Oracle Commerce

Content Acquisition System Migration Guide

Version 11.2 • October 2015

Contents

Copyright and disclaimer

for any loss, costs, or damages incurred due to your access to or use of third-party content, products, or services, except as set forth in an applicable agreement between you and Oracle.

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 that have purchased support 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

Oracle Commerce 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, Guided Search enables businesses to influence customers in each step of their search experience. At the core of Guided Search is the MDEX Engine™, a hybrid search-analytical database specifically designed for high-performance exploration and discovery. The Oracle Commerce 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. The Oracle Commerce Assembler dynamically assembles content from any resource and seamlessly combines it into results that can be rendered for display.

Oracle Commerce Experience Manager enables non-technical users to create, manage, and deliver targeted, relevant content to customers. With Experience Manager, you can combine unlimited variations of virtual product and customer data into personalized assortments of relevant products, promotions, and other content and display it to buyers in response to any search or facet refinement. Out-of-the-box templates and experience cartridges are provided for the most common use cases; technical teams can also use a software developer's kit to create custom cartridges.

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 Oracle Commerce 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 answers to implementation questions, product and solution help, and important news and updates about Guided Search software.

You can contact Oracle Support through the My Oracle Support site 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 Notes

Refer to the release notes for information about known issues for this release. You can download the *Oracle Commerce Guided Search and Experience Manager Release Notes* from the Oracle Technology Network.

CAS Quick Start Guide

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

Migration path

CAS supports migrating from 11.1 to 11.2.

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, and 3.1.1 or lower 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 11.1.

Upgrading CAS 11.1 to 11.2

This section describes migrating from CAS 11.1 to version 11.2. 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 11.1

This procedure describes how to back up CAS configurations from 11.1 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 Configura¬
   tionFileName.xml [-h HostName] [-n] [-p PortNumber]
   ```
   c) To migrate existing 11.1 data to 11.2, 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 RecordStoreuOt¬
   putFileName.xml [-h HostName] [-p PortNumber]
   ```
3. Stop the 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`.

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 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.

7. 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 11.2

To upgrade, uninstall the older version, and install 11.2 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 Oracle Commerce Guided Search Platform Services and Oracle Commerce Tools and Frameworks. For details, see the *Platform Services Migration Guide* and the *Oracle Commerce 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 **Oracle Commerce 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 11.2. For details, see the *CAS Installation Guide*.

Restoring a CAS 11.1 backup into CAS 11.2

To restore a CAS backup into CAS 11.2:

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 11.2 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 11.1, 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 11.2 version of `jetty.xml`.
d) Save and close the 11.2 version of `jetty.xml`.

3. If you had any CAS plug-ins installed in 11.1, 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. Restart the CAS Service.
6. If you created unmanaged Record Store instances or if you modified the configuration of a Record Store instance in 11.1, import these Record Store instances including their configurations and contents into 11.2 by doing the following:
   a) Open a Command Prompt and navigate to `<install path>\CAS\<new version>\bin` on Windows or `<install path>/CAS/<new 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 11.1 data to 11.2, 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 RecordStoreuOtputFileName.xml [-h HostName] [-p PortNumber]
      ```

7. Import the backed up crawl configurations into 11.2 by doing the following:
   a) Open a Command Prompt and navigate to `<install path>\CAS\<new version>\bin` on Windows or `<install path>/CAS/<new 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 11.1* 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.

8. If you used any custom Web Crawler plug-ins installed in 11.1, copy the directory containing the plug-in JAR and `plugin.xml` from the backup location to `<install path>/CAS/11.2/lib/web-crawler/plugins`.
9. If you used the Web Crawler in 11.1, do the following:
   a) Copy the older version of `default.xml` and `site.xml` from the `workspace` backup into the 11.2 `workspace\conf\web-crawler` locations.
   b) Open `default.xml` in a text editor and modify the path in the `plugin.folders` property to the 11.1 directory structure of `<install path>/CAS/11.2/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 *Oracle Commerce 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 11.1.

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/11.2/lib/recordstore-forge-adapter/recordstore-forge-adapter-11.2.jar`.

You can re-crawl data sources as necessary. The back up task for CAS 11.1 includes the Record Store instances, so a baseline update may not be necessary unless the application requires an update.

Chapter 2

Required Changes

This section describes the required changes for this release.

Enabling script manipulators

This section describes how to enable the Filtering Script and Modifying Script manipulators.

Follow these steps:

1. Shut down the Oracle Commerce Content Acquisition System Server.
2. In the `%CAS_ROOT%\lib\cas-server-plugins\cas\plugin.xml` file. Uncomment the following two plugin extension entries:

```
<!--extension id="com.endeca.cas.manipulator.FilterScriptManipulator"
class="com.endeca.itl.extension.manipulator.script.FilterScriptManipulator"
pointId="com.endeca.cas.extension.Manipulator"/>

<extension id="com.endeca.cas.manipulator.ModifierScriptManipulator"
class="com.endeca.itl.extension.manipulator.script.ModifierScriptManipulator"
pointId="com.endeca.cas.extension.Manipulator"/-->
```

3. Start the Oracle Commerce Content Acquisition System Server.

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.

Improved exception handling

In previous releases of CAS, the following fatal exceptions resulted in commands finishing without errors or warnings. The CAS service log contained reports of fatal exceptions that occurred at the start of a crawl.

- Crawling a source file that did not exist
- Running full or incremental crawl in a state when it was not allowed
- Incorrect or missing attributes
- Incorrect precedence rules or invalid search properties in the MDEX configuration

In CAS 11.2, exception handling has been improved by adding proper messaging when the following tasks are run:

- createCrawls
- updateCrawls
- startCrawls

 **Note:** Many of these exceptions are also invoked in the deployment template.

The following list contains some of the exceptions that generate informative messages on the command line, CAS Console, or CAS Server log.

- Source property in the crawl configuration is missing.
- Seed property specified in the crawl configuration does not exist or is empty.
- `Outputconfig` in crawl configuration is missing.
- `ModuleProperties` in `Outputconfig` in the crawl configuration are missing.
- `Outputconfig` is defined, but the `Outputdirectory` property in the crawl configuration is missing.
- `Ismanaged` property is missing in the crawl configuration.
- If `unavailableIncrementalSwitchesToFullCrawl` is false and any one of the following scenarios is true:
  - Crawl runs for the first time. A full crawl is expected.
  - Repository properties have changed. A full crawl is expected.
  - Existing seed removed.
  - Document conversion options and filters have changed. A full crawl is expected.

    - Manipulator configuration has changed. A full crawl is expected.
- `Outputconfig` is MDEX-compatible but there are no permissions to access the `inputDirectory` or `outputDirectory`.
- The `dimvals.xml` file is empty when `load_baseline_test_data` is run.
- The `recordstore-config.xml` file does not hold any configuration when set-configuration task is performed.
- `Outputconfig` is MDEX-compatible, but `InputDirectory` property value does not exist.
- Record store is empty when `baseline_update` is run.
- MDEX input file is empty.
- The `InputFile path` of the `Delimited` or `EndecaRecordFile` data source does not exist.
- The precedence rules configuration in the `index-config.json` is not valid. For example, it does not have an associated `triggerDimensionValueSpec` to a trigger dimension.

```
"precedence_rule1" : {
 "targetDimension" : "product.category",
 "triggerDimensionValueSpec" : "575",
 "triggerDimension" : "product.price_range",
 "isLeafTrigger" : false
 }
```

Exceptions in an integrated environment

When Oracle Commerce Guided Search has been integrated with Oracle Commerce Platform, the following exception generates a message:

- The record stores are not present in the CAS workspace directory.
- The record stores are not present in the CAS server when `load_baseline_test_data` is run.

  Note that this exception is different from the exception that results when the `dimvals.xml` file is empty when `load_baseline_test_data` is run.

SSL Protocol Changes

In release 11.2 of the Content Acquisition System (CAS), the cryptographic protocols TLSv1.1 and TLSv1.2 are enabled by default.

These protocols provide protection against serious security threats that have emerged recently. The protocols SSL 3.0 and TLS 1.0 do not provide similar protection and are disabled by default. Note that if you enable SSL 3.0 and TLS 1.0 -- for compatibility or any other reason -- you thereby make your application vulnerable to the serious threats against which TLSv1.1 and TLSv1.2 provide protection.

To enable the TLSv1.1 and TLSv1.2 protocols, uncomment the TLSSocketConnector file in `CAS_ROOT/workspace/conf/jetty.xml`:

```
<Call name="addConnector">
      <Arg>
          <New class="com.endeca.itl.jetty.TLSSocketConnector">
           <Set name="Port"><SystemProperty name="com.endeca.cas.ssl.port"/></Set>

             <Set name="maxIdleTime">600000</Set>
             <Set name="keystore"><SystemProperty name="javax.net.ssl.key¬
Store"/></Set>
             <Set name="keyPassword"><SystemProperty name="javax.net.ssl.key¬
StorePassword"/></Set>
```

```
            <Set name="truststore"><SystemProperty name="javax.net.ssl.trust¬
Store"/></Set>
            <Set name="trustPassword"><SystemProperty name="javax.net.ssl.trust¬
StorePassword"/></Set>
            <Set name="needClientAuth">true</Set>
            <Set name="IncludeProtocols">
                <Array type="java.lang.String">
                    <Item>TLSv1.1</Item>
                    <Item>TLSv1.2</Item>
                </Array>
            </Set>
          </New>
        </Arg>
    </Call>
```

Steps to enable the SSL 3.0 and TLS 1.0 protocols

Note: If you enable SSL 3.0 and TLS 1.0 -- for compatibility or any other reason -- you thereby make your application vulnerable to the serious threats against which TLSv1.1 and TLSv1.2 provide protection.

To enable the SSL 3.0 protocol for CAS, follow these steps:

1. Open `jetty.xml` at `%CAS_ROOT%..\workspace\conf`.
2. Set the "`IncludeProtocols`" array property in TLSSocketConnector to SSLv3:

```
<Call name="addConnector">
      <Arg>
          <New class="com.endeca.itl.jetty.TLSSocketConnector">
            <Set name="Port"><SystemProperty name="com.ende¬
ca.cas.ssl.port"/></Set>
            <Set name="maxIdleTime">600000</Set>
            <Set name="keystore"><SystemProperty name="javax.net.ssl.key¬
Store"/></Set>
            <Set name="keyPassword"><SystemProperty name="javax.net.ssl.key¬
StorePassword"/></Set>
            <Set name="truststore"><SystemProperty name="javax.net.ssl.trust¬
Store"/></Set>
            <Set name="trustPassword"><SystemProperty
name="javax.net.ssl.trustStorePassword"/></Set>
            <Set name="needClientAuth">true</Set>
            <Set name="IncludeProtocols">
                <Array type="java.lang.String">
                    <Item>SSLv3</Item>
                </Array>
            </Set>
          </New>
        </Arg>
    </Call>
```

3. Open java.security file in %CAS_ROOT%/java/jre/lib/security.
4. Uncomment the "jdk.tls.disabledAlgorithms" property and disable all other protocols except SSLv3 "jdk.tls.disabledAlgorithms=TLSv1, TLSv1.1, TLSv1.2".
5. Restart the CAS server.

To enable the TLS 1.0 protocol for CAS, follow the same steps that you follow to enable SSL 3.0, with the following exceptions:

1. In `%CAS_ROOT%\..\workspace\conf\jetty.xml` modify the item in `IncludeProtocols` array property in `TLSSocketConnector` to TLSv1:

```
<Set name="IncludeProtocols">
   <Array type="java.lang.String">
   <Item>TLSv1</Item>
   </Array>
   </Set>
```

2. In the `java.security` file, the "`jdk.tls.disabledAlgorithms`" property must be changed to "`jdk.tls.disabledAlgorithms=SSLv3, TLSv1.1, TLSv1.2`".