

# Netwrix Data Classification Installation and Configuration Guide

Version: 5.6.1  
8/12/2021



# Table of Contents

1. Supported Data Sources .....	5
2. Deployment Planning .....	6
2.1. NDC Server .....	6
2.1.1. Configuring NDC Servers Cluster and Load Balancing with DQS Mode .....	6
2.1.1.1. Applying DQS Mode .....	7
2.2. Data Storages and Sizing .....	9
2.2.0.1. NDC SQL database .....	9
2.2.0.2. NDC Index .....	10
2.2.1. Scalability and Performance .....	10
3. Requirements to Install Netwrix Data Classification .....	11
3.1. Hardware Requirements .....	12
3.1.1. Netwrix Data Classification Server .....	12
3.1.2. SQL Server .....	12
3.1.3. Network Access .....	13
3.2. Software Requirements .....	14
3.3. Accounts and Required Permissions .....	16
3.4. Supported Content Types .....	17
4. Install Netwrix Data Classification .....	19
4.1. Configure NDC Database .....	21
5. Upgrade to the Latest Version .....	22
5.1. Preparatory Steps .....	22
5.2. Upgrade Process .....	22
5.3. After the Upgrade .....	22
6. Configure IT Infrastructure .....	24
6.1. Configure Box for Crawling .....	25
6.1.1. Prerequisites .....	25
6.1.2. Procedure Steps .....	25
6.1.2.1. Step 1. Create the App .....	26

6.1.2.2. Step 2. Authorize the App .....	27
6.2. Configure Dropbox for Crawling .....	28
6.3. Configure Microsoft Exchange for Crawling and Classification .....	31
6.3.1. Basic Authentication .....	31
6.3.1.1. Exchange Online .....	31
6.3.2. Exchange Server (On-Premise) .....	32
6.3.3. Modern Authentication .....	33
6.3.4. Create Azure AD app for Modern Authentication .....	33
6.3.4.1. Step 1: Prepare application certificate .....	33
6.3.4.2. Step 2: Create and Register a new app in Azure AD .....	33
6.3.4.3. Step 3: Grant Required Permissions .....	34
6.3.4.4. Step 4: Configure Certificates & secrets .....	35
6.3.4.5. Step 5: Obtain Tenant ID .....	35
6.4. Configure NFS File Share for Crawling .....	36
6.5. Configure G Suite and Google Drive for Crawling .....	36
6.6. Accessing SharePoint Online using modern authentication .....	41
6.6.1. Required roles and permissions .....	41
6.6.2. Configuration steps .....	41
6.7. Set Up MIP Integration .....	42
7. Initial Product Configuration .....	44
7.1. Select Processing Mode .....	44
7.1.1. No Index .....	45
7.1.2. Keyword .....	45
7.1.3. Compound Term .....	45
7.2. Processing Settings .....	45
7.3. Add Taxonomy .....	47
7.4. Security .....	47
7.5. Configure Health Alerting .....	48
7.6. Review Your Configuration .....	49

## 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.

© 2021 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>
Box	<ul style="list-style-type: none"><li>• Enterprise</li><li>• Business / Business Plus</li><li>• Starter</li></ul>
Dropbox	<ul style="list-style-type: none"><li>• Business</li></ul>
Exchange	<ul style="list-style-type: none"><li>• Exchange Server 2010 and above</li><li>• Exchange Online</li></ul>
<b>NOTE:</b> Automatic detection, crawling and classification of multiple Exchange mailboxes from the same Exchange server (and, respectively, <i>Exchange Server</i> content source configuration in the NDC web console) is only supported for Exchange Server 2013 or later due to limitations in the Microsoft APIs. For earlier versions, consider using <i>Exchange Mailbox</i> content source.	
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  $\geq 16$  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](#) for more information.

#### 2.1.1. 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 64 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.

### 2.1.1.1. Applying DQS Mode

DQS mode can be configured via the administrative web console.

If you want to implement DQS configuration for 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.

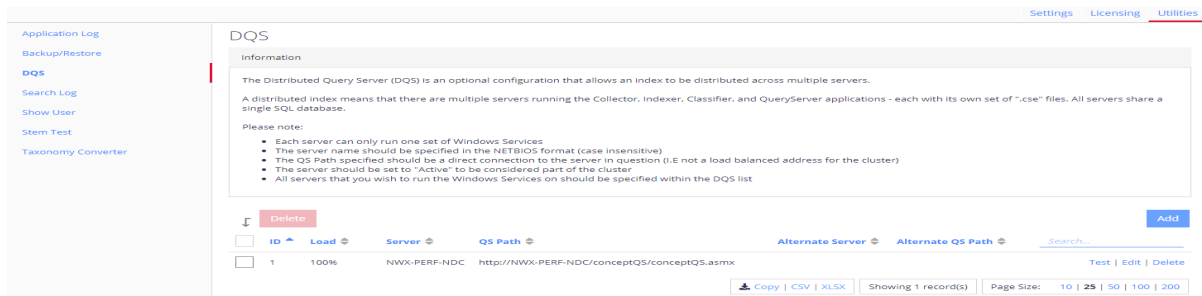
To be able to configure the DQS mode, current account requires a **Superuser** role.

#### *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 **Settings** → **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.



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	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 an optional configuration 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

[Delete](#) [Add](#)

<input type="checkbox"/>	ID	Load	Server	QS Path	Alternate Server	Alternate QS Path	Search...
<input type="checkbox"/>	1	100%	NWX-PERF-NDC	http://NWX-PERF-NDC/conceptQS/conceptQS.asmx			<a href="#">Test</a>   <a href="#">Edit</a>   <a href="#">Delete</a>
<input type="checkbox"/>	2	0%	NWX-PERF-NDC-2	http://NWX-PERF-NDC-2/conceptQS/conceptQS.asmx			<a href="#">Test</a>   <a href="#">Edit</a>   <a href="#">Delete</a>

[Copy](#) | [CSV](#) | [XLSX](#) | Showing 2 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 **Settings > Core > Collector** tab.

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

## 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).

**TIP:** Netwrix recommends using SSD storage for both: database and Netwrix Data Classification servers.

- 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.

## 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 1, 000, 000	Proof-of-concept and small-size environment	
Up to 16, 000, 000	Mid-size environment	
Up to 64, 000, 000	Large-size environment	
More than 64, 000, 000	Extra-large environment	System architect's assistance is required for deployment planning requires

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.

# 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. 16, 000, 000 objects to process.

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

Hardware Component 1 Server per 16 M objects	
Cores	8 Cores
RAM	32 GB
Hard disk	UP TO 35% of all data in scope
Hard drive type	SSD storage (recommended)

### 3.1.2. SQL Server

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

Hardware Component	Up to 16 M objects	Up to 32 M objects and up to 8 M objects for SharePoint	Up to 64 M objects and up to 16 M objects for SharePoint
Processor	8 cores	8 cores	8 cores

Hardware Component	Up to 16 M objects	Up to 32 M objects and up to 8 M objects for SharePoint	Up to 64 M objects and up to 16 M objects for SharePoint
RAM	32 GB	64 GB	128 GB
Hard disk	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> .		
Hard disk type	SSD storage (recommended)		

### 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.

## 3.2. Software Requirements

The table below lists the software requirements for 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> <li>• Named Pipe Activation</li> </ul> <p><b>NOTE:</b> To activate these features, select them under <b>.Net Framework Advanced Services - WCF Services</b> from <b>Windows Features</b>.</p>
SQL Server	<ul style="list-style-type: none"> <li>• <a href="#">SQL Server 2008 R2 Standard Edition</a> (or later).</li> </ul>

Component	Requirements
	<ul style="list-style-type: none"><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>
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 .CSE 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.

## 3.4. Supported Content Types

The table below lists types of content and their default extensions supported out of the box.

**NOTE:** To review the full list of available content types, navigate to **Config → Text Processing → Content Type Extraction Methods**.

Default extension	Content type
.aiff	AIFF
.bmp	Bitmap
.chm	Compiled HTML
.doc	Word
.docx	Word Xml
.dwg	CAD
.eml	Exchange Mail
.flv	FLV
.html	HTML
.java	Java Source
.jpg	JPEG

Default extension	Content type
.mpp	Project
.msg	Message
.pdf	PDF
.png	png
.ppt	Powerpoint
.pptx	Powerpoint Xml
.pub	Publisher
.rar	Archive
.rtf	Rich Text
.tiff	Tiff
.tmp	Unknown
.txt	Text
.vsd	Visio
.vtl	Dictionary / VTL
.wav	WAV
.wp	Word Perfect
.xls	Excel
.xlsx	Excel Xml
.xml	XML
.zip	Archive
.7z	Archive

## 4. 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*.

## 4.1. 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. 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.

### 5.1. Preparatory Steps

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

1. 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. For that, it is recommended to do the following:
  - a. On the computer where Netwrix Data Classification is installed, start Netwrix Data Classification Service Viewer tool. Select **Stop** next to each service.
  - b. Locate the folder containing index files (default location is *C:\Program Files\ConceptSearching\ConceptDB*) and back it up.

**NOTE:** Consider that after the upgrade indexing mode will be set to the **Compound Term**.

### 5.2. Upgrade Process

Seamless upgrade to Netwrix Data Classification 5.6.1 is supported for versions 5.5.4 and 5.5.3.

To upgrade your deployment, after taking the preparatory steps described above, run the product setup and follow the wizard steps. When finished, all solution components will be up and running.

If you need to upgrade from an earlier version, you should perform a staged upgrade, e.g., 5.5.0 → 5.5.1 → 5.5.6.

### 5.3. After the Upgrade

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 several steps you may need to take after upgrading:

1. 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

Global Settings

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

Taxonomies | Backups | 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 BACKUP ADD Search

<input type="checkbox"/>	Name	Group Name	Status	Location	
<input checked="" type="checkbox"/>	CCPA d9f3c9b7-7471-15-4878-995d-d79e3b602671		✓ Online	① SQL	Update   Compare   Backup   Export   Edit   Delete
<input type="checkbox"/>	CMAC a0b6a7b7-26a1-4a96-ad88-ad8f1152a1a63		✓ Online	① SQL	Update   Compare   Backup   Export   Edit   Delete
<input type="checkbox"/>	Financial Records 28a4a23c-5a87-ad9e-9d5e-eed8b0a70367		✓ Online	① SQL	Update   Compare   Backup   Export   Edit   Delete
<input checked="" type="checkbox"/>	GDPR 3c7a7c2a-8a8f-d32f-a166-82a1a1ff562c		✓ Online	① SQL	Update   Compare   Backup   Export   Edit   Delete
<input type="checkbox"/>	GDPR Restricted ba7149f0b-886b-4856-9278-cc593a6234		✓ Online	① SQL	Update   Compare   Backup   Export   Edit   Delete
<input type="checkbox"/>	GLBA 0a803762-197b-46ab-8a66-278a43f12aaf		✓ Online	① SQL	Update   Compare   Backup   Export   Edit   Delete
<input checked="" type="checkbox"/>	HIPAA 0a2b780a-4488-4b0f-bc2c-392296379ca8		✓ Online	① SQL	Update   Compare   Backup   Export   Edit   Delete
<input type="checkbox"/>	PCI DSS 07b0a0fa-30ad-4151-9383-9aee0b9c0188		✓ Online	① SQL	Update   Compare   Backup   Export   Edit   Delete

2. 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: [How to modify Index Processing Mode](#).

## 6. Configure IT Infrastructure

Successful crawling requires a certain configuration of native audit settings in the audited environment. Configuring your IT infrastructure may also include enabling certain built-in Windows services, etc. Proper audit configuration is required to ensure audit data integrity, otherwise your change reports may contain warnings, errors or incomplete audit data.

Review the following for additional information:

- [Configure Box for Crawling](#)
- [Configure Dropbox for Crawling](#)
- [Configure Microsoft Exchange for Crawling and Classification](#)
- [Configure NFS File Share for Crawling](#)
- [Configure G Suite and Google Drive for Crawling](#)
- [Set Up MIP Integration](#)



## 6.1. Configure Box for Crawling

Netwrix Data Classification connects to and crawls a **Box** source via a custom Box app, created within Box management portal.

### 6.1.1. Prerequisites

1. Check if your Box license plan provides the capacity you need. Netwrix Data Classification uses API calls for content crawling (min one API call for a single file). Therefore, if you need to store and crawl a large amount of files in Box (e.g. 100, 000 files), then your selected plan should support a sufficient number of API calls per month. Otherwise, the solution will not be able to crawl your content in one month due to limited number of allowed API calls (e.g. with a *Starter* plan that provides only 25 000 calls per month). So, when selecting a Business subscription plan at <https://www.box.com/pricing>, remember to click **Show more features** in the bottom and examine the information on **API calls per month** supported by each plan.

<div> <div>box</div> <div> <a href="#">Products</a> <a href="#">Solutions</a> <a href="#">Customers</a> <a href="#">Support</a> <a href="#">Pricing</a> <a href="#">Contact</a> +44 808 189 0504 </div> <div> <a href="#">Get Started</a> <a href="#">Login</a> </div> </div>					
	Starter €4.50 <b>€4.27</b> per user/month per annum	Business €19.50 <b>€12.82</b> per user/month per annum	Business Plus €22.50 <b>€21.37</b> per user/month per annum	Enterprise	
	<a href="#">Try It</a>   <a href="#">Buy It</a>	<a href="#">Try It</a>   <a href="#">Buy It</a>	<a href="#">Try It</a>   <a href="#">Buy It</a>	<a href="#">Contact Us</a>	
Optional: Box Governance, Box Relay, Box KeySafe, Box Zones	—	—	✓	✓	
Unlimited integrations, including DLP & eDiscovery	—	—	—	✓	
Relay Lite	—	—	✓	✓	
Device trust (advanced mobile requirements)	—	—	—	✓	
Password policy enforcement	—	—	—	✓	
Document watermarking	—	—	—	✓	
<b>API calls per month</b>	25,000	50,000	50,000	100,000	

[Show less features](#) ⌵

2. Make sure you have configured a valid Google account (with multi-factor authentication supported) and registered it as a *Box Developer Account*. This account is needed to create an app that Netwrix Data Classification will use for interaction with Box API. For more information on the custom apps and Box API, refer to [this article](#).

**NOTE:** Account with multi-factor authentication will be required for private/public keys creation and usage, so if such authentication is not enabled, the program will display a warning message and suggest to configure the necessary settings.

### 6.1.2. Procedure Steps

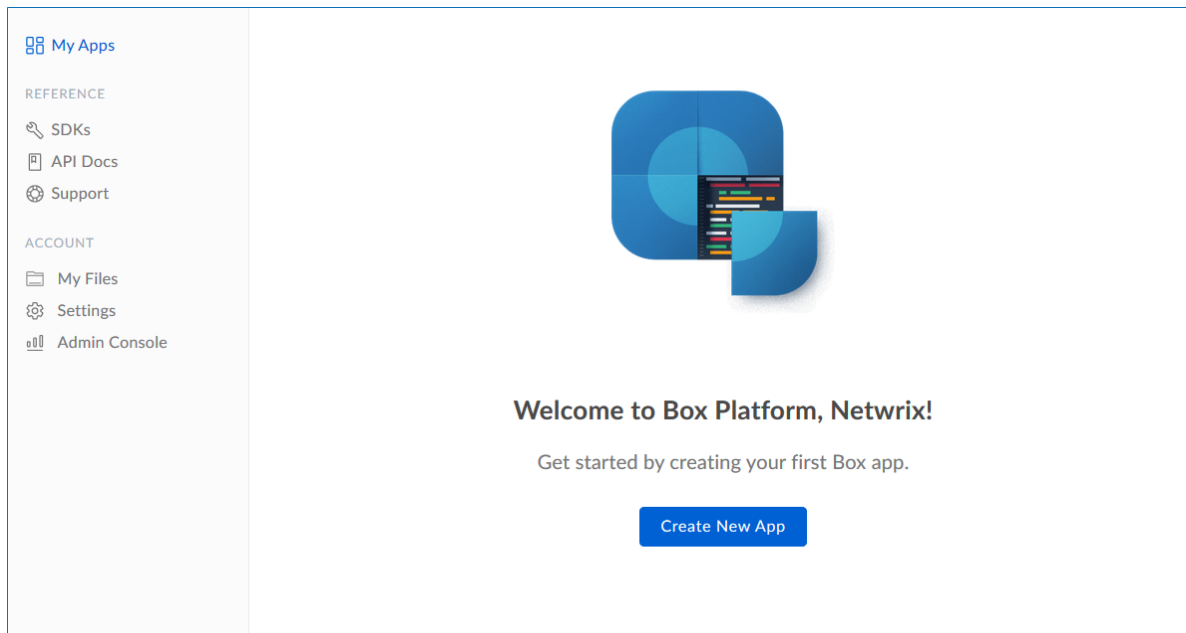
There are four key stages in this procedure:

1. Create an app.
2. Authorize the app to access your organization's data.
3. Register the source with Netwrix Data Classification.
4. Configure content for crawling within Box.

This section describes steps 1 and 2 that are performed on the Box side. Steps 3 and 4 are performed on the Netwrix Data Classification side and described in the [Add Box Source](#) section.

### 6.1.2.1. Step 1. Create the App

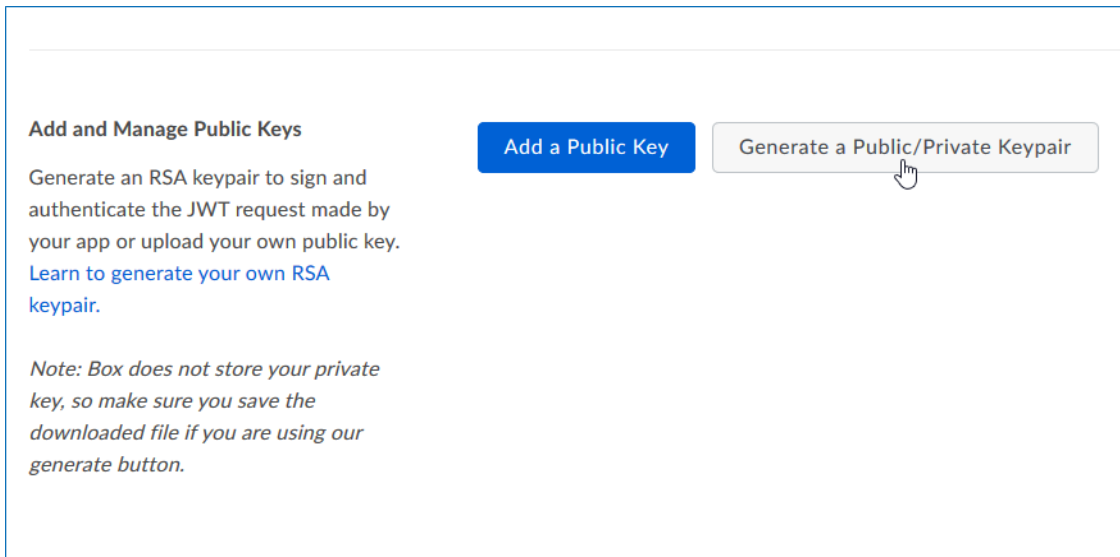
1. Log into your Box cloud-based storage facility using your *Box Developer Account*.
2. Open the Box developer's console endpoint: `https://app.box.com/developers/console`.
3. If you have not created an app before, you will see a screen similar to the one below:



4. Click **Create New App**.
5. Select **Custom App**.
6. Ensure that the **Authentication Method** is selected as **OAuth 2.0 with JWT (Server Authentication)**.
7. Name the app appropriately, typically *Netwrix Data Classification*.
8. Select **View Your App** on the confirmation screen.
9. Open the **Configuration** window by clicking the related item on the left.  
Navigate to **Application Access** and make sure that level is set to **Enterprise**. Click **Save changes**.
10. Navigate to **Advanced Features** and turn ON both switches: **Perform Actions as User** and **Create User Access Token**.

11. Then you should create a public/private key pair to authenticate the JWT requests made by your app. Navigate to **Add and Manage Public Keys** and select **Generate a Public/Private Keypair**.

**NOTE:** If you have not enabled two-factor authentication for the app account in advance, you will be prompted to do it. Click **Settings**, then in the **Account Settings**, navigate to **Authentication**. Select **Require 2-step verification to protect your account**, then provide the necessary information in the **Enable Login Verification** dialog and complete the verification. When finished, get back to the **Configuration** section, clicking the related item in the left pane.



12. You will be notified about downloading a JSON file with all configuration settings of your app.

**IMPORTANT!** Since Box does not store any private keys, this file contains the only copy of your private key, so store it securely.

### 6.1.2.2. Step 2. Authorize the App

1. Go to the **General** section by clicking the item in the left pane.
2. Navigate to **App Authorization** and click **Submit for Authorization**. In the dialog displayed, review the settings.

**NOTE:** If you are a Box administrator, copy the **Client ID** and store it to a safe location.

3. Click **Submit** to send a request to Box administrator.

**IMPORTANT!** If any changes are made to the app configuration later, you will need to re-authorise the app.

If you are a Box administrator, you will receive an email with submitted request. Authorize it, as described in Box documentation. For instance, you can take these steps:

1. Navigate to [box.com](https://box.com) and open the **Admin Console**.
2. Click **Apps** on the left.
3. Navigate to **Custom Apps** and select **Authorise New App**:
4. Enter the **Client ID** of the app you received (the *API Key* in email).
5. Click **Authorize**.

See also:

Box documentation at <https://developer.box.com/guides/authentication/#section-advanced-features>

## 6.2. Configure Dropbox for Crawling

Netwrix Data Classification connects to and crawls a Dropbox source via a custom Dropbox app, created within Dropbox management console.

You will need to create a Dropbox App and authorize it. Do the following:

1. Create a new App - see See [To create a new app](#) for more information.
2. Generate Access token - See [To authorize your app](#) for more information.

### *To create a new app*

To create a new app, you should sign in to Dropbox cloud using a Dropbox Business account with administrative rights. Refer to [Dropbox documentation](#) for more information on the accounts and rights.

1. Navigate to <https://www.dropbox.com/developers/apps/create>
2. On the **Choose an API** step, select **Scoped Access**.
3. On the **Choose the type of access you need** step, select **Full Dropbox** type.
4. Provide a name for your App. For example, *Netwrix Data Classification*.

**NOTE:** Remember to agree with Dropbox API Terms and Conditions.

## 5. Click Create app.

[NDC](#)

### Create a new app on the DBX Platform

1. Choose an API

☐

**Dropbox API**  
For apps that need to access files in Dropbox. [Learn more](#)

☒

**Dropbox Business API**  
For apps that need access to Dropbox Business team info. [Learn more](#)

2. Choose the type of access you need

[Learn more about Business access types](#)

☐ Team information – Information about the team and aggregate usage data.
 ☐ Team auditing – Team information, plus the team's detailed activity log.
 ☒ Team member file access – Team information and auditing, plus the ability to perform any action as any team member.
 ☐ Team member management – Team information, plus the ability to add, edit, and delete team members.

3. Name your app

### To authorize your app

- Once your App has been created, navigate to the **Permissions** tab.
- Select the following permissions and click submit.
  - account\_info.read
  - files.metadata.write
  - files.metadata.read
  - files.content.write
  - files.content.read
  - sharing.read
  - team\_info.read
  - team\_data.member
  - team\_data.team\_space

- files.team\_metadata.write
  - members.read
3. Navigate to the **Settings** tab then scroll down to **OAuth2** option and set the members.read **Access token expiration** to 'No expiration' then click **Generate** under **Generated access token**.

**NOTE:** If you change the app's permissions you will need to regenerate this token.

4. Copy the token to a clipboard. You will need it later when adding a Dropbox source in Netwrix Data Classification administrative web console.

### Netwrix Data Classification

SettingsBrandingAnalytics

Status

Development

Apply for production

Development teams

0 / 1

Enable additional teams

Unlink all teams

Permission type

Team member file access ⓘ

App key

e14z47h81tgg15o

App secret

Show

OAuth 2

Redirect URIs

https:// (http allowed for localhost)

Add

Allow implicit grant ⓘ

Allow ▾

Generated access token ⓘ

Generate

Chooser / Saver / Embedder domains

example.com

Add

If using the [Chooser](#), the [Saver](#), or the [Embedder](#) on a website, add the domain of that site.

Webhooks

Webhook URIs ⓘ

https://

Add

Delete app

Delete app

## 6.3. Configure Microsoft Exchange for Crawling and Classification

When preparing your Exchange Server for data classification, consider that for on-premise Exchange Server, Basic authentication is supported for crawling account, and for Exchange Online you can use either Modern authentication or Basic authentication. Both scenarios are described in the sections below.

### 6.3.1. Basic Authentication

This method is supported for Exchange Online and on-premise Exchange organizations. You should configure sufficient permissions that will allow the crawling account to impersonate the mailboxes that you wish to crawl. This requires the setup of two permissions:

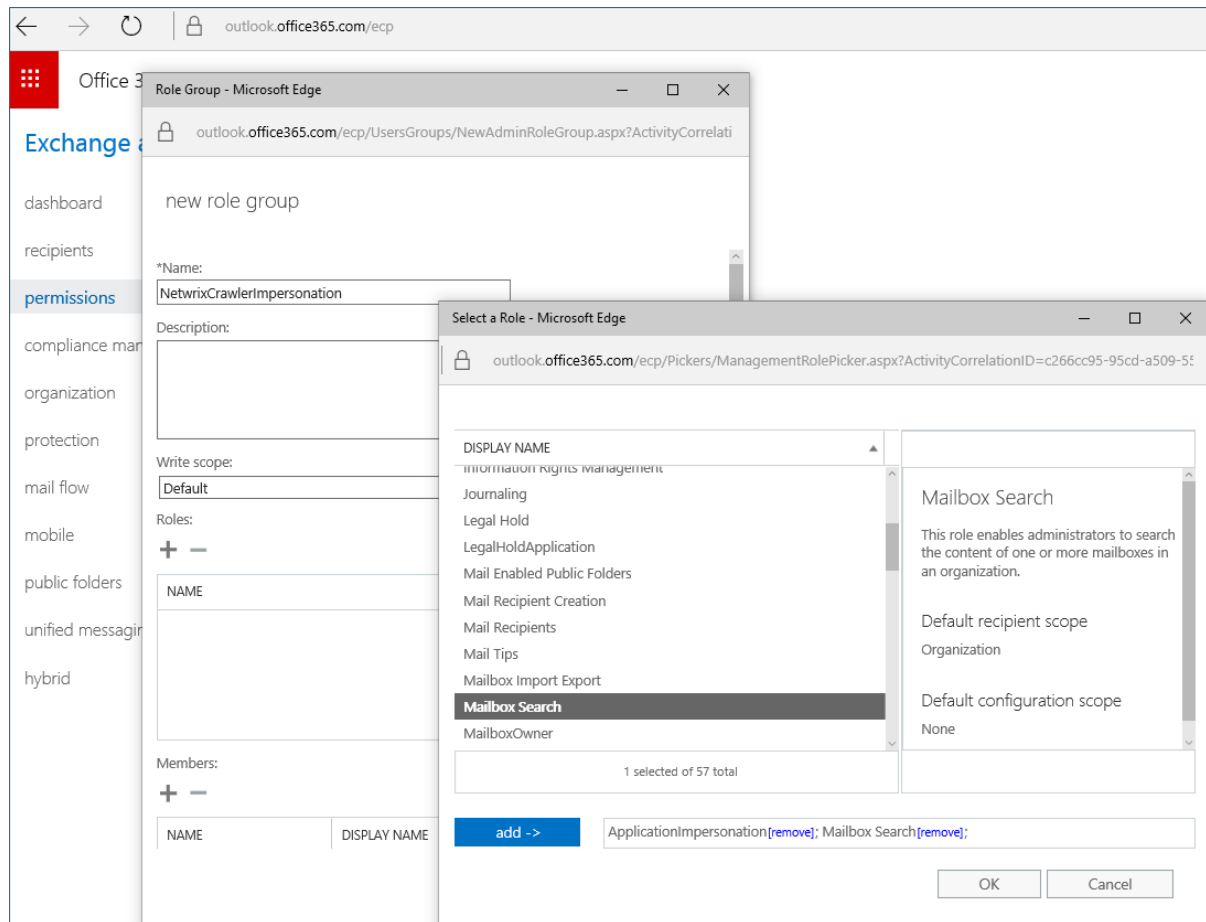
- **ApplicationImpersonation**—Allows the crawling account to impersonate each of the mailboxes / users configured for collection
- **Mailbox Search**—Allows the crawling account to enumerate mailboxes (automatic discovery of mailboxes)

Review the related procedure that corresponds to your Exchange deployment:

- [Exchange Online](#)
- [Exchange Server \(On-Premise\)](#)

#### 6.3.1.1. Exchange Online

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**.

### 6.3.2. Exchange Server (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
```

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



### 6.3.3. Modern Authentication

Starting with version 5.5.3, Netwrix Data Classification allows for crawling Microsoft Exchange Online organization mailboxes using Modern authentication. For that, it uses an Azure AD application which can leverage Microsoft API to connect to Exchange Online organization.

If you plan to implement the scenario that involves modern authentication, you should do the following:

1. [Create Azure AD app for Modern Authentication](#)
2. Configure [Exchange Server](#) source settings.

### 6.3.4. Create Azure AD app for Modern Authentication

To connect to Exchange Online organization that uses Modern authentication, you need to create an Azure AD application, as described in this section.

#### 6.3.4.1. Step 1: Prepare application certificate

Prepare application certificate as follows:

1. Create (or load) an IIS certificate on NDC Server (recommended).

**NOTE:** This certificate should be installed for the local machine so that it can be accessed by Netwrix Data Classification and other services.

2. Export the certificate (.CER file):
  - a. Open the certificate in IIS management console.
  - b. Go to the **Details** tab.
  - c. Select **Copy to File**.

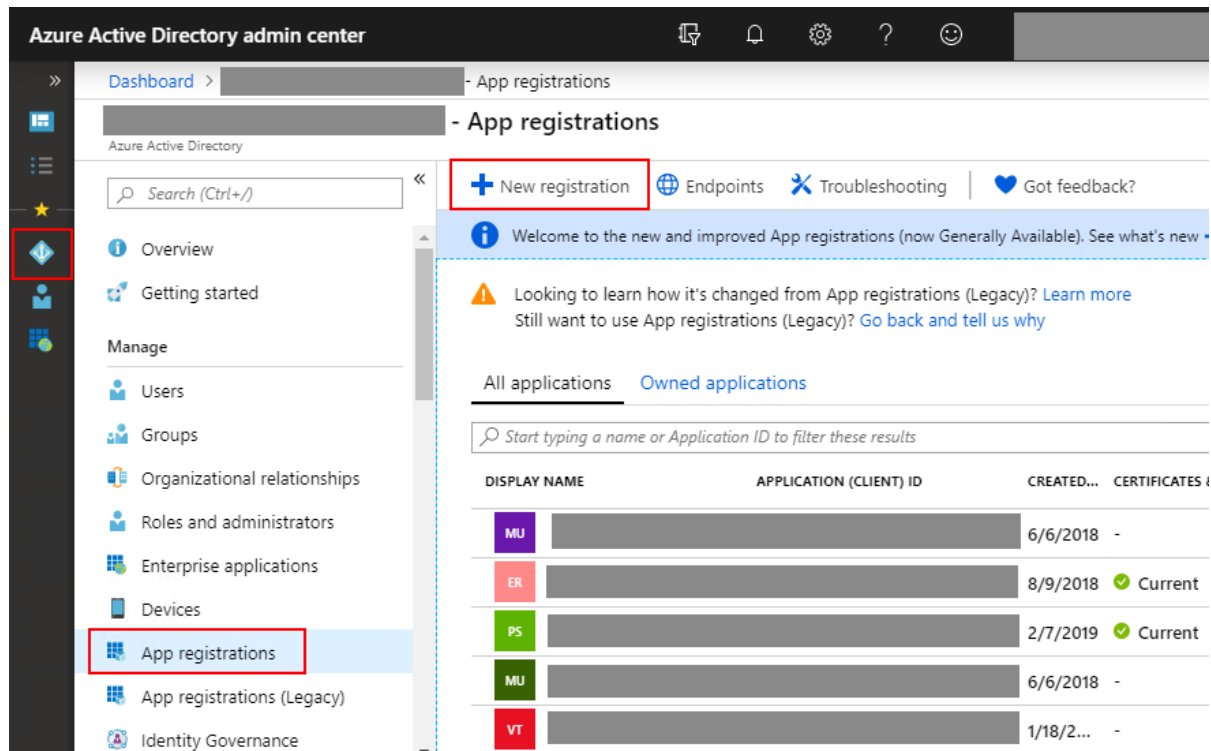
**NOTE:** Do not export private key.

- d. Set file type to *DER-encoded CER*.

#### 6.3.4.2. Step 2: Create and Register a new app in Azure AD

To register a new Azure AD application, do the following:

1. Sign into the **Microsoft 365 Admin Center** (with your *Global Administrator, Application Administrator* or *Cloud Application Administrator* account) and go to the **Azure Active Directory admin center**.
2. Under the **App registrations** section, select **New registration**:



3. In the **Name** field, enter the application name.
4. In the **Supported account types** select who can use this application – use the **Accounts in this organizational directory only** option.
5. Click the **Register** button.

**NOTE:** Application redirect URI is optional, you can leave it blank on this step.

6. Your application ID is now available in the **Overview** section. Copy it to a safe location.

### 6.3.4.3. Step 3: Grant Required Permissions

Next, you need to grant your new application the required API permissions.

Azure AD applications can be assigned *Delegated* or *Application* permissions:

- *Delegated* permissions require a signed-in user present who consents to the permissions every time an API call is sent.
- *Application* permissions are consented by an administrator once granted.

For the newly created app, you should use *Application* permissions.

**NOTE:** By default, a new application is granted one delegated permission for **Microsoft Graph API – User.Read**. It is not required and can be removed.

Do the following:

When found, click on the entry and proceed with adding the necessary permissions. The steps from here on remain the same, so in most cases you would need the Application permissions entry, and the relevant set of permissions therein (such as `full_access_as_app` for EWS OAuth, `Exchange.ManageAsApp` for CBA). Select the relevant entries, hit the Add permissions

1. At the top of the **Request API permissions** pane, click the **APIs my organization uses** tab and search for *Office 365 Exchange Online*.
2. Click on the *Office 365 Exchange Online* entry in the list of apps found.
3. Proceed with adding the permissions for this app: select **Application permissions** and then select **full\_access\_as\_app**.
4. Click **Add permissions**.

Finally, you need to grant admin consent to the tenant (that is, for Exchange organization whose audit data will be collected by the newly registered app).

Do the following:

1. Go to the new app settings > **API permissions** and click **Grant admin consent for <tenant name>**.
2. When prompted to confirm granting, click **Yes**.

#### 6.3.4.4. Step 4: Configure Certificates & secrets

Having configured the app, you can upload its application certificate.

1. In the app settings, click **Certificates & secrets** and select **Upload certificate**.
2. Upload the .CER file you prepared at [Step 1: Prepare application certificate](#).
3. Copy the certificate thumbprint to a safe location.

#### 6.3.4.5. Step 5: Obtain Tenant ID

1. Open **Azure Active Directory admin center**.
2. Select **Azure Active Directory** > **Overview** section for the required Exchange Online organization.
3. Locate the **Tenant ID** and copy it to a safe location.

## 6.4. Configure NFS File Share for Crawling

To enable processing Network File System (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.

**NOTE:** Prior to configuration, consider the following:

- NFS File shares are only supported from servers running Windows Server 2012 or later (or Windows 10)
- Writing classifications to NFS file shares is only supported from Netwrix Data Classification 5.4.8 onwards
- 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**.

Add the Folder source as described in the [File System](#) section.

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

### *To configure Windows Server 2012 Onward*

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**.

## 6.5. 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 Cloud Platform web 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 Cloud Platform web 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 <i>"Internal"</i>.</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 Cloud Platform</b> web console, navigate to <b>Credentials</b> and click <b>Create Credentials</b>.</li> </ol>

To...	Do...
	<ol style="list-style-type: none"> <li>Then, click <b>Service account</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> <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> section, click edit on 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>
Delegate domain-wide authority to the service account	<ol style="list-style-type: none"> <li>On the <b>Service accounts</b> section, select your service account and click <b>Edit</b>.</li> <li>Click the <b>Show Domain-Wide Delegation</b> link and tick the <b>Enable G Suite Domain-wide Delegation</b> checkbox.</li> <li>Click <b>Save</b>.</li> <li>Once completed, review the "<i>Domain wide delegation</i>" column for this account and make sure that the delegation enabled.</li> <li>Click the <b>View Client ID</b> link.</li> <li>Copy your Client ID, you will need it later.</li> </ol>
Enable Google Drive API	<ol style="list-style-type: none"> <li>In <b>Google Cloud Platform</b> web 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 <b>Google Drive API</b> and click <b>Enable</b> (or <b>Manage</b>).</li> <li>Search for <b>Admin SDK API</b> and click <b>Enable</b> (or <b>Manage</b>).</li> <li>Switch to <b>G Suite Admin Console</b>.</li> <li>Navigate to <b>Security</b> → <b>API Controls</b> → <b>Manage Domain-wide</b></li> </ol>

To...	Do...
	<p><b>Delegation</b> within the Google admin portal.</p> <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 scopes and select <b>Authorize</b>: <ul style="list-style-type: none"> <li><a href="https://www.googleapis.com/auth/drive">https://www.googleapis.com/auth/drive</a></li> <li><a href="https://www.googleapis.com/auth/admin.directory.user">https://www.googleapis.com/auth/admin.directory.user</a></li> </ul> </li> </ol>

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

In Google Cloud Platform web 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 Cloud Platform web 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 Google Cloud Platform web 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>
Enable Google Drive API	<ol style="list-style-type: none"> <li>1. In Google Cloud Platform web 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> </ol>
Allow sharing for your files and folders	<ol style="list-style-type: none"> <li>1. Navigate to each Google Drive account that you wish to crawl</li> <li>2. Right click each file / folder you wish to crawl and select <b>Share...</b></li> <li>3. 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 Google API 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>4. 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>



## 6.6. Accessing SharePoint Online using modern authentication

This option is recommended for organizations that use modern authentication as the identity management approach, having multi-factor authentication (MFA) enabled for their user accounts. In this scenario, Netwrix Data Classification will access the cloud-based infrastructure via Microsoft Graph and other modern APIs, being authenticated through a pre-configured Azure AD application with appropriate access permissions.

So, if you plan to implement such scenario, you should register an Azure AD app manually and provide its settings to Netwrix Data Classification when configuring a monitored item.

### 6.6.1. Required roles and permissions

Ensure that the following API permissions are set (and have been granted for the tenant):

- **Graph** – Application permissions (With admin consent granted)
  - **Sites.FullControl.All** (Crawling)
- **SharePoint** – Application permissions (With admin consent granted)
  - **Sites.FullControl.All** (Crawling)
  - **TermStore.ReadWrite.All** (Term Set access)

**NOTE:** for taxonomy manager to full operate you must also make the user “app@sharepoint” a taxonomy admin (or group admin)

### 6.6.2. Configuration steps

1. Set up app registration in Azure (as per Microsoft documentation)
2. Grant required permissions to that application. - Run through the steps described in the ‘Register your application’ section of the Microsoft technical article: [Granting access via Azure AD App-Only](#)
3. Create (or load) certificate in IIS.

**NOTE:** This certificate should be installed for the local machine so that it can be accessed by NDC and the services

4. Export the CER file:
  - Open the certificate in IIS
  - Go to the **Details** tab
  - Select **Copy to File**

- Do not export the private key
  - Set type to "DER-encoded CER"
5. Upload the CER file to Azure
- Navigate to the app (Azure Active Directory → App Registrations)
  - Select **Certificates & Secrets**
  - Upload the certificate
  - Make a note of the certificate thumbprint
6. Create the source using the:
- **ApplicationId/TenantId** can be found in Azure for the App registration
  - **Instance URL** can be left as the default value
  - Enter the certificate thumbprint from Azure

## 6.7. Set Up MIP Integration

To integrate Netwrix Data Classification with MIP technology, you need to perform the following steps:

- Set up application registration in Microsoft Azure
- Load your certificate to Internet Information Services (IIS) Manager
- Export the certificate as .CER file
- Upload the .CER file to Azure

### *To configure IT infrastructure for MIP integration*

Review the following for additional information:

To...	Do...
Set up application registration in Microsoft Azure	<p>Run through the steps described in the following Microsoft article: <a href="#">Register a client application with Azure Active Directory</a>.</p> <p>Make sure that the following API permissions are set (and have been granted for the tenant):</p> <ul style="list-style-type: none"> <li>• Azure Rights Management Services               <ul style="list-style-type: none"> <li>◦ Content.DelegatedWriter</li> <li>◦ Content.Writer</li> </ul> </li> <li>• Microsoft.Graph               <ul style="list-style-type: none"> <li>◦ User.Read</li> </ul> </li> </ul>

To...	Do...
	<ul style="list-style-type: none"> <li>• Microsoft Information Protection Sync Service               <ul style="list-style-type: none"> <li>◦ UnifiedPolicy.Tenant.Read</li> </ul> </li> </ul>
Import certificate in IIS	<p>You can generate custom certificate in IIS which is enough for test and evaluation purposes. However, for production environments, Netwrix recommends importing certificate used by your company. Contact your security administrator to get the certificate.</p> <p><b>NOTE:</b> This certificate need to be installed to the computer where Netwrix Data Classification and all its services run.</p>
Export .CER file	<ol style="list-style-type: none"> <li>1. In <b>Internet Information Services (IIS) Manager</b>, select the certificate you loaded.</li> <li>2. Select <b>View</b> under <b>Actions</b>.</li> <li>3. Go to <b>Details</b> tab and select <b>Copy to File</b>.</li> <li>4. Proceed with Certificate Export wizard.</li> <li>5. On the <b>Export Private Key</b> step, select <b>Do not export the private key</b>.</li> <li>6. On the <b>Export File Format</b> step, select <b>DER encoded binary (.CER)</b>.</li> <li>7. On the <b>File to Export</b> step, select path to store the file.</li> <li>8. Review export settings and click <b>Finish</b>.</li> </ol>
Upload the .CER file to Azure	<ol style="list-style-type: none"> <li>1. Open <a href="#">Microsoft Azure portal</a> and navigate to <b>Azure Active Directory → App Registrations</b>.</li> <li>2. Select application you registered on the <b>Set up application registration in Microsoft Azure</b> step.</li> <li>3. Navigate to <b>Certificates &amp; secrets</b> on the left.</li> <li>4. Click <b>Upload certificate</b>.</li> <li>5. Browse for .CER file you exported and click <b>Add</b>.</li> <li>6. Copy certificate thumbprint to a known location as it will be required later.</li> </ol>


# 7. 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

>

 Security  
Restrict product access

>

 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'.

**Send anonymous usage statistics?**  
When enabled Netwrix Data Classification will upload a small amount of information on the way the product is being used to better improve our offering in the future. Personal/company information/data will not be sent.

☒ Send ☐ Don't Send

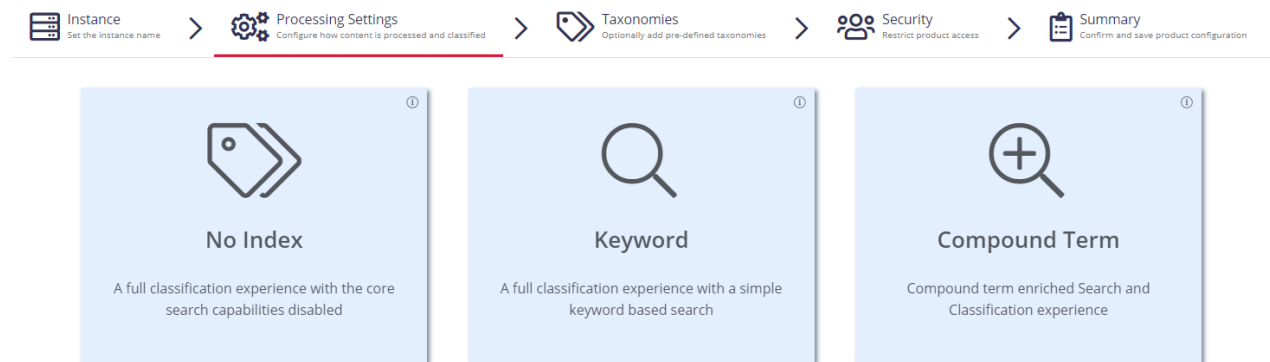
Click **Next** to proceed. See also:

- [Select Processing Mode](#)
- [Processing Settings](#)
- [Add Taxonomy](#)
- [Security](#)
- [Configure Health Alerting](#)
- [Review Your Configuration](#)

## 7.1. Select Processing Mode

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

## Product Configuration Wizard



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

Review the short description below and select mode:

### 7.1.1. 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.

### 7.1.2. 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.

### 7.1.3. 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.

## 7.2. 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.

## Product Configuration Wizard

Instance

Set the instance name

>

Processing Settings

Configure how content is processed and classified

>

Taxonomies

Optionally add pre-defined taxonomies

>

Security

Restrict product access

>

Summary

Confirm and save product configuration

### Text Extraction

**Should OCR be used on image files?**  
 OCR is used to extract text from images. This is useful if the content being collected contains a large number of scanned documents (for example, image file extensions will be automatically added to the list of "Files Included" if this setting is enabled).

☒ Yes ☐ No

**Information**  
 OCR requires the Visual C++ Redistributable for Visual Studio 2015, which is available from the following [link](#).

**Should images embedded in documents be processed?**  
 Images inside office documents (e.g. .DOC and .XLS files) or PDF files can be processed using OCR. Any text extracted will be appended to the document text. Note that this option can dramatically affect the processing speed of content.

☐ Yes ☒ No

**Should the collection process optimise text storage by re-using text offsets?**  
 This reduces the storage requirements for the local database (stored text) by sharing and reusing the stored text when matches are identified. However, this does result in a small increase in sql database demands.

☐ Yes ☒ No

### Classification Configuration

**Should default clues be automatically created?**  
 When enabled a clue will automatically be created when a taxonomy is registered from SharePoint or a term is created. The new clue will either be a standard clue matching the term name or a metadata clue depending on the configuration specified at the taxonomy level settings.

☐ Yes ☒ No

**Should boosted phrasematch scoring be enabled?**  
 When switched on, the score of any phrasematch clues will be boosted if the phrase appears multiple times in the document.

☒ Yes ☐ No

**Should boosted regex scoring be enabled?**

Review the following for additional information:

Option	Description
<b>Text Extraction</b>	
Should OCR be used on image files?	Optical Characters Recognition is a technology used to extract text from images. Enable OCR if the content being collected contains a large number of scanned documents (for example).
	<b>IMPORTANT!</b> OCR requires the Visual C++ Redistributable for Visual Studio 2015. Visit Microsoft <a href="#">website</a> for downloading.
Should images embedded in documents be processed?	Enable this option to recognize documents with integrated images.
Should the collection process optimise text storage by re-using text offsets?	Enable this option to use text offsets.
<b>Classification Configuration</b>	
Should default clues be automatically created?	
Should boosted phrasematch scoring be enabled?	Enable to boost the score of any phrasematch clues if the phrase appears multiple times in the document.

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.

## 7.3. 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

>

Security  
Restrict product access

>

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

☒ HIPAA
 ☒ CCPA

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.

## 7.4. Security

On this step, you are prompted to restrict access to administrative web console by adding users.

Product Configuration Wizard

Instance  
Set the instance name

>

Processing Settings  
Configure how content is processed and classified

>

Taxonomies  
Add pre-defined taxonomies

>

**Security**  
Restrict product access

>

Summary  
Confirm and save product configuration

Do you want to enable user management?

By default any authenticated users have access to the Netwrix Data Classification administration console. We recommend enabling user management in order to restrict access.

☒ Enable user management  
☐ Allow access for all users

**Information**  
 User management is already enabled.

- **Enable user management** – select to add super users and prevent unauthorized access to administrative web console. By default, any authenticated users have access to the console.

**NOTE:** Netwrix recommends enabling this option.

When selected, you are prompted to add super users. Type the name of the new user and click + on the right.

- **Allow access for all users** – select to allow any user access administrative web console.

## 7.5. Configure Health Alerting

On this step, you will be prompted to email settings for health reporting and select immediate health alerts.

Product Configuration Wizard

Instance  
Basic details

Processing Settings  
Configure how content is processed

Taxonomies  
Add pre-defined taxonomies

Security  
Restrict product access

Health  
Configure alerting

Summary  
Confirm configuration

**Would you like to configure Health Reporting?**  
Do you want to set up email notifications for health alerts? It can be done now or at a later date in the communications settings area (requires an SMTP server).

☒ Set up now  
☐ Configure at a different time

**Who should the email be sent from?**  
Choose the email account you wish for this email to be sent from.

Please Select ▼ + ↻

**Specific recipient(s)**  
Enter one or more recipients.

+

**What sort of immediate alerts should be sent?**  
Based on your selection you will be emailed for different levels of health alert when they occur.

☐ Don't send any alerts  
☒ Send Errors only  
☐ Send emails for Errors and Warnings

**Should a daily health summary be sent?**  
By selecting this you will receive an email you every day that there are errors and/or warnings.

☒ Send a daily summary  
☐ Don't send an email daily

Complete the following fields:

Setting	Description
Would you like to configure Health Reporting?	Select <b>Setup now</b> if you want to receive health alerts. You can do it later in the communication settings area. See <a href="#">System Health</a> for more information.
Who should the email be sent from?	Select a user registered in Netwrix Data Classification administrative web console in the field or go to the <b>Specific recipients</b> below and specify one or more email addresses outside your organization.
What sort of immediate alerts should be sent?	Select the appropriate alerting level: do not receive any alerts at all, receive errors only, or get both: emails for errors and warnings. See <a href="#">System Health</a> for more information.
Should a daily health summary be sent?	Select whether you want to receive daily summary on the product health.



## 7.6. 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](#)