Endeca® Content Acquisition System

Installation Guide

Version 3.0.1 • December 2011

Contents

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

Chapter 1: Before you install.....9
Overview of the Content Acquisition System.....9
Compatibility with Endeca components.....9
Prerequisite Endeca components.....9
System requirements.....9
CAS Document Conversion Module.....11
Licensing for CMS connectors.....11
Supported CMS connectors.....11
Recommended reading.....12

Chapter 2: Installing the Content Acquisition System.....13
Creating a user for the Endeca services on Windows.....13
Installing CAS on Windows.....14
Installing CAS on UNIX.....15
Adding the Endeca CAS service to inittab on UNIX.....16
Installing only the CAS Console on UNIX.....17
Installing CAS silently on Windows.....17
Installing CAS silently on UNIX.....18
Installing a plug-in into CAS.....19
Updating the Deployment Template to use the WSDL client stubs and the CAS component.....19
Web Crawler installation.....20
Package contents and directory structure.....20
About changing the role used for the CAS Console extension.....23

Chapter 3: Uninstalling the Content Acquisition System.....25
Uninstalling CAS on Windows.....25
Uninstalling CAS Console on UNIX.....25
Uninstalling CAS on UNIX.....26
Uninstalling a plug-in from CAS.....26
Uninstalling CAS Console if its extension configuration was changed.....26

Copyright and disclaimer

Product specifications are subject to change without notice and do not represent a commitment on the part of Endeca Technologies, Inc. The software described in this document is furnished under a license agreement. The software may not be reverse engineered, decompiled, or otherwise manipulated for purposes of obtaining the source code. The software may be used or copied only in accordance with the terms of the license agreement. It is against the law to copy the software on any medium except as specifically allowed in the license agreement.

No part of this document may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording, for any purpose without the express written permission of Endeca Technologies, Inc.

Copyright © 2003-2011 Endeca Technologies, Inc. All rights reserved. Printed in USA.

Portions of this document and the software are subject to third-party rights, including:

Corda PopChart® and Corda Builder™ Copyright © 1996-2005 Corda Technologies, Inc.

Outside In® Search Export Copyright © 2011 Oracle. All rights reserved.

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

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

Trademarks

Endeca, the Endeca logo, Guided Navigation, MDEX Engine, Find/Analyze/Understand, Guided Summarization, Every Day Discovery, Find Analyze and Understand Information in Ways Never Before Possible, Endeca Latitude, Endeca InFront, Endeca Profind, Endeca Navigation Engine, Don't Stop at Search, and other Endeca product names referenced herein are registered trademarks or trademarks of Endeca Technologies, Inc. in the United States and other jurisdictions. All other product names, company names, marks, logos, and symbols are trademarks of their respective owners.

The software may be covered by one or more of the following patents: US Patent 7035864, US Patent 7062483, US Patent 7325201, US Patent 7428528, US Patent 7567957, US Patent 7617184, US Patent 7856454, US Patent 7912823, US Patent 8005643, US Patent 8019752, US Patent 8024327, US Patent 8051073, US Patent 8051084, Australian Standard Patent 2001268095, Republic of Korea Patent 0797232, Chinese Patent for Invention CN10461159C, Hong Kong Patent HK1072114, European Patent EP1459206, European Patent EP1502205B1, and other patents pending.

Preface

Endeca® InFront enables businesses to deliver targeted experiences for any customer, every time, in any channel. Utilizing all underlying product data and content, businesses are able to influence customer behavior regardless of where or how customers choose to engage — online, in-store, or on-the-go. And with integrated analytics and agile business-user tools, InFront solutions help businesses adapt to changing market needs, influence customer behavior across channels, and dynamically manage a relevant and targeted experience for every customer, every time.

InFront Workbench with Experience Manager provides a single, flexible platform to create, deliver, and manage content-rich, multichannel customer experiences. Experience Manager allows non-technical users to control how, where, when, and what type of content is presented in response to any search, category selection, or facet refinement.

At the core of InFront is the Endeca MDEX Engine,™ a hybrid search-analytical database specifically designed for high-performance exploration and discovery. InFront Integrator provides a set of extensible mechanisms to bring both structured data and unstructured content into the MDEX Engine from a variety of source systems. InFront Assembler dynamically assembles content from any resource and seamlessly combines it with results from the MDEX Engine.

These components — along with additional modules for SEO, Social, and Mobile channel support — make up the core of Endeca InFront, 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 installation tasks for the suite of CAS components.

It assumes that you are familiar with the concepts of the Endeca Content Acquisition System.

Who should use this guide

This guide is intended for application developers who are building applications using the Endeca CAS components and are responsible for installation 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 Endeca Customer Support

The Endeca Support Center provides registered users with important information regarding Endeca software, implementation questions, product and solution help, training and professional services consultation as well as overall news and updates from Endeca.

You can contact Endeca Standard Customer Support through the Support section of the Endeca Developer Network (EDeN) at *http://eden.endeca.com*.

Chapter 1

Before you install

This section provides an overview of the Content Acquisition System, system requirements, and other information you need to know before installing.

Overview of the Content Acquisition System

The Endeca Content Acquisition System is a set of components that add, configure, and crawl data sources for use in an Endeca application. Data sources include file systems, Content Management System crawls, Web servers, and custom data sources. The Endeca Content Acquisition System crawls data sources, converts documents and files to Endeca records and stores them for use in a Forge pipeline.

Compatibility with Endeca components

To determine the compatibility of the Content Acquisition System with other components in the Information Access Platform, see the *Endeca InFront Compatibility Matrix* available on EDeN.

Prerequisite Endeca components

CAS Console depends on Endeca Workbench.

See the *Endeca InFront Compatibility Matrix* for version compatibility. Workbench must be installed on the machine where you install the CAS Console for Endeca Workbench. (Any edition of Endeca Workbench is fine, including Merchandising Workbench, Publishing Workbench, and IAP Workbench.)

System requirements

The Content Acquisition System has the following system requirements. In this guide, the term “x64” refers to any processor compatible with the AMD64/EM64T architecture. Hard disk capacity must be sufficient to store the records written to the Record Store or to record output files. Please contact your Endeca representative if you need more information on sizing your hardware.

Supported operating systems

The Content Acquisition System is supported on the following 64-bit operating systems:

- Windows Server 2003 (For best performance on Windows, Endeca recommends Windows Server 2003 Enterprise Edition Service Pack 2.)
- Windows Server 2008 R2 Enterprise
- Red Hat Enterprise Linux ES (version 4 for x64)
- Red Hat Enterprise Linux AS (version 4 for x64)
- Red Hat Enterprise Linux Server (version 5 for x64)
- Red Hat Enterprise Linux Advanced Platform (version 5 for x64)
- SUSE Enterprise Linux 11

Note: Solaris is not supported.

Note: The CAS Document Conversion Module is not supported on SUSE Enterprise Linux 11.

VMware ESX 3.5 support

The Content Acquisition System is supported in VMware ESX 3.5 environments on the following guest operating systems:

- Windows Server 2003 (For best performance on Windows, Endeca recommends Windows Server 2003 Enterprise Edition Service Pack 2.)
- Red Hat Enterprise Linux Server (version 5 for x64).
- Red Hat Enterprise Linux Advanced Platform (version 5 for x64).

VMware vSphere 4 and 4.1 support

The Content Acquisition System is supported in VMware vSphere 4 and 4.1 environments on the following guest operating systems:

- Windows Server 2008 R2 Enterprise.
- Red Hat Enterprise Linux Server (version 5 for x64).
- Red Hat Enterprise Linux Advanced Platform (version 5 for x64).
- SUSE Enterprise Linux 11

Amazon Elastic Compute Cloud (EC2) support

The Content Acquisition System is supported in Amazon EC2 environments on the following guest operating systems:

- Windows Server 2008 R2 Enterprise.
- Red Hat Enterprise Linux Server (version 5 for x64).
- Red Hat Enterprise Linux Advanced Platform (version 5 for x64).
- SUSE Enterprise Linux 11

Hardware requirements for Linux and Windows on x64

Minimum hardware requirements:

- x64 processor, minimum 1.8 GHz
- 2 GB of RAM
- At least an 80 GB hard drive, depending on the size of your application data set

Note: If you are using RHEL 4 and the virtual address size of your CAS processes exceeds available RAM, it is recommended that you upgrade to the latest version of RHEL 5 to avoid known problems with RHEL 4 performance in this scenario.

CAS Console and Web browser support

CAS Console supports any of the following Web browsers:

- Internet Explorer version 7 or 8
- Mozilla Firefox 3.6

Other Web browsers are not supported. You may also want to see the Web browser support requirements listed in the *Endeca Workbench Installation Guide*.

CAS Console requires Adobe Flash Player 10.1 or higher.

CAS Document Conversion Module

In CAS 3.0.1 and later, the CAS Document Conversion Module is installed and enabled as part of the CAS installation.

Note: The CAS Document Conversion Module is not supported on SUSE Enterprise Linux 11.

For a list of supported file formats, see "Appendix B File Formats Supported by the CAS Document Conversion Module" in the *Endeca CAS Developer's Guide*.

Licensing for CMS connectors

In CAS 3.0.1 and later, all CMS connectors are installed and enabled as part of the CAS installation.

Supported CMS connectors

To run a crawl against a specific CMS data source, the Content Acquisition System requires a CMS connector.

As of this release, the Content Acquisition System supports the following CMS connectors:

- Documentum Content Server
- Documentum eRoom
- FileNet Document and Image Services
- FileNet P8
- Interwoven TeamSite
- JSR-170 compliant repositories
- Lotus Notes (Domino)
- Microsoft SharePoint (supports 2003 and 2007 repositories)
- OpenText LiveLink

For information about configuring each of the CMS connectors, see the corresponding Connector Guide for the specific CMS connector.

Recommended reading

Before installing, Endeca recommends that you read the following documents for important information about the release.

Getting Started Guide

The *Endeca Getting Started Guide* gives an overview of Endeca components and includes information about configuration scenarios. After installing all the components in your Endeca deployment, read this guide for information on verifying your installation. You can download the *Endeca Getting Started Guide* from the Downloads section of the Endeca Developer Network (EDeN) at *http://eden.endeca.com*.

Release Notes

Refer to the Endeca release notes for information about new features, changed features, and bug fixes for this release. You can download the release notes (`README.txt`) from the Knowledge Base section of the Endeca Developer Network (EDeN). After installation, release notes are also available in the following location:

- Windows: `CAS\version`
- UNIX: `CAS/version`

Migration Guide

Refer to the *Endeca CAS Migration Guide* for information about migrating your implementation from a previous version of Endeca software. You can download the *Endeca CAS Migration Guide* from the Knowledge Base section of the Endeca Developer Network (EDeN).

Chapter 2

Installing the Content Acquisition System

This section contains installation procedures and describes the contents of the installation directory.

Creating a user for the Endeca services on Windows

You must run the Endeca services as a specified user, for which you can control permissions.

You are asked to provide information about this Endeca services user during the installation process. Endeca recommends that you create a user account called `endeca` that has the proper file and directory permissions to access all necessary files for your application, and that you set up your Endeca Tools Service to run under this account. However, you can use any user that you prefer, as long as it meets these requirements.

To create the user endeca:

1. Ensure that you have administrator privileges on the local machine.
2. From the Windows Control Panel, select **Computer Management** > **Administrative Tools**.
3. In the tree pane of the Computer Management window, select **System Tools** > **Local Users and Groups**.
4. Right-click **Users** and select **New User**.
5. In the **New User** dialog box, do the following:
   a) Enter `endeca` for both the User name and the Full name.
   b) Optionally, enter a description.
   c) Set a password for user `endeca`.

      Note: The user must have a non-blank password, because the installer will not accept a blank password for the services user.

   d) Uncheck **User must change password at next logon**.
   e) Select **Password never expires**.
6. Click **Create** to create the new user, and then **Close** to exit the dialog box.
7. Close the Computer Management window, but do not exit Administrative Tools.
8. From Administrative Tools, do the following:
   a) Open **Local Security Policy**.
   b) In the Local Security Settings window, select **Local Policies** > **User Rights Assignments**.
   c) From the list of policies on the right, open **Log on as a Service**.

d) Add user `endeca` to the list of users that can register a process as a service.
e) Close the dialog box and exit Administrative Tools.

Installing CAS on Windows

This section provides instructions for installing the Endeca Content Acquisition System on Windows.

Before installing, ensure that you have met the following pre-requisites:

- Ensure you have administrator privileges on the local machine.
- Close all running programs and, if you previously had the Endeca software installed on this machine, refer to the "Upgrading" section of the *Endeca CAS Migration Guide* for information on migration procedures.
- If you are installing CAS Console, the `ENDECA_TOOLS_ROOT` and `ENDECA_TOOLS_CONF` environment variables must be set on the machine running Endeca Workbench before running the CAS installation program. Setting these variables happens automatically when you restart the machine that has Endeca Workbench installed.

Note: Even if you are installing on a drive other than the C drive, you should have approximately 400 MB of space available on the system drive before installing. This space is needed temporarily for the installer files that are unpacked on the C drive during the installation regardless of the drive you specified for your installation. The space is cleared after the installation process completes.

To install CAS on Windows:

1. In your local environment, locate the Endeca CAS software that you downloaded from the Endeca Developer Network (EDeN).
2. Double-click the installer file `cas-`*`version`*`_x86_64pc-win32.exe` to start the wizard.
3. Click **Next** to begin the installation process.
4. In the **Important Information** screen, read the copyright and then click **Next**.
5. In the **Custom Setup** screen, select the program features you want to install and then click **Next**.

   Note: Both installation items are selected by default, but you may choose to install either feature. The remainder of this guide assumes that you selected both options listed below:
   - Content Acquisition System
   - CAS Console as a Workbench Extension

6. In the **Destination Folder** screen, select an installation location or accept the default location of `C:\Endeca\CAS` and then click **Next**.
7. In the **Endeca CAS Service Information** screen, specify the user name, password, and domain information for the user who will run the CAS Service and then click **Next**. (This is typically the `endeca` user you created in the previous procedure.)
8. In the **CAS Server Information** screen, enter the CAS Server port and CAS Server shutdown port, or accept the default values of `8500` and `8506`.
9. In the **Completing the Setup Wizard** screen, click **Next**.

The CAS Service starts automatically.

Related Links

Installing CAS silently on Windows on page 17
The silent installer is useful if you want to add the CAS installation to your own install script, or push out the installation on multiple machines.

Updating the Deployment Template to use the WSDL client stubs and the CAS component on page 19
After installing CAS, you should update the Deployment Template to use the new WSDL client stubs and use the new CAS component that is installed with CAS. Both the WSDL client stubs and the CAS component (the `ContentAcquisitionServerComponent` class) are packaged in `casStubs.jar`.

Creating a user for the Endeca services on Windows on page 13
You must run the Endeca services as a specified user, for which you can control permissions.

Installing CAS on UNIX

The Endeca software is distributed as a self-extracting tar file and install script. It can be installed to any location.

To install CAS on UNIX:

1. Determine an installation location for the Endeca Content Acquisition System. In this procedure, we assume this is the `/usr/local` directory.

   Note: CAS unpacks to approximately 400 MB. Because multiple versions may be stored, a destination in a large partition is recommended.

2. Locate the CAS installation file that you downloaded from EDeN. This procedure assumes the location is `downloads/`*`arch-OS`* . The name of the installation file is `cas-`*`version_arch-OS`*`.sh`
3. Assuming the locations in steps 1 and 2, run the Endeca installation script with the following command:

   ```
   downloads/arch-OS/cas-version_arch-OS.sh --target /usr/local
   [--skip_console_installation] [--endeca_tools_root full path to the Endeca
    Tools
   root directory]
   [--endeca_tools_conf full path to the Endeca Tools /conf directory]
   ```

   If you use the `--skip_console_installation` flag, the installer does not attempt to install CAS Console as an extension to Endeca Workbench. Use this flag if your Endeca Workbench runs on a different server than CAS. The remainder of this procedure assumes you are running the installation script without the `--skip_console_installation` flag.

4. Enter the CAS Service port.
5. Enter the CAS Service Shutdown port.
6. If `ENDECA_TOOLS_ROOT` or `ENDECA_TOOLS_CONF` is not set in the environment, the installer prompts you for these values.

   If running the installation script with the `--skip_console_installation` flag, the installer does not prompt for these values.

7. Enter the fully qualified CAS Server host name.

   If running the installation script with the `--skip_console_installation` flag, the installer does not prompt for a host.

Following installation:

- To start the CAS Service, navigate to `/usr/local/endeca/CAS/version/bin` and run the following script: `cas-service.sh`
- If you chose to install the CAS Console, you must restart the Endeca Tools Service. See the *Endeca Workbench Installation Guide*.

Related Links

Installing CAS silently on UNIX on page 18

The silent installer is useful if you want to add the CAS installation to your own install script, or push out the installation on multiple machines.

Updating the Deployment Template to use the WSDL client stubs and the CAS component on page 19

After installing CAS, you should update the Deployment Template to use the new WSDL client stubs and use the new CAS component that is installed with CAS. Both the WSDL client stubs and the CAS component (the `ContentAcquisitionServerComponent` class) are packaged in `casStubs.jar`.

Adding the Endeca CAS service to inittab on UNIX

In a UNIX development environment, the Endeca CAS Service can be started from the command line. In a UNIX production environment, however, Endeca recommends starting it from `inittab`.

In UNIX you run the CAS Server using `cas-service.sh`, located in `usr/local/endeca/CAS/version/bin`. You can write a script that calls `cas-service.sh` and is referenced in `inittab`.

When writing your script, it is recommended as a best practice that you run the Endeca CAS Service as a user other than root.

This sample script (named `start_cas_service.sh`) sets the ENDECA_USER variable to the “endeca” user and then issues an `su` command to change to the “endeca” user:

```
#!/bin/sh
ENDECA_USER=endeca
CAS_ROOT=/usr/local/endeca/CAS/3.0.0
CAS_WORKSPACE=/usr/local/endeca/CAS/workspace

# change to user endeca
su $ENDECA_USER -c "cd $CAS_ROOT/bin; \
     $CAS_ROOT/bin/cas-service.sh >> $CAS_WORKSPACE/logs/cas-service-wrap¬
per.log 2>&1"
```

The `start_cas_service.sh` script is then referenced in `inittab` with an entry similar to this example:

```
ecas:2345:respawn:/usr/local/endeca/CAS/3.0.0/bin/start_cas_service.sh
```

When writing your startup script, keep in mind that it is server-specific, and therefore its details (such as paths and user names) depend on the configuration of your server.

Installing only the CAS Console on UNIX

You can install just the CAS Console if you used the `--skip_console_installation` flag when you installed the CAS, or if you plan to use the CAS Console with an additional Workbench, or with a CAS Service on a separate machine.

Use the `configure_cas_console.sh` script to install only the CAS Console:

- If you installed CAS but did not install the CAS Console as a Workbench Extension
- If you intend to install the CAS Console with more than one Workbench
- If you intend to run the CAS Service on a separate machine from the one that hosts the CAS Console Workbench Extension

To install only the CAS Console for Endeca Workbench:

- Open a command-line and issue the following:

  ```
  /usr/local/endeca/CAS/version/console/configure_cas_console.sh
  [--endeca_tools_root full path to the Endeca Tools root directory]
  [--endeca_tools_conf full path to the Endeca Tools /conf directory]
  ```

  This registers the CAS Console with the specified Endeca Workbench installation. CAS Console runs in the same Tomcat instance as Endeca Workbench.

  Note: You may omit the flags for `--endeca_tools_root` and `--endeca_tools_conf` if these variables are already set as environment variables.

You must restart the Endeca Tools Service for the CAS Console registration to take effect.

Installing CAS silently on Windows

The silent installer is useful if you want to add the CAS installation to your own install script, or push out the installation on multiple machines.

The silent installer is not interactive.

To launch the silent installer:

1. From a command prompt, navigate to the directory where you downloaded the installer.
2. Issue the following command:

   ```
   start /wait cas-version-arch-OS.exe /s /l=full path to the installer log
    file\installer.log
   [CASSELECTED=TRUE] [CONSOLESELECTED=TRUE] [CASHOST=localhost] [CAS¬
   PORT=8500] [CASSTOPPORT=8506] [TARGETDIR=C:\Endeca] USERNAME=endeca
   PASSWORD=endecapwd CONFIRMPASSWORD=endecapwd DOMMAINNAME=domain
   ```

   If `CASSELECTED=TRUE` the installer installs CAS, and if `CONSOLESELECTED=TRUE` the installer registers CAS Console as a Workbench extension. If you do not want to install one of these components enter `FALSE` for that variable.

   The logging option (`/l=`) provides useful information from silent installer failures. The specified directory must be a full path.

   You can replace `C:\Endeca` with the location to which you want to install. However, if you set the install location to a drive that does not exist, the silent installation fails with a non-zero status code.

Installing CAS silently on UNIX

The silent installer is useful if you want to add the CAS installation to your own install script, or push out the installation on multiple machines.

The silent installer is not interactive.

To run the silent installer:

1. Create a .txt file that contains rows to specify the response value for each prompt on its own line. The order of prompts in the installer is:

   ```
   CAS port
   CAS shutdown port
   CAS host
   ```

   For example:

   ```
   8500
   8506
   localhost
   ```

   Important: The silent installer does not prompt for `$ENDECA_TOOLS_ROOT` or `$ENDECA_TOOLS_CONF`, so if these variables are not set in the environment, or if you want to override what is set, you must use the `--endeca_tools_root` and `--endeca_tools_conf` flags when you run the silent installer (they must be specified after the `--silent` and `--target` flags).

2. From a command prompt, navigate to the location of the installer.
3. Run the install script with the `--silent` flag, the `--target` flag (which specifies the absolute path of the target installation directory), any optional arguments, and the response file as an input. For example:

   - ```
     ./cas-version-arch-OS.sh  --silent --target /usr/local <
     silent.txt
     ```

   - If `$ENDECA_TOOLS_ROOT` and `$ENDECA_TOOLS_CONF` are not set in the environment or you want to override their values, specify their respective flags:

     ```
     ./cas-version-arch-OS.sh  --silent --target /usr/local --ende¬
     ca_tools_root full
     path to the Endeca Tools root directory --endeca_tools_conf /usr/lo¬
     cal/endeca/workbench/2.0.0/conf
      < silent.txt
     ```

   - To skip installation of the CAS Console, you must specify the `--skip_console_installation` flag:

     ```
     ./cas-version-arch-OS.sh  --silent
     --target /usr/local < silent.txt --skip_console_installation
     ```

Following installation:

- To start the CAS Service, navigate to `CAS/version/bin` and run the following command: `cas-service.sh`
- If you chose to install the CAS Console, you must restart the Endeca Tools Service. See the *Endeca Workbench Installation Guide*.
     ```

Installing a plug-in into CAS

After receiving a plug-in (a JAR or set of JAR files) from a plug-in developer, a CAS application developer installs the plug-in into CAS.

The Content Acquisition System detects each plug-in and validates the extensions within it by checking the uniqueness of extension IDs and by checking for the presence of an annotation of either `@CasDataSource` or `@CasManipulator` for each extension.

To install a plug-in into CAS:

1. Stop Endeca CAS Service.
2. Navigate to `<install path>\CAS\`*`version`*`\lib\cas-server-plugins` and create a *`plugin-name`* subdirectory for each plug-in.
   For example: `CAS\`*`version`*`\lib\cas-server-plugins\JDBCDataSourceExt`
3. Copy the plug-in JAR or JARs, and any dependent JAR files, to `<install path>CAS\`*`version`*`\lib\cas-server-plugins\`*`plugin-name`* .
4. Repeat the steps above as necessary for multiple plug-ins.
5. Start Endeca CAS Service.

You can confirm that an extension is installed by runing the `listModules` task of the CAS Server Command-line Utility and specifying a `moduleType` of either `SOURCE` or `MANIPULATOR`. The task returns the installed modules. For example, this task shows that a custom data source named `Sample Data Source for testing` is installed:

```
C:\Endeca\CAS\3.0.0\bin>cas-cmd listModules -t SOURCE
Sample Data Source
 *Id: Sample Data Source
 *Type: SOURCE
 *Description: Sample Data Source for testing

File System
 *Id: File System
 *Type: SOURCE
 *Description: No description available for File System
 *Capabilities:
   *Binary Content Accessible via FileSystem
   *Data Source Filter
   *Has Binary Content
   *Expand Archives
```

Updating the Deployment Template to use the WSDL client stubs and the CAS component

After installing CAS, you should update the Deployment Template to use the new WSDL client stubs and use the new CAS component that is installed with CAS. Both the WSDL client stubs and the CAS component (the `ContentAcquisitionServerComponent` class) are packaged in `casStubs.jar`.

Making these changes means all deployed applications are compatible with CAS 3.0 but not earlier versions of CAS. If you want an application to be compatible with CAS 2.2.x, see the *Endeca Deployment Template Usage Guide* for configuration information.

To update the Deployment Template:

1. Copy the new CAS Deployment Template component into the Deployment Template:
   - On Windows, copy `<install path>\CAS\version\lib\cas-dt\casStubs.jar` to `<install path>\Solutions\deploymentTemplate-3.2\data\eac-java\common\config\lib\java` and replace the existing JAR file.
   - On UNIX, copy `<install path>/CAS/version/lib/cas-dt/casStubs.jar` to `<install path>/Solutions/deploymentTemplate-3.2/data/eac-java/common/config/lib/java`.

   The WSDL client stubs work with CAS 3.0.0 and later. They do not work with CAS 2.2.1 and earlier.
2. In the Deployment Template, modify the global CAS crawl configuration file to point to the new CAS Deployment Template component. This step modifies the configuration file that all new applications are based on when you run the `deploy` script.
   a) Navigate to `<install path>/deploymentTemplate-<version>/data/eac-java/cas_crawl/config/script/`.
   b) In a text editor, open `fetchCasCrawlDataConfig.xml`.
   c) In the `custom-component` element for CAS, change the value of the `class` attribute to `class="com.endeca.eac.toolkit.component.cas.ContentAcquisitionServer¬Component"`.
   d) Save and close the file.

For details on upgrading applications deployed for a previous version of CAS, see the *Endeca CAS Migration Guide*.

Web Crawler installation

The Endeca Web Crawler is installed as part of the CAS installation.

The Web Crawler startup script is in `<install path>\CAS\version\bin` (Windows) and `<install path>/CAS/version/bin` (UNIX). The Web Crawler configuration files are in `<install path>\CAS\conf\web` (Windows) and `<install path>/CAS/conf/web` (UNIX). See the *Endeca Web Crawler Guide* for details and usage.

Package contents and directory structure

The CAS directory is the root directory for the Content Acquisition System.

After installing the Content Acquisition System, you see this directory structure:

```
CAS
   3.0.0
         bin
         ...
   workspace
```

The contents of the CAS directory are described here in detail.

Directory	Contents
version\bin	• Scripts to start the CAS Service (`cas-service.sh` for UNIX and `cas-service.bat` for Windows).

Directory	Contents
	• Scripts to run the CAS Server Command-line Utility (`cas-cmd.sh` for UNIX and `cas-cmd.bat` for Windows). • Scripts to run the Component Instance Manager Command-line Utility (`component-manager-cmd.sh` for UNIX and `component-manager-cmd.bat` for Windows). • Scripts to run the Record Store Command-line Utility (`recordstore-cmd.sh` and `recordstore-cmd.bat`). • Scripts to run the Web Crawler (`web-crawler.sh` for UNIX systems and `web-crawler.bat` for Windows). • Scripts to shut down the CAS Service (`cas-service-shutdown.sh` for UNIX and `cas-service-shutdown.bat` for Windows). • The executable that runs the CAS Service as a Windows service (`cas-service-wrapper.exe`).
version`\cms`	Server-side files for the FileNet Document and Image Services connector. For more information about these files, see the *FileNet Document and Image Services Connector Guide*.
version`\components`	The `RecordStore.war` file, which is the Record Store application.
version`\console`	• The `casconsole.war` file, which is the CAS Console application. • The `casconsole.xml` file, which is the context file for the CAS Console. • The `casconsole.properties` file, which is the configuration file for the CAS Console. These files are copied to the Endeca Workbench installation and used from that location.
version`\doc`	• The reference documentation (Javadoc) for the CAS Server, the Component Instance Manager, the Record Store, and the Web Crawler. • The *Endeca Licensing Guide*.
version`\doc\wsdl`	The Web Service (WSDL) files for the CAS Server, the Component Instance Manager, and the Record Store.
version`\java`	The JDK used to run the CAS components (except CAS Console, which runs in the Endeca Tools Service).
version`\lib`	• Libraries for the CAS command-line utilities including: the CAS Server utility, the Component Instance Manager utility, the Record Store utility. • Libraries for the CAS APIs including: the CAS Server API, the Component Instance Manager API, the Record Store API, and the CAS Extension API.
version`\lib\cas-dt`	CAS WSDL client stub files for the Deployment Template.

Directory	Contents
version\lib\cas-server-plugins	Libraries for CAS plug-ins including CMS connectors and custom extensions (if applicable).
version\lib\oit-sx	The CAS Document Conversion Module libraries.
version\lib\recordstore-forge-adapter	The library for the Forge Record Store adapter.
version\lib\web-crawler	The Web Crawler libraries.
version\sample	Directories for the sample applications and sample extensions. The sample applications are documented in the *CAS Developer's Guide* and also in the *CAS Extension API Guide*.
version\webapps	The `cas.war` and `ComponentInstanceManager.war` files, which are the CAS Server and Component Instance Manager applications.
version\workspace_template	The template for the workspace directory that contains configuration files.
workspace	The working directory for the CAS Server and the Web Crawler.
workspace\conf	• The `commandline.properties` file, which contains the CAS Service settings necessary for the CAS command-line utilities to run. • The `default-log4j.properties` file, which should not be modified. • The `ProductConfig.xml` file, which is the licensing file for CAS components. • Three logging configuration files (`cas-service.log4j.properties` for the CAS Service, `recordstore-cmd.log4j.properties` for the Record Store, and `cas-cmd.log4j.properties` for the Command-line Utility). • The Jetty configuration files.
workspace\conf\web-crawler\default	The default configuration files for the Web Crawler, including the `log4j.properties` logging configuration file.
workspace\conf\web-crawler\non-polite-crawl	Sample crawl configuration files for non-polite crawls. As with the polite version, the settings in these files will override the default settings.
workspace\conf\web-crawler\polite-crawl	Sample crawl configuration files for polite crawls. The settings in `site.xml` will override the same property settings in the `default.xml` file, while the `crawl-urlfilter.txt` file will also override the default version of the file.
workspace\logs	The `cas-service.log` file, which contains the CAS Service log output, and includes log messages from all crawls managed by the CAS Server.
workspace\output	Default destination directory for the crawl output from the Web Crawler. The `output` directory is not present upon installation. It

Directory	Contents
	is created when the Web Crawler writes to output records for a crawl.
`workspace\state`	State files for the CAS Service components. State files can include Record Store instances, state directories for data source extension information, and state directories for manipulator extension information.

Note: There is no `logs` directory for the Web Crawler, because by default the Web Crawler sends its standard output to the console. However, you can modify the `log4j.properties` file to send the output to a file.

About changing the role used for the CAS Console extension

By default, only users with administrative rights can view the CAS Console Workbench extension. You can also make this extension visible to non-administrative users.

See the *Endeca Workbench Administrator's Guide* for details on changing the visibility of the CAS Console Workbench extension for different user roles.

Note: If you change a user role in Endeca Workbench, you must manually remove the CAS Console extension if you choose to unregister it.

Related Links

Uninstalling CAS Console if its extension configuration was changed on page 26

If you need to uninstall the CAS Console extension for Endeca Workbench and you have manually edited its extension configuration (for example to assign the extension to a role other than "admin"), you must manually uninstall the CAS Console as an Endeca Workbench extension.

Chapter 3

Uninstalling the Content Acquisition System

This section contains the procedures for uninstalling the CAS.

Uninstalling CAS on Windows

Follow these steps to uninstall the Content Acquistion System from your Windows machine.

The installer creates a timestamped backup of your workspace prior to uninstalling the Content Acquistion System

To uninstall the Content Acquistion System from a Windows machine:

1. From the Windows Control Panel, select **Add or Remove Programs**.
2. Select the **Endeca Content Acquisition System** from the list of installed software.
3. Click **Remove**.

Uninstalling CAS Console on UNIX

Follow this procedure to remove the CAS Console from Endeca Workbench without uninstalling the Content Acquisition System.

To remove the CAS Console from Endeca Workbench:

1. Open a command-line prompt and run the following command:
   `/endeca/CAS/`*`version`*`/console/configure_cas_console.sh --uninstall_console`
2. If you do not have `ENDECA_TOOLS_ROOT` and `ENDECA_TOOLS_CONF` set in the environment, do one of the following:
   - Enter the value for each at the prompt.
   - Run the `/endeca/CAS/`*`version`*`/console/configure_cas_console.sh --uninstall_console` script with the flags `--endeca_tools_root` *`full path to the Endeca Tools root directory`* `--endeca_tools_conf` *`full path to the Endeca Tools /conf directory`*
3. Restart the Endeca Tools Service

Uninstalling CAS on UNIX

Follow these steps to uninstall the Content Acquistion System from a UNIX machine.

Before you begin the uninstall process, back up any files that you want to retain from the CAS directory.

You must uninstall CAS Console for Endeca Workbench prior to uninstalling the CAS. See the previous topic.

To uninstall the CAS from a UNIX machine:

1. Shutdown CAS Service by navigating to `/usr/local/endeca/CAS/version/bin` and running the following command: `cas-service-shutdown.sh`.
2. In a command prompt, issue an `rm` command as in this example:
   `rm -rf endeca/CAS/version`

Uninstalling a plug-in from CAS

Follow these steps to uninstall a plug-in from the Content Acquisition System.

Before you begin the uninstall process, back up any plug-in JAR files that you want to retain from the `CAS\version\lib\cas-server-plugins` directory.

To uninstall a plug-in from the CAS:

1. Stop Endeca CAS Service.
2. Navigate to `<install path>\CAS\version\lib\cas-server-plugins`.
3. Delete the *`plugin-name/`* subdirectory for each plug-in.
   For example: `<install path>\CAS\version\lib\cas-server-plugins\JDBCDataSourceExt`
4. Repeat the steps above as necessary for multiple plug-ins.
5. Start Endeca CAS Service.

Uninstalling CAS Console if its extension configuration was changed

If you need to uninstall the CAS Console extension for Endeca Workbench and you have manually edited its extension configuration (for example to assign the extension to a role other than "admin"), you must manually uninstall the CAS Console as an Endeca Workbench extension.

To manually remove the CAS Console extension to Endeca Workbench:

1. Navigate to `%ENDECA_TOOLS_CONF%\conf` (on Windows) or `$ENDECA_TOOLS_CONF/conf` (on UNIX).
2. In `ws-extensions.xml`, remove the `<extension>` element with the id `"casconsole"`.
3. In `ws-mainMenu.xml`, remove the `<menuitem>` element with the id `"casconsole"`.
4. Delete `casconsole.properties`.
5. From `%ENDECA_TOOLS_CONF%\conf\Standalone` (on Windows) or `$ENDECA_TOOLS_CONF/conf/Standalone` (on UNIX), delete `casconsole.xml`.

6. From `%ENDECA_TOOLS_ROOT%\server\webapps` (on Windows) or `$ENDECA_TOOLS_ROOT/server/webapps` (on UNIX), delete `casconsole-`*`version`*`.war`.
7. Restart the Endeca Tools Service

Related Links

About changing the role used for the CAS Console extension on page 23

By default, only users with administrative rights can view the CAS Console Workbench extension. You can also make this extension visible to non-administrative users.

Index

B

bin directory
    contents of 20
    location of 20

C

CAS Server API
    location of 21
CAS Service
    starting from inittab 16
cas-server.log4j.properties file 22
CMS connectors
    licensing 11
    supported 11
CMS Connectors
    licensing file 22
configuration files
    for logging 22
    for non-polite crawls 22
    for polite crawls 22

D

Deployment Template
    upgrading to use WSDL stub files 19

I

installing
    CAS Console on UNIX 17
    on UNIX 15
    on Windows 14
    silently on UNIX 18
    silently on Windows 17

O

overview of package 9

P

package
    contents of 20
    directory structure of 20
    overview of 9
prerequisites 10

R

recordstore-server.log4j.properties file 22

S

system requirements 9, 10

U

uninstalling
    a plug-in 26
    CAS Console if extension configuration changed 26
    CAS Console on UNIX 25
    on UNIX 26
    on Windows 25

W

workspace directory
    location of 22
WSDL files
    location of 21
WSDL stub files
    using with Deployment Template 19