Manage Data Classification

NetApp Data Classification

NetApp
January 14, 2026

Table of Contents

Manage Data Classification . . . 1
- Exclude specific directories from NetApp Data Classification scans . . . 1
  - Supported data sources . . . 1
  - Define the directories to exclude from scanning . . . 1
  - Examples . . . 2
  - Escaping special characters in folder names . . . 3
  - View the current exclusion list . . . 4
- Define additional group IDs as open to organization in NetApp Data Classification . . . 4
  - Add the "open to organization" permission to group IDs . . . 4
  - View the current list of group IDs . . . 5
- Customize the stale data definition in NetApp Data Classification . . . 5
- Remove data sources from NetApp Data Classification . . . 6
  - Deactivate scans for a system . . . 6
  - Remove a database from Data Classification . . . 6
  - Remove a group of file shares from Data Classification . . . 6
- Uninstall NetApp Data Classification . . . 7
  - Uninstall Data Classification from a cloud provider . . . 7
  - Uninstall Data Classification from an on-premises deployment . . . 8

Manage Data Classification

Exclude specific directories from NetApp Data Classification scans

If you want NetApp Data Classification to exclude specific directories from scans, you can add these directory names to a configuration file. After you apply this change, the Data Classification engine excludes those directories from scans.

By default, Data Classification scans excludes volume snapshot data, which is identical to its source in the volume.

Supported data sources

Excluding specific directories from Data Classification scans is supported for NFS and CIFS shares in the following data sources:

- On-premises ONTAP
- Cloud Volumes ONTAP
- Amazon FSx for NetApp ONTAP
- Azure NetApp Files
- General file shares

Define the directories to exclude from scanning

Before you can exclude directories from classification scanning, you need to log into the Data Classification system so you can edit a configuration file and run a script. See how to log in to the Data Classification system depending on whether you manually installed the software on a Linux machine or if you deployed the instance in the cloud.

Considerations

- You can exclude a maximum of 50 directory paths per Data Classification system.
- Excluding directory paths can affect scanning times.

Steps

1. On the Data Classification system, go to "/opt/netapp/config/custom_configuration" then open the file `data_provider.yaml`.
2. In the "data_providers" section under the line "exclude:", enter the directory paths to exclude. For example:

```
exclude:
- "folder1"
- "folder2"
```

Do not modify anything else in this file.

3. Save the changes to the file.
4. Go to "/opt/netapp/Datasense/tools/customer_configuration/data_providers" and run the following script:

```
update_data_providers_from_config_file.sh
```

+
This command commits the directories to be excluded from scanning to the classification engine.

Result

All subsequent scans of your data will exclude scanning of those specified directories.

You can add, edit, or delete items from the exclude list using these same steps. The revised exclude list will be updated after you run the script to commit your changes.

Examples

Configuration 1:

Every folder that contains "folder1" anywhere in the name will be excluded from all data sources.

```
data_providers:
   exclude:
   - "folder1"
```

Expected results for paths that will be excluded:

- /CVO1/folder1
- /CVO1/folder1name
- /CVO1/folder10
- /CVO1/*folder1
- /CVO1/+folder1name
- /CVO1/notfolder10
- /CVO22/folder1
- /CVO22/folder1name
- /CVO22/folder10

Examples for paths that will not be excluded:

- /CVO1/*folder
- /CVO1/foldername
- /CVO22/*folder20

Configuration 2:

Every folder that contains "*folder1" only at the start of the name will be excluded.

```
data_providers:
   exclude:
   - "\\*folder1"
```

Expected results for paths that will be excluded:

- /CVO/*folder1
- /CVO/*folder1name
- /CVO/*folder10

Examples for paths that will not be excluded:

- /CVO/folder1
- /CVO/folder1name
- /CVO/not*folder10

Configuration 3:

Every folder in data source "CVO22" that contains "folder1" anywhere in the name will be excluded.

```
data_providers:
   exclude:
   - "CVO22/folder1"
```

Expected results for paths that will be excluded:

- /CVO22/folder1
- /CVO22/folder1name
- /CVO22/folder10

Examples for paths that will not be excluded:

- /CVO1/folder1
- /CVO1/folder1name
- /CVO1/folder10

Escaping special characters in folder names

If you have a folder name that contains one of the following special characters and you want to exclude data in that folder from being scanned, you'll need to use the escape sequence \\ before the folder name.

```
., +, *, ?, ^, $, (, ), [, ], {, }, |
```

For example:

Path in source: `/project/*not_to_scan`

Syntax in exclude file: `"\\*not_to_scan"`

View the current exclusion list

It's possible for the contents of the `data_provider.yaml` configuration file to be different than what has actually been committed after running the `update_data_providers_from_config_file.sh` script. To view the current list of directories that you've excluded from Data Classification scanning, run the following command from "/opt/netapp/Datasense/tools/customer_configuration/data_providers":

```
get_data_providers_configuration.sh
```

Define additional group IDs as open to organization in NetApp Data Classification

When group IDs (GIDs) are attached to files or folders in NFS file shares they define the permissions for the file or folder; such as whether they are "open to the organization". If some GIDs are not initially set up with the "Open to Organization" permission level, you can add that permission to the GID so that any files and folders that have that GID attached will be deemed "open to the organization".

After you make this change and NetApp Data Classification rescans your files and folders, any files and folders that have these group IDs attached will show this permission in the Investigation Details page, and they'll also appear in reports where you are displaying file permissions.

To activate this functionality, you need to log into the Data Classification system so you can edit a configuration file and run a script. See how to log in to the Data Classification system depending on whether you manually installed the software on a Linux machine or if you deployed the instance in the cloud.

Add the "open to organization" permission to group IDs

You need to have the group ID numbers (GIDs) before starting this task.

Steps

1. On the Data Classification system, go to "/opt/netapp/config/custom_configuration" and open the file `data_provider.yaml`.
2. In the line "organization_group_ids: []" add the group IDs. For example:

   ```
   organization_group_ids: [1014, 1015, 21, 2021, 1013, 2020, 1018, 1019]
   ```

   Do not change anything else in this file.

3. Save the changes to the file.
4. Go to "/opt/netapp/Datasense/tools/customer_configuration/data_providers" and run the following script:

   ```
   update_data_providers_from_config_file.sh
   ```

   This command commits the revised group ID permissions to the classification engine.

Result

All subsequent scans of your data will identify files or folders that have these group IDs attached as "open to organization."

You can edit the list of group IDs and delete any group IDs you added in the past using these same steps. The revised list of group IDs will be updated after you run the script to commit your changes.

View the current list of group IDs

It's possible for the contents of the `data_provider.yaml` configuration file to differ from what has actually been committed after running the `update_data_providers_from_config_file.sh` script. To view the current list of group IDs that you've added to Data Classification, run the following command from "/opt/netapp/Datasense/tools/customer_configuration/data_providers":

```
get_data_providers_configuration.sh
```

Customize the stale data definition in NetApp Data Classification

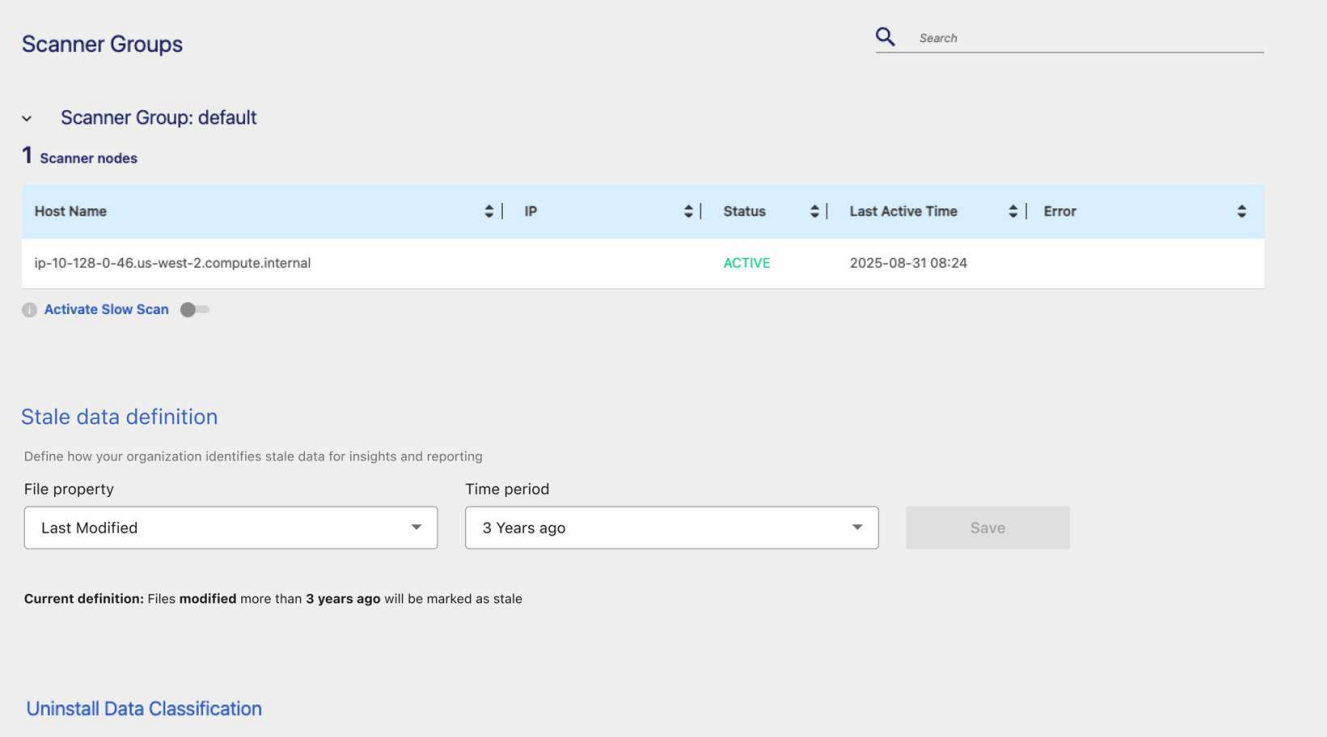
NetApp Data Classification identifies stale data to help you identify saving opportunities and governance risks. Because the definition of stale data can vary across different organizational contexts, you can customize how Data Classification defines stale data.

Stale data can be defined based on when it was *last accessed* or *last modified*. The time period selections range from 6 months ago to 10 years ago.

By default, data is considered stale if it was last modified three years ago.

Define stale data

1. In Ransomware Resilience, select **Configuration**.
2. In the Configuration page, scroll to the **Stale data definition** heading.
3. In the **File properties** dropdown menu, choose if you want to define stale data based on when it was **Last accessed** or **Last modified**.
4. Choose the time period for the stale data definition.



5. Select **Save**.

Remove data sources from NetApp Data Classification

If you need to, you can stop NetApp Data Classification from scanning one or more systems, databases, or file share groups.

Deactivate scans for a system

When you deactivate scans, Data Classification no longer scans the data on the system and it removes the indexed insights from the Data Classification instance. The data from the system itself isn't deleted.

1. From the *Configuration* page, select the button in the row for the system then **Deactivate Data Classification**.

You can also disable scans for a system from the Services panel when you select the system.

Remove a database from Data Classification

If you no longer need to scan a certain database, you can delete it from the Data Classification interface and stop all scans.

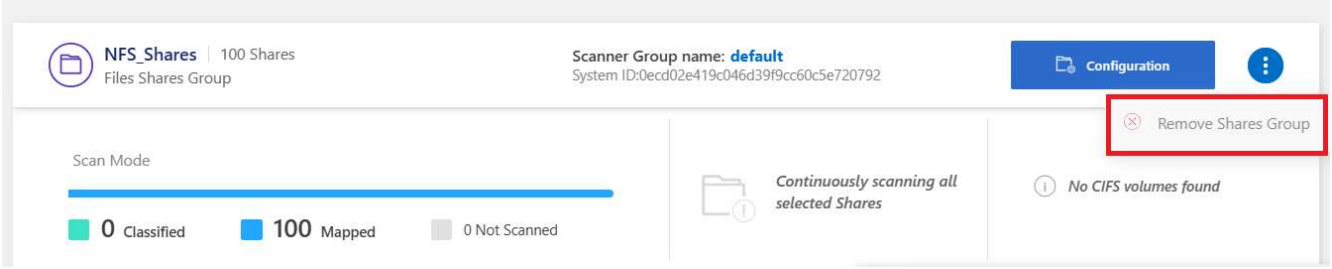
1. From the *Configuration* page, select the button in the row for the database then **Remove DB Server**.

Remove a group of file shares from Data Classification

If you no longer want to scan user files from a file share group, you can delete the File Shares Group from the Data Classification interface and stop all scans.

Steps

1. From the *Configuration* page, select the button in the row for the File Shares Group then **Remove File Shares Group**.



2. Select **Delete Group of Shares** from the confirmation dialog.

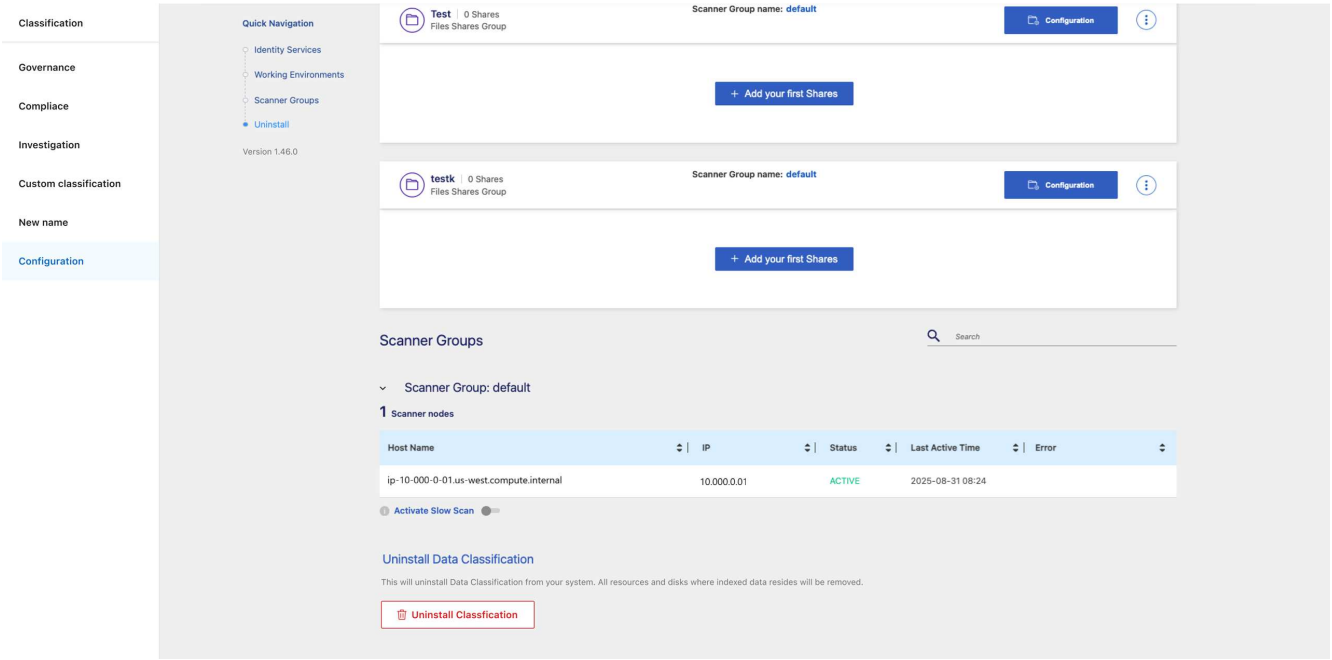
Uninstall NetApp Data Classification

You can uninstall NetApp Data Classification to troubleshoot issues or to permanently remove the software from the host. Deleting the instance also deletes the associated disks where the indexed data resides, meaning all the information Data Classification has scanned will be permanently deleted.

The steps that you need to use depend on whether you deployed Data Classification in the cloud or on an on-premises host.

Uninstall Data Classification from a cloud provider

1. From Data Classification, select **Configuration**.
2. At the bottom of the configuration page, select **Uninstall Classification**.



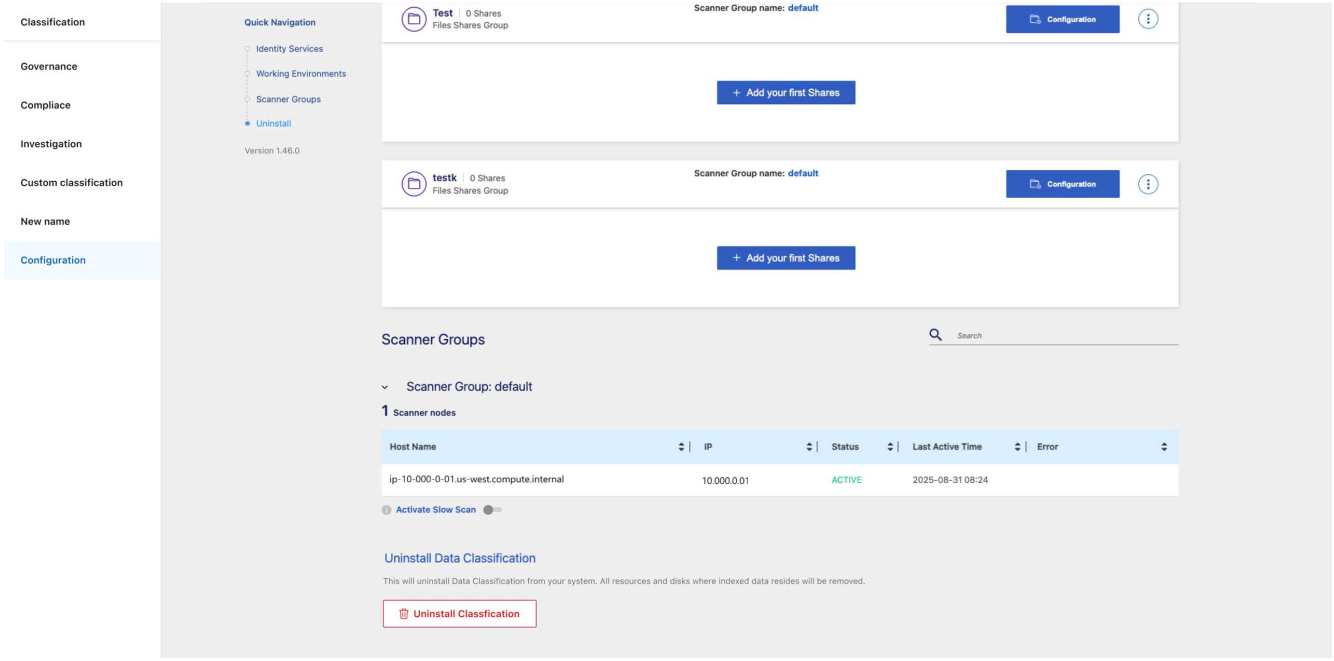
3. In the dialog, enter "uninstall" to proceed with disconnecting the Data Clasification instance from the

Console agent. Select **Uninstall** to confirm.

4. In the *Uninstall Classification* dialog, type **uninstall** to confirm that you want to disconnect the Data Classification instance from the Console agent then select **Uninstall**.
5. To finalize the uninstall process, go to your cloud provider's console and delete the Data Classification instance. The instance is named *CloudCompliance* with a generated hash (UUID) concatenated to it. For example: *CloudCompliance-16bb6564-38ad-4080-9a92-36f5fd2f71c7*

Uninstall Data Classification from an on-premises deployment

1. From Data Classification, select **Configuration**.
2. At the bottom of the configuration page, select **Uninstall Classification**.



3. In the dialog, enter "uninstall" to proceed with disconnecting the Data Clasification instance from the Console agent. Select **Uninstall** to confirm.
4. To uninstall the software from the host, run the `cleanup.sh` script on the Data Classification host machine, for example:

```
cleanup.sh
```

The script is located in the `/install/light_probe/onprem_installer/cleanup.sh` directory. See how to log in to the Data Classification host machine.

Copyright information

Trademark information