

Oracle® Endeca Information Discovery Integrator

Integrator ETL Migration Guide

Version 3.1.1 • December 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 to Oracle Endeca Information Discovery Integrator ETL 3.1.x	6
Required reading	6
Upgrading Integrator ETL	6
Chapter 2: Required Changes in Integrator ETL	8
Change to product name	8
Integrator ETL changes required by Endeca Server changes	8
Required changes to Integrator ETL graphs	8
Chapter 3: Behavioral changes in Integrator ETL	10

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 helps you upgrade your Oracle Endeca Information Discovery Integrator ETL implementation by describing the major changes between versions 3.0.x and 3.1.x.

Who should use this guide

This guide is intended for system administrators and developers who are upgrading Oracle Endeca Information Discovery Integrator ETL 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.

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 to Oracle Endeca Information Discovery Integrator ETL 3.1.x

This section provides instructions for upgrading to Oracle Endeca Information Discovery Integrator ETL 3.1.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 ETL

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 3.1.x.

Release Notes

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

Installation Guide

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

Upgrading Integrator ETL

To upgrade Integrator ETL, uninstall the existing version and install the new one. You can still use your existing workspace after you install Integrator ETL version 3.1.x.

If you have custom plugins deployed to Integrator ETL 3.1.x, you should back up the plugins before beginning the Integrator ETL upgrade. You can reload them after the upgrade.

To upgrade Integrator ETL:

1. Uninstall version 3.0.x. For information, see the *Oracle Endeca Information Discovery Integrator Installation Guide*, version 3.0.x.
After you uninstall, the workspace you used for creating your Integrator ETL graphs should remain in place on your machine.
2. Install Integrator ETL version 3.1.x. For information, see the *Oracle Endeca Information Discovery Integrator ETL Installation Guide*, version 3.1.x.
3. Start Integrator ETL and point it to use the workspace you used for your version 3.0.x graphs.

Now you have upgraded to the new version of Integrator ETL.

Changes between Integrator 3.0.x and Integrator ETL 3.1.x are extensive, as are changes between Endeca Server Version 7.5.x and 7.6.x. Therefore, Oracle recommends that you re-implement your graphs rather than attempting to migrate them.

If you are reloading a custom plugin, keep in mind that the installation path has changed. For example, on Windows, the 3.0.x default installation path has changed from:

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

to this 3.1.x default path:

```
C:\Oracle\Endeca\Discovery\3.1.0\Integrator
```

After the upgrade, Integrator ETL still refers to the `customcomponents.xml` at the previous location. Therefore, you must update the reference to the new 3.1.x location.



Chapter 2

Required Changes in Integrator ETL

The following changes are required for Integrator ETL.

[Change to product name](#)

[Integrator ETL changes required by Endeca Server changes](#)

[Required changes to Integrator ETL graphs](#)

Change to product name

In this version, Oracle Endeca Information Discovery Integrator has been renamed to Oracle Endeca Information Discovery Integrator ETL.

Integrator ETL changes required by Endeca Server changes

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

Web service version changes

In Endeca Server Version 7.6, the version number of some Web services have been changed. For details about the current version of each Web service, see "List of Web Services and their versions" in the *Oracle Endeca Server Installation Guide*.

The functions available in specific Web services have been changed. For details about these changes, see "Endeca Server Interfaces" in the *Oracle Endeca Server Migration Guide*.

Any Endeca Server Web services you use in your implementation should be reviewed and updated to the new version or to use the new functionality.

Required changes to Integrator ETL graphs

This topic describes required changes to Integrator ETL graph implementations.

Loading to a collection is now required

All data must now be loaded to a collection. The following components now include a **Collection key** field:

- Bulk Add/Update Records
- Merge Records
- Delete Records

You must specify the **Collection key** and **Spec attribute** in these components in your graphs.

Re-implement Load Data from Oracle Business Intelligence Server graphs

The graphs used to load data from Oracle Business Intelligence Server have changed extensively in this release. You should recreate your graphs using the Load Data from OBI Server wizard rather than trying to migrate existing graphs.

Text Tagger Regex multiple input patterns

The **Text Tagger Regex** component now supports multiple input patterns. The **Input Pattern** field has been removed from the component. The component now requires an input file. For all details, see "Using the Text Tagger Regex Component" in the *Oracle Endeca Information Discovery Integrator ETL User's Guide*.

Text Enrichment no longer requires a license file

The **Salience license file** property has been removed from the **Text Enrichment** component.



Chapter 3

Behavioral changes in Integrator ETL

The changes listed here do not require any action on your part, but will affect how your application behaves after you upgrade.

In Version 3.1.1, you must download the Eclipse RSE plugin before installing Integrator ETL, and must specify the path to the RSE plugin package during the installation. In Version 3.1.0, the installation script automatically connected to the Internet and downloaded the RSE plugin. This change was implemented to accommodate customers in restricted environments where active Internet connections were not available, as well as customers whose Internet connections are inconsistent.