Installation Cookbook

IBM Content Analytics with Enterprise Search
Version 3.0

Vijai Gandikota
November 2012
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Installing IBM Content Analytics with Enterprise Search 3.0 with the Embedded Application Server</td>
<td>1</td>
</tr>
<tr>
<td>Accessing the administration console</td>
<td>20</td>
</tr>
<tr>
<td>Adding Additional Servers</td>
<td>21</td>
</tr>
<tr>
<td>Installing IBM Content Analytics with Enterprise Search 3.0 with WebSphere Application Server</td>
<td>22</td>
</tr>
<tr>
<td>Know Your Installation</td>
<td>44</td>
</tr>
<tr>
<td>Administration Console</td>
<td>44</td>
</tr>
<tr>
<td>Enterprise Search Application</td>
<td>46</td>
</tr>
<tr>
<td>Content Analytics Miner</td>
<td>47</td>
</tr>
<tr>
<td>ES_INSTALL_ROOT Directory Structure</td>
<td>48</td>
</tr>
<tr>
<td>Troubleshooting</td>
<td>59</td>
</tr>
</tbody>
</table>
Introduction
This article describes how to install IBM® Content Analytics with Enterprise Search and set it up to use either the embedded application server or WebSphere® Application Server. This document shows examples and screenshots from a Microsoft Windows operating system installation. However, you can easily apply these instructions to Linux and AIX®, too.

Installing IBM Content Analytics with Enterprise Search with the Embedded Application Server

1. Locate the folder that contains the installation image and double-click launchpad.exe.
2. Ensure your language of choice for installation is selected and click **OK**.

3. Review the Welcome screen. Visit the product site for an overview of the product.
4. Click **Install Product** and then click **Launch IBM Content Analytics with Enterprise Search installation program**.

![IBM Content Analytics with Enterprise Search Version 3.0](image)

**This is the product installation page**

Launch an installation:

- [Launch IBM Content Analytics with Enterprise Search installation program](#)

Installation documentation:

- [View IBM Content Analytics with Enterprise Search installation documentation](#)

The installation program begins:

![InstallAnywhere](image)

(C) 1997-2011 Flexera Software, Inc. and/or InstallShield Co. Inc.
5. Select the language of your choice and click **OK**.
6. To proceed, review and accept the license agreement terms and click **Next**.

You can also click **Print** to print the license terms before proceeding, or click **Cancel** to exit without installing IBM Content Analytics with Enterprise Search.
7. In this step, you create the administrator user name and password. Select the **Create user account** check box and enter a user name and password. Here **esadmin** is given as the default user name.

If you have the user name already created on the machine, you can clear the **Create user account** check box and provide just the user name and password for the use by the installer.

8. From the drop down select the type of installation that you are planning to do:

   - Master: All on one server
   - Master: Distributed server
   - Additional server

See **Supported system configurations under Planning for installation** in the IBM Content Analytics with Enterprise Search Version 3.0 **Information Center** or **Installation Guide** to learn more about the factors to be considered for each server type.
Let’s take a look at some of the Advanced Options.
Note the installation and data directories and the default server ports. Change the paths to a different directory or drive, if needed. This is especially useful if you do not have enough space on the primary drive for data. You can also change the default ports, if needed for your environment.
Note the ports for the administration console, the enterprise search application, and the content analytics miner. By default, the product is installed to use the embedded web application server.

9. Click **Install** to install the software.
The following screens show the stages the installation goes through before the next user input is required.
Installing IBM Content Analytics with Enterprise Search

- Software License
- Installation Options
  - Installing...
  - Installation is Complete
Please wait, IBM Content Analytics with Enterprise Search is being configured for your system. This may take a moment...
Either save other applications or data before selecting **Yes, restart my system** and clicking **Done**, or select **No, I will restart my system myself** and click **Done**, and then close the other applications and manually restart your system.
IBM Content Analytics with Enterprise Search will be configured for your system and then the machine will be rebooted.

Please wait, IBM Content Analytics with Enterprise Search is being configured for your system. This may take a moment...
After the system is restarted and you log in, a splash screen is shown followed by this First Steps window.

11. Click **Start Server** to start the IBM Content Analytics with Enterprise Search sessions.

12. After all of the services are successfully started, click **Close**. If there are any problems during the installation, review the installation logs folder (logs\install) to troubleshoot the problem or contact IBM Software Support for assistance.
Starting the system...
The Common Communication Layer (CCL) service on OEE84-W2K3-1N is running. These servers are available: master, controller, crawler, search
Starting the system sessions [cmd /c call esadmin system start].
The required sessions are running.
Starting the information center on OEE84-W2K3-1N [eselientexec.bat controller, search invokeHTTPWASCommands.bat startC].
The information center on OEE84-W2K3-1N is running.
All services required to use the system are started.
13. Next, click **Verify Installation** on the First Steps screen. The following two sample screens show the various verification steps.

![Verify Installation](image)

- FFQP0045I The system administrator configured for this installation is esadmin.
- FFQP0046I The installation directory configured for this installation is C:\Program Files\IBM\es\esadmin.
- FFQP0047I The data directory configured for this installation is C:\Program Files\IBM\es\esadmin\data.
- FFQP0048I The common communication port configured for this installation is 6002.
- FFQP0049I The controller component is configured on this server.
- FFQP0050I The search component is configured on this server.
- FFQP0051I The crawler component is configured on this server.
- FFQP0052I A single server configuration was detected for this installation.
- FFQP0055I The name of the server with the controller and index components is OEE84-W2K3-1N.
- FFQP0056I Attempting to access the administration console configured at http://OEE84-W2K3-1N:8390/ESAdmin.
- FFQP0062I The search server on port (8390) is active.
- FFQP0070I Attempting to verify the crawler server installation.
- FFQP0071I Attempting to test the database by creating and dropping a table.
- FFQP0072I Attempting to verify the controller and index server installation.
- FFQP0073I Trying to create a collection named SAMPLE_COLLECTION.
- FFQP0074I A collection was created successfully. Collection ID is col_37567.
- FFQP0075I The name of the server with the controller and index components is OEE84-W2K3-1N.
- FFQP0076I Trying to add a crawler to the collection. Crawler ID is col_37567.
- FFQP0077I Trying to add a file system crawler to crawl C:\Program Files\IBM\es\license.
- FFQP0078I A crawler was added successfully to the collection.
- FFQP0079I Trying to create a crawler configuration file.
- FFQP0080I The crawler configuration file was created successfully.
- FFQP0081I Trying to synchronize the system configuration.
- FFQP0082I The name of the server with the controller and index components is OEE84-W2K3-1N.
- FFQP0083I Trying to start the crawler.
Verify Installation

FFQP0105I The search server on port (8390) is active.
FFQP0072I Attempting to verify the crawler server installation.
FFQP0073I Attempting to test the database by creating and dropping a table.
FFQP0074I Attempting to verify the controller and index server installation.
FFQP0011I Trying to create a collection named SAMPLE_COLLECTION.
FFQP0013I A collection was created successfully. Collection ID is col_37567.
FFQP0055I The name of the server with the controller and index components is OEE84-W2K3-1N.
FFQP0015I Trying to add a crawler to the collection. Crawler ID is col_37567.
FFQP0017I Trying to add a file system crawler to crawl C:\Program Files\IBM\es\license.
FFQP0019I A crawler was added successfully to the collection.
FFQP0020I Trying to create a crawler configuration file.
FFQP0021I The crawler configuration file was created successfully.
FFQP0023I Trying to synchronize the system configuration.
FFQP0055I The name of the server with the controller and index components is OEE84-W2K3-1N.
FFQP0024I Trying to start the crawler.
FFQP0026I The file system crawler is now started.
FFQP0029I Trying to start the parser.
FFQP0030I The parser that tokenizes raw data was started successfully.
FFQP0032I Trying to build the index.
FFQP0077I Waiting for the index session to build the index. This might take several minutes.
FFQP0078I The number of documents in the collection is 20.
FFQP0060I Trying to initialize the search client for col_37567.
FFQP0103I The search server on port (8394) is active.
FFQP0062I The search service was started.
FFQP0079I Attempting to search for the term "IBM" in the index.
FFQP0080I The number of documents returned by the search service is 10.
FFQP0082I The search server session for col_37567 was stopped.
FFQP0069I The crawler was stopped.
FFQP0042I Trying to delete a collection that was created as part of the verification program. The collection ID is col_37567.
FFQP0081I The scenarios to verify installation are now complete. Review the displayed output before you use the server.
14. After all the steps are done, click **Close**. If there are any problems or errors, review the installation logs to troubleshoot the problem or contact IBM Software Support.

Review the installation folder structure and the location of the logs, data, master_config, samples and other folders.

The **Know Your Installation** chapter in this cookbook discusses the installation folders.
Accessing the administration console

To access the administration console, type the URL http://<hostname>:8390/ESAdmin, where <hostname> is the host name of the machine on which you installed IBM Content Analytics with Enterprise Search.

You must log in with the user name and password that you specified for the Content Analytics with Enterprise Search administrator in the installation program. After you log in with this default administrator ID the first time, you can add other administrative users and assign administrative roles.
Adding Additional Servers

After you install the master server, you can install additional servers to support search and document processing on separate machines. When you run the installation program on other machines, choose the option Additional Server when you select the Server Type, as shown on page 6 of this document.

Do the following steps to map the additional server to the system topology:

1. On the master server, log in to the administration console.
2. Navigate to the System page and click the Edit (pencil) icon.
3. Click the Server configuration tab and then click the +Add Server button.
4. Provide the host name and port number, and select the additional server’s role.
Installing IBM Content Analytics with Enterprise Search with WebSphere Application Server

This section presents step by step instructions on how to install IBM Content Analytics with Enterprise Search 3.0 with WebSphere Application server as the web application server for the following web applications:

- ESAdmin
- ESAdminRestServer
- ESRestServer
- analytics
- search

1. Locate the folder that contains the installation image and double-click launchpad.exe.
2. Ensure your language of choice for installation is selected and click **OK**.

3. Review the Welcome screen. Visit the product site for an overview of the product.
4. Click **Install Product** and then click **Launch IBM Content Analytics with Enterprise Search installation program**.

The installation program begins:
5. Select the language of your choice and click **OK**.
6. To proceed, review and accept the license agreement terms and click **Next**.

You can also click **Print** to print the license terms before proceeding, or click **Cancel** to exit without installing IBM Content Analytics with Enterprise Search.
7. In this step, you create the administrator user name and password. Select the **Create user account** check box and enter a user name and password. Here `esadmin` is given as the default user name.

If you have the user name already created on the machine, you can clear the **Create user account** check box and provide just the user name and password for the use by the installer.

8. From the drop down select the type of installation that you are planning to do:

   Master: All on one server
   Master: Distributed server
   Additional server

   See *Supported system configurations* under *Planning for installation* in the IBM Content Analytics with Enterprise Search Version 3.0 *Information Center* or *Installation Guide* to learn more about the factors to be considered for each server type.
To install IBM Content Analytics with Enterprise Search to use WebSphere Application Server, you must click **Advanced Options**.
Note the installation and data directories and the default server ports. Change the paths to a different directory or drive, if needed. This is especially useful if you do not have enough space on the primary drive for data. You can also change the default ports, if needed for your environment.
9. Click **WebSphere Application Server**.

   ![WebSphere Application Server Configuration](image)

   Note the ports and URLs for the administration console, the enterprise search application, and the content analytics miner:

   - **http://<host>:<admin port>/ESAdmin**
   - **http://<host>:<search app port>/search**

   If installing with the embedded application server, the default ports are:

   - Admin port = 8390
   - Search port = 8393
   - Search server port = 8394

   If installing with WebSphere Application Server, the requests will go through the HTTP port, which is the port configured in the httpd.conf file of the IBM HTTP Server. See the IBM HTTP Server documentation to learn more about how to configure the HTTP port to be used for this.

   Click **Next** to configure WebSphere Application Server settings.
10. Verify the installation directories of the WebSphere Application Server and the IBM HTTP Server on your system and click **Next**.
11. Verify the fully qualified server name of the WebSphere Application Server along with the profile name to be used, the cell name, the node name, and the web server name as defined in the WebSphere Application Server. The web server name is important to know where to propagate the plugin with information about the web applications that will be installed. Click Next.
12. Provide the Websphere Application Server administrative user name and password. At this point, we assume that the WebSphere Application Server's administrative security is enabled with the InternalFileRepository as the repository for the usernames and passwords. Application security is not yet enabled and we will do that later when configuring global security and integrating WebSphere Application Server with the LDAP server of your choice.

Here are two views of the IBM WebSphere Application Server 7.0 Global Security Configuration pages.
13. After providing the WebSphere Application Server administrative user name and password, click **Next**.
14. Review the installation options confirmation page and click **Install** to begin installation.

During installation, you will see updates such as:

- Installing… Java Runtime Environment
- Extracting duplicates
- Installing… Samples
- Installing… bin
- Installing… webapps
- Installing… docs
- Installing… ibm_help
Note: This step can take 20 – 30 minutes.

You must restart the system to complete the installation.
15. Either save other applications or data before selecting **Yes, restart my system** and clicking **Done**, or select **No, I will restart my system myself** and click **Done**, and then close the other applications and manually restart your system.
After the system is restarted and you log in, a splash screen is shown followed by this First Steps window.

16. Click **Start Server** to start the IBM Content Analytics with Enterprise Search sessions.

17. After all of the services are successfully started, click **Close**. If there are any problems during the installation, review the installation logs folder (logs\install) to troubleshoot the problem or contact IBM Software Support for assistance.
Starting the system...  
The Common Communication Layer (CCL) service on IBM-F838FD44E88.svl.ibm.com is running. These servers are available: master, controller, crawler, search  
The Web server on IBM-F838FD44E88.svl.ibm.com is running.  
The WebSphere application instance on IBM-F838FD44E88.svl.ibm.com is running.  
Starting the system sessions [cmd /c call esadmin system start].  
The required sessions are running.  
Starting the information center on IBM-F838FD44E88.svl.ibm.com [esclientexec.bat controller,search invokeHTTPWASCommands.bat startIC].  
The information center on IBM-F838FD44E88.svl.ibm.com is running.  
All services required to use the system are started.
18. Next, click **Verify Installation** on the First Steps screen.

![IBM Content Analytics with Enterprise Search](image)

19. After all the steps are done, click **Close** in the Verify Installation window. If there are any problems or errors, review the installation logs to troubleshoot the problem or contact IBM Software Support.

   The following two sample screens show the various verification steps.
Verify Installation

FFQP0041I The installation directory configured for this installation is C:\Program Files\IBM\es\esadmin.
FFQP0047I The data directory configured for this installation is C:\Program Files\IBM\es\esadmin\data.
FFQP0048I The common communication port configured for this installation is 6002.
FFQP0049I The controller component is configured on this server.
FFQP0050I The search component is configured on this server.
FFQP0051I The crawler component is configured on this server.
FFQP0052I A single server configuration was detected for this installation.
FFQP0055I The name of the server with the controller and index components is IBM-F838FD44E88.svl.ibm.com.
FFQP0095I Attempting to access the administration console configured at http://IBM-F838FD44E88.svl.ibm.com:80/ESAdmin.
FFQP0105I The search server on port (80) is active.
FFQP0072I Attempting to verify the crawler server installation.
FFQP0073I Attempting to test the database by creating and dropping a table.
FFQP0074I Attempting to verify the controller and index server installation.
FFQP0011I Trying to create a collection named SAMPLE_COLLECTION.

Verify Installation

FFQP0029I Trying to start the parser.
FFQP0030I The parser that tokenizes raw data was started successfully.
FFQP0032I Trying to build the index.
FFQP0077I Waiting for the index session to build the index. This might take several minutes.
FFQP0078I The number of documents in the collection is 20.
FFQP0060I Trying to initialize the search client for col_2552.
FFQP0105I The search server on port (8394) is active.
FFQP0062I The search service was started.
FFQP0079I Attempting to search for the term "IBM" in the index.
FFQP0080I The number of documents returned by the search service is 10.
FFQP0082I The search server session for col_2552 was stopped.
FFQP0069I The crawler was stopped.
FFQP0042I Trying to delete a collection that was created as part of the verification program. The collection ID is col_2552.
FFQP0081I The scenarios to verify installation are now complete. Review the displayed output before you use the server.

Close
Know Your Installation

Now that the product is installed, let’s examine the installation and understand what was installed, where it was installed, and what its significance is.

Administration Console

The administration console is the place where you create and configure collections and crawlers. This is also where you start and stop sessions and perform a wide variety of administration tasks.

Notice the URL format, http://<hostname>/ESAdmin, where <hostname> is the host name of the machine on which you installed IBM Content Analytics with Enterprise Search.

For WebSphere Application Server, this shows that the administration (ESAdmin) console is being accessed via port 80, which is through the HTTP Server. If the system is installed to use the embedded application server, the default port is 8390.

The first time you access the administration console, you must log in with the user name and password that you specified for the Content Analytics with Enterprise Search administrator in the installation program.

After you log in with this default administrator ID the first time, you can add other administrative users and assign administrative roles.
The first screen immediately after you log in provides links to videos and other educational materials that you can use to learn about common administration tasks.

After you close the Welcome page, you will see the Collections page, which shows the list of collections. There is a section for the Enterprise Search collections and one for the Content Analytics collections.
Enterprise Search Application
To access the enterprise search application, type the URL http://<hostname>/search.

The first screen immediately after you start the application provides links to videos and other educational materials that you can use to learn about common search tasks.

After you close the Welcome page, you can enter query terms to perform searches. You can change the collection that you want to search and specify other preferences.
Content Analytics Miner
To access the content analytics miner, type the URL http://<hostname>/analytics.

The first screen immediately after you start the application provides links to videos and other educational materials that you can use to learn about using the content mining visualizations.

After you close the Welcome page, you can select facets and start drilling down through data to discover trends, deviations, and anomalies. You can change the collection that you want to mine and specify other preferences.
ES_INSTALL_ROOT Directory Structure

The ES_INSTALL_ROOT environment variable is set to the location of the root installation folder “es”. In this example it is C:\Program Files\IBM\es. This folder contains the data, configuration, master configuration, help, documentation, executables, samples and other information.

The esadmin folder is the folder named the same as the administrator user name that was created during installation. This folder contains the data, logs, and master_config folders among others.
nodeinfo

The nodeinfo folder contains bldinfo.txt and es.cfg files.
bldinfo.txt
The bldinfo.txt file provides you with information about the version and fix pack level of the
software. Here, for example, the VRCF field identifies the version as the base or GA version of
the software, 3.0.0.0.
es.cfg
The es.cfg file is an important configuration file that holds the username and encrypted password for the default administrator user ID. It also holds the user name and encrypted password of the WebSphere Application Server user ID that will act as the Search Administrator after global security is configured. Review a copy of this file to understand other settings of the installed system.
uninstall_3.0.0.0
The uninstall_3.0.0.0 folder contains the uninstall_3.0.0.0.exe file that helps you uninstall the Version 3.0 software. With the installation of additional fix packs, additional folders such as uninstall_3.0.0.1 will be created. The latest fix pack must be uninstalled before uninstalling the older fix packs before you can uninstall the GA version.
**bin**

The bin folder contains a number executables essential to run the system, such as the esadmin command. In the following image, notice the esservice.bat and esviewlog.bat files. Ensure that your operating system’s PATH environment variable points to this folder to allow execution of these commands without having to enter this folder.
Default Administrator ID folder
The folder for the default administrator matches the user ID that you specified for the IBM Content Analytics with Enterprise Search administrator during installation. This folder contains the log files, this node’s configuration, the data folder, and so on. On the master server, this folder also contains the master_config folder, which contains the master configuration files that are copied by the configuration manager to each of the additional nodes.
**master_config**

The master_config folder has folders for various system-level sessions and collection-level sessions such as crawlers. Additionally, there are configuration files (.ini or .xml) for various components, such as collections.ini (for collection properties), services.ini (listing various services), nodes.ini (listing all the nodes and their functions), esldap.xml (containing LDAP and Global Security configuration information for the embedded application server), and esltpa.xml (SSO token configuration for the embedded application server).
**Logs**
The logs folder contains the system-level, collection-level and audit-level log files. The audit level logs are present in the audit folder. Installation logs can be found in the install folder. The logs are in an encoded format. To read them, you must use esviewlog.bat to format them into a readable format.

Additionally you can enable detailed traces. The trace files are stored in the folder mentioned in the instructions for the detailed tracing. For more information about detailed tracing, see the technical notes on the IBM Content Analytics with Enterprise Search support site.
**ESService**

The esservice.bat script (on Windows) and esservice.sh script (on Linux and AIX) helps you to gather all the necessary information from each server to send to IBM Software Support for problem determination. It gathers all of the configuration data, log files, and other information and archives the data into a ZIP file. A simple execution of esservice, as shown in the following figure, adds all files in the logs folder to a ZIP file.

To exclude javacore files and heapdump files, use the –nocore –noheapdump option with the esservice command:

```
eeservice.bat --nocore --noheapdump
```
When the esservice zip file is extracted, it contains the folders shown in the above image. The service.log file logs the operations during the archiving process. The os_system.out file contains all the information about the machine and its operating system as well as other things like the Java version.
**Troubleshooting**

1. Search Server cannot be accessed.

To resolve this problem:
   - Verify that WebSphere Application Server is running.
   - Verify that there are no other WebSphere Application Server instances running at the same time.
2. The collection is not available.

To resolve this problem, ensure that there is at least one collection configured in the administration console and that the collection’s search runtime is running.