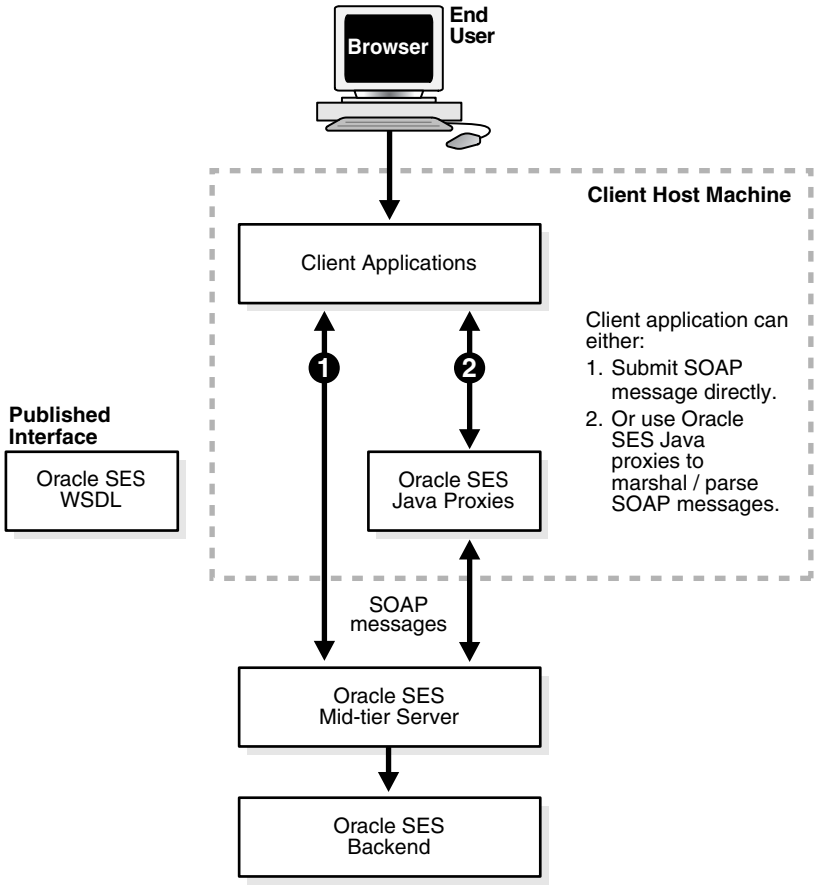
Web Services Architecture

Oracle SES Web Services is powered by Oracle WebLogic Server. The implementation, configuration, and deployment of Oracle SES Web Services follow the procedures and standards provided by Oracle WebLogic Server.

Oracle SES WSDL defines the operations and messages for Oracle SES Web Services. The message exchange of Oracle SES Web Services is RPC style, in which the contents of the SOAP message body conform to a structure that specifies a procedure and includes a set of parameters or a response with a result and any additional parameters.

Oracle SES SOAP messages use HTTP binding where a SOAP message is embedded in the body of a HTTP request and a SOAP message is returned in the HTTP response.

The following diagram illustrates the architecture of Oracle SES Web Services:



Development Platforms

You can implement client applications using platforms that support SOAP, such as Oracle JDeveloper, Microsoft .NET, or Apache Axis. These platforms allow you to automatically create code using the Oracle SES WSDL interface. Include the generated code along with the application logic to create a request, invoke the Web Services, and interpret the response.

Query Web Services Common Data Types

This section contains the following topics:

- Base Data Types
- XML-to-Java Data Type Mappings
- Complex Types
- Array Types

Base Data Types

Oracle Secure Enterprise Search Web Services use the following base data types:

Table 11–1 Base Data Types

Base Type	Description	Example
xsd:Boolean	Boolean	true, false
xsd:dateTime	Date	2005-12-31

Table 11–1 (Cont.) Base Data Types

Base Type	Description	Example
xsd:int	Integer	256
xsd:long	Long integer	12345678900
xsd:string	String	Oracle Secure Enterprise Search

XML-to-Java Data Type Mappings

The mapping between XML schema data types and Java data types depends on the SOAP development environment. The following table shows mappings for the Oracle JDeveloper environment:

Table 11–2 XML-to-Java Type Mappings

XML Schema	Oracle JDeveloper
xsd:Boolean	java.lang.Boolean
xsd:dateTime	java.util.Date
xsd:int	java.lang.Integer
xsd:long	java.lang.Long
xsd:string	java.lang.String

Complex Types

Oracle SES Web Services uses the following complex data types:

- OracleSearchResult
- ResultElement
- SCElement
- DataGroup
- Attribute
- Filter
- Node
- AttributeLOVElement
- SessionContextElement
- Status
- Language

OracleSearchResult The search result container. It has the following elements:

- `returnCount`: A Boolean value indicating whether the result includes the count estimate for the hit list.
- `estimatedHitCount`: The estimated count of the search result; -1 means the search result does not return the estimated hit count.
- `dupRemoved`: A Boolean value indicating whether **near duplicate documents** have been removed from search result.

- `dupMarked`: A Boolean value indicating whether **near duplicate documents** have been marked in the search result. If `dupRemoved` is true, then `dupMarked` is always false.
- `resultElements`: An array of `resultElement`, which represents the actual hit list.
- `suggestedLinks`: An array of `suggestedLink` for the given search.
- `query`: The actual search string, which uses Oracle SES query syntax.
- `altKeywords`: Alternate keywords (suggestions) for a given search. If you set `switch ses.qapp.multiple_alternate_keywords` to `true`, then multiple alternate keywords are returned for a search term. The following terms are returned:
  - **ALTERNATE_SPELLING**: Contains alternate words from the dictionary.
  - **EXPANDED_QUERY**: Contains keywords from the alternate keywords page when Auto-Expand is selected. These keywords are automatically included in the query that is sent to Oracle Text.
  - **ALTERNATE_QUERY**: Contains keywords from the alternate keywords page when Auto-Expand is not selected.
- `startIndex`: The start index of search results.
- `docsReturned`: The number of search matches returned.

ResultElement This is the data type for search result element. It has the following elements:

- `author`: Primary author of the document
- `description`: Description of the document
- `url`: URL of the document
- `snippet`: Keywords in context (KWIC) of the document
- `title`: Title of the document
- `lastModified`: Last modified date of the document
- `mimetype`: Mime type of the document
- `score`: Oracle Text score of the document
- `docID`: Document ID
- `language`: Language of the document
- `contentLength`: Content length of the document
- `signature`: Signature of the document
- `infoSourceID`: InfoSource ID of the document
- `infoSourcePath`: InfoSource path of the document
- `groups`: Array of groups to which the document belongs
- `isDuplicate`: Boolean value indicating whether this document is a near duplicate of another document in the result list
- `hasDuplicate`: Boolean value indicating whether this document has one or more near duplicates in the result list

- `fedID`: Federated instance ID, used to track which federated instance the document is fetched from
- `customAttributes`: Array of custom nondefault attributes extracted from/for the document during crawling

  `CustomAttribute` encapsulates the name and value of the custom (user-defined) attribute. The name of the attribute is represented by actual name and type of the attribute in name_type format. For example, the string attribute foo is represented as `foo_STRING`. All Date attributes use the format `mm/dd/yyyy`.

SCElement Suggested content from a provider. It has the following elements:

- `name`: name of the suggested content provider
- `content`: suggested content from the provider. The content is a byte array of the XML or HTML content

DataGroup The source group. It has the following elements:

- `groupID`: Source group ID
- `groupName`: Source group name
- `groupDisplayName`: Display name for the source group

Attribute The data type for search attribute. It has the following elements:

- `id`: Search attribute ID
- `name`: Internal name of search attribute
- `displayName`: Display name of search attribute
- `type`: The search attribute type. Value is either number, string, or date.

Filter The data type for filter condition (predicate). It has the following elements:

- `attributeId`: Search attribute ID
- `attributeType`: Search attribute type. Value is either number, string, or date.
- `operator`: Operator of the filter condition
  - If `attributeType` is string, then it should be either equals or contains.
  - If `attributeType` is number or date, then it should be either greaterthan, greaterthanequals, lessthan, lessthanequals, or equals.
- `attributeValue`: Value of the filter condition (predicate)
  - For string type attribute, the value is simply the string itself.
  - For number type attribute, the value should be represented by a string consisting of an optional sign, (+) or (-), followed by a sequence of zero or more decimal digits ("the integer"), optionally followed by a fraction. The fraction consists of a decimal point followed by zero or more decimal digits. The string must contain at least one digit in either the integer or the fraction.
  - For date type attribute, the value should be in the format `mm/dd/yyyy`, where `mm` is the month (01~12), `dd` is the date (01~31), `yyyy` is the year (for example, 2005)

Examples:

- If the filter condition is Title contains `Oracle Secure Enterprise Search`, then the client application must look up the attribute ID of search attribute `Title` and include the following (element, value) pairs:
  - `attributeID = 1` (assuming the search attribute id of `Title` is 1)
  - `operator = contains`
  - `attributeValue = Oracle Secure Enterprise Search`
- If the filter condition is Price greater than 1000, then the client application must look up the attribute ID of search attribute `Price` and include the following (element, value) pairs:
  - `attributeID = 2` (assuming the search attribute id of 'Price' is 2)
  - `operator = greaterthan`
  - `attributeValue = 1000`

Node This is the data type for the infosource node. It has the following elements:

- `id`: Infosource node ID
- `fedId`: Federated instance ID, used to track which federated instance the node belongs to
- `name`: Name of the node
- `docCount`: Number of documents under the node. If the value is –1, then there exists documents under the node but the count cannot be shown.
- `hasChildren`: Indicates if the node has any children
- `fullpath`: Full path of the category node
- `fullpathIds`: The IDs of each node in the full path

AttributeLOVElement This is the element of `AttributeLOV`, the list of search attribute values. It has the following elements:

- `value`: Attribute value (internal value)
- `displayValue`: Display value

SessionContextElement This data structure is used to store authentication information for the search user in the form of a name-value pair, which can be used during query-time authorization filtering of the results. It has the following elements:

- `name`: Name of the authentication attribute
- `value`: Value of the authentication attribute

Status This is the status of the request. It has the following elements:

- `status`: Status code. Value is either successful or 'ailed
- `message`: Status message. Value is null, or an error message if the status is 'ailed

Language This is the language data type. It has the following elements:

- `languageName`: Name of the language
- `languageDisplayName`: Display name (translated name) of the language

Array Types

Oracle Secure Enterprise Search Web Services uses the following complex array types:

- `AttributeArray`: Array of `Attribute`
- `AttributeLOVElementArray`: Array of `AttributeLOVElement`
- `CustomAttributeArray`: Array of `CustomAttribute`
- `SCElementArray`: Array of `SCElement`
- `DataGroupArray`: Array of `DataGroup`
- `FilterArray`: Array of `Filter`
- `IntArray`: Array of `int`
- `LanguageArray`: Array of `Language`
- `NodeArray`: Array of `Node`
- `ResultElementArray`: Array of `ResultElement`
- `SessionContextElementArray`: Array of `SessionContextElement`
- `StringArray`: Array of `String`

Query Web Services Operations

This section contains the following topics:

- Overview of Query Web Services Operations
- Authentication Operations
- Search Operations
- Browse Operations
- Metadata Operations
- Search Hit Operations
- User Feedback Operations

Overview of Query Web Services Operations

Oracle Secure Enterprise Search provides the following categories of Web Services operations:

- **Authentication**: Authenticates a user's access to Oracle SES. The operation is only required if the user performs secure search.
- **Search**: Runs a search on Oracle SES and obtains a hit list along with information such as estimated hit count, **near duplicate documents** in the result list, suggested links, and alternate keywords for the search. Gets suggested content from external providers for the given query. You can also customize the default Oracle SES ranking to create a more relevant search result list for your enterprise or configure clustering for customized Oracle SES applications.
- **Metadata**: Obtains the search metadata, such as the list of source groups, the list of supported languages, or the list of search attributes.
- **Search Hit**: Obtains the search result details, such as the cached version of search result and in-links and out-links of the search hit.
- **User Feedback**: Sends user feedback to Oracle SES, such as user-submitted URLs.

See Also: "Query Web Services Operations" on page 11-10

Authentication Operations

This section describes the following authentication operations:

- loginRequest Message
- loginResponse Message
- logoutRequest Message
- logoutResponse Message
- setSessionContextRequest Message
- setSessionContextResponse Message
- proxyLoginRequest Message
- proxyLoginResponse Message

loginRequest Message Requests Oracle SES to authenticate the search user. It consists of the following parameters:

- `username`: User name for the search user. This value is *not* case-sensitive.
- `password`: Password for the search user

```
<message name="loginRequest">
   <part name="username"      type="xsd:string"/>
   <part name="password"      type="xsd:string"/>
</message>
```

loginResponse Message Contains the return status for the `loginRequest` message.

```
<message name="loginResponse">
   <part name="return"      type="typens:Status"/>
</message>
```

logoutRequest Message Used when the user logs out from the search application.

```
<message name="logoutRequest">
</message>
```

logoutResponse Message Contains the return status for the `logoutRequest` message.

```
<message name="logoutResponse">
   <part name="return"      type="typens:Status"/>
</message>
```

setSessionContextRequest Message Passes authentication information for the search user, which can be used during query-time filtering. It consists of the following parameter:

- `sessionContext`: An array of `SessionContextElement`. This array stores the authentication information needed for the query-time authentication filtering in the form of name-value pairs.

```
<message name="setSessionContextRequest">
    <part name="sessionContext"     type="typens:SessionContextElementArray"/>
</message>
```

Note: Login and logout Web Services calls cause Oracle SES to automatically set or reset the `AUTH_USER` value in the session context that is passed to the query-time filter. This session context attribute cannot be overwritten explicitly through the `setSessionContext` call.

setSessionContextResponse Message Contains the return status for the `setSessionContext` message.

```
<message name="setSessionContextResponse">
  <part name="return"            type="typens:Status"/>
</message>
```

proxyLoginRequest Message Logs in the end user to Oracle SES using proxy authentication. It consists of following parameters:

- `username`: User name of the proxy user
- `password`: Password of the proxy user
- `searchUser`: User name of the end user

```
<message name="proxyLoginRequest">
  <part name="username"           type="xsd:string"/>
  <part name="password"           type="xsd:string"/>
  <part name="searchUser"         type="xsd:string"/>
</message>
```

The proxy user must be a federation trusted entity created on the Oracle SES instance.

See Also: "Federation Trusted Entities" on page 5-13

proxyLoginResponse Message This message contains the return status for the `proxyLoginRequest` message.

```
<message name="proxyLoginResponse">
  <part name="return"        type="typens:Status"/>
</message>
```

Search Operations

This section describes the following search operations:

- doOracleAdvancedSearch Message
- doOracleFetchSearch Message
- doOracleOrganizedSearch Message
- doOracleSearch Message
- doOracleSearchResponse Message
- doOracleBrowseSearch Message
- doOracleBrowseSearchResponse Message
- doOracleSimpleSearch Message
- doOracleSimpleSearchResponse Message
- getSuggestedContent Message
- getSuggestedContentResponse Message

doOracleAdvancedSearch Message Invokes Oracle SES advanced search and returns search results. It consists of the following parameters:

- `query`: The search string. This should follow Oracle SES query syntax. See "Query Web Services Query Syntax" on page 11-24 for details.
- `startIndex`: Index of the first document in the hitlist to be returned. The default is 1 if not set explicitly.
- `docsRequested`: The maximum number of documents in the hitlist to be returned. The default is 10 if not set explicitly.
- `dupRemoved`: Boolean flag to enable or disable duplicate removal. If turned on, then duplicate documents in the hitlist are removed. The default is false if not set explicitly. Note: The `dupMarked` switch has no effect when `dupRemoved` is turned on.
- `dupMarked`: Boolean flag to enable or disable duplicate detection. The default is false if not set explicitly. Note: The `dupMarked` switch has no effect when `dupRemoved` is turned on.
- `groups`: Data source groups that the search is restricted to. The default is all groups if not set explicitly.
- `queryLang`: Language of the query. This is equivalent to Locale. The default is English (`en`) if not set explicitly. This is used in relevancy boosting.
- `docLang`: Language of the documents to restrict the search. If the value is not set explicitly, then search is performed against documents of all the languages.
- `returnCount`: Boolean flag to fetch the total hit count with the result. The default is false if not set explicitly.
- `filterConnector`: Connector between all the filters: "and" indicates that the documents in the hitlist must satisfy all the filters, and "or" indicates that the documents in the hitlist must satisfy at least one filter. The default is "and" if not set explicitly.
- `filters`: An array of filters. Each filter is a restriction on the search result. Filters are connected by the `filterConnector`. The default is null (no filter applies to the search result) if not set explicitly.
- `fetchAttributes`: Array of integers representing the IDs of custom or nondefault attributes to be fetched with the search result
- `searchControls`: XML string to specify advanced filter conditions and ranking parameters

> **Note:** The attribute filter `LastModifiedDate` uses the format `mm/dd/yyyy`.

```
public OracleSearchResult doOracleAdvancedSearch(
        String query,
        Integer startIndex,
        Integer docsRequested,
        Boolean dupRemoved,
        Boolean dupMarked,
        DataGroup[] groups,
        String queryLang,
        String docLang,
        Boolean returnCount,
```

```
        String filterConnector,
        Filter[] filters,
        Integer[] fetchAttributes,
        String searchControls)
    throws Exception
```

doOracleFetchSearch Message Invokes Oracle SES fetch search and returns fetch results. It consists of the following parameters:

- `query`: The search string. This should follow Oracle SES query syntax. See "Query Web Services Query Syntax" on page 11-24 for details.
- `targetDocIdList`: Target document ID list, most likely from a cluster node.
- `startIndex`: Index of the first document in the hitlist to be returned. The default is 1 if not set explicitly.
- `docsRequested`: Maximum number of documents in the hitlist to be returned. The default is 10 if not set explicitly.
- `queryLang`: Language of the query. This is equivalent to Locale. The default is English (`en`) if not set explicitly. This is used in relevancy boosting.
- `fetchAttributeNames`: Array of names of custom or nondefault attributes to be fetched with the search result.
- `groupAttr`: Attribute used for grouping.
- `sortAttrList`: List of sorting attribute settings.
- `clusterList`: List of cluster configurations.

```
public OracleResultContainer doOracleFetchSearch(
        String query,
        String[] targetDocIdList,
        Integer startIndex,
        Integer docsRequested,
        String queryLang,
        String[] fetchAttributeNames,
        GroupAttribute groupAttr,
        SortAttribute[] sortAttrList,
        ClusterConfig[] clusterList)
    throws Exception
```

doOracleOrganizedSearch Message This invokes Oracle SES organized search and returns search results. It consists of the following parameters:

`query`: The search string. This should follow Oracle SES query syntax. See "Query Web Services Query Syntax" on page 11-24 for details.

`topN`: Top N search result for grouping, sorting, and clustering.

`startIndex`: Index of the first document in the hitlist to be returned. The default is 1 if not set explicitly.

`docsRequested`: Maximum number of documents in the hitlist to be returned. The default is 10 if not set explicitly.

`dupRemoved`: Boolean flag to enable or disable duplicate removal. If turned on, duplicate documents in the hitlist are removed. The default is false if not set explicitly. Note: The `dupMarked` switch has no effect when `dupRemoved` is turned on.

`dupMarked`: Boolean flag to enable or disable duplicate detection. The default is false if not set explicitly. The `dupMarked` switch has no effect when `dupRemoved` is turned on.

`groups`: Data source groups that the search is restricted to. The default is all groups if not set explicitly.

`queryLang`: Language of the query. This is equivalent to Locale. The default is English (`en`) if not set explicitly. This is used in relevancy boosting.

`docLang`: Language of the documents to restrict the search. If the value is not set explicitly, then search is performed against documents of all the languages.

`returnCount`: Boolean flag to fetch the total hit count with the result. The default is false if not set explicitly.

`filterConnector`: Connector between all the filters: "and" indicates that the documents in the hitlist must satisfy all filters, "or" indicates that the documents in the hitlist must satisfy at least one filter. The default is "and" if not set explicitly.

`filters`: An array of filters. Each filter is a restriction on the search result. Filters are connected by the `filterConnector`. The default is null (no filter applies to the search result) if not set explicitly.

`fetchAttributeNames`: Array of names of custom or nondefault attributes to be fetched with the search result.

`searchControls`: XML string to specify advanced filter conditions and ranking parameters.

`groupAttr`: Attribute used for grouping.

`sortAttrList`: List of sorting attribute settings.

`clusterList`: List of cluster configurations.

```
public OracleResultContainer doOracleOrganizedSearch(
        String query,
        Integer topN,
        Integer startIndex,
        Integer docsRequested,
        Boolean dupRemoved,
        Boolean dupMarked,
        DataGroup[] groups,
        String queryLang,
        String docLang,
        Boolean returnCount,
        String filterConnector,
        Filter[] filters,
        String[] fetchAttributeNames,
        String searchControls,
        GroupAttribute groupAttr,
        SortAttribute[] sortAttrList,
        ClusterConfig[] clusterList)
    throws Exception
```

doOracleSearch Message This is the main message for the search application. It consists of the following parameters:

- `query`: A search string. It must be a valid string and it cannot be null. The search string should follow Oracle SES query syntax. See "Query Web Services Query Syntax" on page 11-24 for details.
- `startIndex`: The index of the first result to be returned. For example, if there are 67 results, you might want to start at 20. The default is 1 if not set explicitly.
- `docsRequested`: The maximum number of results to be returned. The default is 10 if not set explicitly.

- `dupRemoved`: Enable or disable duplicate removal. If turned on, then the search result eliminate all **near duplicate documents** from the result list. The `dupMarked` switch has no effect when `dupRemoved` is turned on. The default is false if not set explicitly.
- `dupMarked`: Enable or disable duplicate detection. If `dupRemoved` is turned off and `dupMarked` is turned on, then the search result keeps all **near duplicate documents** from the result list and marks them as duplicates. If `dupRemoved` is turned on, then the `dupMarked` switch has no effect. The default is false if not set explicitly.
- `groups`: Limit the search result to the documents from specified source groups. The default is for all groups if not set explicitly.
- `queryLang`: The query language argument should be a valid ISO language code. These codes are the lower-case, two-letter codes as defined by ISO-639. Examples: "en" for English and "de" for German. The default is English ("en") if not set explicitly. This is used for relevancy boosting.
- `docLang`: Set the language of the documents to limit the search. If the value is not set explicitly, then search is performed against documents of all the languages.
- `returnCount`: Set to true to return total hit count with the result. The default is false if not set explicitly.
- `filterConnector`: The connector between all filters: "and" indicates the search result must satisfy all filters, "or" indicates the search result must satisfy at least one filter. The default is "and" if not set explicitly.
- `filters`: An array of filters. Each filter is a restriction on search results. Filters are connected by `filterConnector`. The default is null (no filter applies to the search result) if not set explicitly.
- `fetchAttributes`: Array of integers representing the nondefault attribute IDs to be fetched in the `resultElements`. The default is null (or set one int value '0'), so no attributes other than default-attributes are fetched in the `resultElements`.

```
<message name="doOracleSearch">
   <part name="query"            type="xsd:string"/>
   <part name="startIndex"       type="xsd:int"/>
   <part name="docsRequested"    type="xsd:int"/>
   <part name="dupRemoved"       type="xsd:boolean"/>
   <part name="dupMarked"        type="xsd:boolean"/>
   <part name="groups"           type="typens:DataGroupArray"/>
   <part name="queryLang"        type="xsd:string"/>
   <part name="docLang"          type="xsd:string"/>
   <part name="returnCount"      type="xsd:boolean"/>
   <part name="filterConnector"  type="xsd:string"/>
   <part name="filters"          type="typens:FilterArray"/>
   <part name="fetchAttributes"  type="typens:IntArray"/>
</message>
```

doOracleSearchResponse Message This message returns the search result in `OracleSearchResult` data type.

```
<message name="doOracleSearchResponse">
    <part name="return"  type="typens:OracleSearchResult"/>
</message>
```

doOracleBrowseSearch Message This message restricts a search to a particular node. It consists of the following parameters:

- `query`: A search string. It must be a valid string, and it cannot be null. The search string should follow Oracle SES query syntax. See "Query Web Services Query Syntax" on page 11-24 for more details.
- `nodeID`: The ID of the node to restrict the search to.
- `fedID`: The ID of the federated instance the parent node belongs to (null for local node).
- `startIndex`: The index of the first result to be returned. For example, if there are 67 results, then you might want to start at 20. The default is 1 if not set explicitly.
- `docsRequested`: The maximum number of results to be returned. The default is 10 if not set explicitly.
- `dupRemoved`: Enable or disable duplicate removal. If turned on, then the search result eliminate all **near duplicate documents** from the result list, and the `dupMarked` switch have no effect when `dupRemoved` is turned on. The default is false if not set explicitly.
- `dupMarked`: Enable or disable duplicate detection. If `dupRemoved` is turned off and `dupMarked` is turned on, then the search result keeps all **near duplicate documents** from the result list and marks them as duplicates. If `dupRemoved` is turned on, then the `dupMarked` switch has no effect. The default is false if not set explicitly.
- `queryLang`: The query language argument should be a valid ISO language code. These codes are the lower-case, two-letter codes as defined by ISO-639. Examples: "en" for English and "de" for German. The default is English ("en") if not set explicitly. This is used for relevancy boosting.
- `docLang`: Set the language of the documents to limit the search. If the value is not set explicitly, then search is performed against documents of all the languages.
- `returnCount`: Set to true to return total hit count with the result. The default is false if not set explicitly.
- `fetchAttributes`: Array of integers representing the nondefault attribute IDs to be fetched in the `resultElements`. The default is null (or set one int value '0'), so no attributes other than default-attributes are fetched in the `resultElements`.

```
<message name="doOracleBrowseSearch">
  <part name="query"              type="xsd:string"/>
  <part name="nodeID"             type="xsd:string"/>
  <part name="fedID"              type="xsd:string"/>
  <part name="startIndex"         type="xsd:int"/>
  <part name="docsRequested"      type="xsd:int"/>
  <part name="dupRemoved"         type="xsd:boolean"/>
  <part name="dupMarked"          type="xsd:boolean"/>
  <part name="queryLang"          type="xsd:string"/>
  <part name="docLang"            type="xsd:string"/>
  <part name="returnCount"        type="xsd:boolean"/>
  <part name="fetchAttributes"    type="typens:IntArray"/>
</message>
```

doOracleBrowseSearchResponse Message Returns the search result in `OracleSearchResult` data type.

```
<message name="doOracleBrowseSearchResponse">
    <part name="return"  type="typens:OracleSearchResult"/>
</message>
```

doOracleSimpleSearch Message A simplified form of the doOracleSearch message. In this message you do not need to specify the advanced search parameters that are specified in the doOracleSearch message. It consists of following parameters:

- `query`: A search string. It must be a valid string and it cannot be null. The search string should follow Oracle SES query syntax. See "Query Web Services Query Syntax" on page 11-24 for details.
- `startIndex`: The index of the first result to be returned. For example, if there are 67 results, you might want to start at 20. The default is 1, if not set explicitly.
- `docsRequested`: The maximum number of results to be returned. The default is 10, if not set explicitly.
- `dupRemoved`: Enable or disable duplicate removal. If turned on, then the search result eliminates all **near duplicate documents** from the result list. The `dupMarked` switch has no effect when `dupRemoved` is turned on. The default is false if not set explicitly.
- `dupMarked`: Enable or disable duplicate detection. If `dupRemoved` is turned off and `dupMarked` is turned on, then the search result keeps all **near duplicate documents** from the result list and marks them as duplicates. If `dupRemoved` is turned on, then the `dupMarked` switch has no effect. The default is false if not set explicitly.
- `returnCount`: Set to true to return total hit count with the result. The default is false if not set explicitly.

```
<message name="doOracleSimpleSearch">
  <part name="query"              type="xsd:string"/>
  <part name="startIndex"         type="xsd:int"/>
  <part name="docsRequested"      type="xsd:int"/>
  <part name="dupRemoved"         type="xsd:boolean"/>
  <part name="dupMarked"          type="xsd:boolean"/>
  <part name="returnCount"        type="xsd:boolean"/>
</message>
```

doOracleSimpleSearchResponse Message Returns the search result in `OracleSearchResult` data type.

```
<message name="doOracleSimpleSearchResponse">
  <part name="return"         type="typens:OracleSearchResult"/>
</message>
```

getSuggestedContent Message Returns the suggested content for the given query. It consists of the following parameters:

- `query`: Query string
- `returnType`: Format in which the content is to be returned, either `html` or `xml`. If no style sheet is configured for a given provider, then the return type is the return type of the content returned by the provider, regardless of whether `html` or `xml` is specified.

```
<message name="getSuggestedContent">
  <part name="query"                  type="xsd:string"/>
  <part name="returnType"             type="xsd:string"/>
</message>
```

getSuggestedContentResponse Message Returns the suggested content for the query.

```
<message name="getSuggestedContentResponse">
  <part name="return"                 type="typens:SCElementArray"/>
```

```
</message>
```

Browse Operations

This section describes the following browse operations:

- getInfoSourceNodesRequest Message
- getInfoSourceNodesResponse Message
- getInfoSourceAncestorNodesRequest Message
- getInfoSourceAncestorNodesResponse Message
- getInfoSourceNodeRequest Message
- getInfoSourceNodeResponse Message

getInfoSourceNodesRequest Message Obtains the list of info source nodes given the parent node ID. It consists of the following parameters:

- `parentNodeID`: The node ID for which all children nodes are returned. If it is not set, then the message returns all the root nodes.
- `fedID`: The ID of the federated instance the parent node belongs to (null for local node).
- `locale`: A two letter representation of locale. The default is English ("en") if not set explicitly.

```
<message name="getInfoSourceNodesRequest">
    <part name="parentNodeID"      type="xsd:string"/>
    <part name="fedID"             type="xsd:string"/>
    <part name="locale"            type="xsd:string"/>
</message>
```

getInfoSourceNodesResponse Message Returns an array of info source nodes.

```
<message name="getInfoSourceNodesResponse">
    <part name="nodes"    type="typens:NodeArray"/>
</message>
```

getInfoSourceAncestorNodesRequest Message Obtains the full path of a node, from root to node, given an info source node. It consists of the following parameters:

- `nodeID`: The node ID for which all the nodes in the path from root to node are returned; nodeID must be set and it cannot be null.
- `locale`: A two letter representation of locale. The default is English ("en") if not set explicitly.

```
<message name="getInfoSourceAncestorNodesRequest">
    <part name="nodeID"        type="xsd:string"/>
    <part name="locale"        type="xsd:string"/>
</message>
```

> **Note:** The `getInfoSourceAncestorNode` messages have been deprecated in Oracle SES.

getInfoSourceAncestorNodesResponse Message Returns an array of info source ancestor nodes.

```
<message name="getInfoSourceAncestorNodesResponse">
```

```
    <part name="nodes"    type="typens:NodeArray"/>
</message>
```

getInfoSourceNodeRequest Message Retrieves a particular node. It consists of the following parameters:

- `nodeID`: The node ID of the node to get, nodeID must be set and it cannot be null.
- `fedID`: The ID of the federated instance the parent node belongs to (null for local node).
- `locale`: A two letter representation of Locale, the default is English ("en") if not set explicitly.

Message format:

```
<message name="getInfoSourceNodeRequest">
   <part name="nodeID        "type="xsd:string"/>
   <part name="fedID"        "type="xsd:string"/>
   <part name="locale        "type="xsd:string"/>
</message>
```

getInfoSourceNodeResponse Message This message returns the node requested.

```
<message name="getInfoSourceNodeResponse">
    <part name="node    "type="typens:Node"/>
</message>
```

Metadata Operations

This section describes the following metadata operations:

- getLanguageRequest Message
- getLanguageResponse Message
- getDataGroupsRequest Message
- getDataGroupsResponse Message
- getAttributesRequest Message
- getAttributesResponse Message
- getAllAttributesRequest Message
- getAllAttributesResponse Message
- getAttributeLOVRequest Message
- getAttributeLOVResponse Message

getLanguageRequest Message Obtains all the languages supported by Oracle SES. It is used by the client application to display the list of languages. It consists of the following parameter:

`locale`: A two letter representation of locale. The default is English (`en`) if not set explicitly.

```
<message name="getLanguagesRequest">
    <part name="locale"       type="xsd:string"/>
</message>
```

getLanguageResponse Message

This message returns all supported languages.

```
<message name="getLanguagesResponse">
    <part name="return"        type="typens:LanguageArray"/>
</message>
```

getDataGroupsRequest Message Requests for all source groups defined in Oracle SES. It is used by the client application to show all source groups in the search page, such that the end user can restrict their search results within one or multiple source groups. It consists of the following parameter:

`locale`: A two letter representation of locale. The default is English (`en`) if not set explicitly.

```
<message name="getDataGroupsRequest">
    <part name="locale"        type="xsd:string"/>
</message>
```

getDataGroupsResponse Message Returns all source groups defined in Oracle SES.

```
<message name="getDataGroupsResponse">
    <part name="groups"        type="typens:DataGroupArray"/>
</message>
```

getAttributesRequest Message Obtains a list of search attributes that applied to the given source groups. It consists of the following parameters:

- `locale`: A two letter representation of locale. The default is English (`en`) if not set explicitly.
- `groups`: Limit the request to the attributes from specified source groups. The default is all groups if not set explicitly.
- `groupConnector`: The connector between all groups: "and" indicates the response is the attributes available in the set of source groups by finding the intersection of each group's attributes, "or" indicates the response is the attributes available in the set of source groups by finding the union of each group's attributes. The default is "or" if not set explicitly.

```
<message name="getAttributesRequest">
    <part name="locale"          type="xsd:string"/>
    <part name="groups"          type="typens:DataGroupArray"/>
    <part name="groupConnector"  type="xsd:string"/>
</message>
```

getAttributesResponse Message Returns an array of search attributes.

```
<message name="getAttributesResponse">
    <part name="return"       type="typens:AttributeArray"/>
</message>
```

getAllAttributesRequest Message Obtains all search attributes defined in Oracle SES. It consists of the following parameter:

`locale`: A two letter representation of locale. The default is English (`en`) if not set explicitly.

```
<message name="getAllAttributesRequest">
  <part name="locale" type="xsd:string"/>
</message>
```

getAllAttributesResponse Message Returns all search attributes defined in Oracle SES.

```
<message name="getAllAttributesResponse">
    <part name="return"       type="typens:AttributeArray"/>
```

```
</message>
```

getAttributeLOVRequest Message Obtains the LOV items given a search attribute. It consists of the following parameters:

- `attribute`: A search attribute for the LOV (list of values) requested.
- `locale`: A two letter representation of locale. The default is English ("en") if not set explicitly.

```
<message name="getAttributeLOVRequest">
  <part name="attribute"     type="typens:Attribute"/>
  <part name="locale"        type="xsd:string"/>
</message>
```

getAttributeLOVResponse Message Returns an array of search attribute LOV elements.

```
<message name="getAttributeLOVResponse">
  <part name="return"    type="typens:AttributeLOVElementArray"/>
</message>
```

Search Hit Operations

This section describes the following search hit operations:

- getCachedPageRequest Message
- getCachedPageResponse Message
- getInLinksRequest Message
- getInLinksResponse Message
- getOutLinksRequest Message
- getOutLinksResponse Message
- logUserClickRequest Message
- logUserClickResponse Message

getCachedPageRequest Message Obtains the cached version of a document given the document ID and the search string. The search string is highlighted in the output. It consists of the following parameters:

- `query`: The search string.
- `docID`: The document ID to be fetched.
- `fedID`: The federated instance ID, used to track which federated instance the document is fetched from.

```
<message name="getCachedPageRequest">
    <part name="query"          type="xsd:string"/>
    <part name="docID"          type="xsd:int"/>
    <part name="fedID"          type="xsd:string"/>
</message>
```

getCachedPageResponse Message Returns the byte array of the cached HTML page.

```
<message name="getCachedPageResponse">
    <part name="return"         type="xsd:base64Binary"/>
</message>
```

getInLinksRequest Message Obtains all the incoming links for a given search hit (document). It consists of the following parameters:

- `docID`: The document ID for the incoming links to be fetched. It must be a valid document ID and it cannot be null.
- `maxNum`: The maximum number of incoming links requested. The default is 25 if not set explicitly.
- `fedID`: The federated instance ID, used to track which federated instance the document is fetched from.

```
<message name="getInLinksRequest">
    <part name="docID"                type="xsd:int"/>
    <part name="maxNum"               type="xsd:int"/>
    <part name="fedID"                type="xsd:string"/>
</message>
```

getInLinksResponse Message Returns an array of incoming link URL strings.

```
<message name="getInLinksResponse">
    <part name="return"      type="typens:StringArray"/>
</message>
```

getOutLinksRequest Message Obtains all the outgoing links for a given search hit (document). It consists of the following parameters:

- `docID`: The document ID for the outgoing links to be fetched. It must be a valid document ID and it cannot be null.
- `maxNum`: The maximum number of outgoing links requested. The default is 25 if not set explicitly.
- `fedID`: The federated instance ID, used to track which federated instance the document is fetched from.

```
<message name="getOutLinksRequest">
    <part name="docID"               type="xsd:int"/>
    <part name="maxNum"              type="xsd:int"/>
    <part name="fedID"               type="xsd:string"/>
</message>
```

getOutLinksResponse Message This message returns an array of outgoing link URL strings.

```
<message name="getOutLinksResponse">
    <part name="return"      type="typens:StringArray"/>
</message>
```

logUserClickRequest Message This message logs the user's click. It consists of the following parameters:

- `queryID`: ID of the submitted search.
- `urlID`: ID of the document that the user clicked.
- `infosourceID`: Infosource ID. If none, then –1 is used as the default value
- `position`: The position of the document in the result list. For example, the first hit on the page or ninth hit on the page.
- `fedID`: Federation ID. Specifies the federated instance on which the document resides.

```
<message name="logUserClickRequest">
```

```
    <part name="queryID"        type="xsd:int"/>
    <part name="urlID"          type="xsd:int"/>
    <part name="infoSourceID"   type="xsd:int"/>
    <part name="position"       type="xsd:int"/>
    <part name="fedID"          type="xsd:string"/>
</message>
```

logUserClickResponse Message Returns the URL of the clicked document.

```
<message name="logUserClickResponse">
  <part name="url"              type="xsd:string"/>
</message>
```

User Feedback Operations

This section describes the following user feedback operations:

- submitUrlRequest Message
- submitUrlResponse Message

submitUrlRequest Message Submits a URL to Oracle SES so that it crawls and indexes the URL. This operation consists of the following parameter:

- `url`: The URL to be submitted to the crawler so it can be crawled next time. It must be a valid URL and it cannot be null.

```
<message name="submitUrlRequest">
     <part name="url"        type="xsd:string"/>
</message>
```

submitUrlResponse Message Returns the status, which consists of two strings. The first is the submission status, which is either `successful` or `failed`. The second string is the error message when the submission status is `failed`.

```
<message name="submitUrlResponse">
     <part name="return"      type="typens:Status"/>
</message>
```

Query Web Services Query Syntax

This section describes the query syntax used in the Oracle Secure Enterprise Search Search API.

Search Term

A search term can be a single word, a phrase, or a special search term. For example, if the search string is `oracle secure enterprise search`, then there are four search terms in the search string: `oracle`, `secure`, `enterprise`, and `search`. If the search string is `oracle "secure enterprise search"`, then there are two search terms in the search string: `oracle` and `"secure enterprise search"`.

Search terms are case insensitive so that different cases are treated the same. For example, searching `oracle`, `Oracle`, or `ORACLE` returns the same search result.

Phrase

A phrase is a string enclosed in double-quotes (`"`). It can contain one or multiple words.

Operators

The following operators are defined in the query syntax:

- Plus [+]: The plus operator specifies that the search term immediately following it must be found in all matching documents. For example, searching for [Oracle +Applications] only finds documents that contain the word "Oracle" and "Applications". In a multiple word search, you can attach a [+] in front of every token including the very first token. You can also attach a [+] in front of a phrase enclosed in double-quotes ("). But there should be no space between the [+] and the search term.
- Minus [-]: The minus operator specifies that the search term immediately following it cannot appear in any document included in the search result. For example, searching for [Oracle -Applications] only finds documents that do not contain the word "Applications". In a multiple word search, you can attach a [-] in front of every token except the very first token. It can be a single word or a phrase, but there should be no space between the [-] and the token.
- Asterisk [*]: The asterisk specifies a wildcard search. For example, searching for the string [Ora*] finds documents that contain all words beginning with "Ora" such as "Oracle" and "Orator". You can also insert an asterisk in the middle of a word. For example, searching for the string [A*e] finds documents that contain words such as "Apple" or "Ape". Wildcards are not enabled by default; you can turn them on by modifying the `queryConfig` object in the Administration API. See the *Oracle Secure Enterprise Search Administration API Guide*.

Default Search: Implicit AND

By default, Oracle SES searches all of your search terms and relevant variations of the terms you entered. You do not need to include any operators (like `AND`) between terms. The order of the terms in the search affects the search results.

Word Separator

Use one or more spaces to separate each of the search terms.

Filter Conditions (Advanced Conditions)

Oracle SES query syntax only supports 'Site' and 'File type' filter conditions. It does not support any other filter conditions (advanced conditions) such as title, author, or last modified date. To restrict your search with other filter conditions, you can specify them in the Web Services API message `doOracleSearch`.

Special Search Terms

Oracle SES supports the use of several special search terms that allow the user or search administrator to access additional capabilities of the Oracle SES. Following is the list of special search terms:

Exclude Search Term You can exclude a word from your search by putting a minus sign [-] immediately in front of the term you want to exclude from the search results. Exclusion does not work with stop words.

Example: `oracle -search`

Negative search is not allowed unless there is another positive search term. For example:

`-search` is an invalid search.

`oracle -search` is a valid search.

Wildcard Search You can use an asterisk to match any number of characters in the middle or the end of a search term. You cannot place it at the beginning, such as searching for *earch.

Example: `Ora*`

Phrase Search Search for complete phrases by enclosing them in quotation marks. Words marked in this way appear together in all results exactly as entered.

Example: `"oracle secure enterprise search"`

Site Restricted Search If you know the specific Web site you want to search, but are not sure where the information is located within that site, then search only within the specific Web site. Enter the search followed by the string `site:` followed by the host name.

Example: `oracle site:example.com`

Notes:

- Domain restriction is not supported, because Oracle SES does not support left-truncated wildcard search (such as *.example.com)
- The exclusion operator (-) can be applied to this search term to remove a Web site from consideration in the search.
- Site restricted search term is implicit AND with other search terms.
- Only one site restriction is allowed. Also, you cannot have both site inclusion and exclusion in the search string. For example, the following search string is invalid:

```
oracle search  site:www.oracle.com  -site:otn.oracle.com
```

File Type Restricted Search The search prefix `filetype:` filters the results returned to include only documents with the extension specified immediately after. There can be no space between `filetype:` and the specified extension.

Example: `oracle filetype:doc`

Notes:

- The exclusion operator (-) can be applied to this search term to remove a file type from consideration in the search.
- Only one file type can be included. The following extensions are supported: doc, htm, html, xml, ps, pdf, txt, rtf, ppt, and xls. doc, html, pdf, txt, rtf, ppt, xls.
- File type restricted search term is implicit AND with other search terms.
- Only one file type restriction is allowed. Also, you cannot have both file type inclusion and exclusion in the search string. For example, the following search string is invalid:

```
oracle search  filetype:doc -filetype:pdf
```

Query Web Services Example: Basic Search

Following is a simple JSP application using Oracle Secure Enterprise Search proxy Java library to provide the basic search functionality:

```
<%@page contentType="text/html; charset=utf-8" %>
<%@page import = "java.util.Vector" %>
```

```
<%@page import = "java.net.URL" %>
<%@page import = "java.util.Properties" %>
<%@page import = "java.util.HashMap" %>
import javax.xml.rpc.Stub;
<%@page import = "oracle.search.query.webservice.client.*" %>

<%
  //
  // Get the search term entered by the user
  //
  String searchTerm = request.getParameter("searchTerm");
  if (searchTerm == null)  searchTerm = "";

  //
  // Define the result element array.
  //
  // ResultElement is a proxy Java class
  ResultElement[] resElemArray = null;
  int estimatedHitCount = 0;

  if (searchTerm != null && !"".equals(searchTerm))
  {
    //
    // Create the Oracle SES Web Services client stub
    //
    OracleSearchService stub = new OracleSearchService();

    //
    // Set the Oracle SES Web Services URL.
    // The URL is http://<host>:<port>/search/query/OracleSearch
    //
    stub.setSoapURL("http://staca19:7777/search/query/OracleSearch");

    //
    // Get the search result by calling OracleSearchService.doOracleSearch()
    //
    OracleSearchResult result = stub.doOracleSearch(searchTerm,
                            new Integer(1),
                            new Integer(10),
                            Boolean.TRUE,
                            Boolean.TRUE,
                            null,
                            "en",
                            "en",
                            Boolean.TRUE,
                            null,
                            null,
                            null);
    //
    // Get the estimated hit count by calling
    estimatedHitCount = result.getEstimatedHitCount().intValue();

    // Get the search results
    resElemArray = result.getResultElements();
  }
%>

<HTML>
<HEAD>
    <TITLE>Oracle SES Web Services Demo </TITLE>
```

```
</HEAD>
<BODY>
<FORM name="searchBox" method="post" action="./DemoWS.jsp">
 <INPUT id="inputMain" type="text" size="40" name="searchTerm"
value="<%=searchTerm%>">
 <INPUT type="hidden"  name="searchTerm" value="<%= searchTerm %>">
 <INPUT type="submit" name="action" value="Search">
</FORM>
<BR><BR><BR>

<%
  //
  // Render the search results
  //
  if (resElemArray == null || resElemArray.length == 0)
  {
%>
  <H3> There are no matches for the search term </H3>
<%
  }
  else
  {
%>
  <H3> There are about <%=estimatedHitCount%> matches </H3>
<%
    for (int i=0; i<resElemArray.length; i++)
    {
      String title = resElemArray[i].getTitle();
      if (title == null) title = "Untitled Document";
%>
  <P>
    <B><A HREF="<%=resElemArray[i].getUrl()%>"><%=title%></A> </B>
    <BR>
     <%=resElemArray[i].getSnippet()%>
    <BR>
  </P>
<%
    }
  }
%>
</BODY>
</HTML>
```

Default-Factor Element

The default-factor element assigns a weight to an attribute.

```
<default-factor>
   <name>title</name>
   <weight>VERY HIGH</weight>
</default-factor>
```

Default factor attribute names are case-insensitive.

When a `default-factor` does not appear in the ranking XML string, Oracle SES takes the default weight for this ranking factor, unless default factors are disabled by `enable-all-default-factor`.

Oracle SES supports the following values for weight element: `empty` (Oracle SES uses the default weight), `none` (this attributes is not used in the ranking query), `very high`, `high`, `medium`, `low`, and `very low`.

Table 11–3 lists the `default-factor` names and weights:

Table 11–3 Oracle SES Default Attributes and Weights

Attribute	Weight
`Title`	High
`Description`	Medium
`Reftext`	High
`Keywords`	Medium
`Subject`	Low
`Author`	Medium
`H1headline`	Low
`H2headline`	Very low
`Url`	Low
`Urldepth`	High
`Language Match`	High
`Linkscore`	High

Query Web Services Example: Customizing Relevancy

The following is the signature of the method for advanced search:

```
public OracleSearchResult doOracleAdvancedSearch (
        String query,
        Integer startIndex,
        Integer docsRequested,
        Boolean dupRemoved,
        Boolean dupMarked,
        DataGroup groups[],
        String queryLang,
        String docLang,
        Boolean returnCount,
        String filterConnector,
        Filter filters[],
        Integer[] fetchAttributes,
        String searchControls)  throws Exception
```

The `searchControls` parameter accepts a XML string, which include the `filter` and `ranking` elements.

```
<searchControls>
     <filter>  ... </filter>
     <ranking> ... </ranking>
</searchControls>
```

Filter Element

Filters for attribute search are passed in the `filter` element. All the various AND and OR conditions on the attributes are specified in the XML. For example:

```
<filter>
   <operator type="and">
   <operator type="or">
      <attributefilter name="xxx" type="string" operation="equals" value="ttt"/>
      <attributefilter name="yyy" type="number"
```

```
         operation="greaterthan" value="22"/>
...
   </operator>
...
      <attributefiler name="aaa" type="number" operation="equals" value="22"/>
...
   </operator>
</filter
```

If the parameter `searchControls` is null, then `filters` and `filterConnector` are used to create advanced search; otherwise, they are ignored.

Ranking Element

The ranking XML string is expressed as `ranking` element in `searchControls`. The following is an example of `ranking` element:

```
<ranking>
   <global-settings>
      <enable-all-default-factor>TRUE</enable-all-default-factor>
   </global-settings>
   <default-factor>
      <!--default ranking factor -- >
      ...
   </default-factor>
   <default-factor>
      <!--default ranking factor -- >
      ...
   </default-factor>
   <custom-factor>
      <!--default ranking factor -- >
      ...
   </custom-factor>
   <custom-factor>
      <!--default ranking factor -- >
      ...
   </custom-factor>
</ranking>
```

The following rules apply to the construction of ranking XML string:

- The whole ranking XML can be null, in which case default ranking is used.
- The ranking XML contains the elements `default-factor` and `custom-factor`. Both can be null or absent at the same time.
- When `default-factor` is null or absent and when `custom-factor` is not null, default ranking is used with the effect of `custom-factor`.
- When `custom-factor` is null or absent, it does not have any impact on the ranking.
- The ranking scheme applies only for the function `doOracleAdvancedSearch` call with none-empty query parameter passed.

Global-Settings Element

The `global-settings` element contains parameter settings across ranking factors. The `ranking` element has an attribute called `enable-all-default-factor`, which accepts two values: `true` or `false`. (When this attribute is absent, `true` is taken as the default value.)

When `enable-all-default-factor` is `true`, all default attributes are included in ranking queries, unless some default attributes are explicitly excluded in `default-factor` elements.

When `enable-all-default-factor` is `false`, all default attributes are excluded in ranking queries, unless some default attributes are explicitly included in `default-factor` elements.

Custom-Factor Element

The `custom-factor` element lets you add more attributes for ranking. Any indexed search attribute can be a custom ranking attribute.

> **Note:** Adding custom attributes for relevancy ranking can downgrade search performance.

The `custom-factor` element has four elements: `attribute-name`, `attribute-type`, `factor-type`, and `weight` (or `match` depending on the `factor-type`).

```
<custom-factor>
     <attribute-name>author manager</attribute-name>
     <attribute-type>STRING</attribute-type>
     <factor-type>QUERY_FACTOR</factor-type>
     <weight>LOW</weight>
</custom-factor>
```

or

```
<custom-factor>
     <attribute-name>document quality</attribute-name>
     <attribute-type>STRING</attribute-type>
     <factor-type>STATIC_FACTOR<factor-type>
     <match>
        <value>good</value>
        <weight>HIGH</weight>
     </match>
     <match>
        <value>fair</value>
        <weight>MEDIUM</weight>
     </match>
     <match>
        <value>bad</value>
        <weight>VERY LOW</weight>
     </match>
</custom-factor>
```

- The `attribute-name` values are literally matched against attribute name in Oracle SES. Any indexed search attribute name can be `attribute-name` value. The value of the `attribute-name` element is case-insensitive.
- The `attribute-type` element defines the type of the attribute. Only String attribute type is supported. Attribute-name and attribute-type in combination define a valid Oracle SES attribute.
- For `factor-type`, Oracle SES supports two types of ranking for custom ranking attributes.

  - `QUERY_FACTOR`: The attribute value is matched against query terms. A positive match boosts the document based on specified weight. `QUERY_FACTOR` is a query-based ranking factor; for example, title and reftext. The `weight` element should appear for this custom ranking factor. For example, with the query "Roger Federer", if a document has a custom attribute publisher with the value "Roger Federer", then it could be relevant.
  - `STATIC_FACTOR`: Attribute value is matched against fixed values specified in the custom ranking factor. (The `match` element should appear for this custom ranking factor.) `STATIC_FACTOR` is not a query-based ranking factor. The fixed values specify qualities of the documents, such as the link score and the sources of documents. For example, assume that documents have been classified based on quality. Well-written documents are classified as good, and poorly-written documents are classified as bad. A good document should be ranked higher than a bad document, even though they are both matched against a query. You can specify in the API that a document having a good quality should be boosted in relevancy by a specified weight.

- The `match` element specifies the match values and corresponding match weights when the `factor-type` is `STATIC_FACTOR`. The following XML string is a example of `match` element:

```
<match>
   <value>bad</value>
   <weight>VERY LOW</weight>
</match>
```

- The `value` element is used to match the corresponding attribute value of this ranking factor. Only alphanumeric letters are allowed in the attribute value. The match is case-insensitive.
- The `weight` element has the identical syntax with `weight` element for default ranking element.

Applying Ranking Factors

The XML ranking text can be applied in two places:

- As a part of the `searchControls` element, the ranking factors can be used as an advanced control for each query execution through the Web services method. This is called **per-query ranking control**.
- The ranking factors specified in the `relevanceRanking` object of the Administration API are applied to all queries. This is called **instance-wide ranking control**.

In federated search, instance-wide ranking controls only applies to one instance. You must configure each instance for ranking customization separately.

If a conflict arises, the per-query ranking control specified in Web services method overrides the settings specified in instance-wide ranking control. That can include the following cases:

- Per-query and instance-wide ranking specify the same factor, the factor set by per-query is taken by Oracle SES.
- Instance-wide ranking control sets a ranking factor, but per-query ranking control does not mention. Oracle SES takes the factor set by instance-wide ranking control.
- Per-query ranking control sets a ranking factor, which instance-wide ranking controls does not mention. Oracle SES takes the factor set by per-query ranking control.

- If instance-wide ranking control sets `enable-all-default-factor` as false and per-query ranking control sets `enable-all-default-factor` as true, then Oracle SES takes the default attributes set explicitly by instance-wide ranking control plus the attributes set by per-query ranking controls, with the latter overriding the former.

Client-Side Query Java Proxy Library

Oracle SES also provides client-side Java proxies for marshalling and parsing Web Services SOAP messages. Client applications can use the library to access Oracle SES Web Services.

The proxy library includes the following Java classes, which are mapped to the corresponding Web Services data types and messages:

- `oracle.search.query.webservice.client.Attribute`
- `oracle.search.query.webservice.client.AttributeLOVElement`
- `oracle.search.query.webservice.client.ClusterAttribute`
- `oracle.search.query.webservice.client.ClusterConfig`
- `oracle.search.query.webservice.client.ClusterTree`
- `oracle.search.query.webservice.client.CustomAttribute`
- `oracle.search.query.webservice.client.DataGroup`
- `oracle.search.query.webservice.client.Filter`
- `oracle.search.query.webservice.client.GroupAttribute`
- `oracle.search.query.webservice.client.GroupingResult`
- `oracle.search.query.webservice.client.Language`
- `oracle.search.query.webservice.client.Node`
- `oracle.search.query.webservice.client.OracleSearchResult`
- `oracle.search.query.webservice.client.OracleSearchService`
- `oracle.search.query.webservice.client.ResultElement`
- `oracle.search.query.webservice.client.SCElement`
- `oracle.search.query.webservice.client.SessionContextElement`
- `oracle.search.query.webservice.client.SortAttribute`
- `oracle.search.query.webservice.client.Status`
- `oracle.search.query.webservice.client.SuggestedLink`

To compile and run your client application using the Oracle SES client-side Java proxy library, you must include the following files in the Java CLASSPATH:

- *MW_HOME*`/wlserver_10.3.4/server/lib/weblogic.jar`
- *MW_HOME*`/wlserver_10.3.4/server/lib/wseeclient.jar`
- *MW_HOME*`/modules/org.apache.ant_1.7.1/lib/ant.jar`
- *ORACLE_HOME*`/search/lib/search_client.jar`
- *ORACLE_HOME*`/search/lib/plugins/cservices/jaxrpc.jar`

A

XML Connector Examples and Schemas

This appendix contains examples and schemas associated with the Oracle SES XML connector framework. This contains the following topics:

- Configuration File XML Schema Definition
- Control Feed Example
- Control Feed XML Schema Definition
- Data Feed Example
- Data Feed XML Schema Definition

See Also: "Overview of XML Connector Framework" on page 3-9

Configuration File XML Schema Definition

The following example shows the XSD for the configuration file.

```
<?xml version="1.0" encoding="windows-1252"?>
<xsd:schema
xmlns:xsd="http://www.w3.org/2001/XMLSchema"xmlns="http://xmlns.oracle.com/search/
rsscrawlerconfig"targetNamespace="http://xmlns.oracle.com/search/rsscrawlerconfig"
elementFormDefault="qualified">

<xsd:element name="rsscrawler">

   <xsd:annotation>
    <xsd:documentation>
     RSS crawler configuration paramters
    </xsd:documentation>
   </xsd:annotation>

<xsd:complexType>
 <xsd:sequence>

  <xsd:element name="sourceName" type="xsd:string" minOccurs="0"/>

  <xsd:element name="feedType" default="dataFeed">
   <xsd:simpleType>
    <xsd:restriction base="xsd:string">
     <xsd:enumeration value="controlFeed"/>
     <xsd:enumeration value="dataFeed"/>
     <xsd:enumeration value="directoryFeed"/>
    </xsd:restriction>
   </xsd:simpleType>
 </xsd:element>
```

```
<xsd:element name="feedLocation">
 <xsd:complexType>
  <xsd:simpleContent>
   <xsd:extension base="xsd:anyURI"/>
  </xsd:simpleContent>
 </xsd:complexType>
</xsd:element>

<xsd:element name="errorFileLocation" type="xsd:string" minOccurs="0">
 <xsd:annotation>
  <xsd:documentation>
Optional. This should be the absolute path of a location to which the status feeds
are uploaded. This location should be in the same computer from where data feeds
are fetched. If not specified, the status feeds are uploaded to the same location
as the data feeds. If HTTP is used to fetch the data feed, the value of this tag
should be the HTTP URL to which the status feed can be posted. If this tag is not
specified, the status feed is posted to the HTTP URL of the data feed.
  </xsd:documentation>
 </xsd:annotation>
</xsd:element>

<xsd:element name="securityType" default="noSecurity" maxOccurs="1" minOccurs="0">
 <xsd:simpleType>
  <xsd:restriction base="xsd:string">
   <xsd:enumeration value="identityBased"/>
   <xsd:enumeration value="attributeBased"/>
   <xsd:enumeration value="noSecurity"/>
  </xsd:restriction>
 </xsd:simpleType>
</xsd:element>

<xsd:element name="securityAttribute" minOccurs="0" maxOccurs="unbounded">
 <xsd:complexType>
  <xsd:simpleContent>
   <xsd:extension base="xsd:string">
    <xsd:attribute name="name" type="xsd:string" use="required"/>
    <xsd:attribute name="grant" type="xsd:boolean" default="true"/>
   </xsd:extension>
  </xsd:simpleContent>
 </xsd:complexType>
</xsd:element>

</xsd:sequence>
</xsd:complexType>

</xsd:element>
</xsd:schema>
```

Control Feed Example

The follow example shows a control feed used in an XML-connector based source.

```
<?xml version="1.0" encoding="windows-1252" ?>
<rss xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" version="2.0"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  xsi:schemaLocation="http://xmlns.oracle.com/orarss
C:\project_drive\SES Application Search\RSS Format
Schema\orarss.xsd">
 <channel>
```

```
   <title>Contacts</title>
   <link>http://my.company.com/rss</link>
   <description>The channel contains feed for contacts</description>
   <lastBuildDate>2006-04-03T12:20:20.00Z</lastBuildDate>
   <channelDesc xmlns="http://xmlns.oracle.com/orarss">
     <feedType>control</feedType></channelDesc>
   <item>
     <link>file://localhost/C:\project\rss_feeds\test.xml</link>
   </item>
   <item>
     <link>file://localhost/C:\project\rss_feeds\test2.xml</link>
   </item>
   <item operation="control">
     <link>http://my.host.com/contacts/control.xml</link></item><item>
   <link>file://localhost/C:\project\rss_feeds\test3.xml</link>
   </item>
 </channel>
</rss>
```

Control Feed XML Schema Definition

The following example shows the XSD for the control feed.

```
<?xml version="1.0" encoding="windows-1252"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
elementFormDefault="qualified">

<xsd:complexType name="descriptionType" abstract="true"/>

<xsd:complexType name="channelDescType">
 <xsd:complexContent>
  <xsd:extension base="descriptionType">
   <xsd:sequence>
    <xsd:element name="sourceName" type="xsd:string" minOccurs="0">
     <xsd:annotation>
      <xsd:documentation>
If the business object for this channel is missing, then the channel contains
information from multiple sources.
      </xsd:documentation>
     </xsd:annotation>
    </xsd:element>
    <xsd:element name="feedType" >
     <xsd:simpleType>
      <xsd:restriction base="xsd:string">
       <xsd:enumeration value="control"/>
      </xsd:restriction>
      </xsd:simpleType>
     </xsd:element>
     <xsd:element name="batchId" type="xsd:string" minOccurs="0"/>
     <xsd:element name="itemCount" type="xsd:positiveInteger" minOccurs="0"/>
    </xsd:sequence>
   </xsd:extension>
  </xsd:complexContent>
 </xsd:complexType>

<xsd:simpleType name="operationType">
 <xsd:restriction base="xsd:string">
  <xsd:enumeration value="control"/>
 </xsd:restriction>
</xsd:simpleType>
```

```
<xsd:complexType name="rssChannelType">
 <xsd:sequence>
  <xsd:element name="title" type="xsd:string"/>
  <xsd:element name="link" type="xsd:anyURI">

   <xsd:annotation>
    <xsd:documentation>same as display URL</xsd:documentation>
   </xsd:annotation>
  </xsd:element>
  <xsd:element name="description" type="xsd:string"/>
  <xsd:element name="lastBuildDate" type="xsd:dateTime">
   <xsd:annotation>
    <xsd:documentation>
    This is the publishing date for this channel
    </xsd:documentation>
   </xsd:annotation>
  </xsd:element>
  <xsd:any namespace="##other" minOccurs="0"/>
  <xsd:element name="channelDesc" type="channelDescType" />
  <xsd:element name="item" type="itemType" maxOccurs="unbounded"/>
 </xsd:sequence>
</xsd:complexType>

<xsd:complexType name="itemType">

 <xsd:sequence>
      <xsd:element name="title" type="xsd:string" minOccurs="0">
        <xsd:annotation>
          <xsd:documentation>
             This is the title for the item.
          </xsd:documentation>
        </xsd:annotation>
      </xsd:element>
      <xsd:element name="link" type="xsd:anyURI"/>
      <xsd:element name="description" type="xsd:string" minOccurs="0">
        <xsd:annotation>
          <xsd:documentation>
            The description is ignored as far as Oracle processing is concerned
          </xsd:documentation>
        </xsd:annotation>
      </xsd:element>
      <xsd:any namespace="##other" minOccurs="0"/>
    </xsd:sequence>
    <xsd:attribute name="operation" type="oss:operationType" />
</xsd:complexType>

  <xsd:element name="rss">
    <xsd:annotation>
      <xsd:documentation>RSS control file</xsd:documentation>
    </xsd:annotation>
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element name="channel" type="rssChannelType"/>
      </xsd:sequence>
      <xsd:attribute name="version" type="xsd:decimal" use="required" fixed="2.0/>
    </xsd:complexType>
  </xsd:element>

</xsd:schema>
```

Data Feed Example

When specifying a date-time value in an attribute for an RSS feed, use this format:

YYYY-MM-DDThh:mm:sstz[+|-*offset*]

Where:

YYYY is the year (4 digits).

MM is the month (2 digits).

DD is the day (2 digits).

T is a literal character used as a separator.

hh is the hour (2 digits in 24-hour time designation).

mm is the minute (2 digits).

ss is the second (2 digits).

tz is an RFC822 time zone,[1] or Z to indicate local time.

offset is the difference between the local time zone and Greenwich Mean Time (GMT) in the format *hh*:*mm*.

All of these examples are correct:

```
2002-06-03T16:06:05.00GMT
2002-06-03T16:06:05.00PDT
2002-06-03T16:06:05.00GMT-07:00
2002-06-03T16:06:05.00JST
2002-06-03T16:06:05.00GMT+09:00
2002-06-03T16:06:05.00Z
```

As shown by these date-time examples, you can specify numeric offsets in the date-time string. Thus, Greenwich Mean Time can be specified as either GMT or GMT+00:00 and Eastern Standard Time can be specified as either EST or GMT-05:00.

Repositories that dispatch data feeds must explicitly specify the content type to avoid the filtering overhead required by Oracle SES to detect the content type. See the `<contentLink>` and `<content>` elements in the example.

The following example shows a data feed containing three documents.

```
<?xml version="1.0" encoding="UTF-8"?>
<rss xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  version="2.0" xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  xsi:schemaLocation="http://xmlns.oracle.com/orarss
     C:\project_drive\SES Application Search\RSS Format Schema\orarss.xsd">
<channel>
<title>Contacts</title>
<link>http://my.company.com/rss</link>
<description>The channel contains feed for contacts</description>
<lastBuildDate>2006-04-03T12:20:20.00Z</lastBuildDate>
<channelDesc xmlns="http://xmlns.oracle.com/orarss" >
     <feedType>full</feedType>
</channelDesc>
<item>
     <link>http://my.company.com/contacts?id=paul</link>
     <itemDesc xmlns="http://xmlns.oracle.com/orarss" operation="insert">
```

[1] RFC822 supports UT and GMT for universal time, and these North American time zones: EST, EDT, CST, CDT, MST, MDT, PST, and PDT.

```
     <documentMetadata>
       <author>Administrator</author>
       <accessURL>http://foo.com</accessURL>
       <lastModifiedDate>2009-12-12T12:22:22.00Z</lastModifiedDate>
       <keywords>Content Contact</keywords>
       <summary>This is the summary of the document.</summary>
       <sourceHierarchy>
           <path>company</path>
           <path>department</path>
           <path>group</path>
       </sourceHierarchy>
       <docAttr name="organization">Reports</docAttr>
       <docAttr name="country">Germany</docAttr>
     </documentMetadata>
     <documentAcl>
       <securityAttr name="EMPLOYEE_ID">0R9NH</securityAttr>
     </documentAcl>
     <documentInfo>
       <status>STATUS_OK_FOR_INDEX</status>
     </documentInfo>
     <documentContent>
       <contentLink
contentType="text/html">http://my.company.com/reports.html</contentLink>
       <content contentType="text/plain">Paul Robinson, A240, Westland
Drive</content>
     </documentContent>
   </itemDesc>
</item>
<item>
   <link>http://my.company.com/contacts?id=tom</link>
   <itemDesc xmlns="http://xmlns.oracle.com/orarss" operation="delete"/>
</item>
<item>
   <link>http://my.company.com/contacts?id=robert</link>
   <itemDesc xmlns="http://xmlns.oracle.com/orarss" operation="insert">
   <documentMetadata>
     <author>Administrator</author>
     <accessURL>http://foo.com</accessURL>
     <lastModifiedDate>2009-12-12T12:22:22.00Z</lastModifiedDate>
     <keywords>Content Contact </keywords>
     <summary>This is the summary of the document</summary>
     <sourceHierarchy>
       <path>company</path>
       <path>department</path>
       <path>group</path>
     </sourceHierarchy>
     <docAttr name="organization">Sales</docAttr>
     <docAttr name="country">China</docAttr>
     </documentMetadata>
     <documentAcl>
       <securityAttr name="EMPLOYEE_ID">I23489</securityAttr>
     </documentAcl>
     <documentInfo>
       <status>STATUS_OK_FOR_INDEX</status>
     </documentInfo>
     <documentContent>
       <contentLink
contentType="text/html">http://my.company.com/sales.html</contentLink >
       <content contentType="text/plain">Robert Mogambo, C318, Lakeside
Avenue</content>
```

```
                    </documentContent>
                  </itemDesc>
                </item>
                </channel>
                </rss>
```

Data Feed XML Schema Definition

Following is the XSD for the data feed.

```
<?xml version="1.0" encoding="windows-1252"?>
<!-- edited with XMLSpy v2005 rel. 3 U (http://www.altova.com) by Oracle XDB (Oracle XDB) -->
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns:oss="http://xmlns.oracle.com/orarss"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
targetNamespace="http://xmlns.oracle.com/orarss" elementFormDefault="qualified">
  <xsd:complexType name="descriptionType" abstract="true"/>

<xsd:complexType name="channelDescType">
    <xsd:complexContent>
      <xsd:extension base="oss:descriptionType">
        <xsd:sequence>
          <xsd:element name="sourceName" type="xsd:string" minOccurs="0">
            <xsd:annotation>
              <xsd:documentation>
The business Object for which this channel corresponds to - if missing then the channel contains
information from multiple sources.
              </xsd:documentation>
            </xsd:annotation>
          </xsd:element>
          <xsd:element name="feedType" default="incremental" maxOccurs="0" minOccurs="0">
            <xsd:simpleType>
              <xsd:restriction base="xsd:string">
                <xsd:enumeration value="incremental"/>
                <xsd:enumeration value="full"/>
              </xsd:restriction>
            </xsd:simpleType>
           </xsd:element>
           <xsd:element name="batchId" type="xsd:string" minOccurs="0"/>
           <xsd:element name="itemCount" type="xsd:positiveInteger" minOccurs="0"/>
        </xsd:sequence>
      </xsd:extension>
    </xsd:complexContent>
  </xsd:complexType>

  <xsd:complexType name="itemType">
    <xsd:sequence>
      <xsd:element name="title" type="xsd:string"/>
      <xsd:element name="link" type="xsd:anyURI">
        <xsd:annotation>
          <xsd:documentation>
            Display URL of the item. This URL should be UTF-8 encoded.
          </xsd:documentation>
        </xsd:annotation>
      </xsd:element>
      <xsd:element name="description" type="xsd:string"/>
      <xsd:element name="itemDesc">
        <xsd:complexType>
          <xsd:complexContent>
            <xsd:extension base="oss:itemDescType">
              <xsd:attribute name="operation" type="oss:operationType" default="insert"/>
```

```
          </xsd:extension>
        </xsd:complexContent>
      </xsd:complexType>
    </xsd:element>
  </xsd:sequence>
 </xsd:complexType>

<xsd:complexType name="rssChannelType">
  <xsd:sequence>
    <xsd:element name="title" type="xsd:string"/>
    <xsd:element name="link" type="xsd:anyURI">
      <xsd:annotation>
        <xsd:documentation>
           display URL
        </xsd:documentation>
      </xsd:annotation>
    </xsd:element>
    <xsd:element name="description" type="xsd:string"/>
    <xsd:element name="lastBuildDate" type="xsd:dateTime">
      <xsd:annotation>
        <xsd:documentation>
          This is the publishing date for this channel
        </xsd:documentation>
      </xsd:annotation>
    </xsd:element>
    <xsd:any namespace="##other" minOccurs="0"/>
    <xsd:element name="channelDesc" type="oss:channelDescType" />
    <xsd:element name="item" maxOccurs="unbounded">
      <xsd:complexType>
        <xsd:complexContent>
          <xsd:extension base="oss:itemType"/>
        </xsd:complexContent>
      </xsd:complexType>
    </xsd:element>
  </xsd:sequence>
</xsd:complexType>

<xsd:complexType name="infoType">
  <xsd:sequence>
    <xsd:element name="status" type="oss:statusType"/>
  </xsd:sequence>
</xsd:complexType>

<xsd:complexType name="itemDescType">
  <xsd:complexContent>
    <xsd:extension base="oss:descriptionType">
      <xsd:sequence>
        <xsd:element name="documentMetadata" type="oss:metadataType" minOccurs="0"/>
        <xsd:element name="documentAcl" type="oss:securityType" minOccurs="0"/>
        <xsd:element name="documentInfo" type="oss:infoType" minOccurs="0"/>
        <xsd:element name="documentContent" type="oss:bodyType" minOccurs="0"/>
      </xsd:sequence>
    </xsd:extension>
  </xsd:complexContent>
</xsd:complexType>

<xsd:complexType name="sourceHierType">
  <xsd:sequence>
    <xsd:element name="path" type="xsd:string" maxOccurs="unbounded"/>
  </xsd:sequence>
```

```
</xsd:complexType>

<xsd:complexType name="metadataType">
  <xsd:sequence>
    <xsd:element name="author" type="xsd:string" minOccurs="0" maxOccurs="unbounded"/>
    <xsd:element name="accessURL" type="xsd:string" minOccurs="0"/>
    <xsd:element name="lastModifiedDate" type="xsd:dateTime" minOccurs="0"/>
    <xsd:element name="keywords" type="xsd:string" minOccurs="0"/>
    <xsd:element name="summary" type="xsd:string" minOccurs="0"/>
    <xsd:element name="language" type="xsd:string" minOccurs="0"/>
    <xsd:element name="sourceHierarchy" type="oss:sourceHierType" minOccurs="0"/>
    <xsd:element name="docAttr" minOccurs="0" maxOccurs="unbounded">
      <xsd:complexType>
        <xsd:simpleContent>
          <xsd:extension base="oss:docAttrType">
            <xsd:attribute name="name" type="xsd:string" use="required"/>
          </xsd:extension>
        </xsd:simpleContent>
      </xsd:complexType>
    </xsd:element>
  </xsd:sequence>
</xsd:complexType>

<xsd:complexType name="securityType">
  <xsd:choice>
    <xsd:element name="principal" minOccurs="0" maxOccurs="unbounded">
      <xsd:complexType>
        <xsd:simpleContent>
          <xsd:extension base="xsd:string">
            <xsd:attribute name="idType" type="oss:idAttrType" use="optional" default="user">
              <xsd:annotation>
                <xsd:documentation>
                  User or group.
                </xsd:documentation>
              </xsd:annotation>
            </xsd:attribute>
            <xsd:attribute name="format" type="xsd:string" use="required"/>
            <xsd:attribute name="grant" type="xsd:boolean" use="optional" default="true"/>
          </xsd:extension>
        </xsd:simpleContent>
      </xsd:complexType>
    </xsd:element>
    <xsd:element name="securityAttr" minOccurs="0" maxOccurs="unbounded">
      <xsd:complexType>
        <xsd:simpleContent>
          <xsd:extension base="xsd:string">
            <xsd:attribute name="name" type="xsd:string"/>
          </xsd:extension>
        </xsd:simpleContent>
      </xsd:complexType>
    </xsd:element>
  </xsd:choice>
  <xsd:attribute name="ownerGuid" type="xsd:string">
    <xsd:annotation>
      <xsd:documentation>
        OwnerGUID - useful if the principal includes OWNER
      </xsd:documentation>
    </xsd:annotation>
  </xsd:attribute>
</xsd:complexType>
```

```
<xsd:simpleType name="statusType">
  <xsd:restriction base="xsd:string">
    <xsd:enumeration value="STATUS_OK_FOR_INDEX"/>
    <xsd:enumeration value="STATUS_BAD_REQUEST"/>
    <xsd:enumeration value="STATUS_AUTH_REQUIRED"/>
    <xsd:enumeration value="STATUS_ACCESS_FORBIDDEN"/>
    <xsd:enumeration value="STATUS_NOTFOUND"/>
    <xsd:enumeration value="STATUS_PROXY_REQUIRED"/>
    <xsd:enumeration value="STATUS_REQUEST_TIMEOUT"/>
    <xsd:enumeration value="STATUS_SERVER_ERROR"/>
    <xsd:enumeration value="STATUS_BAD_GATEWAY"/>
    <xsd:enumeration value="STATUS_FETCH_ERROR"/>
    <xsd:enumeration value="STATUS_READ_TIMEOUT"/>
    <xsd:enumeration value="STATUS_FILTER_ERROR"/>
    <xsd:enumeration value="STATUS_OUT_OF_MEMORY"/>
    <xsd:enumeration value="STATUS_IO_EXCEPTION"/>
    <xsd:enumeration value="STATUS_CONNECTION_REFUSED"/>
    <xsd:enumeration value="STATUS_DUPLICATE_DOC"/>
    <xsd:enumeration value="STATUS_EMPTY_DOC"/>
    <xsd:enumeration value="STATUS_LOGIN_FAILED"/>
    <xsd:enumeration value="STATUS_OK_BUT_NO_INDEX"/>
    <xsd:enumeration value="STATUS_OK_CRAWLED"/>
    <xsd:enumeration value="STATUS_CANNOT_READ"/>
    <xsd:enumeration value="STATUS_DOC_SIZE_TOO_BIG"/>
  </xsd:restriction>
</xsd:simpleType>

<xsd:simpleType name="operationType">
  <xsd:restriction base="xsd:string">
    <xsd:enumeration value="insert"/>
    <xsd:enumeration value="replace"/>
    <xsd:enumeration value="delete"/>
  </xsd:restriction>
</xsd:simpleType>

<xsd:simpleType name="idAttrType">
  <xsd:restriction base="xsd:string">
    <xsd:enumeration value="user"/>
    <xsd:enumeration value="group"/>
    <xsd:enumeration value="owner"/>
  </xsd:restriction>
</xsd:simpleType>

<xsd:complexType name="bodyType">
  <xsd:choice minOccurs="0" maxOccurs="unbounded">
    <xsd:element name="contentLink" minOccurs="0" maxOccurs="unbounded">
      <xsd:complexType>
        <xsd:simpleContent>
          <xsd:extension base="xsd:anyURI">
            <xsd:attribute name="contentType"/>
          </xsd:extension>
        </xsd:simpleContent>
      </xsd:complexType>
    </xsd:element>
    <xsd:element name="content" minOccurs="0">
      <xsd:complexType>
        <xsd:simpleContent>
          <xsd:extension base="xsd:string">
            <xsd:attribute name="contentType" type="xsd:string"/>
```

```
            </xsd:extension>
          </xsd:simpleContent>
        </xsd:complexType>
      </xsd:element>
      <xsd:element name="xhtmlContent" minOccurs="0" maxOccurs="1">
        <xsd:complexType>
          <xsd:complexContent>
            <xsd:extension base="xsd:anyType">
              <xsd:attribute name="lang" type="xsd:string"/>
              <xsd:anyAttribute namespace="##other"/>
            </xsd:extension>
          </xsd:complexContent>
        </xsd:complexType>
      </xsd:element>
    </xsd:choice>
  </xsd:complexType>

  <xsd:simpleType name="docAttrType">
    <xsd:union memberTypes="xsd:dateTime xsd:decimal xsd:string"/>
  </xsd:simpleType>

  <xsd:element name="rss">
    <xsd:annotation>
      <xsd:documentation>RSS data file</xsd:documentation>
    </xsd:annotation>
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element name="channel" type="oss:rssChannelType"/>
      </xsd:sequence>
      <xsd:attribute name="version" type="xsd:decimal" use="required" fixed="2.0/>
    </xsd:complexType>
  </xsd:element>

</xsd:schema>
```

B

URL Crawler Status Codes

The crawler uses a set of codes to indicate the result of the crawled URL. Besides the standard HTTP status code, it uses its own code for non-HTTP related situations.

Only URLs with status 200 are indexed. If the record exists in `EQ$URL` but the status is something other than 200, then the crawler encountered an error trying to fetch the document. A status of less than 600 maps directly to the HTTP status code.

> **See Also:** "Status Code Definitions" in *Hypertext Transfer Protocol -- HTTP/1.1* at
>
> `http://www.w3.org/Protocols/rfc2616/rfc2616-sec10.html`

The following table lists the URL status codes, document container codes used by the crawler plug-in, and EQG codes.

Code	Description	Document Container Code	EQG Codes
0	A URL that has been enqueued but not yet processed		N/A
200	URL OK	`STATUS_OK_FOR_INDEX`	N/A
400	Bad request	`STATUS_BAD_REQUEST`	30009
401	Authorization required	`STATUS_AUTH_REQUIRED`	30007
402	Payment required		30011
403	Access forbidden	`STATUS_ACCESS_FORBIDDEN`	30010
404	Not found	`STATUS_NOTFOUND`	30008
405	Method not allowed		30012
406	Not acceptable		30013
407	Proxy authentication required	`STATUS_PROXY_REQUIRED`	30014
408	Request timeout	`STATUS_REQUEST_TIMEOUT`	30015
409	Conflict		30016
410	Gone		30017
414	Request URI too large		30066
500	Internal server error	`STATUS_SERVER_ERROR`	10018
501	Not implemented		10019

Code	Description	Document Container Code	EQG Codes
502	Bad gateway	STATUS_BAD_GATEWAY	10020
503	Service unavailable	STATUS_FETCH_ERROR	10021
504	Gateway timeout		10022
505	HTTP version not supported		10023
902	Timeout reading document	STATUS_READ_TIMEOUT	30057
903	Filtering failed	STATUS_FILTER_ERROR	30065
904	Out of memory error	STATUS_OUT_OF_MEMORY	30003
905	IOEXCEPTION in processing URL	STATUS_IO_EXCEPTION	30002
906	Connection refused	STATUS_CONNECTION_ REFUSED	30025
907	Socket bind exception		30079
908	Filter not available		30081
909	Duplicate document detected		30082
910	Duplicate document ignored	STATUS_DUPLICATE_DOC	30083
911	Empty document	STATUS_EMPTY_DOC	30106
951	URL not indexed (this can happen if robots.txt specifies that a certain document should not be indexed)	STATUS_OK_BUT_NO_INDEX	N/A
952	URL crawled	STATUS_OK_CRAWLED	N/A
953	Metatag redirection		N/A
954	HTTP redirection		30000
955	Black list URL		N/A
956	URL is not unique		31017
957	Sentry URL (URL as a place holder)		N/A
958	Document read error	STATUS_CANNOT_READ	30173
959	Form login failed	STATUS_LOGIN_FAILED	30183
960	Document size too big, ignored	STATUS_DOC_SIZE_TOO_BIG	30209
962	Document was excluded based on mime type	STATUS_DOC_MIME_TYPE_ EXCLUDED	30041
964	Document was excluded based on boundary rules	STATUS_DOC_BOUNDARY_ RULE_EXCLUDED	30258
1001	Datatype is not TEXT/HTML		30001
1002	Broken network data stream		30004
1003	HTTP redirect location does not exist		30005
1004	Bad relative URL		30006
1005	HTTP error		30024
1006	Error parsing HTTP header		30058
1007	Invalid URL table column name		30067
1009	Binary document reported as text document		30126

Code	Description	Document Container Code	EQG Codes
1010	Invalid display URL		30112
1011	Invalid XML from OracleAS Portal	PORTAL_XMLURL_FAIL	31011
1020-1024	URL is not reachable. The status starts at 1020, and it increases by one with each try. After five tries (if it reaches 1025), the URL is deleted.		N/A
1026-1029	URL cannot be found. The status turns from 404 to 1026 when a URL cannot be found on re-crawl, and it increases by one with each try. After five tries (if it reaches 1030), the URL is deleted.		N/A
1111	URL remained in the queue even after a successful crawl. This indicates that the crawler had a problem processing this document. You could investigate the URL by crawling it in a separate source to check for errors in the crawler log.		N/A

C

Third Party Licenses

This appendix includes the third party licenses for all the third party products included with Oracle Secure Enterprise Search. This appendix includes the following topics:

- Apache Software
- Eclipse Software
- Egothor Software
- Javascript Bubbling Library
- Plug-in Software
- Snowball Software
- Visigoth Software
- Yahoo! Inc.

Apache Software

This program contains code from the Apache Software Foundation ("Apache"). Under the terms of the Apache license, Oracle is required to provide the following notices. Note, however, that the Oracle program license that accompanied this product determines your right to use the Oracle program, including the Apache software, and the terms contained in the following notices do not change those rights. Notwithstanding anything to the contrary in the Oracle program license, the Apache software is provided by Oracle "AS IS" and without any warranty or support of any kind from Oracle or Apache.

```
                 Apache License
           Version 2.0, January 2004
        http://www.apache.org/licenses/

TERMS AND CONDITIONS FOR USE, REPRODUCTION, AND DISTRIBUTION
1. Definitions.

      "License" shall mean the terms and conditions for use, reproduction,
      and distribution as defined by Sections 1 through 9 of this document.

      "Licensor" shall mean the copyright owner or entity authorized by
      the copyright owner that is granting the License.

      "Legal Entity" shall mean the union of the acting entity and all
      other entities that control, are controlled by, or are under common
      control with that entity. For the purposes of this definition,
```

"control" means (i) the power, direct or indirect, to cause the direction or management of such entity, whether by contract or otherwise, or (ii) ownership of fifty percent (50%) or more of the outstanding shares, or (iii) beneficial ownership of such entity.

"You" (or "Your") shall mean an individual or Legal Entity exercising permissions granted by this License.

"Source" form shall mean the preferred form for making modifications, including but not limited to software source code, documentation source, and configuration files.

"Object" form shall mean any form resulting from mechanical transformation or translation of a Source form, including but not limited to compiled object code, generated documentation, and conversions to other media types.

"Work" shall mean the work of authorship, whether in Source or Object form, made available under the License, as indicated by a copyright notice that is included in or attached to the work (an example is provided in the Appendix below).

"Derivative Works" shall mean any work, whether in Source or Object form, that is based on (or derived from) the Work and for which the editorial revisions, annotations, elaborations, or other modifications represent, as a whole, an original work of authorship. For the purposes of this License, Derivative Works shall not include works that remain separable from, or merely link (or bind by name) to the interfaces of, the Work and Derivative Works thereof.

"Contribution" shall mean any work of authorship, including the original version of the Work and any modifications or additions to that Work or Derivative Works thereof, that is intentionally submitted to Licensor for inclusion in the Work by the copyright owner or by an individual or Legal Entity authorized to submit on behalf of the copyright owner. For the purposes of this definition, "submitted" means any form of electronic, verbal, or written communication sent to the Licensor or its representatives, including but not limited to communication on electronic mailing lists, source code control systems, and issue tracking systems that are managed by, or on behalf of, the Licensor for the purpose of discussing and improving the Work, but excluding communication that is conspicuously marked or otherwise designated in writing by the copyright owner as "Not a Contribution."

"Contributor" shall mean Licensor and any individual or Legal Entity on behalf of whom a Contribution has been received by Licensor and subsequently incorporated within the Work.

2. Grant of Copyright License. Subject to the terms and conditions of this License, each Contributor hereby grants to You a perpetual, worldwide, non-exclusive, no-charge, royalty-free, irrevocable copyright license to reproduce, prepare Derivative Works of, publicly display, publicly perform, sublicense, and distribute the Work and such Derivative Works in Source or Object form.

3. Grant of Patent License. Subject to the terms and conditions of this License, each Contributor hereby grants to You a perpetual, worldwide, non-exclusive, no-charge, royalty-free, irrevocable (except as stated in this section) patent license to make, have made, use, offer to sell, sell, import, and otherwise transfer the Work,

where such license applies only to those patent claims licensable by such Contributor that are necessarily infringed by their Contribution(s) alone or by combination of their Contribution(s) with the Work to which such Contribution(s) was submitted. If You institute patent litigation against any entity (including a cross-claim or counterclaim in a lawsuit) alleging that the Work or a Contribution incorporated within the Work constitutes direct or contributory patent infringement, then any patent licenses granted to You under this License for that Work shall terminate as of the date such litigation is filed.

4. Redistribution. You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

   (a) You must give any other recipients of the Work or Derivative Works a copy of this License; and

   (b) You must cause any modified files to carry prominent notices stating that You changed the files; and

   (c) You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and

   (d) If the Work includes a "NOTICE" text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.

   You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

5. Submission of Contributions. Unless You explicitly state otherwise, any Contribution intentionally submitted for inclusion in the Work by You to the Licensor shall be under the terms and conditions of this License, without any additional terms or conditions. Notwithstanding the above, nothing herein shall supersede or modify the terms of any separate license agreement you may have executed with Licensor regarding such Contributions.

6. Trademarks. This License does not grant permission to use the trade
   names, trademarks, service marks, or product names of the Licensor,
   except as required for reasonable and customary use in describing the
   origin of the Work and reproducing the content of the NOTICE file.

7. Disclaimer of Warranty. Unless required by applicable law or
   agreed to in writing, Licensor provides the Work (and each
   Contributor provides its Contributions) on an "AS IS" BASIS,
   WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or
   implied, including, without limitation, any warranties or conditions
   of TITLE, NON-INFRINGEMENT, MERCHANTABILITY, or FITNESS FOR A
   PARTICULAR PURPOSE. You are solely responsible for determining the
   appropriateness of using or redistributing the Work and assume any
   risks associated with Your exercise of permissions under this License.

8. Limitation of Liability. In no event and under no legal theory,
   whether in tort (including negligence), contract, or otherwise,
   unless required by applicable law (such as deliberate and grossly
   negligent acts) or agreed to in writing, shall any Contributor be
   liable to You for damages, including any direct, indirect, special,
   incidental, or consequential damages of any character arising as a
   result of this License or out of the use or inability to use the
   Work (including but not limited to damages for loss of goodwill,
   work stoppage, computer failure or malfunction, or any and all
   other commercial damages or losses), even if such Contributor
   has been advised of the possibility of such damages.

9. Accepting Warranty or Additional Liability. While redistributing
   the Work or Derivative Works thereof, You may choose to offer,
   and charge a fee for, acceptance of support, warranty, indemnity,
   or other liability obligations and/or rights consistent with this
   License. However, in accepting such obligations, You may act only
   on Your own behalf and on Your sole responsibility, not on behalf
   of any other Contributor, and only if You agree to indemnify,
   defend, and hold each Contributor harmless for any liability
   incurred by, or claims asserted against, such Contributor by reason
   of your accepting any such warranty or additional liability.

END OF TERMS AND CONDITION

Eclipse Software

Eclipse Public License v 1.0, 1.0.2, 1.2

THE ACCOMPANYING PROGRAM IS PROVIDED UNDER THE TERMS OF THIS ECLIPSE PUBLIC LICENSE ("AGREEMENT"). ANY USE, REPRODUCTION OR DISTRIBUTION OF THE PROGRAM CONSTITUTES RECIPIENT'S ACCEPTANCE OF THIS AGREEMENT.

1. DEFINITIONS

"Contribution" means:

a) in the case of the initial Contributor, the initial code and documentation distributed under this Agreement, and

b) in the case of each subsequent Contributor:

i) changes to the Program, and

ii) additions to the Program;

where such changes and/or additions to the Program originate from and are distributed by that particular Contributor. A Contribution 'originates' from a Contributor if it was added to the Program by such Contributor itself or anyone acting on such Contributor's behalf. Contributions do not include additions to the Program which: (i) are separate modules of software distributed in conjunction with the Program under their own license agreement, and (ii) are not derivative works of the Program.

"Contributor" means any person or entity that distributes the Program.

"Licensed Patents " mean patent claims licensable by a Contributor which are necessarily infringed by the use or sale of its Contribution alone or when combined with the Program.

"Program" means the Contributions distributed in accordance with this Agreement.

"Recipient" means anyone who receives the Program under this Agreement, including all Contributors.

2. GRANT OF RIGHTS

a) Subject to the terms of this Agreement, each Contributor hereby grants Recipient a non-exclusive, worldwide, royalty-free copyright license to reproduce, prepare derivative works of, publicly display, publicly perform, distribute and sublicense the Contribution of such Contributor, if any, and such derivative works, in source code and object code form.

b) Subject to the terms of this Agreement, each Contributor hereby grants Recipient a non-exclusive, worldwide, royalty-free patent license under Licensed Patents to make, use, sell, offer to sell, import and otherwise transfer the Contribution of such Contributor, if any, in source code and object code form. This patent license shall apply to the combination of the Contribution and the Program if, at the time the Contribution is added by the Contributor, such addition of the Contribution causes such combination to be covered by the Licensed Patents. The patent license shall not apply to any other combinations which include the Contribution. No hardware per se is licensed hereunder.

c) Recipient understands that although each Contributor grants the licenses to its Contributions set forth herein, no assurances are provided by any Contributor that the Program does not infringe the patent or other intellectual property rights of any other entity. Each Contributor disclaims any liability to Recipient for claims brought by any other entity based on infringement of intellectual property rights or otherwise. As a condition to exercising the rights and licenses granted hereunder, each Recipient hereby assumes sole responsibility to secure any other intellectual property rights needed, if any. For example, if a third party patent license is required to allow Recipient to distribute the Program, it is Recipient's responsibility to acquire that license before distributing the Program.

d) Each Contributor represents that to its knowledge it has sufficient copyright rights in its Contribution, if any, to grant the copyright license set forth in this Agreement.

3. REQUIREMENTS

A Contributor may choose to distribute the Program in object code form under its own license agreement, provided that:

a) it complies with the terms and conditions of this Agreement; and

b) its license agreement:

i) effectively disclaims on behalf of all Contributors all warranties and conditions, express and implied, including warranties or conditions of title and non-infringement,

and implied warranties or conditions of merchantability and fitness for a particular purpose;

ii) effectively excludes on behalf of all Contributors all liability for damages, including direct, indirect, special, incidental and consequential damages, such as lost profits;

iii) states that any provisions which differ from this Agreement are offered by that Contributor alone and not by any other party; and

iv) states that source code for the Program is available from such Contributor, and informs licensees how to obtain it in a reasonable manner on or through a medium customarily used for software exchange.

When the Program is made available in source code form:

a) it must be made available under this Agreement; and

b) a copy of this Agreement must be included with each copy of the Program.

Contributors may not remove or alter any copyright notices contained within the Program.

Each Contributor must identify itself as the originator of its Contribution, if any, in a manner that reasonably allows subsequent Recipients to identify the originator of the Contribution.

4. COMMERCIAL DISTRIBUTION

Commercial distributors of software may accept certain responsibilities with respect to end users, business partners and the like. While this license is intended to facilitate the commercial use of the Program, the Contributor who includes the Program in a commercial product offering should do so in a manner which does not create potential liability for other Contributors. Therefore, if a Contributor includes the Program in a commercial product offering, such Contributor ("Commercial Contributor") hereby agrees to defend and indemnify every other Contributor ("Indemnified Contributor") against any losses, damages and costs (collectively "Losses") arising from claims, lawsuits and other legal actions brought by a third party against the Indemnified Contributor to the extent caused by the acts or omissions of such Commercial Contributor in connection with its distribution of the Program in a commercial product offering. The obligations in this section do not apply to any claims or Losses relating to any actual or alleged intellectual property infringement. In order to qualify, an Indemnified Contributor must: a) promptly notify the Commercial Contributor in writing of such claim, and b) allow the Commercial Contributor to control, and cooperate with the Commercial Contributor in, the defense and any related settlement negotiations. The Indemnified Contributor may participate in any such claim at its own expense.

For example, a Contributor might include the Program in a commercial product offering, Product X. That Contributor is then a Commercial Contributor. If that Commercial Contributor then makes performance claims, or offers warranties related to Product X, those performance claims and warranties are such Commercial Contributor's responsibility alone. Under this section, the Commercial Contributor would have to defend claims against the other Contributors related to those performance claims and warranties, and if a court requires any other Contributor to pay any damages as a result, the Commercial Contributor must pay those damages.

5. NO WARRANTY

EXCEPT AS EXPRESSLY SET FORTH IN THIS AGREEMENT, THE PROGRAM IS PROVIDED ON AN "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, EITHER EXPRESS OR IMPLIED INCLUDING, WITHOUT LIMITATION,

ANY WARRANTIES OR CONDITIONS OF TITLE, NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Each Recipient is solely responsible for determining the appropriateness of using and distributing the Program and assumes all risks associated with its exercise of rights under this Agreement, including but not limited to the risks and costs of program errors, compliance with applicable laws, damage to or loss of data, programs or equipment, and unavailability or interruption of operations.

6. DISCLAIMER OF LIABILITY

EXCEPT AS EXPRESSLY SET FORTH IN THIS AGREEMENT, NEITHER RECIPIENT NOR ANY CONTRIBUTORS SHALL HAVE ANY LIABILITY FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING WITHOUT LIMITATION LOST PROFITS), HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OR DISTRIBUTION OF THE PROGRAM OR THE EXERCISE OF ANY RIGHTS GRANTED HEREUNDER, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

7. GENERAL

If any provision of this Agreement is invalid or unenforceable under applicable law, it shall not affect the validity or enforceability of the remainder of the terms of this Agreement, and without further action by the parties hereto, such provision shall be reformed to the minimum extent necessary to make such provision valid and enforceable.

If Recipient institutes patent litigation against any entity (including a cross-claim or counterclaim in a lawsuit) alleging that the Program itself (excluding combinations of the Program with other software or hardware) infringes such Recipient's patent(s), then such Recipient's rights granted under Section 2(b) shall terminate as of the date such litigation is filed.

All Recipient's rights under this Agreement shall terminate if it fails to comply with any of the material terms or conditions of this Agreement and does not cure such failure in a reasonable period of time after becoming aware of such noncompliance. If all Recipient's rights under this Agreement terminate, Recipient agrees to cease use and distribution of the Program as soon as reasonably practicable. However, Recipient's obligations under this Agreement and any licenses granted by Recipient relating to the Program shall continue and survive.

Everyone is permitted to copy and distribute copies of this Agreement, but in order to avoid inconsistency the Agreement is copyrighted and may only be modified in the following manner. The Agreement Steward reserves the right to publish new versions (including revisions) of this Agreement from time to time. No one other than the Agreement Steward has the right to modify this Agreement. The Eclipse Foundation is the initial Agreement Steward. The Eclipse Foundation may assign the responsibility to serve as the Agreement Steward to a suitable separate entity. Each new version of the Agreement will be given a distinguishing version number. The Program (including Contributions) may always be distributed subject to the version of the Agreement under which it was received. In addition, after a new version of the Agreement is published, Contributor may elect to distribute the Program (including its Contributions) under the new version. Except as expressly stated in Sections 2(a) and 2(b) above, Recipient receives no rights or licenses to the intellectual property of any Contributor under this Agreement, whether expressly, by implication, estoppel or otherwise. All rights in the Program not expressly granted under this Agreement are reserved.

This Agreement is governed by the laws of the State of New York and the intellectual property laws of the United States of America. No party to this Agreement will bring a legal action under this Agreement more than one year after the cause of action arose. Each party waives its rights to a jury trial in any resulting litigation.

Egothor Software

This program includes software developed by the Egothor Project. Under the terms of the Egothor license, Oracle is required to provide the following notices. Note, however, that the Oracle program license that accompanied this product determines your right to use the Oracle program, including the Egothor software, and the terms contained in the following notices do not change those rights.

http://egothor.sf.net/

Egothor Software License version 1.00. (C) 1997-2004 Leo Galambos,

(C) 2002-2004 "Egothor developers" on behalf of the Egothor Project

This software is copyrighted by the "Egothor developers". If this license applies to a single file or document, the "Egothor developers" are the people or entities mentioned as copyright holders in that file or document. If this license applies to the Egothor project as a whole, the copyright holders are the people or entities mentioned in the file CREDITS. This file can be found in the same location as this license in the distribution.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

Redistributions of source code must retain the above copyright notice, the list of contributors, this list of conditions, and the following disclaimer.

Redistributions in binary form must reproduce the above copyright notice, the list of contributors, this list of conditions, and the disclaimer that follows these conditions in the documentation and/or other materials provided with the distribution.

The name "Egothor" must not be used to endorse or promote products derived from this software without prior written permission. For written permission, please contact <Leo.G@seznam.cz <mailto:Leo.G@seznam.cz>>.

Products derived from this software may not be called "Egothor", nor may "Egothor" appear in their name, without prior written permission from <Leo.G@seznam.cz <mailto:Leo.G@seznam.cz>>.

In addition, we request that you include in the end-user documentation provided with the redistribution and/or in the software itself an acknowledgement equivalent to the following:

"This product includes software developed by the Egothor Project.

http://egothor.sf.net/"

THIS SOFTWARE IS PROVIDED ``AS IS'' AND ANY EXPRESSED OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE EGOTHOR PROJECT OR ITS CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT

LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

This software consists of voluntary contributions made by many individuals on behalf of the Egothor Project and was originally created by Leo Galambos <Leo.G@seznam.cz <mailto:Leo.G@seznam.cz>>.

Javascript Bubbling Library

Javascript Bubbling Library http://www.bubbling-library.com

Copyright (c) 2007, Caridy Patiño. All rights reserved.

Redistribution and use of this software in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

To get started using Bubbling Library, simply include the two source files into the head of your document:

```
<!-- YUI Core -->
<script src="/PATH/TO/utilities.js" type="text/javascript"></script>
<!-- Bubbling Library Core -->
<script src="/PATH/TO/bubbling.js" type="text/javascript"></script>
```

The documentation can be found here:
http://www.bubbling-library.com/eng/api/docs/

Plug-in Software

Oracle SES ships several *plug-ins* to enterprise sources. (Plug-ins allow Oracle SES to crawl and index content in proprietary systems). For some plug-ins to work, additional software may need to be installed and licensed from the respective vendor; for example, EMC Documentum requires Documentum Foundation Classes (DFC), a Java library, to be installed on the computer running Oracle SES.

The following enterprise sources require additional software to be installed on the computer running Oracle SES:

- EMC Documentum Content Server
- Microsoft Exchange uses the Jakarta Slide libraries for WebDAV
- Microsoft NTFS may require Microsoft .NET 2.0

See Also:

- Chapter 6, "Configuring Access to Content Management Sources"
- Chapter 7, "Configuring Access to Collaboration Sources"

Snowball Software

This program contains software developed by Snowball. Under the terms of the Snowball license, Oracle is required to provide the following notices. Note, however, that the Oracle program license that accompanied this product determines your right to use the Oracle program, including the Snowball software, and the terms contained in the following notices do not change those rights.

Copyright (c) 2001, Dr. Martin Porter, and (for the Java developments) Copyright (c) 2002, Richard Boulton.

All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution. Neither the name Snowball nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Visigoth Software

This product includes software developed by the Visigoth Software Society (`http://www.visigoths.org/`).

Copyright (c) 2003 The Visigoth Software Society. All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
2. The end-user documentation included with the redistribution, if any, must include the following reacknowledgment: "This product includes software developed by the Visigoth Software Society (http://www.visigoths.org/)." Alternately, this reacknowledgment may appear in the software itself, if and wherever such third-party acknowledgments normally appear.
3. Neither the name "FreeMarker", "Visigoth", nor any of the names of the project contributors may be used to endorse or promote products derived from this software without prior written permission. For written permission, please contact visigoths@visigoths.org.
4. Products derived from this software may not be called "FreeMarker" or "Visigoth" nor may "FreeMarker" or "Visigoth" appear in their names without prior written permission of the Visigoth Software Society.

THIS SOFTWARE IS PROVIDED "AS IS" AND ANY EXPRESSED OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE VISIGOTH SOFTWARE SOCIETY OR ITS CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE,EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

This software consists of voluntary contributions made by many individuals on behalf of the Visigoth Software Society. For more information on the Visigoth Software Society, please see http://www.visigoths.org/

Yahoo! Inc.

This program contains software developed by Yahoo! Inc. Under the terms of the Yahoo! Inc. license, Oracle is required to provide the following notices. Note, however, that the Oracle program license that accompanied this product determines your right to use the Oracle program, including the Yahoo! Inc. software, and the terms contained in the following notices do not change those rights.

Software License Agreement (BSD License)

Copyright (c) 2006, Yahoo! Inc.

All rights reserved.

Redistribution and use of this software in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

- Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
- Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

- Neither the name of Yahoo! Inc. nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission of Yahoo! Inc.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE. TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

D

Error Messages

The crawler uses a set of messages to log the crawling activities.

The following table lists the most common crawler error messages.

Message ID	Message	Comment	Action
30025	{0}: Connection refused	The Web site refuses the URL access request.	Check the network setup environment of the computer running the crawler.
30027	Not allowed URL: {0}	A URL link violates boundary rules and is discarded.	Confirm that the URL indeed can be ignored.
30030	Malformed URL: {0}	The URL is not properly formed.	Verify the URL.
30031	Excluded by ROBOTS.TXT: {0}	The robots.txt rule from the Web site of the URL does not allow the URL to be crawled.	Configure the crawler to ignore robots rule only when you are managing the target Web site. Use the Home - Sources - Crawling Parameters page.
30040	Ignore URL: {0}	Redirection to this URL is not allowed by boundary rule.	Confirm that the URL indeed should be ignored.
30041	{0}: excluded by MIME type inclusion rule, URL is {1}	The content type of the URL is not in MIME type inclusion list.	Check if the specified content type should be included.
30054	Excessively long URL: {0}	The URL string is too long, and the URL is ignored.	N/A
30057	{0}: timeout reading document	The target Web site is too slow sending page content.	Increase the crawler timeout threshold from the crawler configuration page. The default is 30 seconds.
30083	{0}: Duplicate document ignored	An identical document has been seen before in the same crawl session. This could be an indication of URL looping; that is, a generation of different URLs pointing back to the same page.	Check if the URL is generated correctly. If necessary, disable indexing dynamic URLs. Use the Home - Sources - Crawling Parameters page.

Message ID	Message	Comment	Action
30126	Binary document reported as text document: "{0}"	A binary file has been sent by the Web site as a text document. In most cases, the URL in question is not a binary format text document, like pdf.	Correct the Web site content type setting for the URL, if possible.
30188	Login form not specified for "{0}"	Unable to perform HTML form login, because the name of the form is not set. In general, the name of the form should be automatically set by the crawler.	Identify the URL of the login page, and check whether this is a regular HTML form login page or an OracleAS Single Sign-on login page. Report the problem to Oracle Support.
30199	Encountered an error while responding to the following HTTP authentication request: [{0}]	Unable to authenticate through the target URL.	Verify if the authentication request is basic authentication or digest authentication. Also confirm the provided authentication credentials.
30201	Missing authentication credentials	Authentication data is not available to access the URL.	Check the type of authentication needed and provide it through the source customization page
30206	Ignoring "{0}" due to host (or redirected host) connection problem	The crawler cannot contact the server of the URL.	Verify that the Web site in question is up and try to re-crawl.
30209	Document size ({0}) too big, ignored: {1}	Document size exceeds the default limit of 10 megabytes.	Increase the document size limit on the Global Settings - Crawler Configuration page.
30215	Excluded by crawling depth limit({0}): {1}	Previously crawled URL is excluded due to newly reduced crawling depth limit.	Confirm that the depth limit is correct.
30782	Invalid document attribute {0} - ignored	Some attribute picked up from the document is not defined for the source. It is ignored.	Most likely this is safe to ignore, unless you know that this particular attribute should be defined for this source. In that case, contact Oracle Support.

Glossary

ACE

An entry in an access control list (ACL). An ACE either grants or denies access to a data source by a particular user.

ACL

A list of access control entries that determines which users have access to a particular data source.

crawl

The process of reading sources and creating the search engine index.

crawler

An Oracle Secure Enterprise Search program that reads sources to create the search engine index.

DN

Distinguished Name. The unique name of a directory entry in Oracle Internet Directory. It includes all the individual names of the parent entries back to the root. The DN tells you exactly where the entry resides in the directory's hierarchy.

DICOM

Digital Imaging and Communications in Medicine (DICOM) is the predominant medical imaging standard for exchanging digital information between medical imaging equipment (such as radiological imaging) and other systems, ensuring inter operability. DICOM images contain rich metadata about the patient, medical equipment, and other medical information. DICOM is a major feature of Oracle Multimedia 11g release.

document

Unit of indexing, returned as one entry in the hitlist. For example, a document could be all the collected information about a person from a human resources system.

duplicate documents

Documents that are identical to each other; that is, they are the exact same size, same content, same title, and so on.

federated search

Oracle SES provides the capability of searching multiple Oracle SES instances with their own document repositories and indexes. It provides a unified framework to search the different document repositories that are crawled, indexed, and maintained

separately. A *federation broker* calls the *federation endpoint* to collect content matching the search criteria for the sources managed at that endpoint.

hitlist

A list of results for a search.

infosource

Data sources that can be browsed based on the source or path of the documents. An infosource hierarchy has this structure: *Source Group* > *host* or *folder* > *folder*. An infosource hierarchy provides a count of documents at levels below the top (that is, the Source Group level). A user's access to the infosource structure relies on the user's access to documents stored in the structure.

Users can search documents under a particular path or level by selecting the corresponding node in the infosource browse hierarchy.

index

An Oracle SES structure that is updated after a crawl. It is used to improve performance of searches.

JDBC

The programming API that enables Java applications to access a database through the SQL language. JDBC drivers are written in Java for platform independence but are specific to each database.

LDAP

Lightweight Directory Access Protocol. A standard for representing and accessing user and group profile information.

lexer

A program that breaks the source text into tokens, usually words, in accordance with a specified language.

LOV

List of values.

near duplicate documents

Documents that are similar to each other. They may or may not be identical to each other.

Oracle Application Server (Oracle AS)

Oracle's integrated application server:

- Is standards compliant (J2EE, Web Services, and XML)
- Delivers a comprehensive set of capabilities, including portal, caching, wireless, integration, and personalization
- Provides a single, unified platform for Java and J2EE, Web Services, XML, SQL, and PL/SQL

OracleAS Portal

A component of Oracle Application Server used for the development, deployment, administration, and configuration of enterprise class portals. OracleAS Portal incorporates a portal building framework with self-service publishing features to enable you to create and manage information accessed within your portal.

OracleAS Single Sign-On

A component of Oracle Application Server that enables users to log in to all features of the Oracle AS product suite and to other Web applications, using a single user name and password.

OracleAS Web Cache

A component of Oracle Application Server that improves the performance, scalability, and availability of frequently used Web sites. By storing frequently accessed URLs in memory, Oracle Application Server Web Cache eliminates the need to repeatedly process requests for those URLs on the Web server.

Oracle Content Database

A consolidated, database-centric content management application that provides a comprehensive, integrated solution for file and document life cycle management. Oracle Content Database also offers a comprehensive set of Web services that developers can use to build and enhance content management applications. This book uses the product name Oracle Content Database to mean *both* Oracle Content Database *and* Oracle Content Services.

Oracle Content Server

Formerly known as Stellent Content Server. Oracle Content Server enables users throughout the organization to contribute content from native desktop applications, manage content through rich library services, publish content to web sites or business applications, and access the content with a browser.

Oracle Content Services

See **Oracle Content Database**.

Oracle HTTP Server

The Web server component of Oracle Application Server, built on Apache Web server technology and used to service HTTP requests. Also referred to as OHS in the guide.

Oracle Internet Directory

A repository for storing user credentials and group memberships. By default, the OracleAS Single Sign-On authenticates user credentials against Oracle Internet Directory information about dispersed users and network resources.

Oracle Secure Enterprise Search application

Application for searching the Oracle Secure Enterprise Search index.

Oracle WebLogic Server

Oracle WebLogic Server is a fast and reliable server that is used to build and run enterprise applications and services. It is the middle tier server on which Oracle SES operates.

relevance

The level of match of the search results to the search string.

schedule

The frequency with which each source is crawled.

search

The process of querying the search engine.

searchctl

A utility for starting and stopping the search engine.

search metadata

Information about the sources, crawls, and schedules.

secure search

A type of search that only returns results that the user is allowed to view based on access privileges.

seed URL

The starting point for a crawl.

SOAP

Simple Object Access Protocol. A lightweight, XML-based protocol for exchanging information in a decentralized, distributed environment. SOAP supports different styles of information exchange, including: Remote Procedure Call style (RPC) and Message-oriented exchange.

sources

A source of data to be searched, such as Web sites, database tables, content management repositories, collaboration repositories, or applications.

WebDAV

Web Distributed Authoring and Versioning (WebDAV) is a standard protocol used to provide users with a file system interface to Oracle XML repository over the Internet. The most popular way of accessing a WebDAV server folder is through WebFolders on Microsoft Windows.

WebDAV is an extension to the HTTP 1.1 protocol. It allows clients to perform remote web content authoring through a coherent set of methods, headers, request body formats, and response body formats. WebDAV provides operations to store and retrieve resources, to create and list contents of resource collections, to lock resources for concurrent access in a coordinated manner, and to set and to retrieve resource properties.

WSDL

A general purpose XML language for describing the interface, protocol bindings, and deployment details of Web services.

Index

Symbols

DREQDOC_PATH_IDX$I, 10-19

A

access URL, 3-2
ACLs
  defined, 9-4
  policies, 9-5, 9-18
    restrictions, 9-6, 9-7
Active Directory
  activating the plug-in, 9-8
  IDM systems, 5-13
Administration GUI, 1-3
administrative user
  SEARCHSYS, 9-2, 9-11
AJP13 protocol, 5-14
  from remote hosts, 9-2
alternate words, 2-3
Apache Axis
  license, C-1
Apache log4j
  license, C-1
APIs
  Authorization Plug-in, 9-6
  Web Services, 11-2
    Administration Web Service, 11-1
    Query Web Service, 11-1
application identity, 1-7, 3-25
Application Server Control Console
  overview, 10-21
Applications Control, 3-25
attributes
  attribute-based security, 3-12
  mapping federated, 5-17
  overview, 3-20
  retrieving federated, 5-17
  tuning the weights of, 4-3
authentication, 10-13
authorization, 10-13
  ACLs, 9-17
  query-time filtering, 9-18
  self service, 9-19
authorization plug-in
  Fusion, 8-1
  WebCenter, 8-3
Authorization Plug-in API, 9-6

B

boundary control of Web crawling, 3-3
boundary rules, 2-4, 3-26
  defined, 3-4
  example using regular expression, 3-5
  exclusion rules, 3-4
  inclusion rules, 3-4
  permanent redirect, 10-9
  tuning, 10-7
  with dynamic pages, 10-8
  with file sources, 10-7
  with Portal sources, 5-11
  with symbolic links, 5-7
buffer cache, 10-19

C

character set detection, 3-7
Chinese, 1-7, 3-7, 3-8
cluster configuration, 10-17
crawler, 3-1
  crawling multimedia files, 3-5
  crawling process, 3-20
  depth, 3-6, 10-8
  document types
    zip files restriction, 3-5
  log file, 3-26, 10-11
    crawler.dat configuration file, 3-26
    enabling character set detection, 3-7
    setting default document titles, 3-7
  maintenance crawls, 3-24
  monitoring the crawling process, 3-25
  overview, 3-1
  URL status codes, B-1
crawler configuration, 2-4
crawler status
  Error Manual Recovery, 10-6
crawling mode, 3-3

D

data files, 10-1

database initialization parameters, 10-18
debug mode, 10-20
DICOM, 3-19
dicom, 3-19
display URL, 3-2
document attributes, 3-20
document service, 9-12
document types
  zip files restriction, 3-5
domain rules, 3-4
duplicate documents, 10-8
  dupMarked, 11-7, 11-16, 11-17, 11-18
  dupRemoved, 11-6, 11-16, 11-17, 11-18
  hasDuplicate, 11-7
  isDuplicate, 11-7
  versus near duplicate documents, 10-8
dynamic pages, 10-8

E

easy connect naming method, 10-21
encryption key, 9-43
Enterprise Manager Applications Control, 3-25
Error Manual Recovery status, 10-6
error messages, D-1
ESSAPP, 3-25

F

faceted navigation, 4-4
failed schedules, 2-2, 10-5
failover support, 10-20
federated search, 1-8
  characteristics, 5-18
  example, 5-15
  limitations, 5-19
  setting up, 5-12
  trusted entities, 5-13
federation trusted entities, 5-13
file sources
  crawling file URLs, 5-7
  multibyte environments, 5-7
  tips, 5-7
  URL boundary rules
    with file sources, 10-7
  with symbolic links, 5-7
Fusion Connector, 8-1
FUSION_APPS_SEARCH_APPID, 1-7, 3-25

G

gif, 3-19
Google Desktop for Enterprise
  integrating with, 10-21

H

hit count
  approximate count, 9-19
  exact count, 9-19
  exact count (adjusted for query-time filtering), 9-19
HTML forms, 9-4
HTTP authentication, 9-4, 9-11
HTTP protocol, 3-2, 5-7, 9-2
HTTP proxy server, 5-1
HTTP proxy servers, 10-6
HTTP status codes, 3-27, 10-9, 10-21, B-1
HTTPS protocol, 3-2, 5-14, 9-2, 9-34, 9-37

I

identity management systems, 2-4, 9-1, 9-4, 9-7, 9-11, 9-17
identity plug-in, 8-1
  Fusion, 8-1
identity plug-ins, 2-4, 9-7
  ACLs, 9-5
  activating, 9-7
  define, 9-1
  re-registering, 9-9
  restrictions, 9-10
  user authentication, 9-4
image format
  dicom, 3-19
  gif, 3-19
  tiff, 3-19
image formats
  jpeg, 3-19
IMAP server, 9-20
  mailing list sources, 5-8
index memory size, 10-16
index optimization, 10-14
indexing
  stopwords, 3-23
indexing batch size, 10-15
indexing parameters, 10-15
initialization parameters, 10-18

J

Japanese, 1-7, 3-7, 3-8, 7-10, 8-7, 8-15
JDBC, 8-6, 9-1
jpeg, 3-19

K

KEEP pool, 10-19
key, master encryption, 9-43
Korean, 1-7, 3-8, 5-7

L

list of values (LOV), 3-21
log files
  crawler log file, 10-11

M

mailing list sources
  tips, 5-8
master encryption key, 9-43

metadata, 3-20
multimedia files
  crawling, 3-5
MW_HOME, xiv

N

navigation tools, 4-4

O

OAM
  See Oracle Access Manager
OC4J server, 11-4
open_cursors parameter, 10-18
optimizing
  index, 10-14
Oracle Access Manager, 9-28
Oracle Calendar sources
  secure, 7-22
Oracle Content Database sources, 6-18
  tips, 6-18
Oracle Content Services, 6-18
Oracle Enterprise Manager, 3-25
Oracle HTTP Server
  channel with Oracle SES, 9-12
  front-ending, 9-11, 9-20
  SSL-protect, 5-14
  with AJP13 port, 5-14
Oracle Internet Directory
  identity plug-in, 9-11
    restrictions, 9-10
  IDM systems, 5-13
  login attribute, 7-22
  overview, 9-11
Oracle RAC
  failover, 10-20
  tuning, 10-17
Oracle Secure Enterprise Search
  accessing Application Server Control
      Console, 10-21
  Administration GUI, 1-3
  components, 1-3
  crawler, 1-4, 3-1
  error messages, D-1
  getting started, 2-1
  global settings, 2-4
  integration with Oracle Internet Directory, 9-11
  overview, 1-1
  security, 9-1
  statistics, 2-3
  third party licenses
    Apache Axis, C-1
    Apache log4j, C-1
  tuning crawl performance, 10-4
  what's new in 10.1.7, xv
ORACLE_HOME, xiv
OracleAS Portal sources, 9-4
  tips, 5-11
  user privileges, 5-11
OracleAS Single Sign-On, 9-4, 9-11

P

parallel querying, 10-14
partitioning, 10-14
passwords
  temporary, 9-4
path rules, 3-4
physical memory, 10-19
processes parameter, 10-18
proxy servers, 10-6

Q

query application, 9-12
  customize results, 4-3
  suggested content, 4-1
query configuration, 2-4
query-time authorization
  comparison with ACLs, 9-6
  configuration, 9-18

R

redo log, 10-10
relevancy boosting, 2-3, 10-16
  limitations, 10-17
result filter, 6-22
ResultFilterPlugin class, 9-19
robots META tag, 3-6, 10-8
robots.txt file, 3-6, 10-8
robots.txt protocol, 3-6, 10-8
rollover_key, 9-43
rules
  domain, 3-4
  path, 3-4

S

schedules, 2-3
  failed, 2-2
  fixing stuck requests, 10-6
  understanding, 10-5
search attributes
  default, 3-20
search performance, 2-3
search results
  narrowing, 4-4
search server configuration, 10-17
SEARCH_DATA tablespace, 10-1
SEARCH_INDEX tablespace, 10-1
SEARCH_TEMP tablespace, 10-1
searchctl rollover_key, 9-43
SEARCHSYS
  administrative user, 9-2, 9-11
secure search, 1-8
  identity plug-ins, 2-4
security filters, 9-4, 9-6, 10-13
self service authorization, 9-19
sessions parameter, 10-18
SOAP, 11-2, 11-3, 11-4
  client applications using, 11-5

development environment, 11-6
message body, 11-4
messages, 11-33
source groups, 2-4
source hierarchy, 2-4
sources
synchronizing, 3-1
types
database, 8-6
e-mail, 1-2
EMC Documentum Content Server, 6-1
EMC Documentum eRoom, 7-1
federated, 5-12
file, 1-2
Lotus Notes, 7-4
mailing list, 1-2
Microsoft Exchange, 7-8
NTFS for UNIX, 7-14
NTFS for Windows, 7-11
Oracle Calendar, 7-22
Oracle Content Database, 6-18
Oracle Mail, 7-23
OracleAS Portal, 1-2
Siebel 7.8, 8-14
Siebel 8, 8-32
table, 1-2
Web, 1-2
spell checking, 2-4
SQL*Plus
connecting using, 9-1
SSL, 9-1, 9-34
certificates, 9-34
crawling Web site with SSL certificates, 9-36
importing certificates, 9-36
in Oracle SES, 9-34
JSSE, 9-34
keystore, 9-34
statistics, 2-3, 10-16
stoplist, 3-23
stopwords, 3-23
storage areas, 10-14
stuck threads, 10-17
suggested content, 4-1
example with Google OneBox, 4-3
security options, 4-2
suggested links, 2-3, 10-12

T

tablespaces, 10-1
temp files, 10-1
temporary passwords, 9-4
threads, stuck, 10-17
tiff, 3-19
time outs, 10-13
tips
using file sources, 5-7
using mailing list sources, 5-8
using Oracle Calendar sources, 7-22
using Oracle Content Database sources, 6-18
using OracleAS Portal sources, 5-11
using user-defined sources, 5-7
titles, changing, 3-7
trusted entities, 5-13

U

undo tablespace, 10-19
UNDO_RETENTION parameter, 10-19
updateCred command (WLST), 10-21
URL boundary rules, 2-4, 3-26
defined, 3-4
permanent redirect, 10-9
tuning, 10-7
with dynamic pages, 10-8
with Portal sources, 5-11
with symbolic links, 5-7
URL crawler status codes, B-1
URL looping, 10-9
URL queue, 3-2
user authentication, 9-4
user authorization, 9-4
user-defined sources, 1-2, 2-3
tips, 5-7

W

Web crawling
boundary control, 3-3
Web Services API, 11-1, 11-2
architecture, 11-4
concepts, 11-3
SOAP, 11-4
WSDL, 11-4
data types, 11-5
example, 11-26
installation, 11-2
operations, 11-10
query syntax, 11-24
URL, 11-2
WebLogic Server Administration Console, 10-21
WebLogic server configuration, 10-17
WSDL specification, 11-4

X

XML connector framework, 3-9
examples, A-1
Oracle E-Business Suite, 8-4
schemas, A-1
Siebel 8, 8-32