

Oracle® Endeca Information Discovery Integrator

Integrator Migration Guide

Version 3.0.0.2 • July 2013

Copyright and disclaimer

Copyright © 2003, 2013, 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 to Oracle Endeca Information Discovery Integrator 3.0.x	6
Required reading	6
Upgrading Integrator	6
Running the upgraded graphs	7
Chapter 2: Required Changes in Integrator	8
Changes to Oracle Endeca Information Discovery packaging	8
Integrator changes required by Endeca Server changes	8
Integrator changes required by Integrator Acquisition System (IAS)	10

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 helps you upgrade your Oracle Endeca Information Discovery Integrator implementation by describing the major changes between versions 2.4.x and 3.0.x.

Who should use this guide

This guide is intended for system administrators and developers who are upgrading Oracle Endeca Information Discovery Integrator on Windows or Linux.

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

Upgrading to Oracle Endeca Information Discovery Integrator 3.0.x

This section provides instructions for upgrading to Oracle Endeca Information Discovery Integrator 3.0.x. Before you start the upgrade processes, check the remaining sections of this guide to learn about the changes that will affect you during or after an upgrade.

Required reading

Upgrading Integrator

Required reading

In addition to reading this document, Oracle recommends that you read the following documents for important information about the release.

Release Announcement

The Release Announcement outlines the new features that were added in version 3.0.x.

Release Notes

The Release Notes for Integrator provide information about known issues and bug fixes for this release.

Installation Guide

The *Oracle Endeca Information Discovery Integrator Installation Guide* provides an overview of Integrator and includes information about configuration scenarios. It contains installation instructions for Integrator, and information on how to verify your installation.

Upgrading Integrator

To upgrade Integrator, uninstall the existing version and install the new one. You can still use your existing workspace after you install Integrator version 3.0.x.

If you have custom plug-ins deployed to Integrator 2.4.x, you should back up the plug-ins before beginning the Integrator upgrade. You can reload them after the upgrade.

To upgrade Integrator:

1. Uninstall version 2.4.x. For information, see the *Oracle Endeca Information Discovery Installation Guide*, version 2.4.x.
After you uninstall, the workspace you used for creating your Integrator graphs should remain in place on your machine.
2. Install Integrator version 3.0.x. For information, see the *Oracle Endeca Information Discovery Integrator Installation Guide*, version 3.0.x.
3. Start Integrator and point it to use the workspace you used for your version 2.4.x graphs.

Now you have upgraded to the new version of Integrator. However, do not run your existing graphs until you have made changes to them that are required for Oracle Endeca Information Discovery, version 3.0.x.

After the upgrade, you can proceed to making changes to your existing Integrator graphs, as described in [Required Changes in Integrator on page 7](#).

If you are reloading a custom plug-in, keep in mind that the installation path has changed. For example, on Windows, the 2.4.x default installation path has changed from:

```
C:\Oracle\Endeca\Discovery\Integrator\2.4.0
```

to this 3.0.x default path:

```
C:\Oracle\Endeca\Discovery\Integrator\3.0.0
```

After the upgrade, Integrator still refers to the `customcomponents.xml` at the previous location. Therefore, you must update the reference to the new 3.0.x location.

[Running the upgraded graphs](#)

Running the upgraded graphs

After you have made changes to the existing graphs, you can run them in Integrator.

After running the graphs and troubleshooting any errors, you can upgrade Studio.



Chapter 2

Required Changes in Integrator

The following changes are required for Integrator.

Changes to Oracle Endeca Information Discovery packaging

Integrator changes required by Endeca Server changes

Integrator changes required by Integrator Acquisition System (IAS)

Changes to Oracle Endeca Information Discovery packaging

In this release, the Oracle Endeca Information Discovery packaging has changed.

Previously, the Oracle Endeca Information Discovery package consisted of Studio and Integrator.

In this release, Information Discovery Studio and Information Discovery Integrator are released as two separate products with separate documentation sets.

Addition to the Oracle Endeca Information Discovery Integrator package

Starting with release 3.0, the Oracle Endeca Information Discovery Integrator product contains both Integrator and the Integrator Acquisition System, or IAS.

Integrator changes required by Endeca Server changes

This section describes changes required in Integrator due to changes in Endeca Server.

PDR changes

The default value of the `mdex-property_IsSingleAssign` in the PDR has changed from `false` to `true`. If you attempt to run graphs from Version 2.4 or earlier and did not explicitly set the value of this property, errors may occur during data processing if you load multiple values into a property that uses the default of the `IsSingleAssign` property.

Property modifications

Several properties on components have been renamed or added, or may require new values.

- **Endeca Server Port**

The default Endeca Server port has changed in Version 7.5.1. The SSL port is 7002. The non-SSL port is 7001. Update the value of this property with the correct new port for your server.

- **Data Store Name**

This property has been renamed to **Data Domain Name**. The **Data Store Name** property has been moved into the Custom Properties set. You can copy the value of the **Data Store Name** property and paste it in the **Data Domain Name** property. Alternatively, in the source XML of the graph, you can change the name of the property from Data Store Name to Data Domain Name.

- **Endeca Server Context Root**

This property has been added in Version 3.0. The default Endeca Server context root is `/endeca-server`. You must enter a value for this property on all components that include it in all graphs. If you do not enter a value for this property in a component in a graph, the graph will fail with a missing property error.

These changes affect the following components:

- Add KVPs
- Add Managed Values
- Bulk Add/Replace Records
- Export Config
- Import Config
- Transaction Run Graph

Reset Data Store renamed

The **Reset Data Store** component has been renamed to **Reset Data Domain**. The change to the name of the component does not modify its functionality.

You can modify the name of the component in the source XML of the graph. Change `ENDECA_RESET_DATA_STORE` to `ENDECA_RESET_DATA_DOMAIN`.

Note that the property changes described in "Property modifications" also affect the Reset Data Domain component.

Add/Update Records component renamed and modified

The **Add/Update Records** component has been renamed to **Merge Records**.

A new option, **Disable Updates**, has been added to this component. When this flag is set to TRUE (in other words, when the box is checked), if an input record matches an existing Endeca record, the operation fails. As a result, the complete batch fails.

Note that in the default PDR configuration, the `IsSingleAssign` property now defaults to `true`. In other words, properties do not default to multi-assign. If you want a property to support multi-assign, you must configure the PDR before loading records.

You can modify the name of the component in the source XML of the graph. Change `ENDECA_ADD_OR_UPDATE_RECORDS` to `ENDECA_MERGE_RECORDS`.

Note that the property changes described in "Property modifications" also affect the **Merge Records** component.

New components

The following Information Discovery components have been added to Integrator:

- **Delete Records**

Use this component to delete complete records from the data domain. Each record input to this component can result in multiple records being deleted from the data domain.

- **Modify Records**

Use this component to:

- Add new assignments to a record
- Modify the value of existing assignments on existing records
- Delete assignments from an existing record.

Each record input to this component can modify multiple records in the data domain.

Re-implementing graphs

You may want to consider re-implementing your graphs to take advantage of the new and modified components.

Web Services updates

Endeca Server Version 7.5.1 includes changes to Web services.

- The Control Web Service has been removed.
- The Cluster and Manage Web Services have been added.
- The version number of the remaining Web services has been changed from 1 to 2.

Any Endeca Server Web services you use in your implementation should be reviewed and updated to current standards.

In particular, backup and restore no longer use the data store snapshot Web services operation. Back and restore now use the `exportDataDomain` and `importDataDomain` operations of the Manage Web Service.

Integrator changes required by Integrator Acquisition System (IAS)

In this version, Integrator Acquisition System (IAS) has replaced Content Acquisition System.

As a result of this change, the following property names have changed:

- **CAS Service Host** is now **IAS Service Host**.
- **CAS Service Port** is now **IAS Service**.
- **CAS Client ID** is now **IAS Client ID**.
- **CAS Read Type** is now **IAS Read Type**.

When you open a graph created before the upgrade, all fields that begin "CAS" have been moved into the Custom Properties set. You can copy the value of each property into the associated field that begins "IAS".

Alternatively, in the source XML of the graph, you can change the name of the property to match the new name. In other words, you can change `casHost` to `iasHost` in the graph XML.