

# Netwrix Data Classification Installation and Configuration Guide

Version: 5.5.2  
3/10/2020



# Table of Contents

1. Supported Data Sources .....	5
2. Deployment Planning .....	6
2.1. NDC Server .....	6
2.2. Data Storages and Sizing .....	6
2.2.0.1. NDC SQL database .....	7
2.2.0.2. NDC Index .....	7
2.2.1. Scalability and Performance .....	7
2.2.2. Recommendations on SQL Database Maintenance .....	8
2.3. Scalability and Performance .....	9
2.3.1. Example: Mid-Size Data Environment .....	10
2.3.1.1. Configuration .....	10
2.3.1.2. Data Set .....	12
2.3.1.3. Data Processing .....	12
2.3.2. Example: Large-Size Environment .....	12
2.3.2.1. Configuration .....	13
2.3.2.2. Data Set .....	14
2.3.2.3. Data Processing .....	15
3. Requirements to Install Netwrix Data Classification .....	16
3.1. Hardware Requirements .....	17
3.1.1. Netwrix Data Classification Server .....	17
3.1.2. SQL Server .....	17
3.1.3. Network Access .....	18
3.1.4. Configuring NDC Servers Cluster and Load Balancing with DQS Mode .....	18
3.1.4.1. Applying DQS Mode .....	19
3.2. Software Requirements .....	22
3.3. Accounts and Required Permissions .....	24
4. Configure NDC Database .....	26

5. Install Netwrix Data Classification .....	27
6. Upgrade to the Latest Version .....	29
6.1. Take Preparatory Steps .....	29
6.2. Considerations and Limitations .....	29
7. Configuring NDC Servers Cluster and Load Balancing with DQS Mode .....	31
7.1. Applying DQS Mode .....	31
8. Configure IT Infrastructure .....	34
8.1. Configure Microsoft Exchange for Crawling and Classification .....	35
8.2. Configure NFS File Share for Crawling .....	38
8.3. Configure G Suite and Google Drive for Crawling .....	38
9. Initial Product Configuration .....	43
9.1. Select Processing Mode .....	43
9.1.1. No Index .....	44
9.1.2. Keyword .....	44
9.1.3. Compound Term .....	44
9.2. Processing Settings .....	44
9.3. Add Taxonomy .....	46
9.4. Review Your Configuration .....	46

## Legal Notice

The information in this publication is furnished for information use only, and does not constitute a commitment from Netwrix Corporation of any features or functions, as this publication may describe features or functionality not applicable to the product release or version you are using. Netwrix makes no representations or warranties about the Software beyond what is provided in the License Agreement. Netwrix Corporation assumes no responsibility or liability for the accuracy of the information presented, which is subject to change without notice. If you believe there is an error in this publication, please report it to us in writing.

Netwrix is a registered trademark of Netwrix Corporation. The Netwrix logo and all other Netwrix product or service names and slogans are registered trademarks or trademarks of Netwrix Corporation. Microsoft, Active Directory, Exchange, Exchange Online, Office 365, SharePoint, SQL Server, Windows, and Windows Server are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries. All other trademarks and registered trademarks are property of their respective owners.

## Disclaimers

This document may contain information regarding the use and installation of non-Netwrix products. Please note that this information is provided as a courtesy to assist you. While Netwrix tries to ensure that this information accurately reflects the information provided by the supplier, please refer to the materials provided with any non-Netwrix product and contact the supplier for confirmation. Netwrix Corporation assumes no responsibility or liability for incorrect or incomplete information provided about non-Netwrix products.

© 2019 Netwrix Corporation.

All rights reserved.

# 1. Supported Data Sources

The table below lists systems that can be crawled with Netwrix Data Classification:

Data Source	Supported Versions
File System	<ul style="list-style-type: none"><li>• CIFS/SMB (Preferred)</li><li>• NFS</li></ul>
SharePoint, SharePoint Online, OneDrive for Business	<ul style="list-style-type: none"><li>• 2010 and above</li></ul>
Database	<ul style="list-style-type: none"><li>• Microsoft SQL Server 2008 and above</li><li>• Oracle 10g and above</li></ul>
Exchange	<ul style="list-style-type: none"><li>• Exchange Server 2010 and above</li><li>• Exchange Online</li></ul>
Google Drive	<ul style="list-style-type: none"><li>• N/A</li></ul>
Outlook Mail Archive	<ul style="list-style-type: none"><li>• Outlook 2010 and above</li></ul>

## 2. Deployment Planning

This section provides recommendations and considerations for Netwrix Data Classification deployment planning. Review these recommendations and choose the most suitable deployment scenario and possible options depending on the IT infrastructure and data sources you are going to process.

In this section:

- [NDC Server](#)
- [Data Storages and Sizing](#)
- [Scalability and Performance](#)

### 2.1. NDC Server

**Netwrix Data Classification Server** can be deployed on a physical server or on a virtual machine in the virtualized environment on VMware or Microsoft Hyper-V platform.

When planning for NDC Server, consider a significant CPU load during data processing. Thus, installing NDC Server on a highly-loaded production machine is not recommended. For more information, refer to [Hardware Requirements](#).

**Web-based client** (management console) is always installed together with the NDC Server, so the IIS server role must be enabled on the target machine. For more information, refer to [Software Requirements](#).

**NOTE:** For evaluation and PoC purposes, Netwrix provides a virtual appliance — a virtual machine image with pre-installed Netwrix Data Classification on Generalized Windows Server 2016 (180-day evaluation version) and Microsoft SQL Server 2017 Express. For details, see [Requirements to Deploy Virtual Appliance](#).

Remember that for production environments, your NDC Server and database server must meet the [Requirements to Install Netwrix Data Classification](#). Virtual appliance configuration is insufficient for production and is not recommended for that purpose.

To balance the load while indexing and classifying data in the large-size and extra-large environments (i.e. with over 8-10 mln objects to process), it is strongly recommended to deploy several NDC Servers and configure **Distributed Query Server** mode for them. See [Configuring NDC Servers Cluster and Load Balancing with DQS Mode](#).

### 2.2. Data Storages and Sizing

Netwrix Data Classification utilizes two data storages:

- NDC SQL database — SQL Server database that stores product configuration and metadata for the data sources.

- NDC Index — a full-text search index that comprises a set of files in the proprietary format (.CSE).

### 2.2.0.1. NDC SQL database

Make sure you have NDC Server and **NDC SQL database** deployed on different machines.

It is recommended to create the **NDC SQL database** on a dedicated SQL Server instance.

- Minimal requirement is SQL Server 2008 R2 Standard Edition.
- Estimate required disk space assuming *10 - 12 KB* per indexed object. For example, for *5,000,000* objects, the database size will be approximately *50 GB*.  
Therefore, SQL Server Express edition will be only suitable for evaluation and PoC environments (up to 1,000,000 documents to process).
- If configuring database settings via SQL Server Management Studio, you will need to set **Autogrowth** / **Maxsize** values for the PRIMARY database files as follows:
  - **File growth:** *128 MB* - recommended value for small to medium environment, *512 MB* - for large environment, i.e. if planning to index data sources containing 1,000,000+ objects.
  - **Maximum File Size** - select *Unlimited*.
- Make sure that the **Recovery model** for this database is set to *Simple*. Do not change the recovery model — to avoid log files growth.

See also [Recommendations on SQL Database Maintenance](#).

### 2.2.0.2. NDC Index

Required disk space for the **NDC Index** file storage will depend, in particular, on the data processing mode you plan to use (*No Index*, *Keyword* or *Compound Term*).

As a rule of thumb, required space can be calculated as 35% of data you plan to be indexed. For example, if you have 45 GB of files, they will require up to 15 GB for the **NDC Index** files.

## 2.2.1. Scalability and Performance

Scalability and performance testing revealed that based on the number of objects to classify, the environments can be ranged as follows:

Number of objects to classify	Environment	Comment
Up to 500,000	Proof-of-concept and small-size environment	

Number of objects to classify	Environment	Comment
Up to 8, 000, 000	Mid-size environment	
Up to 32, 000, 000	Large-size environment	
More than 32, 000, 000	Extra-large environment	System architect's assistance is required for deployment planning requires

The following sections describe related deployment scenarios and provide examples for resource planning:

[Example: Mid-Size Data Environment](#)

[Example: Large-Size Environment](#)

You can use these examples to estimate hardware requirements and plan for scalability of your Data Classification deployment.

Again, consider that for the large-size and extra-large environments, it is strongly recommended to configure a cluster of several NDC Servers and apply DQS mode to these clustered servers. See [Configuring NDC Servers Cluster and Load Balancing with DQS Mode](#) for details.

## 2.2.2. Recommendations on SQL Database Maintenance

Netwrix Data Classification uses SQL Server database as a storage for file metadata prepopulated by **NDC Collector** service. If you are going to crawl more than 1M of objects, you need to pay attention to SQL Server database maintenance procedures, especially during initial collection period. You or your database administrator can perform these tasks according to your company's internal policies, if any, or follow the recommendations below.

To ensure data integrity and performance, maintenance operations recommended by Microsoft should be performed for your NDC SQL database once a day, putting more focus on the **Pages** table.

### *To maintain your SQL database*

**IMPORTANT!** Stop all Netwrix Data Classification services before you start the maintenance procedures. If you are using DQS (Distributed Query Server) mode, you need to stop all services on all instances of Netwrix Data Classification. You can stop and start the services, using the **Netwrix Data Classification: Service Viewer** tool.

Do the following:

1. On the computer where **Netwrix Data Classification** is installed, start the **Netwrix Data Classification Service Viewer** tool. Select **Stop** next to each service.
2. Start **Microsoft SQL Management Studio** and connect to the SQL Server instance hosting NDC SQL database.



3. Right-click the NDC SQL database and select **Reports** → **Standard Reports** → **Index Physical Statistics** report.
4. Based on the report data, perform the recommended operations (*Rebuild* or *Reorganize*) for the indexes of the certain tables. For details, see this Microsoft article: [Reorganize and rebuild indexes](#)

**NOTE:** The following indexes do not influence database performance during the initial data crawling, so you can skip them when performing the initial maintenance procedure:

- **Checksum**
- **IdxPagesFileChecksum**
- **IdxPagesTextChecksum**
- **DocumentChecksum**

After the initial crawling is completed, you can include these indexes in the standard daily database maintenance procedure.

## 2.3. Scalability and Performance

Scalability and performance testing revealed that based on the number of objects to classify, the environments can be ranged as follows:

Number of objects to classify	Environment	Comment
Up to 500, 000	Proof-of-concept and small-size environment	
Up to 8, 000, 000	Mid-size environment	
Up to 32, 000, 000	Large-size environment	
More than 32, 000, 000	Extra-large environment	System architect's assistance is required for deployment planning in such environments.

The following sections describe related deployment scenarios and provide examples for resource planning:

- [Example: Mid-Size Data Environment](#)
- [Example: Large-Size Environment](#)

You can use these examples to estimate hardware requirements and plan for scalability of your Data Classification deployment.

**IMPORTANT!** For the large-size and extra-large environments, it is strongly recommended to configure a cluster of several NDC Servers and apply DQS mode to these clustered servers. See [Configuring NDC Servers Cluster and Load Balancing with DQS Mode](#) for details.

## 2.3.1. Example: Mid-Size Data Environment

This example provides the results of different data processing modes testing for a mid-size environment. The following infrastructure components were deployed as VMware VMs: Netwrix Data Classification server, database server, data source (file server).

### 2.3.1.1. Configuration

Netwrix Data Classification Server

Specification	Settings	Comments
Platform	VMware ESXi 6.0	
Hardware:	<p><i>CPU:</i> Intel Xeon E5-2683 v4 , 2.10 GHz</p> <p><i>Logical Processors:</i> 32 (2 sockets, 16 cores per socket)</p> <p><i>Memory:</i> 256 GB</p>	Using faster processor increases data processing performance.
Virtual machine configuration	<p><i>CPU:</i> 8 vCPU</p> <p><i>Memory:</i> 32 Gb</p> <p><i>Hard disk:</i> SSD storage; thin provisioning enabled</p>	
Guest OS	Windows Server 2012 R2 (64-bit)	
Software	Netwrix Data Classification server with <i>Distributed Query Server Mode</i> configuration.	Used 4 server instances with <i>Distributed Query Server Mode</i> configuration. See "Distributed Query Server Mode" for details.

Database Server

Specification	Settings	Comment
Platform	VMware ESXi 6.0	

Specification	Settings	Comment
<b>Hardware:</b>	<p><i>CPU:</i> Intel Xeon E5-2660 v4 , 2.00 GHz</p> <p><i>Logical Processors:</i> 56 (2 sockets, 14 cores per socket)</p> <p><i>Memory:</i> 512 GB</p>	Using faster processor increases data processing performance.
<b>Virtual machine configuration</b>	<p><i>CPU:</i> 8 vCPU</p> <p><i>Memory:</i> 128 Gb</p> <p><i>Hard disk:</i> SSD storage; thin provisioning enabled</p>	
<b>Guest OS</b>	Windows Server 2012 R2 (64-bit)	
<b>Software</b>	Microsoft SQL Server 2016 SP2 Enterprise Edition	

#### Data Source (File Server)

Specification	Settings	Comments
<b>Platform</b>	VMware ESXi 6.7	
<b>Hardware:</b>	<p><i>CPU:</i> Intel Xeon E5-2620 v4 , 2.10 GHz</p> <p><i>Logical Processors:</i> 32 (2 sockets, 8 cores per socket)</p> <p><i>Memory:</i> 128 GB</p>	Using faster processor increases data processing performance.
<b>Virtual machine configuration</b>	<p><i>CPU:</i> 8 vCPU</p> <p><i>Memory:</i> 32 Gb</p> <p><i>Hard disk:</i> SSD storage; thin provisioning enabled</p>	
<b>Guest OS</b>	Windows Server 2019 Standard (64-bit)	

### 2.3.1.2. Data Set

The file server with the following data set was used as a content source:

- Number of files: 1, 000, 000+
- Number of folders: 65, 000
- File types: PDF, DOCX, HTML, RTF, TXT
- Average file size: 500 K - 1 MB
- Total data set size: 1.8 TB

### 2.3.1.3. Data Processing

Data processing was launched for the file server with **1, 000, 000+** objects (files and folders) in each mode: **No Index, Keyword, Compound Term**. It was set up to use all predefined taxonomies, no OCR.

Automated workflow was configured as follows:

- Workflow condition: a file gets classified with any taxonomy (in addition to the Size, Type and Language standard taxonomies).
- Workflow rule: such file is migrated to the dedicated location.

Data processing results for all modes are listed below.

	No Index	Keyword	Compound Term
Processing time	1 day 07 h 22 m	1 day 10 h 03 m	1 day 12 h 35 m
Files processed per minute (average)	558	514	479
Files with workflow condition triggered (i.e. at least 1 taxonomy applied)	135584	154888	152524
NDC SQL database size (MDF)*	9.5 GB	9.8 GB	7.2 GB
Index size	55 GB	176 GB	321 GB

\* — For overall space estimations, secondary files should also be considered, so please contact your database administrator.

### 2.3.2. Example: Large-Size Environment

This example provides the results of different data processing modes testing for a larger data environment. The following infrastructure components were deployed as VMware VMs: Netwrix Data Classification

server, database server, data source (file server).

### 2.3.2.1. Configuration

#### Netwrix Data Classification Server

Specification	Settings	Comments
Platform	VMware ESXi 6.7	Using faster processor increases data processing performance.
Hardware:	<p><i>CPU:</i> Intel Xeon E5-2660 v4 , 2.00 GHz</p> <p><i>Logical Processors:</i> 56 (2 sockets, 14 cores per socket)</p> <p><i>Memory:</i> 256 GB</p>	
Virtual machine hardware	<p><i>CPU:</i> 8 vCPU</p> <p><i>Memory:</i> 32 Gb</p> <p><i>Hard disk:</i> SSD storage; thin provisioning enabled</p>	
Guest OS	Windows Server 2012 R2 (64-bit)	
Software	Netwrix Data Classification server with <i>Distributed Query Server Mode</i> configuration.	Used 4 server instances with <i>Distributed Query Server Mode</i> configuration. See "Distributed Query Server Mode" for details.

#### Database Server

Specification	Settings	Comment
Platform	VMware ESXi 6.7	Using faster processor increases data processing performance.
	<p><b>Hardware:</b></p> <p><i>CPU:</i> Intel Xeon E5-2620 v4 , 2.10 GHz</p> <p><i>Logical Processors:</i> 32 (2 sockets, 8 cores per socket)</p> <p><i>Memory:</i> 128 GB</p>	
Virtual machine	<i>CPU:</i> 8 vCPU	

Specification	Settings	Comment
hardware	<i>Memory:</i> 128 Gb <i>Hard disk:</i> SSD storage; thin provisioning enabled	
Guest OS	Windows Server 2012 R2 (64-bit)	
Software	Microsoft SQL Server 2016 SP2 Enterprise Edition	

#### Data Source (File Server)

Specification	Settings	Comments
Platform	VMware ESXi 6.7 <b>Hardware:</b> <i>CPU:</i> Intel Xeon E5-2620 v4 , 2.10 GHz <i>Logical Processors:</i> 32 (2 sockets, 8 cores per socket) <i>Memory:</i> 128 GB	Using faster processor increases data processing performance.
Virtual machine hardware	<i>CPU:</i> 8 vCPU <i>Memory:</i> 32 Gb <i>Hard disk:</i> SSD storage; thin provisioning enabled	
Guest OS	Windows Server 2019 Standard (64-bit)	

### 2.3.2.2. Data Set

The file server with the following data set was used as a content source:

- Number of files: 32, 000, 000+
- Number of folder: 2, 000, 000+
- File types: PDF, DOCX, HTML, RTF, TXT
- Average file size: 500K - 1MB
- Total data set size: 57 TB

### 2.3.2.3. Data Processing

Data processing was launched for the file server with **34, 000, 000+** objects (files and folders) in **Keyword** mode. It was set up to use all predefined taxonomies, no OCR.

Automated workflow was configured as follows:

- Workflow condition: a file gets classified with any taxonomy (in addition to the Size, Type and Language standard taxonomies).
- Workflow rule: such file is migrated to the dedicated location.

Data processing results are listed below.

	Keyword
Processing time	62 days 19 hrs 47 min
Files processed per minute (average)	365
Files with workflow condition triggered (i.e. at least 1 taxonomy applied)	4752724
NDC SQL database size (MDF)*	190GB
Index size	4 TB

\* — For overall space estimations, secondary files should also be considered, so please contact your database administrator.

# 3. Requirements to Install Netwrix Data Classification

This section contains the hardware and software requirements and other prerequisites needed to deploy Netwrix Data Classification.

- [Hardware Requirements](#)
- [Software Requirements](#)
- [Accounts and Required Permissions](#)



## 3.1. Hardware Requirements

Review the hardware requirements for the computer where Netwrix Data Classification will be installed.

You can deploy Netwrix Data Classification on a virtual machine running Microsoft Windows guest OS on the corresponding virtualization platform, in particular:

- VMware vSphere
- Microsoft Hyper-V
- Nutanix AHV

Note that Netwrix Data Classification supports only Windows OS versions listed in the [Software Requirements](#) section.

### 3.1.1. Netwrix Data Classification Server

The requirements in this section apply to a single Netwrix Data Classification server.

To deploy a server cluster, make sure all planned cluster nodes meet the requirements listed below. Consider deploying 1 Netwrix Data Classification Server per approx. 1, 000, 000 objects to process.

See [Deployment Planning](#) and [Configuring NDC Servers Cluster and Load Balancing with DQS Mode](#) for more information.

Hardware Component	Minimum Requirements	Recommended
Processor	Any modern.  Consider that greater CPU frequency and number of cores improve overall performance of Netwrix Data Classification Server.	Any multi-core
RAM	8 GB	16 GB

### 3.1.2. SQL Server

Review the hardware requirements for the computer where Netwrix Data Classification SQL Database will be deployed.

Hardware Component	Minimum requirements	Large environment (up to 8 m objects for File Servers and up to 2 m objects for SharePoint)	XLarge environment (up to 32 m objects and up to 8 m objects for SharePoint)
Processor	Any multi-core	8 cores	8 cores
RAM	16 GB	64 GB	128 GB
Disk space	Estimate required disk space assuming 10 - 12 KB per indexed object. For example, for 5,000,000 objects, the database size will be approximately 50 GB. See also <a href="#">Deployment Planning</a> .		

**NOTE:** If you are going to process more than 10,000,000 objects per day, remember to perform database maintenance procedures. See [Recommendations on SQL Database Maintenance](#).

### 3.1.3. Network Access

Specification	Requirement
Network access	Ensure that your Netwrix Data Classification servers are available over the network on a HTTP compliant port from all machines where the client interface (management console) will run.

### 3.1.4. Configuring NDC Servers Cluster and Load Balancing with DQS Mode

The **Distributed Query Server (DQS)** mode allows you to balance the load between multiple Netwrix Data Classification Servers (NDC Servers) while data collection, indexing and classification. This approach is strongly recommended if you need to process large data volumes, for example:

- **File Servers**—Up to 32 m objects per cluster of 4 servers.
- **SharePoint**—Up to 8 m objects per cluster of 4 servers.

To apply **Distributed Query Server** mode, you need to arrange your NDC Servers in a 'cluster' for load distribution, as described below. Each clustered NDC Server will store its own set of .CSE files — that is, **NDC Index** will be a distributed index. To assemble and combine data required for the search results, each NDC Server will automatically communicate with the other clustered servers.

**NOTE:** All NDC Servers in the cluster will share a single NDC SQL database.

This functionality is implemented through the *QueryServer* application installed together with NDC Server.

### 3.1.4.1. Applying DQS Mode

DQS mode can be configured via the administrative web console.

If you want to implement DQS configuration for the your NDC deployment, consider the following:

- This action cannot easily be undone, so before applying the DQS mode, take a full backup of your NDC deployment. Also, read the related documentation sections thoroughly before you start.
- Make sure all servers you plan to add to the DQS cluster have proper network connection and are visible to each other across the network. Adjust firewall settings if necessary.
- Initially, all existing documents will be 'allocated' to the first server in the 'cluster' and then re-distributed across all configured servers.

To be able to configure the DQS mode, current account requires a **Superuser** role.

#### *To arrange NDC Servers cluster and apply DQS mode*

1. Install and configure the first Netwrix Data Classification Server as described in the [Install Netwrix Data Classification](#) section.
2. Open administrative web console.
3. Navigate to **Config** → **Utilities** → **DQS**.
4. Select **Enable DQS**.

**NOTE:** Once the DQS mode is enabled, you cannot roll back your configuration. Netwrix strongly recommends to ensure that you have taken a full backup of your environment. If ready, confirm the DOS enablement operation when prompted.

5. On the **DQS** tab, click **Add** to add servers you prepared, one by one.

ID	Status	Load	Server	QS Path	Alternate Server	Alternate QS Path
1	Active	0%	FM-DDC-2	http://FM-DDC-2/conceptQS/conceptQS.aspx		

Complete the following fields:

The screenshot shows the 'Details' dialog box for configuring a DQS (Data Quality Services) instance. The dialog has the following fields:

- Server:** A text input field.
- QS Path:** A text input field containing the default value: `http://servername/conceptQS/conceptQS.aspx`.
- Active:** A checkbox that is checked.
- Alternate Server:** A text input field.
- Alternate QS Path:** A text input field containing the default value: `http://servername/conceptQS/conceptQS.aspx`.

At the bottom of the dialog are 'Save' and 'Cancel' buttons. The background interface shows a table with columns 'ID', 'Status', and a list of records. There is also a 'Delete' button and an 'Add' button.

Setting	Value
Server	Provide the NDC Server name or IP address (name format is case-insensitive).
QS Path	Path to the solution component responsible for DQS mode, residing on the server being added. Filled in automatically; leave the default value.
Active	Select to enable clustering for the instance being added.
Alternate Server	Netwrix recommends using default values.
Alternate QS Path	Netwrix recommends using default values.

- Click **Save** to close the dialog.
- Prepare to install other Netwrix Data Classification Server instances, assuming each server requires a dedicated machine. Make sure they meet the [Hardware Requirements](#) and general [Software Requirements](#).
- On each server, follow the installation steps as described in the [Install Netwrix Data Classification](#) section until **SQL Database** configuration.
- On the **SQL Database** step, provide the name of the SQL Server instance that hosts **NDC SQL database** you configured for the first NDC Server.

**NOTE:** Ignore the confirmation dialog on the existing schema in the selected SQL database.

- Complete the installation.
- Repeat steps 2 - 6 for every NDC Server, then review the list of servers to make sure the new server was included.

**DQS**

**Information**

The Distributed Query Server (DQS) is a component of conceptClassifier that allows an index to be distributed across multiple servers.

A distributed index means that there are multiple servers running the Collector, Indexer, Classifier, and QueryServer applications - each with its own set of ".cse" files. All servers share a single SQL database.

Please note:

- Each server can only run one set of Windows Services
- The server name should be specified in the NETBIOS format (case insensitive)
- The QS Path specified should be a direct connection to the server in question (I.E not a load balanced address for the cluster)
- The server should be set to "Active" to be considered part of the cluster
- All servers that you wish to run the Windows Services on should be specified within the DQS list

ID	Status	Load	Server	QS Path	Alternate Server	Alternate QS Path
1	Active	0%	FM-DDC-2	http://FM-DDC-2/conceptQS/conceptQS.aspx		
2	Active	0%	fm-ddc-4	http://fm-ddc-4/conceptQS/conceptQS.aspx		
3	Active	0%	fm-ddc-5	http://fm-ddc-5/conceptQS/conceptQS.aspx		
4	Active	0%	fm-ddc-3	http://fm-ddc-3/conceptQS/conceptQS.aspx		

Showing 4 record(s) | Page Size: 10 | 25 | 50 | 100 | 200

12. If you were configuring the DQS mode for the existing NDC deployment, you will be prompted to re-collect data from the data sources—in order to re-distribute the content index across all NDC Servers in the cluster.

**NOTE:** To force re-distribution when necessary, you can use the **Re-Collect** command available after clicking **Run Cleaner** button on the **Config > Settings > Collector** tab.

To review system health and check your configuration, use the product dashboards. See [Dashboards](#) for more information.

## 3.2. Software Requirements

The table below lists the software requirements for the Netwrix Data Classification installation:

Component	Requirements
Operating system	Windows 2012 R2 and above Server Operating System Software.
Windows Features	<div>Web Server Role (IIS)</div> <hr/> <div>Common HTTP Features</div> <ul style="list-style-type: none"> <li>• Default Document</li> <li>• HTTP Errors</li> <li>• Static Content</li> <li>• HTTP Redirection</li> </ul> <hr/> <div>Security</div> <ul style="list-style-type: none"> <li>• Windows Authentication</li> <li>• Anonymous Authentication</li> </ul> <p><b>NOTE:</b> The <b>Anonymous Authentication</b> element is included in the default installation of IIS 7. Make sure you use IIS 7 and above.</p> <hr/> <div>Application</div> <ul style="list-style-type: none"> <li>• ISAPI Extensions</li> </ul> <div>Development</div> <ul style="list-style-type: none"> <li>• ISAPI Filters</li> </ul> <hr/> <div>Other features</div> <hr/> <div>.NET Framework</div> <ul style="list-style-type: none"> <li>• .NET Framework 4.7.2</li> </ul> <div>Features</div> <ul style="list-style-type: none"> <li>• ASP.NET</li> </ul> <hr/> <div>WCF Services</div> <ul style="list-style-type: none"> <li>• HTTP Activation</li> </ul>
SQL Server	<ul style="list-style-type: none"> <li>• <a href="#">SQL Server 2008 R2 Standard Edition</a> (or later).</li> <li>• SQL Server 2016 SP2 recommended (for better performance).</li> </ul> <p><b>NOTE:</b> For large environments, SQL Server Enterprise edition may be needed; see needed. See <a href="#">Deployment Planning</a>.</p>

Component	Requirements
Microsoft IFilters	<ul style="list-style-type: none"><li>• <a href="#">Microsoft Office 2010 Filter Packs</a> and above, 64-x edition.</li></ul>
Visual Studio	<ul style="list-style-type: none"><li>• <a href="#">Visual C++ Redistributable Packages for Visual Studio 2015</a> and above.</li></ul>

Other software	
Antivirus	Netwrix recommends adding Netwrix Data Classification Index files to the list of exclusions (white list) of any installed antivirus. These files have <i>.CSE</i> extension.

## 3.3. Accounts and Required Permissions

Netwrix Data Classification uses the following accounts:

Account	Description
Service Account	<p>This account is specified during the product setup.</p> <p>Windows domain account that you plan to use as a service account will need the following:</p> <ul style="list-style-type: none"> <li>• <b>Local Administrator</b> rights on the server where Netwrix Data Classification will be installed.</li> <li>• Permissions to run the <b>Windows Services</b> and <b>IIS Application pool</b>.</li> </ul> <p>After installation, this account will be automatically granted the <b>Logon as a service privilege</b> on the Netwrix Data Classification server.</p> <ul style="list-style-type: none"> <li>• SQL Server <b>DBO</b> permissions to the NDC SQL database (if using Windows Authentication to access SQL Server).</li> </ul> <p><b>NOTE:</b> Optionally, you can use local account instead of domain account.</p>
Crawl content	<p>Ensure the availability of accounts with sufficient permissions to access your content sources:</p> <ul style="list-style-type: none"> <li>• SharePoint, SharePoint Online site collection— <b>Site Collection Administrator</b> role.</li> <li>• Exchange mailboxes:             <ol style="list-style-type: none"> <li>1. <b>ApplicationImpersonation</b> —allows the crawling account to impersonate each of the mailboxes / users configured for collection.</li> <li>2. <b>Mailbox Search</b> —allows the crawling account to enumerate mailboxes, i.e. automatic discovery of mailboxes.</li> </ol> </li> </ul> <p>See <a href="#">Configure Microsoft Exchange for Crawling and Classification</a> for detailed information on configuring these permissions.</p> <ul style="list-style-type: none"> <li>• Outlook Mail Archive (PST file)— <b>Read</b> permission.</li> <li>• File System (SMB, NFS) — <b>Read</b> permission for the folders and files you need to crawl.</li> </ul>



Account	Description
	<ul style="list-style-type: none"><li>G Suite and Google Drive —service account needs permissions to read data in the individual and shared Drives on behalf of users using the Google Drive API.</li></ul> <p>See <a href="#">Configure G Suite and Google Drive for Crawling</a> for detailed information.</p> <ul style="list-style-type: none"><li>Database— <b>Read</b> permission for the database schema and data.</li></ul>
Apply tagging	To use tagging, i.e. to write classification attributes back to the content file, service account will need the appropriate <b>Modify</b> permissions on the content source.

## 4. Configure NDC Database

Netwrix Data Classification uses Microsoft SQL Server database as data storage. During installation, you have been prompted to create a dedicated **NDC SQL database** on your SQL Server instance. Upon installation completion, you need to configure it as shown below for the product to function properly. You can create the database manually prior to the product installation—Using **SQL Server Management Studio** or **Transact-SQL**. Refer to the following Microsoft article for detailed instructions on how to create a new database: [Create a Database](#).

**NOTE:** For performance purposes, Netwrix strongly recommends to separate NDC and SQL Server machine.

For certain product features, SQL Server Standard or Enterprise edition is required.

### *To configure NDC database*

**NOTE:** The account used to create the NDC SQL database must be granted the **dbcreator** server-level role.

1. On the computer where SQL Server instance with the **NDC SQL database** resides, navigate to **Start → All Programs → Microsoft SQL Server → SQL Server Management Studio**.
2. Connect to the server.
3. Locate the **NDC\_Database**, right-click it and select **Properties**.
4. Select the **Files** page and set the **Initial Size (MB)** parameter for PRIMARY file group to **512 MB**.
5. Click **Expand** next to **PRIMARY** file group and set **Autogrowth / Maxsize** as follows:

Option	Description
File Growth	<ul style="list-style-type: none"><li>• Recommended—<b>128 MB</b>.</li><li>• Large environment— <b>512 MB</b>.</li></ul>
Maximum File Size	Select <b>Unlimited</b> .

6. Go to **Options** page and make sure that the **Recovery model** parameter is set to *"Simple"*.

**NOTE:** Netwrix recommends that you do not change the recovery model to avoid log files growth.

## 5. Install Netwrix Data Classification

1. Run **Netwrix\_Data\_Classification.exe**.
2. Review minimum system requirements and then read the License Agreement. Click **Next**.
3. Follow the instructions of the setup wizard. When prompted, accept the license agreement.
4. On the **Product Settings** step, specify path to install Netwrix Data Classification. For example, *C:\Program Files\NDC\*.
5. On the **Configuration** step, specify the directory where **Index files** reside. For example, *C:\Program Files\NDC\Index*.
6. On the **SQL Database** step, provide SQL Server database connection details.

Complete the following fields:

Option	Description
Server Name	Provide the name of the SQL Server instance that hosts your NDC SQL database. For example, "WORKSTATIONS\SQLSERVER".
Authentication Method	Select Windows or SQL Server authentication method.
Username	Specify the account name.
Password	Provide your password.
Database Name	Enter the name of the SQL Server database. Netwrix recommends using <b>NDC_database</b> name.

7. On the **Licensing** step, add license. You can add license as follows:
  - Click the **Import** button and browse for your license file  
*OR*
  - Open your license file with any text editor, e.g., **Notepad** and paste the license text to the **License** field.
8. On the **Administration Web Application** step, review default IIS configuration.
9. On the **Services** step, configure Netwrix Data Classification services:

- Select all services to be installed.
- **File System Path**—Use default path or provide a custom one to store Netwrix Data Classification's Services files. For example, *C:\Program Files\NDC Services*.
- Provide user name and password for the product services service account.

**NOTE:** This account is granted the **Logon as a service** privilege automatically on the computer where NDC is going to be installed.

- Select additional service options, if necessary.

10. On the **Pre-Installation Tasks and Checks** step, review your configuration and select **Install**.
11. When the installation completes, open a web browser and navigate to the following URL: *http://localhost/conceptQS* where **localhost** is the name or IP address of the computer where Netwrix Data Classification is installed. For example, *http://workstationndc/conceptQS*.

## 6. Upgrade to the Latest Version

Netwrix recommends that you upgrade from the older versions of Netwrix Data Classification to the latest version available in order to take advantage of the new features.

**NOTE:** Seamless upgrade to Netwrix Data Classification 5.5.2 is supported for versions 5.5.1. If you need to upgrade from an earlier version, please perform staged upgrade, e.g., 5.5.0 → 5.5.1 → 5.5.2.

### 6.1. Take Preparatory Steps

Before you start the upgrade, it is strongly recommended to take the following steps:

1. **IMPORTANT!** Make sure you have **.NET Framework 4.7.2** installed on the computer where Netwrix Data Classification resides. If not, download it from Microsoft website: [Download .NET Framework 4.7.2](#).
2. Back up NDC SQL database. For that:
  - a. Start **Microsoft SQL Server Management Studio** and connect to SQL Server instance hosting this database.
  - b. In **Object Explorer**, right-click the database and select **Tasks** → **Back Up**.
  - c. Wait for the process to complete.
3. Back up the Index files.
4. Finally, close administrative web console.

### 6.2. Considerations and Limitations

During the seamless upgrade from previous versions, Netwrix Data Classification preserves its configuration, so you will be able to classify your data right after finishing the upgrade. However, there are some considerations you should examine - they refer to product operation after upgrading from version 5.5.1:

- After the upgrade, you will have to update taxonomies manually. For that:
  - a. In administrative web console, navigate to **Taxonomies** → **Global Settings**.
  - b. Click **Update** in the right corner next to each taxonomy

Netwrix Data Classification 5.5.2

Sources Taxonomies Workflows Config Users Reports Dashboard Help

Term Management Graph User Edits Bulk Updates Taxonomy Settings Global Settings Help

### Global Settings

Taxonomies Boosts

Note: Deleting external taxonomy registrations (SharePoint) does not delete the source taxonomy. The only effect is that the taxonomy is de-registered from the environment.

Delete Export Add

<input type="checkbox"/>	Name	Group Name	Status	Location	Username	
<input checked="" type="checkbox"/>	CCPA	af8c3eb7-4115-4878-9f0d-d79e3d63baf7	Online	SQL		Update   Compare   Edit   Delete
<input checked="" type="checkbox"/>	File Size	c35a73d9-4ba1-432a-976c-6993ae18b702	Online	SQL		Update   Compare   Edit   Delete
<input checked="" type="checkbox"/>	File Type	ff5278a7-0ff7-4a37-bd619-3d3a4a4b911e	Online	SQL		Update   Compare   Edit   Delete
<input checked="" type="checkbox"/>	Financial Records	2844622c-5c81-459e-805e-ea8f80a70a07	Online	SQL		Update   Compare   Edit   Delete
<input checked="" type="checkbox"/>	GDPR Restricted	ba744958-8862-4856-8278-ccc5f33a6234	Online	SQL		Update   Compare   Edit   Delete
<input checked="" type="checkbox"/>	GDPR	3c7e922a-8b46-432f-a166-82a61d9f5d2c	Online	SQL		Update   Compare   Edit   Delete
<input checked="" type="checkbox"/>	GLBA	0a803762-197b-46db-9a8a-278a6f312aaf	Online	SQL		Update   Compare   Edit   Delete
<input checked="" type="checkbox"/>	HIPAA	9a2b780e-4d88-4bca-8a2c-3f322ba379a8	Online	SQL		Update   Compare   Edit   Delete
<input checked="" type="checkbox"/>	Language	d78878a4-3f80-4966-8cc9-b380262d0d43	Online	SQL		Update   Compare   Edit   Delete
<input checked="" type="checkbox"/>	PCI DSS	0fca8b39e-30a6-4181-9383-awcc8ba05188	Online	SQL		Update   Compare   Edit   Delete

- After the upgrade, indexing mode will be set to **Compound Term** mode. Refer to the following Netwrix knowledge base article for instructions on how to modify default Index Processing Mode:

## 7. Configuring NDC Servers Cluster and Load Balancing with DQS Mode

The **Distributed Query Server (DQS)** mode allows you to balance the load between multiple Netwrix Data Classification Servers (NDC Servers) while data collection, indexing and classification. This approach is strongly recommended if you need to process large data volumes, for example:

- **File Servers**—Up to 32 m objects per cluster of 4 servers.
- **SharePoint**—Up to 8 m objects per cluster of 4 servers.

To apply **Distributed Query Server** mode, you need to arrange your NDC Servers in a 'cluster' for load distribution, as described below. Each clustered NDC Server will store its own set of .CSE files — that is, **NDC Index** will be a distributed index. To assemble and combine data required for the search results, each NDC Server will automatically communicate with the other clustered servers.

**NOTE:** All NDC Servers in the cluster will share a single NDC SQL database.

This functionality is implemented through the *QueryServer* application installed together with NDC Server.

### 7.1. Applying DQS Mode

DQS mode can be configured via the administrative web console.

If you want to implement DQS configuration for the your NDC deployment, consider the following:

- This action cannot easily be undone, so before applying the DQS mode, take a full backup of your NDC deployment. Also, read the related documentation sections thoroughly before you start.
- Make sure all servers you plan to add to the DQS cluster have proper network connection and are visible to each other across the network. Adjust firewall settings if necessary.
- Initially, all existing documents will be 'allocated' to the first server in the 'cluster' and then re-distributed across all configured servers.

To be able to configure the DQS mode, current account requires a **Superuser** role.

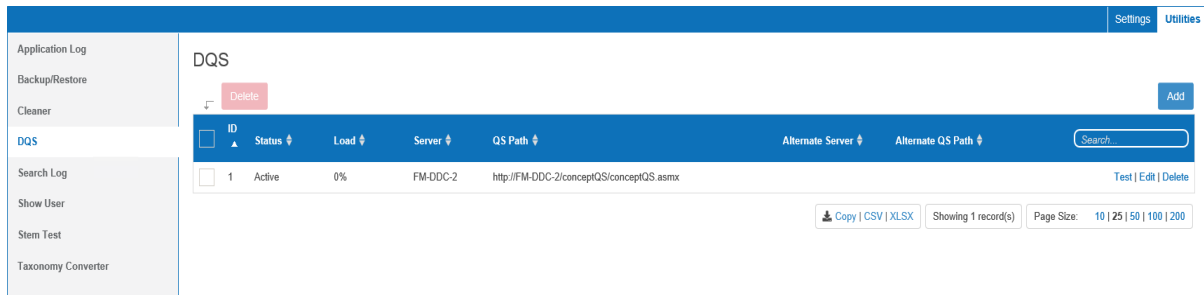
*To arrange NDC Servers cluster and apply DQS mode*

1. Install and configure the first Netwrix Data Classification Server as described in the [Install Netwrix Data Classification](#) section.
2. Open administrative web console.

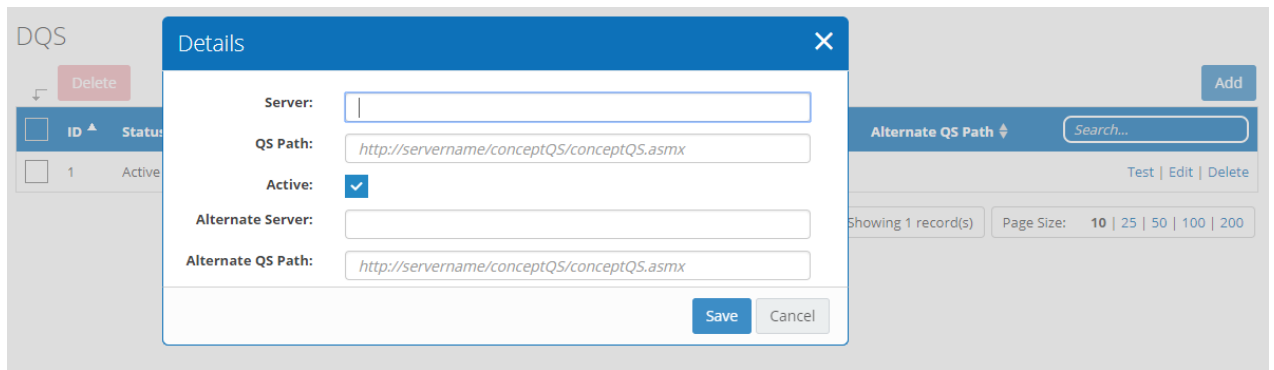
3. Navigate to **Config** → **Utilities** → **DQS**.
4. Select **Enable DQS**.

**NOTE:** Once the DQS mode is enabled, you cannot roll back your configuration. Netrix strongly recommends to ensure that you have taken a full backup of your environment. If ready, confirm the DOS enablement operation when prompted.

5. On the **DQS** tab, click **Add** to add servers you prepared, one by one.



Complete the following fields:



Setting	Value
Server	Provide the NDC Server name or IP address (name format is case-insensitive).
QS Path	Path to the solution component responsible for DQS mode, residing on the server being added. Filled in automatically; leave the default value.
Active	Select to enable clustering for the instance being added.
Alternate Server	Netrix recommends using default values.
Alternate QS Path	Netrix recommends using default values.

6. Click **Save** to close the dialog.



7. Prepare to install other Netrix Data Classification Server instances, assuming each server requires a dedicated machine. Make sure they meet the [Hardware Requirements](#) and general [Software Requirements](#)
8. On each server, follow the installation steps as described in the [Install Netrix Data Classification](#) section until **SQL Database** configuration.
9. On the **SQL Database** step, provide the name of the SQL Server instance that hosts **NDC SQL database** you configured for the first NDC Server.

**NOTE:** Ignore the confirmation dialog on the existing schema in the selected SQL database.

10. Complete the installation.
11. Repeat steps 2 - 6 for every NDC Server, then review the list of servers to make sure the new server was included.

**DQS**

**Information**

The Distributed Query Server (DQS) is a component of conceptClassifier that allows an index to be distributed across multiple servers.

A distributed index means that there are multiple servers running the Collector, Indexer, Classifier, and QueryServer applications - each with its own set of ".cse" files. All servers share a single SQL database.

Please note:

- Each server can only run one set of Windows Services
- The server name should be specified in the NETBIOS format (case insensitive)
- The QS Path specified should be a direct connection to the server in question (I.E not a load balanced address for the cluster)
- The server should be set to "Active" to be considered part of the cluster
- All servers that you wish to run the Windows Services on should be specified within the DQS list

ID	Status	Load	Server	QS Path	Alternate Server	Alternate QS Path
1	Active	0%	FM-DDC-2	http://FM-DDC-2/conceptQS/conceptQS.aspx		
2	Active	0%	fm-ddc-4	http://fm-ddc-4/conceptQS/conceptQS.aspx		
3	Active	0%	fm-ddc-5	http://fm-ddc-5/conceptQS/conceptQS.aspx		
4	Active	0%	fm-ddc-3	http://fm-ddc-3/conceptQS/conceptQS.aspx		

Showing 4 record(s) | Page Size: 10 | 25 | 50 | 100 | 200

12. If you were configuring the DQS mode for the existing NDC deployment, you will be prompted to re-collect data from the data sources —in order to re-distribute the content index across all NDC Servers in the cluster.

**NOTE:** To force re-distribution when necessary, you can use the **Re-Collect** command available after clicking **Run Cleaner** button on the **Config > Settings > Collector** tab.

To review system health and check your configuration, use the product dashboards. See [Dashboards](#) for more information.

## 8. Configure IT Infrastructure

Successful crawling requires certain configuration of your IT infrastructure, which may include enabling Windows services, etc.

Review the following for additional information:

- [Configure Microsoft Exchange for Crawling and Classification](#)
- [Configure NFS File Share for Crawling](#)
- [Configure G Suite and Google Drive for Crawling](#)

## 8.1. Configure Microsoft Exchange for Crawling and Classification

When crawling an Exchange Server, it is necessary to configure sufficient permissions to allow the crawling account to impersonate the mailboxes that you wish to crawl. This requires the setup of two permissions:

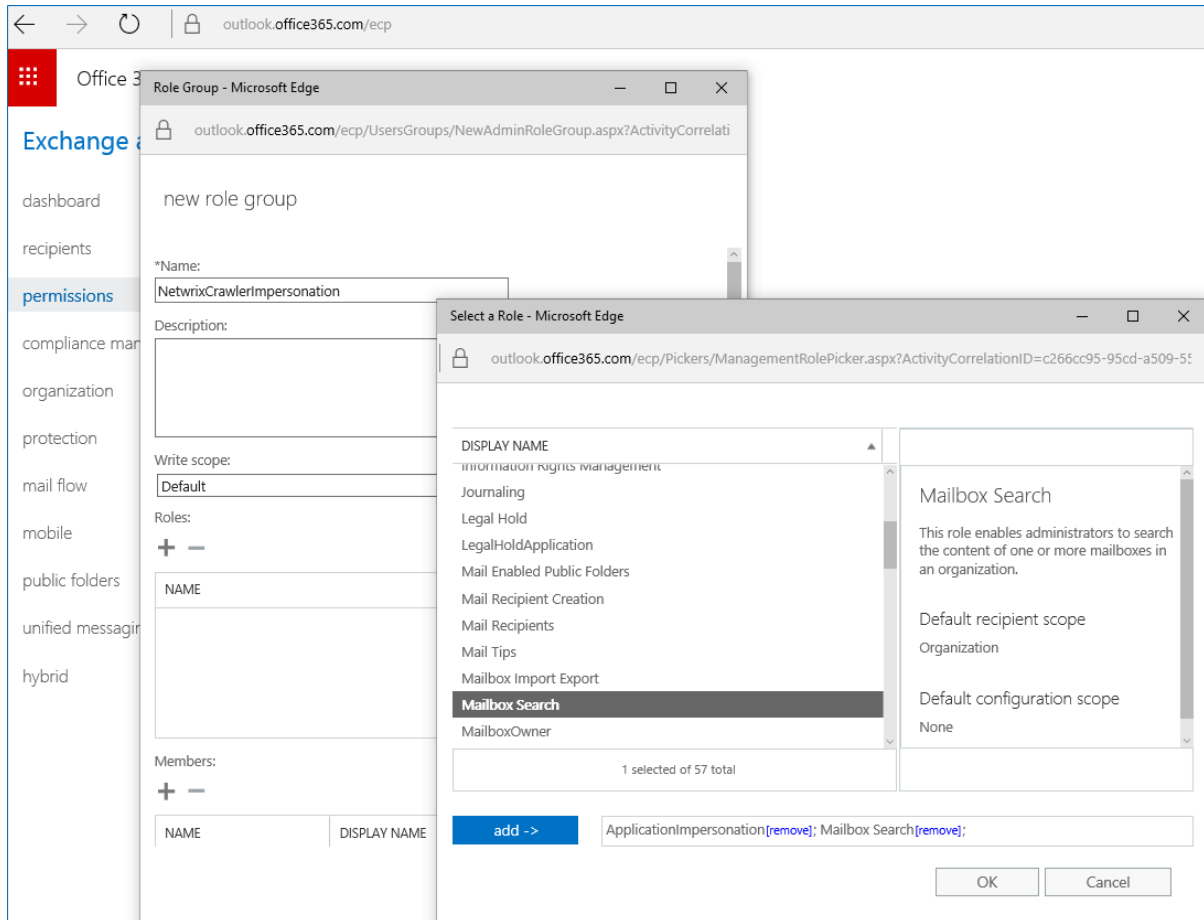
- **ApplicationImpersonation**—Allows the crawling account to impersonate each of the mailboxes / users configured for collection
- **Mailbox Search**—Allows the crawling account to enumerate mailboxes (automatic discovery of mailboxes)

Review the following for additional information:

- [To configure using Office 365 Exchange Admin Portal](#)
- [To configure using Exchange 2010 or later \(on-premise\)](#)
- [To configure match rules](#)

### *To configure using Office 365 Exchange Admin Portal*

1. Login to the [Office 365 Exchange Admin Portal](#)
2. Go to **Permissions**, then under **admin roles** click the '+' symbol to add a new role and enter the Name and Description '*NetwrixCrawlerImpersonation*'.
3. Click the '+' symbol under **Roles**; select **ApplicationImpersonation** and **Mailbox Search** roles.



4. Click **add →** and then **OK**.
5. Click the '+' symbol under **Members:** and select your Admin User.
6. Click **add →** then **OK**.

#### *To configure using Exchange 2010 or later (on-premise)*

1. Login to one of the **Exchange** servers (RDP)
2. Open a **Powershell** window
3. Run the following commands (replacing **ADMINUSERNAME** with the username of your crawling account):

```
New-ManagementRoleAssignment -Name "NetwrixCrawlerImpersonation" -Role
"ApplicationImpersonation" -User ADMINUSERNAME
```

```
New-ManagementRoleAssignment -Name "NetwrixCrawlerSearch" -Role "Mailbox Search" -User
ADMINUSERNAME
```

If crawling **Microsoft Office 365 for Small Business** or many hosted Exchange systems, then it is not possible to setup **Application Impersonation**.

#### *To configure match rules*

The **Match Rules** are an important configuration option, defining which mailboxes will be crawled as part of an Exchange Server source. Here are some example match rules that may be required:

1. `.*@netwrix.com`— Identifies the domain (netwrix.com) within the mailbox name, restricts crawling to a specific set of mailboxes
2. `.*`—Identifies any mailbox, ensuring that all mailboxes will be crawled.

## 8.2. Configure NFS File Share for Crawling

To process NFS file shares, it is necessary to enable specific Windows features. The steps to enable these features differ depending on operating system of the computer where Netwrix Data Classification is installed.

Consider the following:

- NFS File shares are only supported for the machines running Windows Server 2012 and later (server OS) or Windows 10 and later (workstation OS)
- Changes made to files (including adding new files) will not be automatically detected until the source is **re-indexed**—Netwrix recommends setting the **re-index** period for NFS file shares to **1 day**.

### *To configure Windows Server 2012 or later*

1. On the Windows desktop, start **Server Manager**.
2. On the **Manage** menu, click **Add Roles and Features**.
3. Progress to the **Features** step.
4. Ensure that **Client for NFS** option enabled.
5. Complete the wizard.

### *To configure Windows 10*

1. Navigate to Control Panel and select **Programs**.
2. Select **Turn Windows features on or off**.
3. Expand **Services for NFS** and enable the **Client for NFS** option.
4. Click **OK**.

After configuring your NFS share, you will be able to add the **Folder** content source, as described in the [File System](#) section.

**NOTE:** Do not specify username and password while adding data source.

## 8.3. Configure G Suite and Google Drive for Crawling

Netwrix Data Classification can crawl both: Personal Google Drives and G Suite domains. Netwrix Data Classification for Google Drive uses the **OAuth 2.0** protocol to authenticate to your G Suite domain. You will need to create a service account and authorize it to access data in individual and shared Drives on behalf of users using the Google Drive API. Depending on your drive type, do the following: Do the following:

- [To configure G Suite for crawling](#)
- [To configure Personal Google Drive for crawling](#)

### *To configure G Suite for crawling*

#### **In Google API Console:**

1. Create a new project
2. Select Application type
3. Create a new service account
4. Create a service account key (JSON, save a copy for the data source configuration)
5. Enable G Suite domain-wide delegation for the service account (write down the Client ID)
6. Enable Google Drive API

#### **In G Suite Admin Console:**

1. Authorize service account to access the Google Drive API

### *To configure G Suite for crawling*

**IMPORTANT!** Google administrative interfaces tend to change over time, so refer to the following guide for up-to-date instructions on creating OAuth 2.0 service accounts: [Using OAuth 2.0 for Server to Server Applications](#).

Review the following for additional information:

To...	Do...
Create a new project	<ol style="list-style-type: none"> <li>1. Navigate to <a href="https://console.developers.google.com">https://console.developers.google.com</a> (Google API Console) while logged in as a G-Suite administrator within the domain to be crawled (if the user is not added within the correct domain then the correct data will not be identified).</li> <li>2. Create a new project.</li> </ol>
Select Application type	<ol style="list-style-type: none"> <li>1. Once a new project has been created, navigate to <b>APIs&amp;Services</b> → <b>OAuth consent screen</b>.</li> <li>2. Set <b>User type</b> to "Internal".</li> <li>3. Provide the name for new application.</li> <li>4. Click <b>Save</b>.</li> </ol>
Create a new service account	<ol style="list-style-type: none"> <li>1. In <b>Google API</b> console, navigate to <b>IAM &amp; Admin</b> → <b>Service Accounts</b>.</li> </ol>

To...	Do...
	<ol style="list-style-type: none"> <li>2. Create service account as described in Google official <a href="#">article</a>.</li> <li>3. On the <b>Grant this service account access to project (optional)</b> step, do not select any roles.</li> <li>4. On the <b>Grant users access to this service account (optional)</b> step, do not grant any user access. Click <b>Done</b>.</li> </ol>
Create a service account key	<ol style="list-style-type: none"> <li>1. On the <b>Service accounts</b> page, select the account you want to create a key for.</li> <li>2. Click  icon under <b>Actions</b> and select <b>Create key</b>.</li> <li>3. In the <b>Create private key for &lt;Service account name&gt;</b> dialog, select <b>JSON</b> format, and download the file to a known location as it will be required later.</li> </ol> <p><b>NOTE:</b> Your new public / private keypair is generated and downloaded to your machine; it serves as the only copy of this key. You are responsible for storing it securely. If you lose this keypair, you will need to generate a new one.</p>
Delegate domain-wide authority to the service account	<ol style="list-style-type: none"> <li>1. On the <b>Service accounts</b> page, select your service account and click <b>Edit</b>.</li> <li>2. Click the <b>Show Domain-Wide Delegation</b> link and tick the <b>Enable G Suite Domain-wide Delegation</b> checkbox.</li> <li>3. Click <b>Save</b>.</li> <li>4. Once completed, review the "<i>Domain wide delegation</i>" column for this account and make sure that it enabled.</li> <li>5. Click the <b>View Client ID</b> link.</li> <li>6. Copy your Client ID, you will need it later.</li> </ol>
Enable Google Drive API	<ol style="list-style-type: none"> <li>1. In <b>Google API</b> console, navigate to the <b>API Dashboard</b> and select <b>Enable APIs and Services</b> (if APIs have not previously been enabled).</li> <li>2. Search for Google Drive API and click <b>Enable</b> (or <b>Manage</b>).</li> <li>3. Switch to <b>G Suite Admin Console</b>.</li> <li>4. Navigate to <b>Security</b> → <b>Advanced Settings</b> → <b>Manage API Client Access</b> within the Google admin portal.</li> </ol>



To...	Do...
	<ol style="list-style-type: none"> <li>Set the client name to the <b>Client ID</b> you copied on the previous step.</li> <li>Set the API scope to <code>"https://www.googleapis.com/auth/drive"</code> and select <b>Authorize</b>.</li> </ol>

### *To configure Personal Google Drive for crawling*

#### **In Google API Console:**

- Create a new project
- Select Application type
- Create a new service account
- Create a service account key (JSON, save a copy for the data source configuration)
- Enable Google Drive API

#### **In your Google Drive:**

- Allow sharing for your files and folders

Review the following for additional information:

To...	Do...
Create a new project	<ol style="list-style-type: none"> <li>Navigate to <a href="https://console.developers.google.com">https://console.developers.google.com</a> (Google API Console) while logged in as a G-Suite administrator within the domain to be crawled (if the user is not added within the correct domain then the correct data will not be identified).</li> <li>Create a new project.</li> </ol>
Select Application type	<ol style="list-style-type: none"> <li>Once a new project has been created, navigate to <b>APIs&amp;Services</b> → <b>OAuth consent screen</b>.</li> <li>Set <b>User type</b> to <i>"Internal"</i>.</li> <li>Provide the name for new application.</li> <li>Click <b>Save</b>.</li> </ol>
Create a new service account	<ol style="list-style-type: none"> <li>In <b>Google API</b> console, navigate to <b>IAM &amp; Admin</b> → <b>Service Accounts</b>.</li> <li>Create service account as described in Google official <a href="#">article</a>.</li> <li>On the <b>Grant this service account access to project (optional)</b> step, do not select any roles.</li> </ol>

To...	Do...
	<ol style="list-style-type: none"> <li>On the <b>Grant users access to this service account (optional)</b> step, do not grant any user access. Click <b>Done</b>.</li> </ol>
Create a service account key	<ol style="list-style-type: none"> <li>On the <b>Service accounts</b> page, select the account you want to create a key for.</li> <li>Click  icon under <b>Actions</b> and select <b>Create key</b>.</li> <li>In the <b>Create private key for &lt;Service account name&gt;</b> dialog, select <b>JSON</b> format, and download the file to a known location as it will be required later.</li> </ol> <p><b>NOTE:</b> Your new public/private keypair is generated and downloaded to your machine; it serves as the only copy of this key. You are responsible for storing it securely. If you lose this keypair, you will need to generate a new one.</p>
Enable Google Drive API	<ol style="list-style-type: none"> <li>In <b>Google APIs</b> console, navigate to the <b>API Dashboard</b> and select <b>Enable APIs and Services</b> (if APIs have not previously been enabled).</li> <li>Search for Google Drive API and click <b>Enable</b> (or <b>Manage</b>).</li> </ol>
Allow sharing for your files and folders	<ol style="list-style-type: none"> <li>Navigate to each Google Drive account that you wish to crawl</li> <li>Right click each file / folder you wish to crawl and select <b>Share...</b></li> <li>Enter email address of the service account you created on the <b>Create a new service account</b> step. To view email address, do the following: <ul style="list-style-type: none"> <li>In <b>Google API</b> console, navigate to <b>IAM &amp; Admin</b> → <b>Service Accounts</b>.</li> <li>Select your service account and click <b>Edit</b>.</li> <li>Review email address in the <b>Email</b> field.</li> </ul> </li> <li>If you wish to write classifications or apply workflows, ensure that <b>Can organize, add, &amp;edit</b> option is selected (expand the menu to the right of <b>People</b> field).</li> </ol>


## 9. Initial Product Configuration

The **Product Configuration Wizard** allows you quickly configure basic Netwrix Data Classification settings such as processing mode, taxonomies, etc.


In your web browser, navigate to the following URL: `http://hostname/conceptQS` where **hostname** is the name or IP address of the computer where Netwrix Data Classification is installed and perform initial configuration steps.

On the **Instance** step, provide the unique name for your Netwrix Data Classification instance. For example, *"Production"*.


### Product Configuration Wizard

**Instance**  
Set the instance name


>

**Processing Settings**  
Configure how content is processed and classified

>

**Taxonomies**  
Optionally add pre-defined taxonomies

>

**Summary**  
Confirm and save product configuration

---

**What should this instance be called?**

Choose a unique name for this instance. For example: 'Production', 'Development', or 'UAT'.

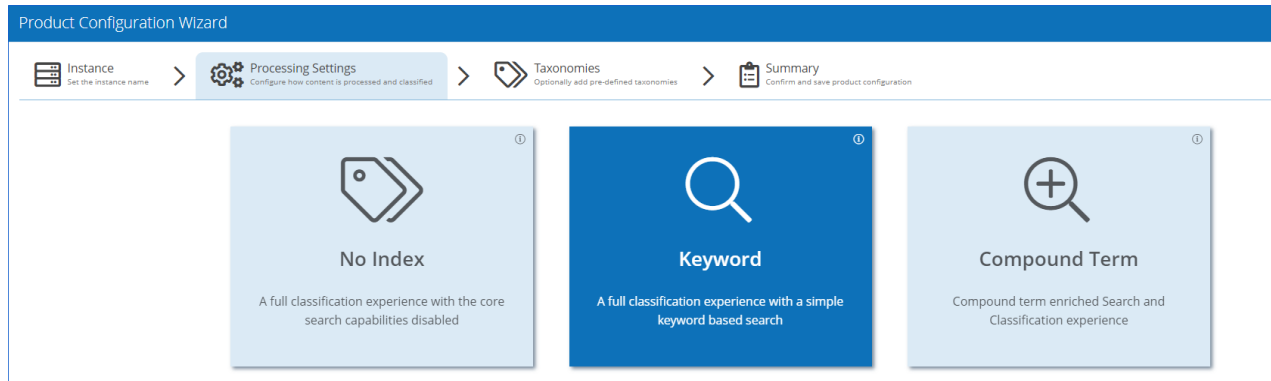
Production|

Click **Next** to proceed. See also:

- [Select Processing Mode](#)
- [Processing Settings](#)
- [Add Taxonomy](#)
- [Review Your Configuration](#)

### 9.1. Select Processing Mode

At this step of the wizard, select processing (indexing) mode for your environment.



For starter and evaluation purpose, select **Keyword** mode.

Review the short description below and select mode:

### 9.1.1. No Index

In this mode, the core search index will be disabled, heavily reducing the disk space requirements for the CSE files and improving overall document throughput for classification. Under this mode **Search** is not available and **Browse** functionality is not subject to security trimming. Recommended for data discovery, data security governance and compliance use cases.

### 9.1.2. Keyword

In this mode the search index will be created; however, disk space required for the core search index will be of medium size. Both **Browse** and **Search** by keyword will be supported. Overall throughput is capable of supporting large number of documents (> 1M). Recommended for compliance, data discovery and classification rules tuning.

### 9.1.3. Compound Term

In this mode you will get a fully featured index, supporting **Search** by compound term. Consider that data storage will require more space, and overall throughput may decrease (compared to the Keyword mode). Recommended for knowledge management, data storage optimization, legal search, other content services.

Proceed with configuring processing settings. See [Processing Settings](#) next.

## 9.2. Processing Settings

On the **Processing Settings** step, select review options for data processing and classification. For test and evaluation purposes, Netwrix recommends use default values.

Review the following for additional information:

45/46

Option	Description
Should boosted regex scoring be enabled?	Enable to boost the score of any regex clues if the regular expression appears multiple times in the document.
How should regular expressions be processed?	Enables and disables case sensitivity when processing regular expressions.
Store trimmed classifications to improve reclassification performance?	Enable to store trimmed classifications to SQL database (trimmed due to the maximum number of classifications being hit for a document). This improves classification performance, however, this may lead to additional data in the SQL database.

Proceed with adding taxonomies.

## 9.3. Add Taxonomy

On this step, you are prompted to load predefined taxonomies.

Product Configuration Wizard

Instance  
Set the instance name

>

Processing Settings  
Configure how content is processed and classified

>

Taxonomies  
Optionally add pre-defined taxonomies

>

Summary  
Confirm and save product configuration

Which preloaded taxonomies would you like to load?

These taxonomies come pre-populated with terms/clues and can be deleted and reloaded as required

☒ GDPR
☒ HIPAA

Click the search bar and select one or several taxonomies you want to add. See [Built-in Taxonomies Overview](#) for the full list of built-in taxonomies supported by Netwrix Data Classification.

## 9.4. Review Your Configuration

On this step, review your configuration. Once you complete the wizard, you can:

- Add a Source
- Add a Taxonomy
- Take the Product Tour
- Get Help