

Content Manager

Software Version 9.4

Elasticsearch Integration Installation and Configuration Guide



Document Release Date: March 2021
Software Release Date: July 2022

Legal notices

Copyright notice

© Copyright 2017-2022 Micro Focus or one of its affiliates.

The only warranties for products and services of Micro Focus and its affiliates and licensors (“Micro Focus”) are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Micro Focus shall not be liable for technical or editorial errors or omissions contained herein. The information contained herein is subject to change without notice.

Adobe™ is a trademark of Adobe Systems Incorporated.

Microsoft® and Windows® are U.S. registered trademarks of Microsoft Corporation.

UNIX® is a registered trademark of The Open Group.

This product includes an interface of the 'zlib' general purpose compression library, which is Copyright © 1995-2002 Jean-loup Gailly and Mark Adler.

Documentation updates

The title page of this document contains the following identifying information:

- Software Version number, which indicates the software version.
- Document Release Date, which changes each time the document is updated.
- Software Release Date, which indicates the release date of this version of the software.

You can check for more recent versions of a document through the [MySupport portal](#). Many areas of the portal, including the one for documentation, require you to sign in with a Software Passport. If you need a Passport, you can create one when prompted to sign in.

Additionally, if you subscribe to the appropriate product support service, you will receive new or updated editions of documentation. Contact your Micro Focus sales representative for details.

Support

Visit the [MySupport portal](#) to access contact information and details about the products, services, and support that Micro Focus offers.

This portal also provides customer self-solve capabilities. It gives you a fast and efficient way to access interactive technical support tools needed to manage your business. As a valued support customer, you can benefit by using the MySupport portal to:

- Search for knowledge documents of interest
- Access product documentation
- View software vulnerability alerts
- Enter into discussions with other software customers
- Download software patches
- Manage software licenses, downloads, and support contracts
- Submit and track service requests
- Contact customer support
- View information about all services that Support offers

Many areas of the portal require you to sign in with a Software Passport. If you need a Passport, you can create one when prompted to sign in. To learn about the different access levels the portal uses, see the [Access Levels descriptions](#).

Contents

Getting started with Elasticsearch	4
Environment overview	4
Install Content Manager 9.4	6
Upgrading from earlier versions	6
New Installations	6
Install Elasticsearch	7
Install Elasticsearch on your nominated servers	7
Edit the Elasticsearch configuration file	7
Memory Management and configuring the Java Virtual Machine (JVM)	10
Configure Elasticsearch to run as a service	12
Run Elasticsearch	12
Adjusting the JVM heap size after installing Elasticsearch as a service	13
Confirm the Elasticsearch service and cluster are operational	14
Configure an Elasticsearch Content Index	15
Reindex your document store with Elasticsearch	15
Configure Content Manager Enterprise Studio to use Elasticsearch	20
Remove existing IDOL content indexes	20
Create a new Elasticsearch index	21
Confirm your Elasticsearch content index is operational	26
Troubleshooting common issues	27
Logging	27
Diagnosing “operation timed out” errors	27
Proxy servers and firewalls	28
Useful Information	29
Kibana	29
cURL	29
Elasticsearch Reference	29
Hardware requirements	29

Getting started with Elasticsearch

Content Manager 9.4 provides support for integration with Elasticsearch for document content indexing. Customers have the option to choose either IDOL or Elasticsearch for this functionality.

Elasticsearch is an open source search and content analytics engine built on top of Apache Lucene™.

Unlike IDOL, Elasticsearch is a third-party product not shipped with Content Manager (i.e. non-OEM). Instead, organisations choosing to use Elasticsearch will be required to download it from elastic.co.

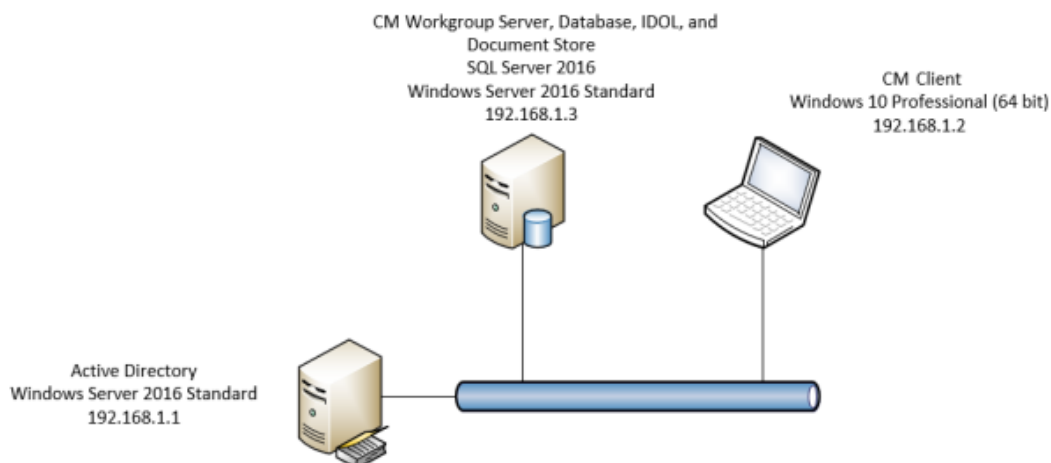
This guide will help you understand what's required to establish a basic Content Manager 9.4 and Elasticsearch instance after upgrading from a previously supported release or installing Content Manager 9.4 for the first time.

Before you begin, it's important to note Elasticsearch can be configured and scaled in many different ways. Specialist Elasticsearch configuration advice and support cannot be provided by the Content Manager team. If, after reading this guide, you believe your organization requires specialist Elasticsearch configuration advice and support, you should consult the Elastic team directly regarding your requirements.

Environment overview

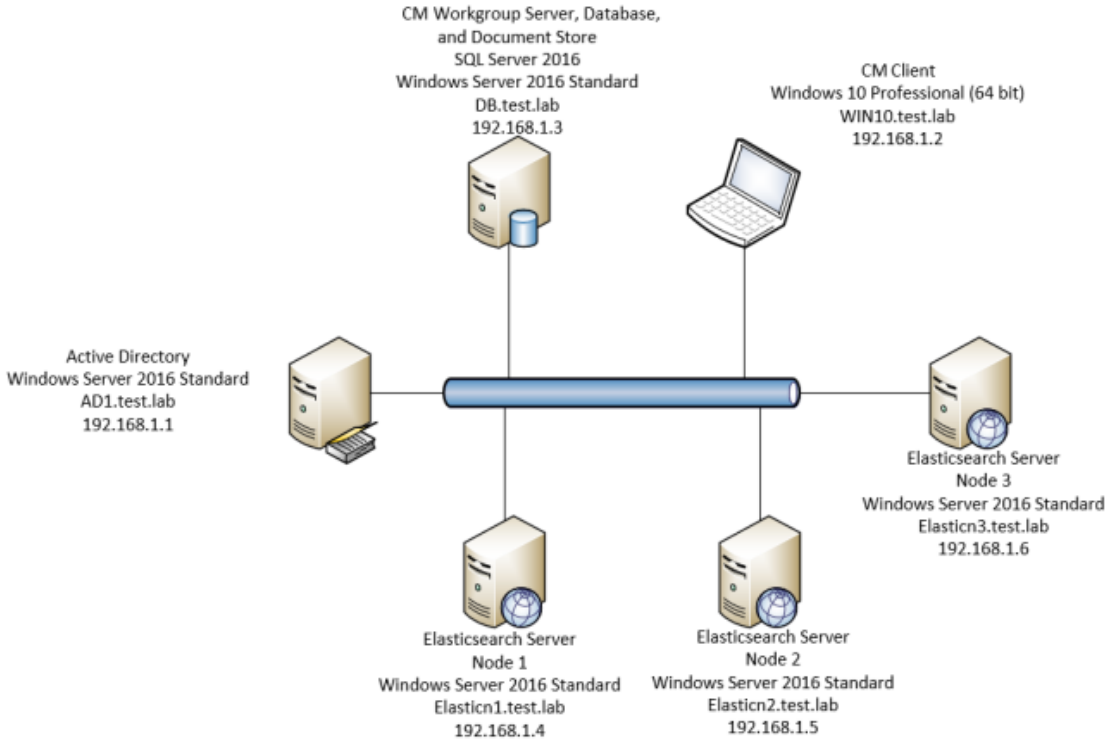
This guide will begin by running you through a scenario where a customer is upgrading from Content Manager 9.3 and IDOL to Content Manager 9.4 and Elasticsearch.

The starting environment will be a simple Content Manager 9.3 and IDOL implementation. We assume you already know how to install Content Manager 9.4 as outlined in Installation Guide (**CM9.4_Install.pdf**), for example:



Whether you're upgrading or doing a new installation the resulting Content Manager 9.4 and Elasticsearch environment, in this example scenario, will include three additional "node" servers in an

Elasticsearch cluster. This is inline with the Elastic team's best practice guidelines regarding the minimum number of nodes you should have in a production cluster:



Install Content Manager 9.4

Upgrading from earlier versions

If you are running an earlier version of Content Manager we recommend upgrading to Content Manager 9.4 and IDOL environment before attempting to install and configure Elasticsearch. This will allow you to confirm your environment and content searches are working as expected prior to reconfiguring your environment to use Elasticsearch integration instead of IDOL.

Follow the instructions outlined in the Installation and Setup Guide (**CM9.4_Install.pdf**) to upgrade your environment to Content Manager 9.4.

New Installations

If you're installing Content Manager 9.4 for the first time you should do so as outlined in the Installation and Setup Guide (**CM9.4_Install.pdf**) before you implement Elasticsearch.

Install Elasticsearch

When you have upgraded to or installed Content Manager 9.4 you can proceed to installing Elasticsearch on each of the servers that will make up the cluster.

We do not recommend installing Elasticsearch on a Workgroup Server as it can use ports that are used by Content Manager or IDOL, and adds unnecessary complexity when you're trying to troubleshoot a problem in the future. The Elastic team are very clear on keeping it simple. We suggest you follow this approach when designing and setting up your environment.

Elastic.co recommend you dedicate a separate server for each Elasticsearch node as we've done in the above network diagram.

The Elastic team recommend a minimum of three nodes for production environments so that if any single node fails no data will be lost.

Install Elasticsearch on your nominated servers

Download Elasticsearch 7.x from elastic.co and unzip it your preferred directory.

In this example, we downloaded the zip package and unzipped it to C:\Elasticsearch\ on Elasticn1, Elasticn2, and Elasticn3.

Edit the Elasticsearch configuration file

Elasticsearch is configured via a configuration file located in the directory you've just unzipped it to. In this scenario it's located at C:\Elasticsearch\elasticsearch-7.x\config\elasticsearch.yml

You can open, edit, and save this file using a text editor like TextPad or Notepad.

The most common settings are already in this file and you'll find they're commented out via the # symbol at the start of each line. The following table lists the ones that are most commonly used:

node.name - Name of the server node. This name is just for reference. If you leave this blank Elasticsearch will choose a random default name for you.

cluster.name - For multi node systems, each node needs to have the same cluster name. All nodes with the same cluster will use a network discovery protocol to find the other nodes in the cluster, and automatically configure themselves. The default name is *elasticsearch*, but you should change it to an appropriate name which describes the purpose of the cluster. A cluster does not become active until this is set.

path.data - Path where Elasticsearch stores the index data. The default is the 'data' directory in the Elasticsearch folder.

path.logs - Path where Elasticsearch stores the log files. The default is the 'log' directory in the Elasticsearch folder.

network.host - The node will bind to this hostname or IP address and publish (advertise) this host to other nodes in the cluster. Accepts an IP address, hostname, a special value, or an array of any combination of these.

discovery.seed_hosts - Out of the box, without any network configuration, Elasticsearch will bind to the available loopback addresses and will scan local ports 9300 to 9305 to try to connect to other nodes running on the same server. This provides an auto-clustering experience without having to do any configuration. When you want to form a cluster with nodes on other hosts, you must use the `discovery.seed_hosts` setting to provide a list of other nodes in the cluster that are master-eligible and likely to be live and contactable in order to seed the discovery process. This setting should normally contain the addresses of all the master-eligible nodes in the cluster. This setting contains either an array of hosts or a comma-delimited string.

cluster.initial_master_nodes - When you start a brand new Elasticsearch cluster for the very first time, there is a cluster bootstrapping step, which determines the set of master-eligible nodes whose votes are counted in the very first election. In development mode, with no discovery settings configured, this step is automatically performed by the nodes themselves. As this auto-bootstrapping is inherently unsafe, when you start a brand new cluster in production mode, you must explicitly list the master-eligible nodes whose votes should be counted in the very first election. This list is set using the `cluster.initial_master_nodes` setting.

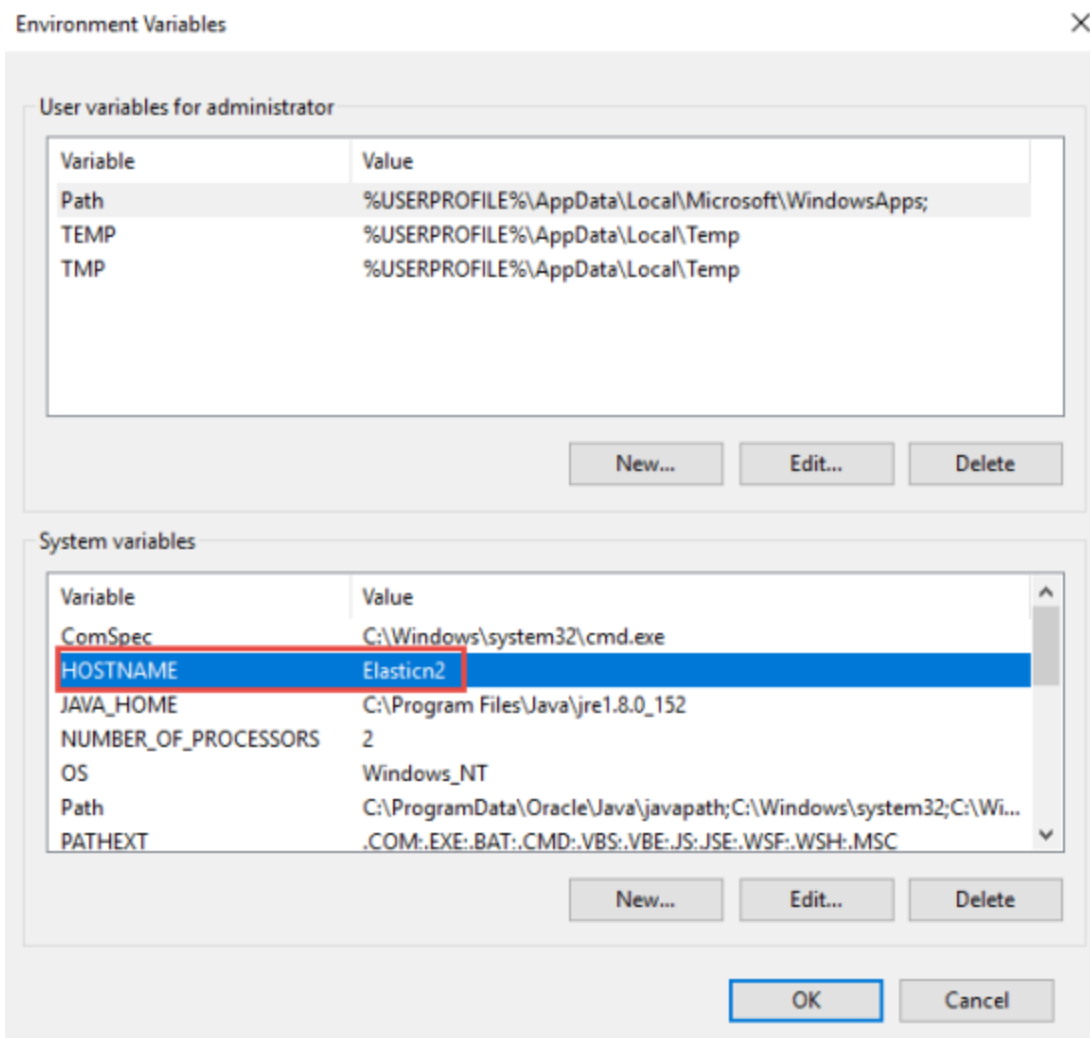
To setup the three-node cluster in our environment we'll be changing these settings in the configuration file on each server:

node.name - We'll set this to be the server's hostname by removing the # and using the entry `${HOSTNAME}`

```
# ----- Node -----  
#  
# Use a descriptive name for the node:  
#  
node.name: ${HOSTNAME}  
#  
# Add custom attributes to the node:  
#  
#node.attr.rack: r1  
#
```

To use this entry, you'll also have to enter the `HOSTNAME` environment variable like you did for `JAVA_HOME`.

This will be `Elasticn1`, `Elasticn2`, and `Elasticn3` on each of the node servers.



cluster.name - We'll set this by removing the # and using the entry contentmanager

```
# ----- Cluster -----  
#  
# Use a descriptive name for your cluster:  
#  
cluster.name: contentmanager
```

network.host - We'll set this by removing the # and using the entry `_site_` which will set the value to any site-local addresses on the system, for example 192.168.1.4

```
# ----- Network -----  
#  
# Set the bind address to a specific IP (IPv4 or IPv6):  
#  
network.host: _site_  
#  
# Set a custom port for HTTP:  
#  
#http.port: 9200  
#  
# For more information, consult the network module documentation.  
#
```

discovery.seed_hosts - We'll set this by removing the # and replacing the host entries with the IP addresses of the three node servers that will make up the cluster (192.168.1.4, 192.168.1.5, and 192.168.1.6). We could have also used hostnames instead of IP addresses (Elasticn1, Elasticn2, and Elasticn3).

```
# ----- Discovery -----  
#  
# Pass an initial list of hosts to perform discovery when this node is started:  
# The default list of hosts is ["127.0.0.1", "[::1]"]  
#  
discovery.seed_hosts: ["192.168.1.4", "192.168.1.5", "192.168.1.6"]  
#
```

TIP: If you change these settings after your environment is operational you'll have to stop and start the Elasticsearch service to implement the changes.

Memory Management and configuring the Java Virtual Machine (JVM)

Getting your memory allocation configured correctly is critical to the successful operation of Elasticsearch. You must size your hardware and configure your environment appropriately to handle the indexing load your organization will be generating.

There are three memory variables you have to consider when setting up Elasticsearch:

- **Installed RAM**

The Elastic team advise a server that has 64 GB of RAM is the ideal amount in a production environment. The Elastic team do not recommend having less than 8 GB of RAM as it can lead to a need for many small servers.

- **Heap memory**

This is the memory allocated to the JVM in which Elasticsearch runs. The Elastic team recommend giving 50% of the available memory to Elasticsearch heap, while leaving the other 50% free.

When setting the heap memory, it is also recommended you do not allocate more than 26 GB. This is due to a Java limitation that becomes apparent at 32 GB and can have a negative impact on performance. This limitation is described in detail in the **Heap: Sizing and Swapping** section of **Elasticsearch: The Definitive Guide** on the elastic.co web page.

- **Off heap memory**

This is the portion of memory used by the operating system and Lucene (the underlying engine used by Elasticsearch) outside of the JVM heap.

When configuring the JVM heap you should ensure there's enough memory left to accommodate off heap processes. In line with the above recommendation, this should be around 50% of the available memory.

Elasticsearch provides a way to manage JVM settings via the `jvm.options` file which has the default heap size set to 1 GB. You can open, edit, and save this file using text editor like TextPad or Notepad. You'll find this file in:

`<Drive>:\<Elasticsearch installation directory>\config\`

In our test environment, we'll be setting the initial and maximum heap size to 1g (1 GB).

```
#####
## IMPORTANT: JVM heap size
#####
##
## You should always set the min and max JVM heap
## size to the same value. For example, to set
## the heap to 4 GB, set:
##
## -Xms4g
## -Xmx4g
##
## See https://www.elastic.co/guide/en/elasticsearch/reference/current/heap-size.html
## for more information
##
#####

# Xms represents the initial size of total heap space
# Xmx represents the maximum size of total heap space
```

```
-Xms1g
-Xmx1g
```

Because of a bug in Elasticsearch 7.2 that doesn't set up the Java configuration properly when installing Elasticsearch as a service at the next step, we'll also explicitly specify a path for the Java IO temp location. Also in the `jvm.options` file set the following parameter to your choice of temporary location, for example

Change this line from the default value

```
-Djava.io.tmpdir=${ES_TMPDIR}
```

To

```
-Djava.io.tmpdir=C:\Windows\Temp
```

NOTE: Path cannot contain spaces.

```
# log4j 2
-Dlog4j.shutdownHookEnabled=false
-Dlog4j2.disable.jmx=true
```

```
-Djava.io.tmpdir=C:\Windows\Temp
```

```
## heap dumps
```

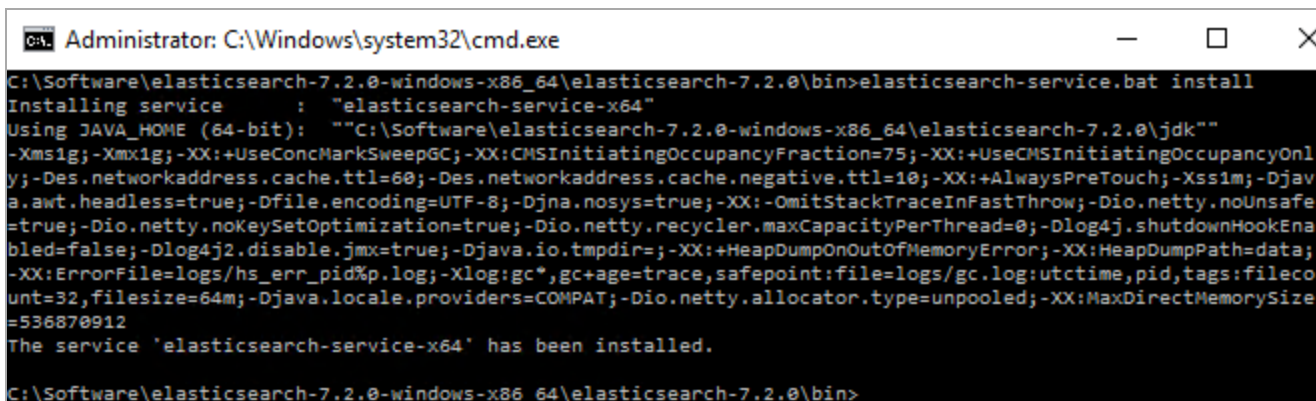
Save your changes to `jvm.options` before proceeding.

Configure Elasticsearch to run as a service

You can use the supplied batch script `<Drive>:\<Elasticsearch installation directory>\bin\elasticsearch-service.bat` to install Elasticsearch as a Windows service.

To do this, open an administrative command prompt, navigate to the `bin` directory, and run the following command on each Elasticsearch server:

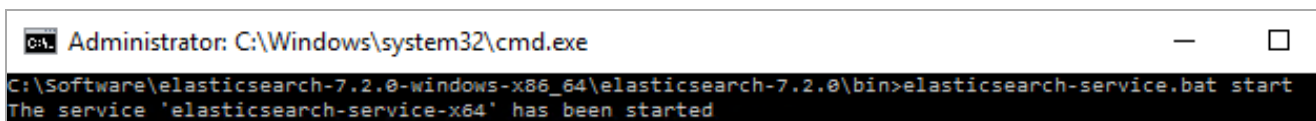
```
elasticsearch-service.bat install
```



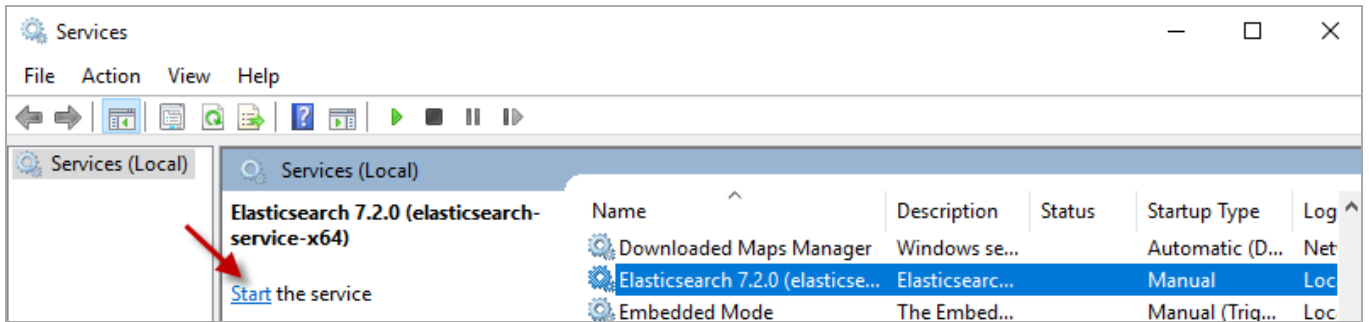
Run Elasticsearch

Start the Elasticsearch service on each server via an administrative command prompt using the following command:

```
elasticsearch-service.bat start
```



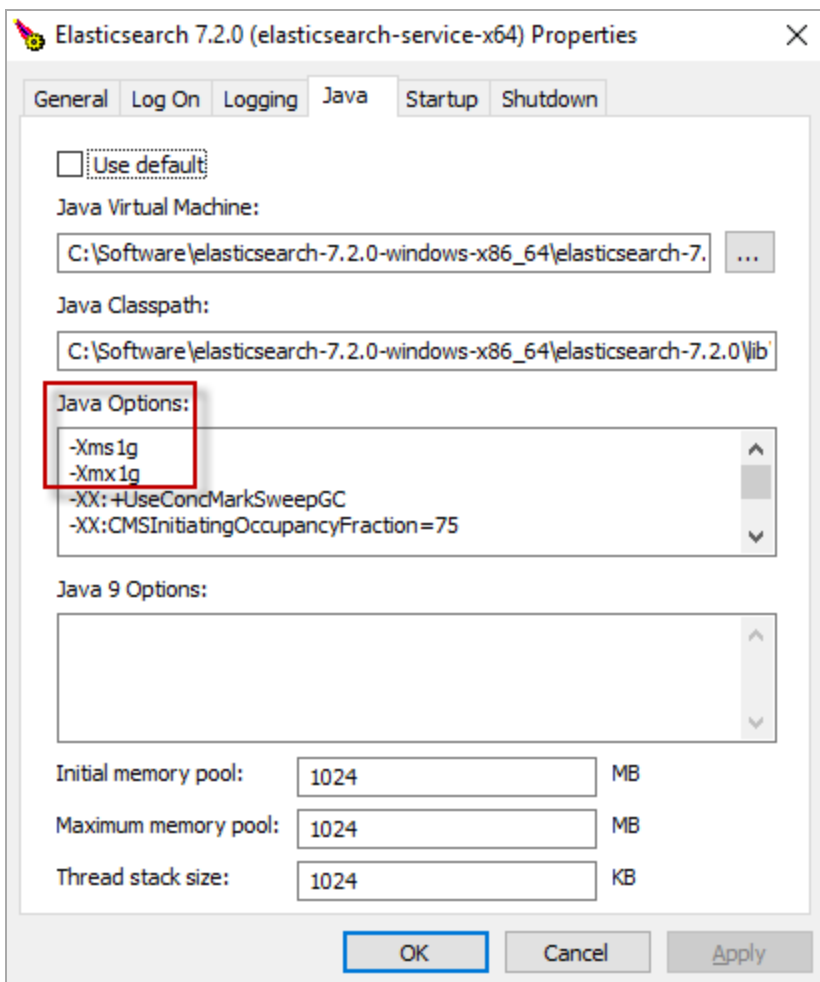
You can also start the Elasticsearch service via the Windows Services Management Console (Start Button > Run > `services.msc`)



Adjusting the JVM heap size after installing Elasticsearch as a service

To adjust the JVM heap size after installing Elasticsearch as a service you'll need to use the service manager GUI located in the ...\\bin directory. Editing `jvm.options` will not pass on heap changes to the service once it's installed.

You can open the service manager window by invoking `elasticsearch-service.bat` manager from the command line:



Any changes you make via the service manager will require you to re-start the service which can be done from the **General** tab.

Confirm the Elasticsearch service and cluster are operational

Check the health of your cluster by opening `http://<ip address or hostname>:9200/_cat/health?v` in a browser. The IP address or hostname used can be any Elasticsearch server in your environment.

Elasticsearch uses a traffic-light scheme for server health: green, yellow if there are warnings (e.g. disk space more than 85% full), and red for serious errors.

The Elasticsearch service can take some time to start if you've had a reason to restart it e.g. after changing configuration values. During this period, you may see the health status as **red**. If this occurs, you should wait until the service is finished its start-up sequence and try the health check again.



Configure an Elasticsearch Content Index

Content Manager 9.4 will only allow you to have one type of content indexing engine configured. If you upgrade from earlier versions your existing IDOL content index will continue to be available in 9.4.

However, if you decide to use Elasticsearch as your content index your IDOL content index will no longer be available.

NOTE: If you decide to go back to an IDOL content index, you will need to delete your Elasticsearch content index and create a new IDOL index. See **Setting up document content indexing and searching** in the Content Manager Enterprise Studio help file.

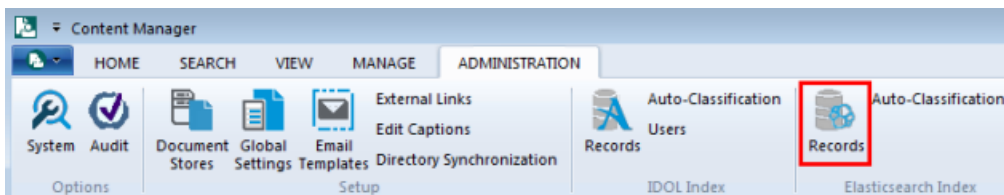
Now we have a working Elasticsearch environment in place we can move on to enabling it within Content Manager 9.4.

Reindex your document store with Elasticsearch

If you're setting up a new instance of Content Manager there's no need to perform this step yet. You can go straight to [Create a new Elasticsearch index, on page 21](#) and follow the steps to setup a new Content Index in Enterprise Studio.

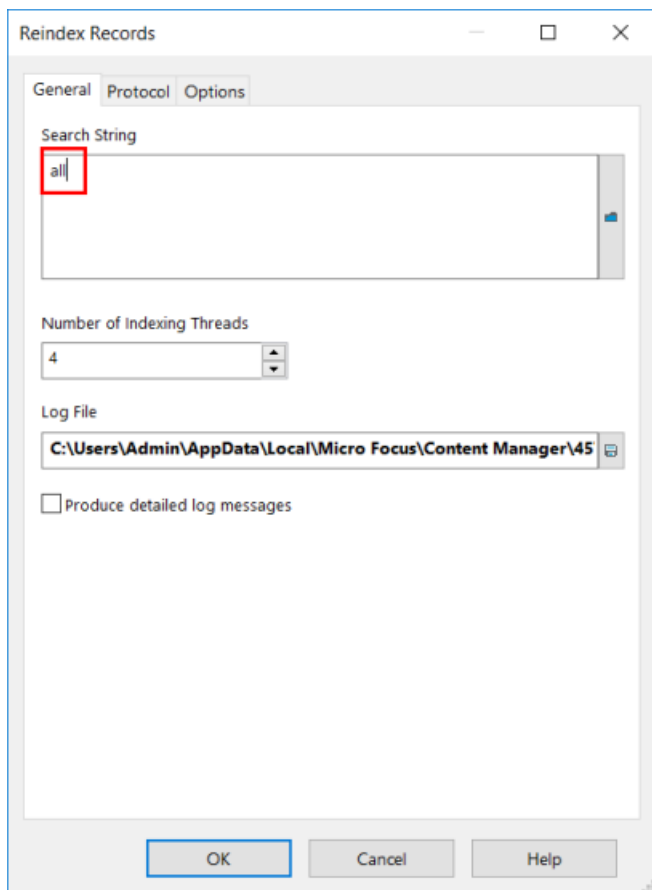
Open the Content Manager client and select the **Administration** tab.

From the **Administration** tab, in the **Elasticsearch Index** group, click **Records**.



The **Reindex Records** dialog will appear.

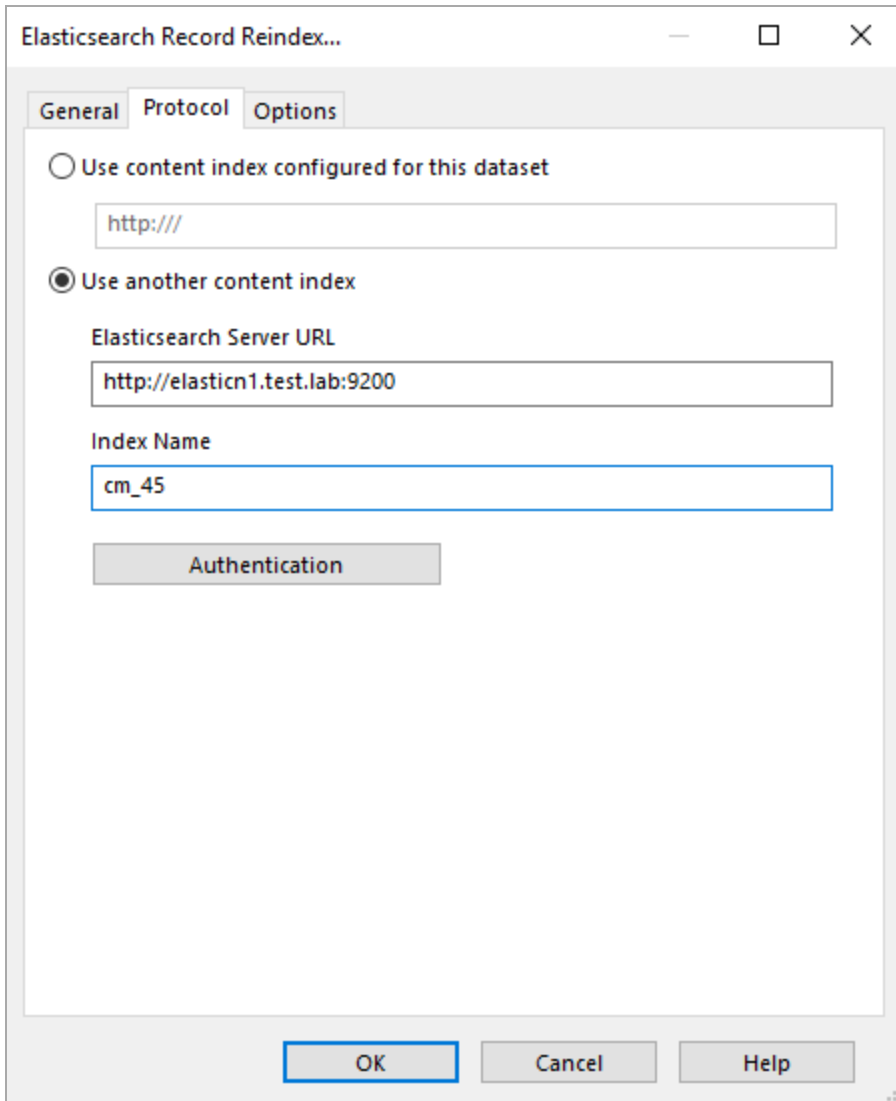
In the **Search String** field on the **General** tab, type all.



IMPORTANT: To successfully complete a full re-index, make sure you are using a Content Manager location profile that has sufficient access to bypass all current security settings you may have in place.

You also have the opportunity to potentially improve performance by increasing the **Number of Indexing Threads**. This is largely dependent on your hardware. You'll have to trial a number of different thread settings to find out exactly how many threads it's capable of handling. For this scenario we'll use the default value of 4.

Because we haven't configured an Elasticsearch index yet in CM Enterprise Studio, select **Use another content index** on the **Protocol tab** and in the **Elasticsearch server URL** field type the URL for the Elasticsearch server. This can be any Elasticsearch server in the cluster.



Click **OK** and the re-indexing process will start.

The screenshot shows a window titled "Reindex Records" with a close button in the top right corner. Below the title bar, the text "Indexing threads" is displayed. A table lists the progress of four indexing threads. The table has columns for Thread, Status, Processed, Bytes Processed, Errors, Warnings, and Current Item. Below the table is a horizontal scrollbar. Underneath the table, there is a log entry with columns for Thread, Date/Time, and Message. At the bottom of the window, there are four buttons: "View Log", "Pause", "Cancel", and "Help".

Thread	Status	Processed	Bytes Processed	Errors	Warnings	Current Item
1	Running	40	13.353Kb	0	0	92/4920(132)
2	Running	40	13.872Kb	0	0	92/6935(142)
3	Running	40	13.279Kb	0	0	93/2670(152)
4	Running	40	13.767Kb	0	0	94/292(162)
TOTAL		44%	160	54.271Kb	0	0

Thread	Date/Time	Message
	26/07/2019 4:22:46 PM	Start processing

View Log Pause Cancel Help

When the re-indexing process is complete the Complete status will be displayed along with any errors it encountered.

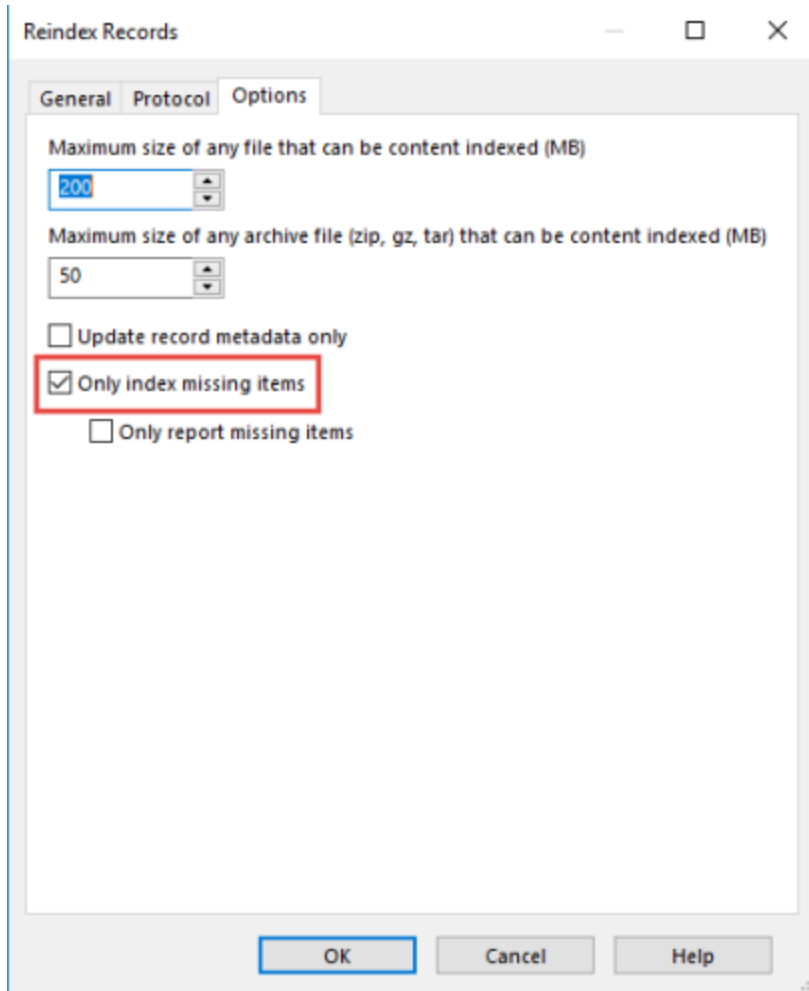
The screenshot shows a window titled "Reindex Records" with a close button (X) in the top right corner. Below the title bar, the text "Indexing threads" is displayed. A table summarizes the progress of four threads. Below this table is a horizontal scrollbar. Underneath is a log of messages for each thread, including completion and start messages. At the bottom of the window are four buttons: "View Log", "Pause", "Close", and "Help".

Thread	Status	Processed	Bytes Processed	Errors	Warnings	Current Item
1	Complete	89	36.750Kb	0	0	
2	Complete	100	41.716Kb	0	0	
3	Complete	79	73.171Kb	0	0	
4	Complete	89	40.917Kb	0	0	
TOTAL	100%	357	192.554Kb	0	0	

Thread	Date/Time	Message
	26/07/2019 4:23:03 PM	Processing complete.
	26/07/2019 4:23:03 PM	All threads stopped.
3	26/07/2019 4:23:02 PM	Bulk transaction completed. (89123 bytes)
2	26/07/2019 4:23:02 PM	Bulk transaction completed. (59998 bytes)
4	26/07/2019 4:23:01 PM	Bulk transaction completed. (57747 bytes)
1	26/07/2019 4:23:01 PM	Bulk transaction completed. (52943 bytes)
	26/07/2019 4:22:46 PM	Start processing

If there are errors, click **View Log** to get additional information.

To rectify this, we can re-run the re-index but this time only index those records that are missing by selecting the **Only index missing items** on the **Options** tab.

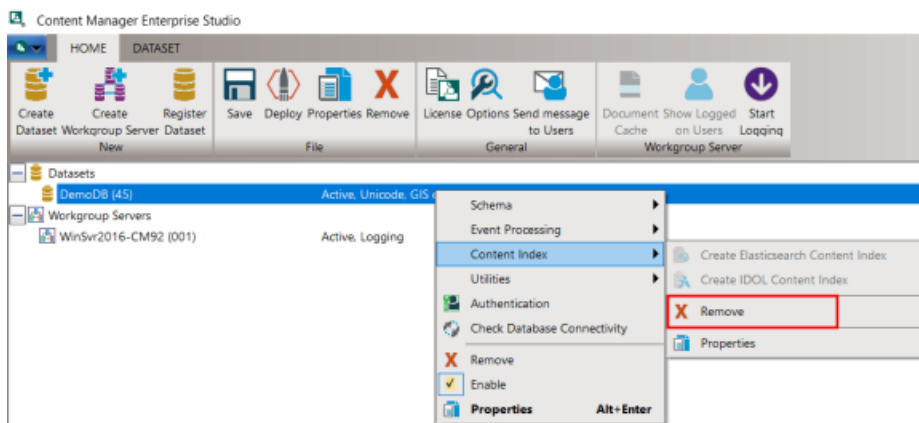


Configure Content Manager Enterprise Studio to use Elasticsearch

Remove existing IDOL content indexes

Now we've successfully created an Elasticsearch content index the existing IDOL index can be removed.

On your Workgroup Server open Enterprise Studio -> right mouse click on your Dataset -> **Content Index** -> **Remove**.



Select **Delete Content Index from Server** then select **OK**.

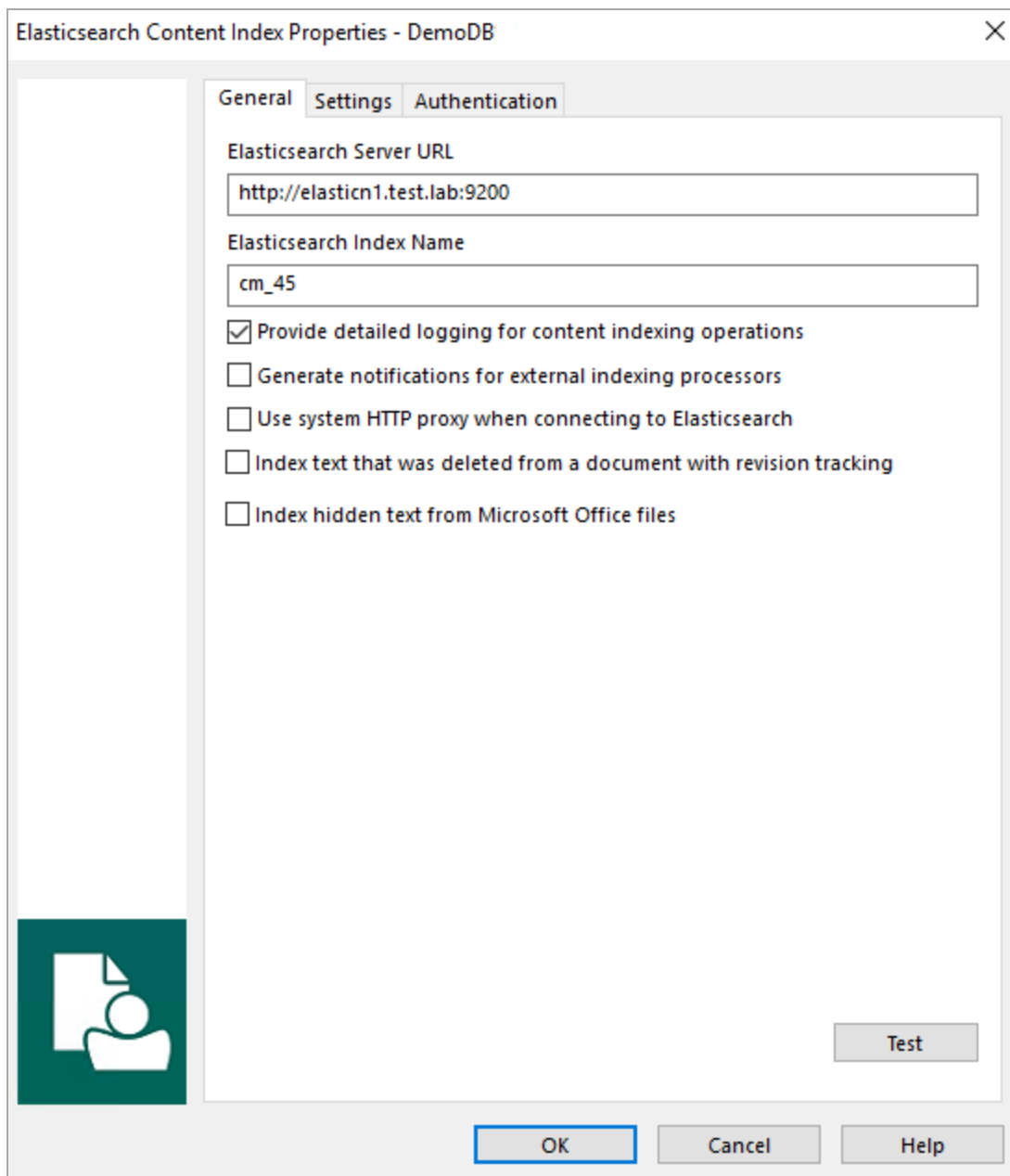
Create a new Elasticsearch index

On your Workgroup Server open Enterprise Studio -> right mouse click on your Dataset -> **Content Index** -> **Create Elasticsearch**.

Fill in the fields on the Content index properties dialog:

On the **General** tab:

- **Elasticsearch Server URL** - enter the Elasticsearch Server URL. If using the X-Pack authentication, specify the https-based URL for the ES server here.
- **Elasticsearch Index Name** - Name of the Elasticsearch index to use for this CM dataset. The characters in the name must be alphanumeric characters or underscores. If you performed a reindex at an earlier step this would be the same value used for the Index name during that process.
- **Provide detailed logging for content indexing operations** - select this option to generate detailed log files.
- **Generate notifications for external indexing processors** - if you are configured to use Elasticsearch but you wish for a CFS connector to be notified of indexing events, then enable this option. This is for installations that use ControlPoint.
- **Use system HTTP proxy when connecting to Elasticsearch** - select this option if your organization needs to connect to Elasticsearch using a proxy.
- **Index text that was deleted from a document with revision tracking** - select this option to index content that was deleted from a document that was edited using revision tracking/ "tracked changes".
- **Index hidden text from Microsoft Office files** - select this option to index content that is marked 'hidden' in documents, for example, hidden cells in Microsoft Excel.

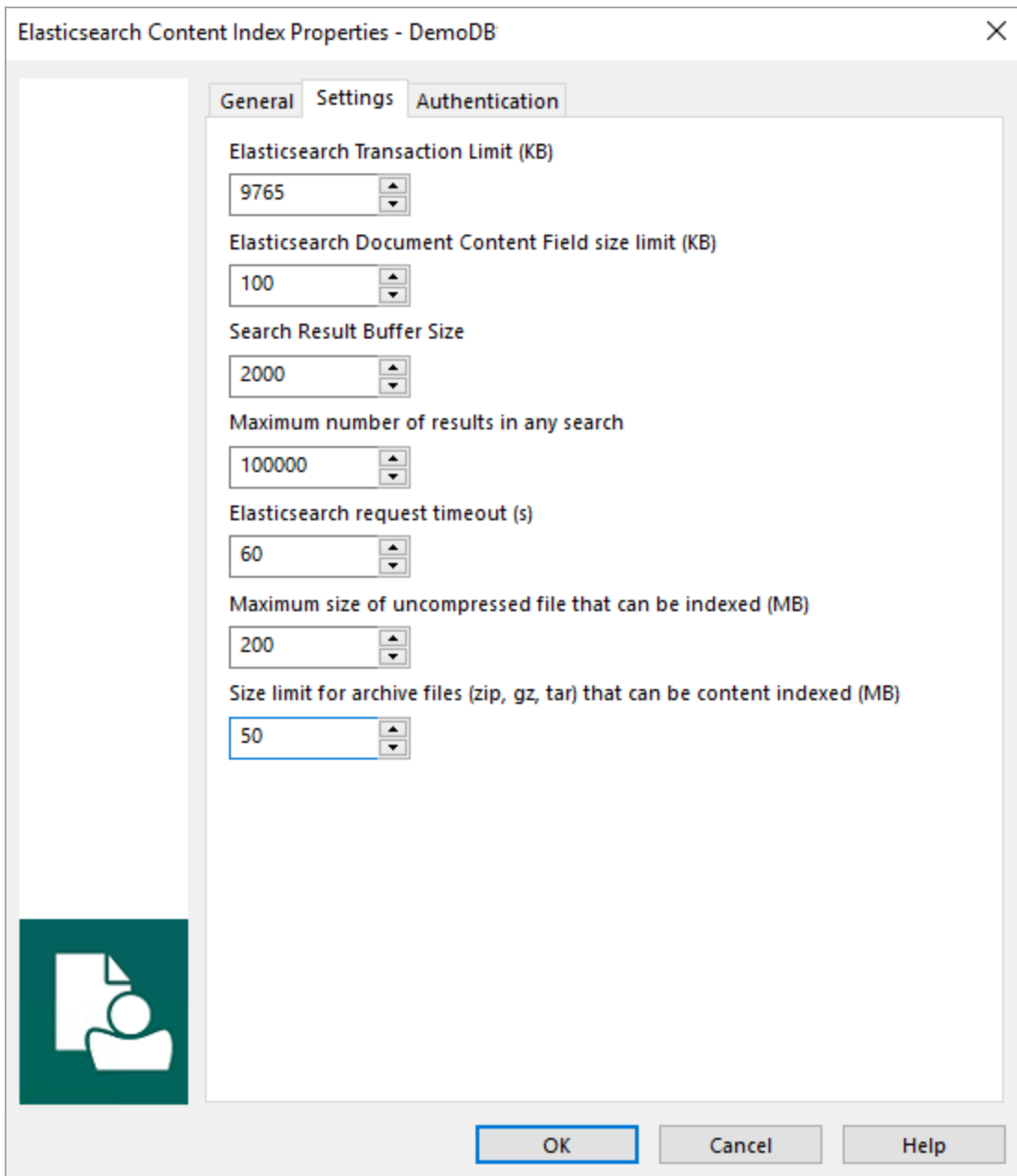


On the **Settings** tab:

- **Elasticsearch Transaction Limit (MB)** - Limits how much indexing data can be sent to Elasticsearch in a single request.
- **Elastic Document Content Field size limit (KB)** – The maximum amount of textual data that can be indexed as the content of a single document.
- **Search Result Buffer Size** - default value of 2000. For customers running searches that retrieve many results, increasing the buffer size will reduce the number of service calls to the Elasticsearch server to retrieve all the results.
- **Maximum number of results in any search** - The maximum number of results Content Manager

will get back from Elasticsearch, by default, this is set to 100,000. If the number of results found by a search exceeds this number, a warning displays and the user only gets this number. The maximum value that can be set is 10,000,000.

- **Elasticsearch Request Timeout (s)** - Limits how long CM will wait for Elasticsearch to process a single request.
- **Maximum size of uncompressed file that can be indexed (MB)** - default: 200. Content Manager does not index the content of documents whose file size is greater than this number in megabytes.
- **Size limit for archive files (zip, gz, tar) that can be content indexed (MB)** - default: 50. Content Manager does not index the content of archive files whose file size is greater than this number in megabytes.

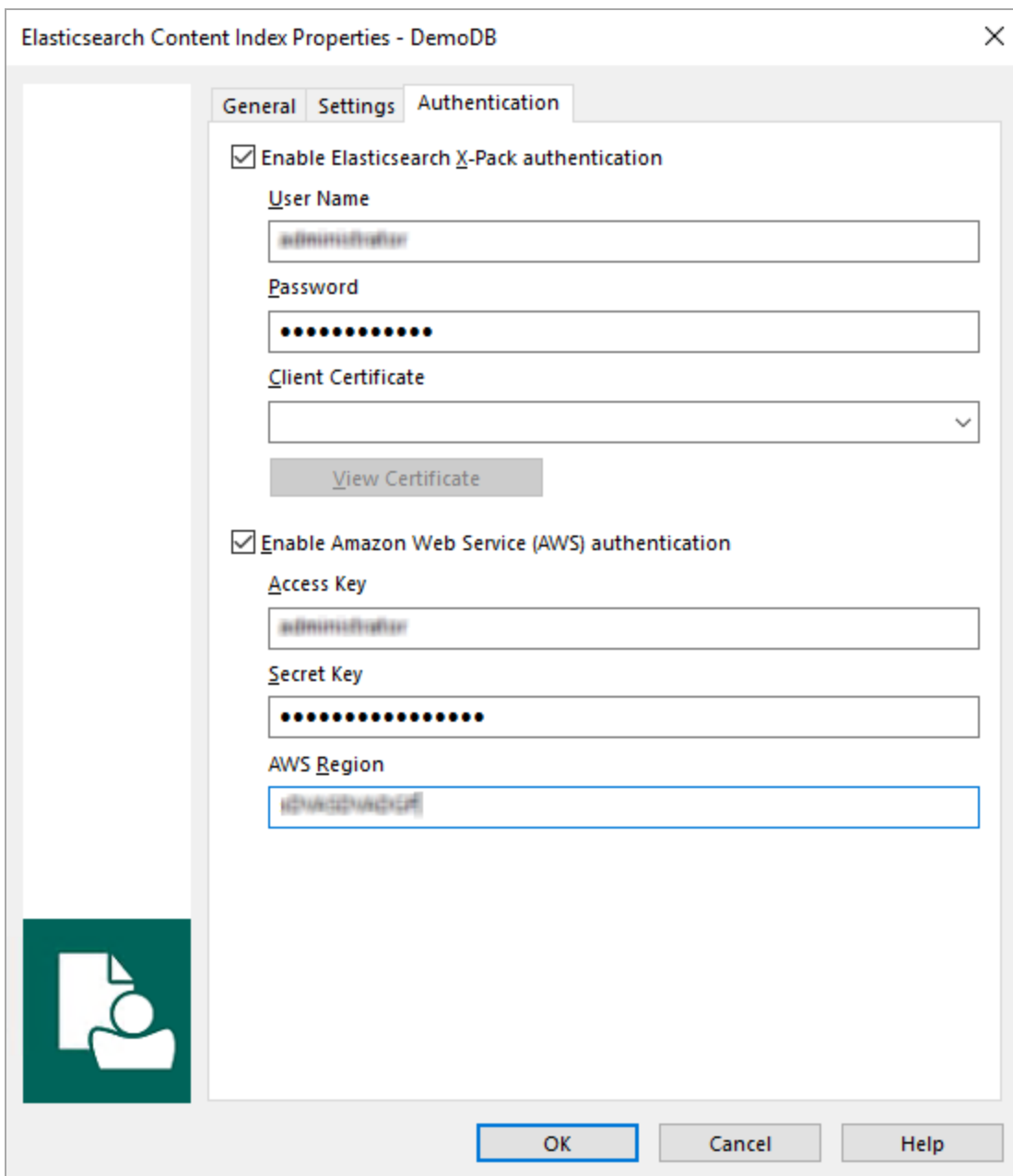


On the **Authentication** tab, fill in the fields:

- **Enable Elasticsearch X-Pack authentication** - select this option to enable X-Pack authentication if X-Pack authentication is being used. You can either specify a username and password or a certificate:
 - **User Name** - enter the username to connect to the Elasticsearch server.
 - **Password** - enter the password for the defined user.
 - **Client Certificate** - if using a certificate for the X-Pack authentication, the certificate must be installed to the Personal store of the Local Computer account. Once installed, the certificate will appear in the Client Certificate drop-down list. Select the required certificate from the drop-down list.
 - **View Certificate** - click to view the selected Client Certificate.

TIP: If you get certification validation errors, ensure that the local computer trusts the certificate authority of the certificate that the Elasticsearch server is presenting.

- **Enable Amazon Web Service (AWS) authentication** - select this option to use the Amazon Web Service (AWS) version of the Elasticsearch service.
 - **Access Key** - enter the AWS Access Key. This is the username for the AWS IAM user that has access to the Elasticsearch services inside AWS.
 - **Secret Key** - enter the AWS Secret Key. This is the password for the Access Key user.
 - **AWS Region** - enter the AWS Region name.

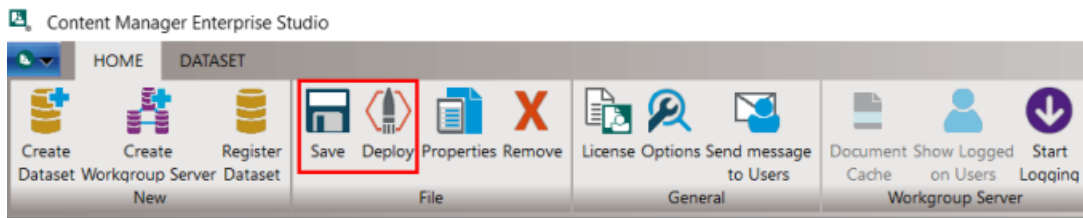


If you've setup your environment correctly the **Test** button should return a **Succeeded** response.

Click **OK** to close the **Content index properties** dialog.

If you encounter the following prompt: **The index 'CM_45' already exists on the Elasticsearch server. Do you want to delete it and create and new one?** Select **No**.

Save and then **Deploy** your new settings in Enterprise Studio.



Confirm your Elasticsearch content index is operational

Performing a simple document content search from a Content Manager client is the best way to confirm your content index is operational.

In this case, we selected a document with unique content so only one result was returned. The use of unique content confirms the searching mechanism can interrogate the entire index and return the correct result.

Search for		Search by	Matching criteria
Records	Document Content	Polyfluoroalkyl	
Record Type	Record Number	Title	Date Created
Document	G17/19	An Overview of Perfluoroalkyl and Polyfluoroalkyl Substances and Interim ...	9/10/2017 at 9:15 AM

Troubleshooting common issues

Logging

When it comes to troubleshooting general content indexing issues there are three key logs you should be looking at:

Log file name	Default Log file location	Useful for...
Elasticsearch cluster log	ES_HOME\logs\[cluster_name].log	Identifying issues to do with the health of the Elasticsearch environment. Located on each Elasticsearch server in your cluster.
Workgroup server document content indexing event log	<Drive>\Micro Focus Content Manager\ServerLocalData\TRIM\Log\ DCI_<DB ID>_<YYYY>_<MM>_<DD>.log	Whenever an electronic record is created, a new version is committed or an electronic document is destroyed, deleted or removed, an event is created to indicate that the document content index should be updated. This log file may help you narrow down where a particular problem is occurring when compared to other logs listed in this table.
Client Reindex log	<Drive>\Users\<username>\AppData\Local\Micro Focus\Content Manager\<DB ID>\Log\ElasticReindex_<YYYY>_<MM>_<DD>.log	Identifying problematic documents or timeout issues when reindexing from the Content Manager client.

Diagnosing “operation timed out” errors

The most common issue you’ll probably come across when indexing content is **operation timed out** errors in the client **Reindex** log.

This error doesn’t always mean there’s an issue. Sometimes, it can just mean the client has waited for a response from the Elasticsearch server beyond the **Request Timeout** value you’ve set in Enterprise Studio. Often, you’ll find Elasticsearch has continued to index the content after the client has reported a timeout event. You can check if the content has been indexed by doing a search based on content. If it doesn’t return a result you should then attempt to run a re-index on only missing items.

If you continue to observe timeout errors and content isn't being indexed you can then proceed to examining the logs identified in the previous section. Key areas you should be looking at in this scenario are:

- **Enterprise Studio – Transaction Limit**

Content Manager sends documents to Elasticsearch using a bulk-index API, this value sets the maximum size of the memory buffer before we send the bulk import request to Elasticsearch. When performing re-indexing, using a larger value will speed up indexing. However, you don't want it to be too large as it can result in Elasticsearch taking longer to return a result and/or the JVM running out of heap memory if resources are limited. We recommend you keep the transaction limit set to 5-15 MB (the default transaction limit value is 9 MB).

- **Enterprise Studio – Request Timeout**

The request timeout value sets the time Content Manager will wait for Elasticsearch to respond every time a HTTP request is sent. Every environment will perform differently depending on the resources it has available. If there's adequate resources simply extending the timeout value may resolve many of the timeout errors observed (the default timeout value is 60 seconds).

- **Elasticsearch – available heap and available nodes**

Timeout errors may also indicate there's isn't enough memory allocated to the JVM, your servers aren't large enough to handle the load, or you don't have enough nodes in your cluster. If the cluster log indicates this you should revisit the memory configuration and sizing recommendations outlined in [Memory Management and configuring the Java Virtual Machine \(JVM\)](#).

Proxy servers and firewalls

Content Manager Clients and Workgroup Servers need direct access to your Elasticsearch servers. Make sure your firewall has the appropriate ports open (default port range is 9300-9400) and your proxy server is configured to not cache traffic bound for Elasticsearch hosts.

Useful Information

Kibana

The Elastic team recommend that as your Elasticsearch environment grows to include more nodes and clusters you should consider using Kibana to manage it. Kibana not only allows you to manage your clusters more efficiently it also gives you the ability to analyze and make sense of your data. Visit elastic.co to get more information on Kibana.

cURL

The Elastic team recommend cURL, which is an open source command line utility that can be used to directly query and manage Elasticsearch via its API. If you don't decide to use Kibana you can use cURL commands to perform a large range of tasks e.g. delete an index, get node and cluster health data. The Elasticsearch documentation contains many examples of cURL scripts.

cURL can be downloaded, free of charge, from the official cURL site.

Elasticsearch Reference

The Elasticsearch Reference is a detailed knowledge base of all aspects of Elasticsearch. You should consult this when considering the use of Elasticsearch in your organization.

You can find the Elasticsearch Reference in the Docs section on elastic.co

Hardware requirements

The Content Manager Support team regularly gets asked for assistance to size a customer's environment for content indexing. The Elastic team do a good job of explaining how to approach this problem in the **Hardware** section of **Elasticsearch: The Definitive Guide** on elastic.co

Microsoft also present an excellent overview of Elasticsearch architecture and how to scale it on Azure which can be applied to any environment. You can find this information in the **Elasticsearch on Azure** section of the Azure online documentation.