

Endeca® Content Acquisition System

Migration Guide

Version 3.1.2 • April 2013



Contents

Preface.....	7
About this guide.....	7
Who should use this guide.....	7
Conventions used in this guide.....	7
Contacting Oracle Support.....	8
Chapter 1: Upgrading the Content Acquisition System.....	9
Recommended reading.....	9
Migration path.....	9
Upgrading CAS 3.1.1 to 3.1.2.....	10
Backing up CAS 3.1.1	10
Upgrading to CAS 3.1.2.....	11
Restoring a CAS 3.1.1 backup into CAS 3.1.2	12
Upgrading CAS client applications that use the CAS APIs.....	13
Updating the CAS Deployment Template Component.....	13
Updating the Forge pipeline and re-crawling data sources.....	13
Chapter 2: Required Changes.....	15
Chapter 3: Behavioral Changes.....	17
Changes to record store configuration properties.....	17

Copyright and disclaimer

Copyright © 2003, 2013, Oracle and/or its affiliates. All rights reserved.

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.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group.

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.

For information about Oracle's commitment to accessibility, visit the Oracle Accessibility Program website at <http://www.oracle.com/pls/topic/lookup?ctx=acc&id=docacc>.

Oracle customers have access to electronic support through My Oracle Support. For information, visit <http://www.oracle.com/pls/topic/lookup?ctx=acc&id=info> or visit <http://www.oracle.com/pls/topic/lookup?ctx=acc&id=trs> if you are hearing impaired.

Preface

The Oracle Endeca Commerce solution enables your company to deliver a personalized, consistent customer buying experience across all channels — online, in-store, mobile, or social. Whenever and wherever customers engage with your business, the Oracle Endeca Commerce solution delivers, analyzes, and targets just the right content to just the right customer to encourage clicks and drive business results.

Oracle Endeca Commerce is the most effective way for your customers to dynamically explore your storefront and find relevant and desired items quickly. An industry-leading faceted search and Guided Navigation solution, Oracle Endeca Commerce enables businesses to help guide and influence customers in each step of their search experience. At the core of Oracle Endeca Commerce is the MDEX Engine™, a hybrid search-analytical database specifically designed for high-performance exploration and discovery. The Endeca Content Acquisition System provides a set of extensible mechanisms to bring both structured data and unstructured content into the MDEX Engine from a variety of source systems. Endeca Assembler dynamically assembles content from any resource and seamlessly combines it with results from the MDEX Engine.

Oracle Endeca Experience Manager is a single, flexible solution that enables you to create, deliver, and manage content-rich, cross-channel customer experiences. It also enables non-technical business users to deliver targeted, user-centric online experiences in a scalable way — creating always-relevant customer interactions that increase conversion rates and accelerate cross-channel sales. Non-technical users can control how, where, when, and what type of content is presented in response to any search, category selection, or facet refinement.

These components — along with additional modules for SEO, Social, and Mobile channel support — make up the core of Oracle Endeca Experience Manager, a customer experience management platform focused on delivering the most relevant, targeted, and optimized experience for every customer, at every step, across all customer touch points.

About this guide

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

Who should use this guide

This guide is intended for application developers who are building applications using the Endeca Content Acquisition System and are responsible for migration tasks.

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 Support

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

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

Chapter 1

Upgrading the Content Acquisition System

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

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 CAS.

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:

- Windows: <install path>\CAS\<version>\README.txt
- UNIX: <install path>/CAS/<version>/README.txt

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

CAS Quick Start Guide

The *Endeca CAS Quick Start Guide* provides high-level procedures to guide you through setting up and running CAS.

Migration path

CAS supports migrating from 3.1.1 to 3.1.2.

Skipping major releases is not supported

You must migrate the Content Acquisition System from one major release to the next major release without skipping releases in between. Upgrades from CAS 1.x, 2.x, 3.0.x are not supported and not documented in this guide. To migrate from any previous releases, see the *CAS Migration Guide* for that release.

Variables in paths

To simplify path examples in this guide, some procedures use the variable <old version> to mean 3.1.1.

Upgrading CAS 3.1.1 to 3.1.2

This section describes migrating from CAS 3.1.1 to version 3.1.2. 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 CAS 3.1.1

This procedure describes how to back up CAS configurations from 3.1.1 before upgrading. CAS configuration includes crawl configurations, Record Store configurations, Web crawler configuration, CAS extensions, and so on.

To back up CAS:

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

The syntax for the `getAllCrawls` task is:

```
cas-cmd getAllCrawls [-f FileName.xml] [-h HostName] [-p PortNumber]
```

(You later import this XML file 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 export these Record Store instances including configurations and contents by doing the following:

- a) Open a Command Prompt and navigate to <install path>\CAS\<old version>\bin on Windows or <install path>/CAS/<old version>/bin on UNIX.
 - b) For each Record Store instance, 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 ConfigurationFileName.xml [-h HostName] [-n] [-p PortNumber]
```

- c) To migrate existing 3.1.1 data to 3.1.2, run `recordstore-cmd` and the `read-baseline` task for each Record Store instance. Specify a path to an XML file for the Record Store instance contents.

The syntax for the `read-baseline` task is:

```
recordstore-cmd read-baseline -a RecordStoreInstanceName -f RecordStoreOutputFileName.xml [-h HostName] [-p PortNumber]
```

3. Stop the Endeca CAS Service.
4. If you have any CAS plug-ins installed in <install path>/CAS/<old version>/lib/cas-server-plugins, copy the directory containing the plug-in JAR or JARs, to a location outside the CAS installation.



Note: Copy only your plug-in directory. You do not need to back up <install path>/CAS/<old version>/lib/cas-server-plugins/cas or <install path>/CAS/<old version>/lib/cas-server-plugins/entropysoft

5. If you have any custom Web Crawler plug-ins installed in <install path>/CAS/<old version>/lib/web-crawler/plugins, copy the directory containing the JAR and plugin.xml, to a location outside the CAS installation.
6. If you have any CMS connectors that required you to add additional JAR files as part of connector set up to <install path>/CAS/<old version>/lib/cas-server-plugins/entropysoft, copy the JAR files to a location outside the CAS installation.
7. If you made any changes to <install path>/CAS/<old version>/bin/cas-service, or cas-service-wrapper.conf, copy the files to a location outside the CAS installation.
This is typically necessary if you customized the JVM settings that CAS uses.
8. Back up the workspace directory or leave it in place for the installation program to back up automatically.
 - On Windows, the CAS installation program automatically backs up and time stamps the workspace when you uninstall.
 - On UNIX, the CAS installation program automatically backs up and time stamps the workspace when you install.

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

Upgrading to CAS 3.1.2

To upgrade, uninstall the older version, and install 3.1.2 as described in this topic.

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

To upgrade to CAS:

1. If you haven't already, upgrade to the latest versions of Oracle Endeca Platform Services and Oracle Endeca Tools and Frameworks. For details, see the *Platform Services Migration Guide* and the *Oracle Endeca Tools and Frameworks Migration Guide*.
2. Uninstall the older version of the Content Acquisition System.
 - On Windows, go to the Windows Control Panel, select **Programs and Features**, select **Endeca Content Acquisition System** and click **Remove**. (The Windows uninstall creates a time-stamped backup of workspace in CAS.)
 - On UNIX, run the following command to remove CAS Console:


```
CAS/<version>/console/configure_cas_console.sh --uninstall_console
```

 and then run the following rm command:


```
rm -rf CAS/<version>
```
3. Install CAS 3.1.2. For details, see the *CAS Installation Guide*.

Restoring a CAS 3.1.1 backup into CAS 3.1.2

To restore a CAS backup into CAS 3.1.2:

1. If you modified <install path>\CAS\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 CAS 3.1.2 installation in a text editor and add the include and exclude filters.
 - d) Save and close DocumentConversionFilters.xml.
2. If you modified <install path>\CAS\workspace\conf\jetty.xml in 3.1.1, 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 3.1.2 version of jetty.xml.
 - d) Save and close the 3.1.2 version of jetty.xml.
3. If you had any CAS plug-ins installed in 3.1.1, copy the directory containing the plug-in JAR or JARs from the backup location to <install path>/CAS/<version>/lib/cas-server-plugins.
4. If a crawl configuration contains custom references to CAS version numbers, for example, in paths to output files, then modify the paths as appropriate, and save and close the crawl configuration.
5. If you had any CMS connectors in 3.1.1 that required you to add additional JAR files as part of connector set up to <install path>/CAS/<old version>/lib/cas-server-plugins/entropysoft, copy the JAR files from the backup location to <install path>/CAS/3.1.2/lib/cas-server-plugins/entropysoft
6. Restart the Endeca CAS Service.
7. If you created unmanaged Record Store instances or if you modified the configuration of a Record Store instance in 3.1.1, import these Record Store instances including their configurations and contents into 3.1.2 by doing the following:
 - a) Open a Command Prompt and navigate to <install path>\CAS\<old version>\bin on Windows or <install path>/CAS/<old version>/bin on UNIX.
 - b) Re-create 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 created during backup for the Record Store instance configuration. The syntax for the set-configuration task is:

```
recordstore-cmd set-configuration -a RecordStoreInstanceName
-f ConfigurationFileName.xml [-h HostName] [-n] [-p PortNumber]
```
 - d) To restore existing 3.1.1 data to 3.1.2, run recordstore-cmd and the write task for each Record Store instance. Specify the path to an XML file with the exported contents of the Record Store. The syntax for the write task is:

```
recordstore-cmd.bat write -a RecordStoreInstanceName
-f RecordStoreOutputFileName.xml [-h HostName] [-p PortNumber]
```

8. Import the backed up crawl configurations into 3.1.2 by doing the following:
 - a) Open a Command Prompt and navigate to <install path>\CAS\<old version>\bin on Windows or <install path>/CAS/<old version>/bin on UNIX.
 - b) Run cas-cmd and the createCrawls task. Specify the path to the XML file you created in [Backing up CAS 3.1.1](#) on page 10.

The syntax for the createCrawls task is:

```
cas-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.

9. If you used any custom Web Crawler plug-ins installed in 3.1.1, copy the directory containing the plug-in JAR and plugin.xml from the backup location to <install path>/CAS/3.1.2/lib/web-crawler/plugins.
10. If you used the Web Crawler in 3.1.1, do the following:
 - a) Copy the older version of default.xml and site.xml from the workspace backup into the 3.1.1 workspace\conf\web-crawler locations.
 - b) Open default.xml in a text editor and modify the path in the plugin.folders property to the 3.1.1 directory structure of <install path>/CAS/3.1.2/lib/web-crawler/plugins.

Upgrading CAS client applications that use the CAS APIs

If you are using the CAS WSDL client stubs provided with the Content 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 CAS Web Services" in the *Endeca CAS API Guide*.

Updating the CAS Deployment Template Component

For *existing applications*, you must update the application with the latest version of the CAS Deployment Template Component. This update is necessary if you have already deployed an application and have CAS integrated into your Deployment Template environment.

To update the CAS Deployment Template Component:

```
Copy casStubs.jar from <install path>\CAS\<version>\lib\cas-dt into the
<installpath>/<appDir>/config/lib/java directory of each application that uses CAS.
```

Updating the Forge pipeline and re-crawling data sources

This topic describes how to update your Forge pipeline and then re-crawl data sources, and process records with a baseline update when migrating from CAS 3.1.1.

To configure and test the upgrade:

In your Forge pipeline, modify the record adapters that read from a Record Store instance to use the newer JAR files. In particular, in the **Classpath** field of Java Properties, specify the path to <install path>/CAS/3.1.2/lib/recordstore-forge-adapter/recordstore-forge-adapter-3.1.2.jar.

You can re-crawl data sources as necessary. The back up task for CAS 3.1.1 includes the Record Store instances, so a baseline update may not be necessary unless the application requires an update.

Chapter 2

Required Changes

There are no required changes for this release.

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.

Changes to record store configuration properties

Two properties are no longer used in the configuration properties of a record store instance.

maxDataFileSize removed

The `maxDataFileSize` configuration property has been removed. In previous releases, this property determined the maximum size of the data file for a record store instance. In 3.1.2, the record store has been optimized to use less disk space. Now, each generation of a record store instance has a separate file, and this generation file only stores records that have changed from previous generations. Moreover, the cleaning process further optimizes the record store by periodically merging old generations into a single file and removing unnecessary records.

duplicateRecordCompressionEnabled has been deprecated.

In previous releases, the `duplicateRecordCompressionEnabled` configuration property specified whether or not to store new versions of records whose change properties have not changed. In 3.1.2, new versions of records whose change properties have not changed are never stored, so the `duplicateRecordCompressionEnabled` property is ignored.

