Activate scanning on your data sources

NetApp Data Classification

NetApp
February 11, 2026

Table of Contents

Activate scanning on your data sources

Scan data sources with NetApp Data Classification

NetApp Data Classification scanns the data in the repositories (the volumes, database schemas, or other user data) that you select to identify personal and sensitive data. Data Classification then maps your organizational data, categorizes each file, and identifies predefined patterns in the data. The result of the scan is an index of personal information, sensitive personal information, data categories, and file types.

After the initial scan, Data Classification continuously scans your data in a round-robin fashion to detect incremental changes. This is why it's important to keep the instance running.

You can enable and disable scans at the volume level or at the database schema level.

What's the difference between Mapping and Classification scans

You can conduct two types of scans in Data Classification:

- **Mapping-only scans** provide only a high-level overview of your data and are performed on selected data sources. Mapping-only scans take less time than map and classify scans because the do not access files to see the data inside. You might want to do this initially to identify areas of research and then perform a Map & Classify scan on those areas.
- **Map & Classify scans** provide deep-level scanning of your data.

The table below shows some of the differences:

Feature	Map & classify scans	Mapping-only scans
Scan speed	Slow	Fast
Pricing	Free	Free
Capacity	Limited to 500 TiB*	Limited to 500 TiB*
List of file types and used capacity	Yes	Yes
Number of files and used capacity	Yes	Yes
Age and size of files	Yes	Yes
Ability to run a Data Mapping Report	Yes	Yes
Data Investigation page to view file details	Yes	No
Search for names within files	Yes	No
Create saved queries that provide custom search results	Yes	No
Ability to run other reports	Yes	No
Ability to see metadata from files**	No	Yes

* Data Classification does not impose a limit on the amount of data it can scan. Each Console agent supports

scanning and displaying 500 TiB of data. To scan more than 500 TiB of data, install another Console agent then deploy another Data Classification instance.
The Console UI displays data from a single connector. For tips on viewing data from multiple Console agents, see Work with multiple Console agents.

** The following metadata is extracted from files during mapping scans:

- System
- System type
- Storage repository
- File type
- Used capacity
- Number of files
- File size
- File creation
- File last access
- File last modified
- File discovered time
- Permissions extraction

Governance dashboard differences:

Feature	Map & Classify	Map
Stale data	Yes	Yes
Non-business data	Yes	Yes
Duplicated files	Yes	Yes
Predefined saved queries	Yes	No
Default saved queries	Yes	Yes
DDA report	Yes	Yes
Mapping report	Yes	Yes
Sensitivity level detection	Yes	No
Sensitive data with wide permissions	Yes	No
Open permissions	Yes	Yes
Age of data	Yes	Yes
Size of data	Yes	Yes
Categories	Yes	No
File types	Yes	Yes

Compliance dashboard differences:

Feature	Map & Classify	Map
Personal information	Yes	No
Sensitive personal information	Yes	No
Privacy risk assessment report	Yes	No
HIPAA report	Yes	No
PCI DSS report	Yes	No

Investigation filters differences:

Feature	Map & Classify	Map
Saved queries	Yes	Yes
System type	Yes	Yes
System	Yes	Yes
Storage repository	Yes	Yes
File type	Yes	Yes
File size	Yes	Yes
Created time	Yes	Yes
Discovered time	Yes	Yes
Last modified	Yes	Yes
Last access	Yes	Yes
Open permissions	Yes	Yes
File directory path	Yes	Yes
Category	Yes	No
Sensitivity level	Yes	No
Number of identifiers	Yes	No
Personal data	Yes	No
Sensitive personal data	Yes	No
Data subject	Yes	No
Duplicates	Yes	Yes
Classification status	Yes	Status is always "Limited insights"
Scan analysis event	Yes	Yes
File hash	Yes	Yes
Number of users with access	Yes	Yes
User/group permissions	Yes	Yes
File owner	Yes	Yes
Directory type	Yes	Yes

Scan Amazon FSx for ONTAP volumes with NetApp Data Classification

Complete a few steps to scan Amazon FSx for ONTAP volumes with NetApp Data Classification.

Before you begin

- You need an active Console agent in AWS to deploy and manage Data Classification.
- The security group you selected when creating the system must allow traffic from the Data Classification instance. You can find the associated security group using the ENI connected to the FSx for ONTAP file system and edit it using the AWS Management Console.

  AWS security groups for Linux instances

  AWS security groups for Windows instances

  AWS elastic network interfaces (ENI)

- Ensure the following ports are open to the Data Classification instance:
  - For NFS – ports 111 and 2049.
  - For CIFS – ports 139 and 445.

Deploy the Data Classification instance

Deploy Data Classification if there isn't already an instance deployed.

You should deploy Data Classification in the same AWS network as the Console agent for AWS and the FSx volumes you wish to scan.

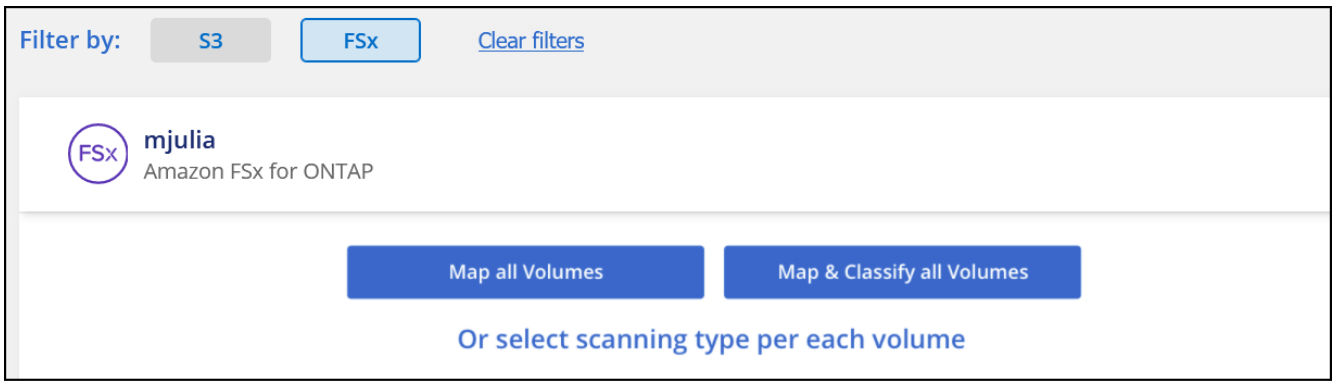
Note: Deploying Data Classification in an on-premises location is not currently supported when scanning FSx volumes.

Upgrades to Data Classification software is automated as long as the instance has internet connectivity.

Enable Data Classification in your systems

You can enable Data Classification for FSx for ONTAP volumes.

1. From NetApp Console, **Governance > Classification**.
2. From the Data Classification menu, select **Configuration**.



3. Select how you want to scan the volumes in each system. Learn about mapping and classification scans:
   - To map all volumes, select **Map all Volumes**.
   - To map and classify all volumes, select **Map & Classify all Volumes**.

- To customize scanning for each volume, select **Or select scanning type for each volume**, and then choose the volumes you want to map and/or classify.

4. In the confirmation dialog box, select **Approve** to have Data Classification start scanning your volumes.

Result

Data Classification starts scanning the volumes you selected in the system. Results will be available in the Compliance dashboard as soon as Data Classification finishes the initial scans. The time that it takes depends on the amount of data—it could be a few minutes or hours. You can track the progress of the initial scan by navigating to the **Configuration** menu then selecting the **System configuration**. Track the progress of each scan in the progress bar; you can hover over the progress bar to see the number of files scanned relative to the total files in the volume.

- By default, if Data Classification doesn't have write attributes permissions in CIFS, or write permissions in NFS, the system won't scan the files in your volumes because Data Classification can't revert the "last access time" to the original timestamp. If you don't care if the last access time is reset, select **Or select scanning type for each volume**. The resulting page has a setting you can enable so that Data Classification will scan the volumes regardless of permissions.
- Data Classification scans only one file share under a volume. If you have multiple shares in your volumes, you'll need to scan those other shares separately as a shares group. See more details about this Data Classification limitation.

Verify that Data Classification has access to volumes

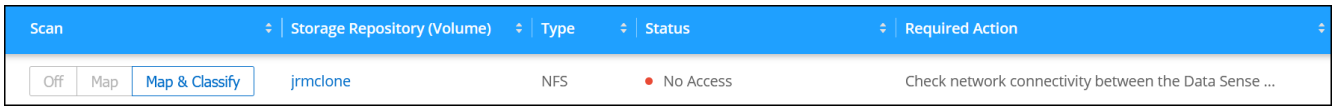
Make sure Data Classification can access volumes by checking your networking, security groups, and export policies.

You'll need to provide Data Classification with CIFS credentials so it can access CIFS volumes.

Steps

1. From the Data Classification menu, select **Configuration**.
2. On the Configuration page, select **View Details** to review the status and correct any errors.

   For example, the following image shows a volume Data Classification can't scan due to network connectivity issues between the Data Classification instance and the volume.



3. Make sure there's a network connection between the Data Classification instance and each network that includes volumes for FSx for ONTAP.

   For FSx for ONTAP, Data Classification can scan volumes only in the same region as the Console.

4. Ensure NFS volume export policies include the IP address of the Data Classification instance so it can access the data on each volume.
5. If you use CIFS, provide Data Classification with Active Directory credentials so it can scan CIFS volumes.
   a. From the Data Classification menu, select **Configuration**.

b. For each system, select **Edit CIFS Credentials** and enter the user name and password that Data Classification needs to access CIFS volumes on the system.

   The credentials can be read-only, but providing admin credentials ensures that Data Classification can read any data that requires elevated permissions. The credentials are stored on the Data Classification instance.

   If you want to make sure your files "last accessed times" are unchanged by Data Classification scans, it's recommended the user has Write Attributes permissions in CIFS or write permissions in NFS. If possible, configure the Active Directory user as part of a parent group in the organization which has permissions to all files.

   After you enter the credentials, you should see a message that all CIFS volumes were authenticated successfully.

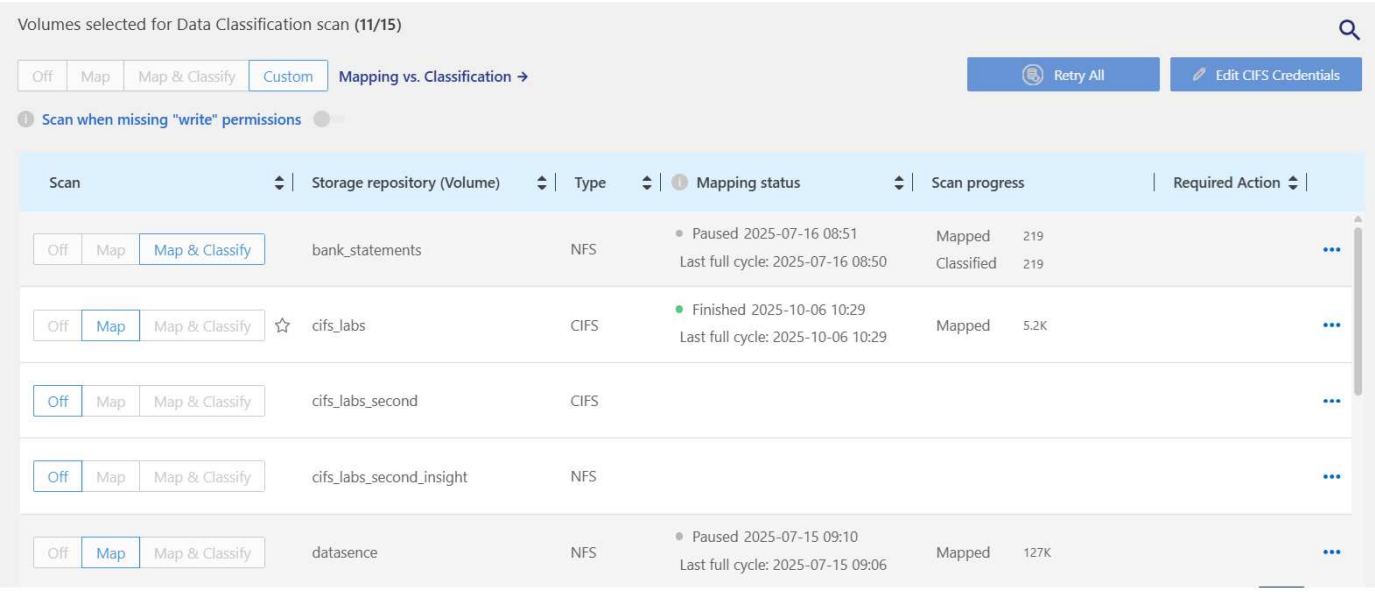
Enable and disable scans on volumes

You can start or stop scans on any system at any time from the Configuration page. You can also switch scans from mapping-only scans to mapping and classification scans, and vice-versa. It's recommended that you scan all volumes in a system.

New volumes added to the system are automatically scanned only when you have selected the **Map** or **Map & Classify** setting in the heading area. When set to **Custom** or **Off** in the heading area, you'll need to activate mapping and/or full scanning on each new volume you add in the system.

The switch at the top of the page for **Scan when missing "write" permissions** is disabled by default. This means that if Data Classification doesn't have write attributes permissions in CIFS or write permissions in NFS, the system won't scan the files because Data Classification can't revert the "last access time" to the original timestamp. If you don't care if the last access time is reset, turn the switch ON and all files are scanned regardless of the permissions. Learn more.

New volumes added to the system are automatically scanned only when you have set the **Map** or **Map & Classify** setting in the heading area. When the setting for all volumes is **Custom** or **Off**, you need to activate scanning manually for each new volume you add.



Steps

1. From the Data Classification menu, select **Configuration**.
2. Choose a system, then select **Configuration**.
3. To enable or disable scans for all volumes, select **Map**, **Map & Classify**, or **Off** in the heading above all volumes.

   To enable or disable scans for individual volumes, find the volumes in the list then select **Map**, **Map & Classify**, or **Off** next to the volume name.

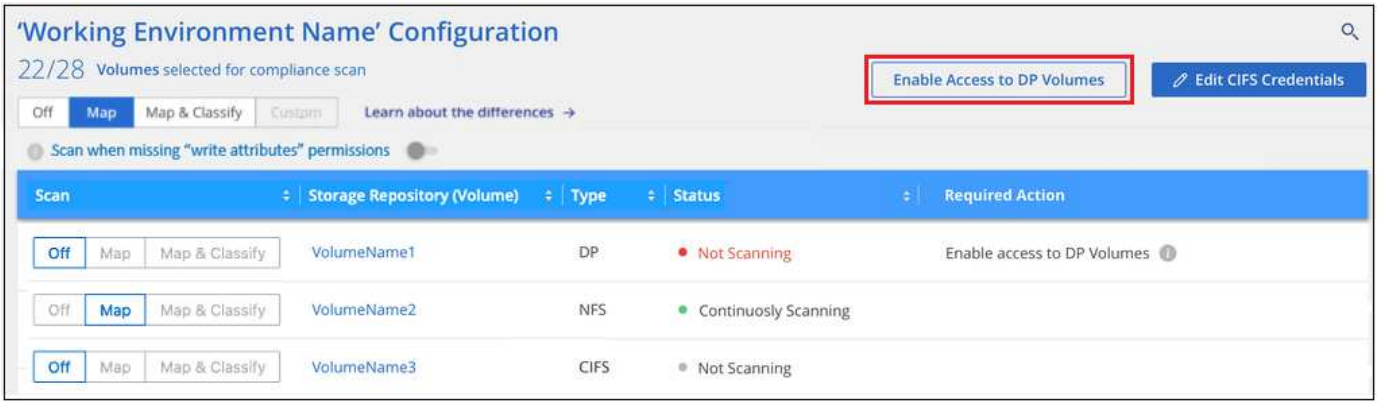
Result

When you enable scanning, Data Classification starts scanning the volumes you selected in the system. Results start to appear in the Compliance dashboard as soon as Data Classification starts the scan. Scan completion time depends on the amount of data, ranging from minutes to hours.

Scan data protection volumes

By default, data protection (DP) volumes are not scanned because they are not exposed externally and Data Classification cannot access them. These are the destination volumes for SnapMirror operations from an FSx for ONTAP file system.

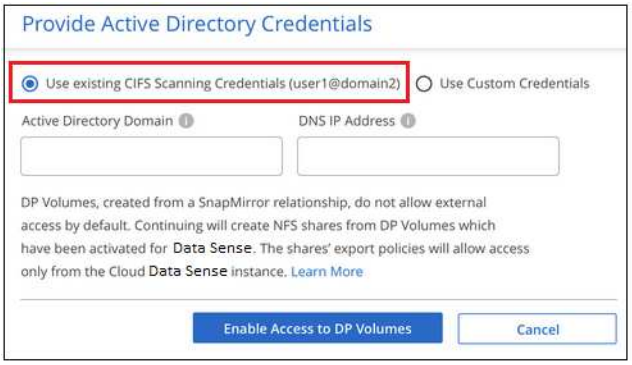
Initially, the volume list identifies these volumes as *Type* **DP** with the *Status* **Not Scanning** and the *Required Action* **Enable Access to DP volumes**.

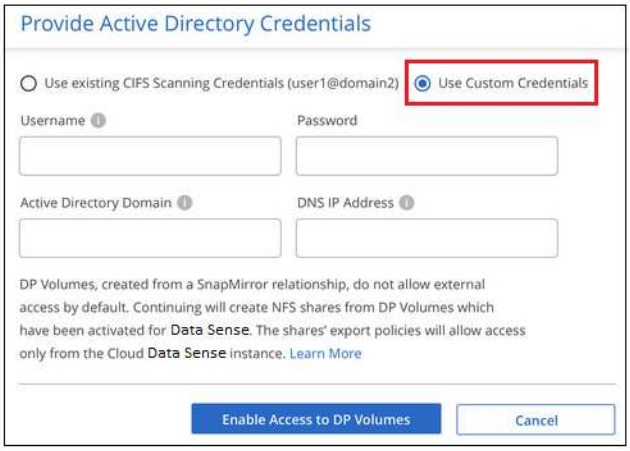


Steps

If you want to scan these data protection volumes:

1. From the Data Classification menu, select **Configuration**.
2. Select **Enable Access to DP volumes** at the top of the page.
3. Review the confirmation message and select **Enable Access to DP volumes** again.
   - Volumes that were initially created as NFS volumes in the source FSx for ONTAP file system are enabled.
   - Volumes that were initially created as CIFS volumes in the source FSx for ONTAP file system require that you enter CIFS credentials to scan those DP volumes. If you already entered Active Directory credentials so that Data Classification can scan CIFS volumes you can use those credentials, or you can specify a different set of Admin credentials.





4. Activate each DP volume that you want to scan.

Result

Once enabled, Data Classification creates an NFS share from each DP volume that was activated for scanning. The share export policies only allow access from the Data Classification instance.

If you had no CIFS data protection volumes when you initially enabled access to DP volumes, and later add some, the button **Enable Access to CIFS DP** appears at the top of the Configuration page. Select this button and add CIFS credentials to enable access to these CIFS DP volumes.

Active Directory credentials are registered only in the storage VM of the first CIFS DP volume, so all DP volumes on that SVM will be scanned. Any volumes that reside on other SVMs will not have the Active Directory credentials registered, so those DP volumes won't be scanned.

Scan Azure NetApp Files volumes with NetApp Data Classification

Complete a few steps to get started with NetApp Data Classification for Azure NetApp Files.

Discover the Azure NetApp Files system that you want to scan

If the Azure NetApp Files system you want to scan is not already in the NetApp Console as a system, add it in the Systems page.

Deploy the Data Classification instance

Deploy Data Classification if there isn't already an instance deployed.

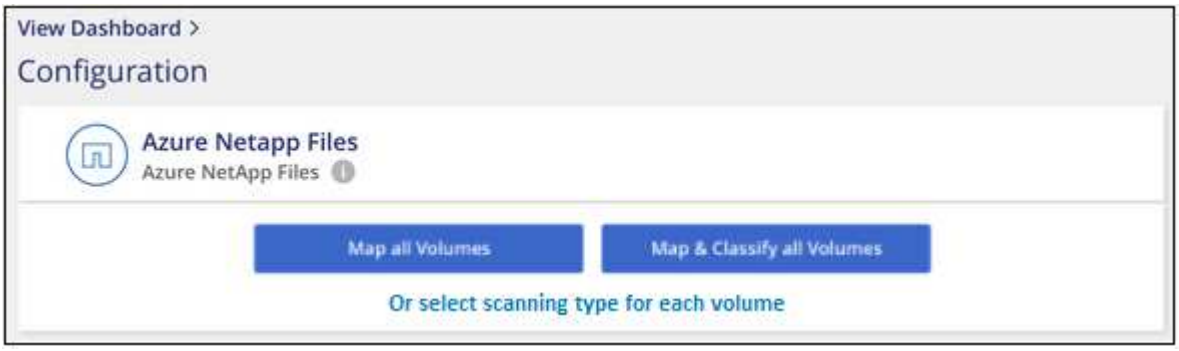
Data Classification must be deployed in the cloud when scanning Azure NetApp Files volumes, and it must be deployed in the same region as the volumes you wish to scan.

Note: Deploying Data Classification in an on-premises location is not currently supported when scanning Azure NetApp Files volumes.

Enable Data Classification in your systems

You can enable Data Classification on your Azure NetApp Files volumes.

1. From the Data Classification menu, select **Configuration**.



2. Select how you want to scan the volumes in each system. Learn about mapping and classification scans:
   - To map all volumes, select **Map all Volumes**.
   - To map and classify all volumes, select **Map & Classify all Volumes**.
   - To customize scanning for each volume, select **Or select scanning type for each volume**, and then choose the volumes you want to map or map and classify.

     See Enable or disable scans on volumes for details.

3. In the confirmation dialog box, select **Approve**.

Result

Data Classification starts scanning the volumes you selected in the system. Results are available in the Compliance dashboard as soon as Data Classification finishes the initial scans. The time that it takes depends on the amount of data—it could be a few minutes or hours. You can track the progress of the initial scan by navigating to the **Configuration** menu then selecting the **System configuration**. Data Classification displays a progress bar for each scan. You can hover over the progress bar to see the number of files scanned relative to the total number of files in the volume.

- By default, if Data Classification doesn't have write attributes permissions in CIFS, or write permissions in NFS, the system won't scan the files in your volumes because Data Classification can't revert the "last access time" to the original timestamp. If you don't care if the last access time is reset, select **Or select scanning type for each volume**. The resulting page has a setting you can enable so that Data Classification will scan the volumes regardless of permissions.
- Data Classification scans only one file share under a volume. If you have multiple shares in your volumes, you'll need to scan those other shares separately as a shares group. Learn about this Data Classification limitation.

Verify that Data Classification has access to volumes

Make sure that Data Classification can access volumes by checking your networking, security groups, and export policies. You need to provide Data Classification with CIFS credentials so it can access CIFS volumes.

For Azure NetApp Files, Data Classification can only scan volumes in the same region as the Console.

Checklist

- Make sure that there's a network connection between the Data Classification instance and each network that includes volumes for Azure NetApp Files.
- Ensure the following ports are open to the Data Classification instance:
  - For NFS – ports 111 and 2049.
  - For CIFS – ports 139 and 445.
- Ensure the NFS volume export policies include the IP address of the Data Classification instance so it can access the data on each volume.

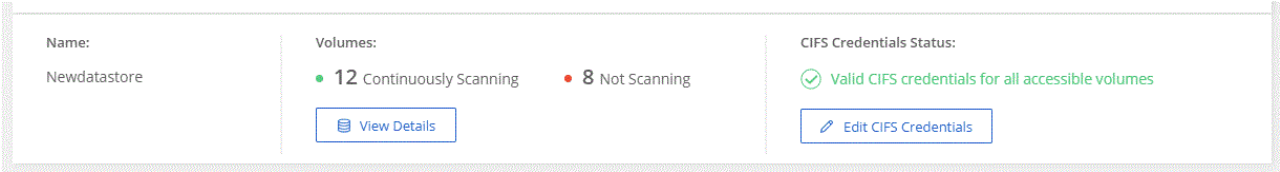
Steps

1. From the Data Classification menu, select **Configuration**.
   a. If you're using CIFS (SMB), ensure the Active Directory credentials are correct. For each system, select **Edit CIFS Credentials** then enter the user name and password that Data Classification needs to access CIFS volumes on the system.

      The credentials can be read-only; providing admin credentials ensures that Data Classification can read any data that requires elevated permissions. The credentials are stored on the Data Classification instance.

      If you want to make sure your files "last accessed times" are unchanged by Data Classification scans, it's recommended the user has Write Attributes permissions in CIFS or write permissions in NFS. If possible, configure the Active Directory user as part of a parent group in the organization which has permissions to all files.

      After you enter the credentials, you should see a message that all CIFS volumes were authenticated successfully.



2. On the Configuration page, select **View Details** to review the status for each CIFS and NFS volume. If necessary, correct any errors such as network connectivity issues.

Enable or disable scans on volumes

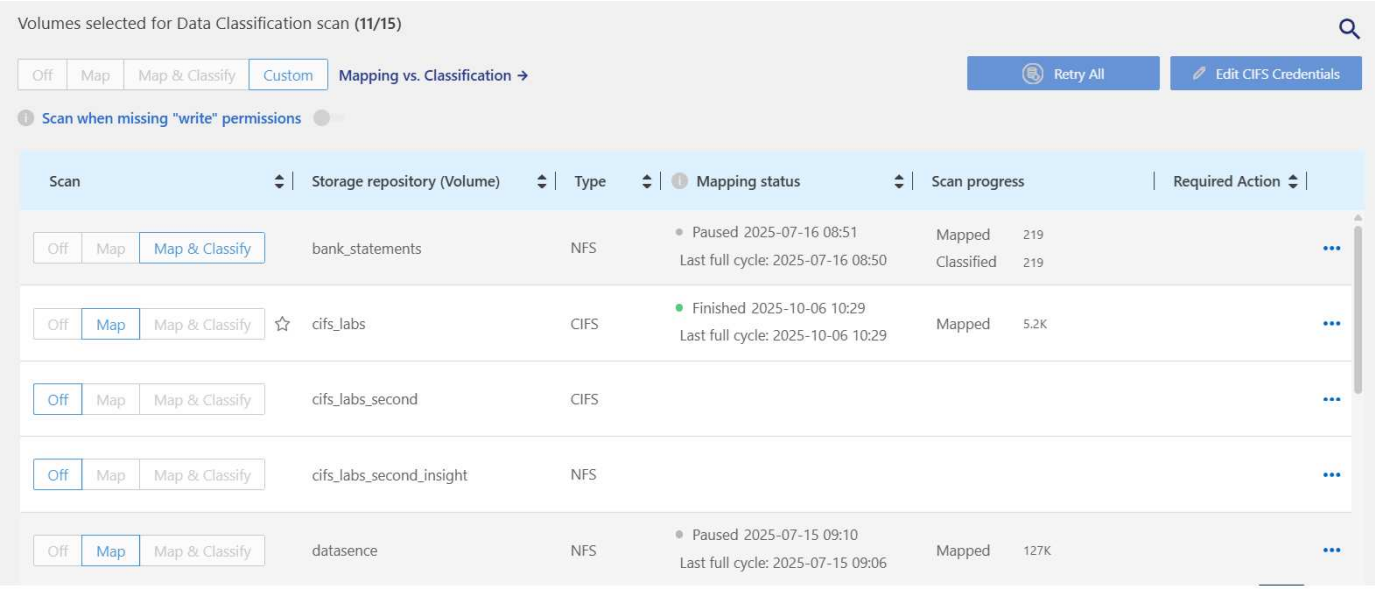
You can start or stop scans on any system at any time from the Configuration page. You can also switch scans from mapping-only scans to mapping and classification scans, and vice-versa. It's recommended that you scan all volumes in a system.

New volumes added to the system are automatically scanned only when you have selected the **Map** or **Map & Classify** setting in the heading area. When set to **Custom** or **Off** in the heading area, you'll need to activate mapping and/or full scanning on each new volume you add in the system.

The switch at the top of the page for **Scan when missing "write" permissions** is disabled by default. This means that if Data Classification doesn't have write attributes permissions in CIFS or write permissions in NFS, the system won't scan the files because Data Classification can't revert the "last access time" to the original timestamp. If you don't care if the last access time is reset, turn the switch ON and all files are scanned

regardless of the permissions. Learn more.

New volumes added to the system are automatically scanned only when you have set the **Map** or **Map & Classify** setting in the heading area. When the setting for all volumes is **Custom** or **Off**, you need to activate scanning manually for each new volume you add.



Steps

1. From the Data Classification menu, select **Configuration**.
2. Choose a system, then select **Configuration**.
3. To enable or disable scans for all volumes, select **Map**, **Map & Classify**, or **Off** in the heading above all volumes.

   To enable or disable scans for individual volumes, find the volumes in the list then select **Map**, **Map & Classify**, or **Off** next to the volume name.

Result

When you enable scanning, Data Classification starts scanning the volumes you selected in the system. Results start to appear in the Compliance dashboard as soon as Data Classification starts the scan. Scan completion time depends on the amount of data, ranging from minutes to hours.

Scan Cloud Volumes ONTAP and on-premises ONTAP volumes with NetApp Data Classification

Complete a few steps to start scanning your Cloud Volumes ONTAP and on-premises ONTAP volumes using NetApp Data Classification.

Prerequisites

Before you enable Data Classification, make sure you have a supported configuration.

- If you are scanning Cloud Volumes ONTAP and on-premises ONTAP systems that are accessible over the internet, you can deploy Data Classification in the cloud or in an on-premises location that has internet

access.

- If you are scanning on-premises ONTAP systems that have been installed in a dark site that has no internet access, you need to deploy Data Classification in the same on-premises location that has no internet access. This requires the Console agent to be deployed in that same on-premises location.

Verify that Data Classification has access to volumes

Make sure that Data Classification can access volumes by checking your networking, security groups, and export policies. You'll need to provide Data Classification with CIFS credentials so it can access CIFS volumes.

Checklist

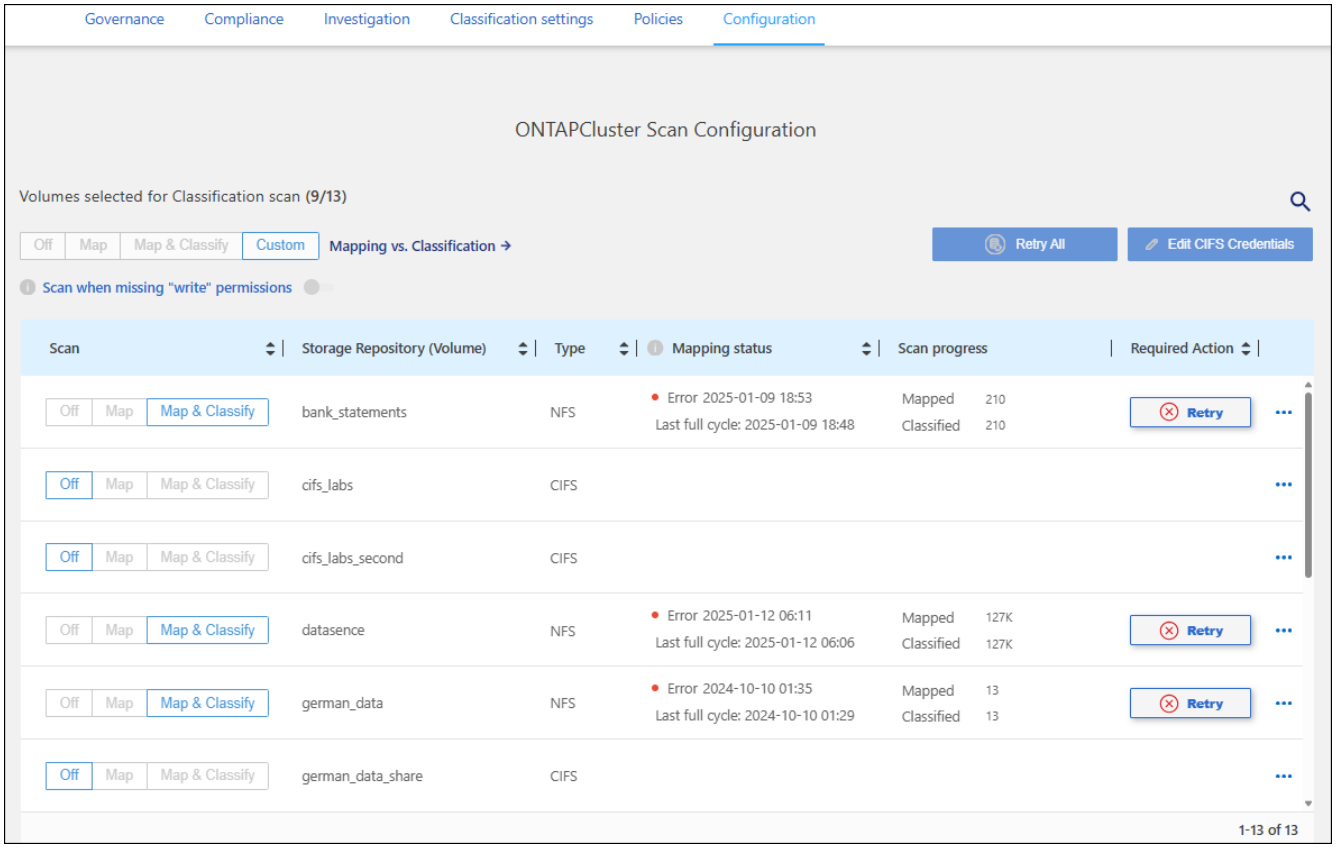
- Make sure that there's a network connection between the Data Classification instance and each network that includes volumes for Cloud Volumes ONTAP or on-prem ONTAP clusters.
- Ensure that the security group for Cloud Volumes ONTAP allows inbound traffic from the Data Classification instance.

  You can either open the security group for traffic from the IP address of the Data Classification instance, or you can open the security group for all traffic from inside the virtual network.

- Ensure that NFS volume export policies include the IP address of the Data Classification instance so it can access the data on each volume.

Steps

1. From the Data Classification menu, select **Configuration**.



2. If you use CIFS, provide Data Classification with Active Directory credentials so it can scan CIFS volumes. For each system, select **Edit CIFS Credentials** and enter the user name and password that Data Classification needs to access CIFS volumes on the system.

The credentials can be read-only, but providing admin credentials ensures that Data Classification can read any data that requires elevated permissions. The credentials are stored on the Data Classification instance.

If you want to make sure your files "last accessed times" are unchanged by Data Classification scans, it's recommended the user has Write Attributes permissions in CIFS or write permissions in NFS. If possible, configure the Active Directory user as part of a parent group in the organization which has permissions to all files.

If you've entered the credentials correctly, a message confirms all CIFS volumes were authenticated successfully.

3. On the Configuration page, select **Configuration** to review the status for each CIFS and NFS volume and correct any errors.

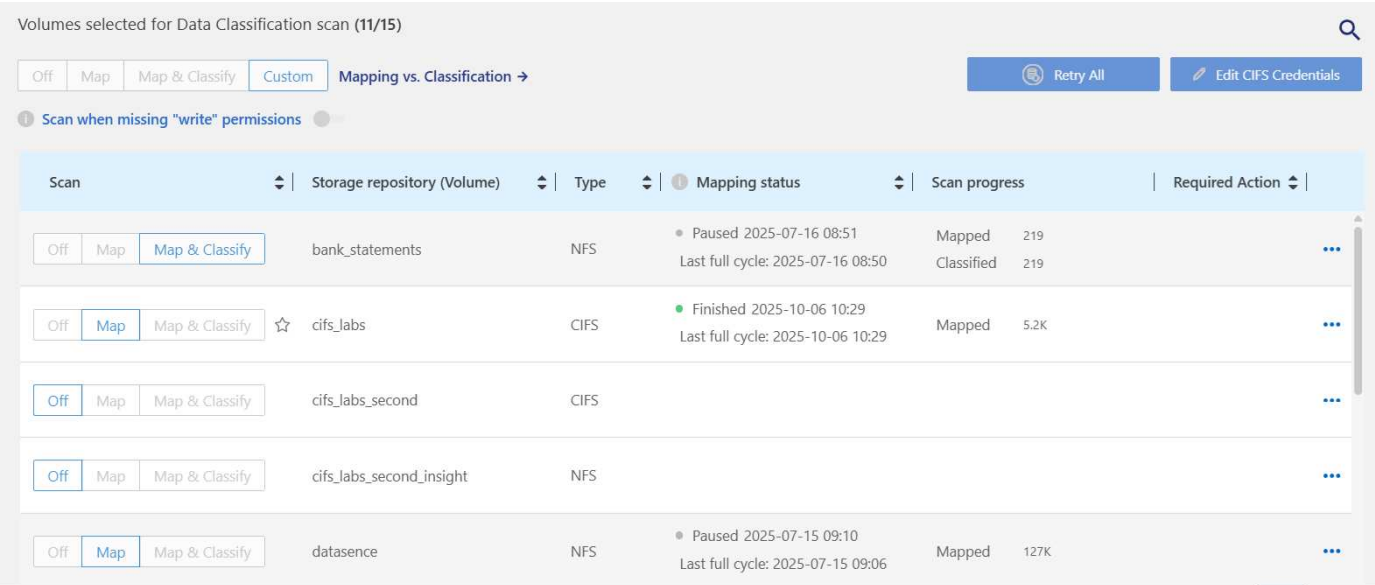
Enable or disable scans on volumes

You can start or stop scans on any system at any time from the Configuration page. You can also switch scans from mapping-only scans to mapping and classification scans, and vice-versa. It's recommended that you scan all volumes in a system.

New volumes added to the system are automatically scanned only when you have selected the **Map** or **Map & Classify** setting in the heading area. When set to **Custom** or **Off** in the heading area, you'll need to activate mapping and/or full scanning on each new volume you add in the system.

The switch at the top of the page for **Scan when missing "write" permissions** is disabled by default. This means that if Data Classification doesn't have write attributes permissions in CIFS or write permissions in NFS, the system won't scan the files because Data Classification can't revert the "last access time" to the original timestamp. If you don't care if the last access time is reset, turn the switch ON and all files are scanned regardless of the permissions. Learn more.

New volumes added to the system are automatically scanned only when you have set the **Map** or **Map & Classify** setting in the heading area. When the setting for all volumes is **Custom** or **Off**, you need to activate scanning manually for each new volume you add.



Steps

1. From the Data Classification menu, select **Configuration**.
2. Choose a system, then select **Configuration**.
3. To enable or disable scans for all volumes, select **Map**, **Map & Classify**, or **Off** in the heading above all volumes.

   To enable or disable scans for individual volumes, find the volumes in the list then select **Map**, **Map & Classify**, or **Off** next to the volume name.

Result

When you enable scanning, Data Classification starts scanning the volumes you selected in the system. Results start to appear in the Compliance dashboard as soon as Data Classification starts the scan. Scan completion time depends on the amount of data, ranging from minutes to hours.

Data Classification scans only one file share under a volume. If you have multiple shares in your volumes, you'll need to scan those other shares separately as a shares group. See more details about this Data Classification limitation.

Scan database schemas with NetApp Data Classification

Complete a few steps to start scanning your database schemas with NetApp Data Classification.

Review prerequisites

Review the following prerequisites to make sure that you have a supported configuration before you enable Data Classification.

Supported databases

Data Classification can scan schemas from the following databases:

- Amazon Relational Database Service (Amazon RDS)
- MongoDB
- MySQL
- Oracle
- PostgreSQL
- SAP HANA
- SQL Server (MSSQL)

The statistics gathering feature **must be enabled** in the database.

Database requirements

Any database with connectivity to the Data Classification instance can be scanned, regardless of where it is hosted. You just need the following information to connect to the database:

- IP address or host name
- Port
- Service name (only for accessing Oracle databases)
- Credentials that allow read access to the schemas

  When choosing a user name and password, it's important to choose one that has full read permissions to all the schemas and tables you want to scan. We recommend that you create a dedicated user for the Data Classification system with all the required permissions.

For MongoDB, a read-only admin role is required.

Deploy the Data Classification instance

Deploy Data Classification if there isn't already an instance deployed.

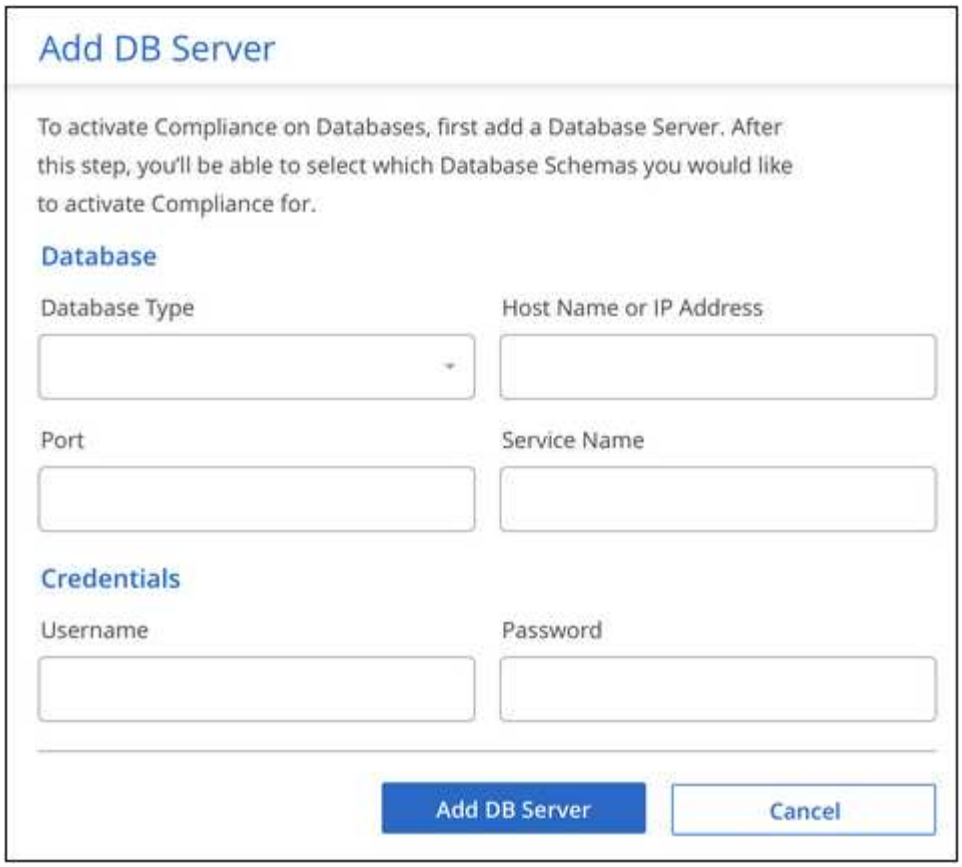
If you are scanning database schemas that are accessible over the internet, you can deploy Data Classification in the cloud or deploy Data Classification in an on-premises location that has internet access.

If you are scanning database schemas that have been installed in a dark site that has no internet access, you need to deploy Data Classification in the same on-premises location that has no internet access. This also requires that the Console agent is deployed in that same on-premises location.

Add the database server

Add the database server where the schemas reside.

1. From the Data Classification menu, select **Configuration**.
2. From the Configuration page, select **Add System** > **Add Database Server**.
3. Enter the required information to identify the database server.
   a. Select the database type.
   b. Enter the port and the host name or IP address to connect to the database.
   c. For Oracle databases, enter the Service name.
   d. Enter the credentials so that Data Classification can access the server.
   e. Select **Add DB Server**.



The database is added to the list of systems.

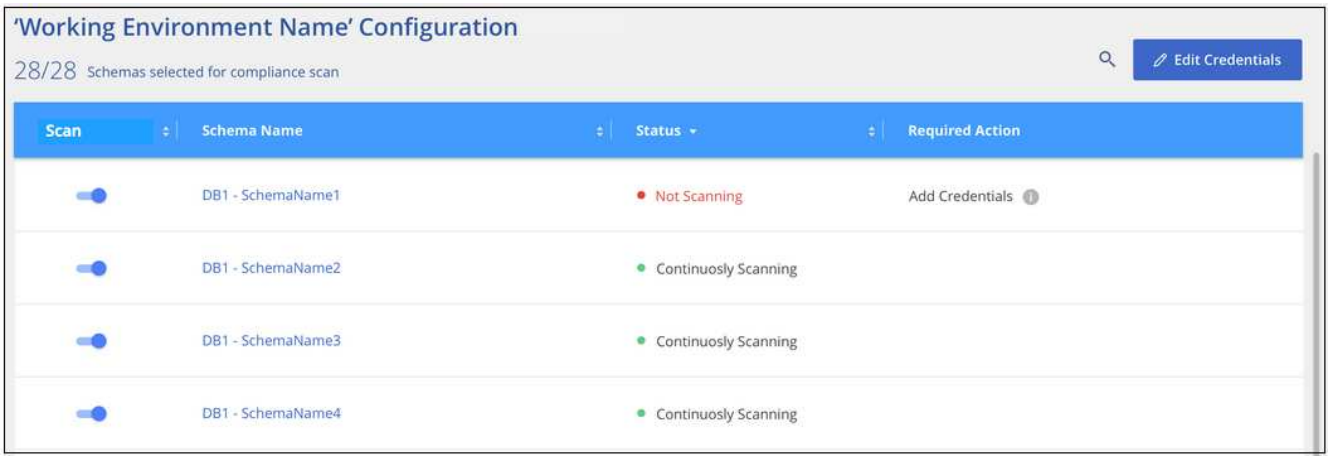
Enable and disable scans on database schemas

You can stop or start full scanning of your schemas at any time.

There is no option to select mapping-only scans for database schemas.

1. From the Configuration page, select the **Configuration** button for the database you want to configure.

2. Select the schemas that you want to scan by moving the slider to the right.



Result

Data Classification starts scanning the database schemas that you enabled. You can track the progress of the initial scan by navigating to the **Configuration** menu then selecting the **System configuration**. The progress of each scan is shown as a progress bar. You can also hover over the progress bar to see the number of files scanned relative to the total number of files in the volume. If there are any errors, they'll appear in the Status column, alongside the required actions to fix the error.

Data Classification scans your databases once per day; databases are not continuously scanned like other data sources.

Scan Google Cloud NetApp Volumes with NetApp Data Classification

NetApp Data Classification supports Google Cloud NetApp Volumes as a system. Learn how to scan your Google Cloud NetApp Volumes system.

Discover the Google Cloud NetApp Volumes system that you want to scan

If the Google Cloud NetApp Volumes system you want to scan is not already in the NetApp Console as a system, add it to the Systems page.

Deploy the Data Classification instance

Deploy Data Classification if there isn't already an instance deployed.

Data Classification must be deployed in the cloud when scanning Google Cloud NetApp Volumes, and it must be deployed in the same region as the volumes you wish to scan.

Note: Deploying Data Classification in an on-premises location is not currently supported when scanning Google Cloud NetApp Volumes.

Enable Data Classification in your systems

You can enable Data Classification on your Google Cloud NetApp Volumes system.

1. From the Data Classification menu, select **Configuration**.
2. Select how you want to scan the volumes in each system. Learn about mapping and classification scans:

- To map all volumes, select **Map all Volumes**.
- To map and classify all volumes, select **Map & Classify all Volumes**.
- To customize scanning for each volume, select **Or select scanning type for each volume**, and then choose the volumes you want to map and/or classify.

  See Enable and disable scans on volumes for details.

3. In the confirmation dialog box, select **Approve**.

Result

Data Classification starts scanning the volumes you selected in the system. Results are available in the Compliance dashboard as soon as Data Classification finishes the initial scans. The time that it takes depends on the amount of data: a few minutes to a few hours. You can track the progress of the initial scan in the **Configuration** menu's **System configuration** section. Data Classification displays a progress bar for each scan. You can also hover over the progress bar to see the number of files scanned relative to the total files in the volume.

- By default, if Data Classification doesn't have write attributes permissions in CIFS, or write permissions in NFS, the system won't scan the files in your volumes because Data Classification can't revert the "last access time" to the original timestamp. If you don't care if the last access time is reset, select **Or select scanning type for each volume**. The resulting page has a setting you can enable so that Data Classification will scan the volumes regardless of permissions.
- Data Classification scans only one file share under a volume. If you have multiple shares in your volumes, you need to scan those other shares separately as a shares group. Learn about this Data Classification limitation.

Verify that Data Classification has access to volumes

Ensure Data Classification can access volumes by checking your networking, security groups, and export policies. For CIFS volumes, you need to provide Data Classification with CIFS credentials.

For Google Cloud NetApp Volumes, Data Classification can only scan volumes in the same region as the Console.

Checklist

- Make sure that there's a network connection between the Data Classification instance and each network that includes volumes for Google Cloud NetApp Volumes.
- Ensure the following ports are open to the Data Classification instance:
  - For NFS – ports 111 and 2049.
  - For CIFS – ports 139 and 445.
- Ensure the NFS volume export policies include the IP address of the Data Classification instance so it can access the data on each volume.

Steps

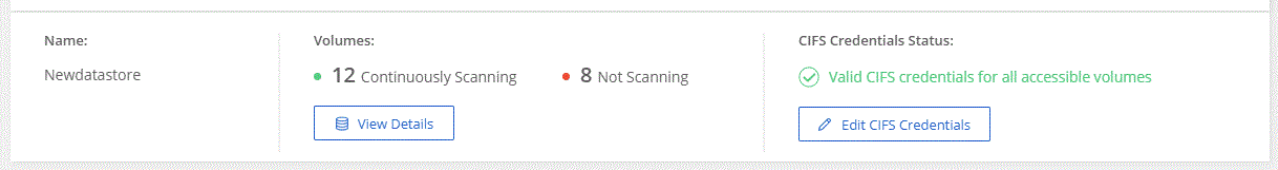
1. From the Data Classification menu, select **Configuration**.
   a. If you're using CIFS (SMB), ensure the Active Directory credentials are correct. For each system, select **Edit CIFS Credentials** then enter the user name and password that Data Classification needs to access CIFS volumes on the system.

      The credentials can be read-only, but providing admin credentials ensures that Data Classification can

read any data that requires elevated permissions. The credentials are stored on the Data Classification instance.

If you want to make sure your files "last accessed times" are unchanged by Data Classification scans, it's recommended the user has Write Attributes permissions in CIFS or write permissions in NFS. If possible, configure the Active Directory user as part of a parent group in the organization which has permissions to all files.

After you enter the credentials, you should see a message that all CIFS volumes were authenticated successfully.



2. On the Configuration page, select **View Details** to review the status for each CIFS and NFS volume and correct any errors.

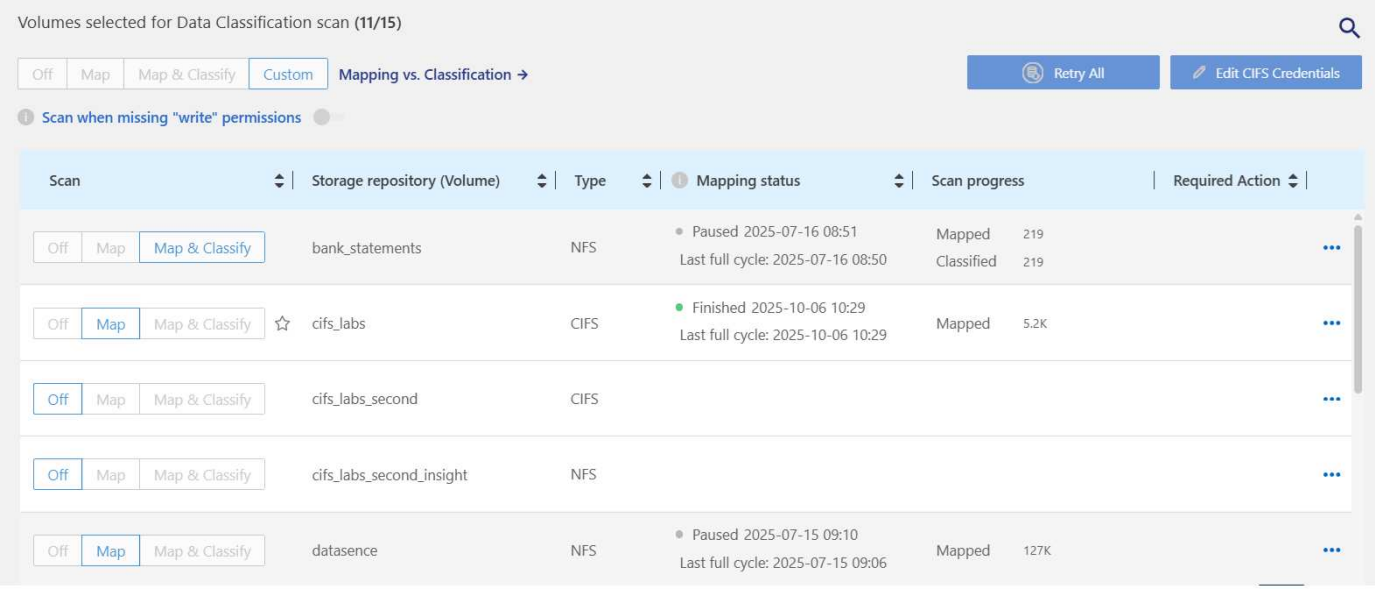
Enable and disable scans on volumes

You can start or stop scans on any system at any time from the Configuration page. You can also switch scans from mapping-only scans to mapping and classification scans, and vice-versa. It's recommended that you scan all volumes in a system.

New volumes added to the system are automatically scanned only when you have selected the **Map** or **Map & Classify** setting in the heading area. When set to **Custom** or **Off** in the heading area, you'll need to activate mapping and/or full scanning on each new volume you add in the system.

The switch at the top of the page for **Scan when missing "write" permissions** is disabled by default. This means that if Data Classification doesn't have write attributes permissions in CIFS or write permissions in NFS, the system won't scan the files because Data Classification can't revert the "last access time" to the original timestamp. If you don't care if the last access time is reset, turn the switch ON and all files are scanned regardless of the permissions. Learn more.

New volumes added to the system are automatically scanned only when you have set the **Map** or **Map & Classify** setting in the heading area. When the setting for all volumes is **Custom** or **Off**, you need to activate scanning manually for each new volume you add.



Steps

1. From the Data Classification menu, select **Configuration**.
2. Choose a system, then select **Configuration**.
3. To enable or disable scans for all volumes, select **Map**, **Map & Classify**, or **Off** in the heading above all volumes.

   To enable or disable scans for individual volumes, find the volumes in the list then select **Map**, **Map & Classify**, or **Off** next to the volume name.

Result

When you enable scanning, Data Classification starts scanning the volumes you selected in the system. Results start to appear in the Compliance dashboard as soon as Data Classification starts the scan. Scan completion time depends on the amount of data, ranging from minutes to hours.

Scan file shares with NetApp Data Classification

To scan file shares, you must first create a file shares group in NetApp Data Classification. File shares groups are for NFS or CIFS (SMB) shares hosted on-premises or in the cloud.

Scanning data from non-NetApp file shares is not supported in the Data Classification core version.

Prerequisites

Review the following prerequisites to make sure that you have a supported configuration before you enable Data Classification.

- The shares can be hosted anywhere, including in the cloud or on-premises. CIFS shares from older NetApp 7-Mode storage systems can be scanned as file shares.
  - Data Classification can’t extract permissions or the "last access time" from 7-Mode systems.

    - Because of a known issue between some Linux versions and CIFS shares on 7-Mode systems, you must configure the share to use only SMBv1 with NTLM authentication enabled.
- There needs to be network connectivity between the Data Classification instance and the shares.
- You can add a DFS (Distributed File System) share as a regular CIFS share. Because Data Classification is unaware that the share is built upon multiple servers/volumes combined as a single CIFS share, you might receive permission or connectivity errors about the share when the message really only applies to one of the folders/shares that is located on a different server/volume.
- For CIFS (SMB) shares, ensure that you have Active Directory credentials that provide read access to the shares. Admin credentials are preferred in case Data Classification needs to scan any data that requires elevated permissions.

  If you want to make sure your files "last accessed times" are unchanged by Data Classification scans, it's recommended the user has Write Attributes permissions in CIFS or write permissions in NFS. If possible, configure the Active Directory user as part of a parent group in the organization which has permissions to all files.

- All CIFS file shares in a group must use the same Active Directory credentials.
- You can mix NFS and CIFS (using either Kerberos or NTLM) shares. You must add the shares to the group separately. That is, you must complete the process twice—once per protocol.
    - You cannot create a file shares group that mixes CIFS authentication types (Kerberos and NTLM).
- If you're using CIFS with Kerberos authentication, ensure the IP address provided is accessible to the Data Classification. The files shares can't be added if the IP address is unreachable.

Create a file shares group

When you add file shares to the group, you must use the format `<host_name>:/<share_path>`.

You can add file shares individually or you can enter a line-separated list of the file shares you want to scan. You can add up to 100 shares at a time.

Steps

1. From the Data Classification menu, select **Configuration**.
2. From the Configuration page, select **Add System** > **Add File Shares Group**.
3. In the Add File Shares Group dialog, enter the name for the group of shares then select **Continue**.
4. Select the protocol for the file shares you are adding.

    a. If you're adding CIFS shares with NTLM authentication, enter the Active Directory credentials to access the CIFS volumes. Although read-only credentials are supported, it's recommended you provide full access with administrator credentials. Select **Save**.
5. Add the file shares that you want to scan (one file share per line). Then select **Continue**.
6. A confirmation dialog displays the number of shares that were added.

    If the dialog lists any shares that could not be added, capture this information so that you can resolve the issue. If the issue pertains to a naming convention, you can re-add the share with a corrected name.

7. Configure scanning on the volume:

- To enable mapping-only scans on file shares, select **Map**.
- To enable full scans on file shares, select **Map & Classify**.
- To disable scanning on file shares, select **Off**.

> The switch at the top of the page for **Scan when missing "write attributes" permissions** is disabled by default. This means that if Data Classification doesn't have write attributes permissions in CIFS or write permissions in NFS, the system won't scan the files because Data Classification can't revert the "last access time" to the original timestamp.
> If you switch **Scan when missing "write attributes" permissions** to **On**, the scan resets the last accessed time and scans all files regardless of permissions.
> To learn more about the last accessed time stamp, see Metadata collected from data sources in Data Classification.

Result

Data Classification starts scanning the files in the file shares you added. You can Track the scanning progress and view the results of the scan in the **Dashboard**.

> If the scan doesn't complete successfully for a CIFS configuration with Kerberos authentication, check the **Configuration** tab for errors.

Edit a file shares group

After you create a file shares group, you can edit the CIFS protocol or add and remove file shares.

Edit the CIFS protocol configuration

1. From the Data Classification menu, select **Configuration**.
2. From the Configuration page, select the file shares group you want to modify.
3. Select **Edit CIFS Credentials**.

4. Choose the authentication method: **NTLM** or **Kerberos**.
5. Enter the Active Directory **Username** and **Password**.
6. Select **Save** to complete the process.

Add file shares to scans

1. From the Data Classification menu, select **Configuration**.
2. From the Configuration page, select the file shares group you want to modify.
3. Select **+ Add shares**.
4. Select the protocol for the file shares you are adding.

   If you're adding file shares to a protocol you've already configured, no changes are required.

   If you're adding file shares with a second protocol, ensure you've properly configured the authentication as detailed in the prerequisites.

5. Add the file shares you want to scan (one file share per line) using the format `<host_name>:/<share_path>`.
6. Select **Continue** to complete adding the file shares.

Remove a file share from scans

1. From the Data Classification menu, select **Configuration**.
2. Select the system you want to remove file shares from.
3. Select **Configuration**.
4. From the Configuration page, select the Actions ••• for the file share you want to remove.
5. From the Actions menu, select **Remove Share**.

Track the scanning progress

You can track the progress of the initial scan.

1. Select the **Configuration** menu.
2. Select the **System Configuration**.
3. For the storage repository, check the Scan progress column to view its status.

Scan StorageGRID data with NetApp Data Classification

Complete a few steps to start scanning data within StorageGRID directly with NetApp Data Classification.

Review StorageGRID requirements

Review the following prerequisites to make sure that you have a supported configuration before you enable Data Classification.

- You need to have the endpoint URL to connect with the object storage service.
- You need to have the Access Key and Secret Key from StorageGRID so that Data Classification can access the buckets.

Deploy the Data Classification instance

Deploy Data Classification if there isn't already an instance deployed.

If you are scanning data from StorageGRID that is accessible over the internet, you can deploy Data Classification in the cloud or deploy Data Classification in an on-premises location that has internet access.

If you are scanning data from StorageGRID that has been installed in a dark site that has no internet access, you need to deploy Data Classification in the same on-premises location that has no internet access. This also requires that the Console agent is deployed in that same on-premises location.

Add the StorageGRID service to Data Classification

Add the StorageGRID service.

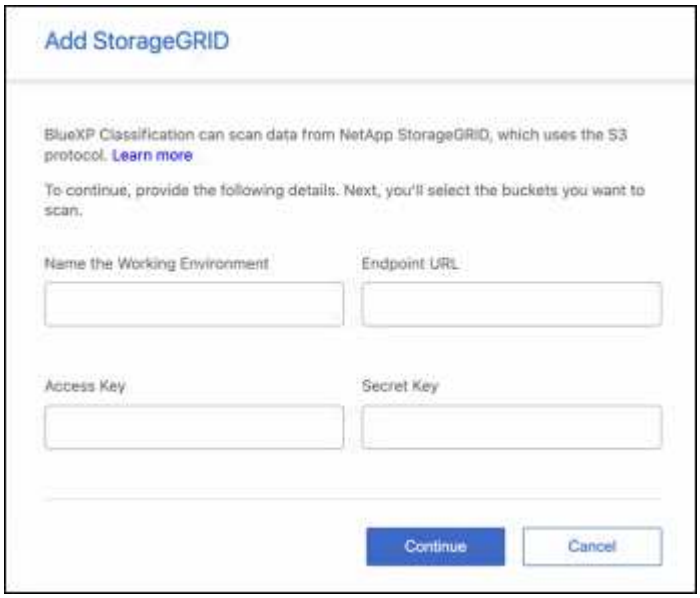
Steps

1. From the Data Classification menu, select the **Configuration** option.
2. From the Configuration page, select **Add System** > **Add StorageGRID**.
3. In the Add StorageGRID Service dialog, enter the details for the StorageGRID service and select **Continue**.

a. Enter the name you want to use for the System. This name should reflect the name of the StorageGRID service to which you are connecting.

b. Enter the Endpoint URL to access the object storage service.

c. Enter the Access Key and Secret Key so that Data Classification can access the buckets in StorageGRID.



Result

StorageGRID is added to the list of systems.

Enable and disable scans on StorageGRID buckets

After you enable Data Classification on StorageGRID, the next step is to configure the buckets that you want to scan. Data Classification discovers those buckets and displays them in the system you created.

Steps

1. In the Configuration page, locate the StorageGRID system.
2. On the StorageGRID system tile, select **Configuration**.
3. Complete one of the following steps to enable or disable scanning:
   - To enable mapping-only scans on a bucket, select **Map**.
   - To enable full scans on a bucket, select **Map & Classify**.
   - To disable scanning on a bucket, select **Off**.

Result

Data Classification starts scanning the buckets that you enabled. You can track the progress of the initial scan by navigating to the **Configuration** menu then selecting the **System configuration**. The progress of each scan is show as a progress bar. You can also hover over the progress bar to see the number of files scanned relative to the total files in the volume. If there are any errors, they'll appear in the Status column, alongside the required action to fix the error.

Copyright information

Trademark information