

Oracle Endeca Information Discovery

CAS Installation Guide

Version 3.0.2 • July 2012

Copyright and disclaimer

Copyright © 2003, 2012, Oracle and/or its affiliates. All rights reserved.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners. UNIX is a registered trademark of The Open Group.

This software and related documentation are provided under a license agreement containing restrictions on use and disclosure and are protected by intellectual property laws. Except as expressly permitted in your license agreement or allowed by law, you may not use, copy, reproduce, translate, broadcast, modify, license, transmit, distribute, exhibit, perform, publish or display any part, in any form, or by any means. Reverse engineering, disassembly, or decompilation of this software, unless required by law for interoperability, is prohibited.

The information contained herein is subject to change without notice and is not warranted to be error-free. If you find any errors, please report them to us in writing.

If this is software or related documentation that is delivered to the U.S. Government or anyone licensing it on behalf of the U.S. Government, the following notice is applicable:

U.S. GOVERNMENT END USERS: Oracle programs, including any operating system, integrated software, any programs installed on the hardware, and/or documentation, delivered to U.S. Government end users are "commercial computer software" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, use, duplication, disclosure, modification, and adaptation of the programs, including any operating system, integrated software, any programs installed on the hardware, and/or documentation, shall be subject to license terms and license restrictions applicable to the programs. No other rights are granted to the U.S. Government.

This software or hardware is developed for general use in a variety of information management applications. It is not developed or intended for use in any inherently dangerous applications, including applications that may create a risk of personal injury. If you use this software or hardware in dangerous applications, then you shall be responsible to take all appropriate fail-safe, backup, redundancy, and other measures to ensure its safe use. Oracle Corporation and its affiliates disclaim any liability for any damages caused by use of this software or hardware in dangerous applications.

This software or hardware and documentation may provide access to or information on content, products and services from third parties. Oracle Corporation and its affiliates are not responsible for and expressly disclaim all warranties of any kind with respect to third-party content, products, and services. Oracle Corporation and its affiliates will not be responsible for any loss, costs, or damages incurred due to your access to or use of third-party content, products, or services.

Rosette® Linguistics Platform Copyright © 2000-2011 Basis Technology Corp. All rights reserved.

Teragram Language Identification Software Copyright © 1997-2005 Teragram Corporation. All rights reserved.

Table of Contents

Copyright and disclaimer	ii
Preface	iv
About this guide	iv
Who should use this guide	iv
Conventions used in this guide	iv
Contacting Oracle Customer Support	iv
CAS and Oracle Endeca Information Discovery	v
Chapter 1: Before you install	1
Overview of the Content Acquisition System	1
System requirements	1
CAS Document Conversion Module	2
Licensing for CMS connectors	2
Supported CMS connectors	2
Recommended reading	3
Chapter 2: Installing the Content Acquisition System	4
Creating a user for the Endeca services on Windows	4
Installing CAS on Windows	5
Installing CAS on UNIX	6
Adding the Endeca CAS service to inittab on UNIX	7
Installing CAS silently on Windows	7
Installing CAS silently on UNIX	8
Installing a plug-in into CAS	9
Web Crawler installation	10
Package contents and directory structure	10
Chapter 3: Uninstalling the Content Acquisition System	13
Uninstalling CAS on Windows	13
Uninstalling CAS on UNIX	13
Uninstalling a plug-in from CAS	13

Preface

Oracle® Endeca Information Discovery is an enterprise data discovery platform for advanced, yet intuitive, exploration and analysis of complex and varied data.

Information is loaded from disparate source systems and stored in a faceted data model that dynamically supports changing data. This integrated and enriched data is made available for search, discovery, and analysis via interactive and configurable applications.

Oracle Endeca Information Discovery enables an iterative “model-as-you-go” approach that simultaneously frees IT from the burdens of traditional data modeling and supports the broad exploration and analysis needs of business users.

About this guide

This guide describes the major installation tasks for the Content Acquisition System (CAS) used with Oracle Endeca Information Discovery.

It assumes that you are familiar with the concepts of Oracle Endeca Information Discovery.

Who should use this guide

This guide is intended for developers who are responsible for loading source data and configuration documents into an Endeca data store.

Conventions used in this guide

This guide uses the following typographical conventions:

Code examples, inline references to code elements, file names, and user input are set in `monospace` font. In the case of long lines of code, or when inline monospace text occurs at the end of a line, the following symbol is used to show that the content continues on to the next line: ↵

When copying and pasting such examples, ensure that any occurrences of the symbol and the corresponding line break are deleted and any remaining space is closed up.

Contacting Oracle Customer Support

Oracle Customer Support provides registered users with important information regarding Oracle software, implementation questions, product and solution help, as well as overall news and updates from Oracle.

You can contact Oracle Customer Support through Oracle's Support portal, My Oracle Support at <https://support.oracle.com>.

CAS and Oracle Endeca Information Discovery

This topic explains which CAS features are supported for use with Oracle Endeca Information Discovery. It also notes features that are not supported and can be ignored.

Your Oracle Endeca Information Discovery license permits you to use the command-line version of the Endeca Content Acquisition System (CAS). The following features are not supported:

- CAS Console
- Forge
- Dgidx

References to these features are not relevant to users of Oracle Endeca Information Discovery, and can be ignored. Specifically, the following documents and topics can be ignored:

- CAS Console Help (CASConsoleHelp.pdf)
 - Entire document
- CAS Developer's Guide (CASDevGuide.pdf)
 - Before you install
 - System Requirements
 - CAS Console and Web browser support
 - Running the CAS sample applications
 - About the sample CAS applications
 - Order of running the Forge applications
 - Writing records from Forge into the Record Store
 - Reading records from the Record Store into Forge
 - Reading records from multiple Record Stores into Forge
 - Creating a Forge pipeline to read from or write to a Record Store
 - Configuring SSL
 - Enabling SSL for CAS Console for Oracle Endeca Workbench
 - Enabling SSL for Forge record adapters
 - Tips and troubleshooting CAS
 - Modifying the CAS Server connection information for the CAS Console
- CAS Installation Guide (as distributed with Oracle Endeca Guided Search documentation)
 - Installing the Content Acquisition System
 - Installing only the CAS Console on Unix
 - About changing the role used for the CAS Console extension
 - Uninstalling the Content Acquisition System
 - Uninstalling CAS Console on UNIX
 - Uninstalling CAS Console if its extension configuration was changed

- CAS Quick Start
 - Using the Endeca Content Acquisition System
 - Adding a data source using CAS Console

In addition, the remaining text of the CAS documentation includes frequent references to CAS Console, Forge, and Dgidx. References to these features can be ignored.



Chapter 1

Before you install

This section provides an overview of the Content Acquisition System, system requirements, and other information you need to know before installing.

[Overview of the Content Acquisition System](#)

[System requirements](#)

[CAS Document Conversion Module](#)

[Licensing for CMS connectors](#)

[Supported CMS connectors](#)

[Recommended reading](#)

Overview of the Content Acquisition System

The Endeca Content Acquisition System (CAS) adds, configures, and crawls data sources for use in by Oracle Endeca Information Discovery. Data sources include file systems, Content Management System crawls, Web servers, and custom data sources. The Endeca Content Acquisition System crawls data sources and stores the results in a Record Store for extraction and loading into an Endeca server data source.

System requirements

The Content Acquisition System has the following system requirements. In this guide, the term “x64” refers to any processor compatible with the AMD64/EM64T architecture. Hard disk capacity must be sufficient to store the records written to the Record Store or to record output files. Please contact your Oracle representative if you need more information on sizing your hardware.

Supported operating systems

The Content Acquisition System is supported on the following 64-bit operating systems:

- Windows Server 2003 (For best performance on Windows, Oracle recommends Windows Server 2003 Enterprise Edition Service Pack 2.)
- Windows Server 2008 R2 Enterprise
- Red Hat Enterprise Linux ES (version 4 for x64)
- Red Hat Enterprise Linux AS (version 4 for x64)
- Red Hat Enterprise Linux Server (version 5 for x64)
- Red Hat Enterprise Linux Advanced Platform (version 5 for x64)

- SUSE Enterprise Linux 11

Hardware requirements for Linux and Windows on x64

Minimum hardware requirements:

- x64 processor, minimum 1.8 GHz
- 2 GB of RAM
- At least an 80 GB hard drive, depending on the size of your application data set



Note: If you are using RHEL 4 and the virtual address size of your CAS processes exceeds available RAM, it is recommended that you upgrade to the latest version of RHEL 5 to avoid known problems with RHEL 4 performance in this scenario.

CAS Document Conversion Module

In CAS 3.0.1 and later, the CAS Document Conversion Module is installed and enabled as part of the CAS installation.



Note: The CAS Document Conversion Module is not supported on SUSE Enterprise Linux 11.

For a list of supported file formats, see "Appendix B File Formats Supported by the CAS Document Conversion Module" in the *Endeca CAS Developer's Guide*.

Licensing for CMS connectors

In CAS 3.0.1 and later, all CMS connectors are installed and enabled as part of the CAS installation.

Supported CMS connectors

To run a crawl against a specific CMS data source, the Content Acquisition System requires a CMS connector.

As of this release, the Content Acquisition System supports the following CMS connectors:

- Documentum Content Server
- Documentum eRoom
- FileNet Document and Image Services
- FileNet P8
- Interwoven TeamSite
- JSR-170 compliant repositories
- Lotus Notes (Domino)
- Microsoft SharePoint (supports 2003 and 2007 repositories)
- OpenText LiveLink

For information about configuring each of the CMS connectors, see the corresponding Connector Guide for the specific CMS connector.

Recommended reading

Before installing, Oracle recommends that you read the following documents for important information about the release.

Quick Start Guide

The *Oracle Endeca Information Discovery Quick Start Guide* guides you through the process of installing and exploring a sample Endeca Information Discovery application.

Getting Started Guide

The *Oracle Endeca Information Discovery Integrator Getting Started Guide* provides an overview of the Integrator and guides you through your first Integrator project.

Release Notes

Refer to the release notes for information about new features, changed features, and bug fixes for this release. The release notes (`README.txt`) are part of the CAS documentation download. After installation, release notes are also available in the following location:

- Windows: `CAS\version`
- UNIX: `CAS/version`



Chapter 2

Installing the Content Acquisition System

This section contains installation procedures and describes the contents of the installation directory.

[Creating a user for the Endeca services on Windows](#)

[Installing CAS on Windows](#)

[Installing CAS on UNIX](#)

[Adding the Endeca CAS service to inittab on UNIX](#)

[Installing CAS silently on Windows](#)

[Installing CAS silently on UNIX](#)

[Installing a plug-in into CAS](#)

[Web Crawler installation](#)

[Package contents and directory structure](#)

Creating a user for the Endeca services on Windows

You must run the Endeca services as a specified user, for which you can control permissions.

You are asked to provide information about this Endeca services user during the installation process. Oracle recommends that you create a user account called `endeca` that has the proper file and directory permissions to access all necessary files for your application, and that you set up your Endeca Tools Service to run under this account. However, you can use any user that you prefer, as long as it meets these requirements.

To create the user `endeca`:

1. Ensure that you have administrator privileges on the local machine.
2. From the Windows Control Panel, select **Computer Management**>**Administrative Tools**.
3. In the tree pane of the Computer Management window, select **System Tools**>**Local Users and Groups**.
4. Right-click **Users** and select **New User**.
5. In the **New User** dialog box, do the following:
 - (a) Enter `endeca` for both the User name and the Full name.
 - (b) Optionally, enter a description.
 - (c) Set a password for user `endeca`.



Note: The user must have a non-blank password, because the installer will not accept a blank password for the services user.

- (d) Uncheck **User must change password at next logon**.
 - (e) Select **Password never expires**.
6. Click **Create** to create the new user, and then **Close** to exit the dialog box.
7. Close the Computer Management window, but do not exit Administrative Tools.
8. From Administrative Tools, do the following:
 - (a) Open **Local Security Policy**.
 - (b) In the Local Security Settings window, select **Local Policies>User Rights Assignments**.
 - (c) From the list of policies on the right, open **Log on as a Service**.
 - (d) Add user `endeca` to the list of users that can register a process as a service.
 - (e) Close the dialog box and exit Administrative Tools.

Installing CAS on Windows

This section provides instructions for installing the Endeca Content Acquisition System on Windows.

Before installing, ensure that you have met the following pre-requisites:

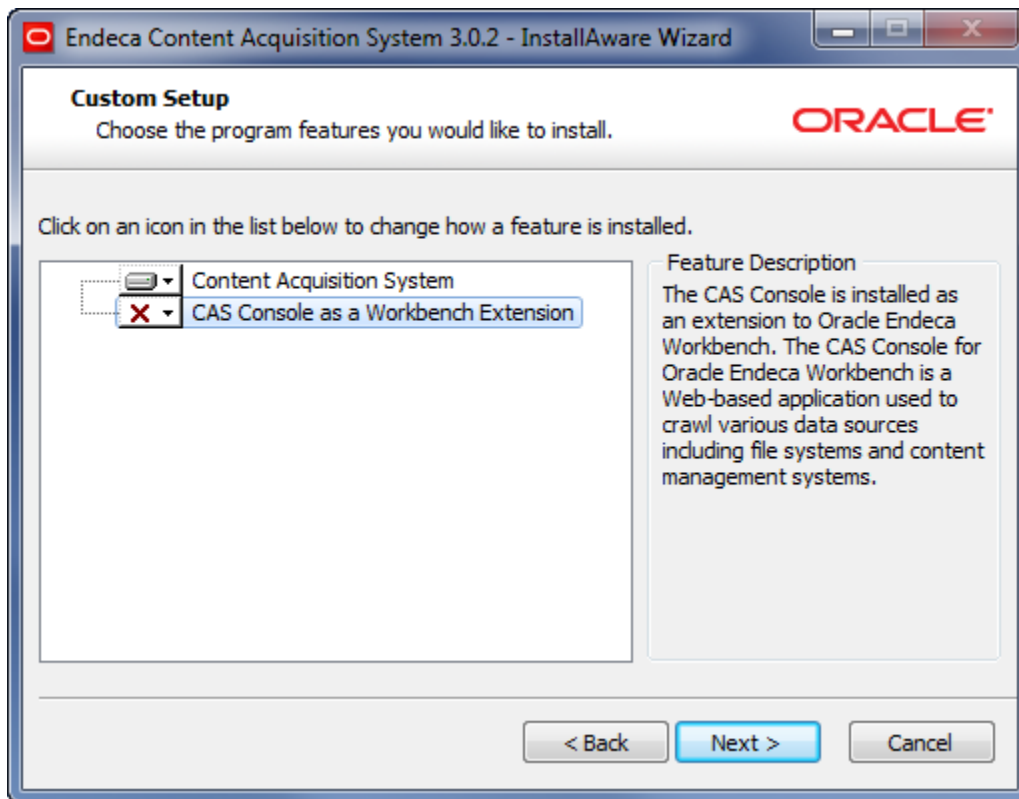
- Ensure you have administrator privileges on the local machine.
- Close all running programs.



Note: Even if you are installing on a drive other than the C drive, you should have approximately 400 MB of space available on the system drive before installing. This space is needed temporarily for the installer files that are unpacked on the C drive during the installation regardless of the drive you specified for your installation. The space is cleared after the installation process completes.

To install CAS on Windows:

1. In your local environment, locate the Endeca CAS software that you downloaded from the Oracle Software Delivery Cloud.
2. Double-click the installer file `cas-version_x86_64pc-win32.exe` to start the wizard.
3. Click **Next** to begin the installation process.
4. In the **Important Information** screen, read the copyright and then click **Next**.
5. In the **Custom Setup** dialog, click on the **CAS Console as a Workbench Extension** option and from the popup menu choose **Entire feature will be unavailable**. Click **Next**.



6. In the **Destination Folder** screen, select an installation location or accept the default location of `C:\Endeca\CAS` and then click **Next**.
7. In the **Endeca CAS Service Information** screen, specify the user name, password, and domain information for the user who will run the CAS Service and then click **Next**. (This is typically the `endeca` user you created in the previous procedure.)
8. In the **CAS Server Information** screen, enter the CAS Server port and CAS Server shutdown port, or accept the default values of 8500 and 8506.
9. In the **Completing the Setup Wizard** screen, click **Next**.

The CAS Service starts automatically.

Installing CAS on UNIX

The Endeca software is distributed as a self-extracting tar file and install script. It can be installed to any location.

To install CAS on UNIX:

1. Determine an installation location for the Endeca Content Acquisition System. In this procedure, we assume this is the `/usr/local` directory.



Note: CAS unpacks to approximately 400 MB. Because multiple versions may be stored, a destination in a large partition is recommended.

2. Locate the CAS installation file that you downloaded from the Oracle Software Delivery Cloud. This procedure assumes the location is `downloads/arch-OS`. The name of the installation file is `cas-version_arch-OS.sh`.
3. Assuming the locations in steps 1 and 2, run the Endeca installation script with the following command:

```
downloads/arch-OS/cas-version_arch-OS.sh --target /usr/local
[--skip_console_installation]
```

You must use the `--skip_console_installation` flag to ensure that CAS is installed without the CAS Console; this feature is not licensed for use with Oracle Endeca Information Discovery.

4. If the installer prompts you for `ENDECA_TOOLS_ROOT` or `ENDECA_TOOLS_CONF` and the CAS Server host name, you have run the installer without the `[--skip_console_installation]` flag. Terminate the installation and run it with the `[--skip_console_installation]` flag.

Following installation, to start the CAS Service, navigate to `/usr/local/endeca/CAS/version/bin` and run the script `cas-service.sh`

Adding the Endeca CAS service to inittab on UNIX

In a UNIX development environment, the Endeca CAS Service can be started from the command line. In a UNIX production environment, however, Oracle recommends starting it from `inittab`.

In UNIX you run the CAS Server using `cas-service.sh`, located in `usr/local/endeca/CAS/<version>/bin`. You can write a script that calls `cas-service.sh` and is referenced in `inittab`.

When writing your script, it is recommended as a best practice that you run the Endeca CAS Service as a user other than root.

This sample script (named `start_cas_service.sh`) sets the `ENDECA_USER` variable to the “endeca” user and then issues an `su` command to change to the “endeca” user:

```
#!/bin/sh
ENDECA_USER=endeca
CAS_ROOT=/usr/local/endeca/CAS/3.0.2
CAS_WORKSPACE=/usr/local/endeca/CAS/workspace

# change to user endeca
su $ENDECA_USER -c "cd $CAS_ROOT/bin; \
    $CAS_ROOT/bin/cas-service.sh >> $CAS_WORKSPACE/logs/cas-service-wrapper.log 2>&1"
```

The `start_cas_service.sh` script is then referenced in `inittab` with an entry similar to this example:

```
ecas:2345:respawn:/usr/local/endeca/CAS/3.0.2/bin/start_cas_service.sh
```

When writing your startup script, keep in mind that it is server-specific, and therefore its details (such as paths and user names) depend on the configuration of your server.

Installing CAS silently on Windows

The silent installer is useful if you want to add the CAS installation to your own install script, or push out the installation on multiple machines.

The silent installer is not interactive.

To launch the silent installer:

1. From a command prompt, navigate to the directory where you downloaded the installer.
2. Issue the following command:

```
start /wait cas-version-arch-OS.exe /s /l=full path to the installer log file\installer.log
[CASSELECTED=TRUE] [CONSOLESELECTED=FALSE] [CASPORT=8500] [CASSTOPPORT=8506] [TARGETDIR
=C:\Endeca] USERNAME=endeca PASSWORD=endecapwd CONFIRMPASSWORD=endecapwd DOMMAINNAME=domain
```

You must specify `CONSOLESELECTED=FALSE`; the CAS Console is not licensed to users of Oracle Endeca Information Discovery, so it should not be installed.

The logging option (`/l=`) provides useful information from silent installer failures. You must specify the full path to the directory where you want to write the log file, as well as the name of the log file itself. The specified directory must be a full path.

You can replace `C:\Endeca` with the location to which you want to install. However, if you set the install location to a drive that does not exist, the silent installation fails with a non-zero status code.

Installing CAS silently on UNIX

The silent installer is useful if you want to add the CAS installation to your own install script, or push out the installation on multiple machines.

The silent installer is not interactive.

To run the silent installer:

1. Create a `.txt` file that contains rows to specify the response value for each prompt on its own line. The order of prompts in the installer is:

```
CAS port
CAS shutdown port
CAS host
```

For example:

```
8500
8506
localhost
```

2. From a command prompt, navigate to the location of the installer.
3. Run the install script with the following flags: `--silent` flag, the `--target` flag (which specifies the absolute path of the target installation directory), and the `--skip_console_installation` flag. For example:

Option	Description
<code>--silent</code>	Runs in the installer in silent mode
<code>--target</code>	Specifies the absolute path to the target installation directory

Option	Description
<code>--skip_console_installation</code>	Prevents the installer from installing the CAS Console. The CAS Console is not licensed to Users of Oracle Endeca Information Discovery, so it should not be installed.

For example:

```
./cas-version-arch-OS.sh --silent
--target /usr/local < silent.txt --skip_console_installation
```

Following installation, to start the CAS Service, navigate to `CAS/version/bin` and run the following command: `cas-service.sh`.

Installing a plug-in into CAS

After receiving a plug-in (a JAR or set of JAR files) from a plug-in developer, a CAS application developer installs the plug-in into CAS.

The Content Acquisition System detects each plug-in and validates the extensions within it by checking the uniqueness of extension IDs and by checking for the presence of an annotation of either `@CasDataSource` or `@CasManipulator` for each extension.

To install a plug-in into CAS:

1. Stop Endeca CAS Service.
2. Navigate to `<install path>\CAS\version\lib\cas-server-plugins` and create a *plugin-name* subdirectory for each plug-in.
For example: `CAS\version\lib\cas-server-plugins\JDBCDataSourceExt`
3. Copy the plug-in JAR or JARs, and any dependent JAR files, to `<install path>CAS\version\lib\cas-server-plugins\plugin-name`.
4. Repeat the steps above as necessary for multiple plug-ins.
5. Start Endeca CAS Service.

You can confirm that an extension is installed by running the `listModules` task of the CAS Server Command-line Utility and specifying a `moduleType` of either `SOURCE` or `MANIPULATOR`. The task returns the installed modules. For example, this task shows that a custom data source named `Sample Data Source for testing` is installed:

```
C:\Endeca\CAS\3.0.0\bin>cas-cmd listModules -t SOURCE
Sample Data Source
 *Id: Sample Data Source
 *Type: SOURCE
 *Description: Sample Data Source for testing

File System
 *Id: File System
 *Type: SOURCE
 *Description: No description available for File System
 *Capabilities:
  *Binary Content Accessible via FileSystem
  *Data Source Filter
  *Has Binary Content
  *Expand Archives
```

Web Crawler installation

The Endeca Web Crawler is installed as part of the CAS installation.

The Web Crawler startup script is in `<install path>\CAS\<version>\bin` (Windows) and `<install path>/CAS/<version>/bin` (UNIX). The Web Crawler configuration files are in `<install path>\CAS\conf\web` (Windows) and `<install path>/CAS/conf/web` (UNIX). See the *Endeca Web Crawler Guide* for details and usage.

Package contents and directory structure

The CAS directory is the root directory for the Content Acquisition System.

After installing the Content Acquisition System, you see this directory structure:

```
CAS
 3.0.2
   bin
   ...
 workspace
```

The contents of the CAS directory are described here in detail.

Directory	Contents
<code>version\bin</code>	<ul style="list-style-type: none"> • Scripts to start the CAS Service (<code>cas-service.sh</code> for UNIX and <code>cas-service.bat</code> for Windows). • Scripts to run the CAS Server Command-line Utility (<code>cas-cmd.sh</code> for UNIX and <code>cas-cmd.bat</code> for Windows). • Scripts to run the Component Instance Manager Command-line Utility (<code>component-manager-cmd.sh</code> for UNIX and <code>component-manager-cmd.bat</code> for Windows). • Scripts to run the Record Store Command-line Utility (<code>recordstore-cmd.sh</code> and <code>recordstore-cmd.bat</code>). • Scripts to run the Web Crawler (<code>web-crawler.sh</code> for UNIX systems and <code>web-crawler.bat</code> for Windows). • Scripts to shut down the CAS Service (<code>cas-service-shutdown.sh</code> for UNIX and <code>cas-service-shutdown.bat</code> for Windows). • The executable that runs the CAS Service as a Windows service (<code>cas-service-wrapper.exe</code>).
<code>version\cms</code>	Server-side files for the FileNet Document and Image Services, Alfresco, and Sharepoint connectors.

Directory	Contents
<code>version\components</code>	<ul style="list-style-type: none"> The <code>RecordStore.war</code> file, which is the Record Store application. The <code>DeminsionIdValueManager.war</code> file, which is the Dimension ID Value Manager application.
<code>version\console</code>	The files in this directory support the CAS Console, which is not license to users of Oracle Endeca Information Discovery.
<code>version\doc</code>	<ul style="list-style-type: none"> The reference documentation (Javadoc) for the CAS Server, the Component Instance Manager, the Record Store, and the Web Crawler. The <i>Oracle Endeca Guided Search Licensing Guide</i>.
<code>version\doc\wsdl</code>	The Web Service (WSDL) files for the CAS Server, the Component Instance Manager, and the Record Store.
<code>version\java</code>	The JDK used to run the CAS components (except CAS Console, which runs in the Endeca Tools Service).
<code>version\lib</code>	<ul style="list-style-type: none"> Libraries for the CAS command-line utilities including: the CAS Server utility, the Component Instance Manager utility, the Record Store utility. Libraries for the CAS APIs including: the CAS Server API, the Component Instance Manager API, the Record Store API, and the CAS Extension API.
<code>version\lib\cas-dt</code>	The feature stored in this directory is not licensed with Oracle Endeca Information Discovery.
<code>version\lib\cas-server-plugins</code>	Libraries for CAS plug-ins including CMS connectors and custom extensions (if applicable).
<code>version\lib\oit-sx</code>	The CAS Document Conversion Module libraries.
<code>version\lib\recordstore - forge-adapter</code>	The feature stored in this directory is not licensed with Oracle Endeca Information Discovery.
<code>version\lib\web-crawler</code>	The Web Crawler libraries.
<code>version\sample</code>	Directories for the sample applications and sample extensions. The sample applications are documented in the <i>CAS Developer's Guide</i> and also in the <i>CAS Extension API Guide</i> .
<code>version\webapps</code>	The <code>cas.war</code> and <code>ComponentInstanceManager.war</code> files, which are the CAS Server and Component Instance Manager applications.

Directory	Contents
<code>version\workspace_template</code>	The template for the workspace directory that contains configuration files.
<code>workspace</code>	The working directory for the CAS Server and the Web Crawler.
<code>workspace\conf</code>	<ul style="list-style-type: none"> • The <code>commandline.properties</code> file, which contains the CAS Service settings necessary for the CAS command-line utilities to run. • The <code>default-log4j.properties</code> file, which should not be modified. • The <code>ProductConfig.xml</code> file, which is the licensing file for CAS components. • Three logging configuration files (<code>cas-service.log4j.properties</code> for the CAS Service, <code>recordstore-cmd.log4j.properties</code> for the Record Store, and <code>cas-cmd.log4j.properties</code> for the Command-line Utility). • The Jetty configuration files.
<code>workspace\conf\web-crawler\default</code>	The default configuration files for the Web Crawler, including the <code>log4j.properties</code> logging configuration file.
<code>workspace\conf\web-crawler\non-polite-crawl</code>	Sample crawl configuration files for non-polite crawls. As with the polite version, the settings in these files will override the default settings.
<code>workspace\conf\web-crawler\polite-crawl</code>	Sample crawl configuration files for polite crawls. The settings in <code>site.xml</code> will override the same property settings in the <code>default.xml</code> file, while the <code>crawl-urlfilter.txt</code> file will also override the default version of the file.
<code>workspace\logs</code>	The <code>cas-service.log</code> file, which contains the CAS Service log output, and includes log messages from all crawls managed by the CAS Server.
<code>workspace\output</code>	Default destination directory for the crawl output from the Web Crawler. The <code>output</code> directory is not present upon installation. It is created when the Web Crawler writes to output records for a crawl.
<code>workspace\state</code>	State files for the CAS Service components. State files can include Record Store instances, state directories for data source extension information, and state directories for manipulator extension information.



Note: There is no `logs` directory for the Web Crawler, because by default the Web Crawler sends its standard output to the console. However, you can modify the `log4j.properties` file to send the output to a file.



Chapter 3

Uninstalling the Content Acquisition System

This section contains the procedures for uninstalling the CAS.

[Uninstalling CAS on Windows](#)

[Uninstalling CAS on UNIX](#)

[Uninstalling a plug-in from CAS](#)

Uninstalling CAS on Windows

Follow these steps to uninstall the Content Acquisition System from your Windows machine.

The installer creates a timestamped backup of your workspace prior to uninstalling the Content Acquisition System

To uninstall the Content Acquisition System from a Windows machine:

1. From the Windows Control Panel, select **Add or Remove Programs**.
2. Select the **Endeca Content Acquisition System** from the list of installed software.
3. Click **Remove**.

Uninstalling CAS on UNIX

Follow these steps to uninstall the Content Acquisition System from a UNIX machine.

Before you begin the uninstall process, back up any files that you want to retain from the CAS directory.

To uninstall the CAS from a UNIX machine:

1. Shutdown CAS Service by navigating to `/usr/local/endeca/CAS/version/bin` and running the following command: `cas-service-shutdown.sh`.
2. In a command prompt, issue an `rm` command as in this example:

```
rm -rf endeca/CAS/version
```

Uninstalling a plug-in from CAS

Follow these steps to uninstall a plug-in from the Content Acquisition System.

Before you begin the uninstall process, back up any plug-in JAR files that you want to retain from the `CAS\version\lib\cas-server-plugins` directory.

To uninstall a plug-in from the CAS:

1. Stop Endeca CAS Service.
2. Navigate to `<install path>\CAS\version\lib\cas-server-plugins`.
3. Delete the *plugin-name*/ subdirectory for each plug-in.
For example: `<install path>\CAS\version\lib\cas-server-plugins\JDBCDataSourceExt`
4. Repeat the steps above as necessary for multiple plug-ins.
5. Start Endeca CAS Service.

Index

B

- bin directory
 - contents of 10
 - location of 10

C

- CAS Console
 - unsupported v
- CAS Server API
 - location of 11
- cas-server.log4j.properties file 12
- CAS Service
 - starting from inittab 7
- CMS connectors
 - licensing 2
 - licensing file 12
 - supported 2
- configuration files
 - for logging 12
 - for non-polite crawls 12
 - for polite crawls 12

D

- Dgidx
 - unsupported v

F

- Forge
 - unsupported v

I

- installing
 - on UNIX 6

- on Windows 5
- silently on UNIX 8
- silently on Windows 7

O

- overview of package 1

P

- package
 - contents of 10
 - directory structure of 10
 - overview of 1
- prerequisites 1

R

- recordstore-server.log4j.properties file 12

S

- system requirements 1

U

- uninstalling
 - a plug-in 13
 - on UNIX 13
 - on Windows 13
- unsupported features v

W

- workspace directory
 - location of 12
- WSDL files
 - location of 11