

Oracle® Endeca Information Discovery Integrator

Integrator Acquisition System Migration Guide

Version 3.1.0 • October 2013

Copyright and disclaimer

Copyright © 2003, 2014, 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.

Table of Contents

Copyright and disclaimer	2
Preface	4
About this guide	4
Who should use this guide	4
Conventions used in this guide	4
Contacting Oracle Customer Support	5
Chapter 1: Upgrading the Integrator Acquisition System	6
Recommended reading	6
Migration path	6
Migrating from 3.0.0 to 3.1.0	7
Backing up 3.0.0	7
Upgrading to 3.1.0	8
Restoring a 3.0.0 backup into 3.1.0	8
Upgrading client applications that use the IAS APIs	10
Updating an Integrator ETL graph and re-crawling data sources	10
Chapter 2: Required Changes	11
IAS API changes	11
IAS port number changes	12
CMS connector support	12
Chapter 3: Behavioral Changes	13
Windows installer changes	13
Platform support	13

Preface

Oracle® Endeca Information Discovery Integrator is a powerful visual data integration environment that includes:

The Information Acquisition System (IAS) for gathering content from delimited files, file systems, JDBC databases, and Web sites.

Integrator ETL, an out-of-the-box ETL purpose-built for incorporating data from a wide array of sources, including Oracle BI Server.

In addition, Oracle Endeca Web Acquisition Toolkit is a Web-based graphical ETL tool, sold as an add-on module. Text Enrichment and Text Enrichment with Sentiment Analysis are also sold as add-on modules. Connectivity to data is also available through Oracle Data Integrator (ODI).

About this guide

This guide describes how to upgrade the Integrator Acquisition System and describes the major changes between versions.

Who should use this guide

This guide is intended for data developers who are using IAS to crawl data sources, manipulate the records if necessary, and incorporate the records into Integrator ETL.

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.

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.

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

Upgrading the Integrator Acquisition System

This section contains upgrade instructions that include backing up, uninstalling, installing, restoring backups, configuring, and testing.

Recommended reading

Migration path

Migrating from 3.0.0 to 3.1.0

Recommended reading

In addition to reading this document, Oracle recommends that you read the following documents.

Release announcement

The Release Announcement provides a brief explanation of the new features that were added in IAS.

Release notes

The Release Notes provide information about new features, changed features, and bug fixes for this release. After installation, release notes are also available in the following location:

- `<install path>\IAS\<version>\README.txt`

You can also download the release notes (`README.txt`) from the Oracle Technology Network.

Migration path

IAS supports migrating from IAS 3.0.0 installed on Jetty Web server to IAS 3.1.0 on Jetty. In IAS 3.1.0, you cannot migrate from IAS running on Jetty Web server to IAS running on WebLogic Server.

Skipping major releases is not supported

You must migrate IAS from one major release to the next major release without skipping releases in between. For example, this means you can migrate from IAS 3.0.0 to 3.1.0.

Variables in paths

To simplify examples in this guide, some procedures may use `<install path>\IAS\<old version>` where `<old version>` represents that last major release.

Path syntax

For simplicity, the paths in this guide use Windows syntax. The UNIX paths are equivalent.

Migrating from 3.0.0 to 3.1.0

This section describes migrating from 3.0.0 to 3.1.0. It describes how to back up the old version, uninstall the old version, install the new version, restore the backup, and configure the new version.

Backing up 3.0.0

This procedure describes how to back up IAS configurations from 3.0.0 before upgrading. IAS configuration includes crawl configurations, Record Store configurations, Web crawler configuration, IAS extensions, and so on.

This procedure does not describe how to migrate the data in a Record Store instance or migrate the crawl history. At the end of the migration process, you must re-crawl data sources to repopulate Record Store instances and recreate crawl history.

To back up 3.0.0:

1. Retrieve and save your crawl configurations by doing the following:
 - (a) Open a Command Prompt and navigate to `<install path>\IAS\<old version>\bin`.
 - (b) Run `ias-cmd` and the `getAllCrawls` task. Specify a path to an XML file to store the crawl configurations.

The syntax for the `getAllCrawls` task is:

```
ias-cmd getAllCrawls -f <FileName.xml> -h <HostName> -p <PortNumber>
```

(You later import this XML file into version 3.1.0 to recreate your crawl configurations.)

2. If you created unmanaged Record Store instances or if you modified the configuration of a Record Store instance, retrieve and save your Record Store instance configuration by doing the following:
 - (a) Open a Command Prompt and navigate to `<install path>\IAS\<old version>\bin`.
 - (b) For each Record Store instance that is unmanaged or has a modified configuration, run `recordstore-cmd` and the `get-configuration` task. Specify a path to an XML file for the Record Store instance configuration.

The syntax for the `get-configuration` task is:

```
recordstore-cmd get-configuration -a <RecordStoreInstanceName>  
-f <FileName.xml> -h <HostName> -n -p <PortNumber>
```

3. Stop the Endeca IAS Service.
4. If you have any plugins installed in `<install path>\IAS\<old version>\lib\ias-server-plugins`, copy the directory containing the plugin JAR or JARs, to a location outside the IAS installation.



Note: Copy only your plugin directory. You do not need to back up `<install path>\IAS\<old version>\lib\ias-server-plugins\ias`.

5. If you have any custom Web Crawler plugins installed in `<install path>\IAS\<old version>\lib\web-crawler\plugins`, copy the directory containing the JAR and `plugin.xml`, to a location outside the IAS installation.
6. If you made any changes to `<install path>\IAS\<old version>\bin\ias-service`, or `ias-service-wrapper.conf`, copy the files to a location outside the IAS installation.

(This is typically necessary if you customized the JVM settings that IAS uses.)

7. Back up the `workspace` directory or leave it in place for the installation program to back up automatically.
 - On Windows, the IAS installation program automatically backs up and time stamps the workspace when you *uninstall*.
 - On UNIX, the IAS installation program automatically backs up and time stamps the workspace when you *install*.

The `workspace` directory contains IAS configuration files and state information such as `DocumentConversionFilters.xml`, logging configuration files, and so on.

Upgrading to 3.1.0

To upgrade, uninstall the older version, and install 3.1.0 as described in this topic.

Be sure you have backed up all IAS configuration according to the previous task before performing this task.

To upgrade to 3.1.0:

1. Uninstall the older version of IAS.
 - On Windows, go to the Windows Control Panel, select **Programs and Features**, select **Endeca Integrator Acquisition System**, and click **Remove**. (The Windows uninstall creates a time-stamped backup of `workspace` in IAS.)
 - On UNIX, run the following command:

```
rm -rf IAS/<version>
```

2. Install IAS 3.1.0. For details, see the *Integrator Acquisition System Installation Guide*.

Restoring a 3.0.0 backup into 3.1.0

After you install 3.1.0, you can restore the 3.0.0 backup into the new installation.

To restore a 3.0.0 backup into 3.1.0:

1. If you modified `<install path>\IAS\workspace\conf\DocumentConversionFilters.xml`, make the following changes:
 - (a) Open the backed up copy of `DocumentConversionFilters.xml` in a text editor.
 - (b) Copy your include and exclude filters from the backup.
 - (c) Open `DocumentConversionFilters.xml` of your 3.1.0 installation in a text editor and add the include and exclude filters.
 - (d) Save and close `DocumentConversionFilters.xml`.
2. If you modified `<install path>\IAS\workspace\conf\jetty.xml` in 3.0.0, make the following changes:
 - (a) Open the backed up copy of `jetty.xml` in a text editor.

- (b) Open the newer version of `jetty.xml` in a text editor.
 - (c) Copy your specific customizations into the 3.1.0 version of `jetty.xml`.
 - (d) Save and close the 3.1.0 version of `jetty.xml`.
3. If you had any IAS plugins installed in 3.0.0, copy the directory containing the plugin JAR or JARs from the backup location to `<install path>\IAS\<version>\lib\ias-server-plugins`.
 4. If a crawl configuration contains custom references to version numbers, for example, in paths to output files, then modify the paths as appropriate, and save and close the crawl configuration.
 5. Restart the Endeca IAS Service.
 6. If you created unmanaged Record Store instances or if you modified the configuration of a Record Store instance in 3.0.0, import the Record Store instance configuration into 3.1.0 by doing the following:
 - (a) Open a Command Prompt and navigate to `<install path>\IAS\<old version>\bin`.
 - (b) Recreate each Record Store instance by running `component-manager-cmd` and the `create-component` task for each Record Store instance. The syntax for the `create-component` task is:


```
component-manager-cmd create-component -n <RecordStoreInstanceName>
-t <RecordStore> -h <HostName> -p <PortNumber>
```
 - (c) For each Record Store instance, run `recordstore-cmd` and the `set-configuration` task. Specify the path to an XML file for the Record Store instance configuration.

The syntax for the `set-configuration` task is:

```
recordstore-cmd set-configuration -a <RecordStoreInstanceName>
-f <FileName.xml> -h <HostName> -n -p <PortNumber>
```
 7. Import the backed up crawl configurations into 3.1.0 by doing the following:
 - (a) Open a Command Prompt and navigate to `<install path>\IAS\<old version>\bin`.
 - (b) Run `ias-cmd` and the `createCrawls` task. Specify the path to the XML file you created in [Backing up 3.0.0 on page 7](#).

The syntax for the `createCrawls` task is:

```
ias-cmd createCrawls -f <FileName.xml> -h <HostName> -p <PortNumber>
```

You will be prompted for the password of any connector or data source that requires a password.
 8. If you used any custom Web Crawler plugins installed in 3.0.0, copy the directory containing the plugin JAR and `plugin.xml` from the backup location to `<install path>\IAS\3.1.0\lib\web-crawler\plugins`.
 9. If you used the Web Crawler in 3.0.0, do the following:
 - (a) Copy the older version of `default.xml` and `site.xml` from the workspace backup into the `3.1.0 workspace\conf\web-crawler` locations.
 - (b) Open `default.xml` in a text editor and modify the path in the `plugin.folders` property to the 3.1.0 directory structure of `<install path>\IAS\3.1.0\lib\web-crawler\plugins`.

Upgrading client applications that use the IAS APIs

If you are using the IAS WSDL client stubs provided with the Integrator Acquisition System, make any changes listed in the Required Changes chapter.

If you are using a WSDL tool to generate stubs, see "Generating client stubs for the IAS Web Services" in the *Integrator Acquisition System API Guide*.

Updating an Integrator ETL graph and re-crawling data sources

You must also update your Integrator ETL graphs to reflect the new installation of IAS.

To update an Integrator ETL graph:

1. In your Integrator ETL graph, modify the Record Store Reader components that read from a Record Store instance to use the new installation paths.

Re-crawl data sources as necessary. The back up task includes backing up the Record Store instances, so a baseline update may not be necessary unless the application requires an update.



Chapter 2

Required Changes

You must make the changes specified in this section if the changes apply to your application.

[IAS API changes](#)

[IAS port number changes](#)

[CMS connector support](#)

IAS API changes

For detailed usage information about the API changes listed below, also see the *Integrator Acquisition System API Guide*.

IAS Server API

A new `ServiceAddress` class has been added to represent the explicit address of the services provided by IAS.

The class is composed of three fields:

- `host`: host name or IP address of the machine running IAS.
- `port`: opened port for IAS.
- `contextPath`: the context path of IAS.

For example, in IAS 3.0.0, you connected to the IAS Server using the sample following code:

```
// Connect to the IAS Server.
IasCrawlerLocator locator = IasCrawlerLocator.create("localhost", 8510);
IasCrawler crawler = locator.getService();
```

For example, in IAS 3.1.0, you can connect to the IAS Server using the following sample code:

```
ServiceAddress address = new ServiceAddress("localhost", 8401, contextPath);
IasCrawlerLocator locator = IasCrawlerLocator.create(address);
IasCrawler crawler = locator.getService();
```

Component Instance Manager API

A new `ServiceAddress` class has been added to represent the explicit address of the services provided by IAS.

For example, in IAS 3.1.0, you can connect to the Component Instance Manager using the following sample code:

```
ServiceAddress address = new ServiceAddress("localhost", 8401, contextPath);
ComponentInstanceManagerLocator locator = ComponentInstanceManagerLocator.create(address);
ComponentInstanceManager cim = locator.getService();
```

Record Store API

A new `ServiceAddress` class has been added to represent the explicit address of the services provided by IAS.

For example, in IAS 3.1.0, you can connect to the Record Store using the following sample code:

```
ServiceAddress address = new ServiceAddress("localhost", 8401, contextPath);
RecordStoreLocator locator = RecordStoreLocator.create(address, "MyCrawl");
RecordStore rs = locator.getService();
```

IAS port number changes

The range of port numbers that IAS uses has changed to be more in line with other Oracle Endeca products. The IAS installation script prompts for the new port values and allows you to accept the defaults. However, if you have port numbers stored in the `outputConfig` element of your crawl configuration XML, you must update the values for 3.1.0.

Port Name	3.0.0 Default Value	3.1.0 Default Value
IAS port	8510	8401
IAS SSL port	8505	8402
IAS shutdown port (applies only to Jetty installations)	8516	8403

CMS connector support

IAS 3.1.0 removes support for the remaining CMS connectors. In particular, the following connectors are no longer supported:

- Documentum Content Server
- Microsoft SharePoint Object Model
- Microsoft SharePoint Web Services



Chapter 3

Behavioral Changes

This section describes changes that do not require action on the developer's part, but will have an effect on how an application behaves after you upgrade.

[Windows installer changes](#)

[Platform support](#)

Windows installer changes

The Windows installation program has been replaced with an installation script named `ias-setup`. It operates the same as the UNIX installation script, and it enables future releases of IAS to be incorporated into a consolidated installation program. For details, see the *Integrator Acquisition System Installation Guide*.

Platform support

This release of IAS adds support for Oracle Linux Release 6 for x64. For details, see the *Integrator Acquisition System Installation Guide*.