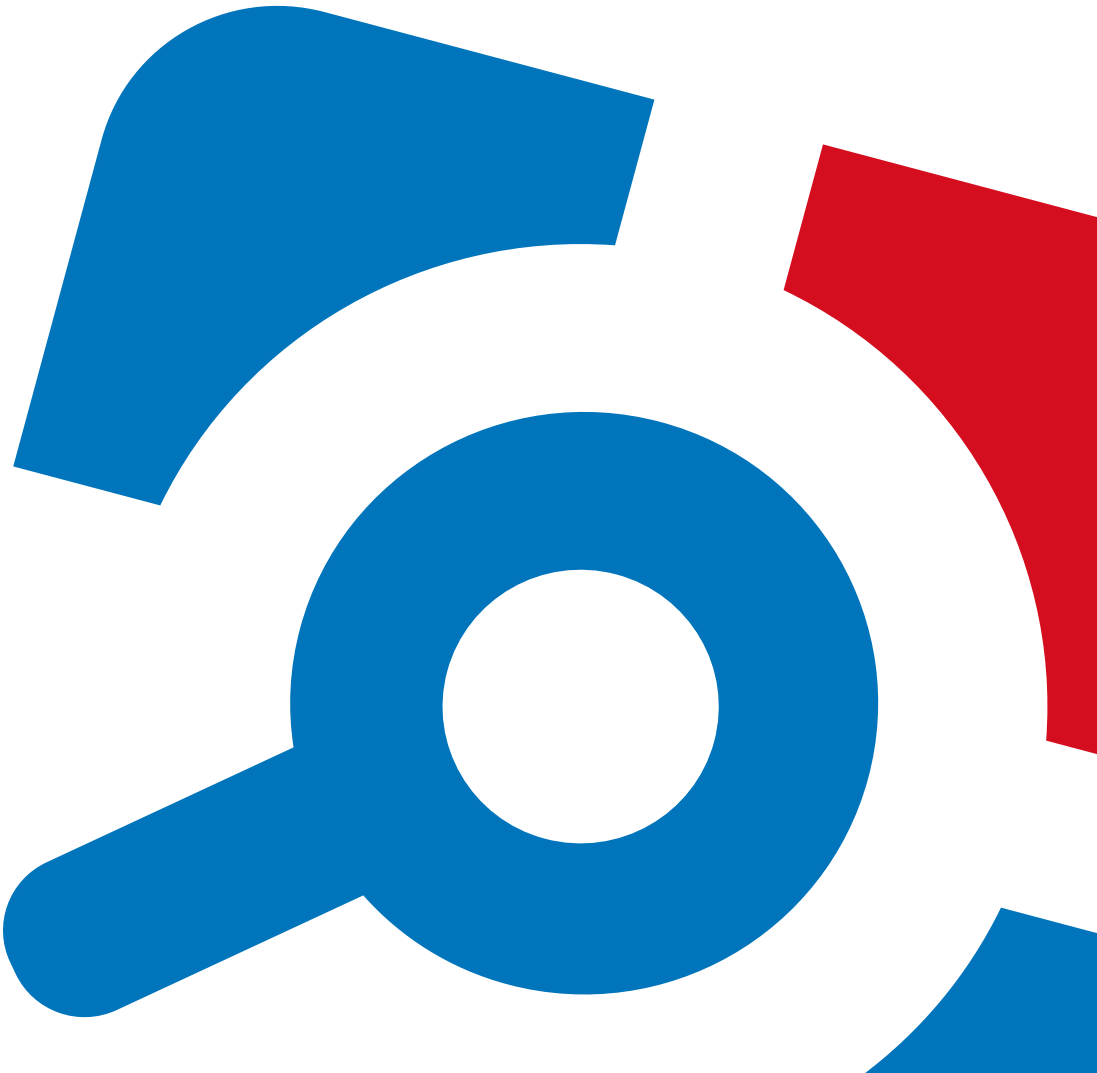
netwrix

Netwrix Auditor Data Discovery and Classification

Quick-Start Guide

Version: 9.9
3/13/2020



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.

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.

Table of Contents

1. Introduction

This guide is intended for the first-time users of Netwrix Auditor Data Discovery and Classification. It can be used for evaluation purposes, therefore, it is recommended to read it sequentially, and follow the instructions in the order they are provided. After reading this guide you will be able to:

- Install and configure DDC Collector
- Configure data sources in Netwrix Auditor
- Enable and configure DDC Provider
- Review Data Discovery and Classification reports

NOTE: The DDC Collector and DDC Provider work only in combination with supported Netwrix Auditor applications; so this guide covers a basic procedure for running the modules and assumes that you have Netwrix Auditor installed and configured in your environment. For installation scenarios, data collection options, as well as detailed information on how Netwrix Auditor works, refer to the following Quick-Start Guides, depending on your data source:

- Netwrix Auditor for Windows File Servers Quick-Start Guide
- Netwrix Auditor for EMC Quick-Start Guide
- Netwrix Auditor for NetApp Quick-Start Guide
- Netwrix Auditor for SharePoint Quick-Start Guide

1.1. Netwrix Auditor Features and Benefits

Netwrix Auditor is a visibility platform for user behavior analysis and risk mitigation that enables control over changes, configurations and access in hybrid IT environments to protect data regardless of its location. The platform provides security analytics to detect anomalies in user behavior and investigate threat patterns before a data breach occurs.

Netwrix Auditor includes applications for Active Directory, Active Directory Federation Services, Azure AD, Exchange, Office 365, Windows file servers, EMC storage devices, NetApp filer appliances, Nutanix Files, network devices, SharePoint, Oracle Database, SQL Server, VMware, Windows Server, and User Activity. Empowered with a RESTful API, the platform delivers visibility and control across all of your on-premises or cloud-based IT systems in a unified way.

Major benefits:

- Detect insider threats—on premises and in the cloud
- Pass compliance audits with less effort and expense
- Increase productivity of IT security and operations teams

To learn how Netwrix Auditor can help your achieve your specific business objectives, refer to Netwrix Auditor Best Practices Guide.

2. Data Discovery and Classification Overview

Netwrix Auditor's Data Discovery and Classification gives you complete visibility into where your sensitive files are, what content is inside them, who can access these files and who actually uses them. With this actionable information, your risk, compliance and data security officers and IT security pros can prioritize their efforts and secure data in accordance with its value or sensitivity. Your organization will be able to mitigate the risk of PII, PHI, PCI and IP being stored outside dedicated locations, and apply controls and policies consistently and accurately, ensuring both data security and regulatory compliance.

With Netwrix Auditor Data Discovery and Classification, you can identify, classify and secure sensitive data on Windows file servers, EMC storage devices and NetApp filer appliances, and SharePoint sites.

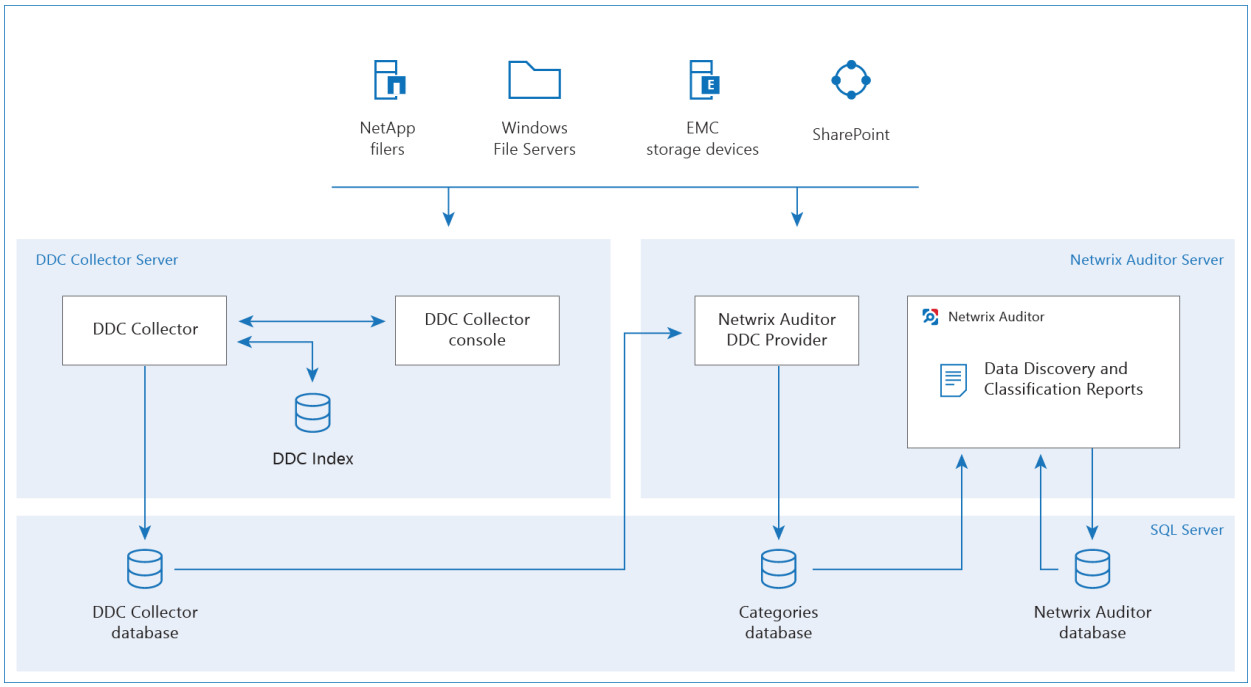
Major benefits:

- Gain a high-level view of the sensitive data you store
- Discover sensitive data stored outside of a secure dedicated location
- Streamline regular attestations of access rights to sensitive data
- Detect unauthorized activity that might threaten your sensitive data

Looking for real-life use cases and walk through examples? Check out Netwrix training materials. Go the Data Remediation Workflows page on Netwrix website.

2.1. How It Works

The following diagram illustrates the data flows in a typical deployment of Netwrix Auditor Data Discovery and Classification:



The **DDC Collector** is a data discovery and classification service that runs on a dedicated server. It scans your various file repositories for supported file content, stores the raw text in the **DDC Index** and indexes that content. It classifies the indexed file content by matching it against predefined third-party taxonomies (rules and patterns for finding, for example, personal data governed by the GDPR or medical records governed by HIPAA) and any custom taxonomies you create. It stores the resulting document classifications in the DDC Collector database. You use the DDC Collector console to monitor and control the DDC Collector service, as well as to select, create, modify and manage taxonomies.

Meanwhile, the DDC Provider service runs on the **Netwrix Auditor Server**. It reads the classification results from the **DDC Collector database** and translates the **DDC Collector taxonomy** format into the **Netwrix Auditor category** format the resulting list of objects and their categories is periodically transferred to the **Categories** database.

Netwrix Auditor merges data from the **Categories** database and other Netwrix Auditor databases (such as the file server State-in-Time database) to generate the **reports** you request.

2.2. PoC Planning

For PoC, evaluation, or testing purposes (up to 100 K files) you can run all components on the same machine. Minimal configuration:

- **Processor**—3 cores
- **RAM**—12 GB
- **SQL Server Edition**—2008 R2 Express Edition and above

If you plan to deploy in bigger environments, consider the following considerations and restrictions:

- DDC Collector Requirements to Install DDC Collector
- Netwrix Auditor Requirements
- Netwrix Auditor Data Discovery and Classification Deployment Matrix

3. DDC Collector

DDC Collector is a web-based configuration module designed to discover potentially sensitive documents and directories and classify them according to specific taxonomy clues.

3.1. Compatibility Notice

Make sure to check your Netwrix Auditor version.

Data source	Netwrix Auditor version
• File Servers • EMC • NetApp	9.5 (build 2591) and later
• SharePoint	9.8 and later

3.2. Requirements to Install DDC Collector

This section contains the hardware and software requirements to flawlessly install DDC Collector.

Review the following for additional information:

- Hardware Requirements
- Software Requirements

3.2.1. Supported Data Sources

The table below lists systems that can be monitored with Netwrix Auditor Data Discovery and Classification:

Data source	Supported Versions

Windows File Servers	• Windows Server OS:
	  • Windows Server 2019
	  • Windows Server 2016
	  • Windows Server 2012/2012 R2
	  • Windows Server 2008/2008 R2

Data source	Supported Versions

	• Windows Desktop OS (32 and 64-bit):
	  • Windows 10
	  • Windows 8.1
	  • Windows 7

EMC	• EMC VNX/VNXe/Celerra families (CIFS configuration only) • EMC Isilon 7.2.0.0 – 7.2.0.4, 7.2.1.0 – 7.2.1.2, 8.0.0.0 , 8.1.0.0 (CIFS configuration only)
NetApp	• NetApp ONTAP 9.0 – 9.5 (CIFS configuration only) • NetApp Clustered Data ONTAP 8.2.1 – 8.2.3, 8.3, 8.3.1, 8.3.2 (CIFS configuration only) • NetApp Data ONTAP 8 in 7-mode (CIFS configuration only) • NetApp Data ONTAP 7 (CIFS configuration only)
SharePoint	• Microsoft SharePoint Server 2019 • Microsoft SharePoint Server 2016 • Microsoft SharePoint Foundation 2013 and SharePoint Server 2013 • Microsoft SharePoint Foundation 2010 and SharePoint Server 2010

3.2.2. Hardware Requirements

Netwrix strongly recommends installing DDC Collector apart from Netwrix Auditor. Review the hardware requirements for the computer where DDC Collector is going to be installed.

You can deploy DDC Collector 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 DDC Collector supports only Windows OS versions listed in the Software Requirements section.

IMPORTANT! Keep in mind that Netwrix Auditor Data Discovery and Classification is an integration between Netwrix Auditor and Netwrix Data Classification. This article contains recommendations

for Netwrix Auditor and the machine reserved for SQL Server. For Netwrix Data Classification requirements, see the corresponding section of Netwrix online helpcenter: Deployment Planning.

Starter / Proof of Concept

up to 500 K files (minimum requirements)

	DDC Collector	SQL Server
Processor	Any modern	Any multi-core
RAM	8 GB	16 GB

Mid-size environments

up to 8 m objects for File Servers and up to 2 m objects for SharePoint

	DDC Collector	SQL Server
Processor	Calculate according to the following example: Mid-Size Data Environment	2 cores
RAM		16 GB

Large environments

up to 32 m objects and up to 8 m objects for SharePoint

	DDC Collector	SQL Server
Processor	Calculate according to the following example: Large-Size Data Environment	8 cores
RAM		128 GB

NOTE: When planning your deployment, consider the following:

- Document throughput and processing results are overly dependent on types and amount of documents you are going to process with Netwrix Auditor Data Discovery and Classification.
- Hardware requirements for SQL Servers listed in the table above apply to each SQL Server instance in your configuration separately.
- The requirements marked with an asterisk (*) apply to a single server. To estimate total hardware requirements, multiply the values above to a number of your nodes. See Distributed Query Server Mode for more information.
- For XLarge environments, each SQL Server instance must be deployed to a separate physical disk. System architect's assistance is required for deployment planning requires.

3.2.2.1. Distributed Query Server Mode

The Distributed Query Server (DQS) mode allows re-balancing load while collection, indexing and classification. The load is distributed evenly over all enabled DDC Collector instances that allows processing large data volumes:

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

General procedure is as follows:

- Install the first DDC Collector instance where you are going to enable DQS mode.
- Install other DDC Collector instances.
- Enable DQS mode on the first instance.
- Add servers to the Distributed Query Servers list.
- Re-collect data sources.

Refer to To enable DQS Mode for detailed instructions on how to re-balance load to your system.

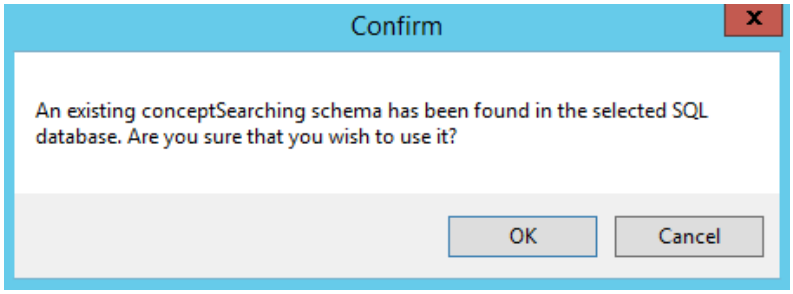
To enable DQS Mode

Recommendations below refer to clear install of DDC Collector in extra-large environment. If you upgraded from the previous version, perform steps 6 - 18.

1. Prepare machines for SQL Server instances where DDC Collector database and Netwrix Auditor (including Categories) databases reside. Consider the following recommendations:
   - Each SQL Server instance must be deployed to a separate physical disk.
   - For better performance, deploy both SQL Server instances to dedicated SSD storages. The configuration where one of the SQL Server instances runs on HDD is acceptable, but not recommended. Using HDD for all instances lead to performance loss and long report generation.
2. Create and configure **DDC Collector database** as described in the DDC Collector Database section.
3. Prepare server to install DDC Collector. Meet the Hardware Requirements hardware requirements and general Software Requirements.
4. Install and configure DDC Collector as described in the Install DDC Collector section.
5. Add license.
6. Prepare servers to install other DDC Collector instances assuming one server per one instance. Each server must meet the XLarge environment (up to 32 m objects) hardware requirements and general Software Requirements.
7. Copy the **Netwrix_Auditor_DDC_Collector.exe** file to the server considered to be the next DDC Collector instance.
8. Run installation package.

9. Proceed with installation as described in the Install DDC Collector section until **SQL Database** configuration.
10. On the **SQL Database** step, provide the name of the SQL Server instance that hosts **DDC Collector database** you configured on the step 2.

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



11. Complete the installation.
12. Repeat the steps 7 - 11 for each successive DDC Collector instance.
13. Add license.
14. On the computer where the first DDC Collector instance installed, open DDC Collector console.
15. Navigate to **Config** → **Utilities** → **DQS**.
16. Select **Enable DQS**.

    NOTE: Once DQS mode enabled you cannot roll back your configuration. Netwrix strongly recommends to ensure that you took a full backup of your environment.

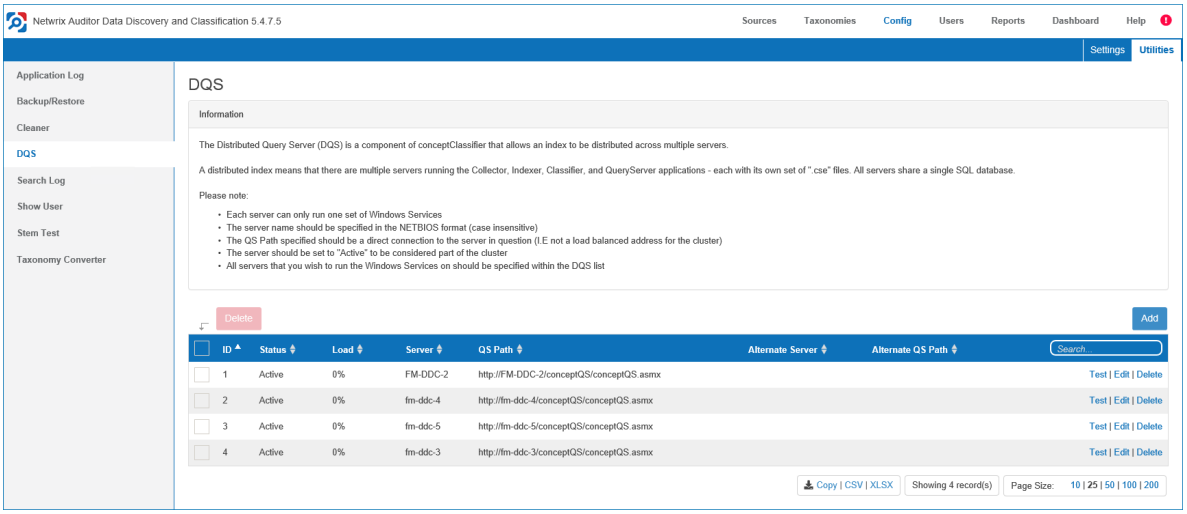
17. On the **DQS** tab, click **Add** to add servers you prepared on the step 6 one by one.

Complete the following fields:

Option	Description
Server	Provide server name. **NOTE:** FQDN format is not supported, use the NetBIOS name.

Option	Description
QS Path	Substituted automatically after adding server name.
Active	Select to enable the instance.
Alternate Server	Netwrix recommends using default values.
Alternate QS Path	Netwrix recommends using default values.

18. Review the list of DQS servers. For example:



19. When prompted, re-collect data sources to re-distribute the content across all of the configured servers.

20. You can review system health and services dashboards to check your configuration. See Review Dashboard for more information.

3.2.3. Software Requirements

The table below lists the software requirements for the DDC Collector installation:

Component	Requirements
Operating system	Windows 2012 R2 and above Server Operating System Software.
Windows Features	**Web Server Role (IIS)**
	Common HTTP Features • Default Document • HTTP Errors

Component	Requirements	
		• Static Content • HTTP Redirection
	Security	• Windows Authentication • Anonymous Authentication **NOTE:** The **Anonymous Authentication** element is included in the default installation of IIS 7. Make sure you use IIS 7 and above.
	Application Development	• ISAPI Extensions • ISAPI Filters
	Other features	
	.NET Framework 4.6 Features	• .NET Framework 4.6 • ASP.NET 4.6
	WCF Services	• HTTP Activation
SQL Server	• SQL Server 2008 R2 Standard Edition (or later). **NOTE:** Required for DDC Collector database. See DDC Collector Database for more information. For better performance, Netwrix recommends using SQL Server 2016 SP2. For earlier versions of SQL Server, consider Microsoft Recommendations to reduce allocation contention in SQL Server tempdb database if you are going to review the "Most Exposed Sensitive Data Objects" report.	
Microsoft IFilters	• Microsoft Office 2010 Filter Packs and above, 64-x edition.	
Visual Studio	• Visual C++ Redistributable Packages for Visual Studio 2015 and above.	

3.2.3.1. DDC Collector Database

DDC Collector uses Microsoft SQL Server database as data storage. You need to create a dedicated **DDC Collector database** on your SQL Server instance and configure it as shown below for the product to

function properly. You can create the database manually—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 DDC Collector and SQL Server machine.

To configure the DDC Collector database

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

1. On the computer where SQL Server instance with the **DDC Collector database** resides, navigate to **Start → All Programs → Microsoft SQL Server → SQL Server Management Studio**.
2. Connect to the server.
3. Locate the **DDC_Collector_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:

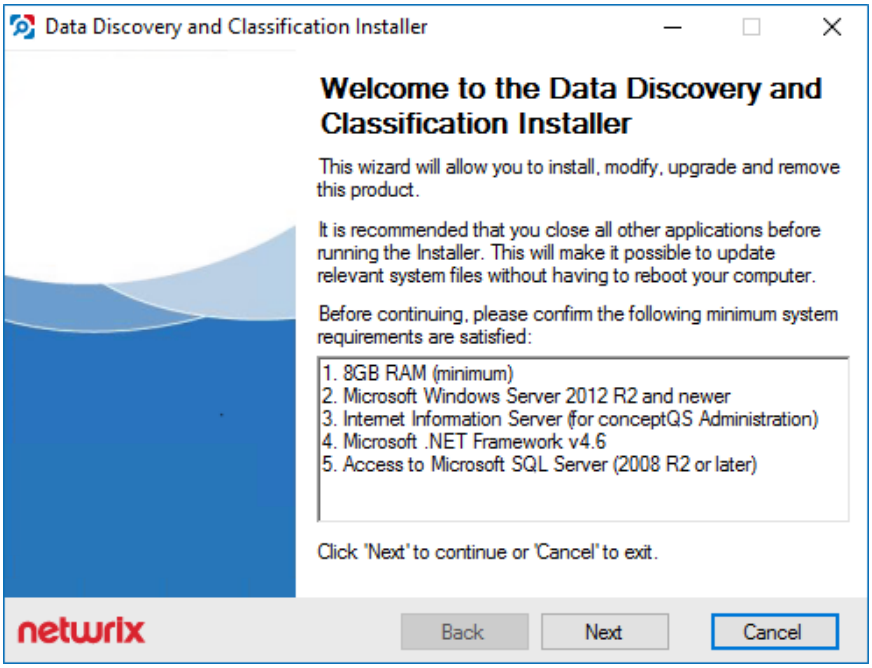
Option	Description
File Growth	• Recommended—**128 MB**. • Large environment— **512 MB**.
Maximum File Size	Select **Unlimited**.

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

3.3. Install DDC Collector

NOTE: DDC Collector uses Microsoft SQL Server database as data storage. You need to create and configure the dedicated **DDC_Collector_database** on your SQL Server instance. Check that the database has been created prior to installation. Refer to DDC Collector Database for detailed instructions on how to configure the database.

1. Run **Netwrix_Auditor_DDC_Collector.exe**.



2. Review minimum system requirements and then read the License Agreement. Click **Next**.
3. On the **Product Settings** step, specify path to install DDC Collector. For example, *C:\Program Files\Netwrix DDC*.
4. On the **License** step, add license. You can add license as follows:
   - Click the **Import** button and browse for you license file

     OR

   - Open your license file with any text editor, e.g., **Notepad** and paste the license text to the **License** field.
5. On the **Configuration** step, specify the directory where **Index files** reside. For example, *C:\Program Files\Netwrix DDC\DDC 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 DDC Collector database. For example, *"WORKSTATIONSQL\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 you created for DDC Collector. Netwrix recommends using **DDC_Collector_database** name.

7. On the **QueryServer Web Application** step, review default IIS configuration.
8. On the **Services** step, configure DDC Collector services:
   - Select all services to be installed.
   - **File System Path**—Provide a path to store DDC Collector's Services files. For example, *C:\Program Files\Netwrix DDC Services.*
   - Provide user name and password for the product services service account.
   - Select additional service options, if necessary.
9. On the **Pre-Installation Tasks and Checks** step, review your configuration and select **Install**.
10. When the installation completes, open a web browser and navigate to the following URL: *http://hostname/conceptQS* where **hostname** is the name or IP address of the computer where DDC Collector is installed.

3.4. Upgrade to the Latest Version

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

This section lists steps required to upgrade DDC Collector to the latest version. Review the following for additional information:

- To take preparatory steps
- To perform upgrade

To take preparatory steps

1. Check that the account under which you plan to run the setup has **local Administrator** rights.
2. Back up **DDC_Collector_Database** database. For that:
    a. Start Microsoft SQL Server Management Studio and connect to SQL Server instance hosting the database.
    b. In **Object Explorer**, right-click **DDC_Collector_Database** database and select **Tasks → Back Up**.
    c. Wait for the process to complete.
3. Stop the following product services:
    - **conceptCollector**
    - **conceptIndexer**
    - **conceptClassifier**
4. Finally, close Netwrix Auditor console.

To perform upgrade

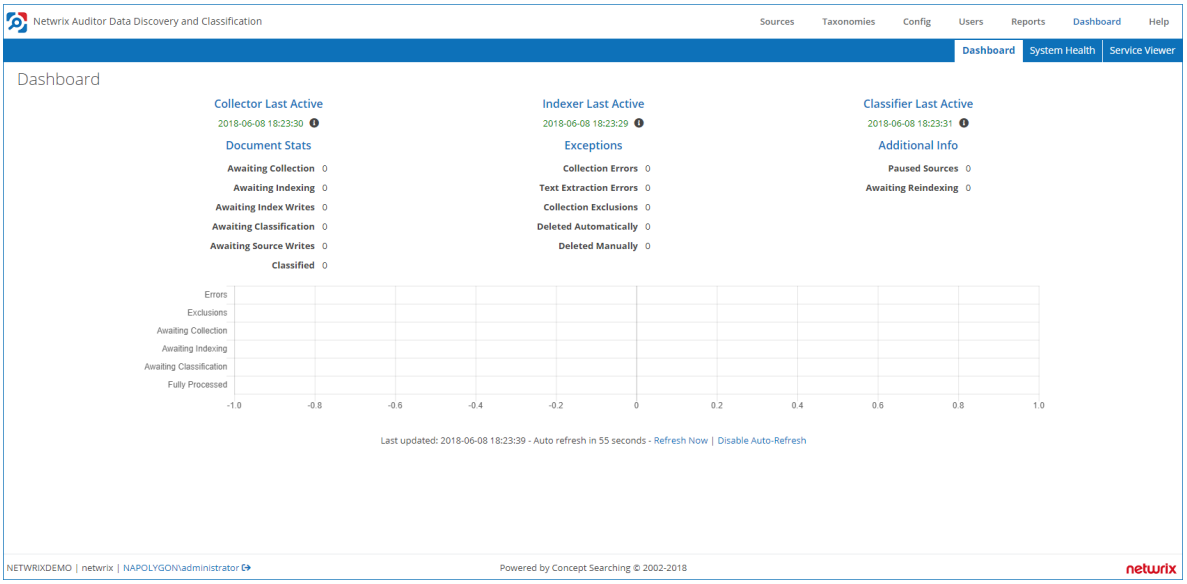
1. Run the DDC Collector installation package.
2. On the **Installation Type** step, select the Upgrade option.
3. Install DDC Collector as described in the Install DDC Collector section.

3.5. Configure DDC Collector

This section contains basic procedures to configure DDC Collector to process your sensitive data. To start configuration procedures, launch DDC Collector console. DDC Collector console is the web-based multitasking console to work with DDC Collector.

To start DDC Collector console

1. In your web browser, navigate to the following URL: *http://hostname/conceptQS* where **hostname** is the name or IP address of the computer where DDC Collector is installed.
2. In the **Instance Details** dialog, enter the unique name for your DDC Collector instance. For example, **Netwrix DDC**. The following window appears:



Review the following for additional information:

- Secure Your Data
- Add Taxonomy
- Add Content Sources
- Review Dashboard
- Enable Optical Character Recognition

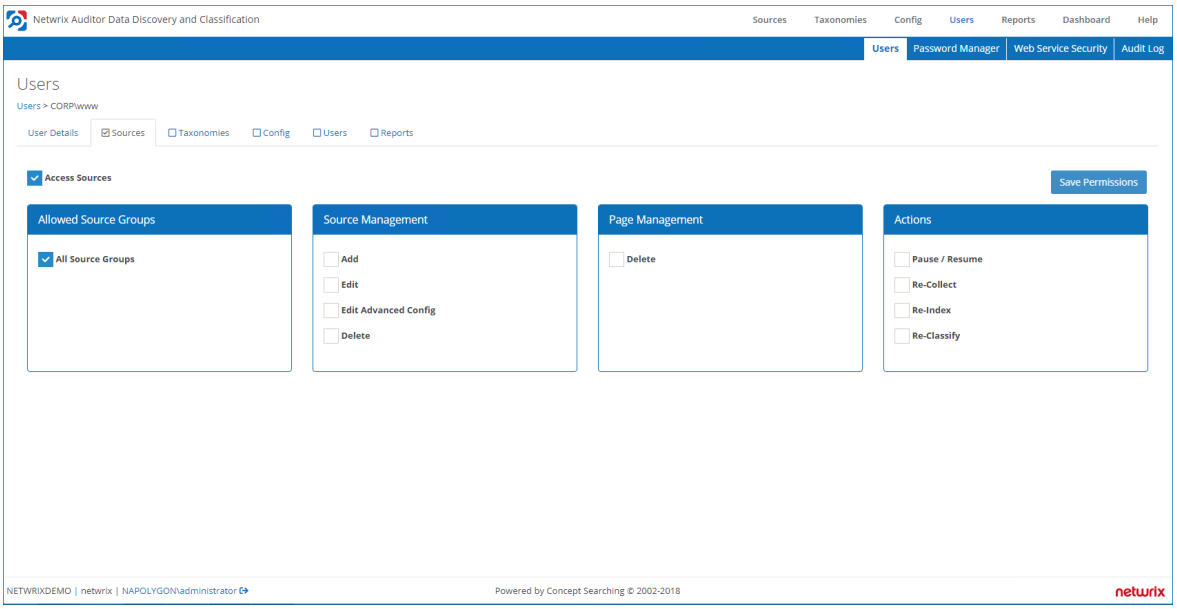
3.5.1. Secure Your Data

Security and awareness of *who* has access to *what* is crucial for every organization. Besides notifying you on who changed *what*, *when* and *where*, and *who* has access to what in your IT infrastructure, Netwrix pays attention to safety of its own configuration and collected data.

Out of the box, you are assigned the *"Super User"* role in DDC Collector console. If you want to provide access to several tabs of DDC Collector console to other users, do the following:

1. In DDC Collector console, navigate to **Sources** → *"file share or folder"* and unselect the **Anonymous Access** checkbox.
2. Navigate to **Users** → **Add** → **Validate** to add a new user.

To...	Do...
Add Super User	Select the **Super User** checkbox. The first user you add will be assigned the *"Super User"* role and he or she will have unrestricted rights in DDC Collector console. Consider to add verified user first.
Restrict access to DDC Collector console components	In the **Users** tab, select DDC Collector components available for this user: • **Sources** • **Taxonomies** • **Config** • **Users** • **Reports** Configure granular permissions for the user, if needed.



3.5.2. Add Taxonomy

Taxonomy is set of parameters to subsume concept of information for purpose of capture, management and presentation. For your convenience, DDC Collector comes with the following predefined taxonomies:

1. Personally identifiable information covering GDPR scope.
2. Medical records covering HIPAA scope.
3. Financial records and payment cards information covering GLBA and PCI DSS scope.

Each taxonomy contains a set of terms. Terms are defined by set of configuration rules (Clues). The most common clue types are the following:

- **Standard**—A single word or multi-word concept. Matched on a fuzzy basis with word stemming enabled. Use quotes around single words to disable stemming. Use double quotes around phrases to invoke exact phrase matching.
- **Case Sensitive**—A case sensitive phrase match clue.
- **RegEx**—A clue based on a Regular Expression.

For the full list of supported taxonomies, refer to Built-in Taxonomies.

Review the following for additional information:

- To upload default taxonomy
- To add custom taxonomy
- To manage taxonomies

To upload default taxonomy

1. In DDC Collector console, navigate to **Taxonomies** → **Global Settings**.
2. Navigate to **Loaded Taxonomies**, select **Add Taxonomies**.
3. Select taxonomies that you want to add in the list.

   NOTE: Multiple taxonomies selection supported. Clicking the search field enables drop-down list of default taxonomies.

4. Click **Load**.

To add custom taxonomy

1. In DDC Collector console, navigate to **Taxonomies** → **Global Settings**.
2. Navigate to **Loaded Taxonomies**, select **Add Taxonomies**.
3. Select the **Load XML file to SQL** option to import an XML file directly into the DDC Collector console; large taxonomies will be imported by the background services.

4. Browse for your custom taxonomy file.
5. Select **Upload**.

To manage taxonomies

1. In DDC Collector console, navigate to **Taxonomies** and locate the taxonomy that you want to manage.

   NOTE: If your taxonomy does not have any terms yet, right-click the taxonomy and select **Add Child Term**. Specify one or several child terms—one term per line.

2. Expand the taxonomy and locate the desired term on the left pane. Review the following for additional information:

To...	Do...
Review predefined clues	Navigate to the **Clues** tab and review available default clues. Clues are used to describe the language found in documents that make them about a particular topic.
Add automatically suggested clues	1. Navigate to the **Suggest** tab and click **Suggest** to add new clues. 2. A score is suggested automatically for the clue, you can change its type.
Search collected and classified files	1. Navigate to the **Search** tab and enter search criteria in the **Find** field. 2. Click **Search** to view search results.
Review all files matching the taxonomy	1. Navigate to the **Browse** tab and review the list of files matching the selected taxonomy. 2. Select a file and click **Calculations** link to see how the classification scores are calculated.

3.5.3. Add Content Sources

Content sources are logical containers for external systems (SharePoint and File Servers) to be crawled and classified. All your content sources are listed in the **Sources** section.

To add a content source

1. In DDC Collector console, navigate to **Sources** → **Add** and select one of the following:
   - **Folder**—Select if you want to monitor Windows File Servers, NetApp Filer or EMC Storage.
   - **SharePoint**—Select if you want to monitor SharePoint farm.
2. Complete the following fields:

Option	Description
	# Folder
	NOTE: Click **Settings** in the at the bottom of the dialog to expand advanced settings.
Folders	Enter the UNC path of the root folder where collection is to start. You can add either windows directories, or NetApp filer or EMC storage devices, to the index. **NOTE:** Multiple folders selection supported. Specify equal UNC paths for both: in Netwrix Auditor and DDC Collector. Any actions made over data sources configured in different way or locally (e.g., *"C:\"*) are out of scope. Otherwise, you need to map to the same server location and then restart the DDC Provider service.
Username	Specify the account used to process the folder. **NOTE:** It is recommended to create a dedicated account for this purpose and specify it here. This will, in particular, simplify filtering file crawling activity (*read* operations) by the service account name when reporting on file and folder access. To read more about configuring exclusions for Netwrix Auditor reports, refer to Configure Data Sources in Netwrix Auditor section.
Password	Provide a password for the account specified above.
Depth Limit	Select if you want to process data in sub-folders and set depth limit.
Text Patterns	Netwrix recommends using default values.
Allow anonymous access	This option is used to disable security filtering for selected sources. If unselected, the indexing processes will collect Windows Access Control

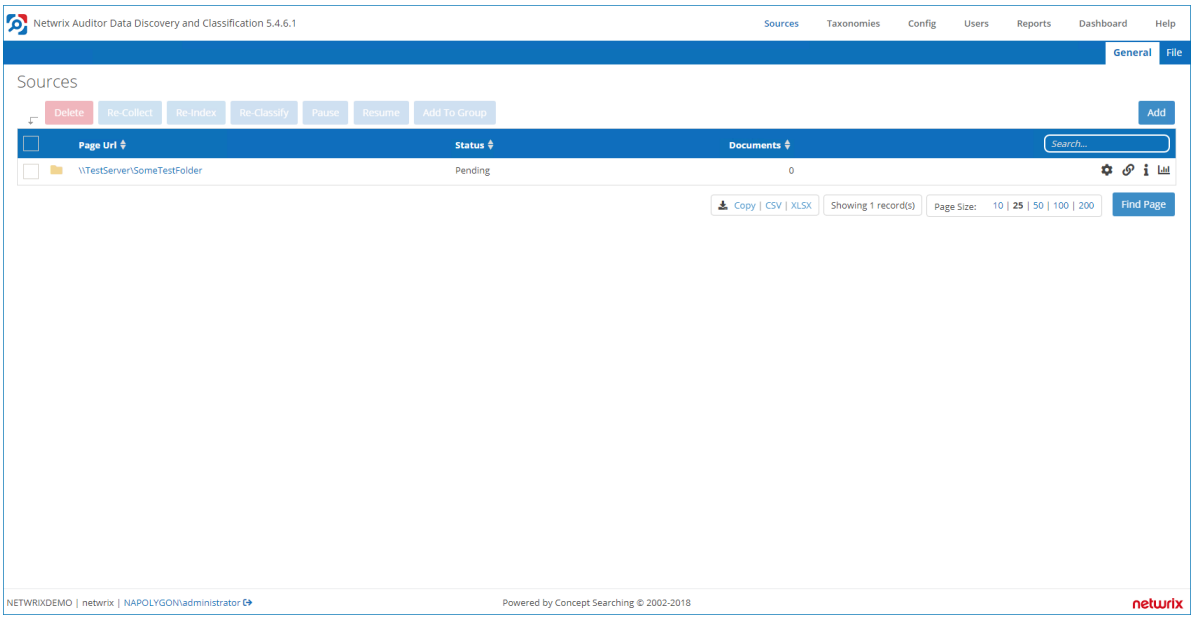
Option	Description
	Lists (ACLs) for the files and search results will be filtered based upon the end user's Windows identity. Netwrix recommends unselect this option. See Secure Your Data for more information.
Duplicate Detection Enabled	Select to exclude documents that contain the same text content from the index.
Write classifications	Netwrix recommends using default values.
Re-Index Period	Specifies how often the source should be checked for changes. Netwrix recommends using default values.
Priority	Netwrix recommends using default values.
Max Collector Retries	Netwrix recommends using default values.
Document Type	Specify a value that can be used to restrict queries when utilising the DDC Collector search index.
Source Group	Netwrix recommends using default values.

SharePoint

URL	Enter the SharePoint Site Collection URL. **NOTE:** Multiple farms selection supported. List your SharePoint Site Collection URLs line by line, for example, in Notepad. Copy the list and paste it to the **URL** field. Specify equal URLs for both: in Netwrix Auditor and DDC Collector.
Username	Specify the account used to connect to a SharePoint farm.
Password	Provide a password for the account specified above.
Write classifications	Netwrix recommends using default values.
Source Group	Netwrix recommends using default values.

Option	Description
Pause source on creation	Netwrix recommends using default values.

3. Select **Index Folder** to start indexing process. You will see an information popup window on successful indexing.



3.5.4. Review Dashboard

Upon data classification completion, check your files processing progress. In DDC Collector console, navigate to the **Dashboards** section.

The default screen (Dashboard) shows a high level overview of Netwrix Auditor Data Discovery and Classification service statistics. You can review all processing stages of every component:

- Collector
- Indexer
- Classifier

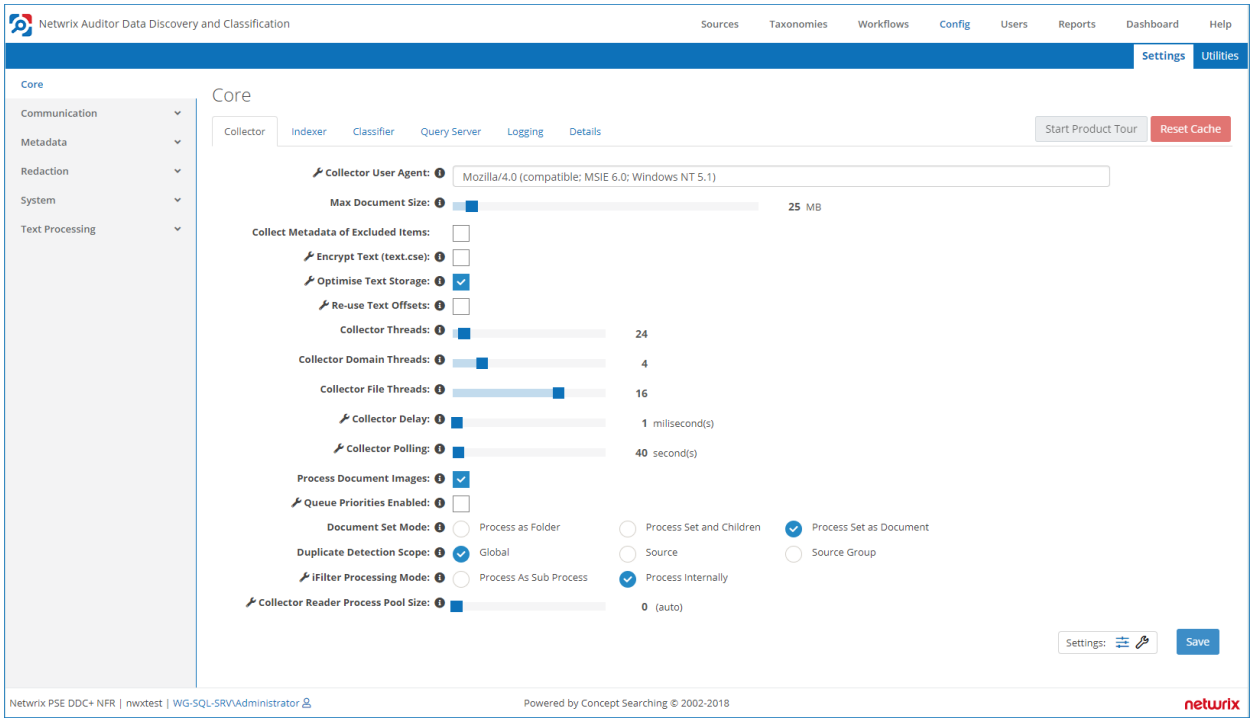
For the full list of the potential issues, refer to System Health and Troubleshooting

3.5.5. Enable Optical Character Recognition

Optical Character Recognition, or OCR, is a technology that enables you to convert different types of files, such as stand-alone images, PDF files and Microsoft Office documents with integrated images into discoverable data. By default, this option is disabled to avoid loss of performance.

Out of the box, DDC Collector processes JPEG, png, TIFF, and Bitmap images. For the full list of supported content types, refer to Supported Content Types section. If you want to enable OCR, configure the product as follows:

To...	Do...
Recognize stand-alone images	Do the following to enable OCR for image files having specific extension: 1. In DDC Collector console, navigate to **Sources** → **File**. 2. Select **Files Included** on the left. 3. Click **Add Inclusion** on the right pane to add desired extension.
Recognize documents with integrated images	1. In DDC Collector console, navigate to **Config** → **Settings** → **Core** → **Collector**. 2. Select the **Process Document Images** option.



The settings will be applied in an hour after configuration. If you want to start process images and documents earlier, navigate to the **Services** snap-in and restart the following services:

- conceptIndexer
- ConceptCollector
- conceptClassifier

NOTE: Make sure that DDC Collector does not process any files, otherwise service restart may fail data classification process.

4. Configure Data Sources in Netwrix Auditor

To see your sensitive data in Data Discovery and Classification reports, you need to create a monitoring plan in Netwrix Auditor and configure data sources. The following data sources are available:

- Windows File Servers
- EMC
- NetApp
- SharePoint

Check your monitoring plan and items:

Option	Description
	File Servers
Item	Specify file shares that you want to process with DDC Collector. **NOTE:** Specify equal UNC paths for both: item in Netwrix Auditor and DDC Collector. Any actions made over data sources configured in different way or locally (e.g., *"C:\"*) are out of scope. Otherwise, you need to map to the same server location and then restart the DDC Provider service.
Additional options	1. Enable the **Collect data for state-in-time reports** option for each item that you want to process. 2. Enable the **Include details on effective permissions** option to review the following reports: • Most Accessible Sensitive Files and Folders • Overexposed Files and Folders • Sensitive Folder and File Permission Details

Option	Description
SharePoint	
Item	Specify SharePoint farm— Enter the SharePoint Central Administration website URL. **NOTE:** Specify equal URL for both: item in Netwrix Auditor and DDC Collector.
Additional options	Enable the **Collect data for state-in-time reports** option for each item that you want to process.

NOTE: Refer to the Create a New Plan section in **Netwrix Auditor Online Help Center** for detailed instructions on how to create a new monitoring plan.

When DDC Collector processes ("crawls") specified files and folders, it performs *read* operation under the dedicated DDC Collector account (described in the Add Content Sources section). Netwrix Auditor that monitors your file storage system (Windows File Server, NetApp Filer or EMC Storage) or SharePoint farm, will report these *read* operations by default. To avoid this excessive reporting, it is recommended to include the dedicated DDC Collector account and its *read* operations in the *omitreportlist.txt* and *omitstorelist.txt* files for Netwrix Auditor. Review the following for additional information:

- Exclude Data from File Servers Monitoring Scope
- Exclude Data from SharePoint Monitoring Scope

5. DDC Provider

DDC Provider is the integration module used to aggregate data inventory, access permissions, activity and classification information into a single reporting database for Netwrix Data Discovery and Classification reports.

5.1. Hardware and Software Requirements

DDC Provider and Netwrix Auditor must be installed on the same computer. Refer to the Requirements to Install Netwrix Auditor section in **Netwrix Auditor Online Help Center** for detailed list of hardware and software requirements.

NOTE: If you plan to use Microsoft SQL Server 2016, make sure it has SP2 installed.

5.2. Account Requirements

This section lists the requirements for the accounts used by DDC Provider. The accounts must be granted the following rights and permissions:

- A member of the **local Administrators** group on the computer where Netwrix Auditor Server and DDC Provider are installed.
- The **Database datareader** server role must be assigned to the account on the SQL Server instance where the **DDC Collector database** resides.

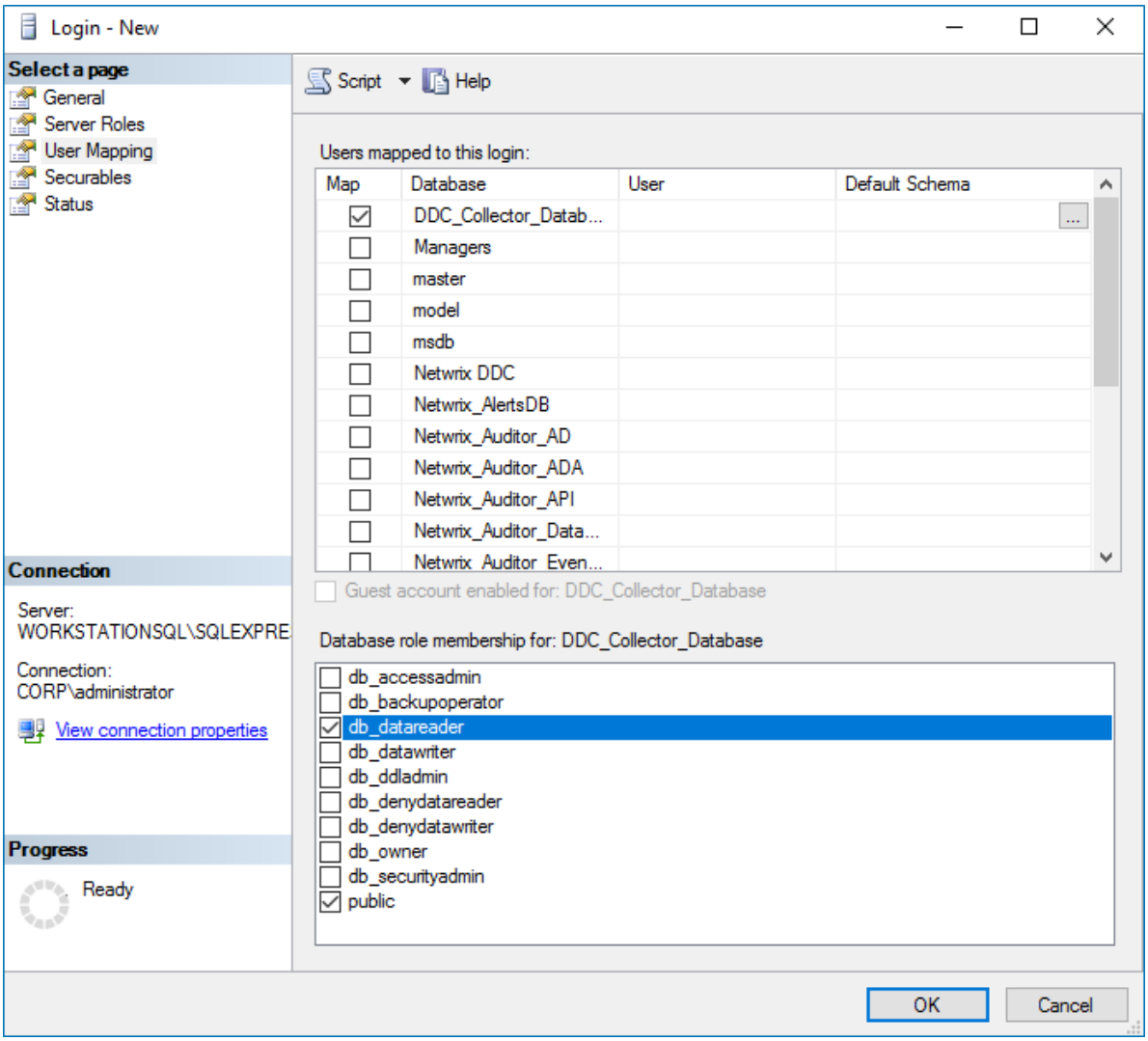
NOTE: Netwrix recommends using different accounts to connect to the SQL Server instances where **DDC Collector database** and **Categories database** reside.

Review the following for additional information:

- To assign the Database datareader server role

To assign the Database datareader server role

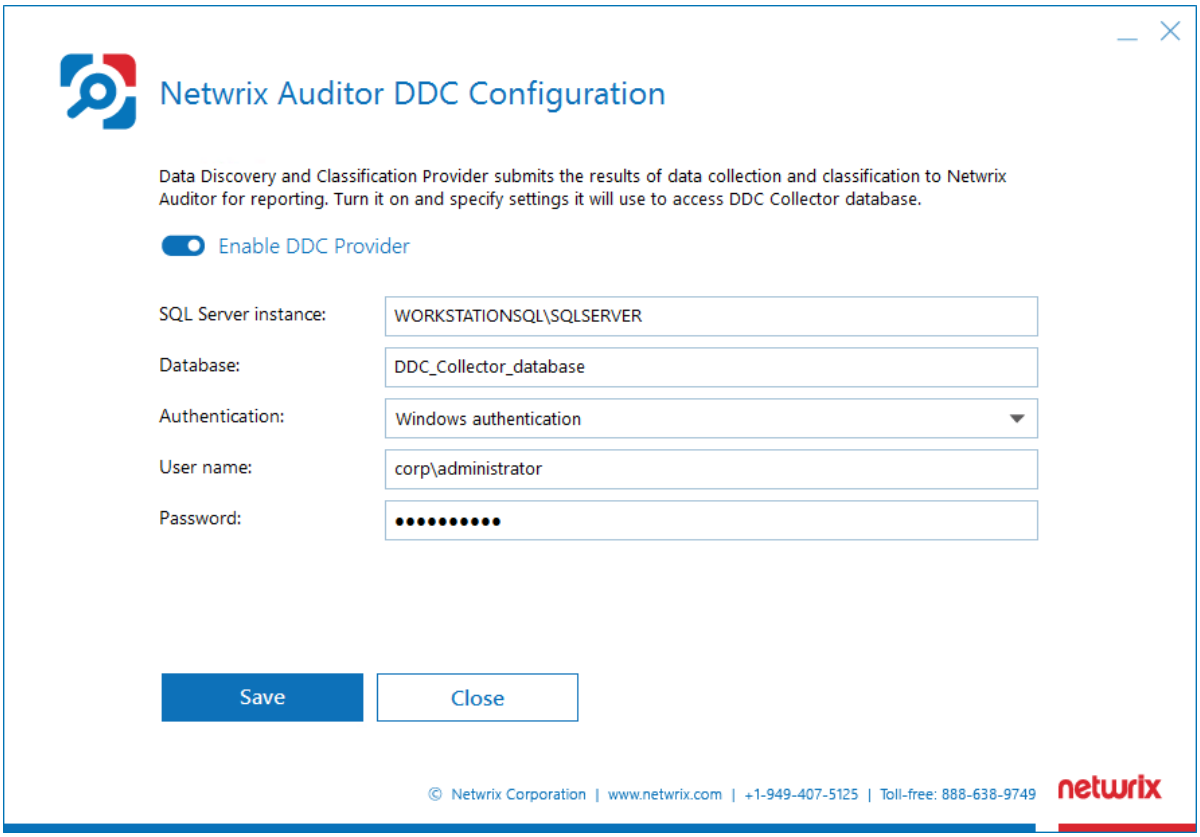
1. On the computer where SQL Server instance with **DDC_Collector_Database** resides, navigate to **Start → All Programs → Microsoft SQL Server → SQL Server Management Studio**.
2. Connect to the server.
3. In the left pane, expand the **Security** node. Right-click the **Logins** node and select **New Login** from the pop-up menu.



4. Select **User mapping** on the left and select the **DDC_Collector_database** for which you want to assign the role.
5. In the **Database role membership for: DDC_Collector_database** list, select the **db_datareader** role.

5.3. Enable and Configure DDC Provider

1. On the computer where Netwrix Auditor is installed, navigate to *Netwrix Auditor installation folder\AuditIntelligence* (by default, Netwrix Auditor is installed to *C:\Program Files (x86)\Netwrix Auditor*) and run the **DDCConfiguration** tool.
2. Complete the fields as shown below:



Option	Description
Enable DDC Provider	Starts the Netwrix Auditor DDC Provider service and changes it's **Startup Type** to *"Automatic"*.
SQL Server instance	Provide the name of the SQL Server where **DDC Collector database** resides (e.g., *WORKSTATIONSQL\SQLEXPRESS* for *SQLEXPRESS* instance). See DDC Collector Database for more information.
Database	Provide the name of the database you created for DDC Collector.
Authentication	Select Windows or SQL Server authentication method to connect to **DDC Collector database**.
User name	Specify the account to be used to connect to the SQL Server instance.
Password	Provide password for the account.

3. Click **Save** to save your configuration.

NOTE: Mind that DDC Provider is a part of Netwrix Auditor Data Discovery and Classification. For the solution to function properly, install and configure DDC Collector as described in the DDC Collector section.

If you have any issues while using DDC Provider, See System Health and Troubleshooting for more information.

5.4. Upgrade to the Latest Version

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

This section lists steps required to upgrade DDC Provider to the latest version. Review the following for additional information:

- To take preparatory steps
- To perform upgrade

To take preparatory steps

1. Check that the account under which you plan to run the setup has **local Administrator** rights.
2. Back up **Categories** database. For that:
    a. Start Microsoft SQL Server Management Studio and connect to SQL Server instance hosting the database.
    b. In **Object Explorer**, right-click **Categories** database and select **Tasks → Back Up**.
    c. Wait for the process to complete.
3. Stop the following product services:
    - **conceptCollector**
    - **conceptIndexer**
    - **conceptClassifier**
4. Finally, close Netwrix Auditor console.

To perform upgrade

You can upgrade DDC Provider by running the Netwrix Auditor installation package.

6. Review Data Discovery and Classification Reports

Looking for real-life use cases and walk through examples? Check out Netwrix training materials. Go the Data Discovery and Classification Reports page on Netwrix website.

NOTE: Re-open Netwrix Auditor after DDC Provider installation.

In Netwrix Auditor, navigate to **Reports** → and select a report you are interested in and click **View**.

Data Discovery and Classification reports grouped by data sources.

The table below lists the reports available for Data Discovery and Classification:

Report	Description
File Servers	
Activity reports	
Activity Related to Sensitive Files and Folders	This report lists all access attempts to files and folders that contain certain categories of sensitive data at the moment.
State-in-time reports	
Most Accessible Sensitive Files and Folders	This report shows the number of users that effectively have access to sensitive files or folders, sorted in descending order. Use this report to identify data at high risk and plan for corrective actions accordingly.
Overexposed Files and Folders	This report shows sensitive files and folders accessible by the specified users or groups, based on the combination of folder and share permissions. Use this report to identify data at high risk and plan for corrective actions accordingly.
Sensitive Files and Folders by Owner	This report shows ownership of files and folders that are stored in the specified file share and contain selected categories of sensitive data. Use this report to determine the owners of particular sensitive data.
Files and Folders Categories by	This report shows files and folders that contain specific categories of sensitive data. Use this report to see whether a specific file or folder contains sensitive data.

Report	Description
Object	
Sensitive Files Count by Source	This report shows the number of files that contain specific categories of sensitive data. Use this report to estimate amount of your sensitive data in each category, plan for data protection measures and control their implementation.
Sensitive File and Folder Permissions Details	This report shows permissions granted on files and folders that contain certain categories of sensitive data. Use this report to see who has access to a particular file or folder, via either group membership or direct assignment. Reveal sensitive content that has permissions different from the parent folder.

SharePoint

Activity reports	
Activity Related to Sensitive Data Objects	This report shows changes and read operations on SharePoint sites and documents that contain sensitive information. Use this report to detect suspicious activity around your sensitive data.
State-in-time reports	
Sensitive Data Objects by Site Collection	For each SharePoint site collection listed, this report shows the categories of sensitive data stored there and the number of documents in each category. Use this report to reveal the number of sensitive files stored in your SharePoint site collections.
Sensitive Data Objects	For each site collection listed, this report shows the SharePoint objects (sites, lists and documents) that have been classified as containing sensitive information. Use this report to plan and control data protection measures for sensitive information stored on your SharePoint.
Sensitive Data Object Permissions	For each SharePoint object (site, list or document) listed, this report shows the user accounts that have access to this object, their effective permissions and how those permissions were granted (for example, permissions can be granted directly, via group membership or using SharePoint policy). Use this report to control access to SharePoint objects that contain sensitive data.
Overexposed Sensitive Data Objects	For each user account listed, this report shows the SharePoint objects (sites, lists and documents) containing sensitive data that the user can access based on their effective permissions. Use this report to identify overexposed data and plan

Report	Description
	measures to mitigate your risk.
Most Exposed Sensitive Data Objects	Lists the SharePoint objects (sites, lists and documents) containing sensitive data that can be accessed by the most users (or even Everyone), based on effective permissions. Use this report to identify data at high risk and plan corrective actions.

6.1. Leverage Filtering Capabilities

Report filters allow you to display changes matching certain criteria. For example, you can filter changes by source or object type. Filtering does not delete changes, but modifies the report view allowing you to see changes you are interested in. Filters can be found in the upper part of the **Preview Report** page.

To apply filters

1. Navigate to **Reports** and generate a report.
2. Apply filters to the report and click **View Report**. For example, you can update report timeframe, select specific values for *Who* and *Where*, apply sorting, etc.

Wildcards are supported. For example, type *%admin%* in the **Who (domain\user)** field if you want to view changes made by users with the name containing "administrator" (e.g., *enterprise\administrator, corp\administrator, sqladmin*).

Do not use % in the exclusive filters (e.g., *Who (Exclude domain\user)*). Otherwise, you will receive an empty report.

6.2. Subscribe to Report

Subscriptions enable you to schedule email delivery of a variety of reports. Subscriptions are helpful if you are a rare guest of Netwrix Auditor and you only need to get statistics based on individual criteria.

To create report subscription

1. On the main Netwrix Auditor page, navigate to **Reports**. Specify the report that you want to subscribe to and click **Subscribe**.
2. On the **Add Subscription to a Report** page, complete the following fields:

Option	Description
	General

Option	Description
Subscription name	Enter the name for the subscription.
Report name	You cannot edit report name.
Send empty subscriptions when no activity occurred	Slide the switch to **Yes** if you want to receive a report even if no changes occurred.

Specify delivery options	• **File format**—Configure reports to be delivered as the doc or xls files.
	• **File delivery**—Select one of the following:
	  • **Attach report to email**—Select this option to receive reports as email attachments.
	  The maximum size of the attachment file is 50 MB.
	  • **Upload to a file share**—Select this option to save reports on the selected file share. Click **Browse** to select a folder on the computer that hosts Netwrix Auditor Server or specify a UNC path to a shared network resource.
	  NOTE: Make sure that the recipients have sufficient rights to access it and the Long-Term Archive service account has sufficient rights to upload reports. Refer to the Configure Long-Term Account section in **Netwrix Auditor Online Help Center** for the full list of required account rights and permissions.
	• **File delivery**—Select report delivery method:
	  • **Attach report to email**—Select this option to receive reports as email attachments.
	  The maximum size of the attachment file is 50 MB. If the limit exceeded, the product creates a shared folder *"netwrix_ report_ subscriptions"* to upload the attachment. The attachment files will be available for 7 days. Check the subscription email to get the files.
	  • **Upload to a file share**—Select this option to save reports on the selected file share. Click **Browse** to select a folder on the computer that hosts Netwrix Auditor Server or specify a UNC path to a shared

Option	Description
	network resource.
	Other tabs
Recipients	Shows the number of recipients selected and allows specifying emails where reports are to be sent. Expand the **Recipients** list and click **Add Recipient** to add more recipients.
Schedule	Allows specifying report delivery schedule (daily, certain days of week, a certain day of a certain month). **NOTE:** By default, the product emails reports daily at 8.00 am.
Filters	Specify the report filters, which vary depending on the selected report.

7. System Health and Troubleshooting

This section provides instructions on how to troubleshoot issues that you may encounter while using DDC Collector. Review the following for additional information:

- System Health and Services
- Troubleshooting Issues
- DDC Provider Issues

7.1. System Health and Services

Navigate to the **Dashboards** section to check Netwrix Auditor Data Discovery and Classification health state. Review the following for additional information:

Dashboard	Description
System Health	Review health statuses of every service. If an issue occurs, you can expand it and review details and suggested resolution.
Service Viewer	Shows real-time activity of all services. Once all work is complete “Idle ...” will be displayed. It is possible to use this to check which sources are currently being processed, as well as to ensure that the services are currently running.

7.2. Troubleshooting Issues

Issue	Resolution
DDC Collector installation completes with warnings.	On the computer where DDC Collector is installed, navigate to the **Services** snap-in and restart the following services manually: • conceptIndexer • ConceptCollector • conceptClassifier
The **Service Viewer** dashboard cannot load the **Indexer** service status.	
The **Classifier** service highlighted as inactive on the **Service Viewer** dashboard.	

7.3. DDC Provider Issues

Issue	Resolution
Upgrade completes with warnings and errors.	On the computer where DDC Provider is installed, navigate to DDC Provider logs. By default, they are stored to *"C:\ProgramData\Netwrix Auditor\Logs\Data Discovery and Classification\Tracing"* and review the corresponding log.
DDC Provider configuration completes with warnings.	
reports do not show data.	

8. Glossary

The table below contains basic glossary terms:

Term	Description	Map to Reports
Source	External system being processed.	Object path / UNC path
Taxonomy / Termset	Taxonomy is set of parameters to subsume concept of information for purpose of capture, management and presentation.	Category
Clues	Clues are used to describe the language found in documents that make them about a particular topic.	Not reflected in reports.
Class / Term	Synonymous, used to describe a node in a taxonomy / termset.	Not reflected in reports.

9. Related Documents

The table below lists all documents available to support Netwrix Auditor Data Discovery and Classification:

Document	Description
Netwrix Auditor Online Help Center	Gathers information about Netwrix Auditor from multiple sources and stores it in one place, so you can easily search and access any data you need for your business. Read on for details about the product configuration and administration, its security intelligence features, such as interactive search and alerts, and Integration API capabilities.
Netwrix Auditor Installation and Configuration Guide	Provides detailed instructions on how to install Netwrix Auditor, and explains how to configure your environment for auditing.
Netwrix Auditor Administration Guide	Provides step-by-step instructions on how to configure and use the product.
Netwrix Auditor Intelligence Guide	Provides detailed instructions on how to enable complete visibility with Netwrix Auditor interactive search, report, and alert functionality.
Netwrix Auditor Integration API Guide	Provides step-by-step instructions on how to leverage Netwrix Auditor audit data with on-premises and cloud auditing solutions using RESTful API.
Netwrix Auditor Release Notes	Lists the known issues that customers may experience with Netwrix Auditor 9.6, and suggests workarounds for these issues.