

Endeca Content Acquisition System

Migration Guide

Version 3.0.2 • March 2012

ORACLE®

ENDECA

Contents

- Preface.....7**
- About this guide.....7
- Who should use this guide.....7
- Conventions used in this guide.....7
- Contacting Oracle Endeca Customer Support.....8

- Chapter 1: Upgrading the Content Acquisition System.....9**
- Recommended reading.....9
- Migration paths.....9
- Migrating CAS 2.2.x to 3.0.2.....10
 - Backing up CAS 2.2.x10
 - Upgrading to CAS 3.0.2.....11
 - Restoring a CAS 2.2.x backup into CAS 3.0.2.....12
 - Updating the CAS Deployment Template Component from 2.2.x to 3.0.2.....13
 - Updating the Forge pipeline and re-crawling data sources.....14
 - Upgrading CAS 2.2.x client applications that use the CAS APIs.....14
- Migrating CAS 3.0.0 or 3.0.1 to 3.0.2.....15
 - Backing up CAS 3.0.0 or 3.0.1.....15
 - Upgrading to CAS 3.0.2.....15
 - Restoring a CAS 3.0.0 or 3.0.1 backup into CAS 3.0.2.....16
 - Updating the Forge pipeline and re-crawling data sources.....17
 - Upgrading CAS 3.0.0 and 3.0.1 client applications that use the CAS APIs.....17

- Chapter 2: Required Changes for upgrades from CAS 2.2.x to 3.0.2...19**
- System requirements changes.....19
- CMS connector changes.....19
- Record property changes.....21
- Record Store output changes.....22

- Chapter 3: Behavioral Changes in 3.0.0 and 3.0.1.....23**
- Enablement changes23
- Packaging changes for the JUnit Testing Framework.....23
- Packaging changes for the Document Conversion Module.....23



Copyright and disclaimer

Copyright © 2003, 2012, 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.

Rosette® Linguistics Platform Copyright © 2000-2011 Basis Technology Corp. All rights reserved.

Teragram Language Identification Software Copyright © 1997-2005 Teragram Corporation. All rights reserved.

Preface

Oracle Endeca's Web 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 Web commerce solution delivers, analyzes, and targets just the right content to just the right customer to encourage clicks and drive business results.

Oracle Endeca Guided Search 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 Guided Search enables businesses to help guide and influence customers in each step of their search experience. At the core of Oracle Endeca Guided Search 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 the major migration tasks for the Content Acquisition System.

It assumes that you are familiar with the concepts of the Endeca Content Acquisition System and the Endeca Information Transformation Layer.

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 Endeca Customer Support

Oracle Endeca Customer 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 Endeca Customer 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 customers through setting up and running CAS.

Migration paths

In all migration scenarios, the migration tasks are similar: back up the old version, uninstall the old version, install the new version, restore the backup, configure, and test.

The following migration paths are supported:

- **From 2.2.x to 3.0.2** - As part of migrating from 2.2.x to 3.0.2, you may also have to make configuration changes from the 2.2.1 or 2.2.2 release if they apply to your application.
- **From 3.0.0 or 3.0.1 to 3.0.2** - This is a simpler migration scenario with little or no configuration changes.

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 2.2.1 to 3.0.2 but you cannot skip from 2.0.0 to 3.0.2. Upgrades from CAS 1.0, 1.1, 1.2, 2.0, and 2.1 to 3.0 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, the procedures may use `<install path>\CAS\OldVersion\` to mean 2.2.0, 2.2.1, 3.0.0, or 3.0.1 depending on which release you are upgrading from.

Migrating CAS 2.2.x to 3.0.2

This section describes migrating from CAS 2.2.0, 2.2.1, or 2.2.2 to 3.0.2.

Backing up CAS 2.2.x

This procedure describes how to back up CAS configurations from 2.2.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 2.2.x:

1. Retrieve and save your crawl configurations by doing the following:
 - a) Open a Command Prompt and navigate to `<install path>\CAS\OldVersion\bin` on Windows or `<install path>/CAS/OldVersion/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 3.0.2 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\OldVersion\bin` on Windows or `<install path>/CAS/OldVersion/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.

- If you have any CAS plug-ins installed in `<install path>/CAS/OldVersion/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/OldVersion/lib/cas-server-plugins/cas` or `<install path>/CAS/OldVersion/lib/cas-server-plugins/entropysoft`

- If you have any custom Web Crawler plug-ins installed in `<install path>/CAS/OldVersion/lib/web-crawler/plugins`, copy the directory containing the JAR and `plugin.xml`, to a location outside the CAS installation.
- If you have any CMS connectors in 2.2.x that required you to add additional JAR files as part of connector set up to `<install path>/CAS/OldVersion/lib/cas-server-plugins/entropysoft`, copy the JAR files to a location outside the CAS installation.
- If you made any changes to `<install path>/CAS/OldVersion/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.
- 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 3.0.2

To upgrade, back up all CAS configuration, uninstall the older version, and install the version 3.0.2 of CAS as described in this topic.

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

To upgrade to CAS 3.0.2:

- If you haven't already, upgrade to Platform Services 6.1 and Oracle Endeca Workbench 2.1. For details, see the *Platform Services Migration Guide* and the *Oracle Endeca Workbench Migration Guide*.
- 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 version 3.0.2 of the Content Acquisition System. For details, see the *CAS Installation Guide*.

Restoring a CAS 2.2.x backup into CAS 3.0.2

This procedure discusses how to restore a CAS 2.2.x backup.

To restore a CAS 2.2.x backup into CAS 3.0.2:

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 3.0.2 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 2.2.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 3.0.2 version of `jetty.xml`.
 - d) Save and close the 3.0.2 version of `jetty.xml`.
3. If you had any CAS plug-ins installed in 2.2.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 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 2.2.x that required you to add additional JAR files as part of connector set up to `<install path>/CAS/OldVersion/lib/cas-server-plugins/entropysoft`, copy the JAR files from the backup location to `<install path>/CAS/3.0.2/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 2.2.x, import the Record Store instance configuration into 3.0.2 by doing the following:
 - a) Open a Command Prompt and navigate to `<install path>\CAS\3.0.2\bin` on Windows or `<install path>/CAS/3.0.2/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 3.0.2 by doing the following:

- a) Open a Command Prompt and navigate to `<install path>\CAS\3.0.2\bin` on Windows or `<install path>/CAS/3.0.2/bin` on UNIX.
 - b) Check [CMS connector changes](#) on page 19 to see whether there are additional crawl configuration changes for your CMS connector.
 - c) Run `cas-cmd` and the `createCrawls` task. Specify the path to the XML file you created in [Backing up CAS 2.2.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 2.2.x, copy the directory containing the plug-in JAR and `plugin.xml` from the backup location to `<install path>/CAS/version/lib/web-crawler/plugins`.
 10. If you used the Web Crawler in 2.2.x, do the following:
 - a) Copy the older version of `default.xml` and `site.xml` from the workspace backup into the 3.0.2 `workspace\conf\web-crawler` locations.
 - b) Open `default.xml` in a text editor and modify the path in the `plugin.folders` property to the 3.0.2 directory structure of `<install path>/CAS/3.0.2/lib/web-crawler/plugins`.

Updating the CAS Deployment Template Component from 2.2.x to 3.0.2

In CAS 3.0.x, the CAS Deployment Template Component is now installed with CAS and not installed with the Deployment Template. Also, the CAS Deployment Template Component is packaged in a Java package that has a different fully qualified name than in CAS 2.2.x releases.

The name change affects compatibility with CAS versions. If you are using CAS 2.2.x, the CAS Deployment Template Component is implemented in `com.endeca.soleng.eac.toolkit.component.ContentAcquisitionServerComponent`. If you are using CAS 3.0.x, the CAS Deployment Template Component is implemented in `com.endeca.eac.toolkit.component.cas.ContentAcquisitionServerComponent`.

There are two scenarios to update Deployment Template configuration:

- For *new applications*, you can update the Deployment Template files on which all new applications are based. In this scenario, you specify new configuration settings in `fetchCasCrawlDataConfig.xml` to point to the new CAS Deployment Template component and you update the Deployment Template with the CAS WSDL client stubs. After the modification, any new applications that you create with the `deploy` script will specify the class name for the CAS 3.0.x Deployment Template Component.
- For *existing applications*, you can update a specific application. This scenario is typically appropriate if you have already deployed an application and have CAS integrated into your Deployment Template environment. In this scenario, you modify the `AppConfig.xml` settings to specify the new class name for the CAS 3.0.x Deployment Template Component.

To update the Deployment Template files:

1. On Windows, copy `<install path>\CAS<version>\lib\cas-dt\casStubs.jar` to `<install path>\Solutions\deploymentTemplate-<version>\data\eac-java\common\config\lib\java`. Or, on UNIX, copy `<install path>/CAS/<version>/lib/cas-dt/casStubs.jar` to `<install path>/Solutions/deploymentTemplate-<version>/data/eac-java/common/config/lib/java`.

2. Navigate to `<installpath>/deploymentTemplate-<version>/data/eac-java/cas_crawl/config/script/`.
3. In a text editor, open `fetchCasCrawlDataConfig.xml`.
4. In the `custom-component` element for CAS, change the value of the `class` attribute to `class="com.endeca.eac.toolkit.component.cas.ContentAcquisitionServerComponent"`.
5. Save and close the file.

To update a specific application after deployment:

1. Navigate to `<installpath>/<appDir>/config/script`.
2. In a text editor, open `AppConfig.xml`.
3. In the `custom-component` element for CAS, change the value of the `class` attribute to `class="com.endeca.eac.toolkit.component.cas.ContentAcquisitionServerComponent"`.
4. Save and close the file. Here is an example of an updated `AppConfig.xml`:

```
#####
# Content Acquisition System Server
#
<custom-component id="CAS" host-id="CASHost" class="com.endeca.eac.toolkit.component.cas.ContentAcquisitionServerComponent">
  <properties>
    <property name="casHost" value="localhost" />
    <property name="casPort" value="8500" />
  </properties>
</custom-component>
```

5. Modify the Deployment Template 3.2 to use the CAS 3.0.2 WSDL stub files by copying `<installpath>\CAS\<version>\lib\cas-dt\casStubs.jar` into each application that uses CAS, for example, `<installpath>/<appDir>/config/lib/java`.

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 if you are migrating from CAS 2.2.x.

To configure and test the upgrade:

1. 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 `<installpath>/CAS/3.0.2/lib/recordstore-forge-adapter/recordstore-forge-adapter-3.0.2.jar`.
2. Re-crawl all data sources and run a baseline update to process the Endeca records.

Upgrading CAS 2.2.x 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.

The WSDL namespace has changed in this release. If you are using a WSDL tool to generate stubs, you need to re-generate new CAS WSDL client stubs for `CasCrawler`, `RecordStore`, and the `ComponentInstanceManager`, and then make any changes listed in the Required Changes chapter. For details, see "Generating client stubs for the CAS Web Services" in the *Endeca CAS API Guide*.

Migrating CAS 3.0.0 or 3.0.1 to 3.0.2

This section describes migrating from either CAS 3.0.0 or 3.0.1 to version 3.0.2.

Backing up CAS 3.0.0 or 3.0.1

This procedure describes how to back up CAS configurations from 3.0.0 or 3.0.1 before upgrading to version 3.0.2.

To back up CAS 3.0.0 or 3.0.1:

1. Stop the Endeca CAS Service.
2. If you have any CAS plug-ins installed in `<install path>/CAS/OldVersion/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/OldVersion/lib/cas-server-plugins/cas` or `<install path>/CAS/OldVersion/lib/cas-server-plugins/entropysoft`

3. If you have any custom Web Crawler plug-ins installed in `<install path>/CAS/OldVersion/lib/web-crawler/plugins`, copy the directory containing the JAR and `plugin.xml` to a location outside the CAS installation.
4. If you have any CMS connectors that required JAR files in `<install path>/CAS/OldVersion/lib/cas-server-plugins/entropysoft`, copy the JAR files to a location outside the CAS installation.
5. If you made any changes to `<install path>/CAS/OldVersion/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.
6. 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, `ProductConfig.xml`, `DocumentConversionFilters.xml`, logging configuration files, and so on.

Upgrading to CAS 3.0.2

To upgrade, back up all CAS configuration, uninstall the older version, and install the version 3.0.2 of CAS as described in this topic.

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

To upgrade to CAS 3.0.2:

1. If you haven't already, upgrade to Platform Services 6.1 and Oracle Endeca Workbench 2.1. For details, see the *Platform Services Migration Guide* and the *Oracle Endeca Workbench 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 version 3.0.2 of the Content Acquisition System. For details, see the *CAS Installation Guide*.

Restoring a CAS 3.0.0 or 3.0.1 backup into CAS 3.0.2

This procedure discusses how to restore a CAS 3.0.0 or 3.0.1 backup into version 3.0.2.

To restore a backup into CAS 3.0.2:

1. Stop the Endeca CAS Service.
2. In the new 3.0.2 installation, move `<install path>\CAS\workspace` to a location outside the `<install path>` directory. (You could delete the `workspace` directory, but it is safer to move it.)
3. Copy the `workspace` directory from the older version of CAS into the 3.0.2 installation of `<install path>\CAS`.
4. If you had any CAS plug-ins installed, copy the directory containing the plug-in JAR or JARs from the backup location to `<install path>/CAS/version/lib/cas-server-plugins`.
5. If a crawl configuration contains 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.
6. If you modified `<install path>/CAS/OldVersion/bin/cas-service`, or `cas-service-wrapper.conf`, do the following:
 - a) Open the backed up copy of `cas-service` or `cas-service-wrapper.conf` in a text editor.
 - b) Copy your custom JVM changes from the backup.
 - c) Open `cas-service` or `cas-service-wrapper.conf` of your CAS 3.0.2 installation in a text editor and replace the defaults with your custom JVM changes.
 - d) Save and close files.
7. If you had any custom Web Crawler plug-ins installed in 3.0.0 or 3.0.1, copy the directory containing the plug-in JAR and `plugin.xml` from the backup location to `CAS/version/lib/web-crawler/plugins`.
8. If you used the Web Crawler in 3.0.0 or 3.0.1, open `default.xml` in a text editor and modify the path in the `plugin.folders` property to the 3.0.2 directory structure of `<install path>\CAS\3.0.2\lib\web-crawler\plugins`.
9. If you had any CMS connectors in 3.0.0 or 3.0.1 that required you to add additional JAR files as part of connector set up to `<install path>/CAS/OldVersion/lib/cas-server-plugins/entropysoft`, copy the JAR files from

```
the backup location to <install  
path>/CAS/3.0.2/lib/cas-server-plugins/entropysoft
```

10. Start the Endeca CAS Service.

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 if you are migrating from CAS 3.0.0 or 3.0.1.

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/3.0.2/lib/recordstore-forge-adapter/recordstore-forge-adapter-3.0.2.jar.
```

You can re-crawl data sources as necessary. The back up task for CAS 3.0.0 and 3.0.1 includes the Record Store instances, so a baseline update may not be necessary unless the application requires an update.

Upgrading CAS 3.0.0 and 3.0.1 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*.



Chapter 2

Required Changes for upgrades from CAS 2.2.x to 3.0.2

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

System requirements changes

CAS Console no longer supports Internet Explorer 6.

CMS connector changes

This topic lists changes to various CMS connectors.

FileNet P8 Connector changes

In CAS 2.2.1, the FileNet P8 Connector has a new **Workplace URL** configuration property. The property is optional. For details, see the *FileNet P8 Connector Guide*.

The Microsoft SharePoint connector has been renamed

In CAS 2.2.1, the Microsoft SharePoint connector has been renamed to the Microsoft SharePoint Web Services connector.

To update the `crawlConfig`, change the `moduleId` from `Microsoft SharePoint` to `Microsoft SharePoint Web Services`.

For example, the following `crawlConfig` snippet shows the `moduleId` in CAS 2.2.0:

```
<configurations xmlns="http://endeca.com/it1/cas/2010-03">
  <crawlConfig>
    <crawlId>
      <id>MyCrawl</id>
    </crawlId>
    <crawlThreads>3</crawlThreads>
    <sourceConfig>
      <moduleId>
        <id>Microsoft SharePoint</id>
      </moduleId>
    </sourceConfig>
  </crawlConfig>
</configurations>
```

...

In CAS 3.0.0, the `moduleId` is now as follows:

```
<configurations xmlns="http://endeca.com/itl/cas/2010-03">
  <crawlConfig>
    <crawlId>
      <id>MyCrawl</id>
    </crawlId>
    <crawlThreads>3</crawlThreads>
    <sourceConfig>
    <moduleId>
      <id>Microsoft SharePoint Web Services</id>
    </moduleId>
    ...
```

Open Text Livelink changes

In CAS 2.2.1, the Open Text Livelink Connector has a new **Search Broker [All Versions]** configuration property. For details, see the *Open Text Livelink Connector Guide*.

Also, the **Display URI** configuration property, as it displays in CAS Console, has been renamed to **Display URL**. However the configuration property is still named `displayURI`.

Documentum Content Server Connector changes

In CAS 2.2.1, the Documentum Content Server Connector has a new **Webtop URL** configuration property. The property is optional. For details, see the *Documentum Content Server Connector Guide*.

Also, the `docbase` configuration property, as it displays in CAS Console, has been renamed to **Docbase Name**. However the configuration property is still named `docbase`.

Setting up the CAS Server for Interwoven TeamSite

In previous releases, you set up the CAS Server by copying JAR files from your TeamSite installation into `<install path>\CAS\version\lib\cas-server-plugins\entropysoft`. In CAS 2.2.1, that set up is no longer required.

Interwoven TeamSite configuration property changes

In CAS 2.2.1, the syntax of the `serviceBaseURL` configuration property has changed.

In CAS 2.2.0 and earlier, the syntax of this property is `http://server name:port`.

In CAS 2.2.1, the syntax of this property is `http://server name:port/iw/services/cm/2.0`.

FileNet P8 Connector changes

In CAS 3.0.0, the following configuration properties are new:

- **Remote Server URL**. Required.
- **Check SSL Certificate**. Optional.
- **Remote file transfer URL**. Optional.

For details about these properties, see the *FileNet P8 Connector Guide* or the *CAS Console Help*.

CAS Server configuration for the FileNet P8 connector

In CAS 3.0.0, there are no longer any tasks necessary to set up or configure the CAS Server for FileNet P8. You now simply enable the connector and specify configuration properties as necessary.

Documentum eRoom connector changes

In CAS 3.0.0, there is a new optional **Check SSL Certificate** configuration property. For details, see the *Documentum eRoom Connector Guide* or the *CAS Console Help*.

SharePoint Object Model Connector and Web Services connector

In CAS 3.0.0, there is a new optional **Check SSL Certificate** configuration property. For details, see the *Microsoft SharePoint Connector Guide* or the *CAS Console Help*.

JSR-170 changes

In CAS 3.0.0, the **JNDI Name** configuration property has been modified to become the **JNDI name or RMI address** configuration property.

There is also a new **JCR Repository retrieval method** configuration property that determines how the **JNDI Name** property is interpreted. The default value is JNDI.

For details, see the *JSR-170 Compliant Repositories Connector Guide* or the *CAS Console Help*.

Record property changes

This topic describes newly added record properties and changes to them.

Endeca.CMS.UpdatedBy has been added to CMS crawls

In CAS 2.2.1, the `Endeca.CMS.UpdatedBy` has been added to CMS crawls. Previously, a similar property appeared with a different name as a `Endeca.CMS.Misc` property namespace.

Endeca.CMS.Uri has been added to CMS crawls

In CAS 2.2.1, the `Endeca.CMS.Uri` has been added to CMS crawls. Previously, a similar property appeared with a different name as a `Endeca.CMS.Misc` property namespace.

Changes to Endeca.CMS.Misc

In CAS 2.2.1, the `Endeca.CMS.Misc` properties are passed through from the underlying CMS connector without modification and may have changed.

Changes to the format of user, group, or role names in CMS properties

In CAS 2.2.1, the format of user, group, or role names in the record properties `Endeca.CMS.AllowReadContent`, `Endeca.CMS.AllowReadProperties`, `Endeca.CMS.DenyReadContent`, and `Endeca.CMS.DenyReadProperties` has changed for some CMS connectors.

Previously, CMS connectors were inconsistent in using the formats: user display name, user display name@domain, domain\user name, domain\group name, and domain\role name.

CAS now makes its best effort to return names in the form [domain\](user name), [domain\](group name), and [domain\](role name) but CAS is can be limited by the capabilities of the underlying CMS and the values the CMS returns in ACLs.

Record Store output changes

The following changes affect the Record Store output.

The `useSsl` module property for Record Store output configuration has been renamed to `isPortSsl`

In CAS 3.0.0, the `useSsl` module property has been renamed to `isPortSsl`. This property is used in the `outputConfig` specification of a `crawlConfig`.

For example, the CAS 2.2.1 configuration is as follows:

```
...
<outputConfig>
  <moduleId>
    <id>Record Store</id>
  </moduleId>
  <moduleProperties>
    <moduleProperty>
      <key>useSsl</key>
      <value>>false</value>
    </moduleProperty>
  </moduleProperties>
</outputConfig>
....
```

The 3.0.0 configuration is as follows:

```
...
<outputConfig>
  <moduleId>
    <id>Record Store</id>
  </moduleId>
  <moduleProperties>
    <moduleProperty>
      <key>isPortSsl</key>
      <value>>false</value>
    </moduleProperty>
  </moduleProperties>
</outputConfig>
....
```

There has been no change to the operation of the module property. (A value of `true` uses HTTPS and treats the `port` property as an SSL port.)

If you upgrade `crawl` configurations according to the process described in Chapter 1, then the CAS Server Command Line Utility maps the old property name to the new name and changes it automatically.

If you are using the CAS Server API to configure this property, you should change the string you provide to the `ModuleProperty` key from "useSsl" to "isPortSsl".

For additional details about how this property is used, see "Configuring a crawl to write to a Record Store instance" in the *Content Acquisition System Developer's Guide*.

The `-rs` suffix has been removed from new Record Store instance names

In CAS 2.2.1, the Component Instance Manager created new Record Store instances with the name `<crawl ID>-rs`. The `-rs` suffix was always part of the Record Store instance name.

In CAS 3.0.0, the `-rs` suffix has been removed from new Record Store instance names. The name of a new Record Store instance now corresponds directly to the crawl identifier.

The CAS Server Command Line Utility (`cas-cmd`) preserves the old name during the migration process. Only newly created names drop the suffix.



Chapter 3

Behavioral Changes in 3.0.0 and 3.0.1

This section describes changes that do not require action on your part, but do have an effect on how your Endeca application behaves after you upgrade. There are no behavioral changes in version 3.0.2.

Enablement changes

CAS Document Conversion module and CMS Connectors are enabled by default.

In CAS 3.0.1, the CAS Document Conversion module and all CMS Connectors are enabled by default. If desired, you can still disable a CMS connector.

To disable a connector:

1. Navigate to `<installation_path>\CAS\workspace\conf\ProductConfig.xml`.
2. In a text editor, open `ProductConfig.xml`.
3. Locate the name of the CMS connector you want to disable, and delete (or comment out) the `<FEATURE>` tag for the connector.
4. Save and close the file.
5. Restart the Endeca CAS Service.

Packaging changes for the JUnit Testing Framework

In CAS 3.0.1, the JUnit Testing Framework has been removed from this software distribution. Please download and install JUnit version 3.8.1 before building `cas-sample-extensions` using Ant. To download JUnit, see

<http://sourceforge.net/projects/junit/files/junit/3.8.1/junit3.8.1.zip/download>.
Once you download JUnit, extract the JAR, and place it in `cas-extensions\lib\test` directory.

Packaging changes for the Document Conversion Module

CAS 3.0.0 installs its own instance of the Document Conversion Module.

CAS installs the CAS Document Conversion Module

You no longer need to download and install the standalone Document Conversion Module.

Upgraded file format and version support in the CAS Document Conversion Module

CAS 3.0.0 contains an upgrade to the CAS Document Conversion Module. The module supports additional file formats and additional versions of file formats from earlier releases of the separately installed Document Conversion Module. For the complete list of file format and version support, see "File Formats Supported by the CAS Document Conversion Module" in the *CAS Developer's Guide*.