Oracle® Endeca Information Discovery Integrator

Integrator Acquisition System Installation Guide

Version 3.0.0 • March 2013

ORACLE

Copyright and disclaimer

Table of Contents

Preface

Oracle® Endeca Information Discovery Integrator provides a suite of products to load data from disparate source systems and store it for use in an Endeca Server data domain. The Integrator products include:

- Integrator ETL - Integrator ETL is a high-performance data integration platform that lets you extract source records from a variety of sources and sends that data to the Data Ingest Web Service, which in turn loads the records into the Oracle Endeca Server.

- Integrator Acquisition System - The Integrator Acquisition System, or IAS, is a set of components that crawl source data stored in a variety of formats including: file systems, delimited files, JDBC databases, and custom data sources. IAS transforms the data, if necessary, and outputs the data to an XML file or a Record Store instance that can be accessed by Integrator ETL for use in the Endeca Server.

- IKM SQL to Endeca Server - provides integration and loading modules that enable writing source data to an Endeca Server target within Oracle Data Integrator.

About this guide

This guide describes how to install the Integrator Acquisition System and set up IAS components after installation on Windows and UNIX.

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 domain.

Conventions used in this guide

The following conventions are used in this document.

Typographic conventions

The following table describes the typographic conventions used in this document.

Table 0.1: Typographic conventions

Typeface	Meaning
User Interface Elements	This formatting is used for graphical user interface elements such as pages, dialog boxes, buttons, and fields.
Code Sample	This formatting is used for sample code phrases within a paragraph.
<Variable Name>	This formatting is used for variable values, such as <install path>.
File Path	This formatting is used for file names and paths.

Symbol conventions

The following table describes symbol conventions used in this document.

Table 0.2: Symbol conventions

Symbol	Description	Example	Meaning
>	The right angle bracket, or greater-than sign, indicates menu item selections in a graphic user interface.	File > New > Project	From the File menu, choose New, then from the New submenu, choose Project.

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*.

Chapter 1

Before You Install

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

Overview of the Integrator Acquisition System

System requirements

IAS Document Conversion Module

Supported data sources

Recommended reading

Overview of the Integrator Acquisition System

The Integrator Acquisition System, or IAS, is a set of components that crawl source data stored in a variety of formats including: file systems, Content Management Systems, Web servers, and custom data sources. IAS transforms the data, if necessary, and outputs the data to an XML file or a Record Store that can be accessed by Integrator for use in the Endeca Server.

System requirements

The Integrator 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 hardware sizing.

Supported operating systems

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

Operating System	Description
Oracle Linux 5	Only the Red Hat Compatible Kernel is supported
Linux RHEL 5	• Red Hat Enterprise Linux Server (version 5 for x64) • Red Hat Enterprise Linux Advanced Platform (version 5 for x64) For best performance on Red Hat Linux version 5 (Server and Advanced), Endeca recommends the latest version of RHEL 5.

Operating System	Description
Windows 2008	Windows Server 2008 R2 Enterprise.

Hardware requirements for Linux and Windows on x64

Minimum hardware requirements:

- x64 processor, minimum 1.8 GHz
- 4 GB of RAM
- At least an 500 GB hard drive, depending on the size of your application data set

IAS Document Conversion Module

The IAS Document Conversion Module is installed and enabled by default as part of the IAS installation.

> **Note:** The IAS 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 IAS Document Conversion Module" in the *Integrator Acquisition System Developer's Guide*.

Supported data sources

For the full list of all supported data sources and information about configuring each, see the *Integrator Acquisition System Developer's Guide.*

Recommended reading

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

Getting Started Guide

The *Oracle Endeca Information Discovery Getting Started Guide* guides you through the process of exploring a sample Endeca Information Discovery application.

Integrator User's Guide

The *Oracle Endeca Information Discovery Integrator User's Guide* describes how to use the Oracle Endeca Information Discovery Integrator to add data into an Endeca data domain.

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 IAS documentation download. After installation, release notes are also available in the following location:

- Windows: `<install path>\Oracle\Endeca\IAS\<version>`

- UNIX: `<install path>/oracle/endeca/IAS/<version>`

Chapter 2

Installing the Integrator Acquisition System

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

Installing on Windows

Installing on UNIX

Web Crawler installation

Package contents and directory structure

Installing on Windows

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

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. 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**.

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 IAS 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 IAS on Windows:

1.   In your local environment, locate the Endeca IAS software that you downloaded from the Oracle Software Delivery Cloud.

2.   Double-click the installer file `ias-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 **Destination Folder** screen, select an installation location or accept the default location of `C:\Oracle\Endeca\IAS` and then click **Next**.

6.   In the **Endeca IAS Service Information** screen, specify the user name, password, and domain information for the user who will run the IAS Service and then click **Next**. (This is typically the `endeca` user you created in the previous procedure.)

7.   In the **IAS Server Information** screen, enter the IAS Server port, IAS Server shutdown port, and IAS Server host, or accept the default values of `8510`, `8516`, and the default server name.

8.   Click **Next**.

9.   In the **Completing the Setup Wizard** screen, click **Next**.

10. Click **Finish**.

The IAS Service starts automatically after installation.

Installing IAS silently on Windows

Creating a user for the Endeca services on Windows

Installing IAS silently on Windows

The silent installer is useful if you want to add the IAS installation to an install script, or push out the installation on multiple machines. The silent installer is not interactive.

The syntax to run the silent installation is the following:

```
start /wait ias-version-arch-OS.exe /s
/l=full path to the installer log file\installer.log
[IASHOST=localhost] [IASPORT=8510] [IASSTOPPORT=8516]
[TARGETDIR=C:\Oracle\Endeca\IAS] USERNAME=endeca
PASSWORD=endecapwd DOMAINNAME=domain
```

where the options have the following values:

Option	Description
ias-version-arch-OS.exe	Required. Specifies the installation program to run in silent mode.
/s	Required. Specifies that the IAS installer runs in silent mode.
/l	Optional. The logging option (/l=) provides useful information from silent installer failures. The specified directory must be a full path.
IASHOST	Optional. Specifies the host name of the machine where you want to install and run IAS. If the omitted, the default value is localhost.
IASPORT	Optional. Specifies the port on which IAS Service is listening. If the omitted, the default value is 8510.
IASSTOPPORT	Optional. Specifies the port to stop the IAS Service. If the omitted, the default value is 8516.
TARGETDIR	Optional. You can replace C:\Oracle\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.
USERNAME	Required. Species the user running the Endeca IAS Service. This is typically the specified as the endeca user.
PASSWORD	Required. Specifies the corresponding password for the USERNAME option.
DOMAINNAME	Required. Specifies the domain of the USERNAME option.

To launch the silent installer:

1. From a command prompt, navigate to the directory where you downloaded the installer.
2. Issue the start command and at a minimum specify the required options described above.

Example

```
start /wait F:\endeca-installers\ias-3.0.0-x86_64pc-win32.exe  /s /l
=F:\endeca-installers\logs\ias-install.log
IASHOST=hostname.us.oracle.com  TARGETDIR=F:\oracle\endeca\IAS USERNAME=JSMITH
PASSWORD=endecapwd DOMAINNAME=eastcoast
```

Installing on UNIX

This section provides instructions for installing the Endeca Integrator Acquisition System on UNIX.

Installing IAS on UNIX

The Endeca software is distributed as a self-extracting tar file and install script. The script is named `ias-version_arch-OS`.`sh`. The IAS can be installed to any location, but this procedure assumes IAS will be installed into `/usr/local`. IAS unpacks to approximately 400 MB. Because multiple versions may be stored, a destination in a large partition is recommended.

To install IAS on UNIX:

1.   Locate the installation file that you downloaded from the Oracle Software Delivery Cloud. This procedure assumes the location is in a `downloads` directory.

2.   Run the Endeca installation script with the following command:

     ```
     downloads/ias-version_arch-OS.sh --target /usr/local
     ```

3.   Enter the IAS Service port.

4.   Enter the IAS Service Shutdown port.

5.   Enter the fully qualified IAS Server host name.

Following installation:

- To start the IAS Service, navigate to `/usr/local/oracle/endeca/IAS/<version>/bin` and run `ias-service.sh`

Installing IAS silently on UNIX

Adding the Endeca IAS Service to inittab on UNIX

In a UNIX development environment, the Endeca IAS 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 IAS Server using `ias-service.sh`, located in `usr/local/oracle/endeca/IAS/<version>/bin`. You can write a script that calls `ias-service.sh` and is referenced in `inittab`.

When writing your script, Oracle recommends that you run the Endeca IAS Service as a user other than root.

This sample script (named `start_ias_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
IAS_ROOT=/usr/local/oracle/endeca/IAS/3.0.0
```

```
IAS_WORKSPACE=/usr/local/oracle/endeca/IAS/workspace

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

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

```
eias:2345:respawn:/usr/local/oracle/endeca/IAS/3.0.0/bin/start_ias_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 IAS silently on UNIX

The silent installer is useful if you want to add the IAS 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 on UNIX:

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:

     ```
     IAS port
     IAS shutdown port
     ```

     For example:

     ```
     8510
     8516
     ```

2.   From a command prompt, navigate to the location of the installer.

3.   Run the install script with the `--silent` flag, the `--target` flag (which specifies the absolute path of the target installation directory), any optional arguments, and the response file as an input.
     For example:

     ```
     ./ias-3.0.0-x86_64pc-linux.sh  --silent --target /usr/local <
     silent.txt
     ```

Following installation:

• To start the IAS Service, navigate to `IAS/<version>/bin` and run the following command: `ias-service.sh`

Web Crawler installation

The Endeca Web Crawler is installed as part of IAS.

The Web Crawler startup script is in `IAS\<version>\bin` (and in the equivalent path on UNIX). The Web Crawler configuration files are in `IAS\workspace\conf\web-crawler` (and in the equivalent path on UNIX). See the *Integrator Acquisition System Web Crawler Guide* for details and usage.

Package contents and directory structure

The `IAS` directory is the root directory for the Integrator Acquisition System.

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

```
IAS
   3.0.0
         bin
          ...
   workspace
```

The contents of the `IAS` directory are described here in detail.

Directory	Contents
version\bin	- Scripts to start the IAS Service (`ias-service.sh` for UNIX and `ias-service.bat` for Windows). - Scripts to run the IAS Server Command-line Utility (`ias-cmd.sh` for UNIX and `ias-cmd.bat` for Windows). - Scripts to run the Component Instance Manager Command-line Utility (`component-manager-cmd.sh` for UNIX and `component-manager-cmd.bat` for Windows). - Scripts to run the Record Store Command-line Utility (`recordstore-cmd.sh` and `recordstore-cmd.bat`). - Scripts to run the Web Crawler (`web-crawler.sh` for UNIX systems and `web-crawler.bat` for Windows). - Scripts to shut down the IAS Service (`ias-service-shutdown.sh` for UNIX and `ias-service-shutdown.bat` for Windows). - The executable that runs the IAS Service as a Windows service (`ias-service-wrapper.exe`).
version\cms	Server-side files for the SharePoint Object Model connector.
version\doc	- The reference documentation (Javadoc) for the IAS Server, the Component Instance Manager, the Record Store, the IAS Extension API, and the Web Crawler. - The *IAS Third-Party Software Usage and Licenses*.
version\doc\wsdl	The Web Service (WSDL) files for the IAS Server, the Component Instance Manager, and the Record Store.
version\java	The JDK used to run the IAS components.

Directory	Contents
version\lib	• Libraries for the IAS command-line utilities including: the IAS Server utility, the Component Instance Manager utility, the Record Store utility. • Libraries for the IAS APIs including: the IAS Server API, the Component Instance Manager API, the Record Store API, and the IAS Extension API.
version\lib\ias-server-plugins	Libraries for IAS plug-ins including CMS connectors and custom extensions (if applicable).
version\lib\oit-sx	The IAS Document Conversion Module libraries.
version\lib\web-crawler	The Web Crawler libraries.
version\sample	Directories for the sample applications and sample extensions. The sample applications are documented in the *IAS Developer's Guide* and also in the *IAS Extension API Guide*.
version\webapps	The `root.war` file, which is the IAS Server and Component Instance Manager as standalone applications.
version\workspace_template	The template for the workspace directory that contains configuration files.
workspace	The working directory for the IAS Server and the Web Crawler.
workspace\conf	• The `commandline.properties` file, which contains the IAS Service settings necessary for the IAS command-line utilities to run. • Three logging configuration files (`ias-service.log4j.properties` for the IAS Service, `recordstore-cmd.log4j.properties` for the Record Store, and `ias-cmd.log4j.properties` for the Command-line Utility). • The Jetty configuration files.
workspace\conf\web-crawler\default	The default configuration files for the Web Crawler, including the `log4j.properties` logging configuration file.
workspace\conf\web-crawler\non-polite-crawl	Sample crawl configuration files for non-polite crawls. As with the polite version, the settings in these files will override the default settings.

Directory	Contents
`workspace\conf\web-crawler\polite-crawl`	Sample crawl configuration files for polite crawls. The settings in `site.xml` will override the same property settings in the `default.xml` file, while the `crawl-urlfilter.txt` file will also override the default version of the file.
`workspace\logs`	The `ias-service.log` file, which contains the IAS Service log output, and includes log messages from all crawls managed by the IAS Server.
`workspace\output`	Default destination directory for the crawl output from the Web Crawler. The `output` directory is not present upon installation. It is created when the Web Crawler writes to output records for a crawl.
`workspace\state`	State files for the IAS 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 Integrator Acquisition System

This section contains the procedures for uninstalling the Integrator Acquisition System.

Uninstalling IAS on Windows

Uninstalling IAS on UNIX

Uninstalling IAS on Windows

The installer creates a time stamped backup of your workspace prior to uninstalling the Integrator Acquisition System.

To uninstall the Integrator Acquisition System from a Windows machine:

1. From the Windows Control Panel, select **Add or Remove Programs**.

2. Select the **Endeca Integrator Acquisition System** from the list of installed software.

3. Click **Uninstall**.

4. Click **Next** as prompted by the uninstall wizard.

Uninstalling IAS on UNIX

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

To uninstall IAS from a UNIX machine:

1. Shutdown the IAS Service by navigating to `/usr/local/oracle/endeca/IAS/<version>/bin` and running the following command: `ias-service-shutdown.sh`.

2. In a command prompt, issue an `rm` command as in this example:
   `rm -rf oracle/endeca/IAS/3.0.0`

Index