NetApp Data Classification documentation

NetApp Data Classification

NetApp
February 06, 2026

Table of Contents

NetApp Data Classification documentation

Release notes

What's new in NetApp Data Classification

Learn what's new in NetApp Data Classification.

14 January 2026

Version 1.50

This release of Data Classification includes bug fixes and the following updates:

Custom classification improvements

Data Classification now supports creating custom categories for your data. You can upload files to fine-tune an AI model that Data Classification uses to apply the category marker to data. The interface for all custom classifications has been improved.

For more information, see Create a custom classification.

Custom stale data definition

Data Classification now allows you to customize the definition of stale data so it suits your organizational needs. Previously, stale data was defined as any data that was last modified three years ago. Now, stale data can be identified based on when it was last accessed *or* last modified; the time period can range from 6 months ago to 10 years ago.

For more information, see Customize the stale data definition.

Improved performance

Loading times for all pages in Data Classification, the data mapping report, and filters on the Investigation page have been shortened.

Estimated time for investigation reports

When you download an investigation report, Data Classification now displays the estimated time for the download to complete.

08 December 2025

Version 1.49

This release of Data Classification includes bug fixes and the following updates:

Monitor metrics and performance in the Health Monitoring dashboard

Data Classification now provides a health monitoring dashboard, providing real-time monitoring of your resources and insights into memory usage, disk usage, disk utilization, and more. With insights from the health monitoring dashboard, you can review the infrastructure of your deployment and gain insights to optimize storage and performance.

For more information, see Monitor the health of Data Classification.

Improved loading performance

The loading performance for all pages in Data Classification has been improved to create a more efficient user

experience.

10 November 2025

Version 1.48

This release of Data Classification includes bug fixes, security improvements, and performance enhancements.

Enhanced scan progress clarity

Scan configurations now include improved insights into scan completion. Previously, a progress bar only displayed while the scan was in progress. Now, the progress bar remains visible after completion to confirm scans were completed successfully. You're also able to view the number of files mapped and scanned.

For more information about scan settings, see Change the NetApp Data Classification scan settings for your repositories.

06 October 2025

Version 1.47

BlueXP classification is now NetApp Data Classification

BlueXP classification has been renamed NetApp Data Classification. In addition to the rename, the user interface has been enhanced.

BlueXP is now NetApp Console

BlueXP has been renamed and redesigned to better reflect its role in managing your data infrastructure.

The NetApp Console provides centralized management of storage and data services across on-premises and cloud environments at enterprise grade—delivering real-time insights, faster workflows, and simplified administration.

For details on what has changed, see the NetApp Console release notes.

Enhanced Investigation experience

Find and understand your data faster with new searchable filters, per-value result counts, real-time insights summarizing key findings, and a refreshed results table with customizable columns and a slide-out details pane.

For more information, see Investigate data.

New Governance & Compliance dashboards

Gain critical insights faster with intuitive widgets, clearer visuals, and improved loading performance. For more information, see Review governance information about your data and View compliance information about your data.

Policies for saved queries (preview)

Data Classification now enables you to automate governance with conditional actions. You can create retention rules with automatic deletion set up periodic email notifications, all managed from an updated saved queries page.

For more information, see Create policies.

Actions (preview)

Take direct control from the Investigation page - delete, move, copy, or tag files individually or in bulk, for efficient data management and remediation.

For more information, see Investigate data.

Support for Google Cloud NetApp Volumes

Data Classification now supports scanning on Google Cloud NetApp Volumes. Easily add Google Cloud NetApp Volumes from the NetApp Console for seamless data scanning and classification. For more information, see Scan Google Cloud NetApp Volumes.

11 August 2025

Version 1.46

This Data Classification release includes bug fixes and the following updates:

Enhanced scan event insights in the audit page

The Audit page now supports enhanced insights into scan events for BlueXP classification. The Audit page now displays when the scan of a system begins, statuses of systems, and any issues. Statuses for shares and systems are only available for mapping scans.

For more information about the Audit page, see Monitor NetApp Console operations.

Support for RHEL 9.6

This release adds support for Red Hat Enterprise Linux v9.6 for manual on-premises installation of BlueXP classification, including dark site deployments.

The following operating systems require using the Podman container engine, and they require BlueXP classification version 1.30 or greater: Red Hat Enterprise Linux version 8.8, 8.10, 9.0, 9.1, 9.2, 9.3, 9.4, and 9.5.

14 July 2025

Version 1.45

This BlueXP classification release includes code changes that optimize resource utilization and:

Improved workflow to add file shares for scanning

The workflow to add files shares to a file share group has been simplified. The process also now differentiates CIFS protocol support based on authentication type (Kerberos or NTLM).

For more information, see Scan file shares.

Enhanced file owner information

You can now view more information about file owners for files captured in the Investigation tab. When viewing metadata for a file in the Investigation tab, locate the file owner then select **View details** to see the username, email, and SAM account name. You can also view other items owned by this user. This feature is only available for working environments with Active Directory.

For more information, see Investigate the data stored in your organization.

10 June 2025

Version 1.44

This BlueXP classification release includes:

Improved update times for the Governance dashboard

Update times for individual components of the Governance dashboard have been improved. The following table displays the frequency of updates for each component.

Component	Update times
Age of Data	24 hours
Categories	24 hours
Data Overview	5 minutes
Duplicate Files	2 hours
File Types	24 hours
Non-Business Data	2 hours
Open Permissions	24 hours
Saved Searches	2 hours
Sensitive Data and Wide Permissions	24 hours
Size of Data	24 hours
Stale Data	2 hours
Top Data Repositories by Sensitivity Level	2 hours

You can view the time of the last update and manually update the Duplicate Files, Non-Business Data, Saved Searches, Stale Data, and Top Data Repositories by Sensitivity Level components. For more information about the Governance dashboard, see View governance details about the data stored in your organization.

Performance and security improvements

Enhancements have been made to improve BlueXP classification's performance, memory consumption, and security.

Bug fixes

Redis has been upgraded to improve the reliability of BlueXP classification. BlueXP classification now uses Elasticsearch to improve the accuracy of file count reporting during scans.

12 May 2025

Version 1.43

This BlueXP Classification release includes:

Prioritize classification scans

Data Classification supports the ability to prioritize Map & Classify scans in addition to Mapping-only scans, enabling you to select which scans are completed first. Prioritization of Map & Classify scans is supported during and before the scans begin. If you choose to prioritize a scan while it's in progress, both the mapping

and classification scans are prioritized.

For more information, see Prioritize scans.

Support for Canadian personally identifiable information (PII) data categories

Data Classification scans identify Canadian PII data categories. These categories include banking information, passport numbers, social insurance numbers, driver's license numbers and health card numbers for all Canadian provinces and territories.

For more information, see Personal data categories.

Custom classification (preview)

Data Classification supports custom classifications for Map & Classify scans. With custom classifications, you can tailor Data Classification scans to capture data specific to your organization using regular expressions. This feature is currently in preview.

For more information, see Add custom classifications.

Saved searches tab

The **Policies** tab has been renamed **Saved searches**. The functionality is unchanged.

Send scan events to the Audit page

Data Classification supports sending classification events (when a scan is initiated and when it ends) to the NetApp Consle Audit page.

Security updates

- The Keras package has been updated, mitigating vulnerabilities (BDSA-2025-0107 and BDSA-2025-1984).
- The Docker containers configuration has been updated. The container no longer has access to the host's network interfaces for crafting raw network packets. By reducing unnecessary access, the update mitigates potential security risks.

Performance enhancements

Code enhancements have been implemented to reduce RAM usage and improve the overall performance of Data Classification.

Bug fixes

Bugs that caused StorageGRID scans to fail, the investigation page filter options to not load, and the Data Discovery Assessment to not download for high volume assessments have been fixed.

14 April 2025

Version 1.42

This BlueXP classification release includes:

Bulk scanning for working environments

BlueXP classification supports bulk operations for working environments. You can choose to enable Mapping scans, enable Map & Classify scans, disable scans, or create a custom configuration across volumes in working enviroment. If you make a selection for an individual volume, it overrides the bulk selection. To perform a bulk operation, navigate to the **Configuration** page and make your selection.

Download investigation report locally

BlueXP classification supports the ability to download data investigation reports locally to view in the browser. If you choose the local option, the data investigation is only available in the CSV format and only displays the first 10,000 rows of data.

For more information, see Investigate the data stored in your organization with BlueXP classification.

10 March 2025

Version 1.41

This BlueXP classification release includes general improvements and bug fixes. It also includes:

Scan status

BlueXP classification tracks the real time progress of the *initial* mapping and classification scans on a volume. Separate progressive bars track the mapping and classification scans, presenting a percentage of total files scanned. You can also hover over a progress bar to view the number of files scanned and the total files. Tracking the status of your scans creates deeper insights into the scan progress, enabling you to better plan your scans and understand resource allocation.

To view the status of your scans, navigate to **Configuration** in BlueXP classification then select the **Working Environment configuration**. Progress is displayed in line for each volume.

19 February 2025

Version 1.40

This BlueXP classification release includes the following updates.

Support for RHEL 9.5

This release provides support for Red Hat Enterprise Linux v9.5 in addition to previously supported versions. This is applicable to any manual on-premises installation of BlueXP classification, including dark site deployments.

The following operating systems require using the Podman container engine, and they require BlueXP classification version 1.30 or greater: Red Hat Enterprise Linux version 8.8, 8.10, 9.0, 9.1, 9.2, 9.3, 9.4, and 9.5.

Prioritize mapping-only scans

When conducting Mapping-only scans, you can prioritize the most important scans. This feature helps when you have many working environments and want to ensure high priority scans are completed first.

By default, scans are queued based on the order in which they are initiated. With the ability to prioritize scans, you can move scans to the front of the queue. Multiple scans can be prioritized. Priority is designated in a first-in, first-out order, meaning the first scan you prioritize moves to the front of the queue; the second scan you prioritize becomes second in the queue, and so forth.

Priority is granted on a one-time basis. Automatic rescans of mapping data occur in the default order.

Prioritization is limited to mapping-only scans; it's not available for map and classify scans.

For more information, see Prioritize scans.

Retry all scans

BlueXP classification supports the ability to batch retry all failed scans.

You can reattempt scans in a batch operation with the **Retry all** function. If classification scans are failing due to a temporary issue such as a network outage, you can retry all scans at the same time with one button instead of retrying them individually. Scans can be retried as many times as needed.

To retry all scans:

1. From the BlueXP classification menu, select **Configuration**.
2. To retry all failed scans, select **Retry all scans**.

Improved categorization model accuracy

The accuracy of the machine learning model for predefined categories has improved by 11%.

22 January 2025

Version 1.39

This BlueXP classification release updates the export process for the Data Investigation report. This export update is useful for performing additional analyses on your data, creating additional visualizations on the data, or sharing the results of your data investigation with others.

Previously, the Data Investigation report export was limited to 10,000 rows. With this release, the limit has been removed so that you can export all of your data. This change enables you to export more data from your Data Investigation reports, providing you with more flexibility in your data analysis.

You can choose the working environment, volumes, destination folder, and either JSON or CSV format. The exported filename includes a timestamp to help you identify when the data was exported.

The supported working environments include:

- Cloud Volumes ONTAP
- FSx for ONTAP
- ONTAP
- Share group

Exporting data from the Data Investigation report has the following limitations:

- The maximum number of records to download is 500 million. per type (files, directories, and tables)
- One million records are expected to take about 35 minutes to export.

For details about data investigation and the report, see Investigate data stored in your organization.

16 December 2024

Version 1.38

This BlueXP classification release includes general improvements and bug fixes.

4 November 2024

Version 1.37

This BlueXP classification release includes the following updates.

Support for RHEL 8.10

This release provides support for Red Hat Enterprise Linux v8.10 in addition to previously supported versions. This is applicable to any manual on-premises installation of BlueXP classification, including dark site deployments.

The following operating systems require using the Podman container engine, and they require BlueXP classification version 1.30 or greater: Red Hat Enterprise Linux version 8.8, 8.10, 9.0, 9.1, 9.2, 9.3, and 9.4.

Learn more about BlueXP classification.

Support for NFS v4.1

This release provides support for NFS v4.1 in addition to previously supported versions.

Learn more about BlueXP classification.

10 October 2024

Version 1.36

Support for RHEL 9.4

This release provides support for Red Hat Enterprise Linux v9.4 in addition to previously supported versions. This is applicable to any manual on-premises installation of BlueXP classification, including dark site deployments.

The following operating systems require using the Podman container engine, and they require BlueXP classification version 1.30 or greater: Red Hat Enterprise Linux version 8.8, 9.0, 9.1, 9.2, 9.3, and 9.4.

Learn more about BlueXP classification deployments overview.

Improved scan performance

This release provides improved scan performance.

2 September 2024

Version 1.35

Scan StorageGRID data

BlueXP classification supports scanning data in StorageGRID.

For details, refer to Scan StorageGRID data.

05 August 2024

Version 1.34

This BlueXP classification release includes the following update.

Change from CentOS to Ubuntu

BlueXP classification has updated its Linux operating system for Microsoft Azure and Google Cloud Platform

(GCP) from CentOS 7.9 to Ubuntu 22.04.

For deployment details, refer to Install on a Linux host with internet access and prepare the Linux host system.

01 July 2024

Version 1.33

Ubuntu supported
This release supports the Ubuntu 24.04 Linux platform.

Mapping scans gather metadata
The following metadata is extracted from files during mapping scans and is displayed on the Governance, Compliance, and Investigation dashboards:

- Working environment
- Working environment 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

Additional data in dashboards
This release updates which data appears in the Governance, Compliance, and Investigation dashboards during mapping scans.

For details, see What's the difference between mapping and classification scans.

05 June 2024

Version 1.32

New Mapping status column in the Configuration page
This release now shows a new Mapping status column in the Configuration page. The new column helps you identify if the mapping is running, queued, paused or more.

For explanations of the statuses, see Change scan settings.

15 May 2024

Version 1.31

Classification is available as a core service within BlueXP

BlueXP classification is now available as a core capability within BlueXP at no additional charge for up to 500 TiB of scanned data per connector. No Classification license or paid subscription is required. As we focus BlueXP classification functionality on scanning NetApp storage systems with this new version, some legacy functionality will only be available to customers who had previously paid for a license. The use of those legacy features will expire when the paid contract reaches its end date.

> (i) 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.

01 April 2024

Version 1.30

Support added for RHEL v8.8 and v9.3 BlueXP classification

This release provides support for Red Hat Enterprise Linux v8.8 and v9.3 in addition to previously supported 9.x, which requires Podman, rather than the Docker engine. This is applicable to any manual on-premises installation of BlueXP classification.

The following operating systems require using the Podman container engine, and they require BlueXP classification version 1.30 or greater: Red Hat Enterprise Linux version 8.8, 9.0, 9.1, 9.2, and 9.3.

Learn more about BlueXP classification deployments overview.

BlueXP classification is supported if you install the Connector on a RHEL 8 or 9 host that resides on-premises. It's not supported if the RHEL 8 or 9 host resides in AWS, Azure, or Google Cloud.

Option to activate audit log collection removed

The option to activate audit log collection has been disabled.

Scan speed improved

Scan performance on secondary scanner nodes has been improved. You can add more scanner nodes if you need additional processing power for your scans. For details, refer to Install BlueXP classification on a host that has internet access.

Automatic upgrades

If you deployed BlueXP classification on a system with internet access, the system upgrades automatically. Previously, the upgrade occurred after a specific time elapsed since the last user activity. With this release, BlueXP classification upgrades automatically if the local time is between 1:00 AM and 5:00 AM. If the local time is outside of these hours, the upgrade occurs after a specific time elapses since the last user activity. For details, refer to Install on a Linux host with internet access.

If you deployed BlueXP classification without internet access, you'll need to upgrade manually. For details, refer to Install BlueXP classification on a Linux host with no internet access.

04 March 2024

Version 1.29

Now you can exclude scanning data that resides in certain data source directories

If you want BlueXP classification to exclude scanning data that resides in certain data source directories, you can add these directory names to a configuration file that BlueXP classification processes. This feature enables you to avoid scanning directories that are unnecessary, or that would result in returning false positive personal data results.

Learn more.

Extra Large instance support is now qualified

If you need BlueXP classification to scan more than 250 million files, you can use an Extra Large instance in your cloud deployment or on-premises installation. This type of system can scan up to 500 million files.

Learn more.

10 January 2024

Version 1.27

Investigation page results display the total size in addition to total number of items

The filtered results in the Investigation page display the total size of the items in addition to the total number of files. This can help when moving files, deleting files, and more.

Configure additional Group IDs as "Open to Organization"

Now you can configure Group IDs in NFS to be considered as "Open to Organization" directly from BlueXP classification if the group had not initially been set with that permission. Any files and folders that have these group IDs attached will show as "Open to Organization" in the Investigation Details page. See how to add additional Group IDs as "open to organization".

14 December 2023

Version 1.26.6

This release included some minor enhancements.

The release also removed the following options:

- The option to activate audit log collection has been disabled.
- During the Directories investigation, the option to calculate the number of personal identifiable information (PII) data by Directories is not available. Refer to Investigate the data stored in your organization.
- The option to integrate data using Azure Information Protection (AIP) labels has been disabled.

06 November 2023

Version 1.26.3

The following issues have been fixed in this release

- Fixed an inconsistency when presenting the number of files scanned by the system in dashboards.

- Improved the scanning behavior by handling and reporting on files and directories with special characters in the name and metadata.

04 October 2023

Version 1.26

Support for on-premises installations of BlueXP classification on RHEL version 9

Red Hat Enterprise Linux versions 8 and 9 do not support the Docker engine; which was required for the BlueXP classification installation. We now support BlueXP classification installation on RHEL 9.0, 9.1, and 9.2 using Podman version 4 or greater as the container infrastructure. If your environment requires using the newest versions of RHEL, now you can install BlueXP classification (version 1.26 or greater) when using Podman.

At this time we don't supported dark site installations or distributed scanning environments (using a master and remote scanner nodes) when using RHEL 9.x.

05 September 2023

Version 1.25

Small and medium deployments temporarily unavailable

When you deploy an instance of BlueXP classification in AWS, the option to select **Deploy > Configuration** and choose a small or medium-sized instance is unavailable at this time. You can still deploy the instance using the large instance size by selecting **Deploy > Deploy**.

Apply tags on up to 100,000 items from the Investigation Results page

In the past you could only apply tags to a single page at a time in the Investigation Results page (20 items). Now you can select **all** items in the Investigation Results pages and apply tags to all the items - up to 100,000 items at a time.

Identify duplicated files with a minimum file size of 1 MB

BlueXP classification used to identify duplicated files only when files were 50 MB or larger. Now duplicated files starting with 1 MB can be identified. You can use the Investigation page filters "File Size" along with "Duplicates" to see which files of a certain size are duplicated in your environment.

17 July 2023

Version 1.24

Two new types of German personal data are identified by BlueXP classification

BlueXP classification can identify and categorize files that contain the following types of data:

- German ID (Personalausweisnummer)
- German Social Security Number (Sozialversicherungsnummer)

See all the types of personal data that BlueXP classification can identify in your data.

BlueXP classification is fully supported in Restricted mode and Private mode

BlueXP classification is now fully supported in sites with no internet access (Private mode) and with limited outbound internet access (Restricted mode). Learn more about BlueXP deployment modes for the Connector.

Ability to skip versions when upgrading a Private mode installation of BlueXP classification

Now you can upgrade to a newer version of BlueXP classification even if it is not sequential. This means that the current limitation of upgrading BlueXP classification by one version at a time is no longer required. This feature is relevant starting from version 1.24 onwards.

The BlueXP classification API is now available

The BlueXP classification API enables you to perform actions, create queries, and export information about the data you are scanning. The interactive documentation is available using Swagger. The documentation is separated into multiple categories, including Investigation, Compliance, Governance, and Configuration. Each category is a reference to the tabs in the BlueXP classification UI.

Learn more about the BlueXP classification APIs.

06 June 2023

Version 1.23

Japanese is now supported when searching for data subject names

Japanese names can now be entered when searching for a subject's name in response to a Data Subject Access Request (DSAR). You can generate a Data Subject Access Request report with the resulting information. You can also enter Japanese names in the "Data Subject" filter in the Data Investigation page to identify files that contain the subject's name.

Ubuntu is now a supported Linux distribution on which you can install BlueXP classification

Ubuntu 22.04 has been qualified as a supported operating system for BlueXP classification. You can install BlueXP classification on a Ubuntu Linux host in your network, or on a Linux host in the cloud when using version 1.23 of the installer. See how to install BlueXP classification on a host with Ubuntu installed.

Red Hat Enterprise Linux 8.6 and 8.7 are no longer supported with new BlueXP classification installations

These versions are not supported with new deployments because Red Hat no longer supports Docker, which is a prerequisite. If you have an existing BlueXP classification machine running on RHEL 8.6 or 8.7, NetApp will continue to support your configuration.

BlueXP classification can be configured as an FPolicy Collector to receive FPolicy events from ONTAP systems

You can enable file access audit logs to be collected on your BlueXP classification system for file access events detected on volumes in your working environments. BlueXP classification can capture the following types of FPolicy events and the users who performed the actions on your files: Create, Read, Write, Delete, Rename, Change owner/permissions, and Change SACL/DACL.

Data Sense BYOL licenses are now supported in dark sites

Now you can upload your Data Sense BYOL license into the BlueXP digital wallet in a dark site so that you are notified when your license is getting low.

03 April 2023

Version 1.22

New Data Discovery Assessment Report

The Data Discovery Assessment Report provides a high-level analysis of your scanned environment to highlight the system's findings and to show areas of concern and potential remediation steps. The goal of this report is to raise awareness of data governance concerns, data security exposures, and data compliance gaps

of your data set. See how to generate and use the Data Discovery Assessment Report.

Ability to deploy BlueXP classification on smaller instances in the cloud

When deploying BlueXP classification from a BlueXP Connector in an AWS environment, now you can select from two smaller instance types than what is available with the default instance. If you are scanning a small environment this can help you save on cloud costs. However, there are some restrictions when using the smaller instance. See the available instance types and limitations.

Standalone script is now available to qualify your Linux system prior to BlueXP classification installation

If you would like to verify that your Linux system meets all prerequisites independently of running the BlueXP classification installation, there is a separate script you can download that only tests for the prerequisites. See how to check if your Linux host is ready to install BlueXP classification.

07 March 2023

Version 1.21

New functionality to add your own custom categories from the BlueXP classification UI

BlueXP classification now enables you to add your own custom categories so that BlueXP classification will identify the files that fit into those categories. BlueXP classification has many predefined categories, so this feature enables you to add custom categories to identify where information that is unique to your organization are found in your data.

Now you can add custom keywords from the BlueXP classification UI

BlueXP classification has had the ability to add custom keywords that BlueXP classification will identify in future scans for a while. However, you needed to log into the BlueXP classification Linux host and use a command line interface to add the keywords. In this release, the ability to add custom keywords is in the BlueXP classification UI, making it very easy to add and edit these keywords.

Ability to have BlueXP classification not scan files when the "last access time" will be changed

By default, if BlueXP classification doesn't have adequate "write" permissions, the system won't scan files in your volumes because BlueXP classification can't revert the "last access time" to the original timestamp. However, if you don't care if the last access time is reset to the original time in your files, you can override this behavior in the Configuration page so that BlueXP classification will scan the volumes regardless of permissions.

In conjunction with this capability, and new filter named "Scan Analysis Event" has been added so you can view the files that were not classified because BlueXP classification couldn't revert last accessed time, or the files that were classified even though BlueXP classification couldn't revert last accessed time.

Learn more about the "Last access time timestamp" and the permissions BlueXP classification requires.

Three new types of personal data are identified by BlueXP classification

BlueXP classification can identify and categorize files that contain the following types of data:

- Botswana Identity Card (Omang) Number
- Botswana Passport Number
- Singapore National Registration Identity Card (NRIC)

See all the types of personal data that BlueXP classification can identify in your data.

Updated functionality for directories

- The "Light CSV Report" option for Data Investigation Reports now includes information from directories.

- The "Last Accessed" time filter now shows the last accessed time for both files and directories.

Installation enhancements

- The BlueXP classification installer for sites without internet access (dark sites) now performs a pre-check to make sure your system and networking requirements are in place for a successful installation.

- Installation audit log files are saved now; they are written to `/ops/netapp/install_logs`.

05 February 2023

Version 1.20

Ability to send Policy-based notification emails to any email address

In earlier versions of BlueXP classification you could send email alerts to the BlueXP users in your account when certain critical Policies return results. This feature enables you to get notifications to protect your data when you're not online. Now you can also send email alerts from Policies to any other users - up to 20 email addresses - who are not in your BlueXP account.

Learn more about sending email alerts based on Policy results.

Now you can add personal patterns from the BlueXP classification UI

BlueXP classification has had the ability to add custom "personal data" that BlueXP classification will identify in future scans for a while. However, you needed to log into the BlueXP classification Linux host and use a command line to add the custom patterns. In this release, the ability to add personal patterns using a regex is in the BlueXP classification UI, making it very easy to add and edit these custom patterns.

Ability to move 15 million files using BlueXP classification

In the past you could have BlueXP classification move a maximum of 100,000 source files to any NFS share. Now you can move up to 15 million files at a time.

Ability to see the number of users who have access to SharePoint Online files

The filter "Number of users with access" now supports files stored in SharePoint Online repositories. In the past only files on CIFS shares were supported. Note that SharePoint groups that are not active directory based will not be counted in this filter at this time.

New "Partial Success" status has been added to the Action Status panel

The new "Partial Success" status indicates that a BlueXP classification action is finished and some items failed and some items succeeded, for example, when you are moving or deleting 100 files. Additionally, the "Finished" status has been renamed to "Success". In the past, the "Finished" status might list actions that succeeded and that failed. Now the "Success" status means that all actions succeeded on all items. See how to view the Actions Status panel.

09 January 2023

Version 1.19

Ability to view a chart of files that contain sensitive data and that are overly permissive

The Governance dashboard has added a new *Sensitive Data and Wide Permissions* area that provides a heatmap of files that contain sensitive data (including both sensitive and sensitive personal data) and that are overly permissive. This can help you to see where you may have some risks with sensitive data. Learn more.

Three new filters are available in the Data Investigation page

New filters are available to refine the results that display in the Data Investigation page:

- The "Number of users with access" filter shows which files and folders are open to a certain number of users. You can choose a number range to refine the results - for example, to see which files are accessible by 51-100 users.

- The "Created Time", "Discovered Time", "Last Modified", and "Last Accessed" filters now allow you to create a custom date range instead of just selecting a pre-defined range of days. For example, you can look for files with a "Created Time" "older than 6 months", or with a "Last Modified" date within the "last 10 days".

- The "File Path" filter now enables you to specify paths that you want to exclude from the filtered query results. If you enter paths to both include and exclude certain data, BlueXP classification finds all files in the included paths first, then it removes files from excluded paths, and then it displays the results.

See the list of all the filters you can use to investigate your data.

BlueXP classification can identify the Japanese Individual Number

BlueXP classification can identify and categorize files that contain the Japanese Individual Number (also known as My Number). This includes both the Personal and Corporate My Number. See all the types of personal data that BlueXP classification can identify in your data.

Known limitations in NetApp Data Classification

Known limitations identify functions that are not supported or do not interoperate correctly in this release. Review these limitations carefully.

NetApp Data Classification disabled options

The December 2023 (version 1.26.6) release removed the following options:

- The option to activate audit log collection has been disabled.
- During the Directories investigation, the option to calculate the number of personally identifiable information (PII) data by Directories is not available.
- The option to integrate data using Azure Information Protection (AIP) labels has been disabled.

Data Classification scanning

The following limitations occur with Data Classification scans.

Data Classification scans only one share under a volume

If you have multiple file shares under a single volume, Data Classification scans the share with the highest hierarchy. For example, if you have shares like the following:

- /A
- /A/B
- /C
- /D/E

In this configuration, only the data in /A is scanned. The data in /C and /D is not scanned.

Workaround

There is a workaround to make sure you are scanning data from all the shares in your volume. Follow these steps:

1. In the system, add the volume to be scanned.

2. After Data Classification has completed scanning the volume, go to the *Data Investigation* page and create a filter to see which share is being scanned:

   Filter the data by "system Name" and "Directory Type = Share" to see which share is being scanned.

3. Get the complete list of shares that exist in the volume so you can see which shares are not being scanned.

4. Add the remaining shares to a share group.

   Add all the shares individually, for example:

   ```
   /C
   /D
   ```

5. Perform these steps for each volume in the system that has multiple shares.

Last accessed timestamp

When Data Classification conducts a scan of a directory, the scan impacts the directory's **Last accessed** field. When you view the **Last accessed** field, that metadata reflects either the date and time of the scan or the last time a user accessed the directory.

Get started

Learn about NetApp Data Classification

NetApp Data Classification is a data governance service for the NetApp Console that scans your corporate on-premises and cloud data sources to map and classify data, and to identify private information. This can help reduce your security and compliance risk, decrease storage costs, and assist with your data migration projects.

> ⓘ Beginning with version 1.31, Data Classification is available as a core capability within the NetApp Console. There's no additional charge. No Classification license or subscription is required.
> If you've been using legacy version 1.30 or earlier, that version is available until your subscription expires.

NetApp Console

Data Classification is accessible through the NetApp Console.

The NetApp Console provides centralized management of NetApp storage and data services across on-premises and cloud environments at enterprise grade. The Console is required to access and use NetApp data services. As a management interface, it enables you to manage many storage resources from one interface. Console administrators can control access to storage and services for all systems within the enterprise.

You don't need a license or subscription to start using NetApp Console and you only incur charges when you need to deploy Console agents in your cloud to ensure connectivity to your storage systems or NetApp data services. However, some NetApp data services accessible from the Console are licensed or subscription-based.

Learn more about the NetApp Console.

Features

Data Classification uses artificial intelligence (AI), natural language processing (NLP), and machine learning (ML) to understand the content that it scans in order to extract entities and categorize the content accordingly. This allows Data Classification to provide the following areas of functionality.

Learn about use cases for Data Classification.

Maintain compliance

Data Classification provides several tools that can help with your compliance efforts. You can use Data Classification to:

- Identify Personal Identifiable Information (PII).
- Identify a wide scope of sensitive personal information as required by GDPR, CCPA, PCI, and HIPAA privacy regulations.
- Respond to Data Subject Access Requests (DSAR) based on name or email address.

Strengthen security

Data Classification can identify data that is potentially at risk for being accessed for criminal purposes. You can

use Data Classification to:

- Identify all the files and directories (shares and folders) with open permissions that are exposed to your entire organization or to the public.
- Identify sensitive data that resides outside of the initial, dedicated location.
- Comply with data retention policies.
- Use *Policies* to automatically detect new security issues so security staff can take action immediately.

Optimize storage usage

Data Classification provides tools that can help with your storage total cost of ownership (TCO). You can use Data Classification to:

- Increase storage efficiency by identifying duplicate or non-business-related data.
- Save storage costs by identifying inactive data that you can tier to less expensive object storage. Learn more about tiering from Cloud Volumes ONTAP systems. Learn more about tiering from on-premises ONTAP systems.

Supported systems and data sources

Data Classification can scan and analyze structured and unstructured data from the following types of systems and data sources:

Systems

- Amazon FSx for NetApp ONTAP management
- Azure NetApp Files
- Cloud Volumes ONTAP (deployed in AWS, Azure, or GCP)
- On-premises ONTAP clusters
- StorageGRID
- Google Cloud NetApp Volumes

Data sources

- NetApp file shares
- Databases:
  - Amazon Relational Database Service (Amazon RDS)
  - MongoDB
  - MySQL
  - Oracle
  - PostgreSQL
  - SAP HANA
  - SQL Server (MSSQL)

Data Classification supports NFS versions 3.x, 4.0, and 4.1, and CIFS versions 1.x, 2.0, 2.1, and 3.0.

Cost

Data Classification is free to use. No Classification license or paid subscription is required.

Infrastructure costs

- Installing Data Classification in the cloud requires deploying a cloud instance, which results in charges from the cloud provider where it is deployed. See the type of instance that is deployed for each cloud provider. There is no cost if you install Data Classification on an on-premises system.

- Data Classification requires that you have deployed a Console agent. In many cases you already have a Console agent because of other storage and services you are using in the Console. The Console agent instance results in charges from the cloud provider where it's deployed. See the type of instance that is deployed for each cloud provider. There is no cost if you install the Console agent on an on-premises system.

Data transfer costs

Data transfer costs depend on your setup. If the Data Classification instance and data source are in the same Availability Zone and region, then there are no data transfer costs. But if the data source, such as a Cloud Volumes ONTAP system, is in a *different* Availability Zone or region, then you'll be charged by your cloud provider for data transfer costs. See these links for more details:

- AWS: Amazon Elastic Compute Cloud (Amazon EC2) Pricing

- Microsoft Azure: Bandwidth Pricing Details

- Google Cloud: Storage Transfer Service pricing

The Data Classification instance

When you deploy Data Classification in the cloud, the Console deploys the instance in the same subnet as the Console agent. Learn more about the Console agent.
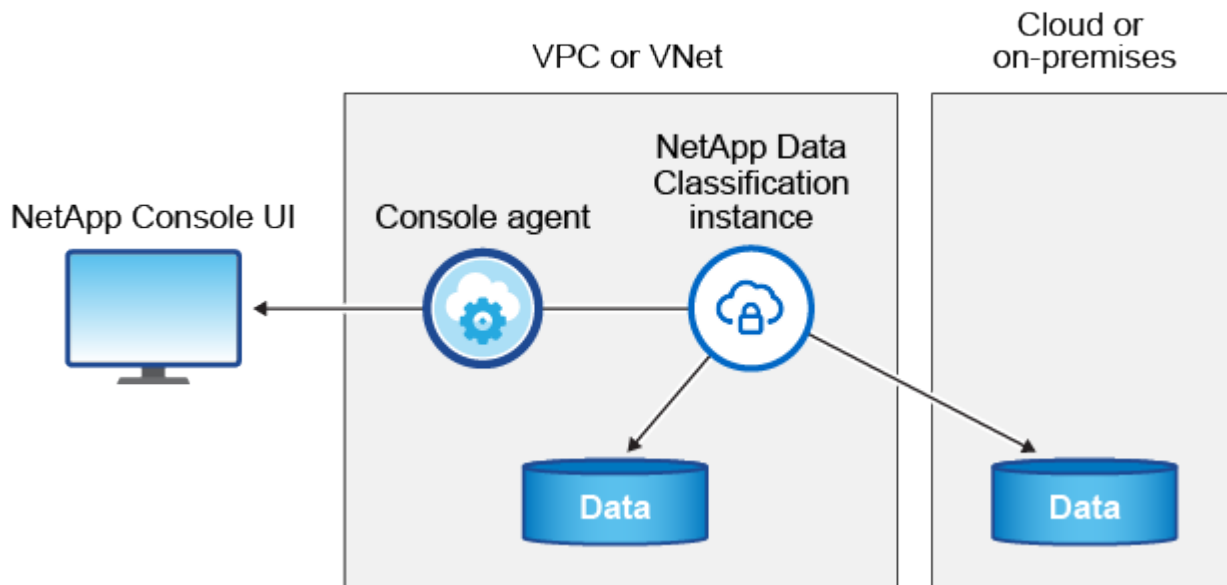
Note the following about the default instance:

- In AWS, Data Classification runs on an m6i.4xlarge instance with a 500 GiB GP2 disk. The operating

system image is Amazon Linux 2. When deployed in AWS, you can choose a smaller instance size if you are scanning a small amount of data.

- In Azure, Data Classification runs on a Standard_D16s_v3 VM with a 500 GiB disk. The operating system image is Ubuntu 22.04.

- In GCP, Data Classification runs on an n2-standard-16 VM with a 500 GiB Standard persistent disk. The operating system image is Ubuntu 22.04.

- In regions where the default instance isn't available, Data Classification runs on an alternate instance. See the alternate instance types.

- The instance is named *CloudCompliance* with a generated hash (UUID) concatenated to it. For example: *CloudCompliance-16bb6564-38ad-4080-9a92-36f5fd2f71c7*

- Only one Data Classification instance is deployed per Console Agent.

You can also deploy Data Classification on a Linux host on your premises or on a host in your preferred cloud provider. The software functions exactly the same way regardless of which installation method you choose. Upgrades of Data Classification software are automated as long as the instance has internet access.

> The instance should remain running at all times because Data Classification continuously scans the data.

Deploy on different instance types

Review the following specifications for instance types:

System size	Specs	Limitations
Extra Large	32 CPUs, 128 GB RAM, 1 TiB SSD	Can scan up to 500 million files.
Large (default)	16 CPUs, 64 GB RAM, 500 GiB SSD	Can scan up to 250 million files.

When deploying Data Classification in Azure or GCP, email ng-contact-data-sense@netapp.com for assistance if you want to use a smaller instance type.

How Data Classification scanning works

At a high-level, Data Classification scanning works like this:

1. You deploy an instance of Data Classification in the Console.

2. You enable high-level mapping (called *Mapping only* scans) or deep-level scanning (called *Map & Classify* scans) on one or more data sources.

3. Data Classification scans data using an AI learning process.

4. You use the provided dashboards and reporting tools to help in your compliance and governance efforts.

After you enable Data Classification and select the repositories that you want to scan (these are the volumes, database schemas, or other user data), it immediately starts scanning the data to identify personal and sensitive data. You should focus on scanning live production data in most cases instead of backups, mirrors, or DR sites. Then Data Classification maps your organizational data, categorizes each file, and identifies and extracts entities and predefined patterns in the data. The result of the scan is an index of personal information, sensitive personal information, data categories, and file types.

Data Classification connects to the data like any other client by mounting NFS and CIFS volumes. NFS volumes are automatically accessed as read-only, while you need to provide Active Directory credentials to scan CIFS volumes.
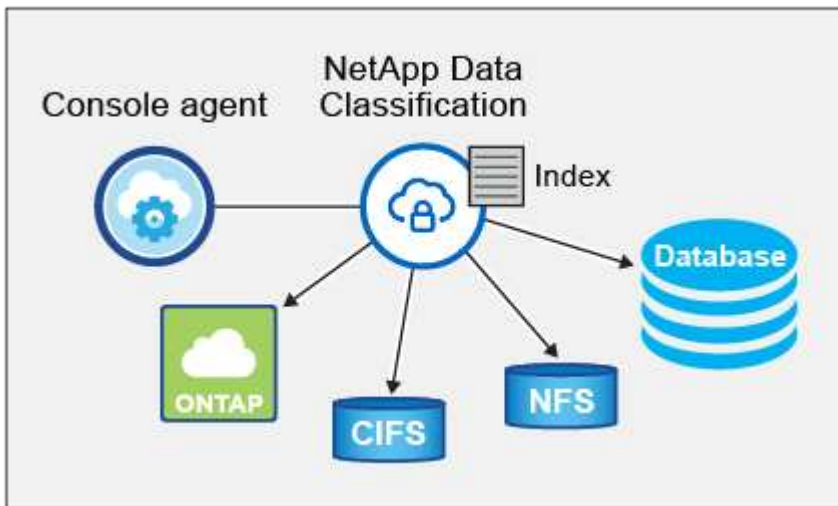
After the initial scan, Data Classification continuously scans your data in a round-robin fashion to detect incremental changes. This is why it's important to keep the instance running.

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

> ℹ️ 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.

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 they 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.

For details about the differences between Mapping and Classification scans, see What's the difference between Mapping and Classification scans?.

Information that Data Classification categorizes

Data Classification collects, indexes, and assigns categories to the following data:

- **Standard metadata** about files: the file type, its size, creation and modification dates, and so on.
- **Personal data**: Personally identifiable information (PII) such as email addresses, identification numbers, or credit card numbers, which Data Classification identifies using specific words, strings, and patterns in the files. Learn more about personal data.

- **Sensitive personal data**: Special types of sensitive personal information (SPII), such as health data, ethnic origin, or political opinions, as defined by General Data Protection Regulation (GDPR) and other privacy regulations. Learn more about sensitive personal data.

- **Categories**: Data Classification takes the data that it scanned and divides it into different types of categories. Categories are topics based on AI analysis of the content and metadata of each file. Learn more about categories.

- **Name entity recognition**: Data Classification uses AI to extract people's natural names from documents. Learn about responding to Data Subject Access Requests.

Networking overview

Data Classification deploys a single server, or cluster, wherever you choose: in the cloud or on premises. The servers connect via standard protocols to the data sources and index the findings in an Elasticsearch cluster, which is also deployed on the same servers. This enables support for multi-cloud, cross-cloud, private cloud, and on-premises environments.

The Console deploys the Data Classification instance with a security group that enables inbound HTTP connections from the Console agent.

When you use the Console in SaaS mode, the connection to the Console is served over HTTPS, and the private data sent between your browser and the Data Classification instance are secured with end-to-end encryption using TLS 1.2, which means NetApp and third parties can't read it.

Outbound rules are completely open. Internet access is needed to install and upgrade the Data Classification software and to send usage metrics.

If you have strict networking requirements, learn about the endpoints that Data Classification contacts.

Access NetApp Data Classification

You can access the NetApp Data Classification through the NetApp Console.

To sign in to the Console, you can use your NetApp Support Site credentials or you can sign up for a NetApp Console login using your email and a password. Learn more about logging in to the Console.

Specific tasks require specific Console user roles. Learn about Console access roles for all services.

Before you begin
- You should add a Console agent.
- Understand which Data Classification deployment style suits your workload.

Steps
1. In a web browser, navigate to the Console.
2. Log in to the Console.
3. From the main page of the NetApp Console, select **Governance** > **Data Classification**.
4. If this is your first time accessing Data Classification, the landing page appears.

   Select **Deploy Classification On-Premises or Cloud** to begin deploying your classification instance. For more information, see Which Data Classification deployment should you use?
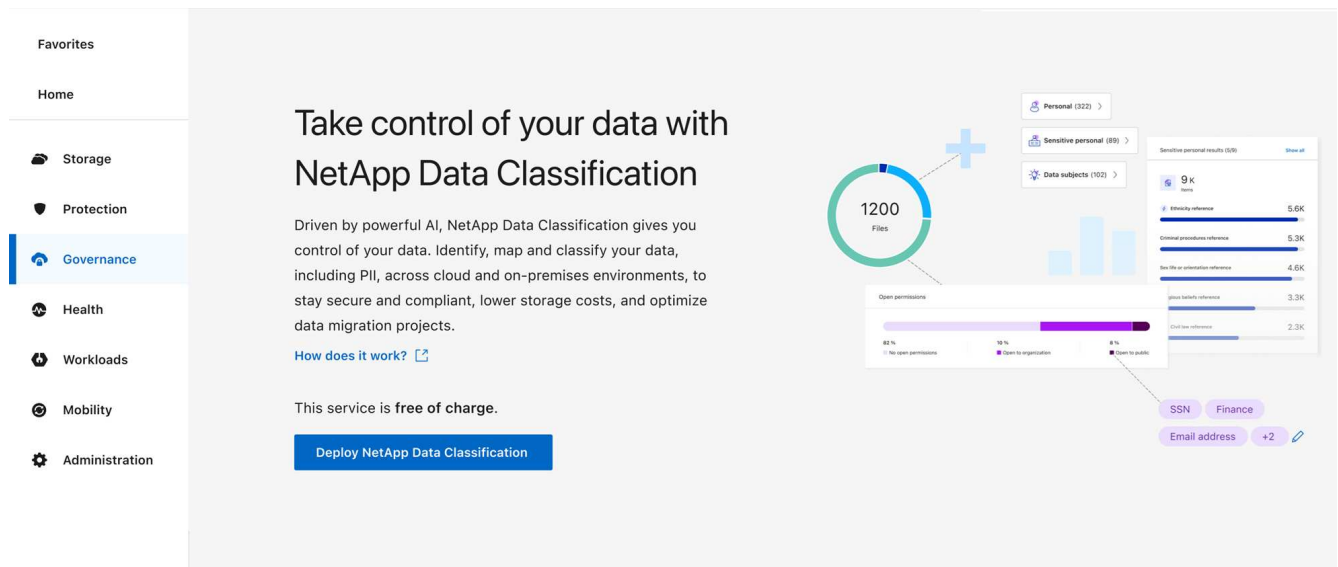
Otherwise, the Data Classification Dashboard appears.

Deploy Data Classification

Which NetApp Data Classification deployment should you use?

You can deploy NetApp Data Classification in different ways. Learn which method meets your needs.

Data Classification can be deployed in the following ways:

- Deploy in the cloud using the Console. The Console deploys the Data Classification instance in the same cloud provider network as the Console agent.

- Install on a Linux host with internet access. Install Data Classification on a Linux host in your network, or on a Linux host in the cloud, that has internet access. This type of installation may be a good option if you prefer to scan on-premises ONTAP systems using a Data Classification instance that's also located on premises, though this isn't a requirement.

- Install on a Linux host in an on-premises site without internet access, also known as *private mode.* This type of installation, which uses an installation script, has no connectivity to the Console SaaS layer.

> ℹ️ BlueXP private mode (legacy BlueXP interface) is typically used with on-premises environments that have no internet connection and with secure cloud regions, which includes AWS Secret Cloud, AWS Top Secret Cloud, and Azure IL6. NetApp continues to support these environments with the legacy BlueXP interface. For private mode documentation in the legacy BlueXP interface, see PDF documentation for BlueXP private mode.

Both the installation on a Linux host with internet access and the on-premises installation on a Linux host without internet access use an installation script. The script starts by checking if the system and environment meet the prerequisites. If the prerequisites are met, the installation starts. If you would like to verify the prerequisites independently of running the Data Classification installation, there is a separate software package you can download that only tests for the prerequisites.

Refer to Check that your Linux host is ready to install Data Classification.

Deploy NetApp Data Classification in the cloud using the NetApp Console

You can deploy NetApp Data Classification in the cloud with the NetApp Console. The Console deploys the Data Classification instance in the same cloud provider network as the Console agent.

Note that you can also install Data Classification on a Linux host that has internet access. This type of installation may be a good option if you prefer to scan on-premises ONTAP systems using a Data Classification instance that's also located on premises — but this is not a requirement. The software functions exactly the same way regardless of which installation method you choose.

Quick start

Get started quickly by following these steps, or scroll down to the remaining sections for full details.

1   **Create a Console agent**

If you don't already have a Console agent, create one. See creating a Console agent in AWS, creating a Console agent in Azure, or creating a Console agent in GCP.

You can also install the Console agent on-premises on a Linux host in your network or on a Linux host in the cloud.

2   **Prerequisites**

Ensure that your environment can meet the prerequisites. This includes outbound internet access for the instance, connectivity between the Console agent and Data Classification over port 443, and more. See the complete list.

3   **Deploy Data Classification**

Launch the installation wizard to deploy the Data Classification instance in the cloud.

Create a Console agent

If you don't already have a Console agent, create a Console agent in your cloud provider. See creating a Console agent in AWS or creating a Console agent in Azure, or creating a Console agent in GCP. In most cases you will probably have a Console agent set up before you attempt to activate Data Classification because most Console features require a Console agent, but there are cases where you'll you need to set one up now.

There are some scenarios where you have to use a Console agent that's deployed in a specific cloud provider:

- When scanning data in Cloud Volumes ONTAP in AWS or Amazon FSx for ONTAP buckets, you use a Console agent in AWS.
- When scanning data in Cloud Volumes ONTAP in Azure or in Azure NetApp Files, you use a Console agent in Azure.
  - For Azure NetApp Files, it must be deployed in the same region as the volumes you wish to scan.
- When scanning data in Cloud Volumes ONTAP in GCP, you use a Console agent in GCP.

On-prem ONTAP systems, NetApp file shares, and databases can be scanned when using any of these cloud Console agents.

Note that you can also install the Console agent on-premises on a Linux host in your network or in the cloud. Some users planning to install Data Classification on-prem may also choose to install the Console agent on-premises.

There might be situations where you need to use multiple Console agents.

> 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.

Government region support

Data Classification is supported when the Console agent is deployed in a Government region (AWS GovCloud, Azure Gov, or Azure DoD). When deployed in this manner, Data Classification has the following restrictions:

Learn about deploying the Console agent in a Government region.

Prerequisites

Review the following prerequisites to make sure that you have a supported configuration before you deploy Data Classification in the cloud. When you deploy Data Classification in the cloud, it's located in the same subnet as the Console agent.

Enable outbound internet access from Data Classification

Data Classification requires outbound internet access. If your virtual or physical network uses a proxy server for internet access, ensure that the Data Classification instance has outbound internet access to contact the following endpoints. The proxy must be non-transparent. Transparent proxies are not currently supported.

Review the appropriate table below depending on whether you are deploying Data Classification in AWS, Azure, or GCP.

Required endpoints for AWS

Endpoints	Purpose
https://api.console.netapp.com	Communication with the Console service, which includes NetApp accounts.
https://netapp-cloud-account.auth0.com https://auth0.com	Communication with the Console website for centralized user authentication.
https://cloud-compliance-support-netapp.s3.us-west-2.amazonaws.com https://hub.docker.com https://auth.docker.io https://registry-1.docker.io https://index.docker.io/ https://dseasb33srnrn.cloudfront.net/ https://production.cloudflare.docker.com/	Provides access to software images, manifests, and templates.
https://kinesis.us-east-1.amazonaws.com	Enables NetApp to stream data from audit records.
https://cognito-idp.us-east-1.amazonaws.com https://cognito-identity.us-east-1.amazonaws.com https://user-feedback-store-prod.s3.us-west-2.amazonaws.com https://customer-data-production.s3.us-west-2.amazonaws.com	Enables Data Classification to access and download manifests and templates, and to send logs and metrics.

Required endpoints for Azure

Endpoints	Purpose
https://api.console.netapp.com	Communication with the Console service, which includes NetApp accounts.
https://netapp-cloud-account.auth0.com https://auth0.com	Communication with the Console website for centralized user authentication.
https://support.compliance.api.console.netapp.com/ https://hub.docker.com https://auth.docker.io https://registry-1.docker.io https://index.docker.io/ https://dseasb33srnrn.cloudfront.net/ https://production.cloudflare.docker.com/	Provides access to software images, manifests, templates, and to send logs and metrics.
https://support.compliance.api.console.netapp.com/	Enables NetApp to stream data from audit records.

Required endpoints for GCP

Endpoints	Purpose
https://api.console.netapp.com	Communication with the Console service, which includes NetApp accounts.

Endpoints	Purpose
https://netapp-cloud-account.auth0.com https://auth0.com	Communication with the Console website for centralized user authentication.
https://support.compliance.api.console.netapp.com/ https://hub.docker.com https://auth.docker.io https://registry-1.docker.io https://index.docker.io/ https://dseasb33srnrn.cloudfront.net/ https://production.cloudflare.docker.com/	Provides access to software images, manifests, templates, and to send logs and metrics.
https://support.compliance.api.console.netapp.com/	Enables NetApp to stream data from audit records.

Ensure that Data Classification has the required permissions

Ensure that Data Classification has permissions to deploy resources and create security groups for the Data Classification instance.

- Google Cloud permissions
- AWS permissions
- Azure permissions

Ensure that the Console agent can access Data Classification

Ensure connectivity between the Console agent and the Data Classification instance. The security group for the Console agent must allow inbound and outbound traffic over port 443 to and from the Data Classification instance. This connection enables deployment of the Data Classification instance and enables you to view information in the Compliance and Governance tabs. Data Classification is supported in Government regions in AWS and Azure.

Additional inbound and outbound security group rules are required for AWS and AWS GovCloud deployments. See Rules for the Console agent in AWS for details.

Additional inbound and outbound security group rules are required for Azure and Azure Government deployments. See Rules for the Console agent in Azure for details.

Ensure you can keep Data Classification running

The Data Classification instance needs to stay on to continuously scan your data.

Ensure web browser connectivity to Data Classification

After Data Classification is enabled, ensure that users access the Console interface from a host that has a connection to the Data Classification instance.

The Data Classification instance uses a private IP address to ensure that the indexed data isn't accessible to the internet. As a result, the web browser that you use to access the Console must have a connection to that private IP address. That connection can come from a direct connection to your cloud provider (for example, a VPN), or from a host that's inside the same network as the Data Classification instance.

Check your vCPU limits

Ensure that your cloud provider's vCPU limit allows for the deployment of an instance with the necessary number of cores. You'll need to verify the vCPU limit for the relevant instance family in the region where the

Console is running. See the required instance types.

See the following links for more details on vCPU limits:

- AWS documentation: Amazon EC2 service quotas
- Azure documentation: Virtual machine vCPU quotas
- Google Cloud documentation: Resource quotas

Deploy Data Classification in the cloud

Follow these steps to deploy an instance of Data Classification in the cloud. The Console agent will deploy the instance in the cloud, and then install Data Classification software on that instance.

In regions where the default instance type isn't available, Data Classification runs on an alternate instance type.

Deploy in AWS

Steps

1. From the main page of Data Classification, select **Deploy Classification On-Premises or Cloud**.
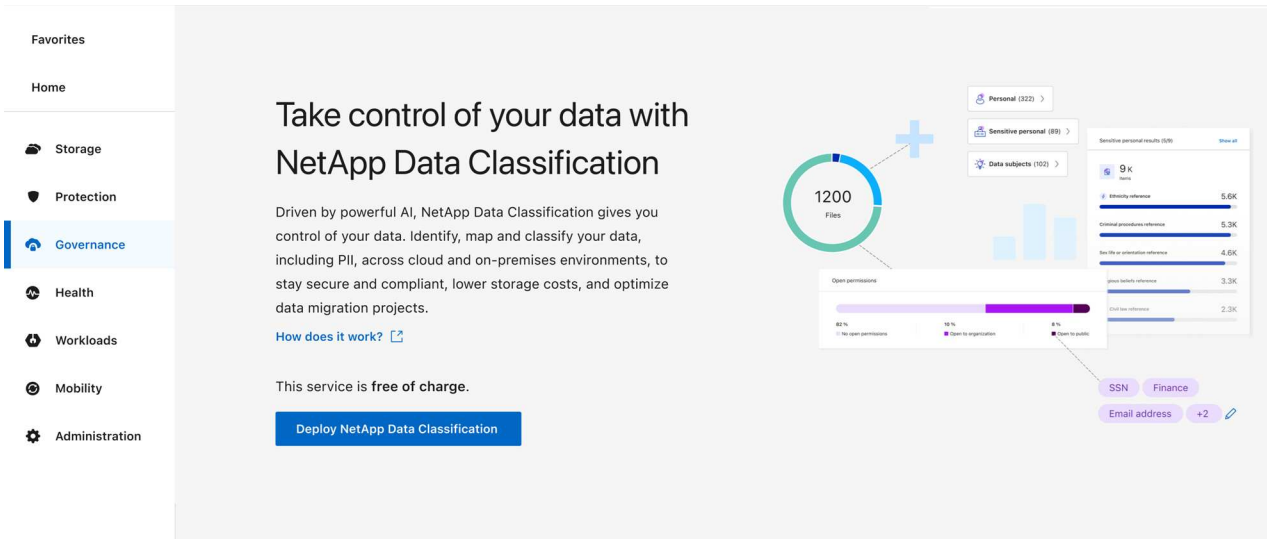
2. From the *Installation* page, select **Deploy > Deploy** to use the "Large" instance size and start the cloud deployment wizard.

3. The wizard displays progress as it goes through the deployment steps. When inputs are required or if it encounters issues, you are prompted.

4. When the instance is deployed and Data Classification is installed, select **Continue to configuration** to go to the *Configuration* page.

Deploy in Azure

Steps

1. From the main page of Data Classification, select **Deploy Classification On-Premises or Cloud**.
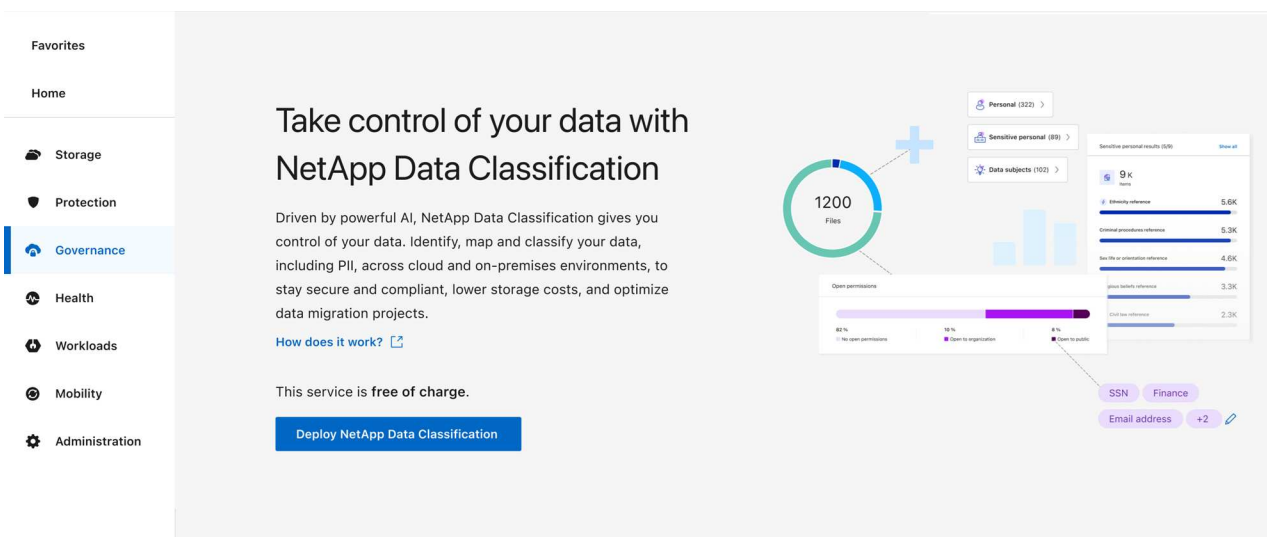
2. Select **Deploy** to start the cloud deployment wizard.

3. The wizard displays progress as it goes through the deployment steps. It will stop and prompt for input if it runs into any issues.

4. When the instance is deployed and Data Classification is installed, select **Continue to configuration** to go to the *Configuration* page.

Deploy in Google Cloud

Steps

1. From the main page of Data Classification, select **Governance > Classification**.

2. Select **Deploy Classification On-Premises or Cloud**.
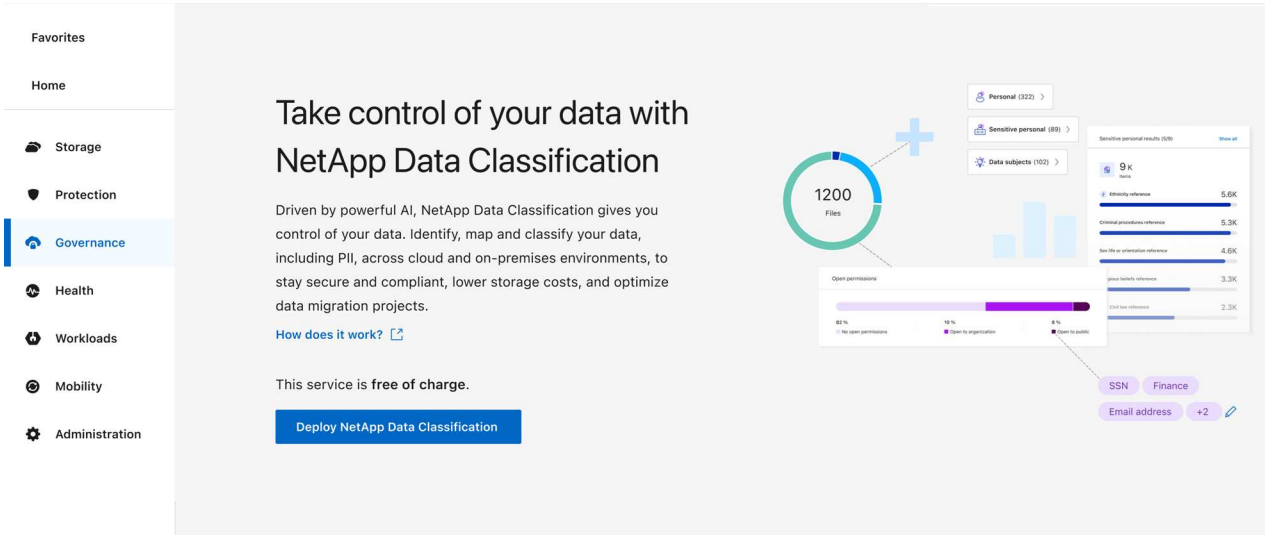
3. Select **Deploy** to start the cloud deployment wizard.

4. The wizard displays progress as it goes through the deployment steps. It will stop and prompt for input if it runs into any issues.

5. When the instance is deployed and Data Classification is installed, select **Continue to configuration** to go to the *Configuration* page.

Result

The Console deploys the Data Classification instance in your cloud provider.

Upgrades to the Console agent and Data Classification software is automated as long as the instances have internet connectivity.

What's Next

From the Configuration page you can select the data sources that you want to scan.

Install NetApp Data Classification on a host that has internet access

To deploy NetApp Data Classification on a Linux host in your network or on a Linux host in the cloud that has internet access, you need deploy the Linux host manually in your network or in the cloud.

The on-premises installation is a good option if you prefer to scan on-premises ONTAP systems using a Data Classification instance that's also located on premises. This is not a requirement. The software functions the same regardless of which installation method you choose.

The Data Classification installation script starts by checking if the system and environment meet the required

prerequisites. If the prerequisites are all met, then the installation starts. If you would like to verify the prerequisites independently of running the Data Classification installation, there is a separate software package you can download that only tests for the prerequisites. See how to check if your Linux host is ready to install Data Classification.

The typical installation on a Linux host *in your premises* has the following components and connections.

The typical installation on a Linux host *in the cloud* has the following components and connections.

Quick start

Get started quickly by following these steps, or scroll down to the remaining sections for full details.

1   **Create a Console agent**

If you don't already have a Console agent , deploy the Console agent on-premises on a Linux host in your network, or on a Linux host in the cloud.

You can also create a Console agent with your cloud provider. See creating a Console agent in AWS, creating a Console agent in Azure, or creating a Console agent in GCP.

2   **Review prerequisites**

Ensure that your environment can meet the prerequisites. This includes outbound internet access for the instance, connectivity between the Console agent and Data Classification over port 443, and more. See the complete list.

You also need a Linux system that meets the following requirements.

3   **Download and deploy Data Classification**

Download the Cloud Data Classification software from the NetApp Support Site and copy the installer file to the Linux host you plan to use. Then launch the installation wizard and follow the prompts to deploy the Data Classification instance.

Create a Console agent

A Console agent is required before you can install and use Data Classification. In most cases you'll probably have a Console agent set up before you attempt to activate Data Classification because most Console features require a Console agent, but there are cases where you'll you need to set one up now.

To create one in your cloud provider environment, see creating a Console agent in AWS, creating a Console agent in Azure, or creating a Console agent in GCP.

There are some scenarios where you have to use a Console agent that's deployed in a specific cloud provider:

- When scanning data in Cloud Volumes ONTAP in AWS or Amazon FSx for ONTAP, you use a Console agent in AWS.

- When scanning data in Cloud Volumes ONTAP in Azure or in Azure NetApp Files, you use a Console agent in Azure.

  For Azure NetApp Files, it must be deployed in the same region as the volumes you wish to scan.

- When scanning data in Cloud Volumes ONTAP in GCP, you use a Console agent in GCP.

On-prem ONTAP systems, NetApp file shares and database accounts can be scanned using any of these cloud Console agents.

Note that you can also deploy the Console agent on-premises on a Linux host in your network or on a Linux host in the cloud. Some users planning to install Data Classification on-prem may also choose to install the Console agent on-prem.

You'll need the IP address or host name of the Console agent system when installing Data Classification. You'll have this information if you installed the Console agent in your premises. If the Console agent is deployed in the cloud, you can find this information from the Console: select the Help icon then **Support** then **Console agent**.

Prepare the Linux host system

Data Classification software must run on a host that meets specific operating system requirements, RAM requirements, software requirements, and so on. The Linux host can be in your network, or in the cloud.

Ensure that you can keep Data Classification running. The Data Classification machine needs to stay on to continuously scan your data.

- Data Classification must be on a dedicated host. The host can't be shared with other applications or third-party software such as antivirus.
- Choose the size that aligns with the data set you plan to scan with Data Classification.

System size	CPU	RAM (swap memory must be disabled)	Disk
Extra Large	32 CPUs	128 GB RAM	1 TiB SSD on /, or 100 GiB available on /opt 895 GiB available on /var/lib/docker 5 GiB on /tmp **For Podman, 30 GB on /var/tmp**
Large	16 CPUs	64 GB RAM	500 GiB SSD on /, or 100 GiB available on /opt 400 GiB available on /var/lib/docker or for Podman /var/lib/containers 5 GiB on /tmp **For Podman, 30 GB on /var/tmp**

- When deploying a compute instance in the cloud for your Data Classification installation, it's recommended you use a system that meets the "Large" system requirements above:
  - **Amazon Elastic Compute Cloud (Amazon EC2) instance type**: "m6i.4xlarge". See additional AWS instance types.
  - **Azure VM size**: "Standard_D16s_v3". See additional Azure instance types.
  - **GCP machine type**: "n2-standard-16". See additional GCP instance types.
- **UNIX folder permissions**: The following minimum UNIX permissions are required:

Folder	Minimum permissions
/tmp	`rwxrwxrwt`
/opt	`rwxr-xr-x`
/var/lib/docker	`rwx------`
/usr/lib/systemd/system	`rwxr-xr-x`

- **Operating system**:
  - The following operating systems require using the Docker container engine:
    - Red Hat Enterprise Linux version 7.8 and 7.9
    - Ubuntu 22.04 (requires Data Classification version 1.23 or greater)
    - Ubuntu 24.04 (requires Data Classification version 1.23 or greater)
  - The following operating systems require using the Podman container engine, and they require Data Classification version 1.30 or greater:
    - Red Hat Enterprise Linux version 8.8, 8.10, 9.0, 9.1, 9.2, 9.3, 9.4, 9.5, and 9.6.
  - Advanced Vector Extensions (AVX2) must be enabled on the host system.
- **Red Hat Subscription Management**: The host must be registered with Red Hat Subscription Management. If it's not registered, the system can't access repositories to update required 3rd-party software during installation.
- **Additional software**: You must install the following software on the host before you install Data Classification:
  - Depending on the OS you are using, you need to install one of the container engines:
    - Docker Engine version 19.3.1 or greater. View installation instructions.
    - Podman version 4 or greater. To install Podman, enter (`sudo yum install podman netavark -y`).
- Python version 3.6 or greater. View installation instructions.
  - **NTP considerations**: NetApp recommends configuring the Data Classification system to use a Network Time Protocol (NTP) service. The time must be synchronized between the Data Classification system and the Console agent system.
- **Firewalld considerations**: If you are planning to use `firewalld`, we recommend that you enable it before installing Data Classification. Run the following commands to configure `firewalld` so that it is compatible with Data Classification:

```
firewall-cmd --permanent --add-service=http
firewall-cmd --permanent --add-service=https
firewall-cmd --permanent --add-port=80/tcp
firewall-cmd --permanent --add-port=8080/tcp
firewall-cmd --permanent --add-port=443/tcp
firewall-cmd --reload
```

If you're planning to use additional Data Classification hosts as scanner nodes, add these rules to your primary system at this time:

```
firewall-cmd --permanent --add-port=2377/tcp
firewall-cmd --permanent --add-port=7946/udp
firewall-cmd --permanent --add-port=7946/tcp
firewall-cmd --permanent --add-port=4789/udp
```

Note that you must restart Docker or Podman whenever you enable or update `firewalld` settings.

> ⓘ   The IP address of the Data Classification host system can't be changed after installation.

Enable outbound internet access from Data Classification

Data Classification requires outbound internet access. If your virtual or physical network uses a proxy server for internet access, ensure that the Data Classification instance has outbound internet access to contact the following endpoints.

| Endpoints	Purpose
https://api.console.netapp.com	Communication with the Console, which includes NetApp accounts.
https://netapp-cloud-account.auth0.com https://auth0.com	Communication with the Console website for centralized user authentication.
https://support.compliance.api.bluexp.netapp.com/ https://hub.docker.com https://auth.docker.io https://registry-1.docker.io https://index.docker.io/ https://dseasb33srnrn.cloudfront.net/ https://production.cloudflare.docker.com/	Provides access to software images, manifests, templates, and to send logs and metrics.
https://support.compliance.api.bluexp.netapp.com/	Enables NetApp to stream data from audit records.
https://github.com/docker https://download.docker.com	Provides prerequisite packages for docker installation.
http://packages.ubuntu.com/ http://archive.ubuntu.com	Provides prerequisite packages for Ubuntu installation.

Verify that all required ports are enabled

You must ensure that all required ports are open for communication between the Console agent, Data Classification, Active Directory, and your data sources.

Connection Type	Ports	Description
Console agent <> Data Classification	8080 (TCP), 443 (TCP), and 80. 9000	The firewall or routing rules for the Console agent must allow inbound and outbound traffic over port 443 to and from the Data Classification instance. Make sure port 8080 is open so you can see the installation progress in the Console. If a firewall is used on the Linux host, port 9000 is required for internal processes within an Ubuntu server.
Console agent <> ONTAP cluster (NAS)	443 (TCP)	The Console discovers ONTAP clusters using HTTPS. If you use custom firewall policies, they must meet the following requirements: • The Console agent host must allow outbound HTTPS access through port 443. If the Console agent is in the cloud, all outbound communication is allowed by the predefined firewall or routing rules. • The ONTAP cluster must allow inbound HTTPS access through port 443. The default "mgmt" firewall policy allows inbound HTTPS access from all IP addresses. If you modified this default policy, or if you created your own firewall policy, you must associate the HTTPS protocol with that policy and enable access from the Console agent host.
Data Classification <> ONTAP cluster	• For NFS - 111 (TCP\UDP) and 2049 (TCP\UDP) • For CIFS - 139 (TCP\UDP) and 445 (TCP\UDP)	Data Classification needs a network connection to each Cloud Volumes ONTAP subnet or on-prem ONTAP system. Firewalls or routing rules for Cloud Volumes ONTAP must allow inbound connections from the Data Classification instance. Make sure these ports are open to the Data Classification instance: • For NFS - 111 and 2049 • For CIFS - 139 and 445 NFS volume export policies must allow access from the Data Classification instance.

Connection Type	Ports	Description
Data Classification <> Active Directory	389 (TCP & UDP), 636 (TCP), 3268 (TCP), and 3269 (TCP)	You must have an Active Directory already set up for the users in your company. Additionally, Data Classification needs Active Directory credentials to scan CIFS volumes. You must have the information for the Active Directory: • DNS Server IP Address, or multiple IP Addresses • User Name and Password for the server • Domain Name (Active Directory Name) • Whether you are using secure LDAP (LDAPS) or not • LDAP Server Port (typically 389 for LDAP, and 636 for secure LDAP)

Install Data Classification on the Linux host

For typical configurations you'll install the software on a single host system. See those steps here.
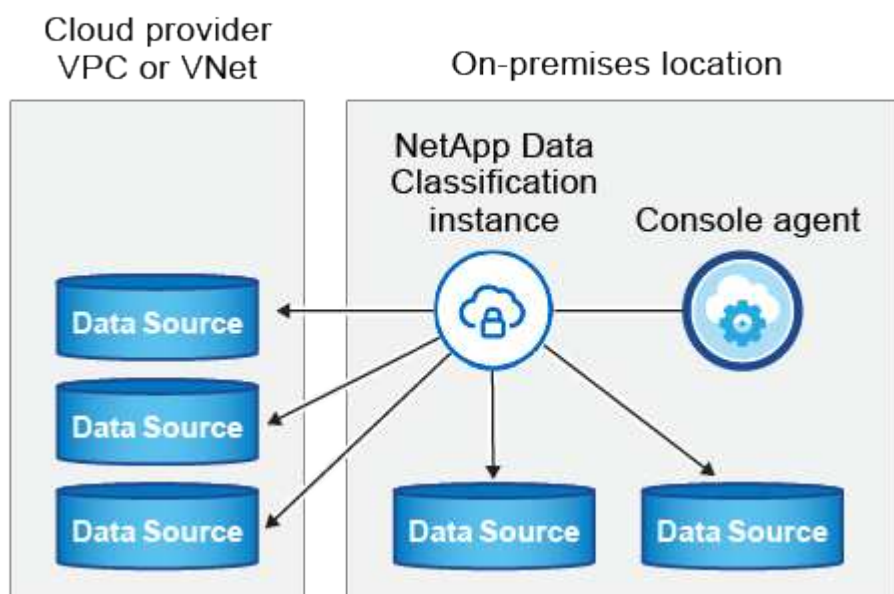
See Preparing the Linux host system and Reviewing prerequisites for the full list of requirements before you deploy Data Classification.

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

> ⓘ Data Classification is currently unable to scan S3 buckets, Azure NetApp Files, or FSx for ONTAP when the software is installed on premises. In these cases you'll need to deploy a separate Console agent and instance of Data Classification in the cloud and switch between Connectors for your different data sources.

Single-host installation for typical configurations

Review the requirements and follow these steps when installing Data Classification software on a single on-premises host.

Watch this video to see how to install Data Classification.

Note that all installation activities are logged when installing Data Classification. If you run into any issues during installation, you can view the contents of the installation audit log. It is written to `/opt/netapp/install_logs/`.

Before you begin

- Verify that your Linux system meets the host requirements.
- Verify that the system has the two prerequisite software packages installed (Docker Engine or Podman, and Python 3).
- Make sure you have root privileges on the Linux system.
- If you're using a proxy for access to the internet:
    - You'll need the proxy server information (IP address or host name, connection port, connection scheme: https or http, user name and password).
    - If the proxy is performing TLS interception, you'll need to know the path on the Data Classification Linux system where the TLS CA certificates are stored.
    - The proxy must be non-transparent. Data Classification does not currently support transparent proxies.
    - The user must be a local user. Domain users are not supported.
- Verify that your offline environment meets the required permissions and connectivity.

Steps

1. Download the Data Classification software from the NetApp Support Site. The file you should select is named **DATASENSE-INSTALLER-<version>.tar.gz**.
2. Copy the installer file to the Linux host you plan to use (using `scp` or some other method).
3. Unzip the installer file on the host machine, for example:

    ```
    tar -xzf DATASENSE-INSTALLER-V1.25.0.tar.gz
    ```

4. In the Console, select **Governance > Classification**.
5. Select **Deploy Classification On-Premises or Cloud**.

6. Depending on whether you are installing Data Classification on an instance you prepared in the cloud or on an instance you prepared in your premises, select the appropriate **Deploy** option to start the Data Classification installation.

7. The *Deploy Data Classification On Premises* dialog is displayed. Copy the provided command (for example: `sudo ./install.sh -a 12345 -c 27AG75 -t 2198qq`) and paste it in a text file so you can use it later. Then select **Close** to dismiss the dialog.

8. On the host machine, enter the command you copied and then follow a series of prompts, or you can provide the full command including all required parameters as command line arguments.

   Note that the installer performs a pre-check to make sure your system and networking requirements are in place for a successful installation. Watch this video to understand the pre-check messages and implications.

Enter parameters as prompted:	Enter the full command:

1. Paste the command you copied from step 7:
   `sudo ./install.sh -a <account_id>`
   `-c <client_id> -t <user_token>`
 	
   If you are installing on a cloud instance (not on your premises), add `--manual-cloud` `-install <cloud_provider>`.
 	
2. Enter the IP address or host name of the Data Classification host machine so it can be accessed by the Console agent system.
 	
3. Enter the IP address or host name of the Console agent host machine so it can be accessed by the Data Classification system.
 	
4. Enter proxy details as prompted. If your Console agent already uses a proxy, there is no need to enter this information again here since Data Classification will automatically use the proxy used by the Console agent.
	Alternatively, you can create the whole command in advance, providing the necessary host and proxy parameters:
	`sudo ./install.sh -a <account_id> -c <client_id> -t <user_token> --host <ds_host> --manager-host <cm_host> --manual-cloud-install <cloud_provider> --proxy-host <proxy_host> --proxy-port <proxy_port> --proxy-scheme <proxy_scheme> --proxy-user <proxy_user> --proxy-password <proxy_password> --cacert-folder-path <ca_cert_dir>`

Variable values:

- *account_id* = NetApp Account ID
- *client_id* = Console agent Client ID (add the suffix "clients" to the client ID if it not already there)
- *user_token* = JWT user access token
- *ds_host* = IP address or host name of the Data Classification Linux system.
- *cm_host* = IP address or host name of the Console agent system.
- *cloud_provider* = When installing on a cloud instance, enter "AWS", "Azure", or "Gcp" depending on cloud provider.
- *proxy_host* = IP or host name of the proxy server if the host is behind a proxy server.
- *proxy_port* = Port to connect to the proxy server (default 80).
- *proxy_scheme* = Connection scheme: https or http (default http).
- *proxy_user* = Authenticated user to connect to the proxy server, if basic authentication is required. The user must be a local user - domain users are not supported.
- *proxy_password* = Password for the user name that you specified.
- *ca_cert_dir* = Path on the Data Classification Linux system containing additional TLS CA certificate bundles. Only required if the proxy is performing TLS interception.

Result

The Data Classification installer installs packages, registers the installation, and installs Data Classification. Installation can take 10 to 20 minutes.

If there is connectivity over port 8080 between the host machine and the Console agent instance, you'll see the installation progress in the Data Classification tab in the Console.

What's Next

From the Configuration page you can select the data sources that you want to scan.

Install NetApp Data Classification on a Linux host with no internet access

Installing NetApp Data Classification on a Linux host in an on-premises site that doesn't have internet access is known as *private mode*. This type of installation, which uses an installation script, has no connectivity to the NetApp Console SaaS layer.

> (i) BlueXP private mode (legacy BlueXP interface) is typically used with on-premises environments that have no internet connection and with secure cloud regions, which includes AWS Secret Cloud, AWS Top Secret Cloud, and Azure IL6. NetApp continues to support these environments with the legacy BlueXP interface. For private mode documentation in the legacy BlueXP interface, see PDF documentation for BlueXP private mode.

Check that your Linux host is ready to install NetApp Data Classification

Before installing NetApp Data Classification manually on a Linux host, optionally run a script on the host to verify that all the prerequisites are in place for installing Data Classification. You can run this script on a Linux host in your network, or on a Linux host in the cloud. The host can be connected to the internet, or the host can reside in a site

that doesn't have internet access (a *dark site*).

The Data Classification installation script encompasses a test script to ensure your environment meets the requirements. You can run this script separately to verify the Linux host's readiness before run the installation script.

Getting Started

You'll perform the following tasks.

- Optionally, install a Console agent if you don't already have one installed. You can run the test script without having a Console agent installed, but the script checks for connectivity between the Console agent and the Data Classification host machine - so it is recommended that you have a Console agent.
- Prepare the host machine and verify that it meets all the requirements.
- Enable outbound internet access from the Data Classification host machine.
- Verify that all required ports are enabled on all systems.
- Download and run the Prerequisite test script.

Create a Console agent

A Console agent is required before you can install and use Data Classification. You can, however, run the Prerequisites script without a Console agent.

You can install the Console agent on-premises on a Linux host in your network or on a Linux host in the cloud. You can also install Data Classification on-premises if the Console agent is installed on-premises.

To create a Console agent in your cloud provider environment, see:

- creating a Console agent in AWS
- creating a Console agent in Azure
- creating a Console agent in GCP

You need the IP address or host name of the Console agent system when running the Prerequisites script. You have this information if you installed the Console agent in your premises. If the Console agent is deployed in the cloud, you can find this information from the Console: select the Help icon then **Support**; in the Agent and Audit section, select **Go to the agent**.

Verify host requirements

Data Classification software must run on a host that meets specific operating system requirements, RAM requirements, and software requirements.

- Data Classification must be on a dedicated host. The host can't be shared with other applications or third-party software such as antivirus.
- Choose the size that aligns with the data set you plan to scan with Data Classification.

System size	CPU	RAM (swap memory must be disabled)	Disk
Extra Large	32 CPUs	128 GB RAM	• 1 TiB SSD on /, or 100 GiB available on /opt • 895 GiB available on /var/lib/docker • 5 GiB on /tmp • **For Podman, 30 GB on /var/tmp**
Large	16 CPUs	64 GB RAM	• 500 GiB SSD on /, or 100 GiB available on /opt • 400 GiB available on /var/lib/docker or for Podman /var/lib/containers • 5 GiB on /tmp • **For Podman, 30 GB on /var/tmp**

- When deploying a compute instance in the cloud for your Data Classification installation, it's recommended you use a system that meets the "Large" system requirements above:
    - **Amazon Elastic Compute Cloud (Amazon EC2) instance type**: "m6i.4xlarge". See additional AWS instance types.
    - **Azure VM size**: "Standard_D16s_v3". See additional Azure instance types.
    - **GCP machine type**: "n2-standard-16". See additional GCP instance types.
- **UNIX folder permissions**: The following minimum UNIX permissions are required:

Folder	Minimum permissions
/tmp	`rwxrwxrwt`
/opt	`rwxr-xr-x`
/var/lib/docker	`rwx------`
/usr/lib/systemd/system	`rwxr-xr-x`

- **Operating system**:
    - The following operating systems require using the Docker container engine:
        - Red Hat Enterprise Linux version 7.8 and 7.9
        - Ubuntu 22.04 (requires Data Classification version 1.23 or greater)
        - Ubuntu 24.04 (requires Data Classification version 1.23 or greater)
    - The following operating systems require using the Podman container engine, and they require Data Classification version 1.30 or greater:
        - Red Hat Enterprise Linux version 8.8, 8.10, 9.0, 9.1, 9.2, 9.3, 9.4, 9.5, and 9.6.

- Advanced Vector Extensions (AVX2) must be enabled on the host system.

- **Red Hat Subscription Management**: The host must be registered with Red Hat Subscription Management. If it's not registered, the system can't access repositories to update required 3rd-party software during installation.

- **Additional software**: You must install the following software on the host before you install Data Classification:

  - Depending on the OS you are using, you need to install one of the container engines:

    - Docker Engine version 19.3.1 or greater. [View installation instructions](#).

    - Podman version 4 or greater. To install Podman, enter (`sudo yum install podman netavark -y`).

- Python version 3.6 or greater. [View installation instructions](#).

  - **NTP considerations**: NetApp recommends configuring the Data Classification system to use a Network Time Protocol (NTP) service. The time must be synchronized between the Data Classification system and the Console agent system.

- **Firewalld considerations**: If you are planning to use `firewalld`, we recommend that you enable it before installing Data Classification. Run the following commands to configure `firewalld` so that it is compatible with Data Classification:

```
firewall-cmd --permanent --add-service=http
firewall-cmd --permanent --add-service=https
firewall-cmd --permanent --add-port=80/tcp
firewall-cmd --permanent --add-port=8080/tcp
firewall-cmd --permanent --add-port=443/tcp
firewall-cmd --reload
```

If you're planning to use additional Data Classification hosts as scanner nodes (in a distributed model), add these rules to your primary system at this time:

```
firewall-cmd --permanent --add-port=2377/tcp
firewall-cmd --permanent --add-port=7946/udp
firewall-cmd --permanent --add-port=7946/tcp
firewall-cmd --permanent --add-port=4789/udp
```

Note that you must restart Docker or Podman whenever you enable or update `firewalld` settings.

Enable outbound internet access from Data Classification

Data Classification requires outbound internet access. If your virtual or physical network uses a proxy server for internet access, ensure that the Data Classification instance has outbound internet access to contact the following endpoints.

> This section is not required for host systems installed in sites without internet connectivity.

Endpoints	Purpose
https://api.console.netapp.com	Communication with the Console service, which includes NetApp accounts.
https://netapp-cloud-account.auth0.com https://auth0.com	Communication with the Console website for centralized user authentication.
https://support.compliance.api.console.netapp. com/ https://hub.docker.com https://auth.docker.io https://registry-1.docker.io https://index.docker.io/ https://dseasb33srnrn.cloudfront.net/ https://production.cloudflare.docker.com/	Provides access to software images, manifests, templates, and to send logs and metrics.
https://support.compliance.api.console.netapp. com/	Enables NetApp to stream data from audit records.
https://github.com/docker https://download.docker.com	Provides prerequisite packages for docker installation.
http://packages.ubuntu.com/ http://archive.ubuntu.com	Provides prerequisite packages for Ubuntu installation.

Verify that all required ports are enabled

You must ensure that all required ports are open for communication between the Console agent, Data Classification, Active Directory, and your data sources.

Connection Type	Ports	Description
Console agent <> Data Classification	8080 (TCP), 443 (TCP), and 80. 9000	The firewall or routing rules for the Console agent must allow inbound and outbound traffic over port 443 to and from the Data Classification instance. Make sure port 8080 is open so you can see the installation progress in the Console. If a firewall is used on the Linux host, port 9000 is required for internal processes within an Ubuntu server.
Console agent <> ONTAP cluster (NAS)	443 (TCP)	The Console discovers ONTAP clusters using HTTPS. If you use custom firewall policies, the Console agent host must allow outbound HTTPS access through port 443. If the Console agent is in the cloud, all outbound communication is allowed by the predefined firewall or routing rules.

Run the Data Classification prerequisites script

Follow these steps to run the Data Classification prerequisites script.

Watch this video to see how to run the Prerequisites script and interpret the results.

Before you begin

- Verify that your Linux system meets the host requirements.

- Verify that the system has the two prerequisite software packages installed (Docker Engine or Podman, and Python 3).

- Make sure you have root privileges on the Linux system.

Steps

1. Download the Data Classification Prerequisites script from the NetApp Support Site. The file you should select is named **standalone-pre-requisite-tester-<version>**.

2. Copy the file to the Linux host you plan to use (using `scp` or some other method).

3. Assign permissions to run the script.

```
chmod +x standalone-pre-requisite-tester-v1.25.0
```

4. Run the script using the following command.

```
./standalone-pre-requisite-tester-v1.25.0 <--darksite>
```

Add the option "--darksite" only if you are running the script on a host that doesn't have internet access. Certain prerequisite tests are skipped when the host is not connected to the internet.

5. The script prompts you for the IP address of the Data Classification host machine.

   - Enter the IP address or host name.

6. The script prompts whether you have an installed Console agent.

   - Enter **N** if you do not have an installed Console agent.

   - Enter **Y** if you do have an installed Console agent. And then enter the IP address or host name of the Console agent so the test script can test this connectivity.

7. The script runs a variety of tests on the system and it displays results as it progresses. When it finishes it writes a log of the session to a file named `prerequisites-test-<timestamp>.log` in the directory `/opt/netapp/install_logs`.

Result

If all the prerequisites tests ran successfully, you can install Data Classification on the host when you are ready.

If any issues were discovered, they are categorized as "Recommended" or "Required" to be fixed. Recommended issues are typically items that would make the Data Classification scanning and categorizing tasks run slower. These items do not need to be corrected - but you may want to address them.

If you have any "Required" issues, you should fix the issues and run the Prerequisites test script again.

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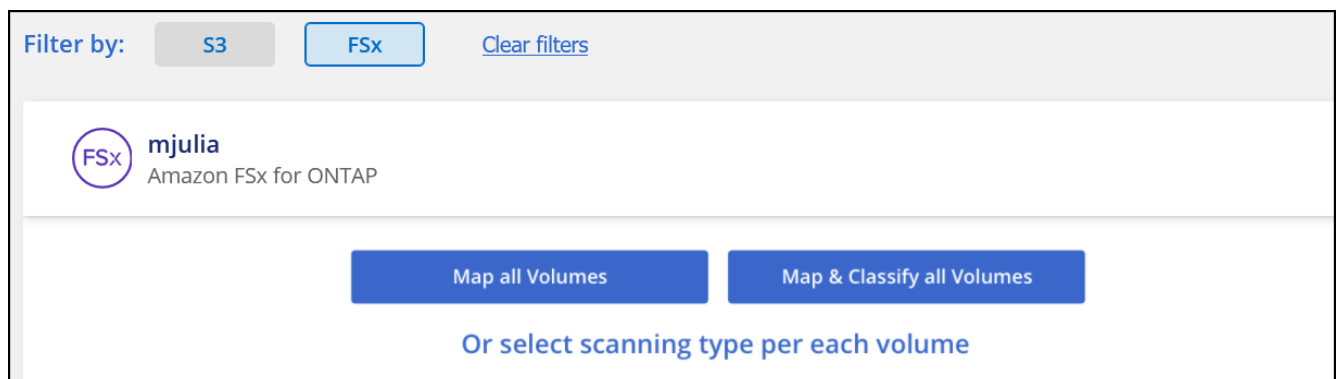
3. Select how you want to scan the volumes in each system. Learn about mapping and classification scans:
   - To map all volumes, select **Map all Volumes**.
   - To map and classify all volumes, select **Map & Classify all Volumes**.
   - To customize scanning for each volume, select **Or select scanning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

Scan	Storage Repository (Volume)	Type	Status	Required Action
Off ⏐ Map ⏐ Map & Classify	jrmclone	NFS	● No Access	Check network connectivity between the Data Sense ...

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.
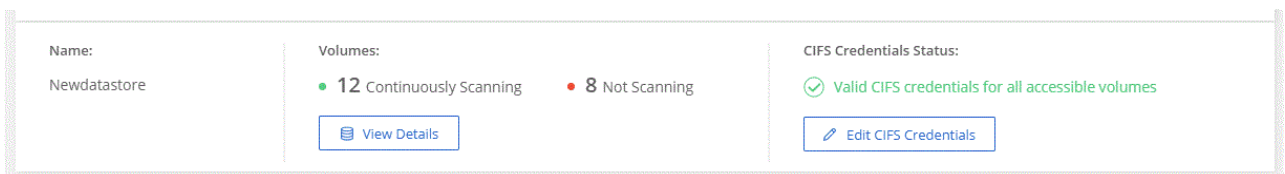
Steps

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.

Name:	Volumes:		CIFS Credentials Status:
Newdatastore	● **12** Continuously Scanning	● **8** Not Scanning	✓ Valid CIFS credentials for all accessible volumes
	▣ View Details		✎ Edit CIFS Credentials

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.

Scan	Storage repository (Volume)	Type	Mapping status	Scan progress	Required Action
Off Map [Map & Classify]	bank_statements	NFS	● Paused 2025-07-16 08:51 Last full cycle: 2025-07-16 08:50	Mapped 219 Classified 219	...
Off [Map] Map & Classify ☆	cifs_labs	CIFS	● Finished 2025-10-06 10:29 Last full cycle: 2025-10-06 10:29	Mapped 5.2K	...
[Off] Map Map & Classify	cifs_labs_second	CIFS			...
[Off] Map Map & Classify	cifs_labs_second_insight	NFS			...
Off [Map] Map & Classify	datasence	NFS	● Paused 2025-07-15 09:10 Last full cycle: 2025-07-15 09:06	Mapped 127K	...

Volumes selected for Data Classification scan (11/15)

Off Map Map & Classify Custom Mapping vs. Classification →   Retry All   Edit CIFS Credentials

ⓘ Scan when missing "write" permissions ⚪

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.

> ℹ️ 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.

> (i) 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.

(i) 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**.

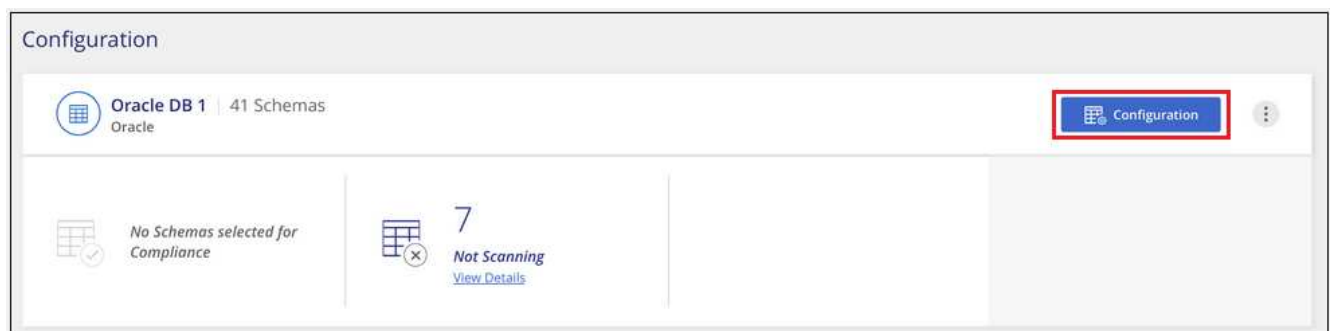The database is added to the list of systems.

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.
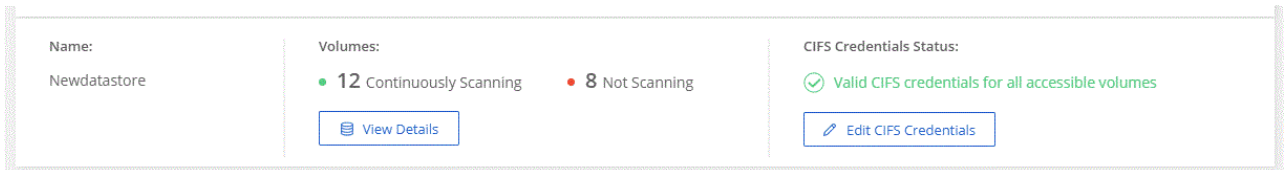
Steps

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.

Name:	Volumes:	CIFS Credentials Status:
Newdatastore	● **12** Continuously Scanning ● **8** Not Scanning	✓ Valid CIFS credentials for all accessible volumes
	🗄 View Details	✎ Edit CIFS Credentials

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.

> ℹ 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.

> (i) 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**.

Edit CIFS Authentication

Classification requires Active Directory credentials to access CIFS Volumes
in Micky.
The credentials can be read-only, but providing admin credentials ensures that
Classification can read any data that requires elevated permissions.

Select Authentication Method

⦿ NTLM

◯ Kerberos

Username ⓘ

domain\user or user@domain

Password

Password

Save   Cancel

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.

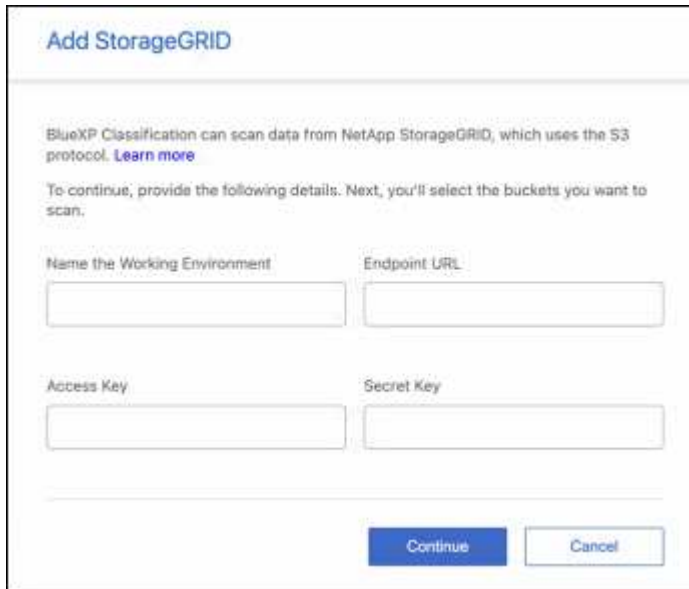### Add the StorageGRID service to Data Classification

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.

Integrate your Active Directory with NetApp Data Classification

You can integrate a global Active Directory with NetApp Data Classification to enhance the results that Data Classification reports about file owners and which users and groups have access to your files.

When you set up certain data sources (listed below), you need to enter Active Directory credentials in order for Data Classification to scan CIFS volumes. This integration provides Data Classification with file owner and permissions details for the data that resides in those data sources. The Active Directory entered for those data sources might differ from the global Active Directory credentials you enter here. Data Classification will look in all integrated Active Directories for user and permission details.

This integration provides additional information in the following locations in Data Classification:

- You can use the "File Owner" filter and see results in the file's metadata in the Investigation pane. Instead of the file owner containing the SID (Security IDentifier), it is populated with the actual user name.

  You can also view more details about the file owner: account name, email address, and SAM account name, or view items owned by that user.

- You can see full file permissions for each file and directory when you click the "View all Permissions" button.

- In the Governance dashboard, the Open Permissions panel will show a greater level of detail about your data.

> ℹ️ Local user SIDs, and SIDs from unknown domains, are not translated to the actual user name.

Supported data sources

An Active Directory integration with Data Classification can identify data from within the following data sources:

- On-premises ONTAP systems
- Cloud Volumes ONTAP
- Azure NetApp Files
- FSx for ONTAP

Connect to your Active Directory server

After you've deployed Data Classification and have activated scanning on your data sources, you can integrate Data Classification with your Active Directory. Active Directory can be accessed using a DNS Server IP address or an LDAP Server IP address.

The Active Directory 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.

For CIFS volumes/file shares, if you want to make sure your files "last accessed times" are unchanged by Data Classification classification scans, the user should have Write Attributes permission. If possible, we recommend making the Active Directory configured user part of a parent group in the organization which has

permissions to all files.

Requirements

- You must have an Active Directory already set up for the users in your company.
- You must have the information for the Active Directory:
  - DNS Server IP address, or multiple IP addresses

    or

    LDAP Server IP address, or multiple IP addresses

  - User Name and Password to access the server
  - Domain Name (Active Directory Name)
  - Whether you are using secure LDAP (LDAPS) or not
  - LDAP Server Port (typically 389 for LDAP, and 636 for secure LDAP)
- The following ports must be open for outbound communication by the Data Classification instance:

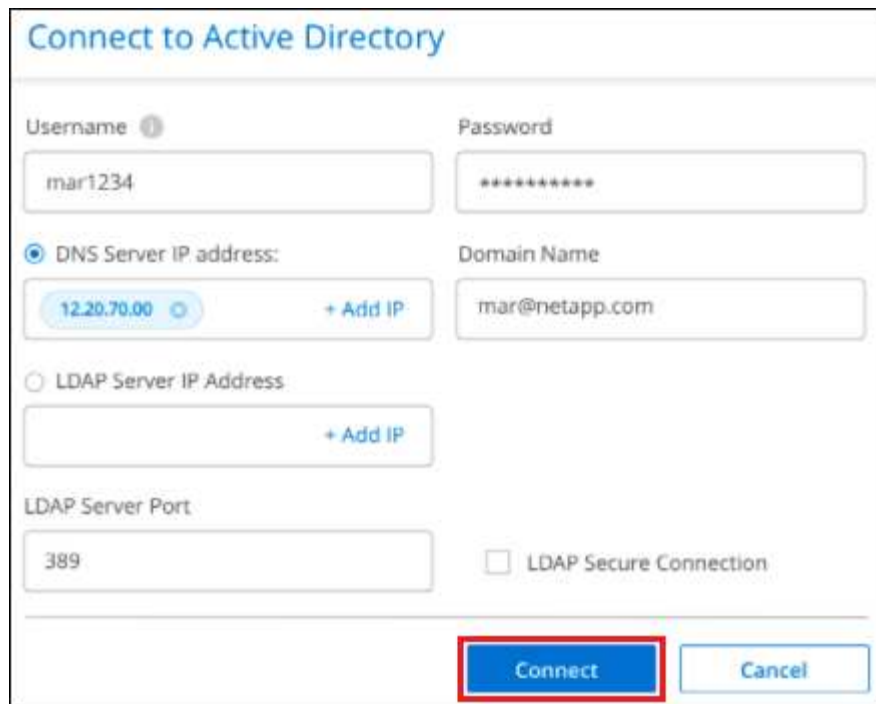| Protocol	Port	Destination	Purpose
TCP & UDP	389	Active Directory	LDAP
TCP	636	Active Directory	LDAP over SSL
TCP	3268	Active Directory	Global Catalog
TCP	3269	Active Directory	Global Catalog over SSL

Steps

1. From the Data Classification Configuration page, click **Add Active Directory**.

2. In the Connect to Active Directory dialog, enter the Active Directory details and click **Connect**.
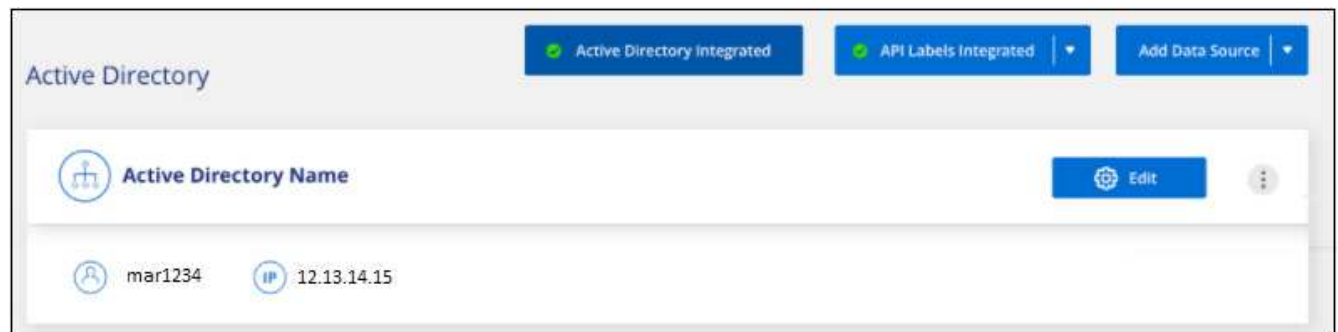
   You can add multiple IP addresses, if required, by selecting **Add IP**.

Data Classification integrates to the Active Directory, and a new section is added to the Configuration page.

Manage your Active Directory integration

If you need to modify any values in your Active Directory integration, click the **Edit** button and make the changes.

You can also delete the integration selecting the ⋮ button then **Remove Active Directory**.

Use Data Classification

View governance details about the data stored in your organization with NetApp Data Classification

Gain control of the costs related to the data on your organization's storage resources. NetApp Data Classification identifies the amount of stale data, duplicate files, and very large files in your systems so you can decide whether you want to remove or tier some files to less expensive object storage.

This is where you should begin your research. From the Governance dashboard, you can select an area for further investigation.

Additionally, if you're planning to migrate data from on-premises locations to the cloud, you can view the size of the data and whether any of the data contains sensitive information before moving it.

Review the Governance dashboard

The Governance dashboard provides information so that you can increase the efficiency and control the costs related to the data stored on your storage resources.

Organization
Org name

Project
Project name

Classification

Governance

Compliace

Investigation

Custom classification

Policies

Configuration

Governance

Monitor data governance metrics and optimize storage   Learn more

Last updated: August 11, 2025, 10:05 AM   ↻ Refresh

260.5 K	265.5 GiB	141	70.6 K
Scanned files count	Scanned files size	Scanned tables count	Identified PII

Sensitive data and wide permissions

Risk zones showing file counts by access level and sensitivity. Click to investigate.

Sensitivity

	1-10 users	11-100 users	Over 100 users
Over 101 identifies	ⓘ 281 files	🛡 110 files	⚠ 82 files
11-100 identifies	✓ 620 files	ⓘ 120 files	🛡 128 files
0-10 identifiers	✓ 220 files	✓ 106 files	ⓘ 109 files

Exposure

✓ 652 files	ⓘ 652 files	ⓘ 238 files	⚠ 82 files
● Low risk	● Medium risk	● High risk	● Critical risk

Savings opportunities

🗑 **Stale data** Files not modified in over 3 years	206.6K Items	227 GiB	View files
🗂 **Duplicate files** Files identified as duplicates of other files	206.6K Items	227 GiB	View files

Open permissions
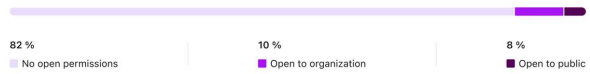
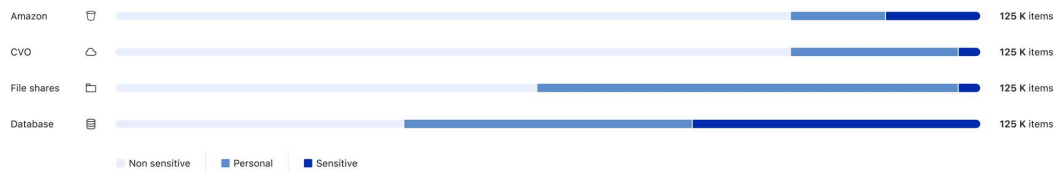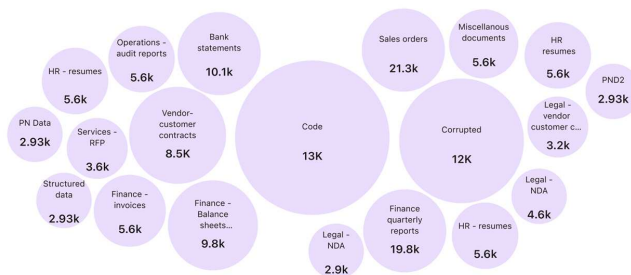| 82 %	10 %	8 %
No open permissions	Open to organization	Open to public

Reports

Data discovery assessment report
Summary of data risks, governance gaps, and compliance findings across scanned systems
↓ Download

Full data mapping overview report
Detailed breakdown of data types, volumes, and storage locations
↓ Download

Top data repositories by sensitivity level

Amazon	🗑		125 K items
CVO	☁		125 K items
File shares	📁		125 K items
Database	🗄		125 K items

● Non sensitive   ● Personal   ● Sensitive

Top document categories (20/40)

Show all

HR - resumes 5.6k
Operations - audit reports 5.6k
Bank statements 10.1k
Sales orders 21.3k
Miscellanous documents 5.6k
HR resumes 5.6k
PND2 2.93k
PN Data 2.93k
Services - RFP 3.6k
Vendor-customer contracts 8.5K
Code 13K
Corrupted 12K
Legal - vendor customer c... 3.2k
Structured data 2.93k
Finance - invoices 5.6k
Finance - Balance sheets... 9.8k
Legal - NDA 2.9k
Finance quarterly reports 19.8k
HR - resumes 5.6k
Legal - NDA 4.6k

Age of data

Last modified ▾

40K	40K	40K	0K	40K	40K	40K	40K	40K
>7 years	3-5 years	1-3 years	181-365 days	91-180 days	31-90 days	<30 days	<30 days	<30 days

Size of data

40K	40K	40K	0K	40K	40K	40K	40K
< 1 Byte	1 Byte - 1KB	1 KB - 1 MB	1 MB - 10 MB	10 MB - 100 MB	100 MB - 1GB	1 GB - 100 GB	> 100 GB
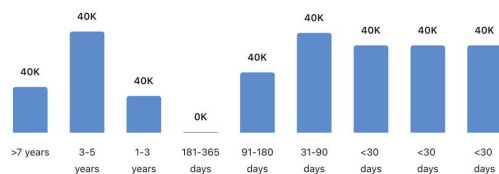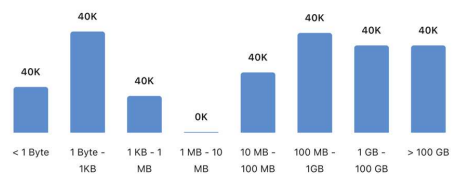
Steps

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

2. Select **Governance**.

   The Governance dashboard appears.

Review savings opportunities

The *Saving Opportunities* component shows data that you can delete or tier to less expensive object storage. The data in *Saving Opportunities* update every 2 hours. You can also manually update the data.

Steps

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

2. Within each Savings Opportunities tile of the Governance dashboard, select **Optimize Storage** to view the filtered results in the Investigation page. To discover any data you should delete or tier to less expensive storage, investigate the the *Saving Opportunities*.

   ◦ **Stale Data** - By default, data is considered stale if it was last modified over 3 years ago. You can [customize the definition of stale data](task-stale-data.html).

   ◦ **Duplicate Files** - Files that are duplicated in other locations in the data sources you are scanning. See what types of duplicate files are displayed.

> ⓘ   If any of your data sources implement data tiering, old data that already resides in object storage can be identified in the *Stale Data* category.

Create the data discovery assessment report

The data discovery assessment report provides a high-level analysis of the scanned environment to show areas of concern and potential remediation steps. The results are based on both mapping and classifying your data. The goal of this report is to raise awareness of three significant aspects of your dataset:

Feature	Description
Data governance concerns	A detailed picture of all the data you own and areas where you may be able to reduce the amount of data to save costs.
Data security exposures	Areas where your data is accessible to internal or external attacks because of broad access permissions.
Data compliance gaps	Where your personal or sensitive personal information is located for both security and for DSARs (data subject access requests).

With the report, you can take the following actions:

- Reduce storage costs by changing your retention policy, or by moving or deleting certain data (stale or duplicate data).

- Protect your data that has broad permissions by revising global group management policies.

- Protect your data that has personal or sensitive personal information by moving PII to more secure data stores.

Steps

1. From Data Classification, select **Governance**.
2. In the reports tile, select **Data Discovery Assessment Report**.

Reports

Data discovery assessment report

Summary of data risks, governance gaps, and compliance findings across scanned systems

⤓ Download

Full data mapping overview report

Detailed breakdown of data types, volumes, and storage locations
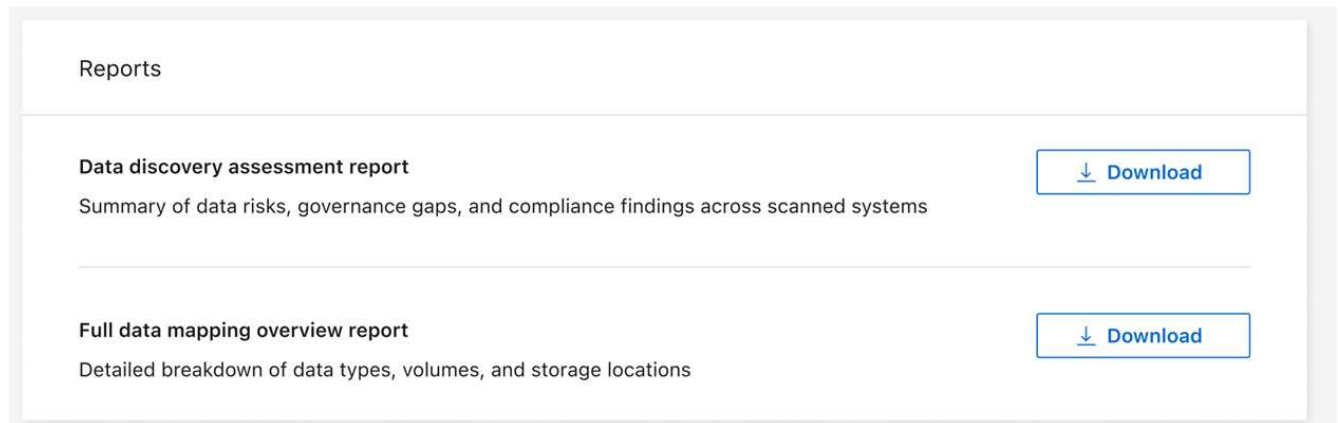
⤓ Download

Result

Data Classification generates a PDF report that you can review and share.

## Create the data mapping overview report

The data mapping overview report provides an overview of the data being stored in your corporate data sources to assist you with decisions of migration, back up, security, and compliance processes. The report summarizes all systems and data sources. It also provides an analysis for each system.

The report includes the following information:

Category	Description
Usage Capacity	For all systems: Lists the number of files and the used capacity for each system. For single systems: Lists the files that are using the most capacity.
Age of Data	Provides three charts and graphs for when files were created, last modified, or last accessed. Lists the number of files, and their used capacity, based on certain date ranges.
Size of Data	Lists the number of files that exist within certain size ranges in your systems.

Steps

1. From Data Classification, select **Governance**.
2. In the reports tile, select **Full data mapping overview report**.
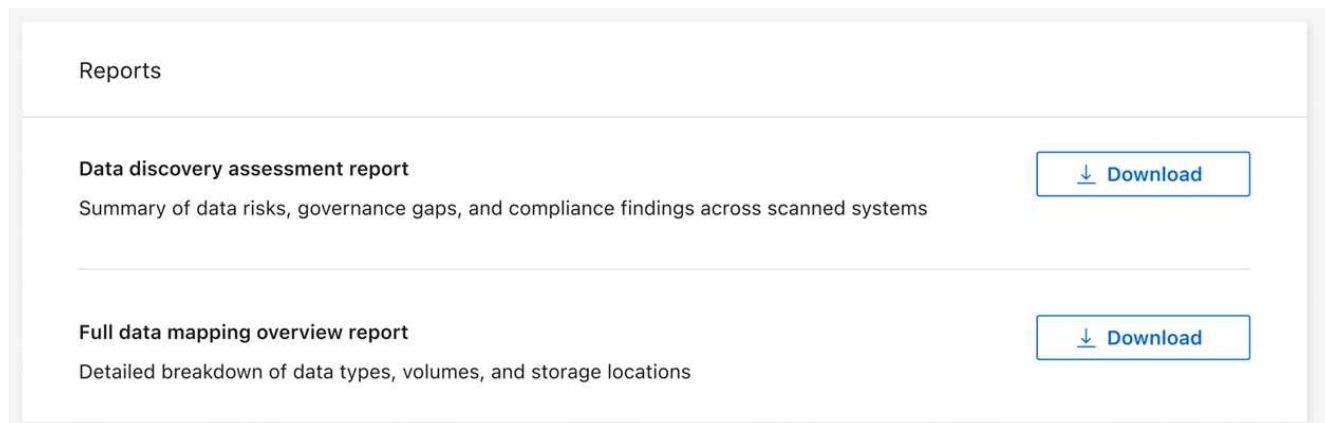
Result

Data Classification generates a PDF report that you can review and send to other groups as needed.

If the report is larger than 1 MB, the PDF file is retained on the Data Classification instance and you'll see a pop-up message about the exact location. When Data Classification is installed on a Linux machine on your premises, or on a Linux machine you deployed in the cloud, you can navigate directly to the PDF file. When Data Classification is deployed in the cloud, you need to authorize with SSH to the Data Classification instance to download the PDF file.

Review the top data repositories listed by data sensitivity

The *Top Data Repositories by Sensitivity Level* area of the Data Mapping Overview report lists the top four data repositories (systems and data sources) that contain the most sensitive items. The bar chart for each system is divided into:

- Non-sensitive data
- Personal data
- Sensitive personal data

This data refreshes every two hours and can be manually refreshed.

Steps

1. To see the total number of items in each category, position your cursor over each section of the bar.
2. To filter results that will appear in the Investigation page, select each area in the bar and investigate further.

Review sensitive data and wide permissions

The *Sensitive Data and Wide Permissions* area of the Governance dashboard shows the counts for files that contain sensitive data and have wide permissions. The table shows the following types of permissions:

- From the most restrictive permissions to the most permissive restrictions on the horizontal axis.
- From the least sensitive data to the most sensitive data on the vertical axis.

Steps

1. To see the total number of files in each category, position your cursor over each box.
2. To filter results that will appear in the Investigation page, select a box and investigate further.

Review data listed by types of open permissions

The *Open Permissions* area of the Data Mapping Overview report shows the percentage for each type of permissions that exist for all files that are being scanned. The chart shows the following types of permissions:

- No Open Permissions
- Open to Organization
- Open to Public
- Unknown Access

Steps

1. To see the total number of files in each category, position your cursor over each box.
2. To filter results that will appear in the Investigation page, select a box and investigate further.

Review the age and size of data

You can investigate the items in the *Age* and *Size* graphs of the Data Mapping Overview report to see if there is any data you should delete or tier to less expensive object storage.

Steps

1. In the Age of Data chart, to see details about the age of the data, position your cursor over a point in the chart.
2. To filter by an age or size range, select that age or size.

   - **Age of Data graph** - Categorizes data based on the time it was created, the last time it was accessed, or the last time it was modified.
   - **Size of Data graph** - Categorizes data based on size.

   > ⓘ If any of your data sources implement data tiering, old data that already resides in object storage might be identified in the *Age of Data* graph.

View compliance details about the private data stored in your organization with NetApp Data Classification

Gain control of your private data by viewing details about the personal data (PII) and sensitive personal data (SPII) in your organization. You can also gain visibility by reviewing the categories and file types that NetApp Data Classification found in your data.

> ⓘ File-level compliance details are only available if you perform a full classification scan. Mapping-only scans don't yield file-level details.

By default, the Data Classification dashboard displays compliance data for all systems and databases. To see data for only some of the systems, select them.
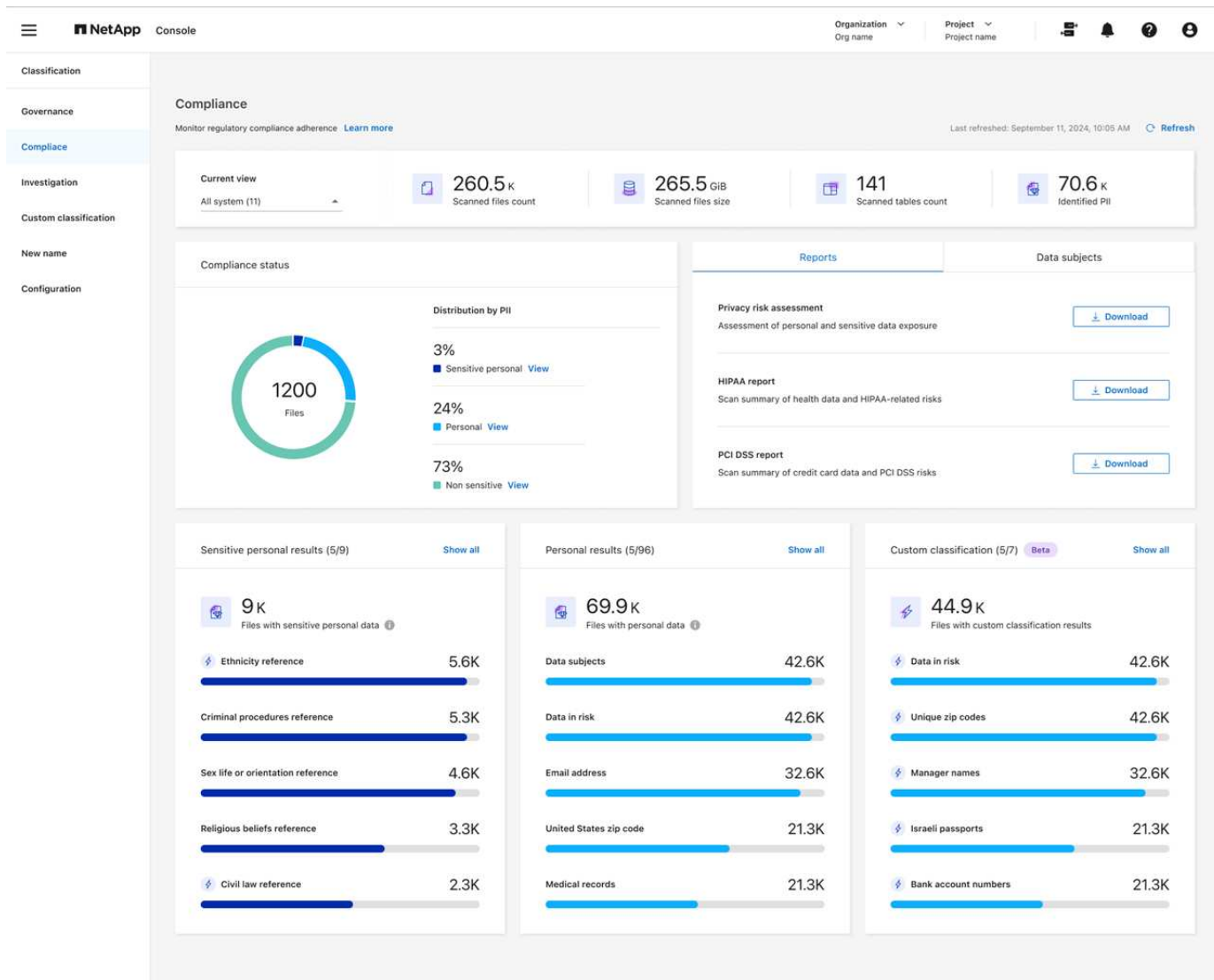
You can filter the results from the Data Investigation page and download a report of the results as a CSV file. See Filtering data in the Data Investigation page for details.

View files that contain personal data

Data Classification automatically identifies specific words, strings, and patterns (Regex) inside the data. For example, credit card numbers, social security numbers, bank account numbers, passwords, and more. Data Classification identifies this type of information in individual files, in files within directories (shares and folders), and in database tables.

You can also create custom search terms to identify personal data specific to your organization. For more information, see Create a custom classification.

For some types of personal data, Data Classification uses *proximity validation* to validate its findings. The validation occurs by looking for one or more predefined keywords in proximity to the personal data that was found. For example, Data Classification identifies a U.S. social security number (SSN) as an SSN if it sees a proximity word next to it—for example, *SSN* or *social security*. The table of personal data shows when Data Classification uses proximity validation.

Steps

1. From the Data Classification menu, select the **Compliance** tab.

2. To investigate the details for all personal data, select the icon next to the personal data percentage.
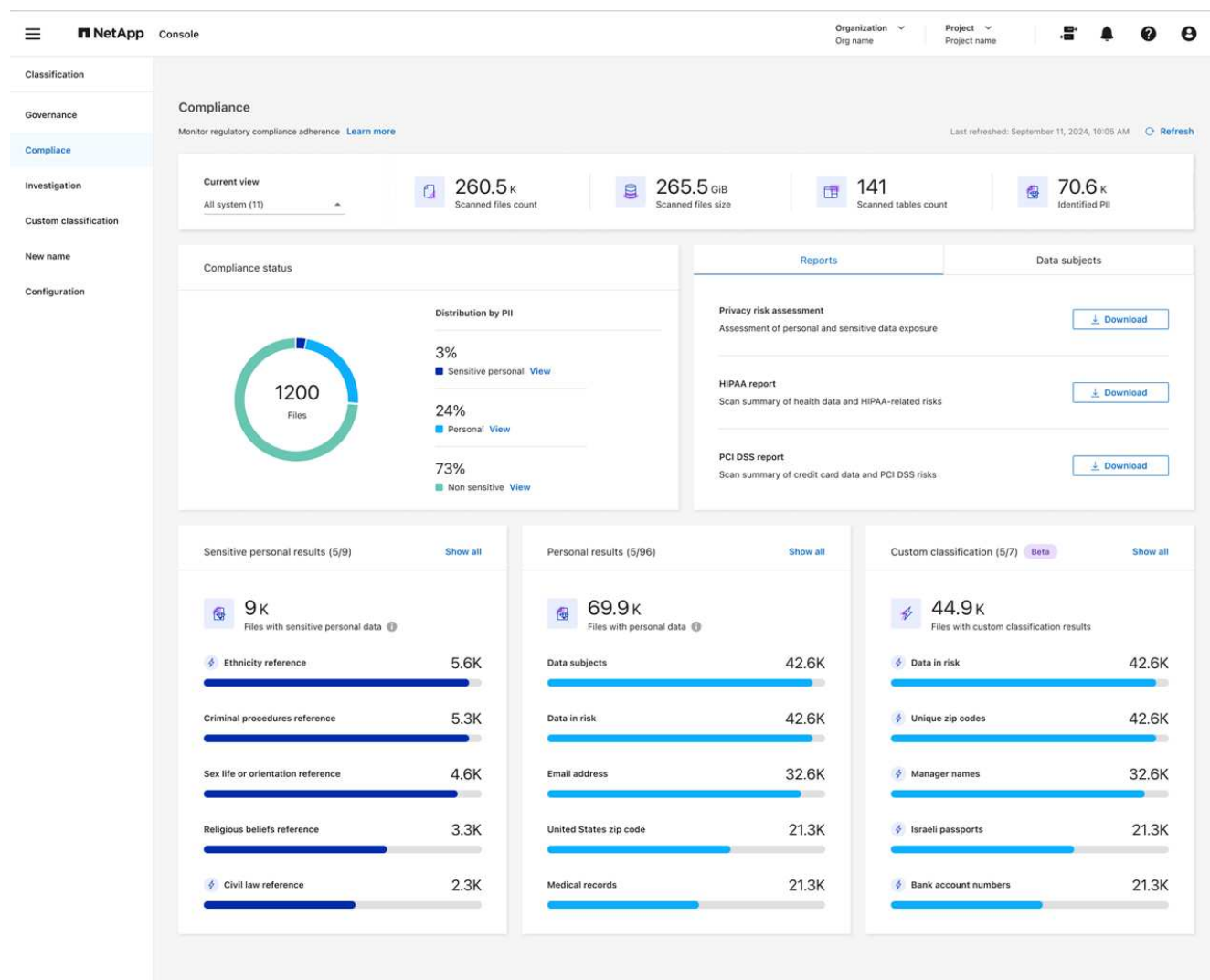
3. To investigate the details for a specific type of personal data, select **View All** and then select the **Investigate Results** arrow icon for a specific type of personal data, for example, email addresses.
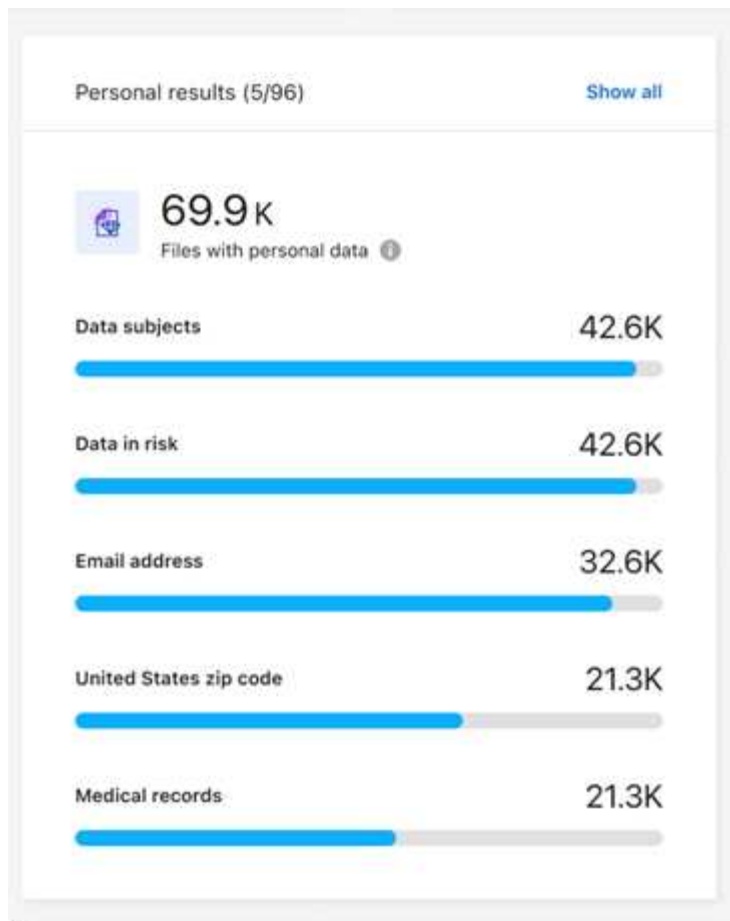
4. Investigate the data by searching, sorting, expanding details for a specific file, selecting the **Investigate Results** arrow to see masked information, or by downloading the file list.

   The following images show personal data found in a directory (shares and folders). In the **Structured** tab, you view personal data found in databases. In the **Unstructured** tab, you can view file-level data.

/benchmark_10TB_nfs_84/share_2_74/share_5/dir2/dir16/dir10/dir20/dir15/dir...   ⋮   ✕

Metadata

Directory type

Folder

System

NFS_Shares

System type

SHARES_GROUP

Storage repository

Path

/benchmark_10TB_nfs_84/share_...

Last accessed

2025-09-03

Last modified

2024-04-20

🏷️ **Tags**   Create tag

Open permissions

Open to organization

Discovered time

2025-10-03

View files that contain sensitive personal data

Data Classification automatically identifies special types of sensitive personal information, as defined by privacy regulations such as articles 9 and 10 of the GDPR. For example, information regarding a person's health, ethnic origin, or sexual orientation. See the full list. Data Classification identifies this type of information in individual files, in files within directories (shares and folders), and in database tables.

Data Classification uses AI, natural language processing (NLP), machine learning (ML), and cognitive computing (CC) to understand the meaning of the content that it scans in order to extract entities and categorize it accordingly.

For example, one sensitive GDPR data category is ethnic origin. Because of its NLP abilities, Data Classification can distinguish the difference between a sentence that reads "George is Mexican" (indicating sensitive data as specified in article 9 of the GDPR), versus "George is eating Mexican food."

	Only English is supported when scanning for sensitive personal data. Support for more languages will be added later.

Steps

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

2. To investigate the details for all sensitive personal data, locate the **Sensitive personal results** card then select **Show all**.
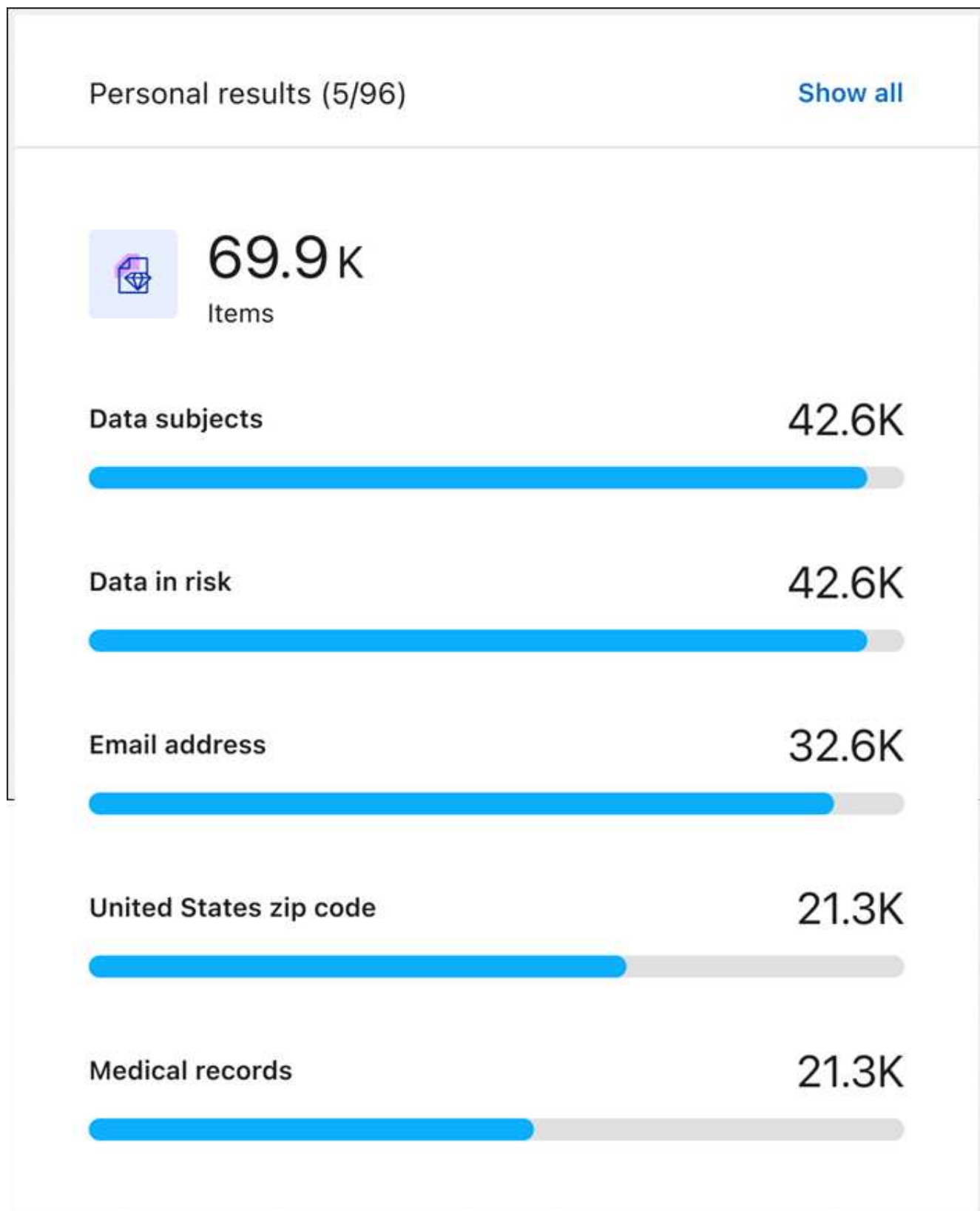
Personal results (5/96)   Show all

69.9 K
Items

Data subjects   42.6K

Data in risk   42.6K

Email address   32.6K

United States zip code   21.3K

Medical records   21.3K

.

3. To investigate the details for a specific type of sensitive personal data, select **View All** then select the **Investigate Results** arrow icon for a specific type of sensitive personal data.

4. Investigate the data by searching, sorting, expanding details for a specific file, clicking **Investigate Results** to see masked information, or by downloading the file list.

Categories of private data in NetApp Data Classification

There are many types of private data that NetApp Data Classification can identify in your volumes and databases.

Data Classification identifies two types of personal data:

- **Personally identifiable information (PII)**
- **Sensitive personal information (SPII)**

> If you need Data Classification to identify other private data types, such as additional national ID numbers or healthcare identifiers, contact your account manager.

Types of personal data

The personal data, or *personally identifiable information* (PII), found in files can be general personal data or national identifiers. The third column in the table below identifies whether Data Classification uses proximity validation to validate its findings for the identifier.

The languages in which these items can be recognized are identified in the table.

Type	Identifier	Proximity validation?	English	German	Spanish	French	Japanese
General	Credit card number	Yes	✓	✓	✓		✓
	Data Subjects	No	✓	✓	✓		
	Email Address	No	✓	✓	✓		✓
	IBAN Number (International Bank Account Number)	No	✓	✓	✓		✓
	IP Address	No	✓	✓	✓		✓
	Password	Yes	✓	✓	✓		✓

Type	Identifier	Proximity validation?	English	German	Spanish	French	Japanese
National Identifiers							

Type	Identifier	Proximity validation?	English	German	Spanish	French	Japanese

Type	Identifier	Proximity validation?	English	German	Spanish	French	Japanese

Type	Identifier	Proximity validation	English	German	Spanish	French	Japanese
	Corporate)						
	Latvian ID	Yes	✓	✓	✓		
	Lithuanian ID	Yes	✓	✓	✓		
	Luxembourg ID	Yes	✓	✓	✓		
	Maltese ID	Yes	✓	✓	✓		
	National Health Service (NHS) Number	Yes	✓	✓	✓		
	New Zealand Bank Account	Yes	✓	✓	✓		
	New Zealand Driver's License	Yes	✓	✓	✓		
	New Zealand IRD Number (Tax ID)	Yes	✓	✓	✓		
	New Zealand NHI (National Health Index) Number	Yes	✓	✓	✓		
	New Zealand Passport Number	Yes	✓	✓	✓		
	Polish ID (PESEL)	Yes	✓	✓	✓		
	Portuguese Tax Identification Number (NIF)	Yes	✓	✓	✓		
	Romanian ID (CNP)	Yes	✓	✓	✓		
	Singapore National Registration Identity Card (NRIC)	Yes	✓	✓	✓		
	Slovenian ID (EMSO)	Yes	✓	✓	✓		
	South African ID	Yes	✓	✓	✓		
	Spanish Tax Identification Number	Yes	✓	✓	✓		
	Swedish ID	Yes	✓	✓	✓		
	UK ID (NINO)	Yes	✓	✓	✓		
	USA California Driver's License	Yes	✓	✓	✓		
	USA Indiana Driver's License	Yes	✓	✓	✓		
	USA New York Driver's License	Yes	✓	✓	✓		
	USA Texas Driver's License	Yes	✓	✓	✓		
	USA Social Security Number (SSN)	Yes	✓	✓	✓		

Types of sensitive personal data

Data Classification can find the following sensitive personal information (SPII) in files.

The following SPII can currently only be recongized in English:

- **Criminal Procedures Reference**: Data concerning a natural person's criminal convictions and offenses.
- **Ethnicity Reference**: Data concerning a natural person's racial or ethnic origin.
- **Health Reference**: Data concerning a natural person's health.
- **ICD-9-CM Medical Codes**: Codes used in the medical and health industry.
- **ICD-10-CM Medical Codes**: Codes used in the medical and health industry.

- **Philosophical Beliefs Reference**: Data concerning a natural person's philosophical beliefs.
- **Political Opinions Reference**: Data concerning a natural person's political opinions.
- **Religious Beliefs Reference**: Data concerning a natural person's religious beliefs.
- **Sex Life or Orientation Reference**: Data concerning a natural person's sex life or sexual orientation.

Types of categories

Data Classification categorizes your data as follows.

Most of these categories can be recognized in English, German, and Spanish.

Category	Type	English	German	Spanish
Finance	Balance Sheets	✓	✓	✓
	Purchase Orders	✓	✓	✓
	Invoices	✓	✓	✓
	Quarterly Reports	✓	✓	✓
HR	Background Checks	✓		✓
	Compensation Plans	✓	✓	✓
	Employee Contracts	✓		✓
	Employee Reviews	✓		✓
	Health	✓		✓
	Resumes	✓	✓	✓
Legal	NDAs	✓	✓	✓
	Vendor-Customer contracts	✓	✓	✓
Marketing	Campaigns	✓	✓	✓
	Conferences	✓	✓	✓
Operations	Audit Reports	✓	✓	✓
Sales	Sales Orders	✓	✓	
Services	RFI	✓		✓
	RFP	✓		✓
	SOW	✓	✓	✓
	Training	✓	✓	✓
Support	Complaints and Tickets	✓	✓	✓

The following metadata is also categorized and identified in the same supported languages:

- Application Data
- Archive Files

- Audio

- Breadcrumbs from Data Classification
  Business Application Data

- CAD Files

- Code

- Corrupted

- Database and index files

- Design Files

- Email Application Data

- Encrypted (files with a high entropy score)

- Executables

- Financial Application Data

- Health Application Data

- Images

- Logs

- Miscellaneous Documents

- Miscellaneous Presentations

- Miscellaneous Spreadsheets

- Miscellaneous "Unknown"

- Password Protected files

- Structured Data

- Videos

- Zero-Byte Files

Types of files

Data Classification scans all files for category and metadata insights and displays all file types in the file types section of the dashboard. When Data Classification detects Personal Identifiable Information (PII) or when it performs a DSAR search, only the following file formats are supported:

```
.CSV, .DCM, .DOC, .DOCX, .JSON, .PDF, .PPTX, .RTF, .TXT, .XLS, .XLSX, Docs,
Sheets, and Slides
```

Accuracy of information found

NetApp can't guarantee 100% accuracy of the personal data and sensitive personal data that Data Classification identifies. You should always validate the information by reviewing the data.

Based on our testing, the table below shows the accuracy of the information that Data Classification finds. We break it down by *precision* and *recall*:

Precision

The probability that what Data Classification finds has been identified correctly. For example, a precision rate of 90% for personal data means that 9 out of 10 files identified as containing personal information,

actually contain personal information. 1 out of 10 files would be a false positive.

Recall

The probability for Data Classification to find what it should. For example, a recall rate of 70% for personal data means that Data Classification can identify 7 out of 10 files that actually contain personal information in your organization. Data Classification would miss 30% of the data and it won't appear in the dashboard.

We are constantly improving the accuracy of our results. Those improvements will be automatically available in future Data Classification releases.

Type	Precision	Recall
Personal data - General	90%-95%	60%-80%
Personal data - Country identifiers	30%-60%	40%-60%
Sensitive personal data	80%-95%	20%-30%
Categories	90%-97%	60%-80%

Create a custom classification in NetApp Data Classification

NetApp Data Classification enables you to create custom categories or personal identifiers to identify data specific to your organization's regulatory and compliance requirements.

Data Classification supports two types of custom classifiers: categories and personal identifiers. Custom categories are created based on a set of files you upload from which Data Classification creates an AI model to identify similar data in your organization (for example, a health research firm might create a clinical analysis category). Custom personal identifiers are created using keyword lists or a regular expression (regex) to identify information specific to your organization that can pose a compliance risk.

All custom classifications are available in the Custom classification dashboard.

Create a custom personal identifier

Data Classification enables you to create a custom personal identifier using either contextual keywords or a regular expression to identify data unique to your organization.
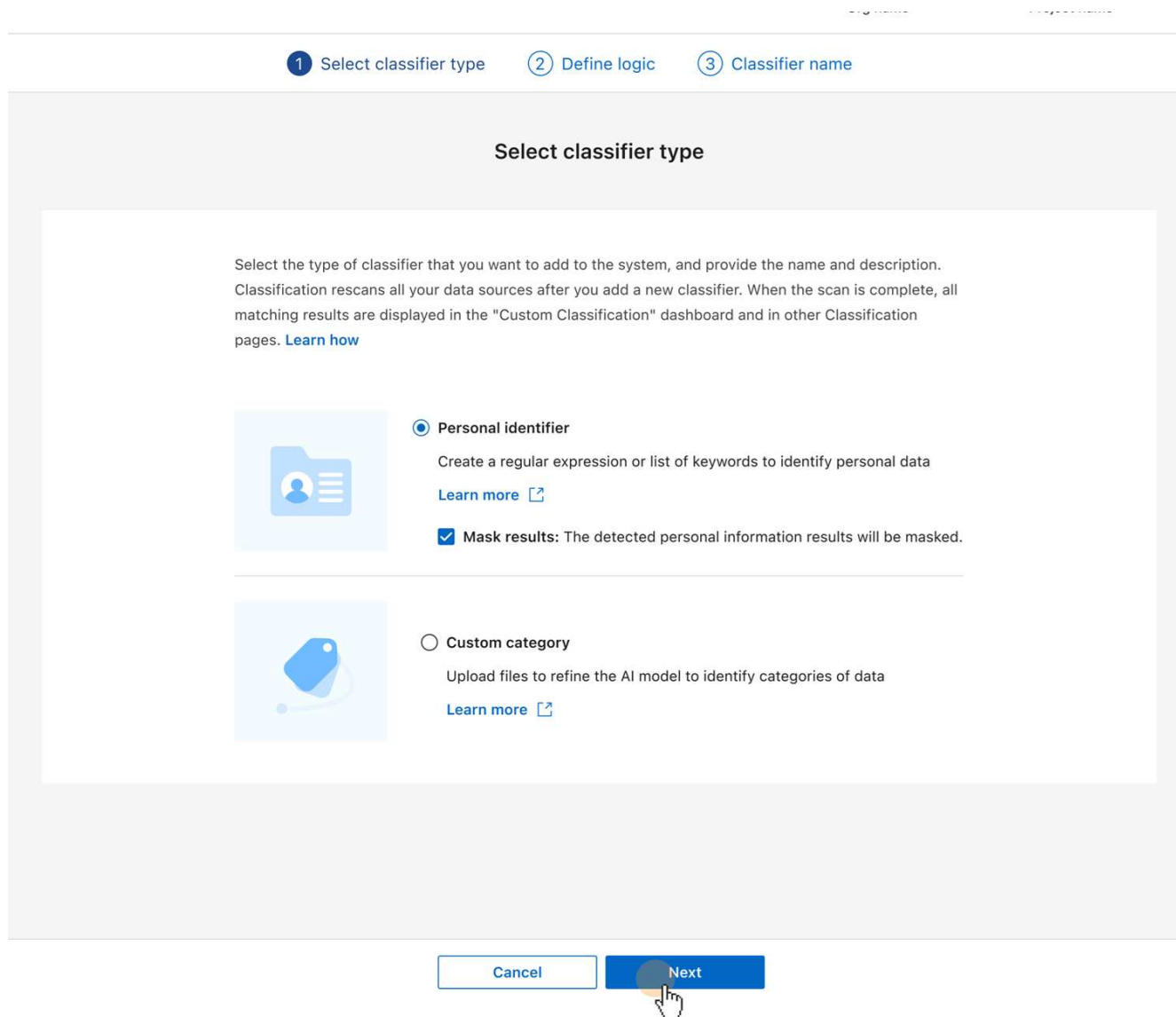
Requirements for keywords

If you're creating your personal identifier with a keyword list, the list must meet the following requirements:

- Keyword entries are case insensitive.
- Keywords must be at least three characters. Any words shorter than three characters are ignored.
- Duplicate words are only added once.
- The total list of keywords can't exceed 500,000 characters. The list must include at least one keyword.

Steps

1. Select the **Custom classification** tab.
2. Select **+ New Classifier** to create the custom classifier.

3. Select **Personal identifier**. Optionally, select **Mask results** to mask detected personal data.

4. Select **Next**.

5. To add the classifier with keywords, select **Keywords**. Enter a list of keywords, with each entry on a separate line. Ensure the keywords adhere to the requirements.

Define logic

Regular expression	Keywords ✓
Define a regular expression to identify patterns in your data.	Create a comprehensive list of keywords to effectively identify personal information.

Define the list of keywords for Data Classification to use for detection.

Custom keywords list

- Enter each keyword or phrase on a new line
- Keywords are not case sensitive
- Each word must be at least 3 characters long, Shorter words are ignored
- Duplicate words are only added once
- The total list of keywords cannot exceed 500,000 characters

Insert keywords

Validate

Cancel   Next

To add the classifier as a regular expression, select **Regular expression** then add a pattern to detect the specific information of your data. Select **Validate** to confirm the syntax of your entry.

Define logic

Regular expression Define a regular expression to identify patterns in your data. ✓	**Keywords** Create a comprehensive list of keywords to effectively identify personal information.

Classifier regular expression

Create the regular expression used to identify data. Optionally, add proximity words to enhance detection. Add the regular expression to identify information in your data

Example: to identify a 12-digit number that begins with 201, the expression is \b201\d{9}\b.

> ^\d{7,9}

[Validate]   ● Regular expression is valid.

Test your regular expression: Enter a string to instantly see if it matches your regex pattern

> 456

[Test]

☐ **Add proximity words**

To improve the detection accuracy, insert phrases that must appear around the regular expression's match.Enter any phrases that must appear adjacent to the regular expression. Separate entries with a line break.

> Insert proximity words (optional)

[Cancel]   [Next]

a. Optionally, enter a sample string that should match your regex pattern then select **Test** to check it.

b. Optionally, add proximity words. If you add proximity words, Data Classification only flags the regex pattern if the proximity words are adjacent to the matching string.

6. Select **Next**.

7. Enter a **Classifier name** and a **Description** to identify the custom category in your dashboard.

8. Select **Save** to create the custom personal identifier.

After you create a custom personal identifier, its results are captured in the next scheduled scan. To capture results sooner, perform an on-demand scan. To view results, see Generate compliance reports.

Create a custom category

With custom categories, you can categorize data specific to your organization. Custom categories are created based on text files that you upload from which Data Classification creates an AI model to identify similar information in other files.

Training data requirements

- The training dataset must contain a minimum of 25 files. The maximum file count is 1,000.

- All files must be located directly in the file path that you provide.

- All files must be larger than 100 bytes.

- Data Classification training data must be one of the following file types: CSV, DOCX, DOC, GZ, JSON, PDF, PPTX, TXT, RTT, XLS, or XLSX. You can upload a combination of all the supported file types.

Steps

1. In NetApp Data Classification, select **Custom classification**.

2. Select **+ New classifier**.

3. Choose **Custom category** as your classifier type then **Next**.

4. Define the logic for your custom category with a collection of text-based files. Provide the IP address of the **Working address** then select the **Volume** from the dropdown menu.

   Enter the **Directory path** for the directory that contains the training data.

5. Select **Load files** for Data Classification to perform a check of the files. You can review the summary of the files, which lists the file name, size, type, and notes if the file was deemed acceptable for training.

Working environment

PWwork_2 ▼

Volume

PWwork_2 ▼

Directory path

NFS: Hostname:/SHARE-PATH (e.g. 172.31.134.172:/jianni_nfs2_150GB   Load files

Items (500)   Change path

● 2 files failed to load   ✕

● 498 files loaded successfully   ✕

File name	Size ⇕	Type ⇕	Reliability ⇕	Included in training ⇕
Contract_v2.docx	415 KB	DOCX	⊘	⊘
RevenueReport_...	256 KB	PDF	⊗	Unsupported file type. Please provide a text file.
Report_Q4_Final...	1.2 MB	TXT	⊗	⊗
Q4_Final_Revised...	89 KB	CSV	⊘	⊘
HRReport_Final_...	640 KB	HTML	⊘	⊘

Cancel   Next

    a. To change the file path or re-upload files, select **Change path** then enter the data and load the files again.

6. When you're satisfied with the files uploaded, select **Next**.

7. Enter a **Classifier name** and a **Description** to identify the custom category in your dashboard.

8. Select **Save** to create the custom category.

Result

After you create a custom category, its results are captured in the next scheduled scan. To capture results sooner, manually initiate the scan.

Edit a custom classifier

You can modify the logic of a personal identifier after you create it. You can't change the type of the personal identifier or the logic type; for example, you can't change a custom category to a custom personal identifier. You also can't change a keyword-based custom identifier to a regex-based custom identifier.

Steps

1. In NetApp Data Classification, select **Custom classification**.

2. Identify the classifier you want to delete, then select the action menu … at the end of its row.

3. Select **Edit logic**.

4. If you're modifying keywords, add, delete, or edit the appropriate keywords. If you're modifying a regular expression, enter the new regular expression and validate it. Optionally, add proximity keywords.

5. Select **Save** to apply the changes.

Delete a custom classifier

1. In NetApp Data Classification, select **Custom classification**.

2. Identify the classifier you want to delete then select the action menu … at the end of its row.

3. Select **Delete classifier**.

Next steps

- Generate compliance reports

Investigate the data stored in your organization with NetApp Data Classification

The Data Investigation dashboard displays file and directory-level insights into your data, enabling you to sort and filter results. The Data Investigation page presents insights into file and directory metadata and permissions as well as identifying duplicate files. With file-, directory-, and database-level insights, you can take actions to improve the compliance of your organization and save storage space. The Data Investigation page also supports moving, copying, and deleting files.

(i)	To gain insights from the Investigation page, you must perform a full classification scan on your data sources. Data sources that have had a mapping-only scan do not show file-level details.

Data investigation structure

The Data Investigation page sorts data into three tabs:

- **Unstructured data**: file data
- **Directories**: folders and file shares
- **Structured**: database

Data filters

The Data Investigation page provides numerous filters to sort through your data so you can what you need. You can use multiple filters in concert.

To add a filter, select the **Add filter** button.

Data investigation

Classifiers scan and tag your items. Use classifiers to identify sensitive data. Learn more

Filters: Sensitivity level: All ✕ | Open permissions: All ✕ | Created time: (Include) Open permissions , +3 ✕

Last accessed : (Includes) 3-5 years , +2 ✕ | File hash : (Includes) 78bb33f1e8d9006595b874a0a75ecf36 ✕ | Last modified : (Includes) 3-5 years , +1 ✕

Save query   Clear filters

＋ Add filters

	120 Items with sensitive data and open permissions		120 Items with sensitive data		50 Recently accessed sensitive data		45 Stale Items

Add as filter   Add as filter   Add as filter   ✓ All results match

Unstructured (500)	Directories (200)	Structured (80)

Items (500) | 3 TiB

	Name	Last modified	Personal	Sensitive personal	Data subjects	File type	
☐	HR_Listworkprogrem.TXT	Feb 2, 2019 07:28 PM	322	89	101	DOC	
☐	Education report.PDF	Mar 20, 2019 11:14 PM	189	12	89	PDF	
☐	Work program>1.PNG	Dec 4, 2019 09:42 PM	956	80	702	TXT	
☐	Ethics consult.DOCX	Dec 4, 2019 09:42 PM	380	0	622	PDF	

Filter sensitivity and content

Use the following filters to view how much sensitive information is contained in your data.

Filter	Details
Category	Select the types of categories.
Sensitivity Level	Select the sensitivity level: Personal, Sensitive personal, or Non sensitive.
Number of identifiers	Select the range of detected sensitive identifiers per file. Includes personal data and sensitive personal data. When filtering in Directories, Data Classification totals the matches from all files in each folder (and sub-folders). NOTE: The December 2023 (version 1.26.6) release removed the option to calculate the number of personal identifiable information (PII) data by Directories.
Personal Data	Select the types of personal data.
Sensitive Personal Data	Select the types of sensitive personal data.
Data Subject	Enter a data subject's full name or known identifier. Learn more about data subjects here.

Filter user owner and user permissions

Use the following filters to view file owners and permissions to access your data.

Filter	Details
Open Permissions	Select the type of permissions within the data and within folders/shares.

Filter	Details
User / Group Permissions	Select one or multiple user names and/or group names, or enter a partial name.
File Owner	Enter the file owner name.
Number of users with access	Select one or multiple category ranges to show which files and folders are open to a certain number of users.

Filter chronologically

Use the following filters to view data based on time criteria.

Filter	Details
Created Time	Select a time range when the file was created. You can also specify a custom time range to further refine the search results.
Discovered Time	Select a time range when Data Classification discovered the file. You can also specify a custom time range to further refine the search results.
Last Modified	Select a time range when the file was last modified. You can also specify a custom time range to further refine the search results.
Last Accessed	Select a time range when the file or directory* was last accessed. You can also specify a custom time range to further refine the search results. For the types of files that Data Classification scans, this is the last time Data Classification scanned the file.

* Last accessed time for a directory is only available for NFS or CIFS shares.

Filter metadata

Use the following filters to view data based on location, size, and directory or file type.

Filter	Details
File Path	Enter up to 20 partial or full paths that you want to include or exclude from the query. If you enter both include paths and exclude paths, Data Classification finds all files in the included paths first, then it removes files from excluded paths, and then it displays the results. Note that using "*" in this filter has no effect, and that you can't exclude specific folders from the scan - all the directories and files under a configured share will be scanned.
Directory Type	Select the directory type; either "Share" or "Folder".
File Type	Select the types of files.
File Size	Select the file size range.
File Hash	Enter the file's hash to find a specific file, even if the name is different.

Filter storage type

Use the following filters to view data by storage type.

Filter	Details
System type	Select the type of system.
System environment name	Select specific systems.
Storage Repository	Select the storage repository, for example, a volume or a schema.

Filter query

Use the following filter to view data by saved queries.

Filter	Details
Saved query	Select one saved query or multiples. Go to the saved queries tab to view the list of existing saved queries and create new ones.
Tags	Select the tag or tags that are assigned to your files.

Filter analysis status

Use the following filter to view data by the Data Classification scan status.

Filter	Details
Analysis Status	Select an option to show the list of files that are Pending First Scan, Completed being scanned, Pending Rescan, or that have Failed to be scanned.
Scan Analysis Event	Select whether you want to view files that were not classified because Data Classification couldn't revert last accessed time, or files that were classified even though Data Classification couldn't revert last accessed time.

See details about the "last accessed time" timestamp for more information about the items that appear in the Investigation page when filtering using the Scan Analysis Event.

Filter data by duplicates

Use the following filter to view files that are duplicated in your storage.

Filter	Details
Duplicates	Select whether the file is duplicated in the repositories.

View file metadata

In addition to showing you the system and volume where the file resides, the metadata shows much more information, including the file permissions, file owner, and whether there are duplicates of this file. This information is useful if you're planning to create saved queries because you can see all the information that you can use to filter your data.

The availability of information depends on the data source. For example, volume name and permissions are not shared for database files.

Steps

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

2. In the Data Investigation list on the right, select the down-caret ﹀ on the right for any single file to view the file metadata.

HR_List Long name for a file that no o... .TXT

Sensitive data

Personal (322) >   Sensitive personal (89) >   Data subjects (102) >

Metadata

Working environment

\\00.000.0.01\cifs_system_name

Storage repository (share)

\\00.000.0.01\cifs_system_name

File path

\\00.000.0.01\cifs_system_name

File size

26.92 KiB

File type

PDF

Created time

2025-10-06 12:34

Storage repository (share)

\\00.000.0.01\cifs_system_name

Last modified

Tags

Reliability   Security

Protection and security

Permissions

No open permissions   View permissions

File owner

\\00.000.0.01\cifs_system_name   View details

Duplicates

1412   View details

3. Optionally, you can create or add a tag to the file with the **Create tag** button. Select an existing tag from the dropdown menu or add a new tag with the **+ Add** button. Tags can be used to filter data.

View user permissions for files and directories

To view a list of all users or groups who have access to a file or to a directory and the types of permissions they have, select **View all Permissions**. This option is available only for data in CIFS shares.

If you security identifiers (SIDs) instead of user and group names, you should integrate your Active Directory into Data Classification. For more information, see add Active Directory to Data Classification.

Steps

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

2. In the Data Investigation list on the right, select the down-caret ⌄ on the right for any single file to view the file metadata.

3. To view a list of all users or groups who have access to a file or to a directory and the types of permissions they have, in the Open Permissions field, select **View all Permissions**.

   > ℹ️ Data Classification shows up to 100 users in the list.

4. Select the down-caret ⌄ button for any group to see the list of users who are part of the group.

   > 💡 You can expand one level of the group to see the users who are part of the group.

5. Select the name of a user or group to refresh the Investigation page so you can see all the files and directories that the user or group has access to.

Check for duplicate files in your storage systems

You can check whether duplicate files are being stored in your storage systems. This is useful if you want to identify areas where you can save storage space. It's also good to ensure certain files that have specific permissions or sensitive information are not unnecessarily duplicated in your storage systems.

Data Classification compares all files (excluding databases) for duplicates if they are:

- 1 MB or lager
- Or contain personal or sensitive personal information

Data Classification uses hashing technology to determine duplicate files. If any file has the same hash code as another file, the files are exact duplicates even if the file names are different.

Steps

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

2. In Filter pane, select "File Size" along with "Duplicates" ("Has duplicates") to see which files of a certain size range are duplicated in your environment.

3. Optionally, download the list of duplicate files and send it to your storage administrator so they can decide which files, if any, can be deleted.

4. Optionally, you can delete, tag, or move the duplicate files. Select the files you want to perform an action on, then select the appropriate action.

View if a specific file is duplicated

You can see if a single file has duplicates.

Steps

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

2. In the Data Investigation list, select ⌄ on the right for any single file to view the file metadata.

   If duplicates exist for a file, this information appears next to the *Duplicates* field.

3. To view the list of duplicate files and where they are located, select **View Details**.

4. In the next page select **View Duplicates** to view the files in the Investigation page.

5. Optionally, you can delete, tag, or move the duplicate files. Select the files you want to perform an action on, then select the appropriate action.

> You can use the "file hash" value provided in this page and enter it directly in the Investigation page to search for a specific duplicate file at any time - or you can use it in a saved query.

Download your report

You can download your filtered results in a CSV or JSON format.

There can be up to three report files downloaded if Data Classification is scanning files (unstructured data), directories (folders and file shares), and databases (structured data).

The files are split into files with a fixed number of rows or records:

- JSON: 100,000 records per report that takes about 5 minutes to generate
- CSV: 200,000 records per report that takes about 4 minutes to generate

> You can download a version of the CSV file to view in this browser. This version is limited to 10,000 records.

What's included in the downloadable report

The **Unstructured Files Data Report** includes the following information about your files:

- File name
- Location type
- System name
- Storage repository (for example, a volume, bucket, shares)
- Repository type
- File path
- File type
- File size (in MB)
- Created time
- Last modified
- Last accessed
- File owner
    - File owner data encompasses account name, SAM account name, and e-mail address when Active

Directory is configured.

- Category

- Personal information

- Sensitive personal information

- Open permissions

- Scan Analysis Error

- Deletion detection date

  The deletion detection date identifies the date that the file was deleted or moved. This enables you to identify when sensitive files have been moved. Deleted files don't contribute to the file number count that appears in the dashboard or on the Investigation page. The files only appear in the CSV reports.

The **Unstructured Directories Data Report** includes the following information about your folders and file shares:

- System type

- System name

- Directory name

- Storage repository (for example, a folder or file shares)

- Directory owner

- Created time

- Discovered time

- Last modified

- Last accessed

- Open permissions

- Directory type

The **Structured Data Report** includes the following information about your database tables:

- DB Table name

- Location type

- System name

- Storage repository (for example, a schema)

- Column count

- Row count

- Personal information

- Sensitive personal information

Steps to generate the report

1. From the Data Investigation page, select the ⬇️ button on the top, right of the page.

2. Choose the report type: CSV or JSON.

3. Enter a **Report name**.

4. To download the complete report, select **System** then choose the **System** and **Volume** from the respective dropdown menus. Provide a **Destination folder path**.

   To download the report in the browser, select **Local** . Note this option limits the report to the first 10,000 rows and is limited to the **CSV** format. You don't need to complete any other fields if you select **Local**.

5. Select **Download Report**.

Download investigation report

Report type

◉ CSV report   ○ JSON report

Report name

investigation_report

Export destination

◉ System   ○ Local (limited to 10K rows)

Working system

PWwork_2 ▼

Volume

PL_D ▼

Destination folder path

NFS: Hostname:/SHARE-PATH (e.g. 172.31.134.172:/jianni_nfs2_150GB) ▼

Estimated report size: 20 MB

ⓘ **Notice:** File is too big and will be spilt into multiple items

Download report   **Cancel**

Result

A dialog displays a message that the reports are being downloaded.

Create a saved query based on selected filters

Steps

1. In the Investigation tab, define a search by selecting the filters you want to use. See Filtering data in the Investigation page for details.

2. Once you have all the filter characteristics set to your liking, select **Save query**.

3. Name the saved query and add a description. The name must be unique.
4. You can optionally save the query as policy:

   a. To save the query as a policy, switch the **Run as a policy** toggle.

   b. Choose to **Delete permanently** or **Send email updates**. If you choose email updates, you can email the query results to *all* Console users at daily, weekly, or monthly. Alternately, you can send the notification to specific email address at the same frequencies.

5. Select **Save**.

Name this query

Name

Stale sensitive date

Description   Optional

Give a short description here

0/500

⬤━ Run as a policy

Select one or more actions for the guardrail to perform on files and objects when conditions are met. **More** ⬀

○ Delete permanently

○ Send email updates

☐ About this query to all console users on this account every Day ▾

☐ Notification emails   Day ▾   to   Enter email here

Save   **Cancel**

Once you've created the search or policy, you can view it in the **Saved queries** tab.

💡 | It can take up to 15 minutes for the results to appear on the Saved Queries page.

Manage saved queries with NetApp Data Classification

NetaApp Data Classification supports saving your search queries. With a saved query, you can create custom filters to sort through frequent queries of your data Investigation page. Data Classification also includes predefined saved queries based on common requests.

The **Saved queries** tab in the Compliance dashboard lists all the predefined and custom saved queries available on this instance of Data Classification.

Saved queries can also be saved as **policies**. Whereas queries filter data, policies allow you to act on the data. With a policy: you can delete discovered data or send email updates about the discovered data.

Saved queries also appear in the list of filters in the Investigation page.

View saved queries results in the Investigation page

To display the results for a saved query in the Investigation page, select the ⋮ button for a specific search then select **Investigate Results**.

Create saved queries and policies

You can create your own custom saved queries that provide results for queries specific to your organization. Results are returned for all files and directories (shares and folders) that match the search criteria.

Steps

1. In the Investigation tab, define a search by selecting the filters you want to use. See Filtering data in the Investigation page for details.

2. Once you have all the filter characteristics set to your liking, select **Save query**.

3. Name the saved query and add a description. The name must be unique.

4. You can optionally save the query as policy:

   a. To save the query as a policy, switch the **Run as a policy** toggle.

   b. Choose to **Delete permanently** or **Send email updates**. If you choose email updates, you can email the query results to *all* Console users at daily, weekly, or monthly. Alternately, you can send the notification to specific email address at the same frequencies.

5. Select **Save**.

Name this query

Beta

Name

Stale sensitive date

Description   Optional

Give a short description here

                                                                            0/500

Run as a policy

Select one or more actions for the guardrail to perform on files and objects when conditions are met. **More**

○ Delete permanently

○ Send email updates

☐ About this query to all console users on this account every Day

☐ Notification emails   Day   to   Enter email here

Save   **Cancel**

Once you've created the search or policy, you can view it in the **Saved queries** tab.

Edit saved queries or policies

You can modify the name and description of a saved query. You can also convert a query to a policy and vice

versa.

You cannot modify default saved queries. You cannot modify the filters of a saved query. You can alternately view the investigation results of a saved query, change or modify the filters, then save it as a new query or policy.

Steps

1. From the Saved queries page, select **Edit Search** for the search that you want to change.

Personal data – High risk	Query	Predefined	Read access	Files with over 20 personal data identifiers, or DB columns with ove...	914.2K	View	⋯
PopPop	Policy	Custom	Email update	popop			⬚ Investigate results
Private data – Stale over 7 years	Query	Predefined	Read access	Files containing personal or sensitive personal information, last mo...			✎ Edit query

2. Make the changes to the name and description fields. To only change the name and description fields.

   You can optionally convert the query to a policy or convert the policy to a saved query. Switch the **Run as a policy** toggle as needed.
   .. If you're converting the query to a policy, choose to **Delete permanently** or **Send email updates**. If you choose email updates, you can email the query results to *all* Console users at daily, weekly, or monthly. Alternately, you can send the notification to specific email address at the same frequencies.

3. Select **Save** to complete the changes.

Delete saved queries

You can delete any custom saved query or policy if you no longer need it. You can't delete default saved queries.

To delete a saved query, select the ⋮ button for a specific search, select **Delete query**, then select **Delete query** again in the confirmation dialog.

Default queries

Data Classification provides the following system-defined search queries:

- **Data Subject names - High risk**

  Files with more than 50 data subject names

- **Email Addresses - High risk**

  Files with more than 50 email addresses or database columns with more than 50% of their rows containing email addresses

- **Personal data - High risk**

  Files with more than 20 personal data identifiers or database columns with more than 50% of their rows containing personal data identifiers

- **Private data - Stale over 7 years**

  Files containing personal or sensitive personal information, last modified more than 7 years ago

- **Protect - High**

  Files or database columns that contain a password, credit card information, IBAN number, or social security number

- **Protect - Low**

  Files that have not been accessed for more than 3 years

- **Protect - Medium**

  Files that contain files or database columns with personal data identifiers including ID numbers, tax identification numbers, drivers license numbers, medicinal IDs, or passport numbers

- **Sensitive Personal data - High risk**

  Files with more than 20 sensitive personal data identifiers or database columns with greater than 50% of their rows containing sensitive personal data

Change the NetApp Data Classification scan settings for your repositories

You can manage how your data is being scanned in each of your systems and data sources. You can make the changes on a "repository" basis; meaning you can make changes for each volume, schema, user, etc. depending on the type of data source you are scanning.

Some of the things you can change are whether a repository is scanned or not, and whether NetApp Data Classification is performing a mapping scan or a mapping & classification scan. You can also pause and resume scanning, for example, if you need to stop scanning a volume for a period of time.

View the scan status for your repositories

You can view the individual repositories that NetApp Data Classification is scanning (volumes, buckets, etc.) for each system and data source. You can also see how many have been "Mapped", and how many have been "Classified". Classification takes longer because the full AI identification is being performed on all data.

You can view the scanning status of each work environment on the Configuration page:

- **Initializing** (light blue dot): The map or classify configuration is activated. This appears briefly before transitioning to the "pending queue" status.
- **Pending queue** (orange dot): The scan task is waiting to be listed in the scanning queue.
- **Queued** (orange dot): The task was successfully added to the scanning queue. The system will start mapping or classifying the volume when its turn in the queue arrives.
- **Running** (green dot): The scan task, which was in the queue, is actively in progress on the selected storage repository.
- **Finished** (green dot): The scan of the storage repository is complete.
- **Paused** (gray dot): You paused scanning. Although the changes in the volume are not displayed in the system, the scanned insights remain available.

- **Error** (red dot): The scan cannot complete because it has encountered issues. If you need to complete an action, the error appears in the tooltip under the "Required action" column. Otherwise, the system shows an "error" status and tries to recover. When it finishes, the status changes.

- **Not scanning**: The volume configuration of "Off" was selected and the system is not scanning the volume.

Steps

1. From the Data Classification menu, select **Configuration**.
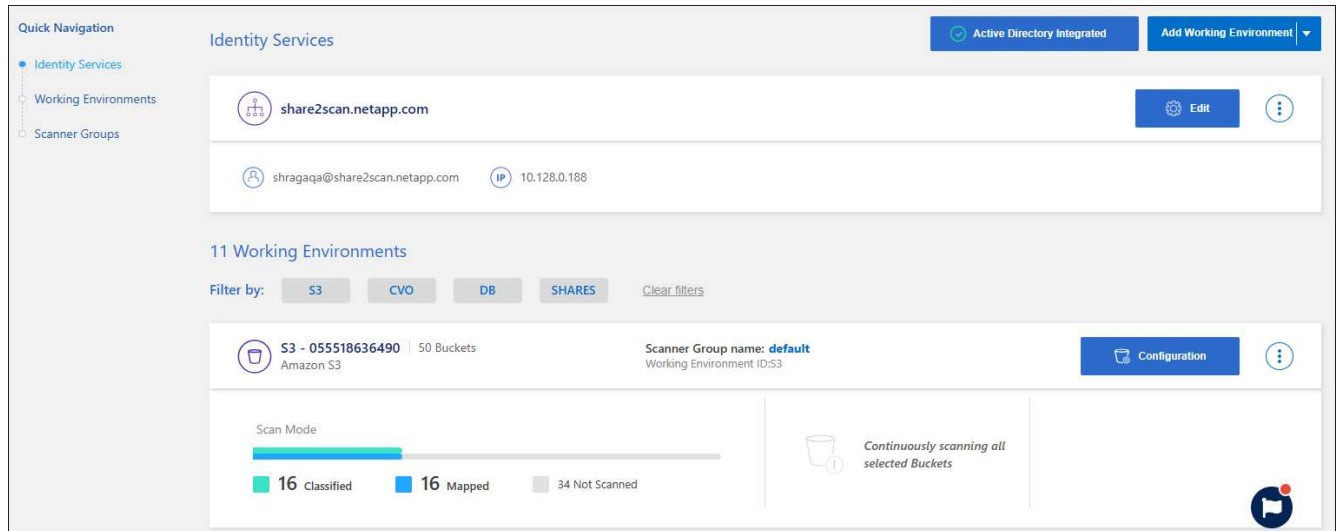
2. From the Configuration tab, select the **Configuration** button for the system.

3. In the Scan Configuration page, view the scan settings for all repositories.
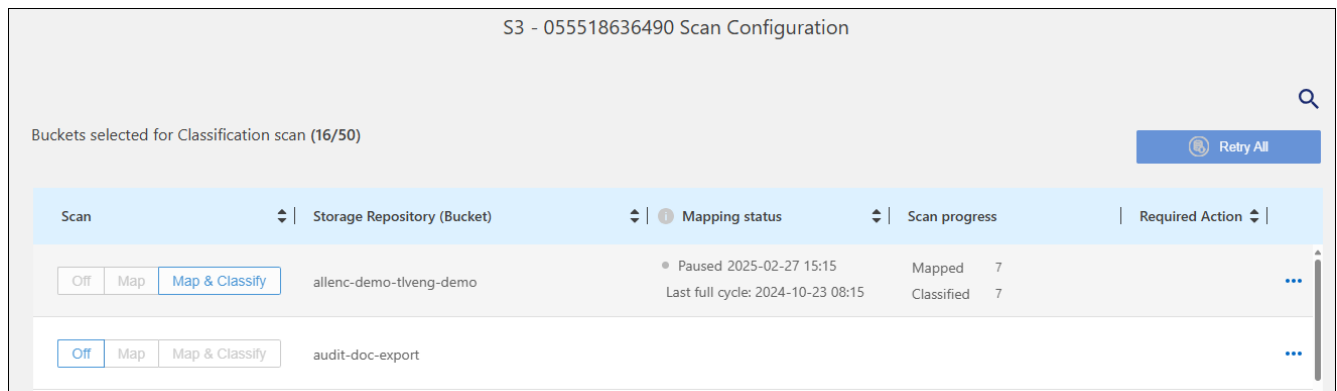
4. During a scan, hover your cursor over the progress bar in the *Mapping status* column to view the number of files in the queue to be mapped or classified for that repository.

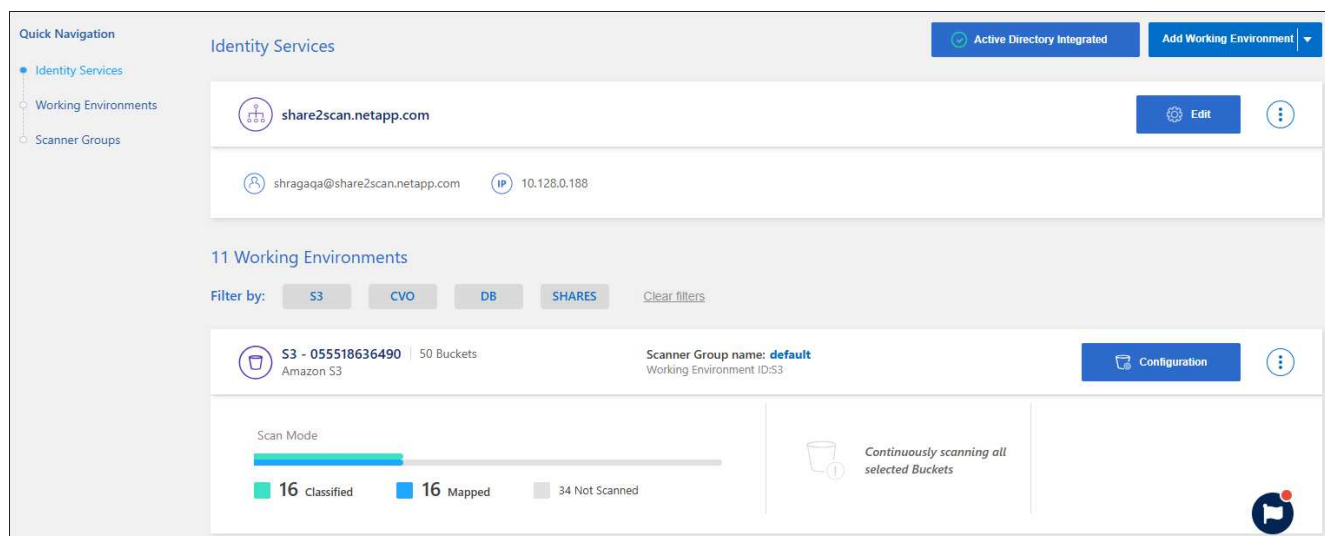## Change the type of scanning for a repository

You can start or stop mapping-only scans, or mapping and classification scans, in a system at any time from the Configuration page. You can also change from mapping-only scans to mapping and classification scans, and vice-versa.

> Databases can't be set to mapping-only scans. Database scanning can be Off or On; where On is equivalent to Map & Classify.

Steps

1. From the Data Classification menu, select **Configuration**.
2. From the Configuration tab, select the **Configuration** button for the system.


3. In the Scan Configuration page, change any of the repositories (buckets in this example) to perform **Map** or **Map & Classify** scans.
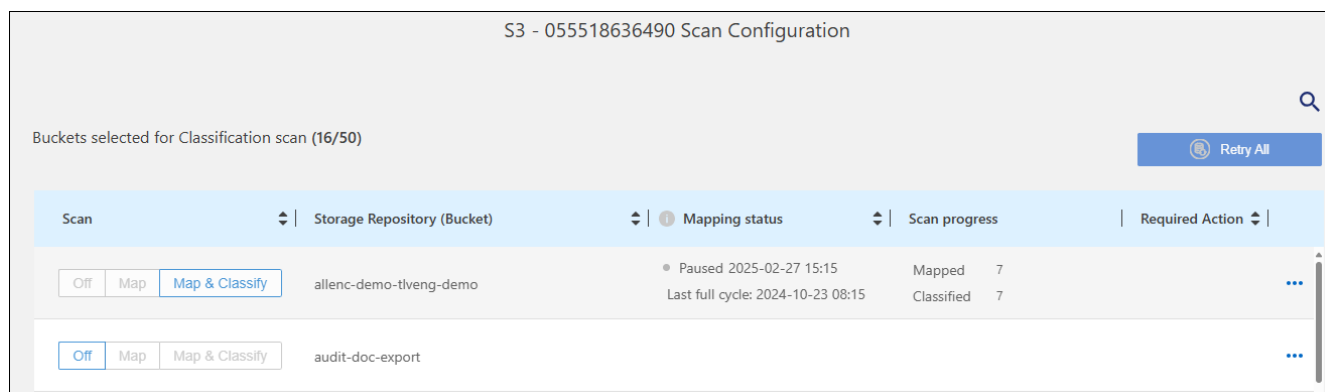
Certain types of systems enable you to change the type of scanning globally for all repositories using a button bar at the top of the page. This is valid for Cloud Volumes ONTAP, on-premises ONTAP, Azure NetApp Files, and Amazon FSx for ONTAP systems.

The example below shows this button bar for an Azure NetApp Files system.
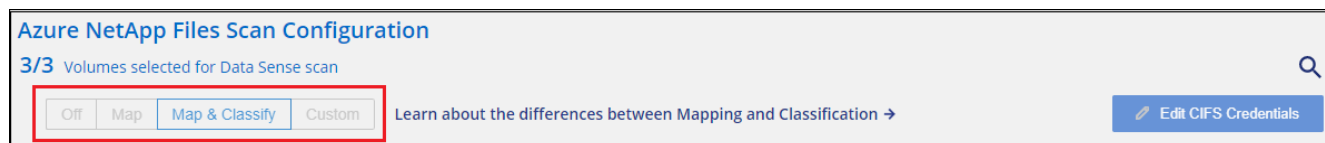
Prioritize scans

You can prioritize the most important mapping-only scans or map & classify scans to ensure high priority scans are completed first.

By default, scans are queued based on the order in which they are initiated. With the ability to prioritize scans, you can move scans to the front of the queue. Multiple scans can be prioritized. Priority is designated in a first-

in, first-out order, meaning the first scan you prioritize moves to the front of the queue; the second scan you prioritize becomes second in the queue, and so forth.

Priority is granted on a one-time basis. Automatic rescans of mapping data occur in the default order.

Steps

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

2. Select the resources you want to prioritize.

3. From the Actions … option, select **Prioritize scan**.
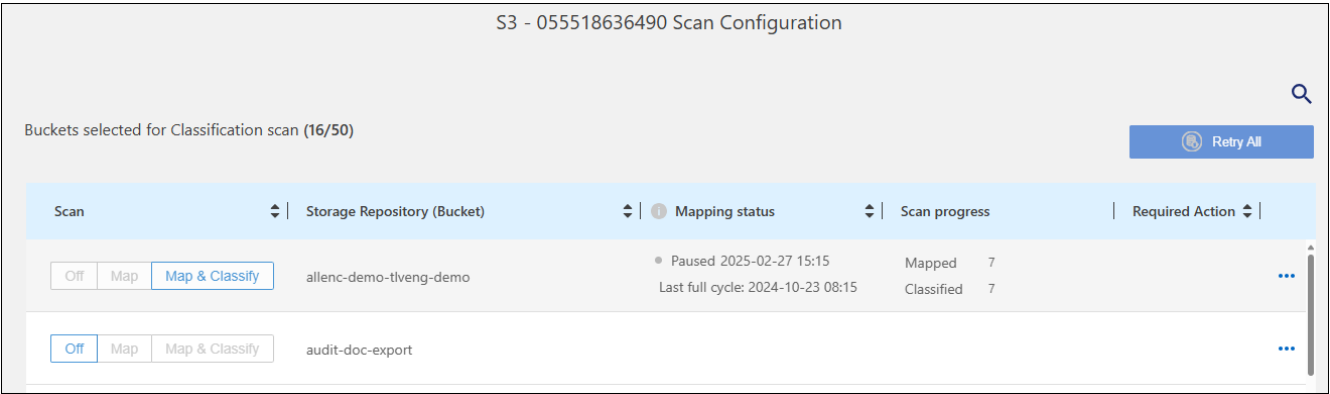
Stop scanning for a repository

You can stop scanning a repository (for example, a volume) if you no longer need to monitor it for compliance. You do this by turning scanning "off". When scanning is turned off, all the indexing and information about that volume is removed from the system, and charging for scanning the data is stopped.

Steps

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

2. From the Configuration tab, select the **Configuration** button for the system.
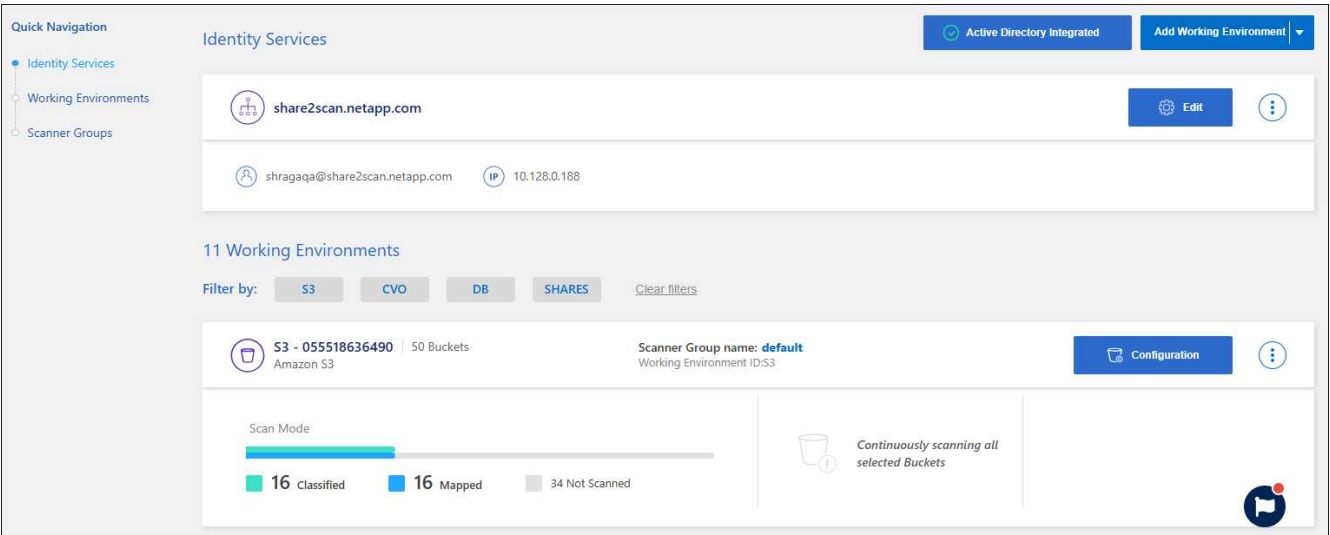
3. In the Scan Configuration page select **Off** to stop scanning for a particular bucket.

Pause and resume scanning for a repository

You can "pause" scanning on a repository if you want to temporarily stop scanning certain content. Pausing scanning means that Data Classification won't perform any future scans for changes or additions to the repository. All current scan results remain accessible in Data Classification.

If you pause scans, it does not eliminate billing charges because the data is still on the system.

You can resume scanning at any time.

Steps

1. From the Data Classification menu, select **Configuration**.
2. From the Configuration tab, select the **Configuration** button for the system.
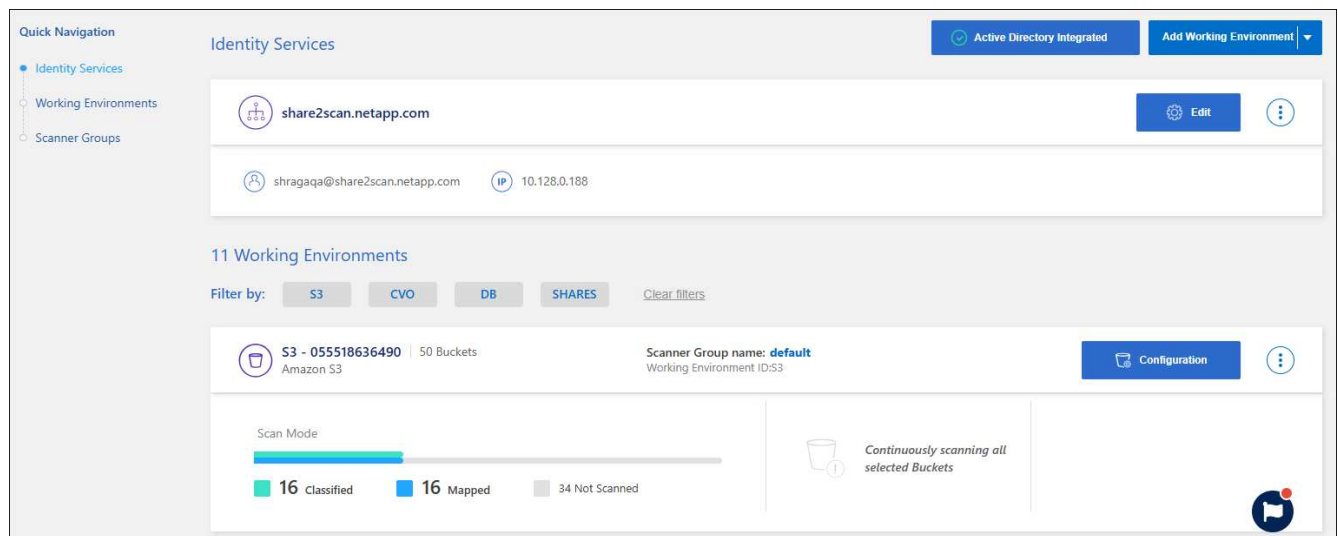
3. In the Scan Configuration page, select the Actions ••• icon.
4. Select **Pause** to pause scanning for a volume, or select **Resume** to resume scanning for a volume that had been previously paused.

View NetApp Data Classification compliance reports

NetApp Data Classification provides reports that you can use to better understand the status of your organization's data privacy program.

By default, the Data Classification dashboards display compliance and governance data for all systems, databases, and data sources. If you want to view reports that contain data for only some of the systems, you can filter to see just them.

> (i)
> • Compliance reports are only available if you perform a full classification scan on your data sources. Data sources that have had a mapping-only scan can only generate the Data Mapping Report.
> • NetApp cannot guarantee 100% accuracy of the personal data and sensitive personal data that Data Classification identifies. You should always validate the information by reviewing the data.

The following reports are available for Data Classification:

- **Data discovery assessment report**: Provides a high-level analysis of the scanned environment to highlight the system's findings and to show areas of concern and potential remediation steps. This report is available in the Governance dashboard.

- **Full data mapping overview report**: Provides information about the size and number of files in your systems. This includes usage capacity, age of data, size of data, and file types. This report is available in the Governance dashboard.

- **Data Subject Access Request report**: Enables you to extract a report of all files that contain information regarding a data subject's specific name or personal identifier. This report is available in the Compliance dashboard.

- **HIPAA report**: Helps you identify the distribution of health information across your files. This report is available in the Compliance dashboard.

- **PCI DSS report**: Helps you identify the distribution of credit card information across your files. This report is available in the Compliance dashboard.

- **Privacy risk assessment report**: Provides privacy insights from your data and a privacy risk score. This report is available in the Compliance dashboard.

- **Reports on a specific information type**: Reports are available that include details about the identified files that contain personal data and sensitive personal data. You can also see files broken down by category and file type.

Select the systems for reports

You can filter the contents of the Data Classification Compliance dashboard to see compliance data for all systems and databases, or for just specific systems.

When you filter the dashboard, Data Classification scopes the compliance data and reports to just those systems that you selected.

Steps

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

2. Select the systems filter drop-down then select the systems.

3. Select **Accept** to confirm your selection.

Data Subject Access Request Report

Privacy regulations such as the European GDPR grant data subjects (such as customers or employees) the right to access their personal data. When a data subject requests this information, this is known as a DSAR (data subject access request). Organizations are required to respond to these requests "without undue delay", and at the latest within one month of receipt.

You can respond to a DSAR by searching for a subject's full name or known identifier (such as an email address) and then downloading a report. The report is designed to aid in your organization's requirement to comply with GDPR or similar data privacy laws.

How can Data Classification help you respond to a DSAR?

When you perform a data subject search, Data Classification finds all of the files that have that person's name or identifier in it. Data Classification checks the latest pre-indexed data for the name or identifier. It doesn't initiate a new scan.

After the search is complete, you can then download the list of files for a Data Subject Access Request report. The report aggregates insights from the data and puts it into legal terms that you can send back to the person.

> (i) Data subject search is not currently supported within databases.

Search for data subjects and download reports

Search for the data subject's full name or known identifier and then download a file list report or DSAR report. You can search by any personal information type.

> (i) English, German, Japanese, and Spanish are supported when searching for the names of data subjects. Support for more languages will be added later.

Steps

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

2. From the Compliance page, locate the **Data Subjects** tab.

3. In the **Data Subjects** section, enter a name or known identifier then select **Search**.

4. When the search completes, select **Download** to access the data subject access request response. Select **Investigate Results** to view more information in the Data Investigation page.

Reports | Data Subjects

← Back

"John Doe"

82

Results Found

Download

Investigate Results ↗

5. Review the results in Data Classification or download them as a report by selecting the download icon.

    a. When you select the download icon, configure your download settings:

- Choose the film format: CSV or JSON

- Enter a **Report name**

- Choose the export destination: **System** or your **Local** machine.

    If you choose system, all data downloads. You must also select the **System**, **Volume**, and **Destination folder path**.

    If you choose **Local**, it limits the report to the first 10,000 rows of unstructured data; 5,000 rows of unstructured data, and 1,000 rows of structured data.

    b. Select **Download Report** to initiate the download.

Download Investigation Report

CSV file ⦿ JSON file ○

Report name

old files

Export destination

⦿ System   ○ Local (limited rows) ⓘ

System ⓘ

ONTAPCluster

Volume

cifs_lab_share

Destination folder path

\\folder\\subfolder

Estimated report size: 35.93 MiB

Download Report   Cancel

Health Insurance Portability and Accountability Act (HIPAA) Report

The Health Insurance Portability and Accountability Act (HIPAA) Report can help you identify files containing health information. It is designed to aid in your organization's requirement to comply with HIPAA data privacy laws. The information Data Classification looks for includes:

- Health reference pattern
- ICD-10-CM Medical code
- ICD-9-CM Medical code
- HR - Health category
- Health Application Data category

The report includes the following information:

- Overview: How many files contain health information and in which systems.
- Encryption: The percentage of files containing health information that are on encrypted or unencrypted systems. This information is specific to Cloud Volumes ONTAP.
- Ransomware Protection: The percentage of files containing health information that are on systems that do or don't have ransomware protection enabled. This information is specific to Cloud Volumes ONTAP.
- Retention: The timeframe in which the files were last modified. This is helpful because you shouldn't keep

health information for longer than you need to process it.

- Distribution of Health Information: The systems where the health information was found and whether encryption and ransomware protection are enabled.

Generate the HIPAA Report

Go to the Compliance tab to generate the report.

Steps

1. From the Data Classification menu, select **Compliance**.
2. Locate the **Reports pane**. Select the download icon next to **HIPAA Report**.

Result

Data Classification generates a PDF report.

Payment Card Industry Data Security Standard (PCI DSS) report

The Payment Card Industry Data Security Standard (PCI DSS) report can help you identify the distribution of credit card information across your files.

The report includes the following information:

- Overview: How many files contain credit card information and in which systems.
- Encryption: The percentage of files containing credit card information that are on encrypted or unencrypted systems. This information is specific to Cloud Volumes ONTAP.
- Ransomware Protection: The percentage of files containing credit card information that are on systems that

do or don't have ransomware protection enabled. This information is specific to Cloud Volumes ONTAP.

- Retention: The timeframe in which the files were last modified. This is helpful because you shouldn't keep credit card information for longer than you need to process it.
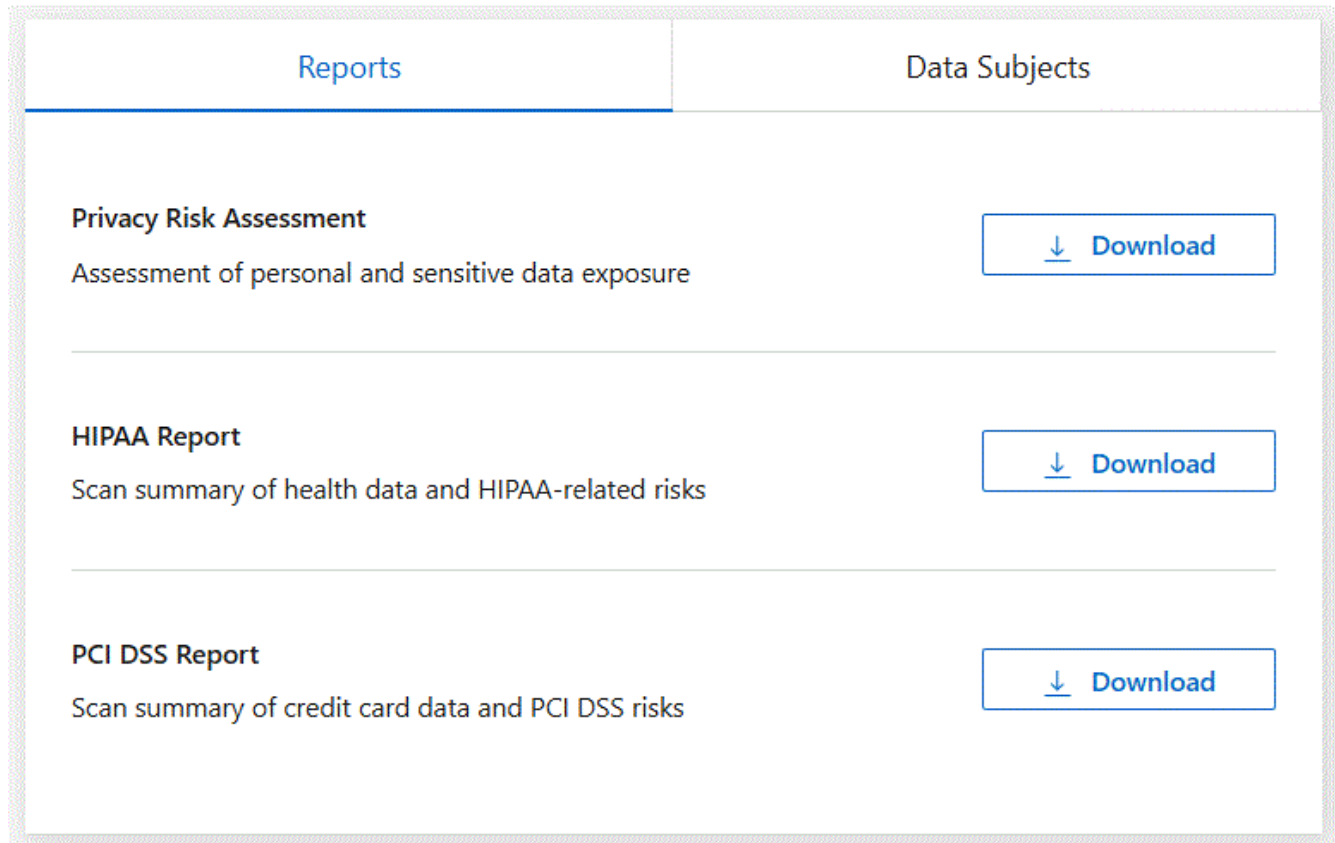
- Distribution of Credit Card Information: The systems where the credit card information was found and whether encryption and ransomware protection are enabled.

Generate the PCI DSS Report

Go to the Compliance tab to generate the report.

Steps

1. From the Data Classification menu, select **Compliance**.
2. Locate the **Reports pane**. Select the download icon next to **PCI DSS Report**.
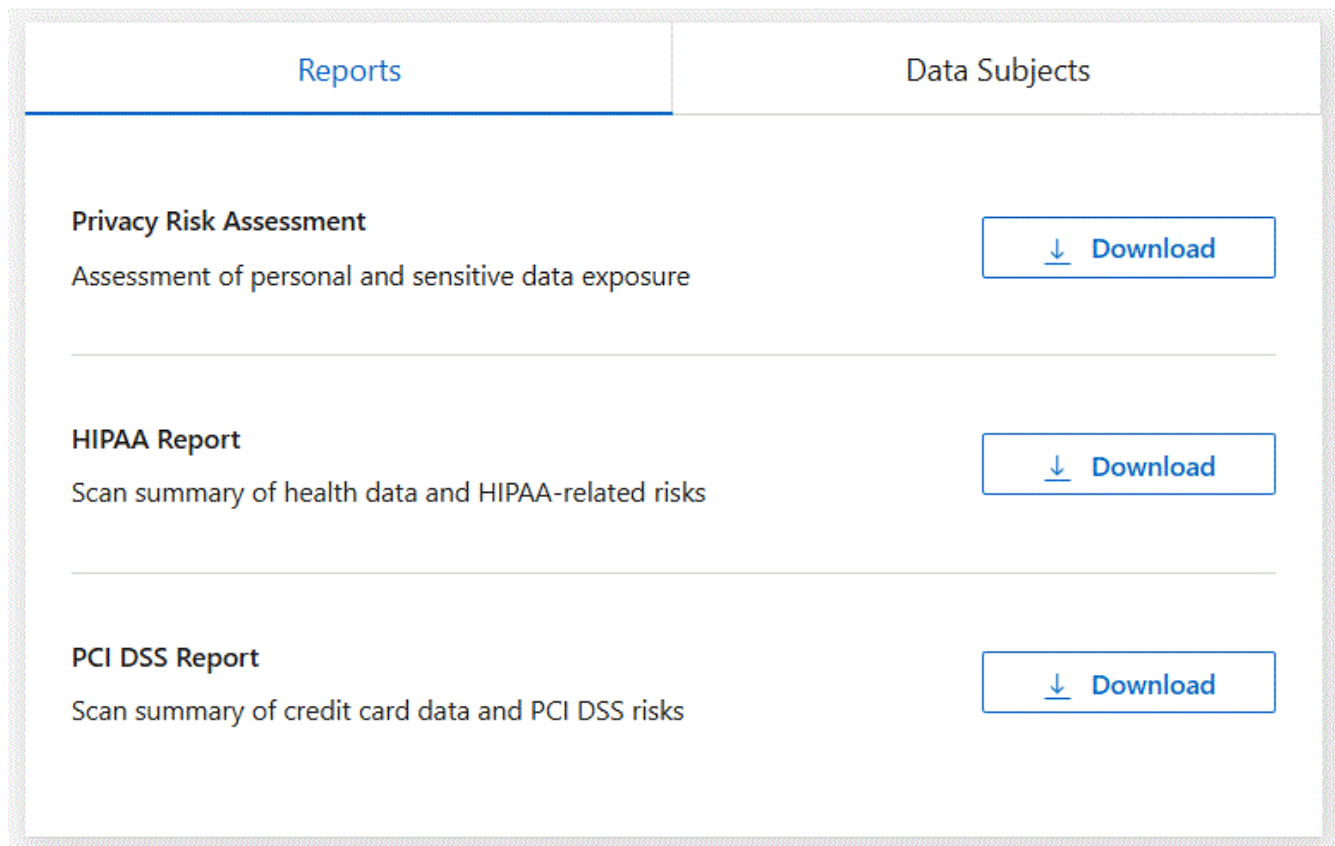
Result

Data Classification generates a PDF report that you can review and send to other groups as needed.

Privacy Risk Assessment Report

The Privacy Risk Assessment Report provides an overview of your organization's privacy risk status, as required by privacy regulations such as GDPR and CCPA.

The report includes the following information:

- Compliance status: A severity score and the distribution of data, whether it's non-sensitive, personal, or sensitive personal.

- Assessment overview: A breakdown of the types of personal data found, as well as the categories of data.
- Data subjects in this assessment: The number of people, by location, for which national identifiers were found.

Generate the Privacy Risk Assessment Report

Go to the Compliance tab to generate the report.

Steps

1. From the Data Classification menu, select **Compliance**.
2. Locate the **Reports pane**. Select the download icon next to **Privacy Risk Assessment Report**.
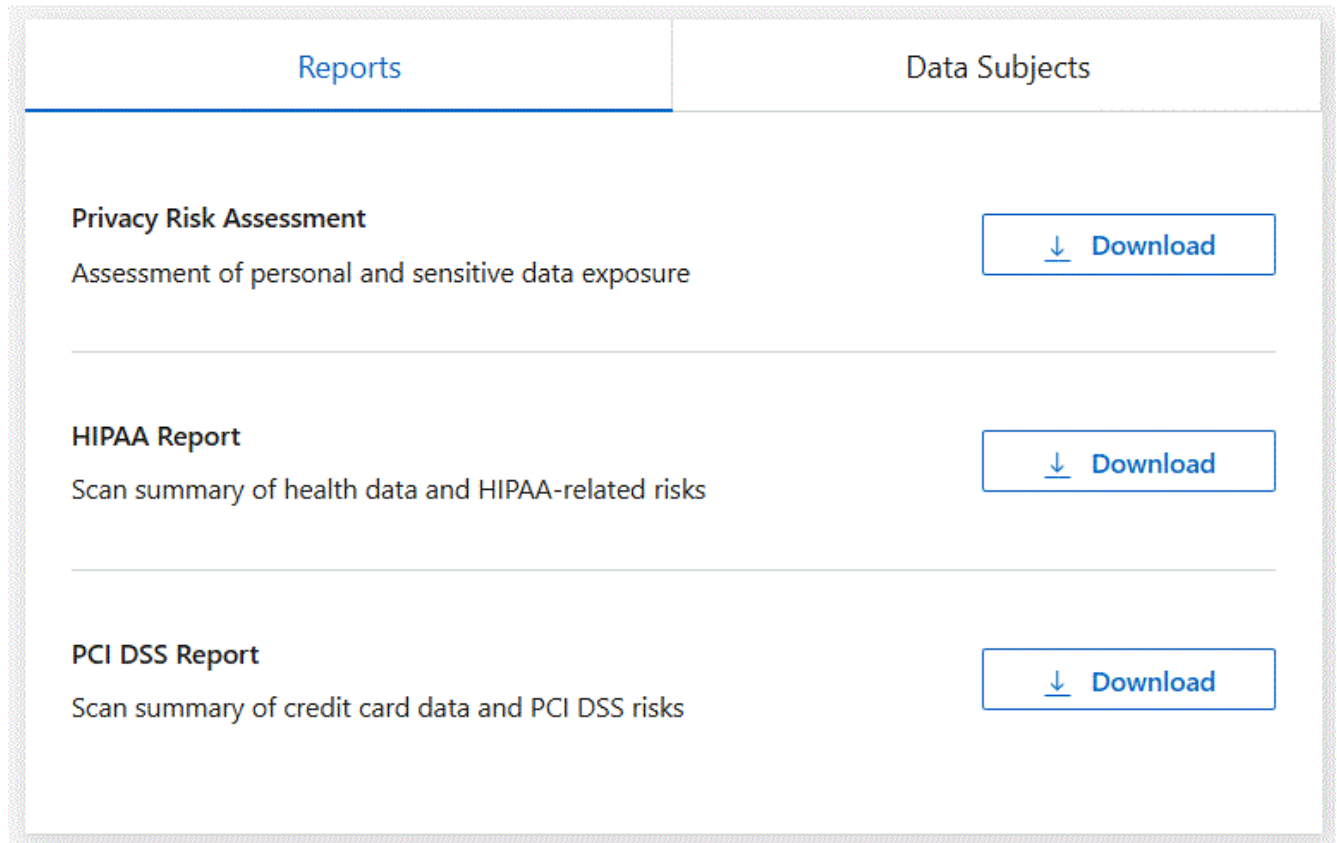
Result

Data Classification generates a PDF report that you can review and send to other groups as needed.

Severity score

Data Classification calculates the severity score for the Privacy Risk Assessment Report on the basis of three variables:

- The percentage of personal data out of all data.
- The percentage of sensitive personal data out of all data.
- The percentage of files that include data subjects, determined by national identifiers such as national IDs, Social Security numbers, and tax ID numbers.

The logic used to determine the score is as follows:

Severity score	Logic
0	All three variables are exactly 0%
1	One of the variables are larger than 0%
2	One of the variables are larger than 3%
3	Two of the variables are larger than 3%
4	Three of the variables are larger than 3%
5	One of the variables are larger than 6%
6	Two of the variables are larger than 6%
7	Three of the variables are larger than 6%
8	One of the variables are larger than 15%
9	Two of the variables are larger than 15%
10	Three of the variables are larger than 15%

Monitor health of NetApp Data Classification

:nofooter
:icons: font
:linkattrs:
:relative_path: ./
:imagesdir: /tmp/d20260206-1555642-bufybq/source/././media/

The NetApp Data Classification Health Monitor dashboard provides real-time monitoring and insights into performance. The Health Monitor captures information about your Data Classification infrastructure, system health, usage metrics, and utilization data, enabling you to identify and remediate issues.

Health Monitor insights

The Health Monitor dashboard presents information in four categories.

- **Infrastructure status**

  View information including the version status, system stability, deployment type, and machine scale.

- **Problematic containers**

  Review the problematic containers field for insights into containers that are stopped or restarting frequently. Use this information to investigate the specific containers.

- **System information**

  The system information panel captures critical information about the NetApp Console and Data Classification such as the public and private IP addresses, host name, operating system, Console version, and Console ID.

- **Usage and utilization**

Review the CPU usage, disk utilization, disk usage, and memory usage. These values are displayed in storage units (GB) or percentages of total use. If any fields display a warning, select the warning for information and remediation recommendations.

Access the Health Monitor dashboard

1. In Data Classification, select **Configuration**.

2. Under the **Configuration** heading, select **Data Classification health monitor**.

3. In the Health Monitor dashboard, you can:

   ◦ Review the usage and utilization. If any usage or utilization metrics display warnings, select the warning for recommendations to resolve the issue.

   ◦ Toggle the graph to display CPU usage, disk utilization, disk usage, and memory usage. You can change the x-axis to display content over hours (6, 12, or 24) or days (2, 7, or 14).

   ◦ Refresh the dashboard to view the most recent data metrics.
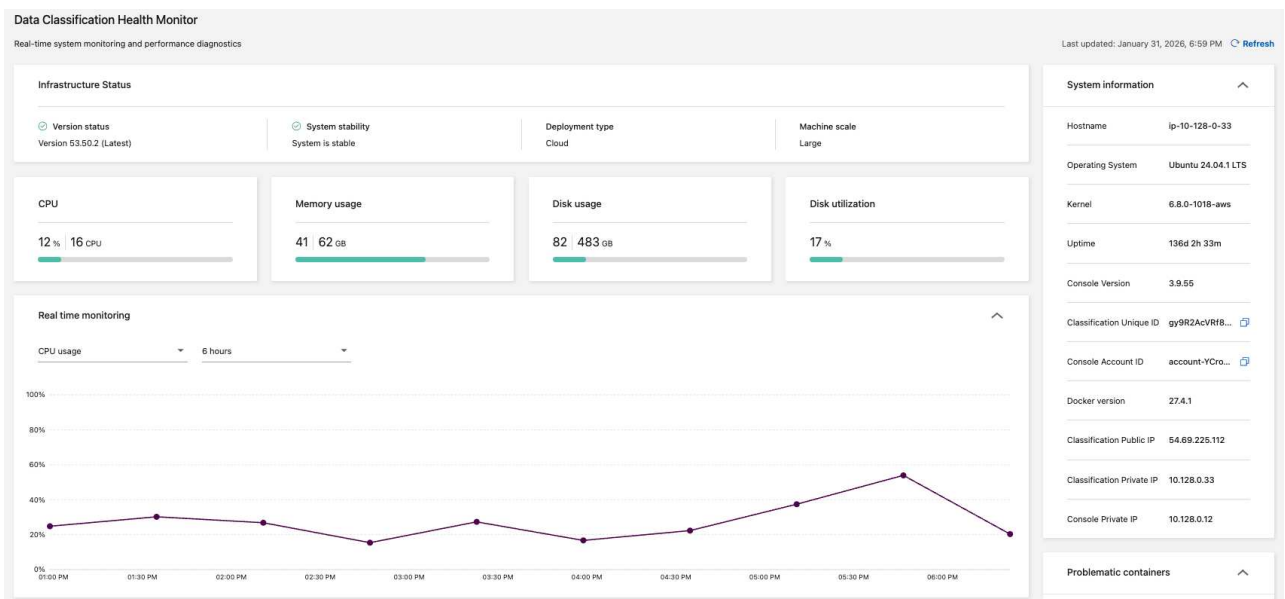
Manage Data Classification

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

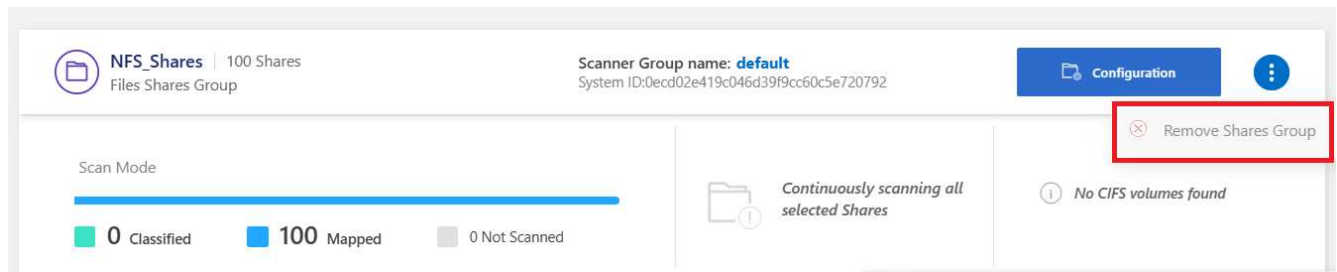If you no longer want to scan user files from a file share group, you can delete the File Shares Group from the Data Classification interface and stop all scans.

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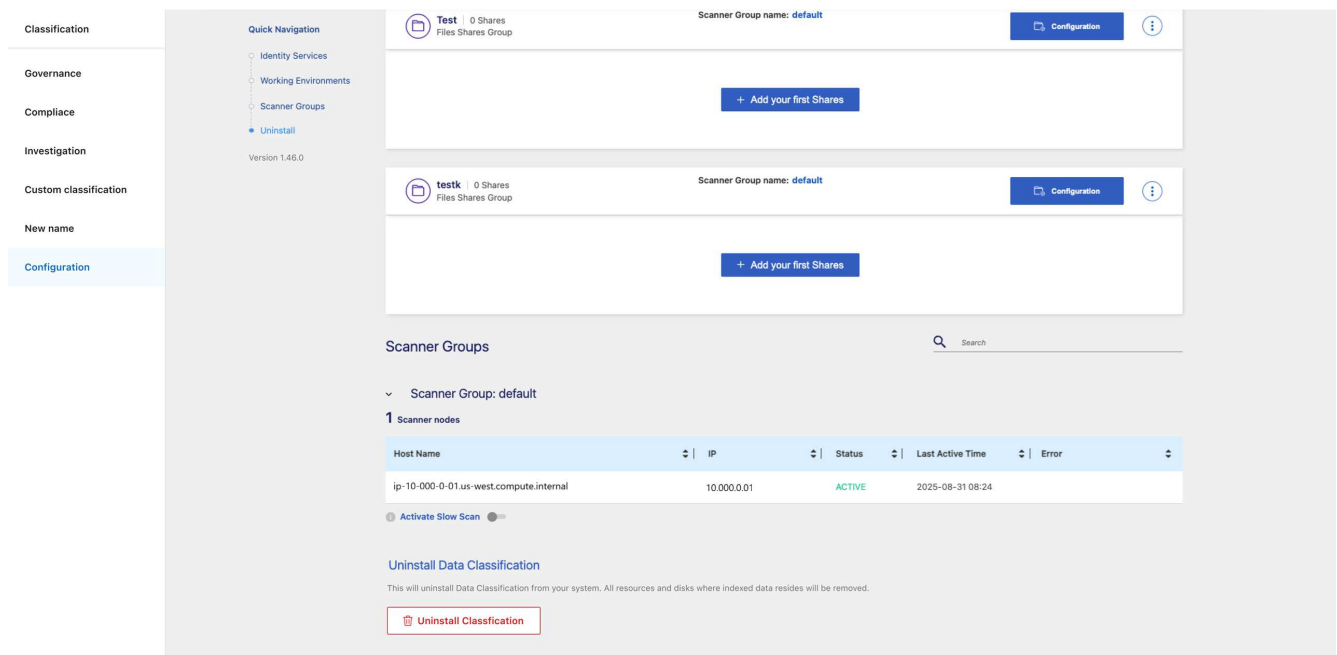
3. In the dialog, enter "uninstall" to proceed with disconnecting the Data Clasification instance from the

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**.
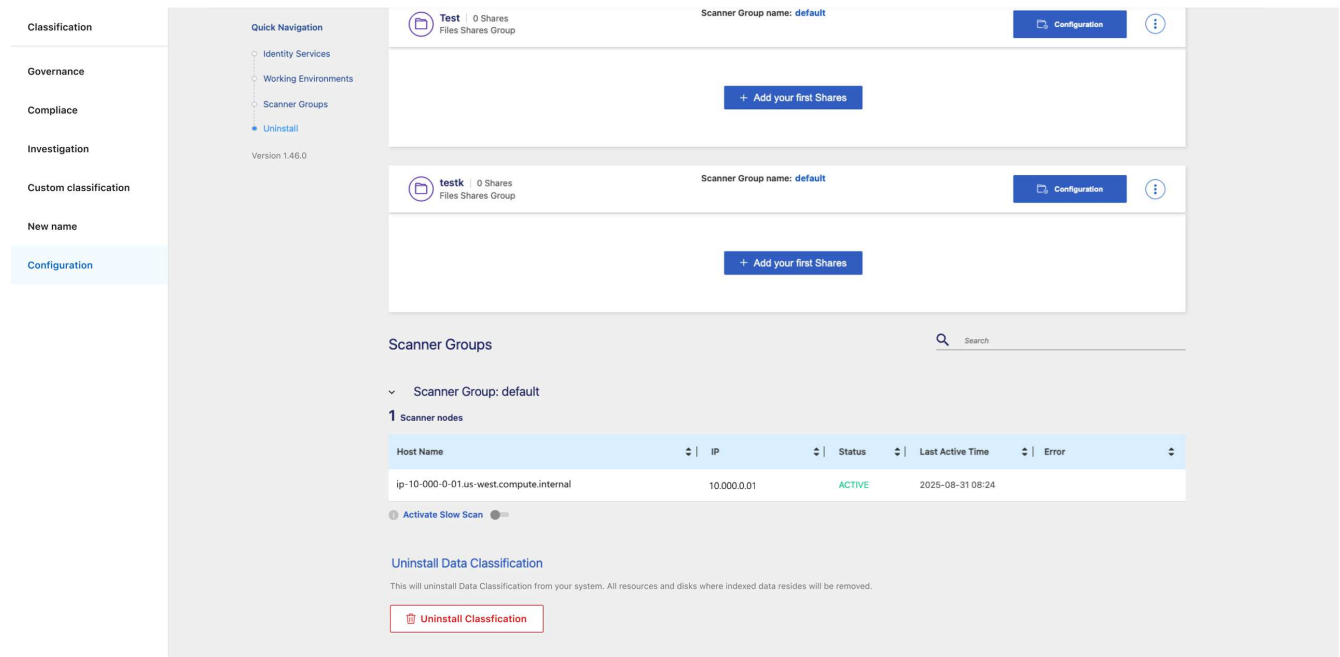
3. In the dialog, enter "uninstall" to proceed with disconnecting the Data Clasification instance from the Console agent. Select **Uninstall** to confirm.

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.

Reference

Supported NetApp Data Classification instance types

NetApp Data Classification software must run on a host that meets specific operating system requirements, RAM requirements, software requirements, and so on. When deploying Data Classification in the cloud, we recommend that you use a system with the "large" characteristics for full functionality.

You can deploy Data Classification on a system with fewer CPUs and less RAM, but there are some limitations when using these less powerful systems. Learn about these limitations.

In the following tables, if the system marked as "default" is not available in the region where you are installing Data Classification, the next system in the table will be deployed.

AWS instance types

System size	Specs	Instance type
Extra Large	32 CPUs, 128 GB RAM, 1 TiB gp3 SSD	m6i.8xlarge (default)
Large	16 CPUs, 64 GB RAM, 500 GiB SSD	m6i.4xlarge (default) m6a.4xlarge m5a.4xlarge m5.4xlarge m4.4xlarge
Medium	8 CPUs, 32 GB RAM, 200 GiB SSD	m6i.2xlarge (default) m6a.2xlarge m5a.2xlarge m5.2xlarge m4.2xlarge
Small	8 CPUs, 16 GB RAM, 100 GiB SSD	c6a.2xlarge (default) c5a.2xlarge c5.2xlarge c4.2xlarge

Azure instance types

System size	Specs	Instance type
Extra Large	32 CPUs, 128 GB RAM, OS Disk (2,048 GiB, min 250 MB/s throughput), and Data Disk (1 TiB SSD, min 750 MB/s throughput)	Standard_D32_v3 (default)
Large	16 CPUs, 64 GB RAM, 500 GiB SSD	Standard_D16s_v3 (default)

GCP instance types

System size	Specs	Instance type
Large	16 CPUs, 64 GB RAM, 500 GiB SSD	n2-standard-16 (default) n2d-standard-16 n1-standard-16

Metadata collected from data sources in NetApp Data Classification

NetApp Data Classification collects certain metadata when performing classification scans on the data from your data sources and systems. Data Classification can access most of the metadata we need to classify your data, but there are some sources where we are unable to access the data we need.

	Metadata	CIFS	NFS
Time stamps	*Creation time*	Available	Not available (Unsupported in Linux)
	Last access time	Available	Available
	Last modify time	Available	Available
Permissions	*Open permissions*	If "EVERYONE" group has access to the file, it is considered "Open to organization"	If "Others" has access to the file, it is considered "Open to organization"
	Users/group access	Users and group information is taken from LDAP	Not available (NFS users are usually managed locally on the server, therefore, the same individual can have a different UID in each server)

> ⓘ
> - Data Classification does not extract the "last accessed time" from the database data sources.
> - Older versions of the Windows OS (for example, Windows 7 and Windows 8) disable the collection of the "last accessed time" attribute by default because it can impact system performance. When this attribute is not collected, Data Classification analytics that are based on "last accessed time" will be impacted. You can enable the collection of the last access time on these older Windows systems if needed.

Last access time timestamp

When Data Classification extracts data from file shares, the operating system considers it as accessing the data and it changes the "last access time" accordingly. After scanning, Data Classification attempts to revert the last access time to the original timestamp. If Data Classification doesn't have write attributes permissions in CIFS, or write permissions in NFS, the system can't revert the last access time to the original timestamp.

ONTAP volumes configured with SnapLock have read-only permissions and also can't revert the last access time to the original timestamp.

By default, if Data Classification doesn't have these permissions, the system won't scan those files in your volumes because Data Classification can't revert the "last access time" to the original timestamp. However, if you don't care if the last access time is reset to the original time in your files, you can select the **Scan when missing "write attributes" permissions** switch at the bottom of the Configuration page so that Data Classification will scan the volumes regardless of permissions.

This functionality is applicable to on-premises ONTAP systems, Cloud Volumes ONTAP, Azure NetApp Files, Amazon FSx for NetApp ONTAP management, and third-party file shares.

There's a filter in the Investigation page called *Scan Analysis Event* that enables you to display either the files that were not classified because Data Classification couldn't revert the last accessed time, or the files that were classified even though Data Classification couldn't revert the last access time.

The filter selections are:

- "Not classified — Cannot revert last access time" - This shows the files that were not classified due to missing write permissions.
- "Classified and updated last access time" - This shows the files that were classified and Data Classification was unable to reset the last access time back to the original date. This filter is relevant only for environments where you turned **Scan when missing "write attributes" permissions** ON.

If needed, you can export these results to a report so you can see which files are, or aren't, being scanned because of permissions. Learn more about Data Investigation reports.

Log in to the NetApp Data Classification system

You need to log into NetApp Data Classification system so you can access log files or edit configuration files.

When Data Classification is installed on a Linux machine on your premises, or on a Linux machine you deployed in the cloud, you can access the configuration file and script directly.

When Data Classification is deployed in the cloud, you need to SSH to the Data Classification instance. You SSH to the system by entering the user and password, or by using the SSH key you provided during the Console agent installation. The SSH command is:

```
ssh -i <path_to_the_ssh_key> <machine_user>@<datasense_ip>
```

- `<path_to_the_ssh_key>` = location of ssh authentication keys

- `<machine_user>`:

  - For AWS: use the <ec2-user>

  - For Azure: use the user created for the Console instance

  - For GCP: use the user created for the Console instance

- `<datasense_ip>` = IP address of the virtual machine instance

You need to modify the security group inbound rules to access the system in the cloud. For details, see:

- Security group rules in AWS
- Security group rules in Azure
- Firewall rules in Google Cloud

NetApp Data Classification APIs

The NetApp Data Classification capabilities available through the web UI are also available through the REST API.

There are four categories defined within Data Classification that correspond to the tabs in the UI:

- Investigation
- Compliance
- Governance
- Configuration

The APIs in the Swagger documentation allow you to search, aggregate data, track your scans, and perform actions including copy, move, and delete.

Overview

The API enables you to perform the following functions:

- Export information

  - Everything that is available in the UI can be exported via the API (with the exception of reports)

  - Data is exported in a JSON format (easy to parse and push to 3rd party applications, like Splunk)

- Create queries using "AND" and "OR" statements, include and exclude information, and more.

  For example, you can locate files *without* specific Personal Identifiable Information (PII) (functionality not available in the UI). You can also exclude specific fields for the export operation.

- Perform actions

  - Update CIFS credentials

  - View and cancel actions

  - Re-scan directories

  - Export data

The API is secure and it uses the same authentication method as the UI. You can find information on the authentication in the REST API documentation.

Accessing the Swagger API reference

To get into Swagger you'll need the IP address of the your Data Classification instance. In the case of a cloud deployment you'll use the public IP address. Then you'll need to get into this endpoint:

https://<classification_ip>/documentation

Example using the APIs

The following example shows an API call to copy files.

API Request
You'll initially need to get all the relevant fields and options for a system to view all of the filters in the investigation tab.

```
curl -X GET "http://{classification_ip}/api/{classification_version}
/search/options?data_mode=ALL_EXTRACTABLE" -H "accept: application/json"
-H "Authorization: Bearer eyJhbGciOiJSUzI1NiIsInR………… " -H "x-agent-id:
hOXsZNvnA5LsthwMILtjL9xZFYBQxAwMclients"
```

Response

```
{
  "options": [
    {
      "active_directory_affected": false,
      "data_mode": "ALL_SCANNED",
      "field": "string",
      "is_rulable": true,
      "name": "string",
      "operators": [
        "EQUALS"
      ],
      "optional_values": [
        {}
      ],
      "secondary": {},
      "server_data": false,
      "type": "TEXT"
    }
  ]
}
{
  "options": [
```

```json
      {
        "active_directory_affected": false,
        "data_mode": "ALL_EXTRACTABLE",
        "field": "POLICIES",
        "name": "Policies",
        "operators": [
          "IN",
          "NOT_IN"
        ],
        "server_data": true,
        "type": "SELECT"
      },
      {
        "active_directory_affected": false,
        "data_mode": "ALL_EXTRACTABLE",
        "field": "EXTRACTION_STATUS_RANGE",
        "name": "Scan Analysis Status",
        "operators": [
          "IN"
        ],
        "server_data": true,
        "type": "SELECT"
      },
      {
        "active_directory_affected": false,
        "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
        "field": "SCAN_ANALYSIS_ERROR",
        "name": "Scan Analysis Event",
        "operators": [
          "IN"
        ],
        "server_data": true,
        "type": "SELECT"
      },
      {
        "active_directory_affected": false,
        "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
        "field": "PUBLIC_ACCESS",
        "name": "Open Permissions",
        "operators": [
          "IN",
          "NOT_IN"
        ],
        "server_data": true,
        "type": "SELECT"
      },
```

```
  {
    "active_directory_affected": true,
    "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
    "field": "USERS_PERMISSIONS_COUNT_RANGE",
    "name": "Number of Users with Access",
    "operators": [
      "IN",
      "NOT_IN"
    ],
    "server_data": true,
    "type": "SELECT"
  },
  {
    "active_directory_affected": true,
    "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
    "field": "USER_GROUP_PERMISSIONS",
    "name": "User / Group Permissions",
    "operators": [
      "IN"
    ],
    "server_data": true,
    "type": "SELECT"
  },
  {
    "active_directory_affected": false,
    "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
    "field": "FILE_OWNER",
    "name": "File Owner",
    "operators": [
      "EQUALS",
      "CONTAINS"
    ],
    "server_data": true,
    "type": "TEXT"
  },
  {
    "active_directory_affected": false,
    "data_mode": "ALL_EXTRACTABLE",
    "field": "ENVIRONMENT_TYPE",
    "name": "system-type",
    "operators": [
      "IN",
      "NOT_IN"
    ],
    "server_data": true,
    "type": "SELECT"
```

```
      },
      {
        "active_directory_affected": false,
        "data_mode": "ALL_EXTRACTABLE",
        "field": "ENVIRONMENT",
        "name": "system",
        "operators": [
          "IN",
          "NOT_IN"
        ],
        "server_data": true,
        "type": "SELECT"
      },
      {
        "active_directory_affected": false,
        "data_mode": "ALL_SCANNED",
        "field": "SCAN_TASK",
        "name": "Storage Repository",
        "operators": [
          "IN",
          "NOT_IN"
        ],
        "server_data": true,
        "type": "SELECT"
      },
      {
        "active_directory_affected": false,
        "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
        "field": "FILE_PATH",
        "name": "File / Directory Path",
        "operators": [
          "MULTI_CONTAINS",
          "MULTI_EXCLUDE"
        ],
        "server_data": true,
        "type": "MULTI_TEXT"
      },
      {
        "active_directory_affected": false,
        "data_mode": "ALL_DASHBOARD_EXTRACTABLE",
        "field": "CATEGORY",
        "name": "Category",
        "operators": [
          "IN",
          "NOT_IN"
        ],
```

```
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "PATTERN_SENSITIVITY_LEVEL",
      "name": "Sensitivity Level",
      "operators": [
        "IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
      "field": "NUMBER_OF_IDENTIFIERS",
      "name": "Number of identifiers",
      "operators": [
        "IN",
        "NOT_IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "PATTERN_PERSONAL",
      "name": "Personal Data",
      "operators": [
        "IN",
        "NOT_IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "PATTERN_SENSITIVE",
      "name": "Sensitive Personal Data",
      "operators": [
        "IN",
        "NOT_IN"
```

```
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "DATA_SUBJECT",
      "name": "Data Subject",
      "operators": [
        "EQUALS",
        "CONTAINS"
      ],
      "server_data": true,
      "type": "TEXT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "DIRECTORIES",
      "field": "DIRECTORY_TYPE",
      "name": "Directory Type",
      "operators": [
        "IN",
        "NOT_IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "FILE_TYPE",
      "name": "File Type",
      "operators": [
        "IN",
        "NOT_IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "FILE_SIZE_RANGE",
      "name": "File Size",
      "operators": [
```

```
        "IN",
        "NOT_IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
      "field": "FILE_CREATION_RANGE_RETENTION",
      "name": "Created Time",
      "operators": [
        "IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "DISCOVERED_TIME_RANGE",
      "name": "Discovered Time",
      "operators": [
        "IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
      "field": "FILE_LAST_MODIFICATION_RETENTION",
      "name": "Last Modified",
      "operators": [
        "IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_FILESYSTEM_EXTRACTABLE",
      "field": "FILE_LAST_ACCESS_RANGE_RETENTION",
      "name": "Last Accessed",
      "operators": [
        "IN"
```

```
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "FILES",
      "field": "IS_DUPLICATE",
      "name": "Duplicates",
      "operators": [
        "EQUALS",
        "IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "FILES",
      "field": "FILE_HASH",
      "name": "File Hash",
      "operators": [
        "EQUALS",
        "IN"
      ],
      "server_data": true,
      "type": "TEXT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "USER_DEFINED_STATUS",
      "name": "Tags",
      "operators": [
        "IN",
        "NOT_IN"
      ],
      "server_data": true,
      "type": "SELECT"
    },
    {
      "active_directory_affected": false,
      "data_mode": "ALL_EXTRACTABLE",
      "field": "ASSIGNED_TO",
      "name": "Assigned to",
      "operators": [
```

```
        "IN",
        "NOT_IN"
      ],
      "server_data": true,
      "type": "SELECT"
    }
  ]
}
```

We will use that response in our request parameters to filter the desired files we want to copy.

You can apply an action on multiple items. Supported action types include: move, delete, and copy.

We will create the copy action:

API Request

This next API is that action API and it allows you to create multiple actions.

```
curl -X POST "http://
{classification_ip}/api//{classification_version}/actions" -H "accept:
application/json" -H "Authorization: Bearer eyJhbGciOiJSUzI1NiIsInR……… "
-H "x-agent-id: hOXsZNvnA5LsthwMILtjL9xZFYBQxAwMclients " -H "Content-
Type: application/json" -d "{ \"action_type\": \"COPY\", \"data_mode\":
\"FILES\", \"policy_id\": 0, \"request_params\": { destination_nfs_path:
"{ontap_ip}:/{share_name} " },
\"requested_query\":{"condition":"AND","rules":[{"field":"ENVIRONMENT_TYPE
","operator":"IN","value":["ONPREM"]},{"field":"CATEGORY","operator":"IN",
"value":["21"]}]}}"
```

Response

The response will return the action object, so you can use the get and delete APIs to get status about the action, or to cancel it.

```
{
  "action_type": "COPY",
  "creation_time": "2023-08-08T12:37:21.705Z",
  "data_mode": "FILES",
  "end_time": "2023-08-08T12:37:21.705Z",
  "estimated_time_to_complete": 0,
  "id": 0,
  "policy_id": 0,
  "policy_name": "string",
  "priority": 0,
  "request_params": {},
  "requested_query": {},
  "result": {
    "error_message": "string",
    "failed": 0,
    "in_progress": 0,
    "succeeded": 0,
    "total": 0
  },
  "start_time": "2023-08-08T12:37:21.705Z",
  "status": "QUEUED",
  "title": "string",
  "user_id": "string"
}
```

Knowledge and support

Register for NetApp Console support

Support registration is required to receive technical support specific to the NetApp Console and its storage solutions and data services. Support registration is also required to enable key workflows for Cloud Volumes ONTAP systems.

Registering for support does not enable NetApp support for a cloud provider file service. For technical support related to a cloud provider file service, its infrastructure, or any solution using the service, refer to "Getting help" in the documentation for that product.

- Amazon FSx for ONTAP
- Azure NetApp Files
- Google Cloud NetApp Volumes

Support registration overview

There are two forms of registration to activate support entitlement:

- Registering your NetApp Console account serial number (your 20 digit 960xxxxxxxxx serial number located on the Support Resources page in the Console).

  This serves as your single support subscription ID for any service within the Console. Each Console account must be registered.

- Registering the Cloud Volumes ONTAP serial numbers associated with a subscription in your cloud provider's marketplace (these are 20 digit 909201xxxxxxxx serial numbers).

  These serial numbers are commonly referred to as *PAYGO serial numbers* and get generated by the NetApp Console at the time of Cloud Volumes ONTAP deployment.

Registering both types of serial numbers enables capabilities like opening support tickets and automatic case generation. Registration is completed by adding NetApp Support Site (NSS) accounts to the Console as described below.

Register NetApp Console for NetApp support

To register for support and activate support entitlement, one user in your NetApp Console account must associate a NetApp Support Site account with their Console login. How you register for NetApp support depends on whether you already have a NetApp Support Site (NSS) account.

Existing customer with an NSS account

If you're a NetApp customer with an NSS account, you simply need to register for support through the Console.

Steps

1. Select **Administration** > **Credentials**.
2. Select **User Credentials**.

3. Select **Add NSS credentials** and follow the NetApp Support Site (NSS) authentication prompt.

4. To confirm that the registration process was successful, select the Help icon, and select **Support**.

   The **Resources** page should show that your Console account is registered for support.

   Note that other Console users will not see this same support registration status if they have not associated a NetApp Support Site account with their login. However, that doesn't mean that your account is not registered for support. As long as one user in the organization has followed these steps, then your account has been registered.

Existing customer but no NSS account

If you're an existing NetApp customer with existing licenses and serial numbers but *no* NSS account, you need to create an NSS account and associate it with your Console login.

Steps

1. Create a NetApp Support Site account by completing the NetApp Support Site User Registration form

   a. Be sure to select the appropriate User Level, which is typically **NetApp Customer/End User**.

   b. Be sure to copy the Console account serial number (960xxxx) used above for the serial number field. This will speed up the account processing.

2. Associate your new NSS account with your Console login by completing the steps under Existing customer with an NSS account.

Brand new to NetApp

If you are brand new to NetApp and you don't have an NSS account, follow each step below.

Steps

1. In the upper right of the Console, select the Help icon, and select **Support**.

2. Locate your account ID serial number from the Support Registration page.
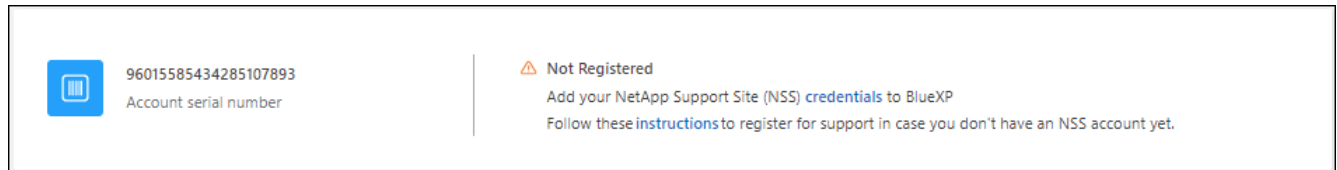
3. Navigate to NetApp's support registration site and select **I am not a registered NetApp Customer**.

4. Fill out the mandatory fields (those with red asterisks).

5. In the **Product Line** field, select **Cloud Manager** and then select your applicable billing provider.

6. Copy your account serial number from step 2 above, complete the security check, and then confirm that you read NetApp's Global Data Privacy Policy.

   An email is immediately sent to the mailbox provided to finalize this secure transaction. Be sure to check your spam folders if the validation email doesn't arrive in few minutes.

7. Confirm the action from within the email.

   Confirming submits your request to NetApp and recommends that you create a NetApp Support Site account.

8. Create a NetApp Support Site account by completing the NetApp Support Site User Registration form

   a. Be sure to select the appropriate User Level, which is typically **NetApp Customer/End User**.

   b. Be sure to copy the account serial number (960xxxx) used above for the serial number field. This will speed up processing.

After you finish

NetApp should reach out to you during this process. This is a one-time onboarding exercise for new users.

Once you have your NetApp Support Site account, associate the account with your Console login by completing the steps under Existing customer with an NSS account.

Associate NSS credentials for Cloud Volumes ONTAP support

Associating NetApp Support Site credentials with your Console account is required to enable the following key workflows for Cloud Volumes ONTAP:

- Registering pay-as-you-go Cloud Volumes ONTAP systems for support

  Providing your NSS account is required to activate support for your system and to gain access to NetApp technical support resources.

- Deploying Cloud Volumes ONTAP when you bring your own license (BYOL)

  Providing your NSS account is required so that the Console can upload your license key and to enable the subscription for the term that you purchased. This includes automatic updates for term renewals.

- Upgrading Cloud Volumes ONTAP software to the latest release

Associating NSS credentials with your NetApp Console account is different than the NSS account that is associated with a Console user login.

These NSS credentials are associated with your specific Console account ID. Users who belong to the Console organization can access these credentials from **Support > NSS Management**.

- If you have a customer-level account, you can add one or more NSS accounts.
- If you have a partner or reseller account, you can add one or more NSS accounts, but they can't be added alongside customer-level accounts.

Steps

1. In the upper right of the Console, select the Help icon, and select **Support**.
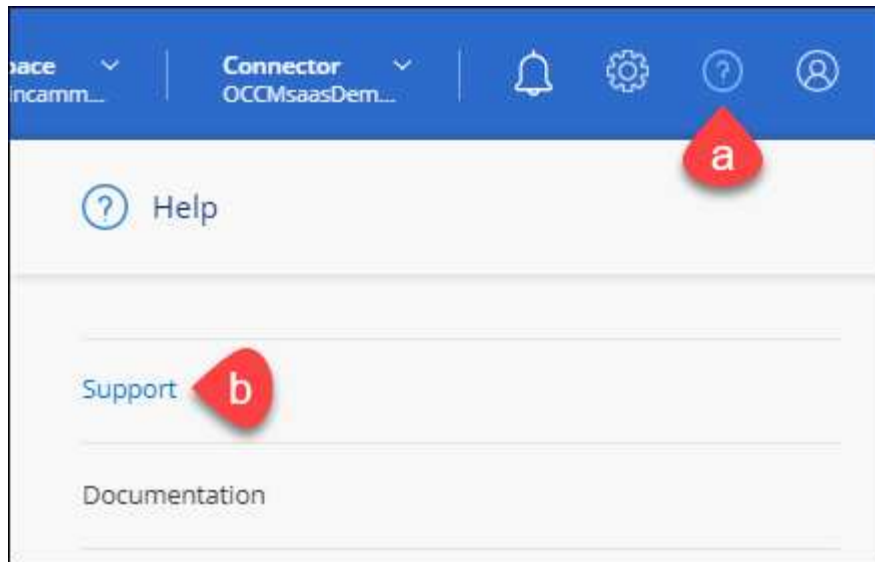
2. Select **NSS Management > Add NSS Account**.

3. When you're prompted, select **Continue** to be redirected to a Microsoft login page.

   NetApp uses Microsoft Entra ID as the identity provider for authentication services specific to support and licensing.

4. At the login page, provide your NetApp Support Site registered email address and password to perform the authentication process.

   These actions enable the Console to use your NSS account for things like license downloads, software upgrade verification, and future support registrations.

   Note the following:

   ◦ The NSS account must be a customer-level account (not a guest or temp account). You can have multiple customer-level NSS accounts.

   ◦ There can be only one NSS account if that account is a partner-level account. If you try to add customer-level NSS accounts and a partner-level account exists, you'll get the following error message:

      "The NSS customer type is not allowed for this account as there are already NSS Users of different type."

      The same is true if you have pre-existing customer-level NSS accounts and try to add a partner-level account.

   ◦ Upon successful login, NetApp will store the NSS user name.

      This is a system-generated ID that maps to your email. On the **NSS Management** page, you can display your email from the ••• menu.

   ◦ If you ever need to refresh your login credential tokens, there is also an **Update Credentials** option in the ••• menu.

      Using this option prompts you to log in again. Note that the token for these accounts expire after 90 days. A notification will be posted to alert you of this.

Get help for NetApp Data Classification

NetApp provides support for NetApp Console and its cloud services in a variety of ways. Extensive free self-support options are available 24/7, such as knowledge base (KB) articles and a community forum. Your support registration includes remote technical support via web ticketing.

Get support for a cloud provider file service

For technical support related to a cloud provider file service, its infrastructure, or any solution using the service, refer to the documentation for that product.

- Amazon FSx for ONTAP
- Azure NetApp Files
- Google Cloud NetApp Volumes

To receive technical support specific to NetApp and its storage solutions and data services, use the support options described below.

Use self-support options

These options are available for free, 24 hours a day, 7 days a week:

- Documentation

  The NetApp Console documentation that you're currently viewing.

- Knowledge base

  Search through the NetApp knowledge base to find helpful articles to troubleshoot issues.

- Communities

  Join the NetApp Console community to follow ongoing discussions or create new ones.

Create a case with NetApp support

In addition to the self-support options above, you can work with a NetApp Support specialist to resolve any issues after you activate support.

Before you get started

- To use the **Create a Case** capability, you must first associate your NetApp Support Site credentials with your Console login. Learn how to manage credentials associated with your Console login.
- If you're opening a case for an ONTAP system that has a serial number, then your NSS account must be associated with the serial number for that system.

Steps

1. In NetApp Console, select **Help > Support**.
2. On the **Resources** page, choose one of the available options under Technical Support:

a. Select **Call Us** if you'd like to speak with someone on the phone. You'll be directed to a page on netapp.com that lists the phone numbers that you can call.

b. Select **Create a Case** to open a ticket with a NetApp Support specialist:

- **Service**: Select the service that the issue is associated with. For example, **NetApp Console** when specific to a technical support issue with workflows or functionality within the Console.

- **System**: If applicable to storage, select **Cloud Volumes ONTAP** or **On-Prem** and then the associated working environment.

  The list of systems are within scope of the Console organization, and Console agent you have selected in the top banner.

- **Case Priority**: Choose the priority for the case, which can be Low, Medium, High, or Critical.

  To learn more details about these priorities, hover your mouse over the information icon next to the field name.

- **Issue Description**: Provide a detailed description of your problem, including any applicable error messages or troubleshooting steps that you performed.

- **Additional Email Addresses**: Enter additional email addresses if you'd like to make someone else aware of this issue.

- **Attachment (Optional)**: Upload up to five attachments, one at a time.

  Attachments are limited to 25 MB per file. The following file extensions are supported: txt, log, pdf, jpg/jpeg, rtf, doc/docx, xls/xlsx, and csv.

After you finish

A pop-up will appear with your support case number. A NetApp Support specialist will review your case and get back to you soon.

For a history of your support cases, you can select **Settings > Timeline** and look for actions named "create support case." A button to the far right lets you expand the action to see details.

It's possible that you might encounter the following error message when trying to create a case:

"You are not authorized to Create a Case against the selected service"

This error could mean that the NSS account and the company of record it's associated with is not the same company of record for the NetApp Console account serial number (ie. 960xxxx) or the working environment serial number. You can seek assistance using one of the following options:

- Submit a non-technical case at https://mysupport.netapp.com/site/help

Manage your support cases

You can view and manage active and resolved support cases directly from the Console. You can manage the

cases associated with your NSS account and with your company.

Note the following:

- The case management dashboard at the top of the page offers two views:

  ◦ The view on the left shows the total cases opened in the past 3 months by the user NSS account you provided.
  ◦ The view on the right shows the total cases opened in the past 3 months at your company level based on your user NSS account.

  The results in the table reflect the cases related to the view that you selected.

- You can add or remove columns of interest and you can filter the contents of columns like Priority and Status. Other columns provide just sorting capabilities.

  View the steps below for more details.

- At a per-case level, we offer the ability to update case notes or close a case that is not already in Closed or Pending Closed status.

Steps

1. In the NetApp Console, select **Help > Support**.

2. Select **Case Management** and if you're prompted, add your NSS account to the Console.

   The **Case management** page shows open cases related to the NSS account that is associated with your Console user account. This is the same NSS account that appears at the top of the **NSS management** page.

3. Optionally modify the information that displays in the table:

   ◦ Under **Organization's cases**, select **View** to view all cases associated with your company.
   ◦ Modify the date range by choosing an exact date range or by choosing a different time frame.
   ◦ Filter the contents of the columns.
   ◦ Change the columns that appear in the table by selecting ⊕ and then choosing the columns that you'd like to display.

4. Manage an existing case by selecting ••• and selecting one of the available options:

   ◦ **View case**: View full details about a specific case.
   ◦ **Update case notes**: Provide additional details about your problem or select **Upload files** to attach up to a maximum of five files.

     Attachments are limited to 25 MB per file. The following file extensions are supported: txt, log, pdf, jpg/jpeg, rtf, doc/docx, xls/xlsx, and csv.

   ◦ **Close case**: Provide details about why you're closing the case and select **Close case**.

Frequently asked questions about NetApp Data Classification

This FAQ can help if you're just looking for a quick answer to a question.

NetApp Data Classification

The following questions provide a general understanding of Data Classification.

How does Data Classification work?

Data Classification deploys another layer of AI alongside your NetApp Console system and storage systems. It then scans the data on volumes, buckets, databases, and other storage accounts and indexes the data insights that are found. Data Classification leverages both artificial intelligence and natural language processing, as opposed to alternative solutions that are commonly built around regular expressions and pattern matching.

Data Classification uses AI to provide contextual understanding of data for accurate detection and classification. It is driven by AI because it is designed for modern data types and scale. It also understands data context in order to provide strong, accurate, discovery and classification.

Learn more about how Data Classification works.

Does Data Classification have a REST API, and does it work with third-party tools?

Yes, Data Classification has a REST API for the supported features in the Data Classification version that is part of the Console core platform. See API documentation.

Is Data Classification available through the cloud marketplaces?

Data Classification is part of the NetApp Console core features, so you do not need to use the marketplaces for this service .

Data Classification scanning and analytics

The following questions relate to Data Classification scanning performance and the analytics.

How often does Data Classification scan my data?

While the initial scan of your data might take a little bit of time, subsequent scans only inspect the incremental changes, which reduces system scan times. Data Classification scans your data continuously in a round-robin fashion, six repositories at a time, so that all changed data is classified very quickly.

Learn how scans work.

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

Data scans have a negligible impact on your storage systems and on your data.

Does scan performance vary?

Scan performance can vary based on the network bandwidth and the average file size in your environment. It can also depend on the size characteristics of the host system (either in the cloud or on-premises). See The Data Classification instance and Deploying Data Classification for more information.

When initially adding new data sources, you can also choose to perform only a "mapping" (Mapping only) scan instead of a full "classification" (Map & Classify) scan. Mapping can be done on your data sources very quickly because it does not access files to see the data inside. See the difference between a mapping and classification scan.

Can I search my data using Data Classification?

Data Classification offers extensive search capabilities that make it easy to search for a specific file or piece of data across all connected sources. Data Classification empowers users to search deeper than just what the metadata reflects. It is a language-agnostic service that can also read the files and analyze a multitude of sensitive data types, such as names and IDs. For example, users can search across both structured and unstructured data stores to find data that may have leaked from databases to user files, in violation of corporate policy. Searches can be saved for later, and policies can be created to search and take action on the results at a set frequency.

Once the files of interest are found, characteristics can be listed, including tags, system account, bucket, file path, category (from classification), file size, last modified, permission status, duplicates, sensitivity level, personal data, sensitive data types within the file, owner, file type, file size, created time, file hash, whether the data was assigned to someone seeking their attention, and more. Filters can be applied to screen out characteristics that are not pertinent.

Data Classification also has role-based access control (RBAC) to allow files to be moved or deleted, if the right permissions are present. If the right permissions are not present, the tasks can be assigned to someone in the organization who does have the right permissions.

Data Classification management and privacy

The following questions provide information on how to manage Data Classification and privacy settings.

How do I enable or disable Data Classification?

First you need to deploy an instance of Data Classification in the Console, or on an on-premises system. Once the instance is running, you can enable the service on existing systems, databases, and other data sources from the **Configuration** tab or by selecting a specific system. Learn how to get started.

> ⓘ   Activating Data Classification on a data source results in an immediate initial scan. Scan results display shortly after.

You can disable Data Classification from scanning an individual system, database, or file share group from the Data Classification Configuration page. See Remove data sources from Data Classification.

To completely remove the Data Classification instance, manually remove the Data Classification instance from your cloud provider's portal or on-prem location.

Can the service exclude scanning data in certain directories?

Yes. If you want Data Classification to exclude scanning data that resides in certain data source directories, you can provide that list to the classification engine. After you apply that change, Data Classification will exclude scanning data in the specified directories. Learn more.

Are snapshots that reside on ONTAP volumes scanned?

No. Data Classification does not scan snapshots because the content is identical to the content in the volume.

What happens if data tiering is enabled on your ONTAP volumes?

When Data Classification scans volumes that have cold data tiered to object storage using the Mapping only scans, it scans all of the data—data that's on local disks and cold data tiered to object storage. This is also true for non-NetApp products that implement tiering.

The Mapping only scan doesn't heat up the cold data—it stays cold and remains in object storage. On the other hand, if you perform the Map & Classify scan, some configurations might heat up the cold data.

Types of source systems and data types

The following questions relate to the types of storage that can be scanned, and the types of data that is scanned.

Are there any restrictions when deployed in a Government region?

Data Classification is supported when the Console agent is deployed in a Government region (AWS GovCloud, Azure Gov, or Azure DoD) - also known as "Restricted mode".

What data sources can I scan if I install Data Classification in a site without internet access?

> ⓘ BlueXP private mode (legacy BlueXP interface) is typically used with on-premises environments that have no internet connection and with secure cloud regions, which includes AWS Secret Cloud, AWS Top Secret Cloud, and Azure IL6. NetApp continues to support these environments with the legacy BlueXP interface. For private mode documentation in the legacy BlueXP interface, see PDF documentation for BlueXP private mode.

Data Classification can only scan data from data sources that are local to the on-premises site. At this time, Data Classification can scan the following local data sources in "Private mode" - also known as a "dark" site:

- On-premises ONTAP systems
- Database schemas
- Object Storage that uses the Simple Storage Service (S3) protocol

Which file types are supported?

Data Classification scans all files for category and metadata insights, and displays all file types in the file types section of the dashboard.

When Data Classification detects Personal Identifiable Information (PII), or when it performs a DSAR search,

only the following file formats are supported:

```
.CSV, .DCM, .DOC, .DOCX, .JSON, .PDF, .PPTX, .RTF, .TXT, .XLS, .XLSX, Docs,
Sheets, and Slides
```

What kinds of data and metadata does Data Classification capture?

Data Classification enables you to run a general "mapping" scan or a full "classification" scan on your data sources. Mapping provides only a high-level overview of your data, whereas Classification provides deep-level scanning of your data. Mapping can be done on your data sources very quickly because it does not access files to see the data inside.

- **Data mapping scan (Mapping only scan)**: Data Classification scans the metadata only. This is useful for overall data management and governance, quick project scoping, very large estates, and prioritization. Data mapping is based on metadata and is considered a **fast** scan.

  After a fast scan, you can generate a Data Mapping Report. This report is an overview of the data stored in your corporate data sources to assist you with decisions about resource utilization, migration, backup, security, and compliance processes.

- **Data classification deep scan (Map & Classify scan)**: Data Classification scans data using standard protocols and read-only permission throughout your environments. Select files are opened and scanned for sensitive business-related data, private information, and issues related to ransomware.

  After a full scan there are many additional Data Classification features you can apply to your data, such as view and refine data in the Data Investigation page, search for names within files, copy, move, and delete source files, and more.

Data Classification captures metadata such as: file name, permissions, creation time, last access, and last modification. This includes all of the metadata that appears in the Data Investigatcdion Details page and in Data Investigation Reports.

Data Classification can identify many types of private data such as personal information (PII) and sensitive personal information (SPII). For details about private data, refer to Categories of private data that Data Classification scans.

Can I limit Data Classification information to specific users?

Yes, Data Classification is fully integrated with the NetApp Console. NetApp Console users can only see information for the systems they are eligible to view according to their permissions.

Additionally, if you want to allow certain users to just view Data Classification scan results without having the ability to manage Data Classification settings, you can assign those users the **Classification viewer** role (when using the NetApp Console in standard mode) or the **Compliance Viewer** role (when using the NetApp Console in restricted mode). Learn more.

Can anyone access the private data sent between my browser and Data Classification?

No. The private data sent between your browser and the Data Classification instance are secured with end-to-end encryption using TLS 1.2, which means NetApp and non-NetApp parties can't read it. Data Classification won't share any data or results with NetApp unless you request and approve access.

The data that is scanned stays within your environment.

How is sensitive data handled?

NetApp does not have access to sensitive data and does not display it in the UI. Sensitive data is masked, for example, the last four numbers are displayed for credit card information.

Where is the data stored?

Scan results are stored in Elasticsearch within your Data Classification instance.

How is the data accessed?

Data Classification accesses data stored in Elasticsearch through API calls, which require authentication and are encrypted using AES-128. Accessing Elasticsearch directly requires root access.

Licenses and costs

The following question relates to licensing and costs to use Data Classification.

How much does Data Classification cost?

Data Classification is a NetApp Console core capability. It's not charged.

Console agent deployment

The following questions relate to the Console agent.

What is the Console agent?

The Console agent is software running on a compute instance either within your cloud account, or on-premises, that enables the NetApp Console to securely manage cloud resources. You must deploy a Console agent to use Data Classification.

Where does the Console agent need to be installed?

When scanning data, the NetApp Console Console agent needs to be installed in the following locations:

- For Cloud Volumes ONTAP in AWS or Amazon FSx for ONTAP: Console agent is in AWS.
- For Cloud Volumes ONTAP in Azure or in Azure NetApp Files: Console agent is in Azure.
- For Cloud Volumes ONTAP in GCP: Console agent is in GCP.
- For on-premises ONTAP systems: Console agent is on-premises.

If you have data in these locations, you may need to use multiple Console agents.

Does Data Classification require access to credentials?

Data Classification itself doesn't retrieve storage credentials. Instead, they are stored within the Console agent.

Data Classification uses data plane credentials, for example, CIFS credentials to mount shares before scanning.

Does communication between the service and the Console agent use HTTP?

Yes, Data Classification communicates with the Console agent using HTTP.

Data Classification deployment

The following questions relate to the separate Data Classification instance.

What deployment models does Data Classification support?

The NetApp Console allows the user to scan and report on systems virtually anywhere, including on-premises, cloud, and hybrid environments. Data Classification is normally deployed using a SaaS model, in which the service is enabled via the Console interface and requires no hardware or software installation. Even in this click-and-run deployment mode, data management can be done regardless of whether the data stores are on premises or in the public cloud.

What type of instance or VM is required for Data Classification?

When deployed in the cloud:

- In AWS, Data Classification runs on an m6i.4xlarge instance with a 500 GiB GP2 disk. You can select a smaller instance type during deployment.
- In Azure, Data Classification runs on a Standard_D16s_v3 VM with a 500 GiB disk.
- In GCP, Data Classification runs on an n2-standard-16 VM with a 500 GiB Standard persistent disk.

Learn more about how Data Classification works.

Can I deploy the Data Classification on my own host?

Yes. You can install Data Classification software on a Linux host that has internet access in your network or in the cloud. Everything works the same and you continue to manage your scan configuration and results through the Console. See Deploying Data Classification on premises for system requirements and installation details.

What about secure sites without internet access?

Yes, that's also supported. You can deploy Data Classification in an on-premises site that doesn't have internet access for completely secure sites.

Legal notices

Legal notices provide access to copyright statements, trademarks, patents, and more.

Copyright

https://www.netapp.com/company/legal/copyright/

Trademarks

NETAPP, the NETAPP logo, and the marks listed on the NetApp Trademarks page are trademarks of NetApp, Inc. Other company and product names may be trademarks of their respective owners.

https://www.netapp.com/company/legal/trademarks/

Patents

A current list of NetApp owned patents can be found at:

https://www.netapp.com/pdf.html?item=/media/11887-patentspage.pdf

Privacy policy

https://www.netapp.com/company/legal/privacy-policy/

Open source

Notice files provide information about third-party copyright and licenses used in NetApp software.

- Notice for NetApp Console
- Notice for NetApp Data Classification