Netwrix Data Classification for Box

Quick-Start Guide

Version: 5.6.1
8/12/2021

Table of Contents

1. Introduction 4
2. Configure Box for Crawling 5
2.1. Prerequisites 5
2.2. Procedure Steps 5
2.2.1. Step 1. Create the App 6
2.2.2. Step 2. Authorize the App 7
3. Install Netwrix Data Classification 9
4. Initial Product Configuration 11
4.1. Select Processing Mode 12
4.2. Processing Settings 12
4.3. Add Taxonomy 13
4.4. Review Your Configuration 13
5. Add Database Source 14
5.1. Database Configuration Wizard 15
5.1.1. Introduction 15
5.1.2. Tables 16
5.1.3. Exceptions 16
5.1.4. Summary 17
6. Add Box 18
7. Review Reports and Browse Classified Documents 21

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.

1. Introduction

This guide is intended for the first-time users of Netwrix Data Classification for Box. 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:

- Prepare your IT infrastructure for scanning
- Install and configure Netwrix Data Classification
- Add a source to start crawling Box
- Review classification results
- Leverage reporting capabilities and export results for custom reports

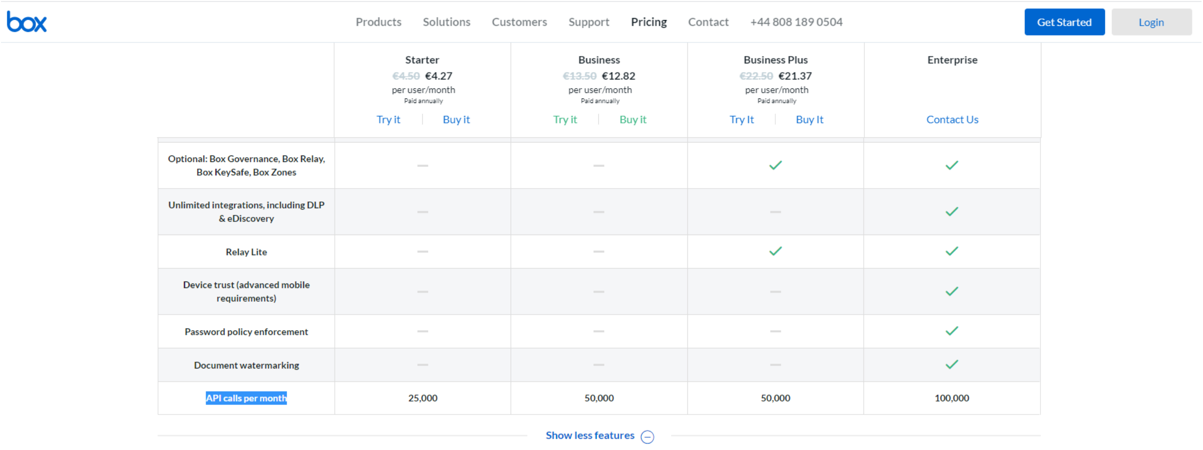
NOTE: This guide only covers the basic configuration and usage options for crawling Box with Netwrix Data Classification. For advanced installation scenarios and configuration options, as well as for information on various reporting possibilities and other product features, refer to Netwrix Data Classification Online Help Center.

2. Configure Box for Crawling

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

2.1. 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.



	Starter ~~€4.50~~ €4.27 per user/month Paid annually Try it | Buy it	Business ~~€13.50~~ €12.82 per user/month Paid annually Try it | Buy it	Business Plus ~~€22.50~~ €21.37 per user/month Paid annually Try It | Buy It	Enterprise Contact Us
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	—	—	—	✓
API calls per month	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.

2.2. Procedure Steps

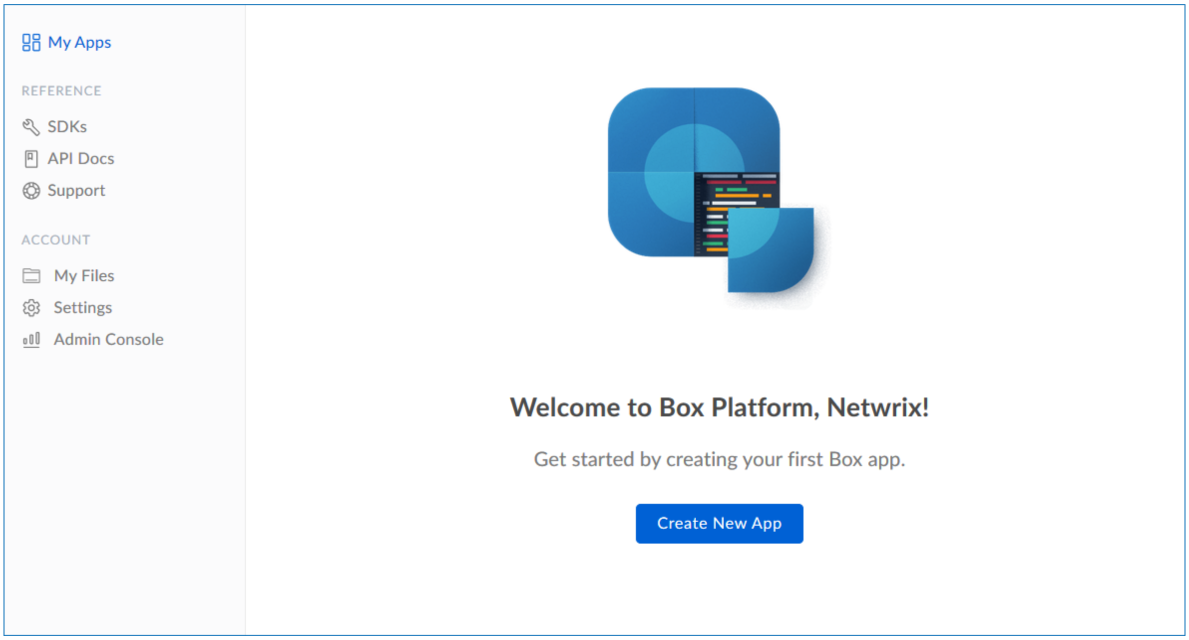
There are four key stages in this procedure:

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 Boxsection.

2.2.1. 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:



Welcome to Box Platform, Netwrix!

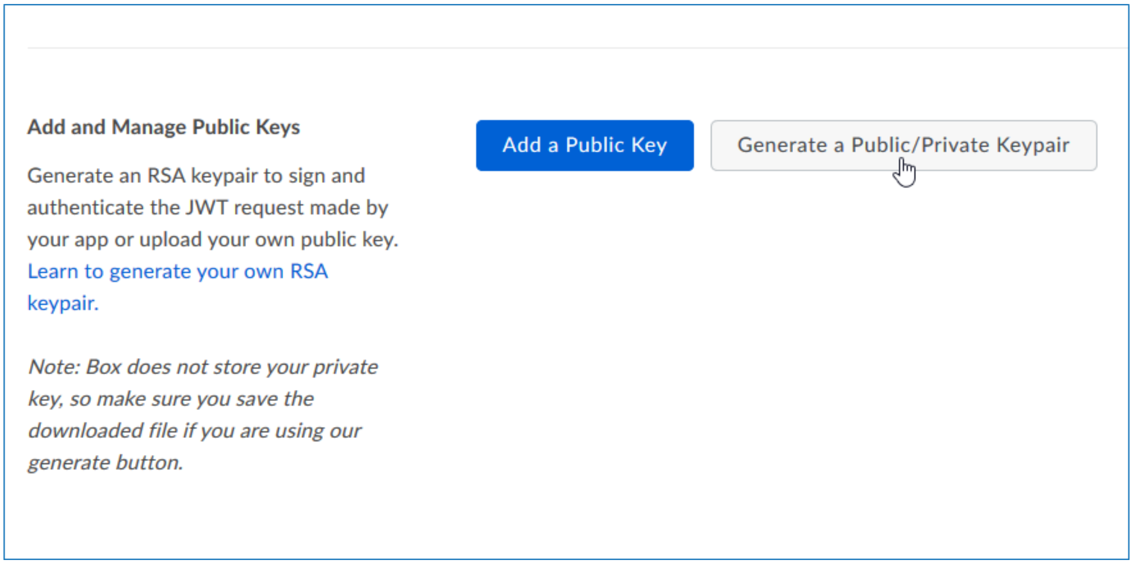
Get started by creating your first Box app.

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.



Add and Manage Public Keys

Generate an RSA keypair to sign and authenticate the JWT request made by your app or upload your own public key. Learn to generate your own RSA keypair.

Note: Box does not store your private key, so make sure you save the downloaded file if you are using our generate button.

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.

2.2.2. 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 decribed in Box documentation. For instance, you can take these steps:

1. Navigate to 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

3. 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, "*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. Netwrix recommends using **NDC_database** 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.

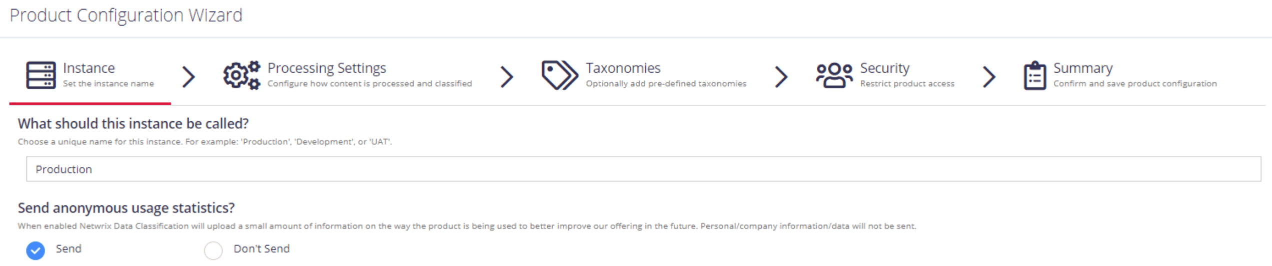
10. On the **Pre-Installation Tasks and Checks** step, review your configuration and select **Install**.
11. When the installation completes, open a web browser and navigate to the following URL: *http://localhost/conceptQS* where **localhost** is the name or IP address of the computer where Netwrix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

4. 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'.

Production

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.

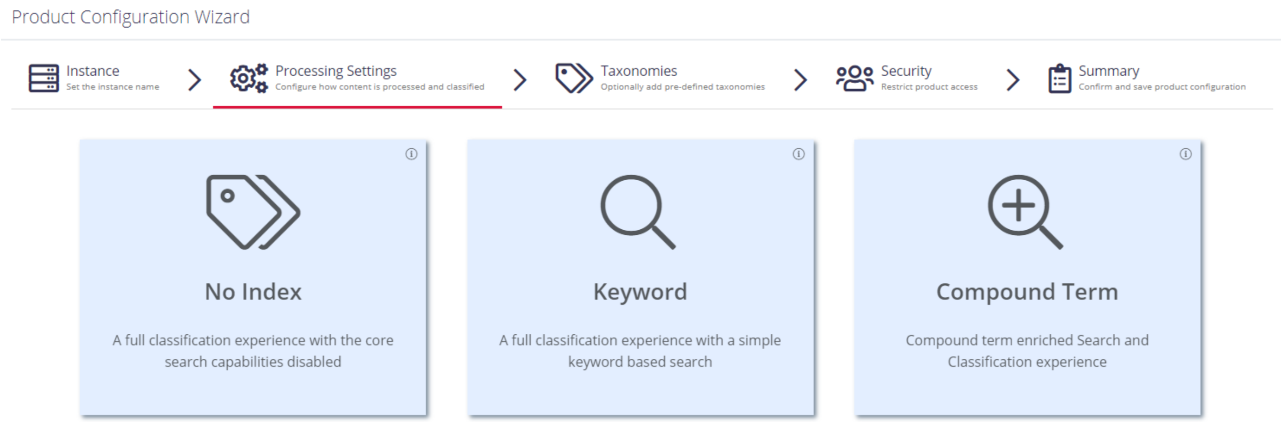
● Send ○ Don't Send

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

- Select Processing Mode
- Processing Settings
- Add Taxonomy
- Security
- Configure Health Alerting
- Review Your Configuration

4.1. Select Processing Mode

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



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

No Index

A full classification experience with the core search capabilities disabled

Keyword

A full classification experience with a simple keyword based search

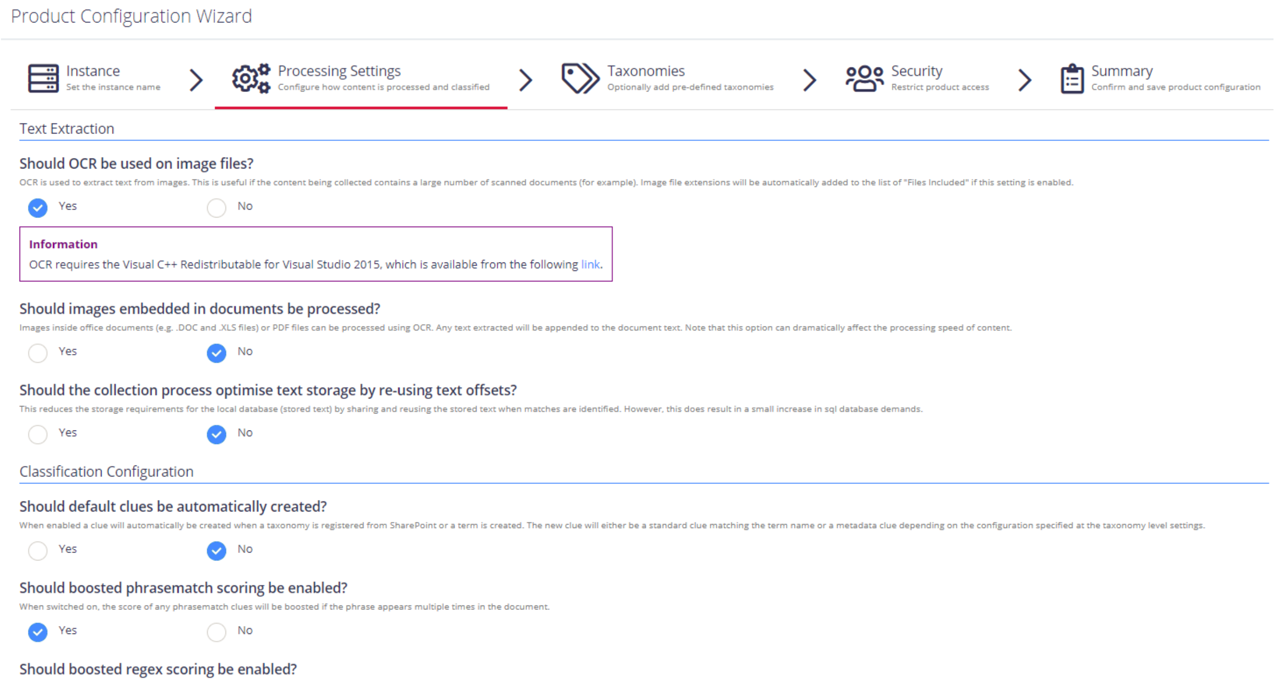
Compound Term

Compound term enriched Search and Classification experience

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

4.2. Processing Settings

On the **Processing Settings** step, 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.

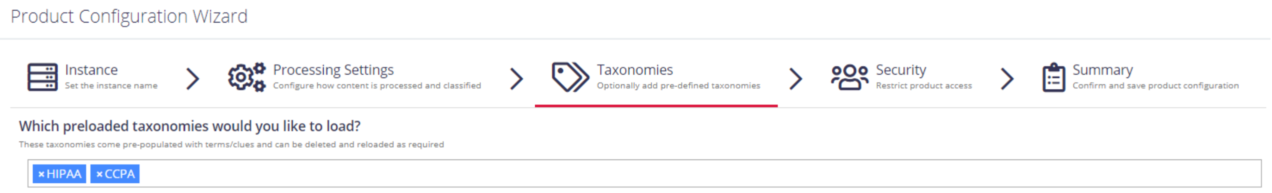
● Yes ○ No

Should boosted regex scoring be enabled?

Proceed with adding taxonomies.

4.3. 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.

4.4. Review Your Configuration

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

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

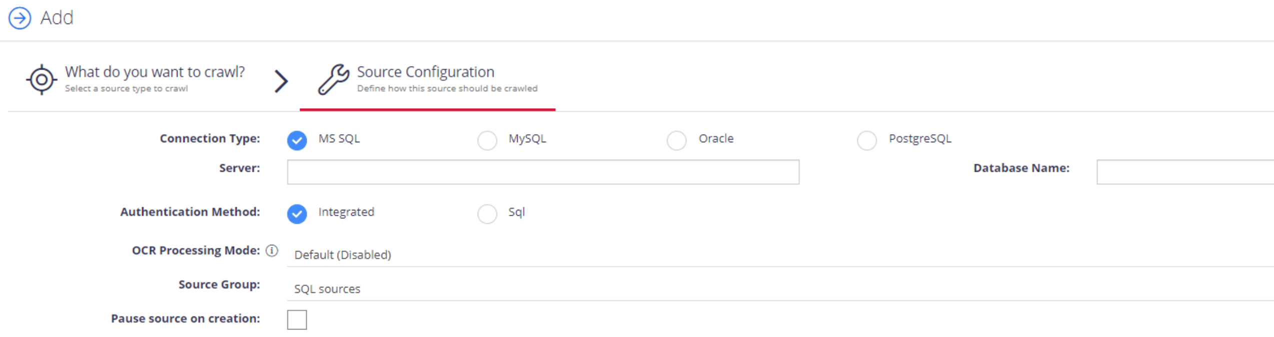
5. Add Database Source

The **Database** source configuration screen allows you to enable the crawling and classification of content stored in your Microsoft SQL Server, MySQL, and Oracle databases.

Content must either be configured / crawled using the configured service accounts (IIS Application Pool User, Windows Services) or by using specific connection details.

Once connected it is possible to create an intelligent content mapping, crawling certain fields as unstructured index text, and other fields as mapped metadata. For more information please see the Database Configuration Wizard section.

If you wish to make other configuration changes before collection of the source occurs ensure you tick the checkbox "*Pause source on creation*".


Add
What do you want to crawl?
Select a source type to crawl
Source Configuration
Define how this source should be crawled

Connection Type: ● MS SQL ○ MySQL ○ Oracle ○ PostgreSQL
Server: Database Name:
Authentication Method: ● Integrated ○ Sql
OCR Processing Mode: Default (Disabled)
Source Group: SQL sources
Pause source on creation: □

Complete the following fields:

Option	Description
Connection Type	Select your connection type: MS SQL, MySQL, or Oracle.
Server	Specify the server name of the database system to be crawled (".") can be used to indicate the local server).
Database Name	Specify the database that will be crawled. It is possible to configure multiple databases from the same server.
Authentication Method	Select authentication method: **Integrated** or **SQL**. • With **Integration** option selected, database will be accessed under the account currently logged on. • With **SQL** option selected, specify user name and password to be used when accessing the database.
OCR Processing Mode	Select processing mode for images in the documents:

Option	Description
	- **Disabled** – documents' images will not be processed. - **Default** – defaults to the source settings if configuring a path or the global setting if configured on a source. - **Normal** – images are processed with normal quality settings. - **Enhanced** – upscale images further to allow more accurate results. This will provide better accuracy but can lead to longer processing time if the images do not contain text.
Source Group	If you want to add database to a source group, select existing, or create a new one.
Pause source on creation	Select to make other configuration changes before the initial data collection starts.

After the source configuration is completed, you will be prompted to lauch SQL crawling configuration wizard. See Database Configuration Wizard for more information.

5.1. Database Configuration Wizard

For the database sources, you can enable security-based crawling, that is, finding sensitive data (which logically will either be stored in text or binary-based columns). It is possible to create an intelligent content mapping, crawling certain fields as unstructured index text, and other fields — as mapped metadata.

This section explains how to use the **Database Configuration Wizard** for configuring the crawling process. You can run this wizard when adding the data source, or you can later open the **Source** tab, select your database source and click **Launch Wizard**.

IMPORTANT! If you want to crawl a target database in your MS SQL replication model, you must backup your database before running the configuration wizard.

See next:

- Introduction
- Tables
- Exceptions
- Summary

5.1.1. Introduction

On this step, provide matching rules to search in the database for data that match exactly or are similar to a specific pattern. You can indicate both: exact or partial matches over the database strings.

5.1.2. Tables

On this step, review the grid of the tables in the database that are not currently enabled for crawling (if already enabled then don't show in this grid) and have at least one text/binary column. Configure your crawling scope considering the following:

Column	Description
Table	Contains the list of all tables in the database, followed by alphabetically.
Text Columns	Contains the number of text/binary columns for each table. Click the number link to review the full list.
Metadata Columns	Contains the number of non-text/binary columns for each table. Click the number link to review the full list.
Primary Key	Contains the primary key for each table. Review the following Microsoft article for more information on SQL Server primary keys: Primary Keys Constraints.
Modified Filter	To improve performance the product performs automatic re-indexing against a field in each table that indicates the last modified date of the row. Where possible, the product will automatically map this based upon the exact match or inclusion of one of the below values within the field name. Additional values can be added below in order to support other naming conventions for modified fields (different language or internal convention).
Include?	Select if you want to disable crawling for this table. **NOTE:** You can disable crawling for all listed tables using the **Include none** option in the upper right corner of the wizard or enable crawling accordingly with the **Include all**.
View Sample	Shows a table of the top 15 rows allowing to view if the table is one to exclude.

5.1.3. Exceptions

On this step, review tables with missing primary keys and/or missing modified filters.

- **Missing primary keys** – only shows if users have tables that are missing primary keys where the user can select the primary key from a dropdown of all the columns. This step does not show if there are no missing primary keys.
- **Missing modified filters** – only shows if there are tables missing modified filters. Here tables are shown that are missing a modified and that have a datetime (or equivalent) typed column to select. If there are none this stage is skipped.

5.1.4. Summary

At this step, review your database configuration.

- **Overview** – review a high-level overview of the number of configured tables and excluded tables with their details.
- **Configured Tables** – double-check the configuration of tables to be crawled.
- **Excluded Tables** – review the full list of the tables to be excluded from classification scope with exclusion reason.

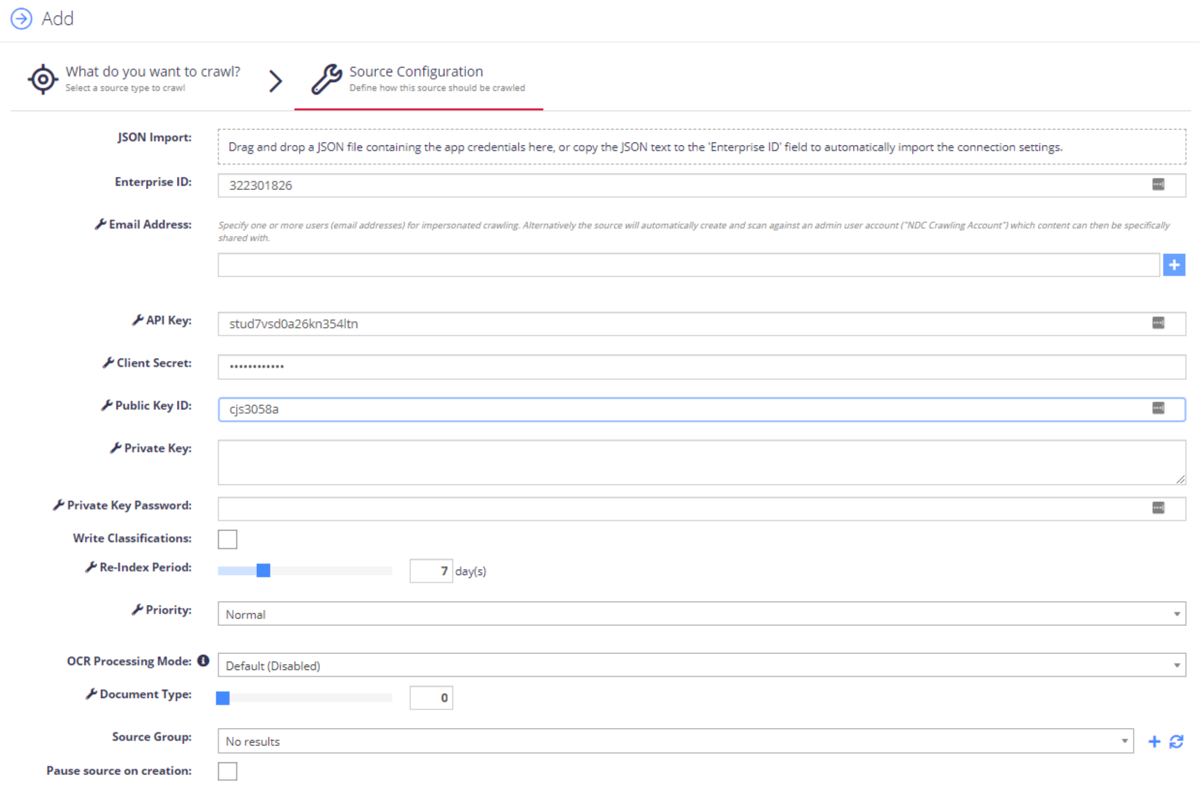
When the database configuration has been completed you will be redirected to the **Advanced Source Configuration**, this allows you to define how the database will be crawled. It is possible to crawl either specific tables, or crawl custom queries (defined select statements, which may use JOIN statements across multiple tables). See Database for more information.

6. Add Box

Use the **Box** source configuration window to set up the crawling and classification operations for content stored in a Box Enterprise account.

By default, configuration window displays basic configuration settings only. It is recommended that you click the "wrench" icon in the bottom left corner to configure advanced settings.

NOTE: To configure advanced settings, your user account may need advanced privileges. See [Security (Users)]() for more information.



Add

What do you want to crawl? Select a source type to crawl > Source Configuration Define how this source should be crawled

JSON Import: Drag and drop a JSON file containing the app credentials here, or copy the JSON text to the 'Enterprise ID' field to automatically import the connection settings.

Enterprise ID: 322301826

🔧 Email Address: *Specify one or more users (email addresses) for impersonated crawling. Alternatively the source will automatically create and scan against an admin user account ("NDC Crawling Account") which content can then be specifically shared with.*

🔧 API Key: stud7vsd0a26kn354ltn

🔧 Client Secret: ••••••••••••

🔧 Public Key ID: cjs3058a

🔧 Private Key:

🔧 Private Key Password:

Write Classifications: □

🔧 Re-Index Period: 7 day(s)

🔧 Priority: Normal

OCR Processing Mode: Default (Disabled)

🔧 Document Type: 0

Source Group: No results

Pause source on creation: □

Configure the following:

Setting	Description
	Basic settings
JSON Import	Drag and drop the JSON file with Box app configuration settings that you downloaded at [Step 1. Create the App]() (see #12). The program then parses this file so that many settings are filled in automatically.
Enterprise ID	Specifies the internal unique identifier for your Box account (filled in

Setting	Description
	automatically).
API Key	*Client ID* of the Box app created at Step 1. Create the App(Filled in automatically.)
Client Secret	Will be generated when allowing access to the Netwrix Data Classification app. Is also known as the “App Key”.
Public Key ID Private Key Private Key Password	Created when generating the trust between your Box account, and the Netwrix Data Classification app – these should be kept secret and secure.
Write Classifications	Identifies whether classifications should be written back to the Box source documents. Classification results can either be written to classification templates or to the generic ‘tags’ property. This is specified using the **Write Configuration** setting of the source.
Source Group	Select the source group (if any).
Pause source on creation	Select if you want to make other configuration changes before collection of the source occurs.
	Advanced settings
Email Address	Specify one or more users (email addresses) for impersonated crawling. ○ If specified explicitly, the crawling engine will impersonate these users when crawling their content as well as shared content where they are the owners. Enter one or several accounts from those listed in the *Managed Users* on the '*Users and Groups*' tab of the Box console. ○ If not specified explicitly, the program will automatically create and use an admin user account (*NDC Crawling Account*) for crawling. Remember to provide this app user account with sufficient permissions for the content you want to index (i.e. share access). To share content for crawling with this account, use group membership.
Re-Index Period	Specifies how often the source should be checked for changes (period in days). Default is **7** days.
Priority	Set priority for this data source to be crawled.

Setting	Description
Document Type	Specify a value which can be used to restrict queries when utilizing the Netwrix Data Classification search index.

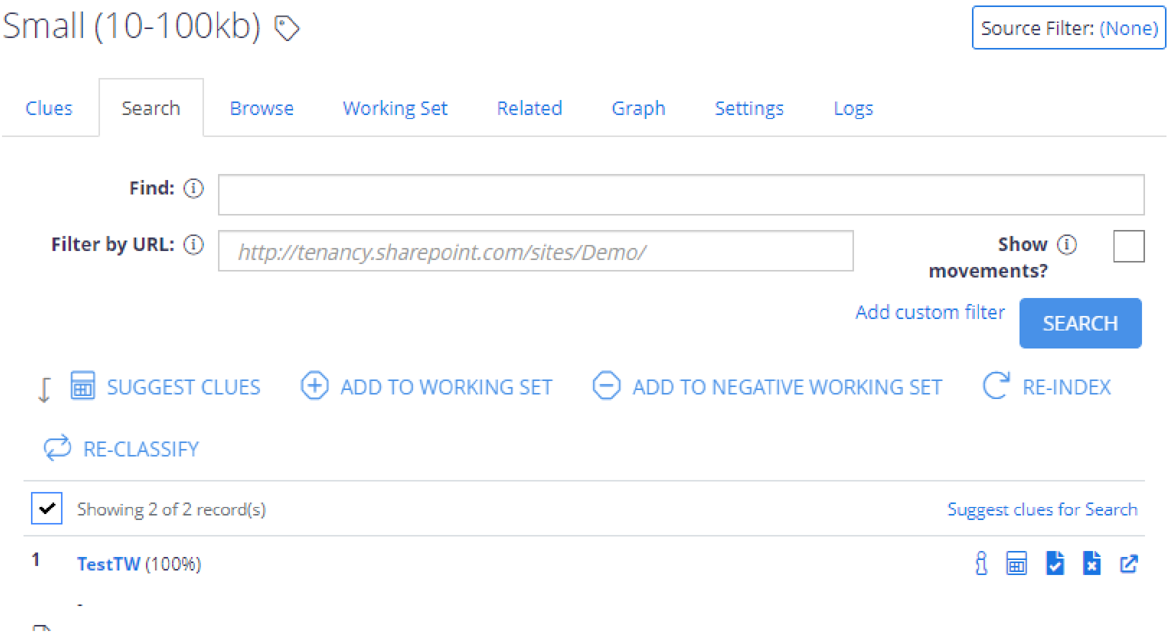
7. Review Reports and Browse Classified Documents

Once your documents are classified, you can identify sensitive information and reduce its exposure. Netwrix recommends starting with the **Document Tagging** report to see automatic and manual classifications of the documents within the reporting set. Further, you can browse your documents to see a list of documents achieving the minimum score set for classification in the term. Review the following for additional information:

- To browse classification results
- To review the Document Tagging report

To browse classification results

1. In administrative web console, navigate to **Taxonomies** → **Term Management**.
2. Select **Taxonomy** in the dropdown on the left and then expand specific term you are interested in.
3. Switch to **Browse** tab:



4. Click **Filter** to start browsing your documents.

To review the Document Tagging report

1. In administrative web console, navigate to **Reports** and expand the **Document Reports** set.
2. Select the **Document Tagging** report and click **Show filters** to narrow report scope.

Filter	Description
Taxonomy	By default, the report shows results for all taxonomies. Select the taxonomy you are interested in to restrict report scope.
Score Range	Select the score. Review Scoring for more information.
Classification	By default, the report shows results for all terms within a taxonomy. Limit your results by selected term.
Page URL	Filter your results by selected page URL.
Source	Select source group you created for Google Drive.

3. Click **Generate** and review results.
4. You can also export displayed page to .csv and .xlsx table or download the whole results.

TIP: Upon export, you will be prompted to include any associated document metadata to the report. It can be useful if you want to generate custom security reports. Specify metadata fields and click **Export** to download report.