Endeca® Content Acquisition System

Migration Guide

Version 3.1.1 • December 2012

ORACLE

ENDECA

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
Migrating CAS 3.0.x to 3.1.1.....10
Backing up CAS 3.0.x 10
Upgrading to CAS 3.1.1.....11
Restoring a CAS 3.0.x backup into CAS 3.1.1 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 in 3.1.1 15
System requirements changes.....15

Chapter 3: Behavioral Changes in 3.1.1.....17
Crawls fail if no records are output.....17
The configure_cas_console.sh script has been renamed and moved.....17
CONFIRMPASSWORD is not required for the silent installer on Windows.....17
Log file encoding is UTF-8.....17

Copyright and disclaimer

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 any 3.0.x point release to 3.1.1.

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. For example, this means you can migrate from CAS 3.0.2 to 3.1.1. but you cannot skip from 2.2.1 to 3.1.1. Upgrades from CAS 1.0, 1.1, 1.2, 2.0, 2.1, and 2.2 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 examples in this guide, some procedures may use `<install path>\CAS\<old version>\` to mean 3.0.0, 3.0.1, or 3.0.2 depending on which release you are upgrading from.

3.1.1 release numbering

The last general availability release of CAS was 3.0.2. There is no CAS 3.1.0 release.

The latest general availability release is numbered CAS 3.1.1 to better align with other Oracle Endeca Commerce products such as Tools and Frameworks 3.1.1.

Migrating CAS 3.0.x to 3.1.1

This section describes migrating from CAS 3.0.x to version 3.1.1. 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 3.0.x

This procedure describes how to back up CAS configurations from 3.0.x before upgrading. CAS configuration includes crawl configurations, Record Store configurations, Web crawler configuration, CAS extensions, and so on.

This procedure does not describe how to migrate the data in a Record Store instance or 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 CAS 3.0.x:

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 into version 3.1.1 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>\CAS\<old version>\bin` on Windows or `<install path>/CAS/<old version>/bin` on UNIX.
   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 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 in 3.0.x 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 3.1.1

To upgrade, uninstall the older version, and install 3.1.1 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 3.1.1:

1. If you haven't already, upgrade to 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 3.1.1. For details, see the *CAS Installation Guide*.

Restoring a CAS 3.0.x backup into CAS 3.1.1

To restore a CAS 3.0.x backup into CAS 3.1.1:

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 3.1.1 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 3.0.x, 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 3.1.1 version of `jetty.xml`.
   d) Save and close the 3.1.1 version of `jetty.xml`.
3. If you had any CAS plug-ins installed in 3.0.x, 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 3.0.x 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/3.1.1/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 3.0.x, import the Record Store instance configuration into 3.1.1 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 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]
      ```

8. Import the backed up crawl configurations into 3.1.1 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 3.0.x* 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 3.0.x, copy the directory containing the plug-in JAR and `plugin.xml` from the backup location to `<install path>/CAS/3.1.1/lib/web-crawler/plugins`.
10. If you used the Web Crawler in 3.0.x, 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/3.1.1/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 3.0.x.

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/3.1.1/lib/recordstore-forge-adapter/recordstore-forge-adapter-3.1.1.jar`.

You can re-crawl data sources as necessary. The back up task for CAS 3.0.x includes the Record Store instances, so a baseline update may not be necessary unless the application requires an update.

Chapter 2

Required Changes in 3.1.1

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

System requirements changes

CAS 3.1.1 removes support for Windows Server 2003 and Red Hat Enterprise Linux ES and AS version 4 for x64. For details about supported platforms, see the *Content Acquisition System Installation Guide*.

Chapter 3

Behavioral Changes in 3.1.1

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.

Crawls fail if no records are output

In previous releases, a crawl would successfully complete even if no records were output. In CAS 3.1.1, a crawl fails if no records are output. This includes full crawls configured for Record File output, Record Store output, and MDEX-compatible record output. In the case of MDEX-compatible record output, this applies to data records only not dimension value records.

The configure_cas_console.sh script has been renamed and moved

In previous releases, the `configure_cas_console.sh` script was stored in `$CAS_ROOT/console`. In CAS 3.1.1, the `configure_cas_console.sh` script has been renamed to `configure_tools_and_frameworks_integration.sh` and it is stored in `$CAS_ROOT/bin`.

For details about how to use `configure_tools_and_frameworks_integration.sh`, see the *CAS Installation Guide*.

CONFIRMPASSWORD is not required for the silent installer on Windows

You no longer need to specify the `CONFIRMPASSWORD` parameter as part of the silent installation configuration. If you specify `CONFIRMPASSWORD`, it is ignored.

Log file encoding is UTF-8

CAS 3.1.1 produces log files encoded as UTF-8.

In previous releases, CAS produced log files encoded according to the default encoding for the platform running CAS. On Linux, the encoding was UTF-8, and on Windows the encoding was typically Win-1252. If CAS logged data that could not be encoded properly in Win-1252, the data displayed as `"?"`. Encoding log files in UTF-8 corrects this problem.