



BlueXP classification documentation

BlueXP classification

NetApp
September 12, 2024

This PDF was generated from <https://docs.netapp.com/us-en/bluexp-classification/index.html> on September 12, 2024. Always check docs.netapp.com for the latest.

Table of Contents

BlueXP classification documentation	1
Release notes	2
What's new with BlueXP classification	2
Known limitations	10
Get started	12
Learn about BlueXP classification	12
Deploy BlueXP classification	21
Activate scanning on your data sources	55
Integrate your Active Directory with BlueXP classification	81
Frequently asked questions about BlueXP classification	84
Use BlueXP classification	92
View governance details about the data stored in your organization	92
View compliance details about the private data stored in your organization	97
Categories of private data	104
Investigate the data stored in your organization	110
Assign policies to your data	119
View compliance reports	125
Manage BlueXP classification	132
Exclude specific directories from BlueXP classification scans	132
Define additional group IDs as open to organization	135
Remove data sources from BlueXP classification	136
Uninstalling BlueXP classification	137
Deprecated features	139
BlueXP classification deprecated features	139
Deploy BlueXP classification deprecations	140
Scan data deprecations	148
Manage data deprecations	169
Reference	209
Supported BlueXP classification instance types	209
Metadata collected from data sources	210
Log in to the BlueXP classification system	211
BlueXP classification APIs	212
Knowledge and support	223
Register for support	223
Get help	227
Legal notices	233
Copyright	233
Trademarks	233
Patents	233
Privacy policy	233
Open source	233

BlueXP classification documentation

Release notes

What's new with BlueXP classification

Learn what's new in BlueXP classification (Cloud Data Sense).

2 September 2024 (Version 1.35)

This BlueXP classification release includes the following update.

Scan StorageGRID data

BlueXP classification can now scan data in StorageGRID.

For details, refer to [Scan StorageGRID data](#).

5 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](#).

1 July 2024 (Version 1.33)

This release includes the following updates.

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](#)

5 June 2024 (Version 1.32)

This release includes the following update.

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. 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.

[Learn more about the deprecated features.](#)

1 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](#).

4 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 now display the total size in addition to total number of items

The filtered results in the Investigation page now shows 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. Refer to [Organize your private data](#).

6 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.

4 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.

5 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. [See how](#).

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.](#)

6 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. [See how to obtain and upload your Data Sense BYOL license.](#)

3 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.](#)

7 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.

[Learn more.](#)

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.

[Learn more about adding custom keywords from the BlueXP classification UI.](#)

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, a 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`.

5 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.

[Learn more about adding custom patterns from the BlueXP classification UI.](#)

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. [Learn more about moving source files using BlueXP classification.](#)

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.](#)

9 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

Known limitations identify functions that are not supported by this release of the product, or that do not interoperate correctly with it. Review these limitations carefully.

BlueXP classification release removed 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 personal identifiable information (PII) data by Directories is not available.
- The option to integrate data using Azure Information Protection (AIP) labels has been disabled.

BlueXP classification scanning limitations

BlueXP classification scans only one share under a volume

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

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

Then the data in /A will be scanned. The data in /C and /D will not be 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 working environment, add the volume to be scanned.
2. After BlueXP classification has completed scanning the volume, go to the *Data Investigation* page and create a filter to see which share is being scanned:

You'll filter the data by "Working Environment 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.](#)

You'll need to add all the shares individually, for example:

/C

/D

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

Get started

Learn about BlueXP classification

BlueXP classification (Cloud Data Sense) is a data governance service for BlueXP 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.

IMPORTANT

Starting in May 2024 with version 1.31, BlueXP classification is now available as a core capability within BlueXP at no additional charge. No Classification license or subscription is required. We have also focused BlueXP classification functionality on NetApp storage systems, so some unused, or underused, features have been deprecated.

[See a list of deprecated features.](#)

Users who have been using legacy versions 1.30 or earlier will continue to be able to use that version until their subscription expires.

Features

BlueXP 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 BlueXP classification to provide the following areas of functionality.

[Learn more about the use cases for BlueXP classification.](#)

Maintain compliance

BlueXP classification provides several tools that can help with your compliance efforts. You can use BlueXP 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

BlueXP classification can identify data that is potentially at risk for being accessed for criminal purposes. You can use BlueXP 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

BlueXP classification provides tools that can help with your storage total cost of ownership (TCO). You can use BlueXP 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 working environments and data sources

BlueXP classification can scan and analyze structured and unstructured data from the following types of working environments and data sources:

Working environments

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

BlueXP classification supports NFS versions 3.x, and CIFS versions 1.x, 2.0, 2.1, and 3.0.

Cost

BlueXP classification is now free to use. No Classification license or paid subscription is required.

Infrastructure costs

- Installing BlueXP 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 BlueXP classification on an on-premises system.
- BlueXP classification requires that you have deployed a BlueXP Connector. In many cases you already

have a Connector because of other storage and services you are using in BlueXP. The Connector instance 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 the Connector on an on-premises system.

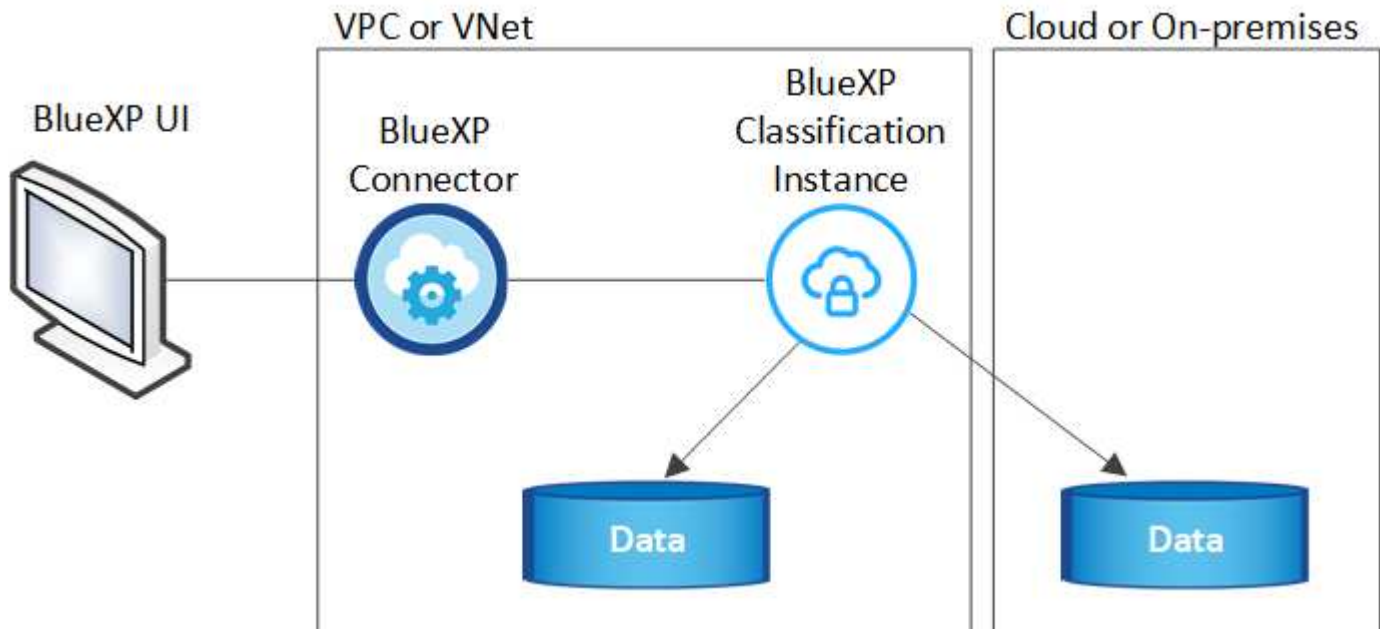
Data transfer costs

Data transfer costs depend on your setup. If the BlueXP 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 BlueXP classification instance

When you deploy BlueXP classification in the cloud, BlueXP deploys the instance in the same subnet as the Connector. [Learn more about Connectors](#).



Note the following about the default instance:

- In AWS, BlueXP 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, BlueXP classification runs on a [Standard_D16s_v3 VM](#) with a 500 GiB disk. The operating system image is CentOS 7.9.
- In GCP, BlueXP classification runs on an [n2-standard-16 VM](#) with a 500 GiB Standard persistent disk. The operating system image is CentOS 7.9.
- In regions where the default instance isn't available, BlueXP 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:

- Only one BlueXP classification instance is deployed per Connector.

You can also deploy BlueXP 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 BlueXP classification software is automated as long as the instance has internet access.



The instance should remain running at all times because BlueXP classification continuously scans the data.

Deploy on different instance types

You can deploy BlueXP classification on a system with fewer CPUs and less RAM.

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

How BlueXP classification works

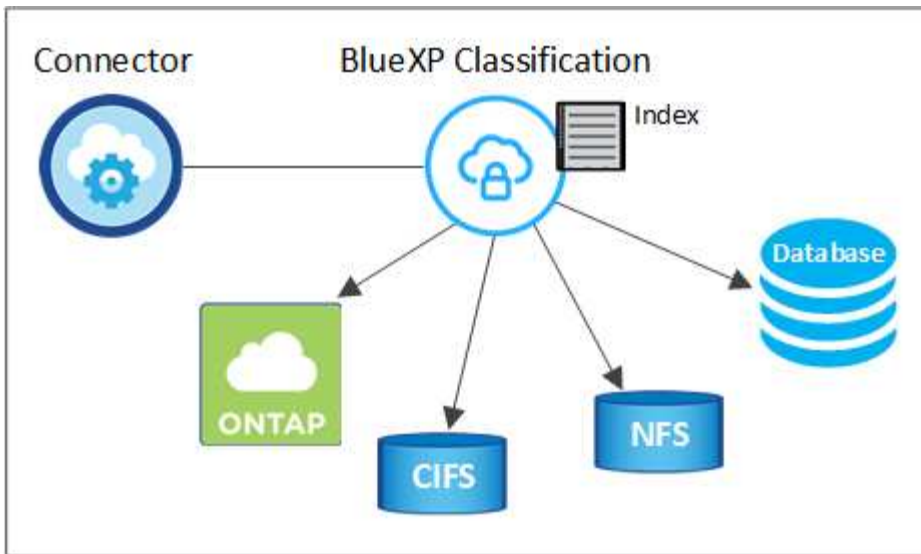
At a high-level, BlueXP classification works like this:

1. You deploy an instance of BlueXP classification in BlueXP.
2. You enable high-level mapping or deep-level scanning on one or more data sources.
3. BlueXP classification scans the data using an AI learning process.
4. You use the provided dashboards and reporting tools to help in your compliance and governance efforts.

How scans work

After you enable BlueXP 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 BlueXP 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.

BlueXP 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, BlueXP 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

BlueXP classification enables you to run a general "mapping" scan on selected 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.

Many users like this functionality because they want to quickly scan their data to identify the data sources that require more research - and then they can enable classification scans only on those required data sources or volumes.

The table below shows some of the differences:

Feature	Classification	Mapping
Scan speed	Slow	Fast
Pricing	Free	Free
Capacity	Limited to 500 TB	Limited to 500 TB
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 policies that provide custom search results	Yes	No
Ability to run other reports	Yes	No

Feature	Classification	Mapping
Ability to see metadata from files*	No	Yes

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

- 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

Governance dashboard differences:

Feature	Map & Classify	Map
Stale data	Yes	Yes
Non-business data	Yes	Yes
Duplicated files	Yes	Yes
Predefined policies	Yes	No
Custom policies	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
Policies	Yes	Yes
Working environment type	Yes	Yes
Working environment	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

How quickly does BlueXP classification scan data

The scan speed is affected by network latency, disk latency, network bandwidth, environment size, and file distribution sizes.

- When performing Mapping scans, BlueXP classification can scan between 100-150 TiBs of data per day.

- When performing Classification scans, BlueXP classification can scan between 15-40 TiBs of data per day.

Information that BlueXP classification categorizes

BlueXP classification collects, indexes, and assigns categories to your data (files). The data that BlueXP classification indexes includes the following:

- **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. [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 GDPR and other privacy regulations. [Learn more about sensitive personal data.](#)
- **Categories:** BlueXP 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.](#)
- **Types:** BlueXP classification takes the data that it scanned and breaks it down by file type. [Learn more about types.](#)
- **Name entity recognition:** BlueXP classification uses AI to extract people's natural names from documents. [Learn about responding to Data Subject Access Requests.](#)

Networking overview

BlueXP deploys the BlueXP classification instance with a security group that enables inbound HTTP connections from the Connector instance.

When using BlueXP in SaaS mode, the connection to BlueXP is served over HTTPS, and the private data sent between your browser and the BlueXP 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 BlueXP classification software and to send usage metrics.

If you have strict networking requirements, [learn about the endpoints that BlueXP classification contacts.](#)

User roles in BlueXP classification

The role each user has been assigned provides different capabilities within BlueXP and within BlueXP classification:

- An **Account Admin** can manage compliance settings and view compliance information for all working environments.
- A **Workspace Admin** can manage compliance settings and view compliance information only for systems that they have permissions to access. If a Workspace Admin can't access a working environment in BlueXP, then they can't see any compliance information for the working environment in the BlueXP classification tab.
- Users with the **Compliance Viewer** role can only view compliance information and generate reports for systems that they have permission to access. These users cannot enable/disable scanning of volumes, buckets, or database schemas.

[Learn more about BlueXP roles](#) and how to [add users with specific roles.](#)

Deploy BlueXP classification

Which BlueXP classification deployment should you use?

You can deploy BlueXP classification in different ways. Learn which method meets your needs.

BlueXP classification can be deployed in the following ways:

- [Deploy in the cloud using BlueXP](#). BlueXP will deploy the BlueXP classification instance in the same cloud provider network as the BlueXP Connector.
- [Install on a Linux host with internet access](#). Install BlueXP 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 BlueXP classification instance that's also located on premises—but this is not 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, is good for your secure sites.

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 BlueXP 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 BlueXP classification](#).

Deploy BlueXP classification in the cloud using BlueXP

Complete a few steps to deploy BlueXP classification in the cloud. BlueXP will deploy the BlueXP classification instance in the same cloud provider network as the BlueXP Connector.

Note that you can also [install BlueXP 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 BlueXP 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 Connector

If you don't already have a Connector, create a Connector now. See [creating a Connector in AWS](#), [creating a Connector in Azure](#), or [creating a Connector in GCP](#).

You can also [install the Connector on-premises](#) on a Linux host in your network or on a Linux host in the cloud.

2

Review prerequisites

Ensure that your environment can meet the prerequisites. This includes outbound internet access for the instance, connectivity between the Connector and BlueXP classification over port 443, and more. [See the complete list](#).



Deploy BlueXP classification

Launch the installation wizard to deploy the BlueXP classification instance in the cloud.

Create a Connector

If you don't already have a Connector, create a Connector in your cloud provider. See [creating a Connector in AWS](#) or [creating a Connector in Azure](#), or [creating a Connector in GCP](#). In most cases you will probably have a Connector set up before you attempt to activate BlueXP classification because most [BlueXP features require a Connector](#), but there are cases where you'll need to set one up now.

There are some scenarios where you have to use a Connector 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 Connector in AWS.
- When scanning data in Cloud Volumes ONTAP in Azure or in Azure NetApp Files, you use a Connector 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 Connector in GCP.

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

Note that you can also [install the Connector on-premises](#) on a Linux host in your network or in the cloud. Some users planning to install BlueXP classification on-prem may also choose to install the Connector on-prem.

As you can see, there may be some situations where you need to use [multiple Connectors](#).

Government region support

BlueXP classification is supported when the Connector is deployed in a Government region (AWS GovCloud, Azure Gov, or Azure DoD). When deployed in this manner, BlueXP classification has the following restrictions:

[See more information about deploying the Connector in a Government region](#).

Review prerequisites

Review the following prerequisites to make sure that you have a supported configuration before you deploy BlueXP classification in the cloud. When you deploy BlueXP classification in the cloud, it's located in the same subnet as the Connector.

Enable outbound internet access from BlueXP classification

BlueXP classification requires outbound internet access. If your virtual or physical network uses a proxy server for internet access, ensure that the BlueXP classification instance has outbound internet access to contact the following endpoints. The proxy must be non-transparent - we don't currently support transparent proxies.

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

Required endpoints for AWS

Endpoints	Purpose
https://api.bluexp.netapp.com	Communication with the BlueXP service, which includes NetApp accounts.
https://netapp-cloud-account.auth0.com https://auth0.com	Communication with the BlueXP 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://dseasb33srrn.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 BlueXP classification to access and download manifests and templates, and to send logs and metrics.

Required endpoints for Azure

Endpoints	Purpose
https://api.bluexp.netapp.com	Communication with the BlueXP service, which includes NetApp accounts.
https://netapp-cloud-account.auth0.com https://auth0.com	Communication with the BlueXP 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://dseasb33srrn.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.

Required endpoints for GCP

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

Endpoints	Purpose
https://netapp-cloud-account.auth0.com https://auth0.com	Communication with the BlueXP 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://dseasb33srrn.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.

Ensure that BlueXP has the required permissions

Ensure that BlueXP has permissions to deploy resources and create security groups for the BlueXP classification instance. You can find the latest BlueXP permissions in [the policies provided by NetApp](#).

Ensure that the BlueXP Connector can access BlueXP classification

Ensure connectivity between the Connector and the BlueXP classification instance. The security group for the Connector must allow inbound and outbound traffic over port 443 to and from the BlueXP classification instance. This connection enables deployment of the BlueXP classification instance and enables you to view information in the Compliance and Governance tabs. BlueXP 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 Connector in AWS](#) for details.

Additional inbound and outbound security group rules are required for Azure and Azure Government deployments. See [Rules for the Connector in Azure](#) for details.

Ensure that you can keep BlueXP classification running

The BlueXP classification instance needs to stay on to continuously scan your data.

Ensure web browser connectivity to BlueXP classification

After BlueXP classification is enabled, ensure that users access the BlueXP interface from a host that has a connection to the BlueXP classification instance.

The BlueXP 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 BlueXP 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 BlueXP 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 BlueXP 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 BlueXP classification in the cloud

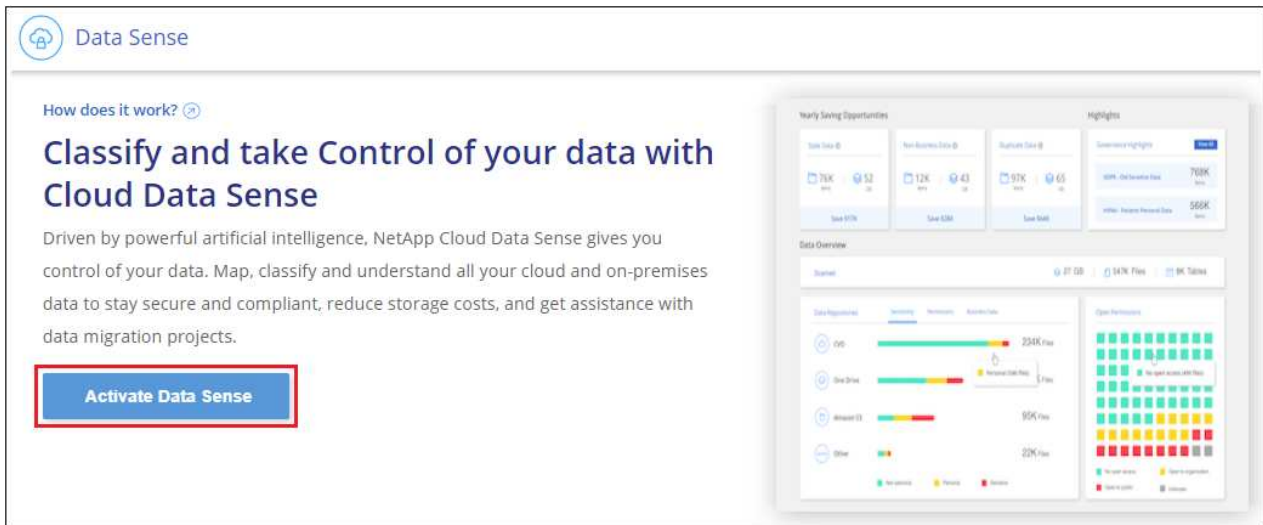
Follow these steps to deploy an instance of BlueXP classification in the cloud. The Connector will deploy the instance in the cloud, and then install BlueXP classification software on that instance.

In regions where the default instance type isn't available, BlueXP classification runs on an [alternate instance type](#).

Deploy in AWS

Steps

1. From the BlueXP left navigation menu, click **Governance > Classification**.



2. Click **Activate Data Sense**.
3. From the *Installation* page, click **Deploy > Deploy** to use the "Large" instance size and 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 BlueXP classification is installed, click **Continue to configuration** to go to the *Configuration* page.

Deploy in Azure

Steps

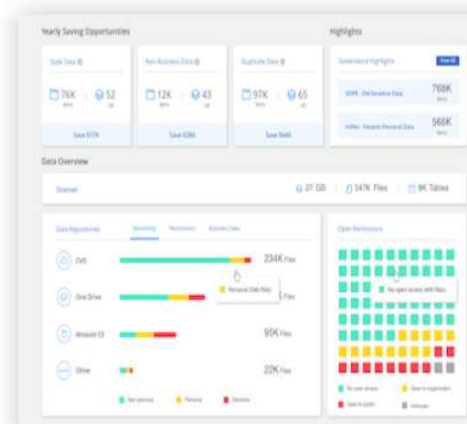
1. From the BlueXP left navigation menu, click **Governance > Classification**.
2. Click **Activate Data Sense**.

How does it work? ⓘ

Classify and take Control of your data with Cloud Data Sense

Driven by powerful artificial intelligence, NetApp Cloud Data Sense gives you control of your data. Map, classify and understand all your cloud and on-premises data to stay secure and compliant, reduce storage costs, and get assistance with data migration projects.

Activate Data Sense



3. Click **Deploy** to start the cloud deployment wizard.

Install your Data Sense instance

Select your preferred deployment location:

[Learn more about deploying Data Sense ⓘ](#)

Cloud Environment

I want BlueXP to deploy the instance and install Data Sense

Deploy

- > BlueXP will deploy a new machine automatically in the chosen cloud environment.
- > You will be taken to an installation wizard where you can configure your Data Sense installation.

I deployed an instance and I'm ready to install Data Sense

Deploy

On Premise

I prepared a local machine and I'm ready to install Data Sense

Deploy

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.

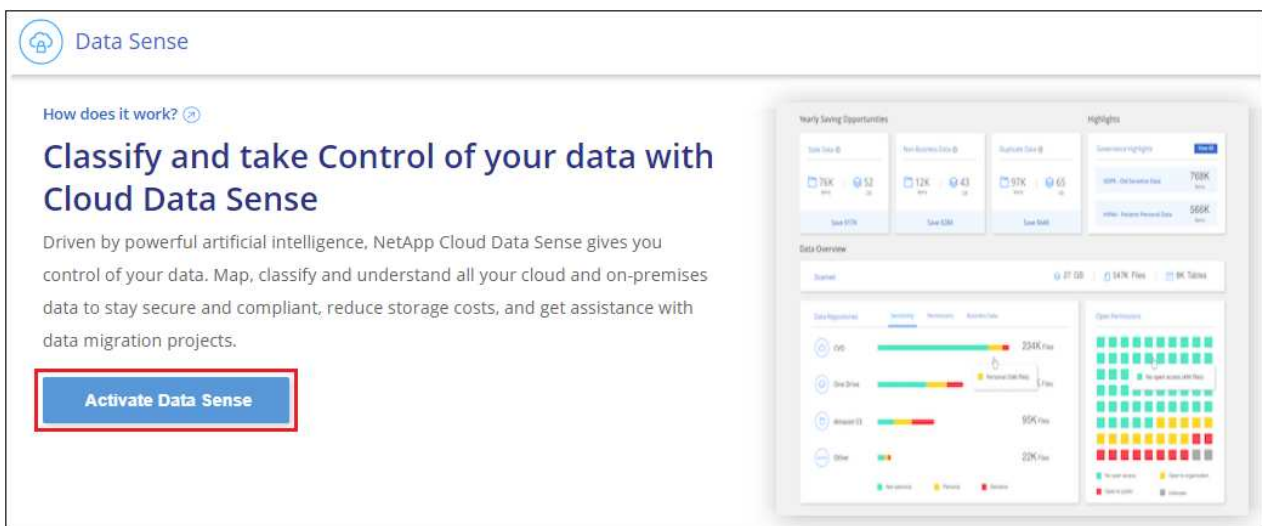


- When the instance is deployed and BlueXP classification is installed, click **Continue to configuration** to go to the *Configuration* page.

Deploy in Google Cloud

Steps

- From the BlueXP left navigation menu, click **Governance > Classification**.
- Click **Activate Data Sense**.





- Click **Deploy** to start the cloud deployment wizard.

Install your Data Sense instance

Select your preferred deployment location:



[Learn more about deploying Data Sense](#)

Cloud Environment



 **I want BlueXP to deploy the instance and install Data Sense** Deploy 

> BlueXP will deploy a new machine automatically in the chosen cloud environment.

> You will be taken to an installation wizard where you can configure your Data Sense installation.

 **I deployed an instance and I'm ready to install Data Sense** Deploy 


On Premise

 **I prepared a local machine and I'm ready to install Data Sense** Deploy 

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.

Deploying Cloud Data Sense

This may take up to 15 minutes. Check this page periodically to make sure the deployment continues successfully.



Deploying Cloud Data Sense instance

Verify connectivity to BlueXP Connector and to the internet

Initializing Cloud Data Sense

[Cancel deployment](#)

5. When the instance is deployed and BlueXP classification is installed, click **Continue to configuration** to go to the *Configuration* page.

Result

BlueXP deploys the BlueXP classification instance in your cloud provider.

Upgrades to the BlueXP Connector and BlueXP 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 BlueXP classification on a host that has internet access

Complete a few steps to install BlueXP classification on a Linux host in your network, or on a Linux host in the cloud, that has internet access. You'll need to deploy the Linux host manually in your network or in the cloud as part of this installation.

The on-prem installation may be a good option if you prefer to scan on-premises ONTAP systems using a BlueXP 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.

The BlueXP 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 BlueXP 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 BlueXP 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.

For very large configurations where you'll be scanning petabytes of data, on versions 1.30 and earlier, you can include multiple hosts to provide additional processing power. When using multiple host systems, the primary system is called the *Manager node*, and the additional systems that provide extra processing power are called *Scanner nodes*.



For legacy versions 1.30 and earlier, if you need to install BlueXP classification on multiple hosts, refer to [Install BlueXP classification on multiple hosts with no internet access.](#)

You can also [install BlueXP classification in an on-premises site that doesn't have internet access](#) for completely secure sites.



For legacy versions 1.30 and earlier, to add scanner nodes, refer to [Add scanner nodes to an existing deployment.](#)

Quick start

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

1

Create a Connector

If you don't already have a Connector, [deploy the Connector on-premises](#) on a Linux host in your network, or on a Linux host in the cloud.

You can also create a Connector with your cloud provider. See [creating a Connector in AWS](#), [creating a Connector in Azure](#), or [creating a Connector in GCP](#).

2

Review prerequisites

Ensure that your environment can meet the prerequisites. This includes outbound internet access for the instance, connectivity between the Connector and BlueXP 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 BlueXP classification

Download the Cloud BlueXP 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 BlueXP classification instance.

Create a Connector

A BlueXP Connector is required before you can install and use BlueXP classification. In most cases you'll probably have a Connector set up before you attempt to activate BlueXP classification because most [BlueXP features require a Connector](#), but there are cases where you'll need to set one up now.

To create one in your cloud provider environment, see [creating a Connector in AWS](#), [creating a Connector in Azure](#), or [creating a Connector in GCP](#).

There are some scenarios where you have to use a Connector 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 Connector in AWS.
- When scanning data in Cloud Volumes ONTAP in Azure or in Azure NetApp Files, you use a Connector 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 Connector in GCP.

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

Note that you can also [deploy the Connector on-premises](#) on a Linux host in your network or on a Linux host in the cloud. Some users planning to install BlueXP classification on-prem may also choose to install the Connector on-prem.

You'll need the IP address or host name of the Connector system when installing BlueXP classification. You'll have this information if you installed the Connector in your premises. If the Connector is deployed in the cloud, you can find this information from the BlueXP console: click the Help icon, select **Support**, and click **BlueXP Connector**.

Prepare the Linux host system

BlueXP 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 BlueXP classification running. The BlueXP classification machine needs to stay on to continuously scan your data.

- BlueXP classification is not supported on a host that is shared with other applications - the host must be a dedicated host.

- When building the host system in your premises, you can choose among these system sizes depending on the size of the dataset that you plan to have BlueXP classification scan.

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
Large	16 CPUs	64 GB RAM	500 GiB SSD on /, or - 100 GiB available on /opt - 395 GiB available on /var/lib/docker or for Podman /var/lib/containers or for Podman /var/lib/containers - 5 GiB on /tmp

- When deploying a compute instance in the cloud for your BlueXP classification installation, we recommend a system that meets the "Large" system requirements above:
 - **Amazon Elastic Compute Cloud (Amazon EC2) instance type:** We recommend "m6i.4xlarge". [See additional AWS instance types.](#)
 - **Azure VM size:** We recommend "Standard_D16s_v3". [See additional Azure instance types.](#)
 - **GCP machine type:** We recommend "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 BlueXP classification version 1.23 or greater)
 - Ubuntu 24.04 (requires BlueXP classification version 1.23 or greater)
 - 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

Note that the following features are not currently supported when using RHEL 8.x and RHEL 9.x:

- Installation in a dark site

- Distributed scanning; using a master scanner node and remote scanner nodes
- **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 BlueXP classification:
 - Depending on the OS you are using, you'll need to install one of the container engines:
 - Docker Engine version 19.3.1 or greater. [View installation instructions](#).
 - [Watch this video](#) for a quick demo of installing Docker on CentOS.
 - 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 BlueXP classification system to use a Network Time Protocol (NTP) service. The time must be synchronized between the BlueXP classification system and the BlueXP Connector system.
 - **Firewalld considerations:** If you are planning to use `firewalld`, we recommend that you enable it before installing BlueXP classification. Run the following commands to configure `firewalld` so that it is compatible with BlueXP 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 BlueXP 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 BlueXP classification host system can't be changed after installation.

Enable outbound internet access from BlueXP classification

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

following endpoints.

Endpoints	Purpose
https://api.bluexp.netapp.com	Communication with the BlueXP service, which includes NetApp accounts.
https://netapp-cloud-account.auth0.com https://auth0.com	Communication with the BlueXP 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://dseasb33srrn.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://mirror.centos.org http://mirrorlist.centos.org http://mirror.centos.org/centos/7/extras/x86_64/Packages/container-selinux-2.107-3.el7.noarch.rpm	Provides prerequisite packages for CentOS installation.
http://packages.ubuntu.com/ http://archive.ubuntu.com	Provides prerequisite packages for Ubuntu installation.

Verify that all required ports are enabled

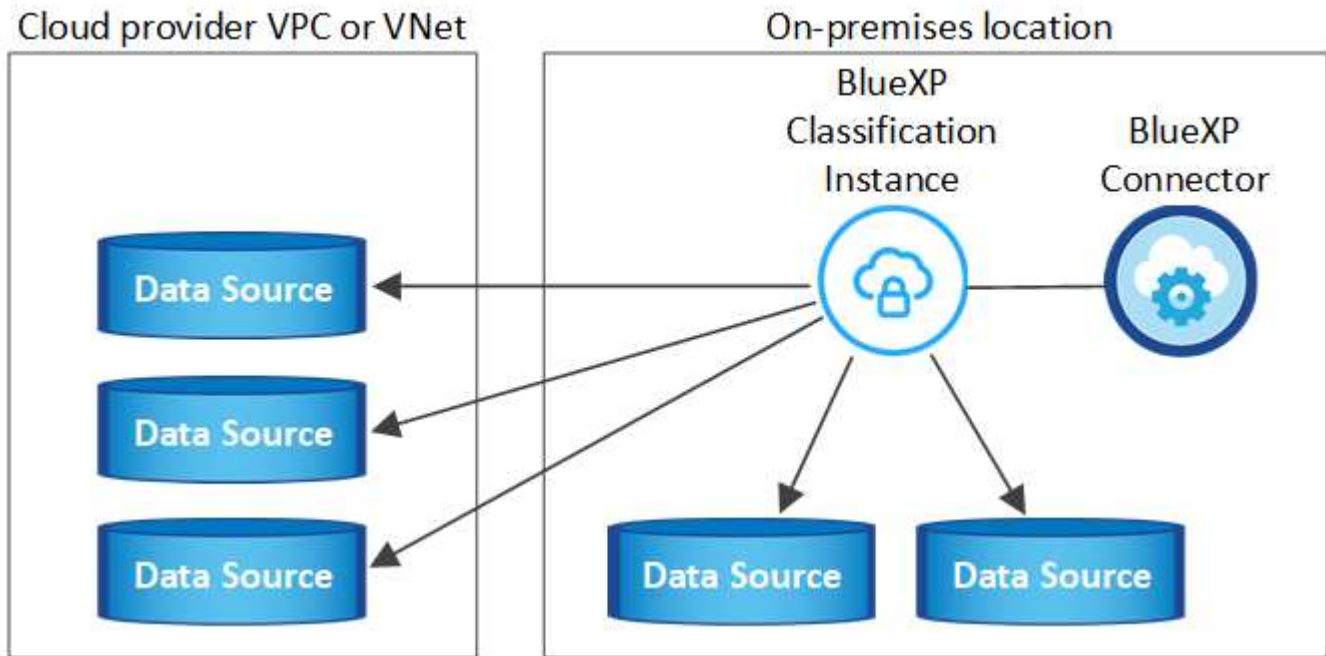
You must ensure that all required ports are open for communication between the Connector, BlueXP classification, Active Directory, and your data sources.

Connection Type	Ports	Description
Connector <> BlueXP classification	8080 (TCP), 443 (TCP), and 80. 9000	<p>The firewall or routing rules for the Connector must allow inbound and outbound traffic over port 443 to and from the BlueXP classification instance.</p> <p>Make sure port 8080 is open so you can see the installation progress in BlueXP.</p> <p>If a firewall is used on the Linux host, port 9000 is required for internal processes within an Ubuntu server.</p>

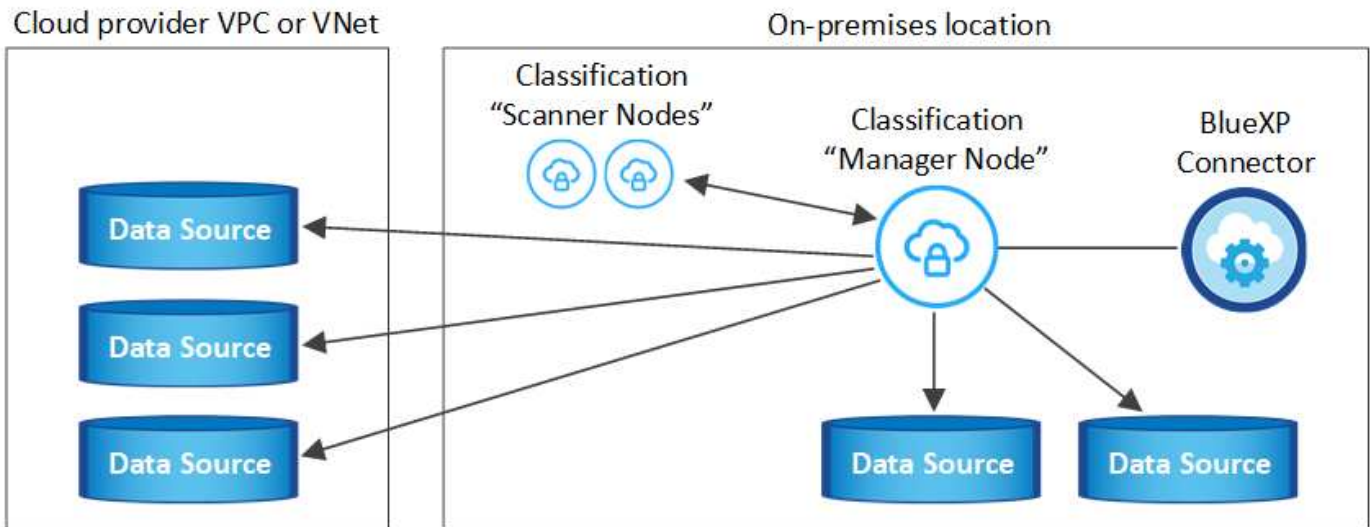
Connection Type	Ports	Description
Connector <> ONTAP cluster (NAS)	443 (TCP)	<p>BlueXP discovers ONTAP clusters using HTTPS. If you use custom firewall policies, they must meet the following requirements:</p> <ul style="list-style-type: none"> • The Connector host must allow outbound HTTPS access through port 443. If the Connector 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 Connector host.
BlueXP classification <> ONTAP cluster	<ul style="list-style-type: none"> • For NFS - 111 (TCP\UDP) and 2049 (TCP\UDP) • For CIFS - 139 (TCP\UDP) and 445 (TCP\UDP) 	<p>BlueXP 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 BlueXP classification instance.</p> <p>Make sure these ports are open to the BlueXP classification instance:</p> <ul style="list-style-type: none"> • For NFS - 111 and 2049 • For CIFS - 139 and 445 <p>NFS volume export policies must allow access from the BlueXP classification instance.</p>
BlueXP classification <> Active Directory	389 (TCP & UDP), 636 (TCP), 3268 (TCP), and 3269 (TCP)	<p>You must have an Active Directory already set up for the users in your company. Additionally, BlueXP classification needs Active Directory credentials to scan CIFS volumes.</p> <p>You must have the information for the Active Directory:</p> <ul style="list-style-type: none"> • 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 BlueXP classification on the Linux host

For typical configurations you'll install the software on a single host system. [See those steps here](#).



For very large configurations where you'll be scanning petabytes of data, you can include multiple hosts to provide additional processing power. Learn more xref.:/task-deploy-multi-host-install-dark-site.html about installing on multiple hosts for large configurations.



See [Preparing the Linux host system](#) and [Reviewing prerequisites](#) for the full list of requirements before you deploy BlueXP classification.

Upgrades to BlueXP classification software is automated as long as the instance has internet connectivity.



BlueXP 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 Connector and instance of BlueXP 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 BlueXP classification software on a single on-premises host.

[Watch this video](#) to see how to install BlueXP classification.

Note that all installation activities are logged when installing BlueXP 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/`. [See more details here](#).

What you'll need

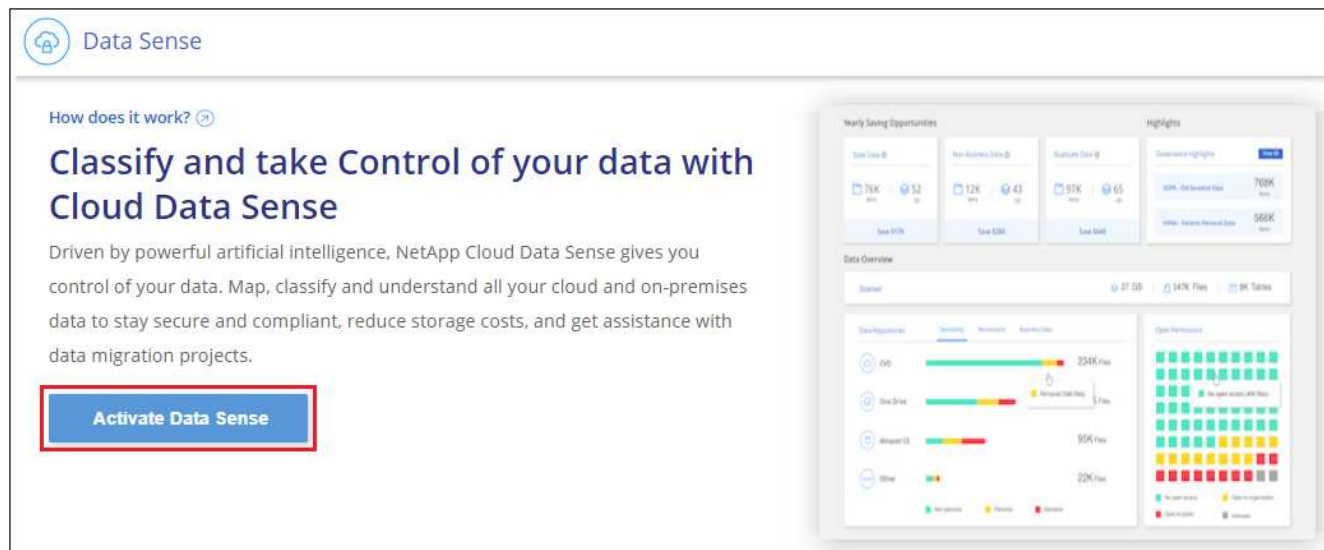
- Verify that your Linux system meets the [host requirements](#).
- Verify that the system has the two prerequisite software packages installed (Docker Engine or Podman, and Python 3).
- Make sure you have root privileges on the Linux system.
- If you're using a proxy for access to the internet:
 - You'll 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 BlueXP classification Linux system where the TLS CA certificates are stored.
 - The proxy must be non-transparent - we don't 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

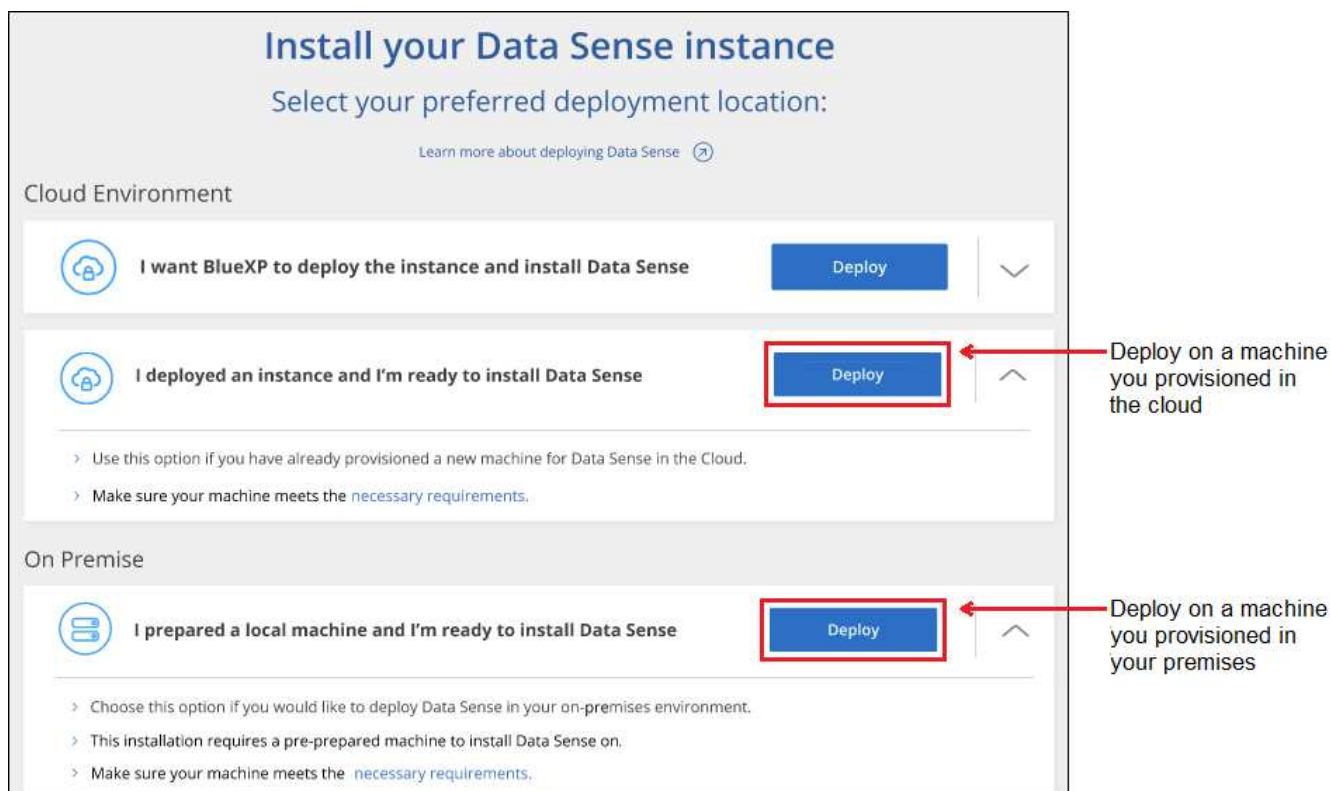
1. Download the BlueXP classification software from the [NetApp Support Site](#). The file you should select is named **DATASENSE-INSTALLER-<version>.tar.gz**.
2. Copy the installer file to the Linux host you plan to use (using `scp` or some other method).
3. Unzip the installer file on the host machine,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

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

4. In BlueXP, select **Governance > Classification**.
5. Click **Activate Data Sense**.



- Depending on whether you are installing BlueXP classification on an instance you prepared in the cloud or on an instance you prepared in your premises, click the appropriate **Deploy** button to start the BlueXP classification installation.



- The *Deploy Data Sense 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 click **Close** to dismiss the dialog.
- 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:
<ol style="list-style-type: none"> 1. Paste the command you copied from step 7: <pre>sudo ./install.sh -a <account_id> -c <client_id> -t <user_token></pre> <p>If you are installing on a cloud instance (not on your premises), add <code>--manual-cloud-install <cloud_provider></code>.</p> 2. Enter the IP address or host name of the BlueXP classification host machine so it can be accessed by the Connector system. 3. Enter the IP address or host name of the BlueXP Connector host machine so it can be accessed by the BlueXP classification system. 4. Enter proxy details as prompted. If your BlueXP Connector already uses a proxy, there is no need to enter this information again here since BlueXP classification will automatically use the proxy used by the Connector. 	<p>Alternatively, you can create the whole command in advance, providing the necessary host and proxy parameters:</p> <pre>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></pre>

Variable values:

- *account_id* = NetApp Account ID
- *client_id* = Connector 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 BlueXP classification Linux system.
- *cm_host* = IP address or host name of the BlueXP Connector 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 BlueXP classification Linux system containing additional TLS CA certificate bundles. Only required if the proxy is performing TLS interception.

Result

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

If there is connectivity over port 8080 between the host machine and the Connector instance, you'll see the installation progress in the BlueXP classification tab in BlueXP.

What's Next

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

Install BlueXP classification on a Linux host with no internet access

Complete a few steps to install BlueXP classification on a Linux host in an on-premises site that doesn't have internet access - also known as *private mode*. This type of installation is perfect for your secure sites.

[Learn about the different deployment modes for the BlueXP Connector and BlueXP classification.](#)

Note that you can also [deploy BlueXP classification in an on-premises site that has internet access.](#)

The BlueXP 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 BlueXP 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 BlueXP classification.](#)



For legacy versions 1.30 and earlier, if you need to install BlueXP classification on multiple hosts, refer to [Install BlueXP classification on multiple hosts with no internet access.](#)

Supported data sources

When installed private mode (sometimes called an "offline" or "dark" site), BlueXP classification can only scan data from data sources that are also local to the on-premises site. At this time, BlueXP classification can scan the following **local** data sources:

- On-premises ONTAP systems
- Database schemas

There is no support currently for scanning Cloud Volumes ONTAP, Azure NetApp Files, or FSx for ONTAP accounts when BlueXP classification is deployed in private mode.

Limitations

Most BlueXP classification features work when it is deployed in a site with no internet access. However, certain features that require internet access are not supported, for example:

- Setting BlueXP roles for different users (for example, Account Admin or Compliance Viewer)
- Copying and synchronizing source files using BlueXP copy and sync
- Automated software upgrades from BlueXP

Both the BlueXP Connector and BlueXP classification will require periodic manual upgrades to enable new features. You can see the BlueXP classification version at the bottom of the BlueXP classification UI pages. Check the [BlueXP classification Release Notes](#) to see the new features in each release and whether you want those features. Then you can follow the steps to [upgrade the BlueXP Connector](#) and [upgrade your BlueXP classification software.](#)

Quick start

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

1

Install the BlueXP Connector

If you don't already have a Connector installed in private mode, [deploy the Connector](#) on a Linux host now.

2

Review BlueXP classification prerequisites

Ensure that your Linux system meets the [host requirements](#), that it has all required software installed, and that your offline environment meets the required [permissions and connectivity](#).

3

Download and deploy BlueXP classification

Download the BlueXP 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 BlueXP classification instance.

Install the BlueXP Connector

If you don't already have a BlueXP Connector installed in private mode, [deploy the Connector](#) on a Linux host in your offline site.

Prepare the Linux host system

BlueXP classification software must run on a host that meets specific operating system requirements, RAM requirements, software requirements, and so on.

- BlueXP classification is not supported on a host that is shared with other applications - the host must be a dedicated host.
- When building the host system in your premises, you can choose among these system sizes depending on the size of the dataset that you plan to have BlueXP classification scan.

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
Large	16 CPUs	64 GB RAM	500 GiB SSD on /, or - 100 GiB available on /opt - 395 GiB available on /var/lib/docker or for Podman /var/lib/containers or for Podman /var/lib/containers - 5 GiB on /tmp

- When deploying a compute instance in the cloud for your BlueXP classification installation, we recommend a system that meets the "Large" system requirements above:
 - **Amazon Elastic Compute Cloud (Amazon EC2) instance type:** We recommend "m6i.4xlarge". [See additional AWS instance types.](#)
 - **Azure VM size:** We recommend "Standard_D16s_v3". [See additional Azure instance types.](#)
 - **GCP machine type:** We recommend "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 BlueXP classification version 1.23 or greater)
 - Ubuntu 24.04 (requires BlueXP classification version 1.23 or greater)
- 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

Note that the following features are not currently supported when using RHEL 8.x and RHEL 9.x:

- Installation in a dark site
- Distributed scanning; using a master scanner node and remote scanner nodes

- **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 BlueXP classification:

- Depending on the OS you are using, you'll need to install one of the container engines:
 - Docker Engine version 19.3.1 or greater. [View installation instructions.](#)

[Watch this video](#) for a quick demo of installing Docker on CentOS.

- 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 BlueXP classification system to use a Network Time Protocol (NTP) service. The time must be synchronized between the BlueXP classification system and the BlueXP Connector system.

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

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



The IP address of the BlueXP classification host system can't be changed after installation.

Verify BlueXP and BlueXP classification prerequisites

Review the following prerequisites to make sure that you have a supported configuration before you deploy BlueXP classification.

- Ensure that the Connector has permissions to deploy resources and create security groups for the BlueXP classification instance. You can find the latest BlueXP permissions in [the policies provided by NetApp](#).
- Ensure that you can keep BlueXP classification running. The BlueXP classification instance needs to stay on to continuously scan your data.
- Ensure web browser connectivity to BlueXP classification. After BlueXP classification is enabled, ensure that users access the BlueXP interface from a host that has a connection to the BlueXP classification instance.

The BlueXP classification instance uses a private IP address to ensure that the indexed data isn't accessible to others. As a result, the web browser that you use to access BlueXP must have a connection to that private IP address. That connection can come from a host that's inside the same network as the BlueXP classification instance.

Verify that all required ports are enabled

You must ensure that all required ports are open for communication between the Connector, BlueXP classification, Active Directory, and your data sources.

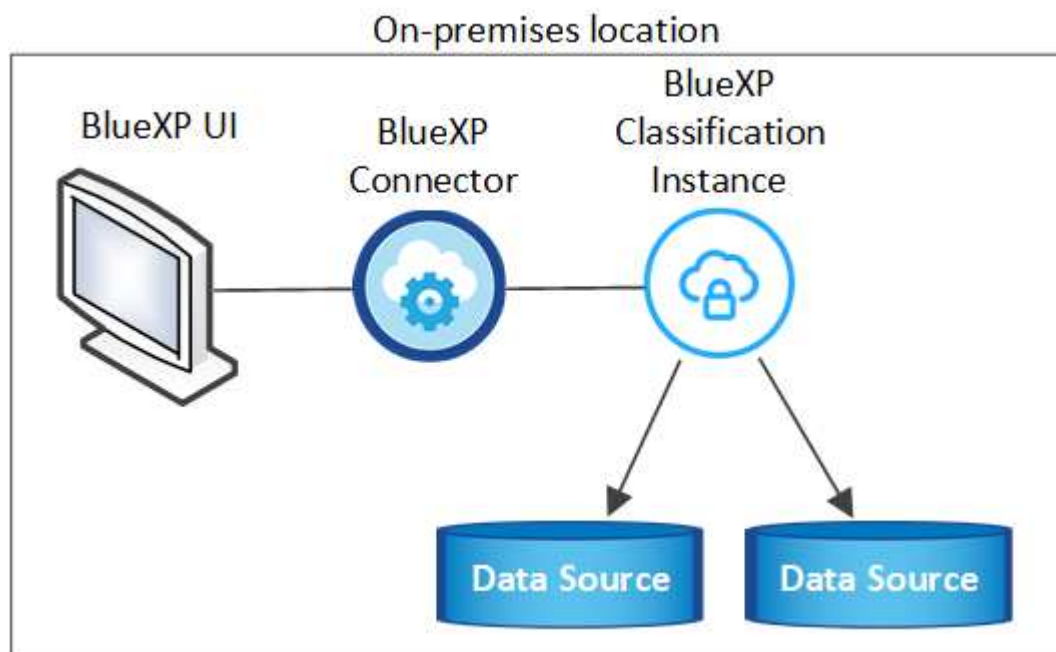
Connection Type	Ports	Description
Connector <> BlueXP classification	8080 (TCP), 6000 (TCP), 443 (TCP), and 80. 9000	<p>The security group for the Connector must allow inbound and outbound traffic over ports 6000 and 443 to and from the BlueXP classification instance.</p> <ul style="list-style-type: none"> • Port 6000 is required so that the BlueXP classification BYOL license works in a dark site. • Port 8080 should be open so you can see the installation progress in BlueXP. • If a firewall is used on the Linux host, port 9000 is required for internal processes within an Ubuntu server.
Connector <> ONTAP cluster (NAS)	443 (TCP)	<p>BlueXP discovers ONTAP clusters using HTTPS. If you use custom firewall policies, they must meet the following requirements:</p> <ul style="list-style-type: none"> • The Connector host must allow outbound HTTPS access through port 443. If the Connector is in the cloud, all outbound communication is allowed by the predefined security group. • 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 Connector host.
BlueXP classification <> ONTAP cluster	<ul style="list-style-type: none"> • For NFS - 111 (TCP\UDP) and 2049 (TCP\UDP) • For CIFS - 139 (TCP\UDP) and 445 (TCP\UDP) 	<p>BlueXP classification needs a network connection to each Cloud Volumes ONTAP subnet or on-prem ONTAP system. Security groups for Cloud Volumes ONTAP must allow inbound connections from the BlueXP classification instance.</p> <p>Make sure these ports are open to the BlueXP classification instance:</p> <ul style="list-style-type: none"> • For NFS - 111 and 2049 • For CIFS - 139 and 445 <p>NFS volume export policies must allow access from the BlueXP classification instance.</p>

Connection Type	Ports	Description
BlueXP classification <> Active Directory	389 (TCP & UDP), 636 (TCP), 3268 (TCP), and 3269 (TCP)	<p>You must have an Active Directory already set up for the users in your company. Additionally, BlueXP classification needs Active Directory credentials to scan CIFS volumes.</p> <p>You must have the information for the Active Directory:</p> <ul style="list-style-type: none"> • 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)
If a firewall used on Linux host	9000	Needed for internal processes within an Ubuntu server.

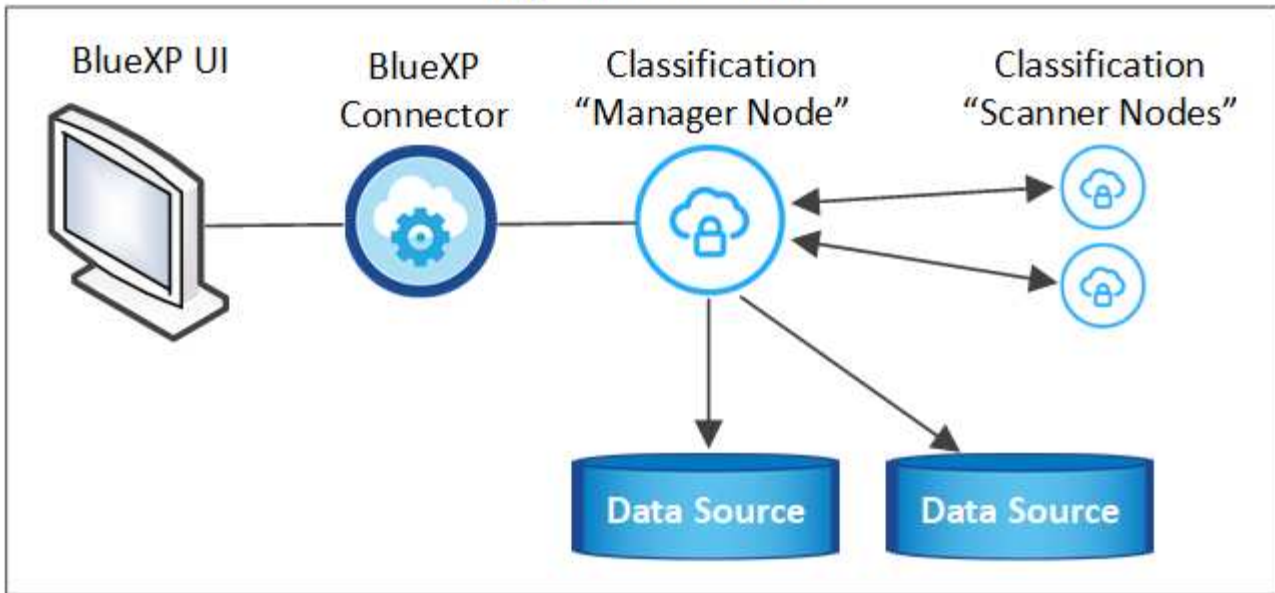
If you are using multiple BlueXP classification hosts to provide additional processing power to scan your data sources, you'll need to enable additional ports/protocols. [See the additional port requirements.](#)

Install BlueXP classification on the on-premises Linux host

For typical configurations you'll install the software on a single host system.



On-premises location



Single-host installation for typical configurations

Follow these steps when installing BlueXP classification software on a single on-premises host in an offline environment.

Note that all installation activities are logged when installing BlueXP 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/`. [See more details here](#).

What you'll need

- Verify that your Linux system meets the [host requirements](#).
- Verify that you have installed the two prerequisite software packages (Docker Engine or Podman, and Python 3).
- Make sure you have root privileges on the Linux system.
- Verify that your offline environment meets the required [permissions and connectivity](#).

Steps

1. On an internet-configured system, download the BlueXP classification software from the [NetApp Support Site](#). The file you should select is named **DataSense-offline-bundle-<version>.tar.gz**.
2. Copy the installer bundle to the Linux host you plan to use in private mode.
3. Unzip the installer bundle on the host machine, for example:

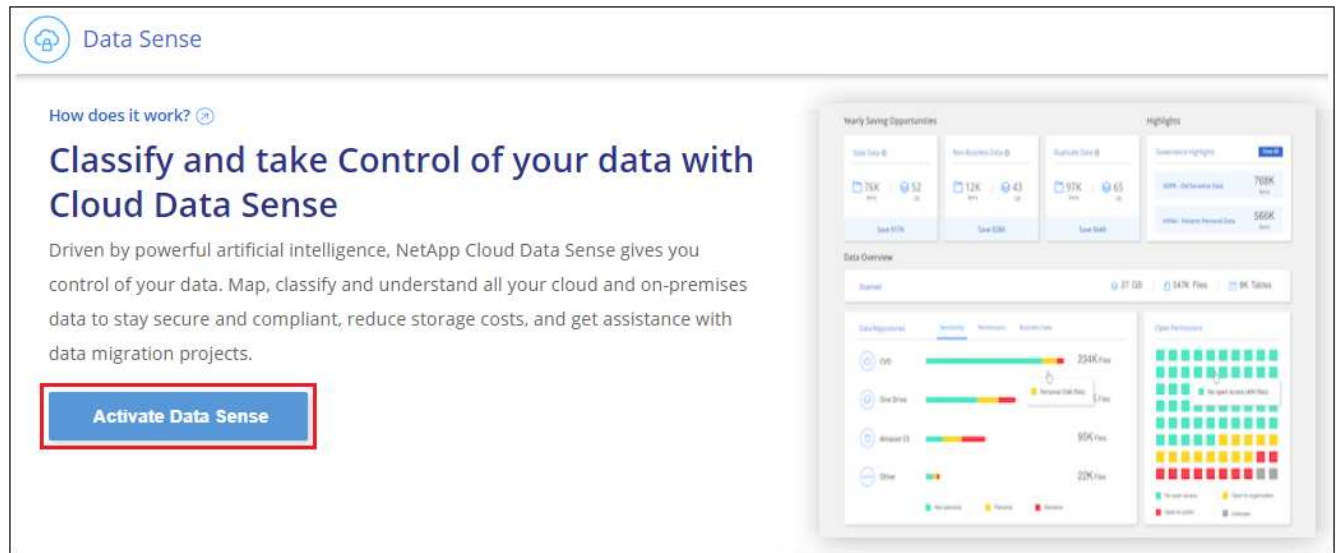
```
tar -xzf DataSense-offline-bundle-v1.25.0.tar.gz
```

This extracts required software and the actual installation file **cc_onprem_installer.tar.gz**.

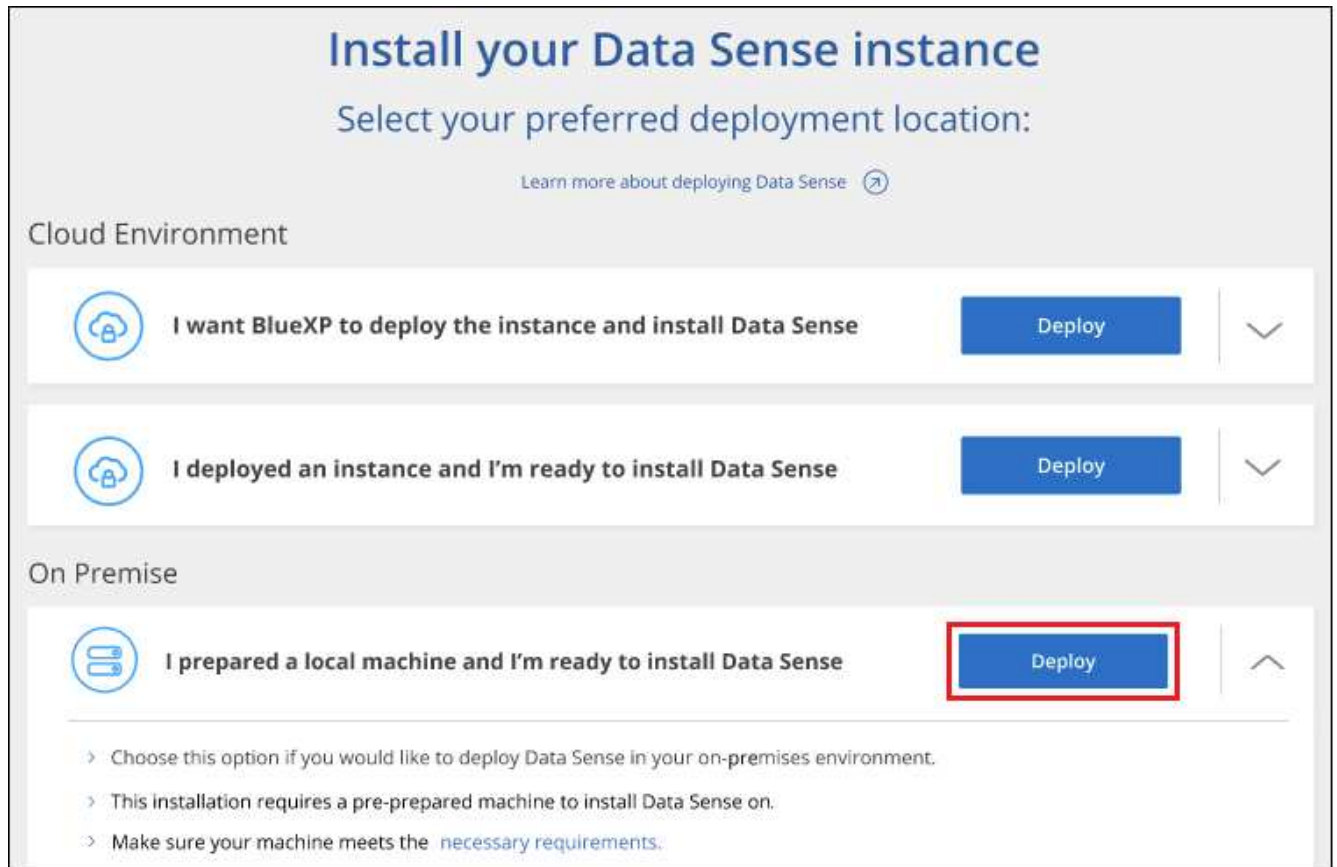
4. Unzip the installation file on the host machine, for example:


```
tar -xzf cc_onprem_installer.tar.gz
```

5. Launch BlueXP and select **Governance > Classification**.
6. Click **Activate Data Sense**.



7. Click **Deploy** to start the on-prem installation.



8. The *Deploy Data Sense On Premises* dialog is displayed. Copy the provided command (for example: `sudo`

`./install.sh -a 12345 -c 27AG75 -t 2198qq --darksite)` and paste it in a text file so you can use it later. Then click **Close** to dismiss the dialog.

9. 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.

Enter parameters as prompted:	Enter the full command:
<ol style="list-style-type: none">1. Paste the information you copied from step 8: <code>sudo ./install.sh -a <account_id> -c <client_id> -t <user_token> --darksite</code>2. Enter the IP address or host name of the BlueXP classification host machine so it can be accessed by the Connector system.3. Enter the IP address or host name of the BlueXP Connector host machine so it can be accessed by the BlueXP classification system.	Alternatively, you can create the whole command in advance, providing the necessary host parameters: <code>sudo ./install.sh -a <account_id> -c <client_id> -t <user_token> --host <ds_host> --manager-host <cm_host> --no-proxy --darksite</code>

Variable values:

- `account_id` = NetApp Account ID
- `client_id` = Connector 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 BlueXP classification system.
- `cm_host` = IP address or host name of the BlueXP Connector system.

Result

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

If there is connectivity over port 8080 between the host machine and the Connector instance, you'll see the installation progress in the BlueXP classification tab in BlueXP.

What's Next

From the Configuration page you can select the local [on-prem ONTAP clusters](#) and [databases](#) that you want to scan.

Upgrade BlueXP classification software

Since BlueXP classification software is updated with new features on a regular basis, you should get into a routine to check for new versions periodically to make sure you're using the newest software and features. You'll need to upgrade BlueXP classification software manually because there's no internet connectivity to perform the upgrade automatically.

Before you begin

- We recommend that your BlueXP Connector software is upgraded to the newest available version. [See the Connector upgrade steps.](#)

- Starting with BlueXP classification version 1.24 you can perform upgrades to any future version of software.

If your BlueXP classification software is running a version prior to 1.24, you can upgrade only one major version at a time. For example, if you have version 1.21.x installed, you can upgrade only to 1.22.x. If you are a few major versions behind, you'll need to upgrade the software multiple times.

Steps

1. On an internet-configured system, download the BlueXP classification software from the [NetApp Support Site](#). The file you should select is named **DataSense-offline-bundle-<version>.tar.gz**.
2. Copy the software bundle to the Linux host where BlueXP classification is installed in the dark site.
3. Unzip the software bundle on the host machine, for example:

```
tar -xvf DataSense-offline-bundle-v1.25.0.tar.gz
```

This extracts the installation file **cc_onprem_installer.tar.gz**.

4. Unzip the installation file on the host machine, for example:

```
tar -xzf cc_onprem_installer.tar.gz
```

This extracts the upgrade script **start_darksite_upgrade.sh** and any required third-party software.

5. Run the upgrade script on the host machine, for example:

```
start_darksite_upgrade.sh
```

Result

The BlueXP classification software is upgraded on your host. The update can take 5 to 10 minutes.

You can verify that the software has been updated by checking the version at the bottom of the BlueXP classification UI pages.

Check that your Linux host is ready to install BlueXP classification

Before installing BlueXP classification manually on a Linux host, you can run a script on the host to verify that all the prerequisites are in place for installing BlueXP 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*).

There is also a prerequisite test script that is part of the BlueXP classification installation script. The script described here is specifically designed for users who want to verify the Linux host independently of running the BlueXP classification installation script.

Getting Started

You'll perform the following tasks.

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

Create a Connector

A BlueXP Connector is required before you can install and use BlueXP classification. You can, however, run the Prerequisites script without a Connector.

You can [install the Connector on-premises](#) on a Linux host in your network or on a Linux host in the cloud. Some users planning to install BlueXP classification on-prem may also choose to install the Connector on-prem.

To create a Connector in your cloud provider environment, see [creating a Connector in AWS](#), [creating a Connector in Azure](#), or [creating a Connector in GCP](#).

You'll need the IP address or host name of the Connector system when running the Prerequisites script. You'll have this information if you installed the Connector in your premises. If the Connector is deployed in the cloud, you can find this information from the BlueXP console: click the Help icon, select **Support**, and click **BlueXP Connector**.

Verify host requirements

BlueXP classification software must run on a host that meets specific operating system requirements, RAM requirements, software requirements, and so on.

- BlueXP classification is not supported on a host that is shared with other applications - the host must be a dedicated host.
- When building the host system in your premises, you can choose among these system sizes depending on the size of the dataset that you plan to have BlueXP classification scan.

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

System size	CPU	RAM (swap memory must be disabled)	Disk
Large	16 CPUs	64 GB RAM	500 GiB SSD on /, or - 100 GiB available on /opt - 395 GiB available on /var/lib/docker or for Podman /var/lib/containers or for Podman /var/lib/containers - 5 GiB on /tmp

- When deploying a compute instance in the cloud for your BlueXP classification installation, we recommend a system that meets the "Large" system requirements above:
 - **Amazon Elastic Compute Cloud (Amazon EC2) instance type:** We recommend "m6i.4xlarge". [See additional AWS instance types.](#)
 - **Azure VM size:** We recommend "Standard_D16s_v3". [See additional Azure instance types.](#)
 - **GCP machine type:** We recommend "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 BlueXP classification version 1.23 or greater)
 - Ubuntu 24.04 (requires BlueXP classification version 1.23 or greater)
 - 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

Note that the following features are not currently supported when using RHEL 8.x and RHEL 9.x:

- Installation in a dark site
- Distributed scanning; using a master scanner node and remote scanner nodes

- **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 BlueXP classification:
 - Depending on the OS you are using, you'll need to install one of the container engines:

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

[Watch this video](#) for a quick demo of installing Docker on CentOS.

- 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 BlueXP classification system to use a Network Time Protocol (NTP) service. The time must be synchronized between the BlueXP classification system and the BlueXP Connector system.
- **Firewalld considerations:** If you are planning to use `firewalld`, we recommend that you enable it before installing BlueXP classification. Run the following commands to configure `firewalld` so that it is compatible with BlueXP 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 BlueXP 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 BlueXP classification

BlueXP classification requires outbound internet access. If your virtual or physical network uses a proxy server for internet access, ensure that the BlueXP 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
<code>https://api.bluexp.netapp.com</code>	Communication with the BlueXP service, which includes NetApp accounts.
<code>https://netapp-cloud-account.auth0.com</code> <code>https://auth0.com</code>	Communication with the BlueXP website for centralized user authentication.

Endpoints	Purpose
https://support.compliance.api.blueexp.netapp.com/ https://hub.docker.com https://auth.docker.io https://registry-1.docker.io https://index.docker.io/ https://dseasb33srrn.cloudfront.net/ https://production.cloudflare.docker.com/	Provides access to software images, manifests, templates, and to send logs and metrics.
https://support.compliance.api.blueexp.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://mirror.centos.org http://mirrorlist.centos.org http://mirror.centos.org/centos/7/extras/x86_64/Packages/container-selinux-2.107-3.el7.noarch.rpm	Provides prerequisite packages for CentOS 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 Connector, BlueXP classification, Active Directory, and your data sources.

Connection Type	Ports	Description
Connector <> BlueXP classification	8080 (TCP), 443 (TCP), and 80.9000	<p>The firewall or routing rules for the Connector must allow inbound and outbound traffic over port 443 to and from the BlueXP classification instance.</p> <p>Make sure port 8080 is open so you can see the installation progress in BlueXP.</p> <p>If a firewall is used on the Linux host, port 9000 is required for internal processes within an Ubuntu server.</p>
Connector <> ONTAP cluster (NAS)	443 (TCP)	BlueXP discovers ONTAP clusters using HTTPS. If you use custom firewall policies, the Connector host must allow outbound HTTPS access through port 443. If the Connector is in the cloud, all outbound communication is allowed by the predefined firewall or routing rules.

Run the BlueXP classification Prerequisites script

Follow these steps to run the BlueXP classification Prerequisites script.

Watch this [video](#) to see how to run the Prerequisites script and interpret the results.

What you'll need

- 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 BlueXP 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 BlueXP classification host machine.
 - Enter the IP address or host name.
6. The script prompts whether you have an installed BlueXP Connector.
 - Enter **N** if you do not have an installed Connector.
 - Enter **Y** if you do have an installed Connector. And then enter the IP address or host name of the BlueXP Connector 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 BlueXP 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 BlueXP 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 Azure NetApp Files volumes with BlueXP classification

Complete a few steps to get started with BlueXP classification for Azure NetApp Files.

Quick start

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

1

Discover the Azure NetApp Files systems you want to scan

Before you can scan Azure NetApp Files volumes, [BlueXP must be set up to discover the configuration](#).

2

Deploy the BlueXP classification instance

[Deploy BlueXP classification in BlueXP](#) if there isn't already an instance deployed.

3

Enable BlueXP classification and select the volumes to scan

Click **Compliance**, select the **Configuration** tab, and activate compliance scans for volumes in specific working environments.

4

Ensure access to volumes

Now that BlueXP classification is enabled, ensure that it can access all volumes.

- The BlueXP classification instance needs a network connection to each Azure NetApp Files subnet.
- Make sure these ports are open to the BlueXP classification instance:
 - For NFS – ports 111 and 2049.
 - For CIFS – ports 139 and 445.
- NFS volume export policies must allow access from the BlueXP classification instance.
- BlueXP classification needs Active Directory credentials to scan CIFS volumes.

Click **Compliance > Configuration > Edit CIFS Credentials** and provide the credentials.

5

Manage the volumes you want to scan

Select or deselect the volumes that you want to scan and BlueXP classification will start or stop scanning them.

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 BlueXP as a working environment, you can add it to the canvas at this time.

[See how to discover the Azure NetApp Files system in BlueXP](#).

Deploy the BlueXP classification instance

[Deploy BlueXP classification](#) if there isn't already an instance deployed.

BlueXP 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 BlueXP classification in an on-premises location is not currently supported when scanning Azure NetApp Files volumes.

Upgrades to BlueXP classification software is automated as long as the instance has internet connectivity.

Enable BlueXP classification in your working environments

You can enable BlueXP classification on your Azure NetApp Files volumes.

1. From the BlueXP left navigation menu, click **Governance > Classification** and then select the **Configuration** tab.



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

See [Enable and disable compliance scans on volumes](#) for details.

3. In the confirmation dialog box, click **Approve** to have BlueXP classification start scanning your volumes.

Result

BlueXP classification starts scanning the volumes you selected in the working environment. Results will be available in the Compliance dashboard as soon as BlueXP classification finishes the initial scans. The time that it takes depends on the amount of data—it could be a few minutes or hours.



- By default, if BlueXP 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 BlueXP classification can't revert the "last access time" to the original timestamp. If you don't care if the last access time is reset, click **Or select scanning type for each volume**. The resulting page has a setting you can enable so that BlueXP classification will scan the volumes regardless of permissions.
- BlueXP 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 BlueXP classification limitation.](#)

Verify that BlueXP classification has access to volumes

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

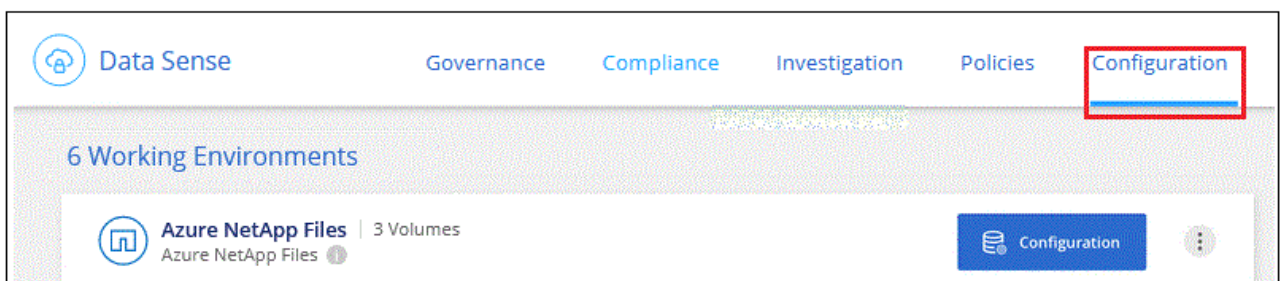
Steps

1. Make sure that there's a network connection between the BlueXP classification instance and each network that includes volumes for Azure NetApp Files.



For Azure NetApp Files, BlueXP classification can only scan volumes that are in the same region as BlueXP.

2. Ensure the following ports are open to the BlueXP classification instance:
 - For NFS – ports 111 and 2049.
 - For CIFS – ports 139 and 445.
3. Ensure that NFS volume export policies include the IP address of the BlueXP classification instance so it can access the data on each volume.
4. If you use CIFS, provide BlueXP classification with Active Directory credentials so it can scan CIFS volumes.
 - a. From the BlueXP left navigation menu, click **Governance > Classification** and then select the **Configuration** tab.



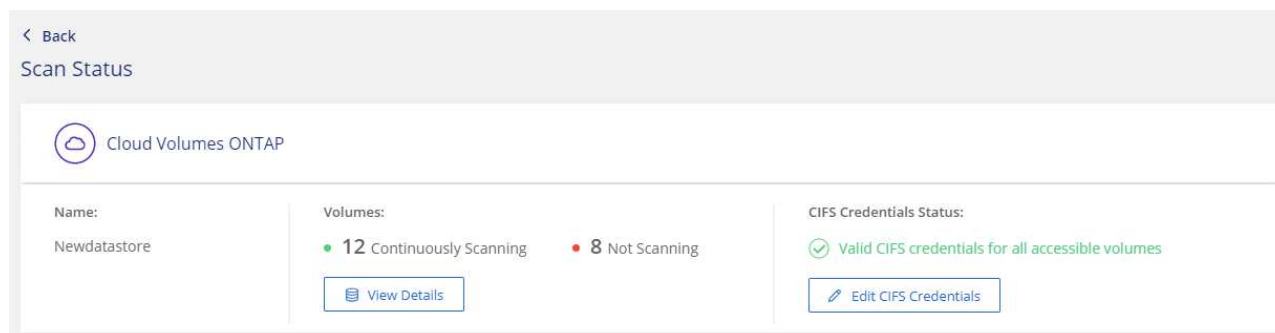
- b. For each working environment, click **Edit CIFS Credentials** and enter the user name and password that BlueXP classification needs to access CIFS volumes on the system.

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

If you want to make sure your files "last accessed times" are unchanged by BlueXP classification

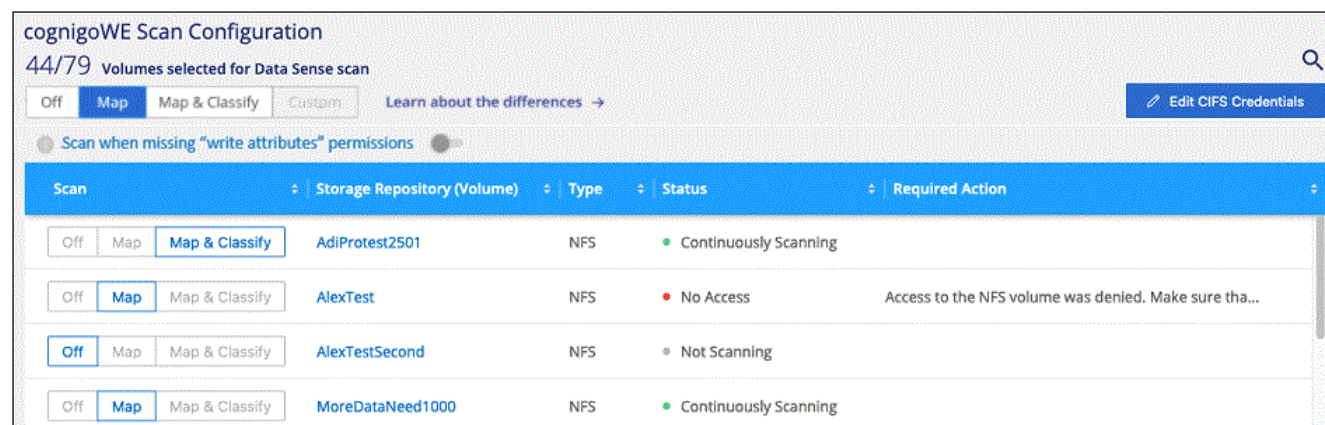
scans, we recommend that the user has Write Attributes permissions in CIFS or write permissions in NFS. If possible, we recommend making the Active Directory configured user part of a parent group in the organization which has permissions to all files.

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



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

For example, the following image shows four volumes; one of which BlueXP classification can't scan due to network connectivity issues between the BlueXP classification instance and the volume.



Enable and disable compliance scans on volumes

You can start or stop mapping-only scans, or mapping and classification scans, in a working environment at any time from the Configuration page. You can also change from mapping-only scans to mapping and classification scans, and vice-versa. We recommend that you scan all volumes.

The switch at the top of the page for **Scan when missing "write attributes" permissions** is disabled by default. This means that if BlueXP classification doesn't have write attributes permissions in CIFS, or write permissions in NFS, that the system won't scan the files because BlueXP 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](#).

cognitoWE Scan Configuration
44/79 Volumes selected for Data Sense scan

Off
Map
Map & Classify
Custom
Learn about the differences →
Edit CIFS Credentials

☐ Scan when missing "write attributes" permissions

Scan	Storage Repository (Volume)	Type	Status	Required Action
Off Map Map & Classify	AdiNFSVol_copy	NFS	No Access	Access to the NFS volume was denied. Make sure tha...
Off Map Map & Classify	AdiProtest2501	NFS	Continuously Scanning	
Off Map Map & Classify	AlexTest	NFS	No Access	Access to the NFS volume was denied. Make sure tha...
Off Map Map & Classify	AlexTestSecond	NFS	Not Scanning	

To:	Do this:
Enable mapping-only scans on a volume	In the volume area, click Map
Enable full scanning on a volume	In the volume area, click Map & Classify
Disable scanning on a volume	In the volume area, click Off
Enable mapping-only scans on all volumes	In the heading area, click Map
Enable full scanning on all volumes	In the heading area, click Map & Classify
Disable scanning on all volumes	In the heading area, click Off



New volumes added to the working environment are automatically scanned only when you have set 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 working environment.

Scan Amazon FSx for ONTAP volumes with BlueXP classification

Complete a few steps to get started scanning Amazon FSx for ONTAP volume with BlueXP classification.

Before you begin

- You need an active Connector in AWS to deploy and manage BlueXP classification.
- The security group you selected when creating the working environment must allow traffic from the BlueXP 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 \(ENI\)](#)

Quick start

Get started quickly by following these steps or scroll down for full details.

1

Discover the FSx for ONTAP file systems you want to scan

Before you can scan FSx for ONTAP volumes, [you must have an FSx working environment with volumes configured](#).

2

Deploy the BlueXP classification instance

[Deploy BlueXP classification in BlueXP](#) if there isn't already an instance deployed.

3

Enable BlueXP classification and select the volumes to scan

Select the **Configuration** tab and activate compliance scans for volumes in specific working environments.

4

Ensure access to volumes

Now that BlueXP classification is enabled, ensure that it can access all volumes.

- The BlueXP classification instance needs a network connection to each FSx for ONTAP subnet.
- Make sure the following ports are open to the BlueXP classification instance:
 - For NFS – ports 111 and 2049.
 - For CIFS – ports 139 and 445.
- NFS volume export policies must allow access from the BlueXP classification instance.
- BlueXP classification needs Active Directory credentials to scan CIFS volumes.

Click **Compliance > Configuration > Edit CIFS Credentials** and provide the credentials.

5

Manage the volumes you want to scan

Select or deselect the volumes you want to scan and BlueXP classification will start or stop scanning them.

Discover the FSx for ONTAP file system that you want to scan

If the FSx for ONTAP file system you want to scan is not already in BlueXP as a working environment, you can add it to the canvas at this time.

[See how to discover or create the FSx for ONTAP file system in BlueXP](#).

Deploy the BlueXP classification instance

[Deploy BlueXP classification](#) if there isn't already an instance deployed.

You should deploy BlueXP classification in the same AWS network as the Connector for AWS and the FSx volumes you wish to scan.

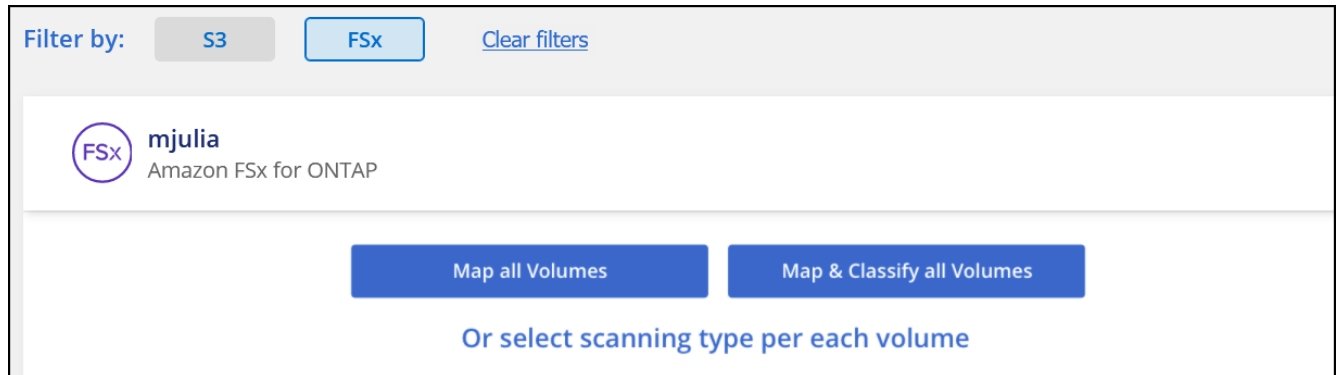
Note: Deploying BlueXP classification in an on-premises location is not currently supported when scanning FSx volumes.

Upgrades to BlueXP classification software is automated as long as the instance has internet connectivity.

Enable BlueXP classification in your working environments

You can enable BlueXP classification for FSx for ONTAP volumes.

1. From the BlueXP left navigation menu, click **Governance > Classification** and then select the **Configuration** tab.



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

See [Enabling and disabling compliance scans on volumes](#) for details.

3. In the confirmation dialog box, click **Approve** to have BlueXP classification start scanning your volumes.

Result

BlueXP classification starts scanning the volumes you selected in the working environment. Results will be available in the Compliance dashboard as soon as BlueXP classification finishes the initial scans. The time that it takes depends on the amount of data—it could be a few minutes or hours.



- By default, if BlueXP 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 BlueXP classification can't revert the "last access time" to the original timestamp. If you don't care if the last access time is reset, click **Or select scanning type for each volume**. The resulting page has a setting you can enable so that BlueXP classification will scan the volumes regardless of permissions.
- BlueXP 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 BlueXP classification limitation](#).

Verify that BlueXP classification has access to volumes

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

You'll need to provide BlueXP classification with CIFS credentials so it can access CIFS volumes.

Steps

1. On the *Configuration* page, click **View Details** to review the status and correct any errors.

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

Scan	Storage Repository (Volume)	Type	Status	Required Action
<input type="button" value="Off"/> <input type="button" value="Map"/> <input type="button" value="Map & Classify"/>	jrmclone	NFS	● No Access	Check network connectivity between the Data Sense ...

2. Make sure there's a network connection between the BlueXP classification instance and each network that includes volumes for FSx for ONTAP.



For FSx for ONTAP, BlueXP classification can scan volumes only in the same region as BlueXP.

3. Ensure the following ports are open to the BlueXP classification instance.
 - For NFS – ports 111 and 2049.
 - For CIFS – ports 139 and 445.
4. Ensure NFS volume export policies include the IP address of the BlueXP classification instance so it can access the data on each volume.
5. If you use CIFS, provide BlueXP classification with Active Directory credentials so it can scan CIFS volumes.
 - a. From the BlueXP left navigation menu, click **Governance > Classification** and then select the **Configuration** tab.
 - b. For each working environment, click **Edit CIFS Credentials** and enter the user name and password that BlueXP classification needs to access CIFS volumes on the system.

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

If you want to make sure your files "last accessed times" are unchanged by BlueXP classification scans, we recommend that the user has Write Attributes permissions in CIFS or write permissions in NFS. If possible, we recommend making the Active Directory configured user 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 compliance scans on volumes

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

classification scans, and vice-versa. We recommend that you scan all volumes.

The switch at the top of the page for **Scan when missing "write attributes" permissions** is disabled by default. This means that if BlueXP classification doesn't have write attributes permissions in CIFS, or write permissions in NFS, that the system won't scan the files because BlueXP 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](#).

cognigoWE Scan Configuration

44/79 Volumes selected for Data Sense scan

OffMapMap & ClassifyCustom

Learn about the differences →

Edit CIFS Credentials

Scan when missing "write attributes" permissions

Scan	Storage Repository (Volume)	Type	Status	Required Action
<div>OffMapMap & Classify</div>	AdiNFSVol_copy	NFS	No Access	Access to the NFS volume was denied. Make sure tha...
<div>OffMapMap & Classify</div>	AdiProtest2501	NFS	Continuously Scanning	
<div>OffMapMap & Classify</div>	AlexTest	NFS	No Access	Access to the NFS volume was denied. Make sure tha...
<div>OffMapMap & Classify</div>	AlexTestSecond	NFS	Not Scanning	

To:	Do this:
Enable mapping-only scans on a volume	In the volume area, click Map
Enable full scanning on a volume	In the volume area, click Map & Classify
Disable scanning on a volume	In the volume area, click Off
Enable mapping-only scans on all volumes	In the heading area, click Map
Enable full scanning on all volumes	In the heading area, click Map & Classify
Disable scanning on all volumes	In the heading area, click Off



New volumes added to the working environment are automatically scanned only when you have set 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 working environment.

Scan data protection volumes

By default, data protection (DP) volumes are not scanned because they are not exposed externally and BlueXP 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*.

'Working Environment Name' Configuration

22/28 Volumes selected for compliance scan

Enable Access to DP Volumes [Edit CIFS Credentials](#)

Off **Map** Map & Classify Custom [Learn about the differences](#) →

Scan when missing "write attributes" permissions ☐

Scan	Storage Repository (Volume)	Type	Status	Required Action
Off Map Map & Classify	VolumeName1	DP	Not Scanning	Enable access to DP Volumes ⓘ
Off Map Map & Classify	VolumeName2	NFS	Continuously Scanning	
Off Map Map & Classify	VolumeName3	CIFS	Not Scanning	

Steps

If you want to scan these data protection volumes:

1. Click **Enable Access to DP volumes** at the top of the page.
2. Review the confirmation message and click **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 BlueXP classification can scan CIFS volumes you can use those credentials, or you can specify a different set of Admin credentials.

Provide Active Directory Credentials

☒ Use existing CIFS Scanning Credentials (user1@domain2) ☐ Use Custom Credentials

Active Directory Domain ⓘ DNS IP Address ⓘ

DP Volumes, created from a SnapMirror relationship, do not allow external access by default. Continuing will create NFS shares from DP Volumes which have been activated for **Data Sense**. The shares' export policies will allow access only from the Cloud **Data Sense** instance. [Learn More](#)

Enable Access to DP Volumes Cancel

Provide Active Directory Credentials

☐ Use existing CIFS Scanning Credentials (user1@domain2) ☒ Use Custom Credentials

Username ⓘ Password

Active Directory Domain ⓘ DNS IP Address ⓘ

DP Volumes, created from a SnapMirror relationship, do not allow external access by default. Continuing will create NFS shares from DP Volumes which have been activated for **Data Sense**. The shares' export policies will allow access only from the Cloud **Data Sense** instance. [Learn More](#)

Enable Access to DP Volumes Cancel

3. Activate each DP volume that you want to scan [the same way you enabled other volumes](#).

Result

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

Note: 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. Click this button and add CIFS credentials to enable access to these CIFS DP volumes.



Active Directory credentials are only registered 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 Cloud Volumes ONTAP and on-premises ONTAP volumes with BlueXP classification

Complete a few steps to start scanning your Cloud Volumes ONTAP and on-premises ONTAP volumes using BlueXP classification.

Quick start

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

1

Discover the data sources that you want to scan

Before you can scan volumes, you must add the systems as working environments in BlueXP:

- For Cloud Volumes ONTAP systems, these working environments should already be available in BlueXP
- For on-premises ONTAP systems, [BlueXP must discover the ONTAP clusters](#)

2

Deploy the BlueXP classification instance

[Deploy BlueXP classification](#) if there isn't already an instance deployed.

3

Enable BlueXP classification and select the volumes to scan

Select the **Configuration** tab and activate compliance scans for volumes in specific working environments.

4

Ensure access to volumes

Now that BlueXP classification is enabled, ensure that it can access all volumes.

- The BlueXP classification instance needs a network connection to each Cloud Volumes ONTAP subnet or on-prem ONTAP system.
- Security groups for Cloud Volumes ONTAP must allow inbound connections from the BlueXP classification instance.
- Make sure these ports are open to the BlueXP classification instance:
 - For NFS - ports 111 and 2049.
 - For CIFS - ports 139 and 445.
- NFS volume export policies must allow access from the BlueXP classification instance.
- BlueXP classification needs Active Directory credentials to scan CIFS volumes.

Click **Compliance > Configuration > Edit CIFS Credentials** and provide the credentials.

5

Manage the volumes you want to scan

Select or deselect the volumes that you want to scan and BlueXP classification will start or stop scanning them.

Discover the data sources that you want to scan

If the data sources you want to scan are not already in your BlueXP environment, you can add them to the canvas at this time.

Your Cloud Volumes ONTAP systems should already be available in the Canvas in BlueXP. For on-premises ONTAP systems, you'll need to have [BlueXP discover these clusters](#).

Deploy the BlueXP classification instance

Deploy BlueXP classification if there isn't already an instance deployed.

If you are scanning Cloud Volumes ONTAP and on-premises ONTAP systems that are accessible over the internet, you can [deploy BlueXP classification in the cloud](#) or [in an on-premises location that has internet access](#).

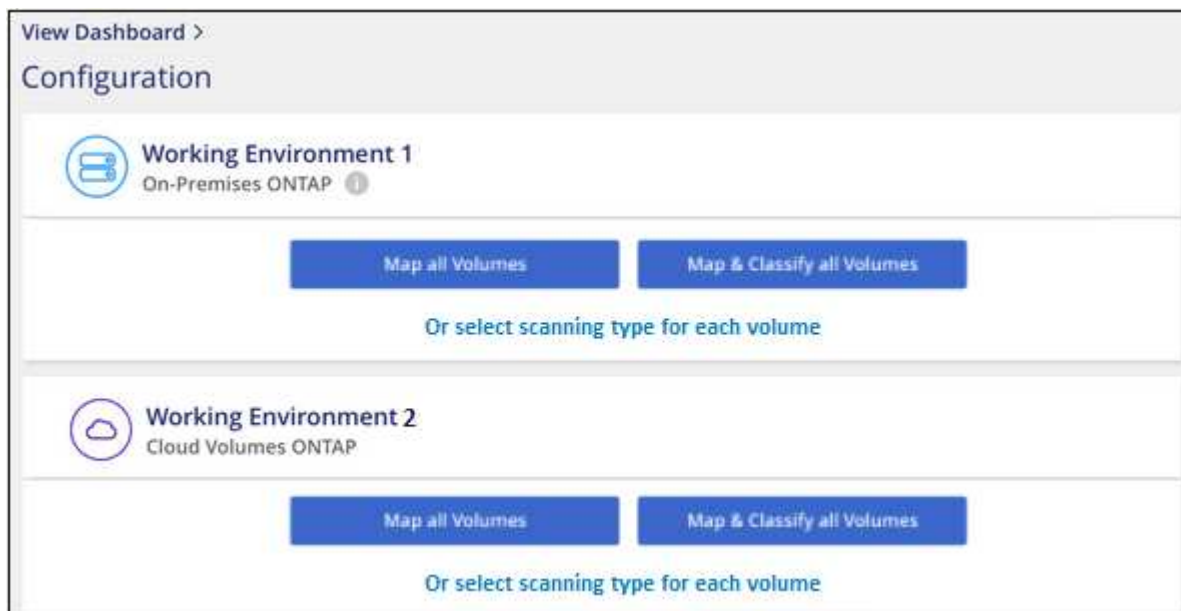
If you are scanning on-premises ONTAP systems that have been installed in a dark site that has no internet access, you need to [deploy BlueXP classification in the same on-premises location that has no internet access](#). This also requires that the BlueXP Connector is deployed in that same on-premises location.

Upgrades to BlueXP classification software is automated as long as the instance has internet connectivity.

Enable BlueXP classification in your working environments

You can enable BlueXP classification on Cloud Volumes ONTAP systems in any supported cloud provider, and on on-premises ONTAP clusters.

1. From the BlueXP left navigation menu, click **Governance > Classification** and then select the **Configuration** tab.



2. Select how you want to scan the volumes in each working environment. [Learn about mapping and classification scans](#):
 - To map all volumes, click **Map all Volumes**.
 - To map and classify all volumes, click **Map & Classify all Volumes**.

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

See [Enable and disable compliance scans on volumes](#) for details.

3. In the confirmation dialog box, click **Approve** to have BlueXP classification start scanning your volumes.

Result

BlueXP classification starts scanning the volumes you selected in the working environment. Results will be available in the Compliance dashboard as soon as BlueXP classification finishes the initial scans. The time that it takes depends on the amount of data—it could be a few minutes or hours.



- By default, if BlueXP 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 BlueXP classification can't revert the "last access time" to the original timestamp. If you don't care if the last access time is reset, click **Or select scanning type for each volume**. The resulting page has a setting you can enable so that BlueXP classification will scan the volumes regardless of permissions.
- BlueXP 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 BlueXP classification limitation.](#)

Verify that BlueXP classification has access to volumes

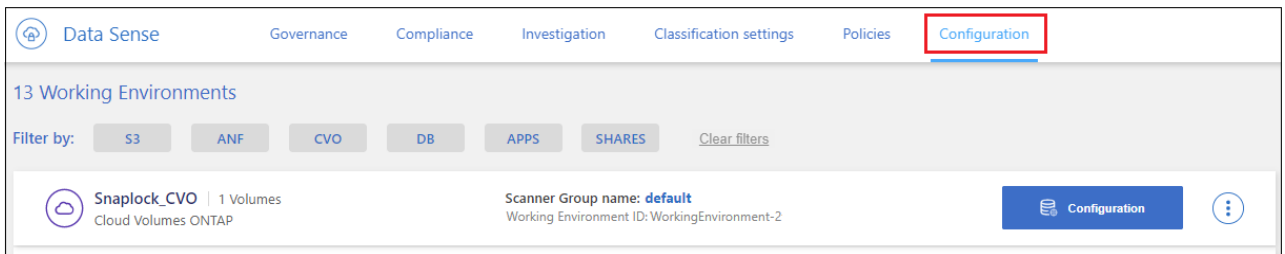
Make sure that BlueXP classification can access volumes by checking your networking, security groups, and export policies. You'll need to provide BlueXP classification with CIFS credentials so it can access CIFS volumes.

Steps

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

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

3. Ensure the following ports are open to the BlueXP classification instance:
 - For NFS - ports 111 and 2049.
 - For CIFS - ports 139 and 445.
4. Ensure that NFS volume export policies include the IP address of the BlueXP classification instance so it can access the data on each volume.
5. If you use CIFS, provide BlueXP classification with Active Directory credentials so it can scan CIFS volumes.
 - a. From the BlueXP left navigation menu, click **Governance > Classification** and then select the **Configuration** tab.

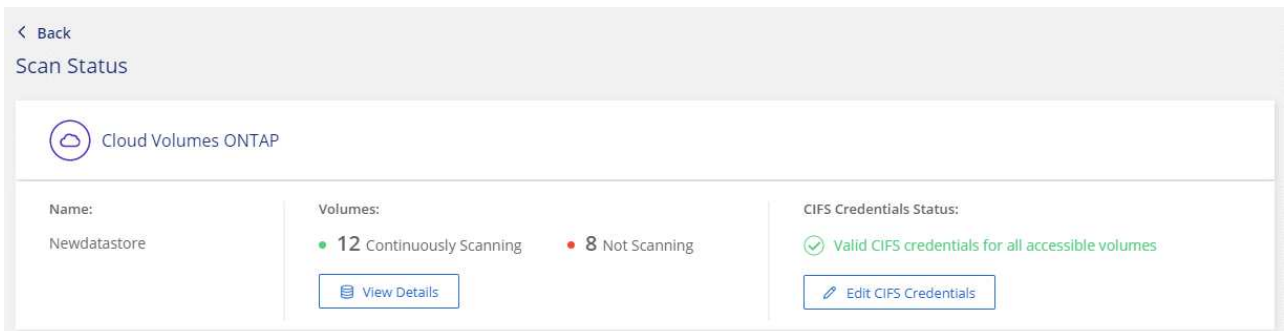


- b. For each working environment, click **Edit CIFS Credentials** and enter the user name and password that BlueXP classification needs to access CIFS volumes on the system.

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

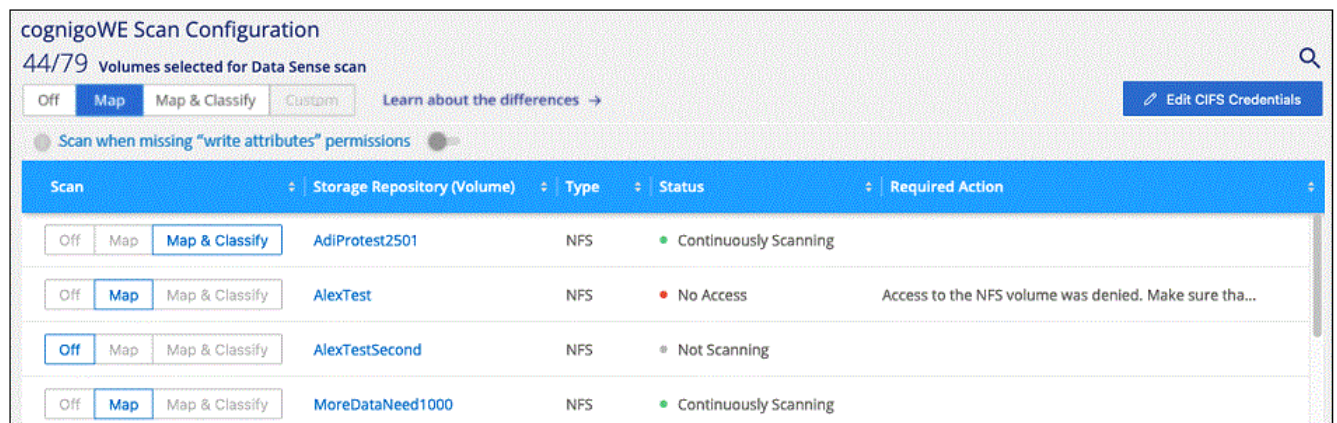
If you want to make sure your files "last accessed times" are unchanged by BlueXP classification scans, we recommend that the user has Write Attributes permissions in CIFS or write permissions in NFS. If possible, we recommend making the Active Directory configured user part of a parent group in the organization which has permissions to all files.

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



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

For example, the following image shows four volumes; one of which BlueXP classification can't scan due to network connectivity issues between the BlueXP classification instance and the volume.



Enable and disable compliance scans on volumes

You can start or stop mapping-only scans, or mapping and classification scans, in a working environment at any time from the Configuration page. You can also change from mapping-only scans to mapping and classification scans, and vice-versa. We recommend that you scan all volumes.

The switch at the top of the page for **Scan when missing "write attributes" permissions** is disabled by default. This means that if BlueXP classification doesn't have write attributes permissions in CIFS, or write permissions in NFS, that the system won't scan the files because BlueXP 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](#).

cognigoWE Scan Configuration

44/79 Volumes selected for Data Sense scan

OffMapMap & ClassifyCustom

Learn about the differences →

Edit CIFS Credentials

☐ Scan when missing "write attributes" permissions

Scan	Storage Repository (Volume)	Type	Status	Required Action
<div>OffMapMap & Classify</div>	AdiNFSVol_copy	NFS	No Access	Access to the NFS volume was denied. Make sure tha...
<div>OffMapMap & Classify</div>	AdiProtest2501	NFS	Continuously Scanning	
<div>OffMapMap & Classify</div>	AlexTest	NFS	No Access	Access to the NFS volume was denied. Make sure tha...
<div>OffMapMap & Classify</div>	AlexTestSecond	NFS	Not Scanning	

To:	Do this:
Enable mapping-only scans on a volume	In the volume area, click Map
Enable full scanning on a volume	In the volume area, click Map & Classify
Disable scanning on a volume	In the volume area, click Off
Enable mapping-only scans on all volumes	In the heading area, click Map
Enable full scanning on all volumes	In the heading area, click Map & Classify
Disable scanning on all volumes	In the heading area, click Off



New volumes added to the working environment are automatically scanned only when you have set 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 working environment.

Scan data protection volumes

By default, data protection (DP) volumes are not scanned because they are not exposed externally and BlueXP classification cannot access them. These are the destination volumes for SnapMirror operations from an on-premises ONTAP system or from a Cloud Volumes ONTAP 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*.

'Working Environment Name' Configuration 22/28 Volumes selected for compliance scan

Enable Access to DP Volumes
Edit CIFS Credentials

[Learn about the differences →](#)

☒ Scan when missing "write attributes" permissions

Scan	Storage Repository (Volume)	Type	Status	Required Action
<input type="button" value="Off"/> <input type="button" value="Map"/> <input type="button" value="Map & Classify"/>	VolumeName1	DP	Not Scanning	Enable access to DP Volumes ⓘ
<input type="button" value="Off"/> <input checked="" type="button" value="Map"/> <input type="button" value="Map & Classify"/>	VolumeName2	NFS	Continuously Scanning	
<input type="button" value="Off"/> <input type="button" value="Map"/> <input type="button" value="Map & Classify"/>	VolumeName3	CIFS	Not Scanning	

Steps

If you want to scan these data protection volumes:

1. Click **Enable Access to DP volumes** at the top of the page.
2. Review the confirmation message and click **Enable Access to DP volumes** again.
 - Volumes that were initially created as NFS volumes in the source ONTAP system are enabled.
 - Volumes that were initially created as CIFS volumes in the source ONTAP system require that you enter CIFS credentials to scan those DP volumes. If you already entered Active Directory credentials so that BlueXP classification can scan CIFS volumes you can use those credentials, or you can specify a different set of Admin credentials.

Provide Active Directory Credentials

☒ Use existing CIFS Scanning Credentials (user1@domain2)
 ☐ Use Custom Credentials

Active Directory Domain ⓘ
 DNS IP Address ⓘ

DP Volumes, created from a SnapMirror relationship, do not allow external access by default. Continuing will create NFS shares from DP Volumes which have been activated for **Data Sense**. The shares' export policies will allow access only from the Cloud **Data Sense** instance. [Learn More](#)

Provide Active Directory Credentials

☐ Use existing CIFS Scanning Credentials (user1@domain2)
 ☒ Use Custom Credentials

Username ⓘ
 Password

Active Directory Domain ⓘ
 DNS IP Address ⓘ

DP Volumes, created from a SnapMirror relationship, do not allow external access by default. Continuing will create NFS shares from DP Volumes which have been activated for **Data Sense**. The shares' export policies will allow access only from the Cloud **Data Sense** instance. [Learn More](#)

3. Activate each DP volume that you want to scan [the same way you enabled other volumes](#).

Result

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

Note: 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. Click this button and add CIFS credentials to enable access to these CIFS DP volumes.



Active Directory credentials are only registered 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 database schemas with BlueXP classification

Complete a few steps to start scanning your database schemas with BlueXP classification.

Quick start

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

1

Review database prerequisites

Ensure that your database is supported and that you have the information necessary to connect to the database.

2

Deploy the BlueXP classification instance

[Deploy BlueXP classification](#) if there isn't already an instance deployed.

3

Add the database server

Add the database server that you want to access.

4

Select the schemas

Select the schemas that you want to scan.

Review prerequisites

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

Supported databases

BlueXP 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 BlueXP 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 BlueXP classification system with all the required permissions.

Note: For MongoDB, a read-only Admin role is required.

Deploy the BlueXP classification instance

Deploy BlueXP classification if there isn't already an instance deployed.

If you are scanning database schemas that are accessible over the internet, you can [deploy BlueXP classification in the cloud](#) or [deploy BlueXP classification in an on-premises location that has internet access](#).

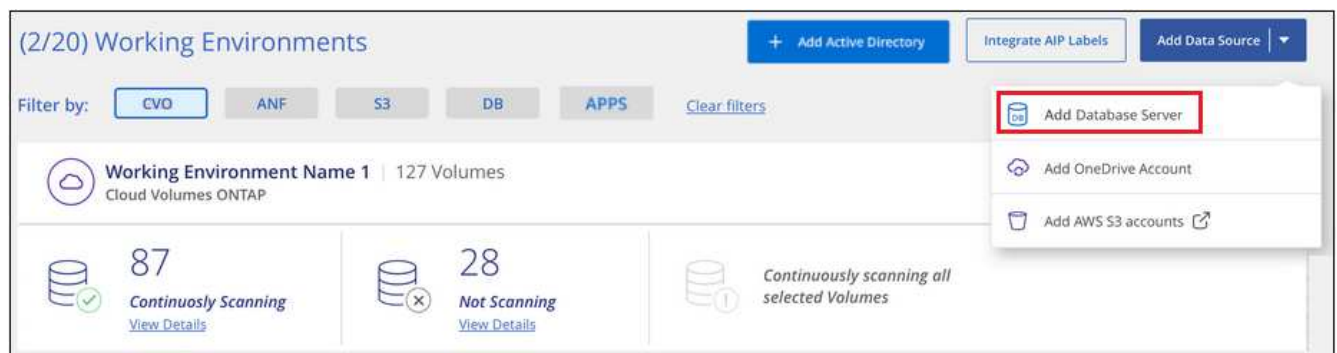
If you are scanning database schemas that have been installed in a dark site that has no internet access, you need to [deploy BlueXP classification in the same on-premises location that has no internet access](#). This also requires that the BlueXP Connector is deployed in that same on-premises location.

Upgrades to BlueXP classification software is automated as long as the instance has internet connectivity.

Add the database server

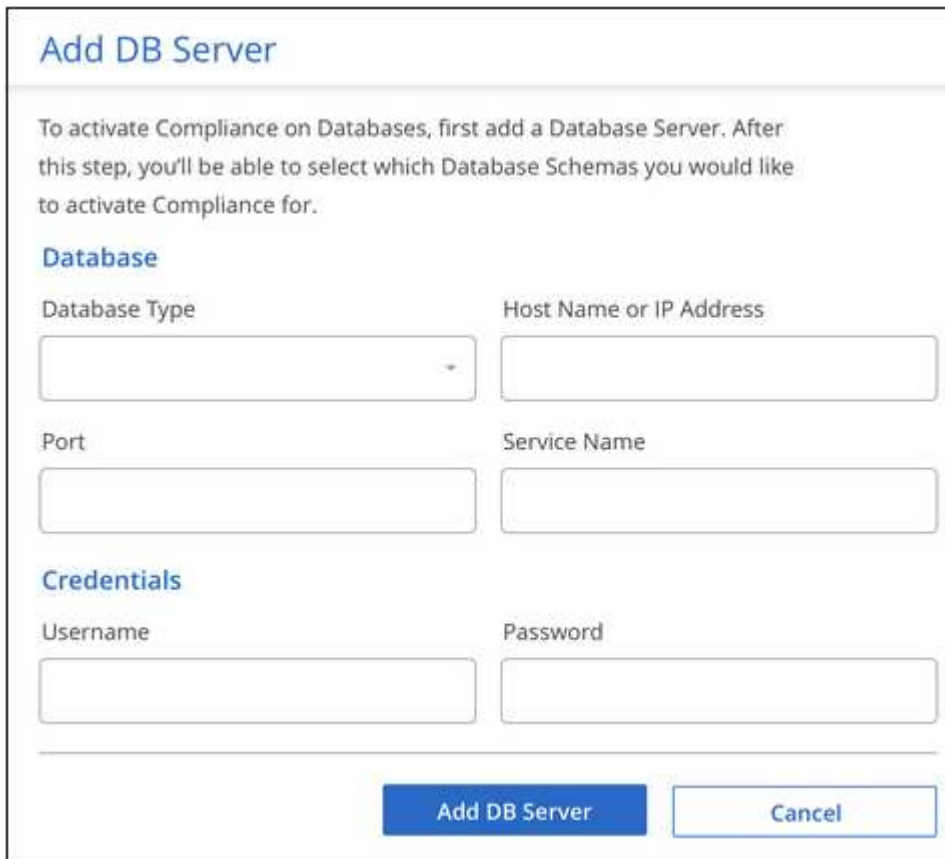
Add the database server where the schemas reside.

1. From the Working Environments Configuration page, click **Add Data Source > Add Database Server**.



2. 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 BlueXP classification can access the server.

e. Click **Add DB Server**.



The 'Add DB Server' dialog box has a title bar with the text 'Add DB Server'. Below the title bar is a paragraph: 'To activate Compliance on Databases, first add a Database Server. After this step, you'll be able to select which Database Schemas you would like to activate Compliance for.' The dialog is divided into two sections: 'Database' and 'Credentials'. The 'Database' section contains four input fields: 'Database Type' (a dropdown menu), 'Host Name or IP Address' (a text box), 'Port' (a text box), and 'Service Name' (a text box). The 'Credentials' section contains two input fields: 'Username' and 'Password'. At the bottom right of the dialog are two buttons: 'Add DB Server' (a blue button) and 'Cancel' (a white button with a blue border).

The database is added to the list of working environments.

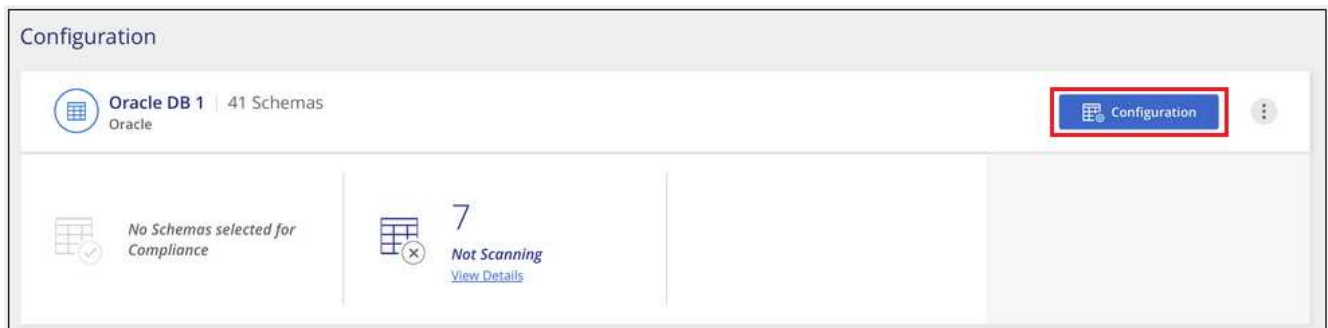
Enable and disable compliance scans on database schemas

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, click the **Configuration** button for the database you want to configure.



The 'Configuration' page shows a header with the title 'Configuration'. Below the header is a section for 'Oracle DB 1 | 41 Schemas'. In the top right corner of this section is a blue button labeled 'Configuration' with a gear icon, which is highlighted with a red rectangle. Below this section are two cards. The first card is titled 'No Schemas selected for Compliance' and has a checkmark icon. The second card is titled '7 Not Scanning' with a red 'x' icon and a 'View Details' link.

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

'Working Environment Name' Configuration			
28/28 Schemas selected for compliance scan		Edit Credentials	
Scan	Schema Name	Status	Required Action
<input type="checkbox"/>	DB1 - SchemaName1	Not Scanning	Add Credentials ⓘ
<input type="checkbox"/>	DB1 - SchemaName2	Continuously Scanning	
<input type="checkbox"/>	DB1 - SchemaName3	Continuously Scanning	
<input type="checkbox"/>	DB1 - SchemaName4	Continuously Scanning	

Result

BlueXP classification starts scanning the database schemas that you enabled. If there are any errors, they'll appear in the Status column, alongside the required action to fix the error.

Note that BlueXP classification scans your databases once per day - databases are not continuously scanned like other data sources.

Scan file shares with BlueXP classification

Complete a few steps to start scanning NFS or CIFS file shares from Google Cloud NetApp Volumes and from older NetApp 7-mode systems. These file shares can reside on-premises or in the cloud.



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

Quick start

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

1

Review file share prerequisites

For CIFS (SMB) shares, ensure that you have credentials to access the shares.

2

Deploy the BlueXP classification instance

[Deploy BlueXP classification](#) if there isn't already an instance deployed.

3

Create a group to hold the file shares

The group is a container for the file shares that you want to scan, and it is used as the working environment name for those file shares.

4

Add the file shares to the group

Add the list of file shares that you want to scan and select the type of scanning. You can add up to 100 file shares at a time.

Review file share requirements

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

Note that BlueXP classification can't extract permissions or the "last access time" from 7-Mode systems. Additionally, because of a known issue between some Linux versions and CIFS shares on 7-Mode systems, you must configure the share to use only SMB v1 with NTLM authentication enabled.

- There needs to be network connectivity between the BlueXP classification instance and the shares.
- Make sure these ports are open to the BlueXP classification instance:
 - For NFS – ports 111 and 2049.
 - For CIFS – ports 139 and 445.
- You can add a DFS (Distributed File System) share as a regular CIFS share. However, because BlueXP classification is not aware 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 BlueXP classification needs to scan any data that requires elevated permissions.

If you want to make sure your files "last accessed times" are unchanged by BlueXP classification scans, we recommend that the user has Write Attributes permissions in CIFS or write permissions in NFS. If possible, we recommend making the Active Directory configured user part of a parent group in the organization which has permissions to all files.

- You will need the list of shares you want to add in the format `<host_name>:/<share_path>`. You can enter the shares individually, or you can supply a line-separated list of the file shares you want to scan.

Deploy the BlueXP classification instance

Deploy BlueXP classification if there isn't already an instance deployed.

Upgrades to BlueXP classification software are automated as long as the instance has internet connectivity.

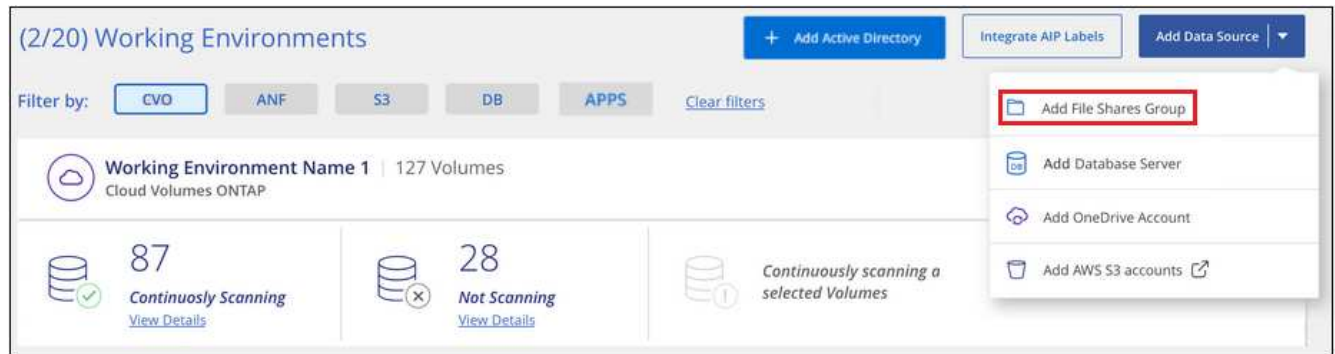
Create the group for the file shares

You must add a file shares "group" before you can add your file shares. The group is a container for the file shares that you want to scan, and the group name is used as the working environment name for those file shares.

You can mix NFS and CIFS shares in the same group, however, all CIFS file shares in a group need to be using the same Active Directory credentials. If you plan to add CIFS shares that use different credentials, you must make a separate group for each unique set of credentials.

Steps

1. From the Working Environments Configuration page, click **Add Data Source > Add File Shares Group**.



2. In the Add Files Shares Group dialog, enter the name for the group of shares and click **Continue**.

The new File Shares Group is added to the list of working environments.

Add file shares to a group

You add file shares to the File Shares Group so that the files in those shares will be scanned by BlueXP classification. You add the shares in the format `<host_name>:/<share_path>`.

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

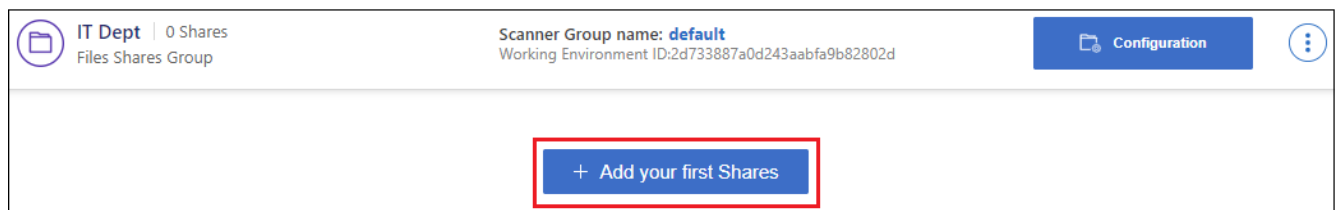
When adding both NFS and CIFS shares in a single group, you'll need to run through the process twice - once adding NFS shares, and then again adding the CIFS shares.

Steps

1. From the *Working Environments* page, click the **Configuration** button for the File Shares Group.



2. If this is the first time adding file shares for this File Shares Group, click **Add your first Shares**.



If you are adding file shares to an existing group, click **Add Shares**.

SMB_Shares Scan Configuration

2 Shares selected for Data Sense scan

Scan when missing "write" permissions ☐

+ Add Shares Edit CIFS Credentials

Scan	Storage Repository (Share)	Protocol	Access	Scan Status	Required Action
Map Map & Classify	.E6	CIFS	Continuously Scanning	<input checked="" type="radio"/> Mapped: 5.8K <input type="radio"/> Classified: 5.8K	...
Map Map & Classify	7	CIFS	Continuously Scanning	<input checked="" type="radio"/> Mapped: 5.8K <input type="radio"/> Classified: 5.8K	...

3. Select the protocol for the file shares you are adding, add the file shares that you want to scan - one file share per line - and click **Continue**.

When adding CIFS (SMB) shares, you need to enter the Active Directory credentials that provide read access to the shares. Admin credentials are preferred.

Adding Shares

Directly add any NFS or CIFS (SMB) File Shares, located in the cloud or on-premises.

Select Protocol

You'll be able to add additional shares from the other protocol later.

☒ NFS ☐ CIFS (SMB)

Type or paste below the Shares to add

Provide a list of shares, line-separated. You can add up to 100 at a time (you can add more later).

Hostname/SHAREPATH
Hostname/SHAREPATH
Hostname/SHAREPATH

Continue Cancel

☐ NFS ☒ CIFS (SMB)

Provide CIFS Credentials

Username Password

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. In some cases you can re-add the share with a corrected host name or share name.

4. Enable mapping-only scans, or mapping and classification scans, on each file share.

To:	Do this:
Enable mapping-only scans on file shares	Click Map
Enable full scans on file shares	Click Map & Classify
Disable scanning on file shares	Click Off

The switch at the top of the page for **Scan when missing "write attributes" permissions** is disabled by default. This means that if BlueXP classification doesn't have write attributes permissions in CIFS, or write

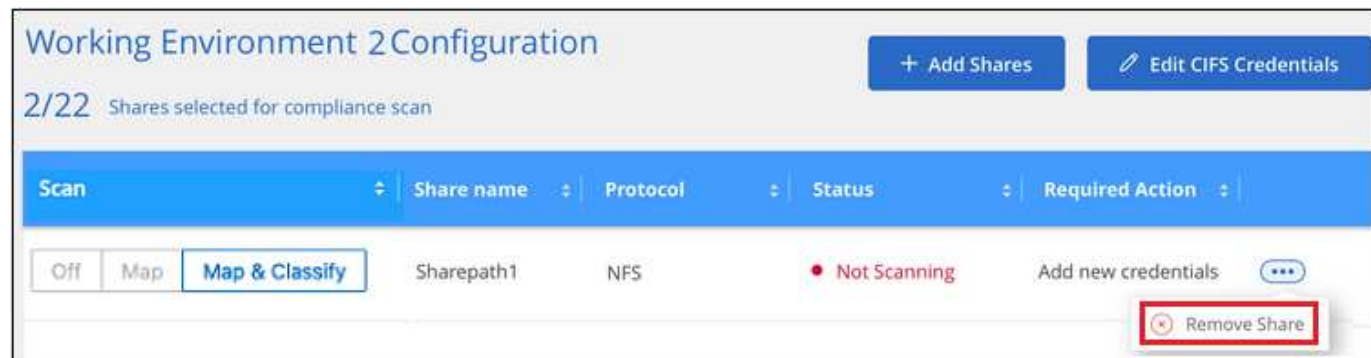
permissions in NFS, that the system won't scan the files because BlueXP classification can't revert the "last access time" to the original timestamp. If you don't care if the last access time is reset, turn the switch ON and all files are scanned regardless of the permissions. [Learn more](#).

Result

BlueXP classification starts scanning the files in the file shares you added, and the results are displayed in the Dashboard and in other locations.

Remove a file share from compliance scans

If you no longer need to scan certain file shares, you can remove individual file shares from having their files scanned at any time. Just click **Remove Share** from the Configuration page.



Scan StorageGRID data with BlueXP classification

Complete a few steps to start scanning data within StorageGRID directly with BlueXP classification.

Quick start

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

1

Review StorageGRID prerequisites

You need to have the endpoint URL to connect with the StorageGRID service.

You need to have the Access Key and Secret Key from StorageGRID so that BlueXP classification can access the buckets.

2

Deploy the BlueXP classification instance

[Deploy BlueXP classification](#) if there isn't already an instance deployed.

3

Add the StorageGRID Service

Add the StorageGRID service to BlueXP classification.

4

Select the buckets to scan

Select the buckets that you'd like to scan and BlueXP classification will start scanning them.

Review StorageGRID requirements

Review the following prerequisites to make sure that you have a supported configuration before you enable BlueXP 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 BlueXP classification can access the buckets.

Deploy the BlueXP classification instance

Deploy BlueXP 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 BlueXP classification in the cloud](#) or [deploy BlueXP 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 BlueXP classification in the same on-premises location that has no internet access](#). This also requires that the BlueXP Connector is deployed in that same on-premises location.

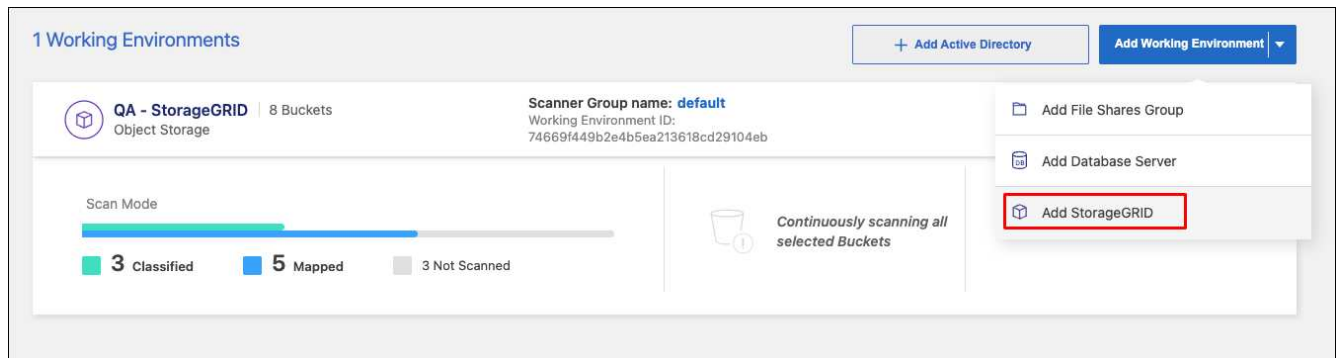
Upgrades to BlueXP classification software are automated as long as the instance has internet connectivity.

Add the StorageGRID service to BlueXP classification

Add the StorageGRID service.

Steps

1. From the Working Environments Configuration page, click **Add Data Source > Add StorageGRID**.



2. In the Add StorageGRID Service dialog, enter the details for the StorageGRID service and click **Continue**.
 - a. Enter the name you want to use for the Working Environment. 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 BlueXP classification can access the buckets in StorageGRID.

Add StorageGRID

BlueXP Classification can scan data from NetApp StorageGRID, which uses the S3 protocol. [Learn more](#)

To continue, provide the following details. Next, you'll select the buckets you want to scan.

Name the Working Environment	Endpoint URL
<input type="text"/>	<input type="text"/>
Access Key	Secret Key
<input type="text"/>	<input type="text"/>

Result

StorageGRID is added to the list of working environments.

Enable and disable compliance scans on StorageGRID buckets

After you enable BlueXP classification on StorageGRID, the next step is to configure the buckets that you want to scan. BlueXP classification discovers those buckets and displays them in the working environment you created.

Steps

1. In the Configuration page, click **Configuration** from the StorageGRID working environment.

1 Working Environments
+ Add Active Directory
Add Working Environment ▼

QA - StorageGRID | 8 Buckets
Object Storage

Scanner Group name: default
Working Environment ID:
74669f449b2e4b5ea213618cd29104eb

Configuration

Scan Mode

3 Classified

5 Mapped

3 Not Scanned

Continuously scanning all selected Buckets

2. Enable mapping-only scans, or mapping and classification scans, on your buckets.

Buckets selected for Classification scan (5/8)

Scan	Storage Repository (Bucket)	Mapping status	Classification status	Required Action
Off Map Map & Classify	bucketadipro	<div>Finished 2024-09-05 10:33</div> <div>Last full cycle: 2024-09-05 10:33</div>	<div> <div>Mapped: 84</div> <div>Classified: 5</div> </div>	...
Off Map Map & Classify	datasense-0-files	<div>Finished 2024-09-05 08:00</div> <div>Last full cycle: 2024-09-05 08:00</div>		...
Off Map Map & Classify	datasense-10tb	<div>Running 2024-09-04 07:25</div>	<div> <div>Mapped: 3.7M</div> <div>Classified: 2.1M</div> </div>	...
Off Map Map & Classify	datasense-1tb	<div>Running 2024-09-05 09:05</div> <div>Last full cycle: 2024-09-05 03:04</div>	<div>Mapped: 1.3M</div>	...
Off Map Map & Classify	datasense-1tb-2	<div>Running 2024-09-05 09:06</div> <div>Last full cycle: 2024-09-05 03:05</div>	<div>Mapped: 1.3M</div>	...
Off Map Map & Classify	datasense-1tb-3	<div>Not scanning</div>		...

To:	Do this:
Enable mapping-only scans on a bucket	Click Map
Enable full scans on a bucket	Click Map & Classify
Disable scanning on a bucket	Click Off

Result

BlueXP classification starts scanning the buckets that you enabled. 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 BlueXP classification

You can integrate a global Active Directory with BlueXP classification to enhance the results that BlueXP 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 BlueXP classification to scan CIFS volumes. This integration provides BlueXP 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. BlueXP classification will look in all integrated Active Directories for user and permission details.

This integration provides additional information in the following locations in BlueXP 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 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.



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

Supported data sources

An Active Directory integration with BlueXP classification can identify data from within the following data sources:

- On-premises ONTAP systems
- Cloud Volumes ONTAP
- Azure NetApp Files
- FSx for ONTAP
- OneDrive accounts and SharePoint accounts (for legacy versions 1.30 and earlier)

There is no support for identifying user and permission information from Database schemas, Google Drive accounts, Amazon S3 accounts, or Object Storage that uses the Simple Storage Service (S3) protocol.

Connect to your Active Directory server

After you've deployed BlueXP classification and have activated scanning on your data sources, you can integrate BlueXP 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 BlueXP classification can read any data that requires elevated permissions. The credentials are stored on the BlueXP classification instance.

For CIFS volumes/file shares, if you want to make sure your files "last accessed times" are unchanged by BlueXP classification scans, we recommend that the user has 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

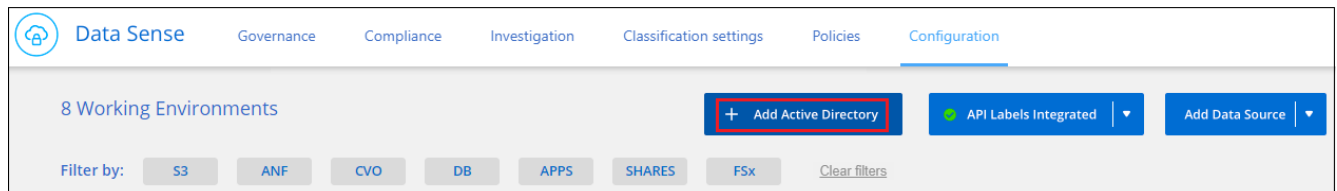
- You must have an Active Directory already set up for the users in your company.
- You must have the information for the Active Directory:
 - DNS Server IP address, or multiple IP addressesor
 - LDAP Server IP address, or multiple IP addresses
 - User Name and Password to access the server
 - Domain 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 BlueXP classification instance:

Protocol	Port	Destination	Purpose
TCP & UDP	389	Active Directory	LDAP

Protocol	Port	Destination	Purpose
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 BlueXP classification Configuration page, click **Add Active Directory**.

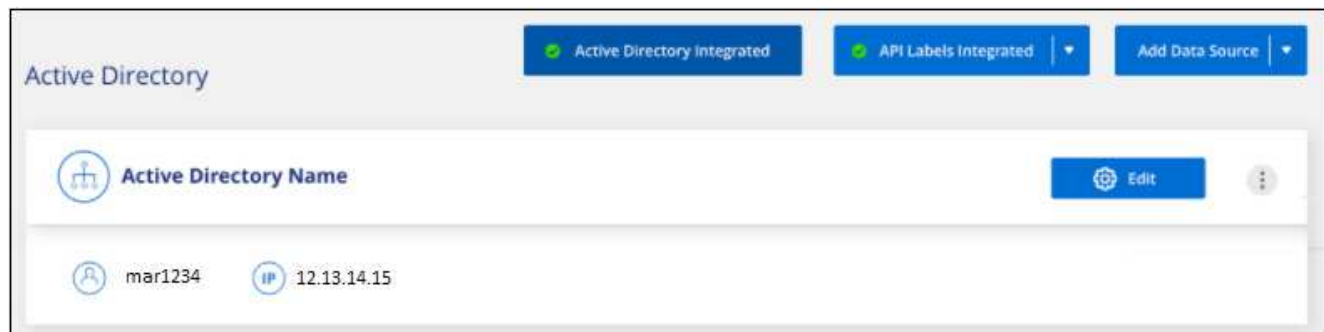


2. In the Connect to Active Directory dialog, enter the Active Directory details and click **Connect**.

You can add multiple IP addresses, if required, by clicking **Add IP**.


 A screenshot of the 'Connect to Active Directory' dialog box. It has a title bar 'Connect to Active Directory'. Below the title bar, there are two input fields: 'Username' (containing 'mar1234') and 'Password' (containing '*****'). Below these, there are two sections. The first section is for 'DNS Server IP address' (selected with a radio button) and 'Domain Name' (containing 'mar@netapp.com'). The 'DNS Server IP address' field contains '12.20.70.00' and has a '+ Add IP' button next to it. The second section is for 'LDAP Server IP Address' (unselected with a radio button) and has a '+ Add IP' button next to its empty input field. Below these sections, there's an 'LDAP Server Port' field containing '389' and a checkbox for 'LDAP Secure Connection' which is unchecked. At the bottom of the dialog, there are two buttons: 'Connect' (highlighted with a red box) and 'Cancel'.

BlueXP 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 if you no longer need it by clicking the  button and then **Remove Active Directory**.

Frequently asked questions about BlueXP classification

This FAQ can help if you're just looking for a quick answer to a question.

BlueXP classification service

The following questions provide a general understanding of BlueXP classification.

What is BlueXP classification?

BlueXP classification is a cloud offering that uses Artificial Intelligence (AI) driven technology to help you understand data context and identify sensitive data across your storage systems. The systems can be working environments that you've added to the BlueXP Canvas and many types of data sources that BlueXP classification can access over your networks. [See the full list below.](#)

BlueXP classification provides pre-defined parameters (such as sensitive information types and categories) to address new data compliance regulations for data privacy and sensitivity, such as GDPR, CCPA, HIPAA, and more.

How does BlueXP classification work?

BlueXP classification deploys another layer of Artificial Intelligence alongside your BlueXP 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. BlueXP classification leverages both artificial intelligence and natural language processing, as opposed to alternative solutions that are commonly built around regular expressions and pattern matching.

BlueXP 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 BlueXP classification works.](#)

[Learn more about the use cases for BlueXP classification.](#)

What about the architecture of BlueXP classification?

BlueXP 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 allows support for multi-cloud, cross-cloud, private cloud, and on-premises environments.

Which cloud providers are supported?

BlueXP classification operates as part of BlueXP and supports AWS, Azure, and GCP. This provides your organization with unified privacy visibility across different cloud providers.

Does BlueXP classification have a REST API, and does it work with third-party tools?

No, BlueXP classification does not have a REST API.

Is BlueXP classification available through the marketplaces?

Yes, BlueXP and BlueXP classification are available from the AWS, Azure, and GCP marketplaces.

BlueXP classification scanning and analytics

The following questions relate to BlueXP classification scanning performance and the analytics available to users.

How often does BlueXP 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. BlueXP 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.](#)

Note that BlueXP 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. However, if you are concerned with even a very small impact, you can configure BlueXP classification to perform "slow" scans. [See how to reduce the scan speed.](#)

Can I search my data using BlueXP classification?

BlueXP classification offers extensive search capabilities that make it easy to search for a specific file or piece of data across all connected sources. BlueXP 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, working environment 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. BlueXP classification also has RBAC controls 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.

Does BlueXP classification offer reports?

Yes. The information offered by BlueXP classification can be relevant to other stakeholders in your organizations, so we enable you to generate reports to share the insights. The following reports are available for BlueXP classification:

Privacy Risk Assessment report

Provides privacy insights from your data and a privacy risk score. [Learn more.](#)

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. [Learn more.](#)

PCI DSS report

Helps you identify the distribution of credit card information across your files. [Learn more.](#)

HIPAA report

Helps you identify the distribution of health information across your files. [Learn more.](#)

Data Mapping report

Provides information about the size and number of files in your working environments. This includes usage capacity, age of data, size of data, and file types. [Learn more.](#)

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. [Learn mode.](#)

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. [Learn more.](#)

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 BlueXP classification instance](#) and [Deploying BlueXP classification](#) for more information.

When initially adding new data sources you can also choose to only perform a "mapping" scan instead of a full "classification" 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.](#)

BlueXP classification management and privacy

The following questions provide information on how to manage BlueXP classification and privacy settings.

How do I enable BlueXP classification?

First you need to deploy an instance of BlueXP classification in BlueXP, or on an on-premises system. Once the instance is running, you can enable the service on existing working environments, databases, and other data sources from the **Configuration** tab or by selecting a specific working environment.

[Learn how to get started.](#)



Activating BlueXP classification on a data source results in an immediate initial scan. Scan results display shortly after.

How do I disable BlueXP classification?

You can disable BlueXP classification from scanning an individual working environment, database, or file share group from the BlueXP classification Configuration page.

[Learn more.](#)



To completely remove the BlueXP classification instance, you can manually remove the BlueXP classification instance from your cloud provider's portal or on-prem location.

Can I customize the service to my organization's needs?

BlueXP classification provides insights to your data. These insights can be extracted and used for your organization's needs.

Additionally, BlueXP classification provides many ways for you to add a custom list of "personal data" that BlueXP classification will identify in scans, giving you the full picture about where potentially sensitive data resides in *all* your organizations' files.

- You can add unique identifiers based on specific columns in databases you are scanning — we call this **Data Fusion**.
- You can add custom keywords from a text file.
- You can add custom patterns using a regular expression (regex).

[Learn more.](#)

Can I instruct the service to exclude scanning data in certain directories?

Yes. If you want BlueXP 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, BlueXP classification will exclude scanning data in the specified directories.

[Learn more.](#)

Are snapshots that reside on ONTAP volumes scanned?

No. BlueXP 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 BlueXP classification scans volumes that have cold data tiered to object storage, 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 scan doesn't heat up the cold data—it stays cold and remains in object storage.

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.

What sources of data can be scanned with BlueXP classification?

BlueXP classification can scan data from working environments that you've added to the BlueXP Canvas and from many types of structured and unstructured data sources that BlueXP classification can access over your networks.

See [Supported working environments and data sources](#).

Are there any restrictions when deployed in a Government region?

BlueXP classification is supported when the Connector is deployed in a Government region (AWS GovCloud, Azure Gov, or Azure DoD) - also known as "Restricted mode". When deployed in this manner, BlueXP classification has the following restrictions:

NOTE This information is relevant only for BlueXP classification legacy versions 1.30 and earlier.

- OneDrive accounts, SharePoint accounts, and Google Drive accounts can't be scanned.
- Microsoft Azure Information Protection (AIP) label functionality can't be integrated.

What data sources can I scan if I install BlueXP classification in a site without internet access?

BlueXP classification can only scan data from data sources that are local to the on-premises site. At this time, BlueXP 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

See [Supported working environments and data sources](#).

Which file types are supported?

BlueXP classification scans all files for category and metadata insights, and displays all file types in the file types section of the dashboard.

When BlueXP classification detects Personal Identifiable Information (PII), or when it performs a DSAR search, only the following file formats are supported:

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

What kinds of data and metadata does BlueXP classification capture?

BlueXP 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:** BlueXP 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:** BlueXP classification scans 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 BlueXP 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.

BlueXP 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 Investigation Details page and in Data Investigation Reports.

BlueXP 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 BlueXP classification scans](#).

Can I limit BlueXP classification information to specific users?

Yes, BlueXP classification is fully integrated with BlueXP. BlueXP users can only see information for the working environments they are eligible to view according to their workspace privileges.

Additionally, if you want to allow certain users to just view BlueXP classification scan results without having the ability to manage BlueXP classification settings, you can assign those users the Cloud Compliance Viewer role.

[Learn more.](#)

Can anyone access the private data sent between my browser and BlueXP classification?

No. The private data sent between your browser and the BlueXP classification instance are secured with end-to-end encryption using TLS 1.2, which means NetApp and non-NetApp parties can't read it. BlueXP 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 BlueXP classification instance.

How is the data accessed?

BlueXP 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 BlueXP classification.

How much does BlueXP classification cost?

BlueXP classification is a BlueXP core capability and is not charged.

Connector deployment

The following questions relate to the BlueXP Connector.

What is the Connector?

The Connector is software running on a compute instance either within your cloud account, or on-premises, that enables BlueXP to securely manage cloud resources. You must deploy a Connector to use BlueXP classification.

Where does the Connector need to be installed?

- When scanning data in Cloud Volumes ONTAP in AWS or Amazon FSx for ONTAP, you use a connector in AWS.
- When scanning data in Cloud Volumes ONTAP in Azure or in Azure NetApp Files, you use a connector in Azure.
- When scanning data in Cloud Volumes ONTAP in GCP, you use a Connector in GCP.
- When scanning data in on-premises ONTAP systems, NetApp file shares, or databases, you can use a connector in any of these cloud locations.

So if you have data in many of these locations, you may need to use [multiple Connectors](#).

Does BlueXP classification require access to credentials?

BlueXP classification itself doesn't retrieve storage credentials. Instead, they are stored within the BlueXP Connector.

BlueXP classification uses data plane credentials, for example, CIFS credentials to mount shares before scanning.

Can I deploy the Connector on my own host?

Yes. You can [deploy the Connector on-premises](#) on a Linux host in your network or on a host in the cloud. If you're planning to deploy BlueXP classification on-premises, then you may want to install the Connector on-premises as well; but it's not required.

Does communication between the service and the Connector use HTTP?

Yes, BlueXP classification communicates with the BlueXP Connector using HTTP.

What about secure sites without internet access?

Yes, that's also supported. You can [deploy the Connector on an on-premises Linux host that doesn't have internet access. This is also known as "Private mode"](#). Then you can discover on-premises ONTAP clusters and other local data sources and scan the data using BlueXP classification.

BlueXP classification deployment

The following questions relate to the separate BlueXP classification instance.

What deployment models does BlueXP classification support?

BlueXP allows the user to scan and report on systems virtually anywhere, including on-premises, cloud, and hybrid environments. BlueXP classification is normally deployed using a SaaS model, in which the service is enabled via the BlueXP 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 BlueXP classification?

When [deployed in the cloud](#):

- In AWS, BlueXP classification runs on an m6i.4xlarge instance with a 500 GiB GP2 disk. You can select a smaller instance type during deployment.
- In Azure, BlueXP classification runs on a Standard_D16s_v3 VM with a 500 GiB disk.
- In GCP, BlueXP classification runs on an n2-standard-16 VM with a 500 GiB Standard persistent disk.

[Learn more about how BlueXP classification works.](#)

Can I deploy the BlueXP classification on my own host?

Yes. You can install BlueXP 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 BlueXP. See [Deploying BlueXP classification on premises](#) for system requirements and installation details.

What about secure sites without internet access?

Yes, that's also supported. You can [deploy BlueXP classification in an on-premises site that doesn't have internet access](#) for completely secure sites.

Use BlueXP classification

View governance details about the data stored in your organization

Gain control of the costs related to the data on your organizations' storage resources. BlueXP classification identifies the amount of stale data, non-business 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.

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.

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.

Save Opportunities

You may want to investigate the items in the *Saving Opportunities* area to see if there is any data you should delete or tier to less expensive object storage. Click each item to view the filtered results in the Investigation page.

- **Stale Data** - Data that was last modified over 3 years ago.
- **Non-Business Data** - Data considered not to be business related, based on their Category or File Type. This includes:
 - Application Data
 - Audio
 - Executables
 - Images
 - Logs
 - Videos
 - Miscellaneous (general "other" category)
- **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 may be identified in the *Stale Data* category.

Policies with the largest number of results

In the *Policies* area, the Policies with the greatest number of results appear at the top of the list. Click the name of a Policy to display the results in the Investigation page. Click **View All** to view the list of all available Policies.

Click [here](#) to learn more about Policies.

Data Overview

The *Data Overview* section provides a quick overview of all the data that is being scanned. Click the button to download a full data mapping report that includes Usage Capacity, Age of Data, Size of Data, and File Types for all of your working environments and data sources. See [Data Mapping Report](#) for complete details about this report.

Top data repositories listed by data sensitivity

The *Top Data Repositories by Sensitivity Level* area lists the top four data repositories (working environments and data sources) that contain the most sensitive items. The bar chart for each working environment is divided into:

- Non-Sensitive data
- Personal data
- Sensitive Personal data

You can position your cursor over each section to see the total number of items in each category.

Click each area to view the filtered results in the Investigation page so that you can investigate further.

Data listed by types of Open Permissions

The *Open Permissions* area 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

You can position your cursor over each section to see the total number of files in each category. Click each area to view the filtered results in the Investigation page so that you can investigate further.

Age of Data and Size of Data graphs

You may want to investigate the items in the *Age* and *Size* graphs to see if there is any data you should delete or tier to less expensive object storage.

You can position your cursor over a point in the charts to see details about the age or size of the data in that category. Click to view all the files filtered by that age or size range.

- **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 may be identified in the *Age of Data* graph.

Most identified data Classifications

The *Classification* area provides a list of the most identified [Categories](#) and [File types](#) in your scanned data.

Categories

Categories can help you understand what's happening with your data by showing you the types of information that you have. For example, a category like "resumes" or "employee contracts" can include sensitive data. When you investigate the results, you might find that employee contracts are stored in a nonsecure location. You can then correct that issue.

See [Viewing files by categories](#) for more information.

File types

Reviewing your file types can help you control your sensitive data because you might find that certain file types are not stored correctly.

See [Viewing file types](#) for more information.

Data Mapping Report

The Data Mapping 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 first lists an overview that summarizes all your working environments and data sources, and then it provides an analysis for

each working environment.

The report includes the following information:

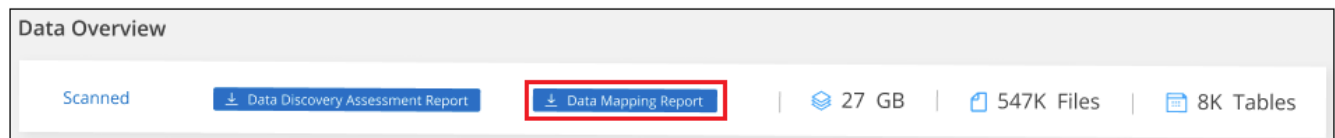
Category	Description
Usage Capacity	For all working environments: Lists the number of files and the used capacity for each working environment. For single working environments: Lists the files that are using the most capacity.
Age of Data	Provides three charts and graphs for when files were created, 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 working environments.
File Types	Lists the total number of files and the used capacity for each type of file being stored in your working environments.

Generate the Data Mapping Report

You generate this report from the Governance tab in BlueXP classification.

Steps


1. From the BlueXP menu, click **Governance > Classification**.
2. Click **Governance**, and then click the **Data Mapping Report** button.



Result

BlueXP 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 BlueXP classification instance and you'll see a pop-up message about the exact location. When BlueXP 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 BlueXP classification is deployed in the cloud, you'll need to SSH to the BlueXP classification instance to download .pdf file. [See how to access data on the Classification instance.](#)

Note that you can customize the company name that appears on the first page of the report from the top of the BlueXP classification page by clicking  and then clicking **Change company name**. The next time you generate the report it will include the new name.

Data Discovery Assessment Report

The 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. 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).

After the assessment, this report identifies areas where you can:

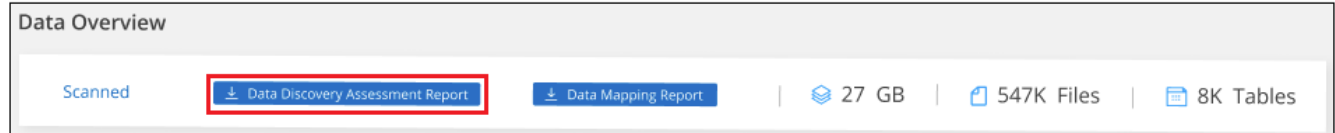
- Reduce storage costs by changing your retention policy, or by moving or deleting certain data (stale, duplicate, or non-business data)
- Protect your data that has broad permissions by revising global group management policies
- Protect your data that has personal or sensitive personal information by moving PII to more secure data stores

Generate the Data Discovery Assessment Report

You generate this report from the Governance tab in BlueXP classification.

Steps

1. From the BlueXP menu, click **Governance > Classification**.
2. Click **Governance**, and then click the **Data Discovery Assessment Report** button.



Result

BlueXP classification generates a .pdf report that you can review and send to other groups as needed.

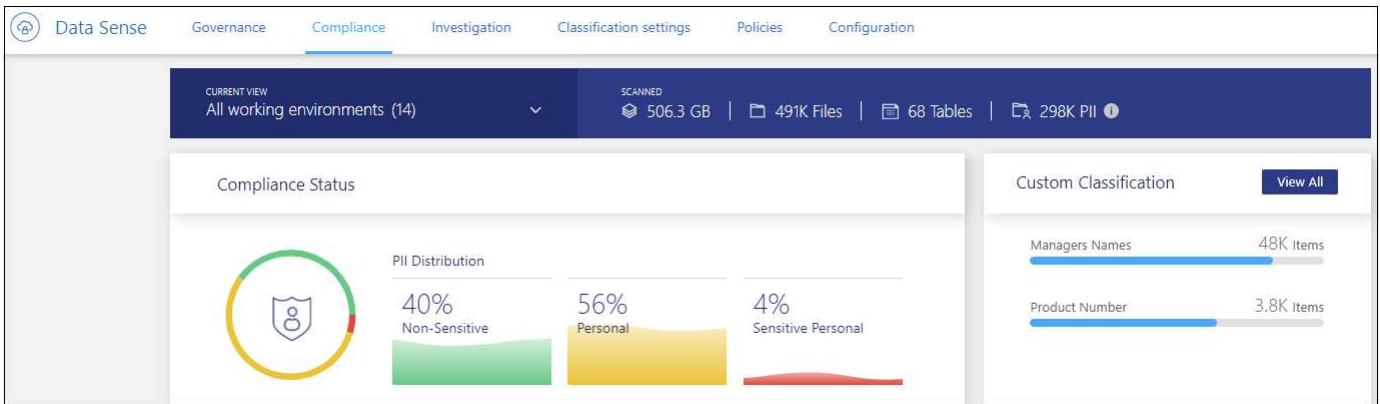
View compliance details about the private data stored in your organization

Gain control of your private data by viewing details about the personal data (Pii) and sensitive personal (SPii) data in your organization. You can also gain visibility by reviewing the categories and file types that BlueXP classification found in your data.



The capabilities described in this section are available only if you have chosen to perform a full classification scan on your data sources. Data sources that have had a mapping-only scan do not show file-level details.

By default, the BlueXP classification dashboard displays compliance data for all working environments and databases.



If you want to see data for only some of the working environments, [select those working environments](#).

You can also 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

BlueXP classification automatically identifies specific words, strings, and patterns (Regex) inside the data. For example, Personal Identification Information (PII), credit card numbers, social security numbers, bank account numbers, passwords, and more. [See the full list](#). BlueXP classification identifies this type of information in individual files, in files within directories (shares and folders), and in database tables.

Additionally, if you have added a database server to be scanned, the *Data Fusion* feature allows you to scan your files to identify whether unique identifiers from your databases are found in those files or other databases. See [Adding personal data identifiers using Data Fusion](#) for details.

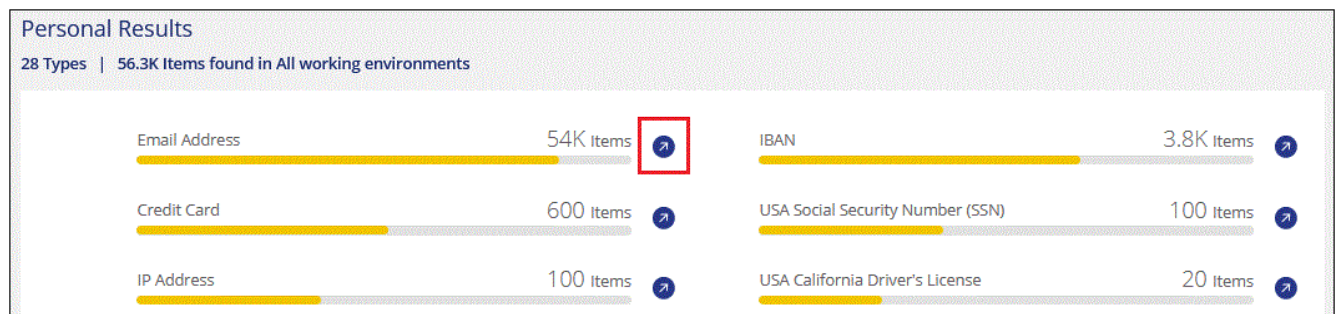
For some types of personal data, BlueXP classification uses *proximity validation* to validate its findings. The validation occurs by looking for one or more predefined keywords in proximity to the personal data that was found. For example, BlueXP classification identifies a U.S. social security number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

Steps

1. From the BlueXP left navigation menu, click **Governance > Classification** and then click the **Compliance** tab.
2. To investigate the details for all personal data, click the icon next to the personal data percentage.



3. To investigate the details for a specific type of personal data, click **View All** and then click the **Investigate Results** 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, clicking **Investigate Results** to see masked information, or by downloading the file list.

The 2 screenshots below show personal data found in individual files, and found in files within directories (shares and folders). You can also select the **Structured** tab to view personal data found in databases.

Unstructured (54.6K Files) | Directories (6 Folders) | Structured (3 Tables) | Search by File Table or location

54.6K items | 1.95 GB

Tags | Assign to | Label | Move | Copy | Delete

File Name | Personal | Sensitive Personal | Data Subjects | File Type

customer-data.xls | S3 | 688 | 0 | **63** | XLS

Tags: Credit Cards | gidi | tartanpion

Working Environment (Account): S3 - 759995470648

Storage Repository (Bucket): compliancedemofiles

File Path: /Patterns/NEW SSN/customer-data.xls

Category: Miscellaneous Spreadsheets

File Size: 142.35 KB

Discovered Time: 2020-11-16 12:40

Created Time: 2019-12-16 12:18 | Last Modified: 2019-12-16 12:18

Open Permissions: NOT PUBLIC

Duplicates: 2 | View Details

Tags: 3 tags

Assigned to: Alona Tyupa

Assign a Label to this file

Copy File

Move File

Delete File

Give feedback on this result

Unstructured (491.4K Files) | Directories (60.7K Folders) | Structured (45 Tables) | Search by File, Table or location

60.7K items | 2.3 GB

Tags | Assign to | Label | Move | Copy | Delete

Directory Name | Storage Repository | Personal | Sensitive Personal | Type

cifs_labs_share | CVO | cifs_labs | 4 | 1 | Share

/datasensecopy/C\$/... | ANF | datasensecopy | 2 | 10 | Folder

Working Environment: Azure NetApp Files

Storage Repository (Volume): datasensecopy

Directory Path: /datasensecopy/copy_63/contextual_data/C\$/Users/shraga.WESTEROS/Desktop/...

Discovered Time: 2022-07-10 22:58

Last Modified: 2020-02-06 09:57

View files that contain sensitive personal data

BlueXP 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](#). BlueXP classification identifies this type of information in individual files, in files within directories (shares and folders), and in database tables.

BlueXP classification uses artificial intelligence (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.

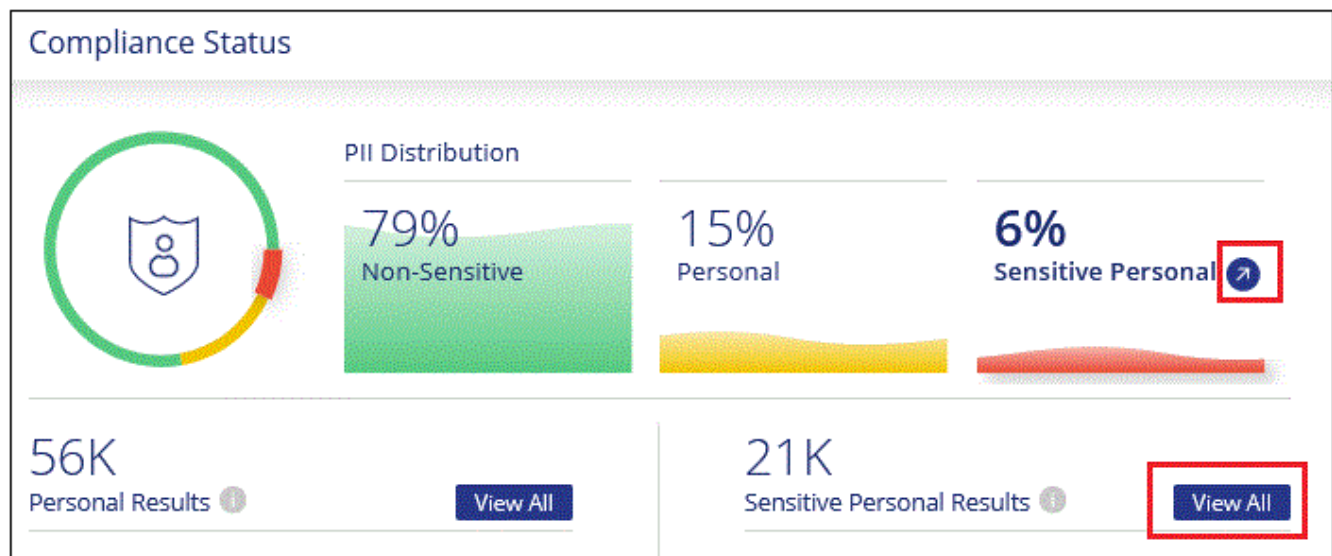
For example, one sensitive GDPR data category is ethnic origin. Because of its NLP abilities, BlueXP classification can distinguish the difference between a sentence that reads "George is Mexican" (indicating sensitive data as specified in article 9 of the GDPR), versus "George is eating Mexican food."



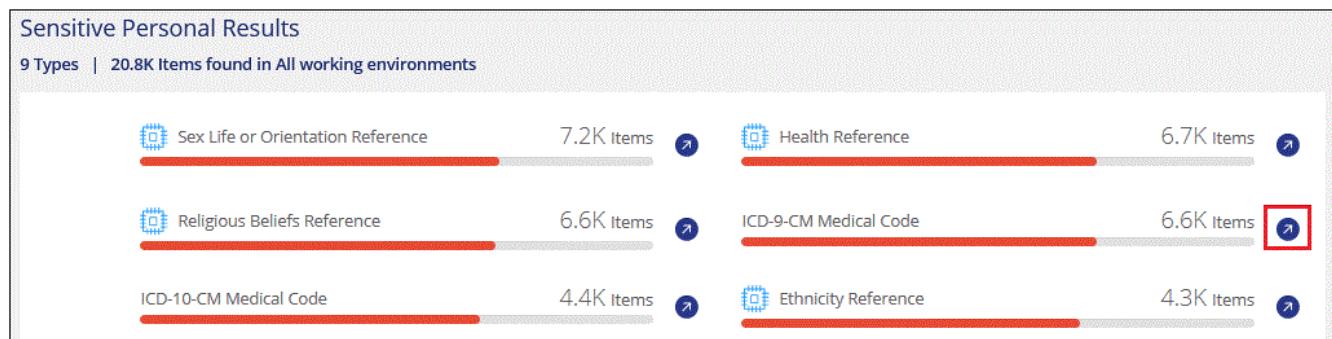
Only English is supported when scanning for sensitive personal data. Support for more languages will be added later.

Steps

1. From the BlueXP left navigation menu, click **Governance > Classification** and then click the **Compliance** tab.
2. To investigate the details for all sensitive personal data, click the icon next to the sensitive personal data percentage.



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

View files by categories

BlueXP 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. [See the list of categories.](#)

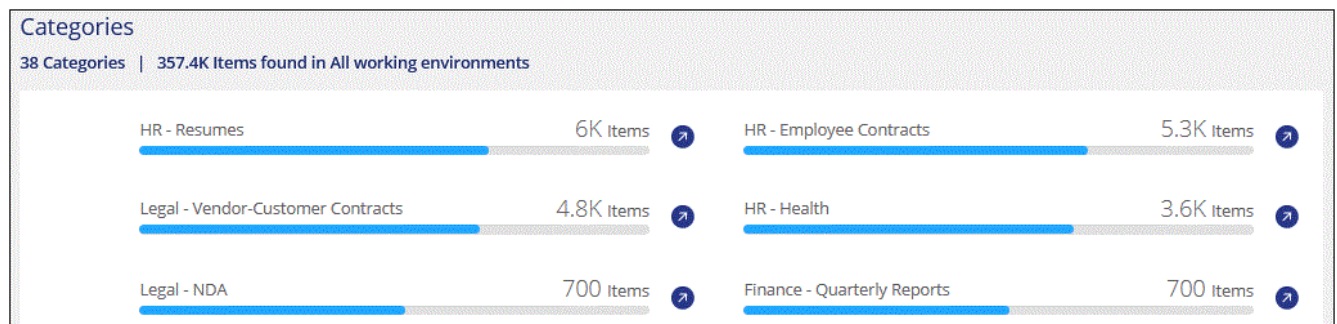
Categories can help you understand what's happening with your data by showing you the types of information that you have. For example, a category like resumes or employee contracts can include sensitive data. When you investigate the results, you might find that employee contracts are stored in a nonsecure location. You can then correct that issue.



English, German, and Spanish are supported for categories. Support for more languages will be added later.

Steps

1. From the BlueXP left navigation menu, click **Governance > Classification** and then click the **Compliance** tab.
2. Click the **Investigate Results** icon for one of the top 4 categories directly from the main screen, or click **View All** and then click the icon for any of the categories.



3. 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.

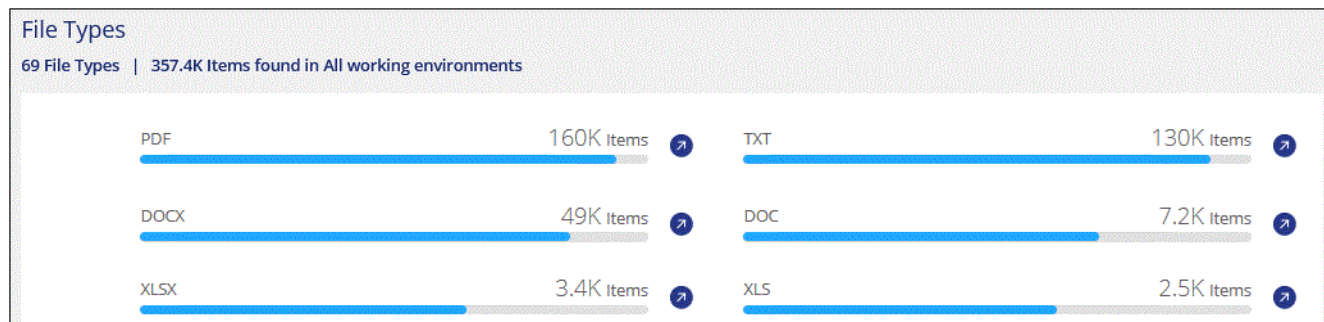
View files by file types

BlueXP classification takes the data that it scanned and breaks it down by file type. Reviewing your file types can help you control your sensitive data because you might find that certain file types are not stored correctly. [See the list of file types.](#)

For example, you might be storing CAD files that include very sensitive information about your organization. If they are unsecured, you can take control of the sensitive data by restricting permissions or moving the files to another location.

Steps

1. From the BlueXP left navigation menu, click **Governance > Classification** and then click the **Compliance** tab.
2. Click the **Investigate Results** icon for one of the top 4 file types directly from the main screen, or click **View All** and then click the icon for any of the file types.



- 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.

View Dashboard data for specific working environments

You can filter the contents of the BlueXP classification dashboard to see compliance data for all working environments and databases, or for just specific working environments.

When you filter the dashboard, BlueXP classification scopes the compliance data and reports to just those working environments that you selected.

Steps

- Click the filter drop-down, select the working environments that you'd like to view data for, and click **View**.



Categories of private data

There are many types of private data that BlueXP classification can identify in your volumes and databases.

BlueXP classification identifies two types of personal data:

- **Personally identifiable information (Pii)**
- **Sensitive personal information (SPii)**



If you need BlueXP classification to identify other private data types, such as additional national ID numbers or healthcare identifiers, email ng-contact-data-sense@netapp.com with your request.

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 BlueXP 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	No	✓	✓	✓		✓
	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
------	------------	--------------------------	---------	--------	---------	--------	----------

	Spanish Tax Identification Number	Yes	✓	✓	✓		
	Swedish ID	Yes	✓	✓	✓		
Type	Texas Driver's License	Yes	✓	✓	✓		
	U.K. ID (NINO)	Yes	✓	✓	✓		
	USA California Driver's License	Yes	✓	✓	✓		
	USA Indiana Driver's License	Yes	✓	✓	✓		
	USA New York Driver's License	Yes	✓	✓	✓		
	USA Social Security Number (SSN)	Yes	✓	✓	✓		

Types of sensitive personal data

BlueXP classification can find the following sensitive personal information (SPii) in files.

The items in this category can be recognized only in English at this time.

- **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

BlueXP 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	✓	✓	✓

Category	Type	English	German	Spanish
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 are identified in the same supported languages:

- Application Data
- Archive Files
- Audio
- Breadcrumbs from BlueXP 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

BlueXP classification scans all files for category and metadata insights and displays all file types in the file types section of the dashboard.

But when BlueXP classification detects Personal Identifiable Information (PII), or when it performs a DSAR search, only the following file formats are supported:

.CSV, .DCM, .DICOM, .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 BlueXP 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 BlueXP classification finds. We break it down by *precision* and *recall*:

Precision

The probability that what BlueXP 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 BlueXP classification to find what it should. For example, a recall rate of 70% for personal data means that BlueXP classification can identify 7 out of 10 files that actually contain personal information in your organization. BlueXP 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 BlueXP 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%

Investigate the data stored in your organization

You can investigate the data from your organization by viewing details in the Data Investigation page. You can navigate to this page from many areas of the BlueXP classification UI, including the Governance and Compliance dashboards.

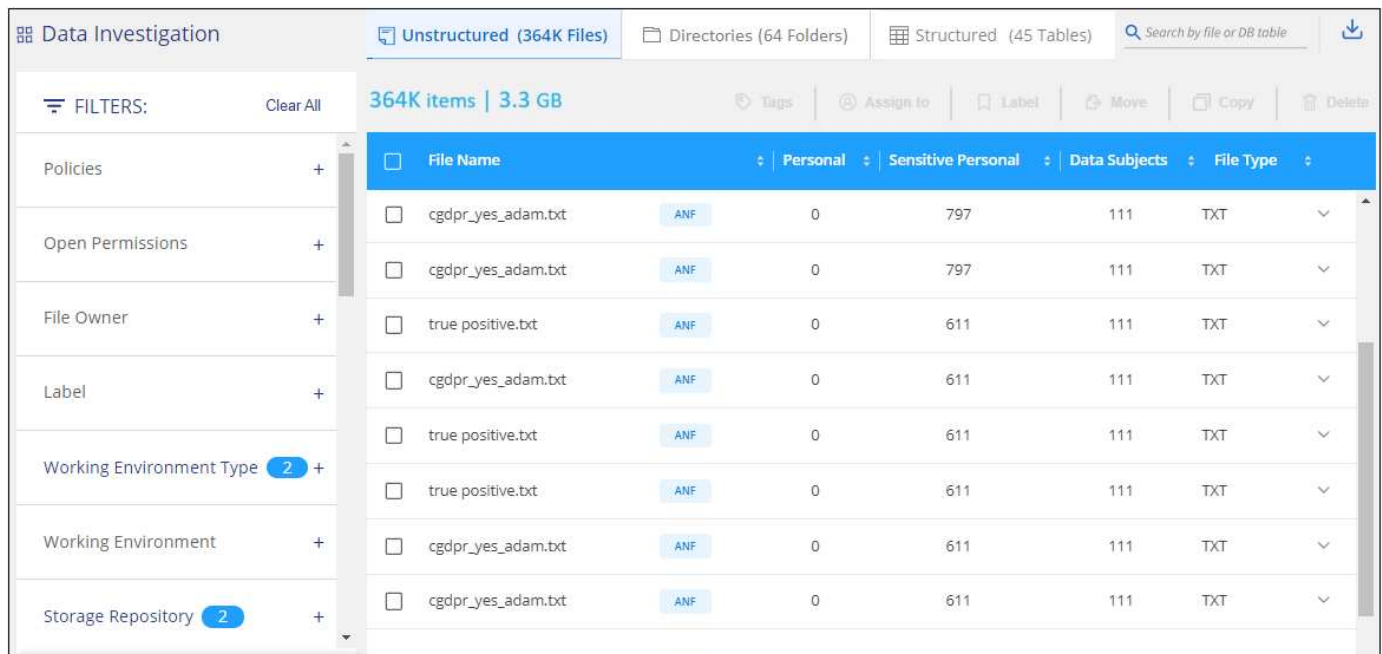


The capabilities described in this section are available only if you have chosen to perform a full classification scan on your data sources. Data sources that have had a mapping-only scan do not show file-level details.

Filter data in the Data Investigation page

You can filter the contents of the investigation page to display only the results you want to see. This is a very powerful feature because after you've refined the data, you can use the button bar at the top of the page to perform a variety of actions, including copying files, moving files, adding a tag or AIP label to the files, and more.

If you want to download the contents of the page as a report after you've refined it, click the  button. [Go here for details about the Data Investigation report.](#)



- The top-level tabs enable you to view data from files (unstructured data), directories (folders and file shares), or from databases (structured data).
- The controls at the top of each column enable you to sort the results in numerical or alphabetical order.
- The left-pane filters enable you to refine the results by selecting the attributes described in the next sections.

Filter data by 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.

Filter	Details
Number of identifiers	Select the range of detected sensitive identifiers per file. Includes personal data and sensitive personal data. When filtering in Directories, BlueXP 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 data by 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.
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 data by time

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 BlueXP 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.

Filter	Details
Last Accessed	<p>Select a time range when the file, or directory (CIFS or NFS only), was last accessed. You can also specify a custom time range to further refine the search results. For the types of files that BlueXP classification scans, this is the last time BlueXP classification scanned the file.</p> <p>Note that BlueXP classification does not extract the "last accessed time" from the following data sources: SharePoint Online, SharePoint On-premises (SharePoint Server), OneDrive, Google Drive, and Amazon S3.</p>

Filter data by 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, BlueXP 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 data by storage type

Use the following filters to view data by storage type.

Filter	Details
Working Environment Type	Select the type of working environment. OneDrive, SharePoint, and Google Drive are categorized under "Apps".
Working Environment name	Select specific working environments.
Storage Repository	Select the storage repository, for example, a volume or a schema.

Filter data by policies

Use the following filter to view data by policies.

Filter	Details
Policies	Select a policy or policies. Go here to view the list of existing policies and to create your own custom policies.

Filter data by analysis status

Use the following filter to view data by the BlueXP 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 BlueXP classification couldn't revert last accessed time, or files that were classified even though BlueXP 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 the Data Investigation results pane you can click  for any single file to view the file metadata.

The screenshot shows a file management interface with a top bar indicating '365K items | 14 GB'. Below this is a navigation bar with tabs: 'File Name', 'Personal', 'Sensitive Personal', 'Data Subjects', and 'File Type'. A table lists files, with the second row selected: 'GM_PD 12-1-09 SP.xls.pdf' (ONDRV, 930, 0, 901, PDF). A red box highlights a chevron icon in the rightmost column of this row. Below the table, a detailed metadata view for the selected file is shown. It includes tags (Decathlon, gidi, IS NOT OK, And 6 more), working environment (OneDrive daylabs.onmicrosoft.com), storage repository (User: ruh@daylabs.onmicrosoft.com), file path (/scattered/26/GM_PD 12-1-09 SP.xls.pdf), category (Miscellaneous Documents), file size (427.46 KB), discovered time (2021-01-12 10:37), created time (2018-05-22 12:38), last modified time (2018-10-22 13:28), and duplicates (None). On the right side of the metadata view, there are buttons for 'Tags: 9 tags', 'Assigned to: Amit Ashbel', 'Assign a Label to this file', 'Copy File', 'Move File', and 'Delete File'. At the bottom right, there is a link 'Give feedback on this result'.

In addition to showing you the working environment 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 Policies](#) because you can see all the information that you can use to filter your data.

Note that not all information is available for all data sources - just what is appropriate for that data source. For example, volume name and permissions are not relevant for database files.

View 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, click **View all Permissions**. This button is available only for data in CIFS shares.

Note that if you see SIDs (Security IDentifiers) instead of user and group names, you should integrate your Active Directory into BlueXP classification. [See how to do this](#).

File Name

Personal

Sensitive Personal

Data Subjects

File Type

Expense Report TPO-1060.pdf

cvo

6

3

16

PDF

Working Environment: WorkingEnvironment1

Repository: Volume Name

File Path: /Prod/labs-base/Expense Report TPO-1060.pdf

Category: Legal

File Size: 22 MB

Last Modified: 2019-08-06 07:51

Open Permissions: NO OPEN PERMISSIONS

File Owner: Avy

View all Permissions

Assign a Label to this file

Delete this file

Permissions list for "Expense Report TPO-1060.pdf"

User / Group	Name	Read	Write
	User Name	✓	✓
	Group Name	✓	⌵
	Group Name	✓	✓
	John L	✓	✓
	George H	✓	✓
	Paul M	✓	✓
	Ringo S	✓	✓

You can click for any group to see the list of users who are part of the group.

Additionally, you can click the name of a user or a group and the Investigation page is displayed with the name of that user or group populated in the “User / Group Permissions” filter 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 view if 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 can also be helpful to make sure certain files that have specific permissions or sensitive information are not unnecessarily duplicated in your storage systems.

All of your files (not including databases) that are 1 MB or larger, and that contain personal or sensitive personal information, are compared to see if there are duplicates. You can use the Investigation page filters “File Size” along with “Duplicates” to see which files of a certain size range are duplicated in your environment.

BlueXP classification uses hashing technology to determine duplicate files. If any file has the same hash code as another file, we can be 100% sure that the files are exact duplicates — even if the file names are different.


You can download the list of duplicate files and send it to your storage admin so they can decide which files, if any, can be deleted. Or you can [delete the file](#) yourself if you are confident that a specific version of the file is not needed.

View all duplicated files

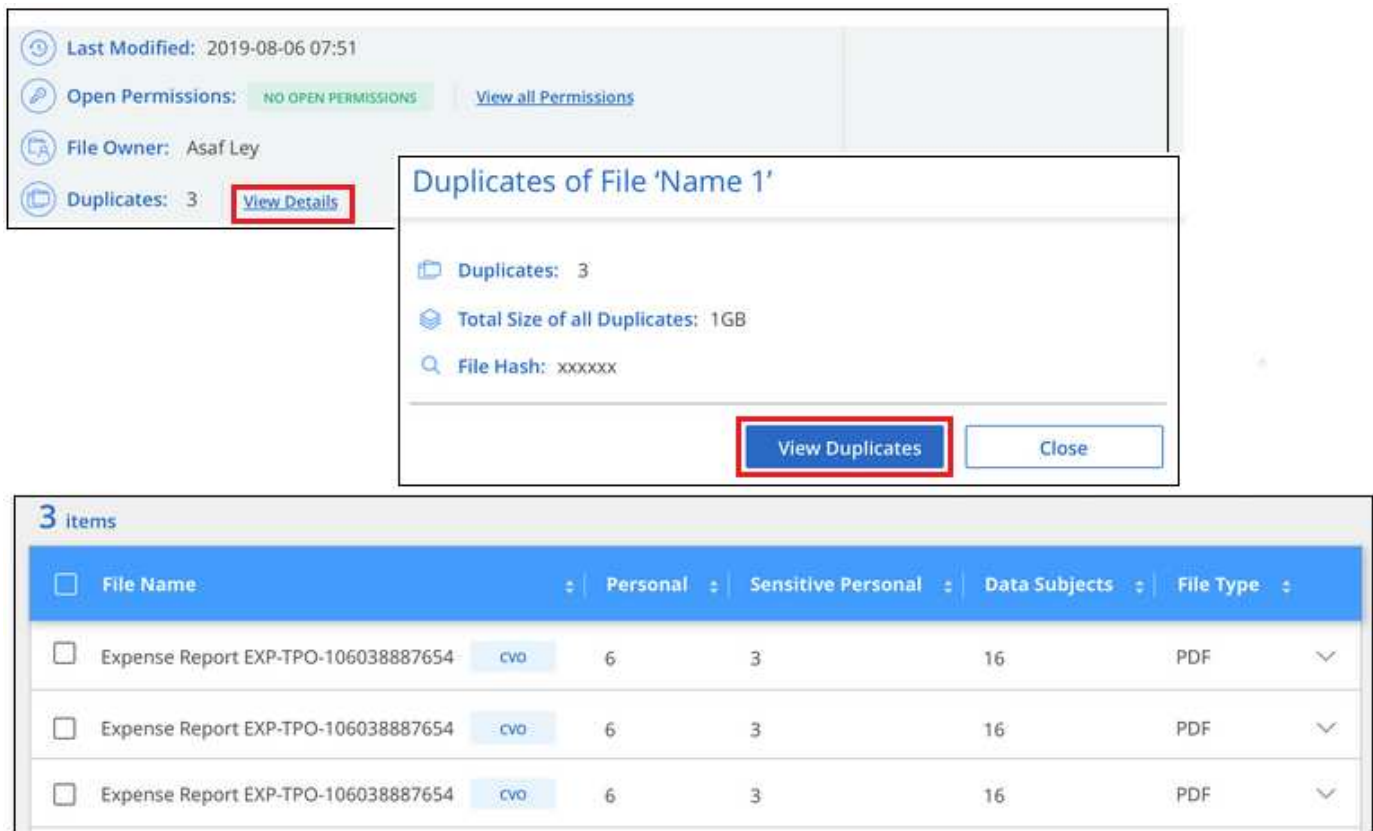
If you want a list of all files that are duplicated in the working environments and data sources you are scanning, you can use the filter called **Duplicates > Has duplicates** in the Data Investigation page.

All duplicated files are displayed in the Results page.

View if a specific file is duplicated

If you want to see if a single file has duplicates, in the Data Investigation results pane you can click  for any single file to view the file metadata. If there are duplicates of a certain file, this information appears next to the *Duplicates* field.

To view the list of duplicate files and where they are located, click **View Details**. In the next page click **View Duplicates** to view the files in the Investigation page.



The screenshot shows the Data Investigation results pane with file metadata: Last Modified: 2019-08-06 07:51, Open Permissions: NO OPEN PERMISSIONS, File Owner: Asaf Ley, and Duplicates: 3. A **View Details** button is highlighted. A modal titled "Duplicates of File 'Name 1'" is open, showing Duplicates: 3, Total Size of all Duplicates: 1GB, and File Hash: xxxxxx. A **View Duplicates** button is highlighted in the modal. Below the modal, a table titled "3 Items" displays the list of duplicate files.

<input type="checkbox"/>	File Name	Personal	Sensitive Personal	Data Subjects	File Type
<input type="checkbox"/>	Expense Report EXP-TPO-106038887654 - cvo	6	3	16	PDF
<input type="checkbox"/>	Expense Report EXP-TPO-106038887654 - cvo	6	3	16	PDF
<input type="checkbox"/>	Expense Report EXP-TPO-106038887654 - cvo	6	3	16	PDF



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 Policy.

Data Investigation Report


The Data Investigation Report is a download of the filtered contents of the Data Investigation page.

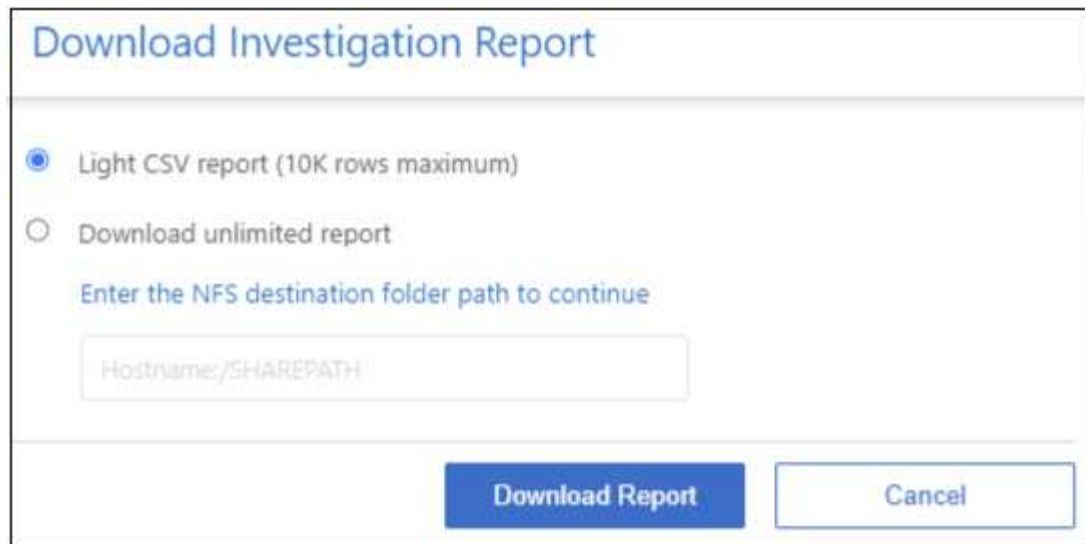
The report is available as a .CSV file that you can save to the local machine.

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

Generate the Data Investigation Report

Steps

1. From the Data Investigation page, click the  button on the top, right of the page.
2. Select to download a .CSV report of the data, and click **Download Report**.

A dialog box titled "Download Investigation Report" with a light blue header. It contains two radio button options: "Light CSV report (10K rows maximum)" which is selected, and "Download unlimited report". Below these is a text prompt "Enter the NFS destination folder path to continue" followed by a text input field containing the placeholder "Hostname:/SHAREPATH". At the bottom right are two buttons: "Download Report" in blue and "Cancel" in white with a blue border.

Download Investigation Report

☒ Light CSV report (10K rows maximum)

☐ Download unlimited report

Enter the NFS destination folder path to continue

Hostname:/SHAREPATH

Download Report Cancel

Result

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

What's included in the Data Investigation Report

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

- File name
- Location type
- Working environment 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
- Category
- Personal information
- Sensitive personal information
- Open permissions
- Scan Analysis Error
- Deletion detection date

A 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 aren't part of 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:

- Working environment type
- Working environment 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
- Working environment name
- Storage repository (for example, a schema)
- Column count
- Row count
- Personal information
- Sensitive personal information

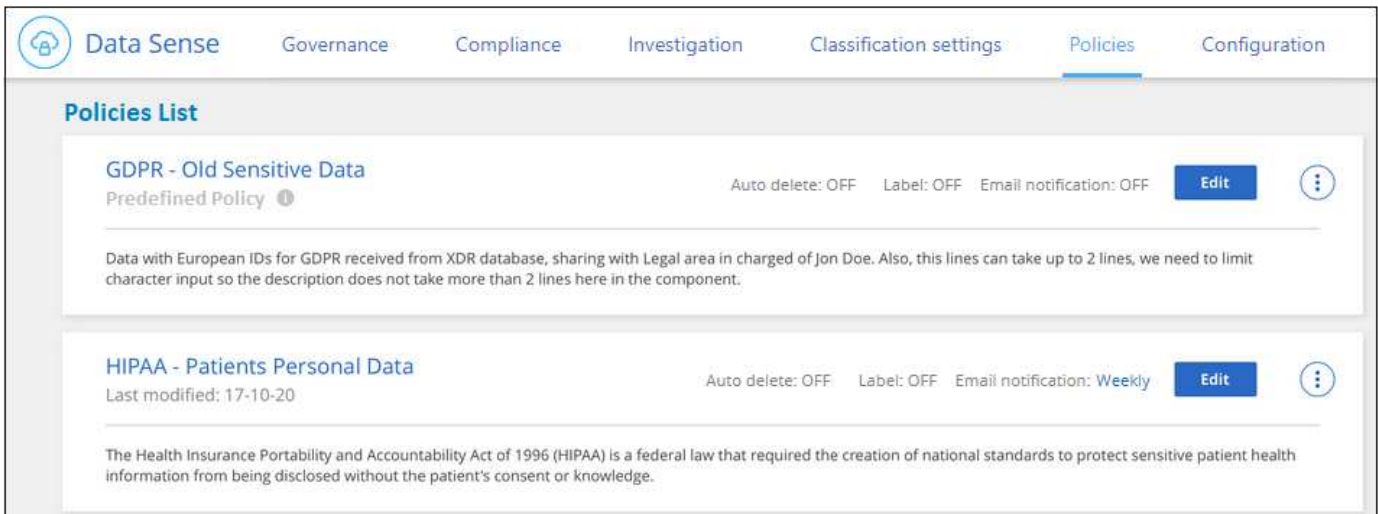
Assign policies to your data

Policies are like a favorites list of custom filters that provide search results in the Investigation page for commonly requested compliance queries. BlueXP classification provides a set of predefined policies based on common customer requests. You can create custom policies that provide results for searches specific to your organization.

Policies provide the following functionality:

- [Predefined policies](#) from NetApp based on user requests
- Ability to create your own custom policies
- Launch the Investigation page with the results from your Policies in one click

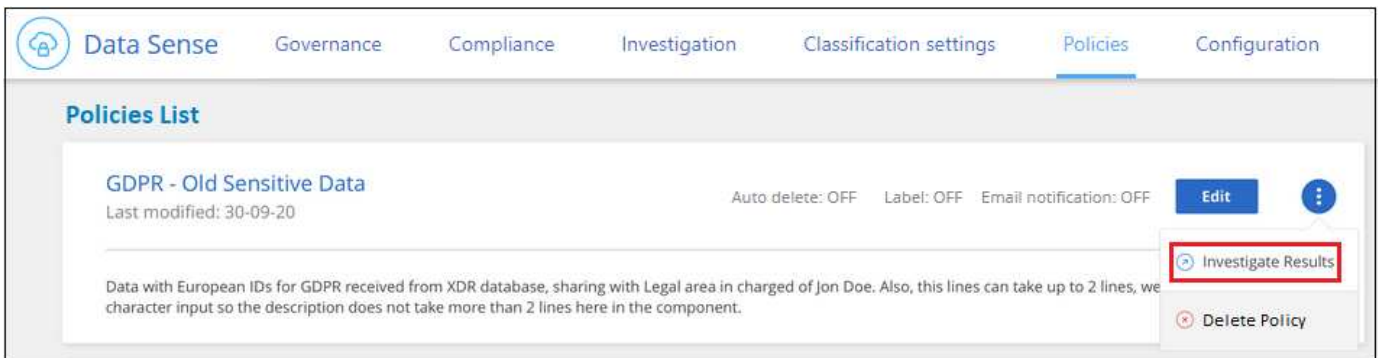
The **Policies** tab in the Compliance Dashboard lists all the predefined and custom policies available on this instance of BlueXP classification.



In addition, policies appear in the list of filters in the Investigation page.

View Policy results in the Investigation page

To display the results for a policy in the Investigation page, click the  button for a specific policy, and then select **Investigate Results**.

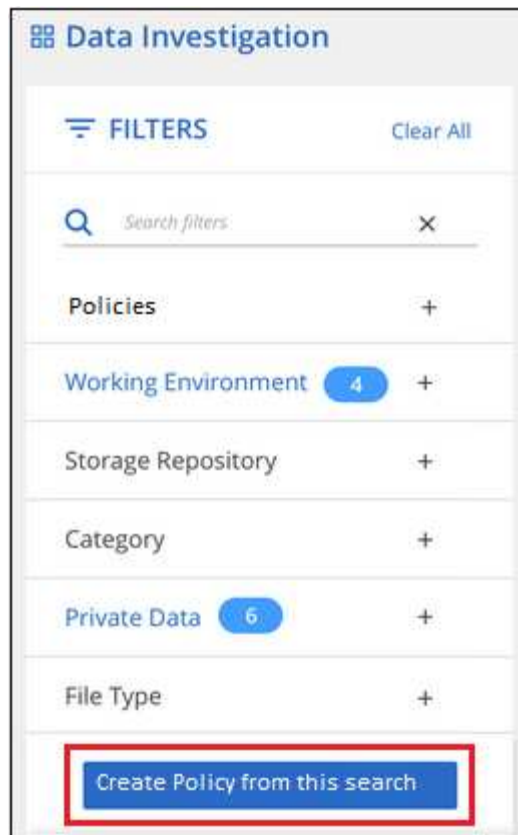


Create custom policies

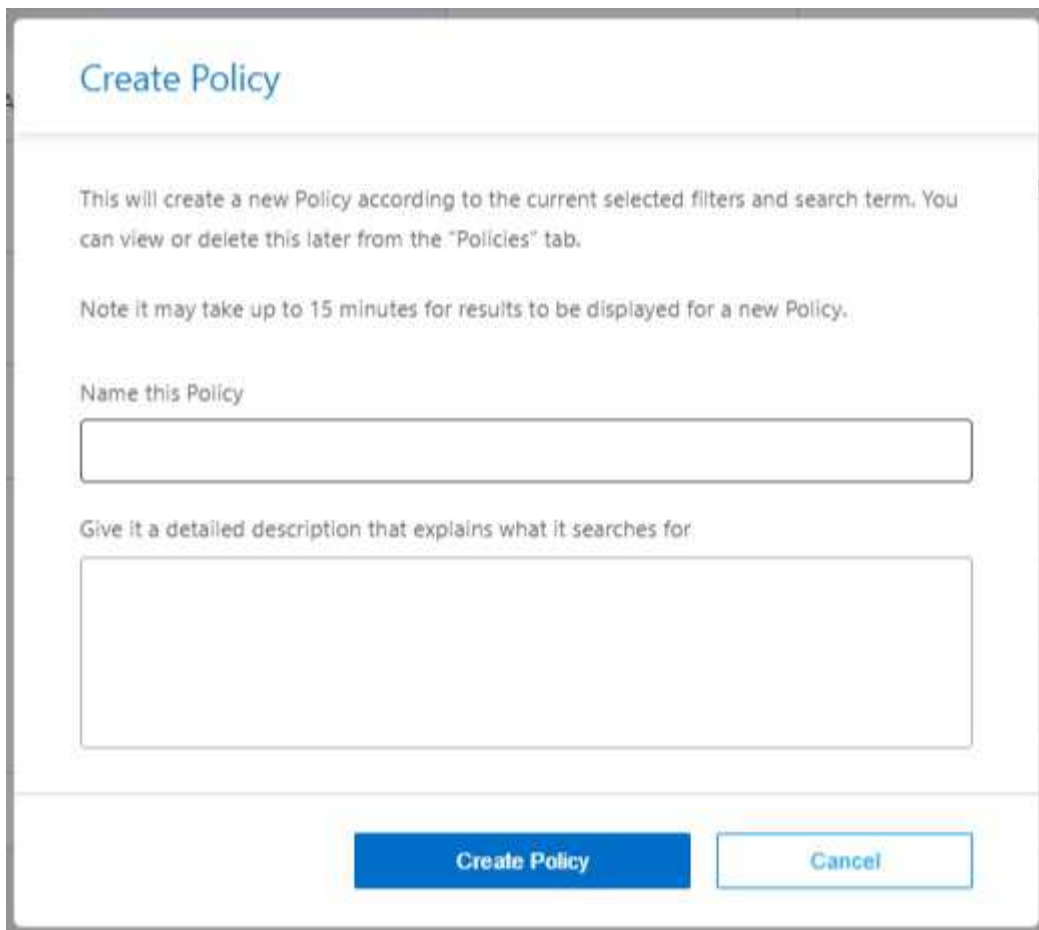
You can create your own custom Policies that provide results for searches specific to your organization. Results are returned for all files and directories (shares and folders) that match the search criteria.

Steps

1. From the Data Investigation page, define your search by selecting all the filters you want to use. See [Filtering data in the Data Investigation page](#) for details.
2. Once you have all the filter characteristics just the way you want them, click **Create Policy from this search**.



3. Name the policy and select other actions that can be performed by the policy:
 - a. Enter a unique name and description.
 - b. Optionally, check the box to automatically delete files that match the policy parameters.
 - c. Click **Create Policy**.



Create Policy

This will create a new Policy according to the current selected filters and search term. You can view or delete this later from the "Policies" tab.

Note it may take up to 15 minutes for results to be displayed for a new Policy.

Name this Policy

Give it a detailed description that explains what it searches for

Create Policy **Cancel**

Result

The new Policy appears in the Policies tab.

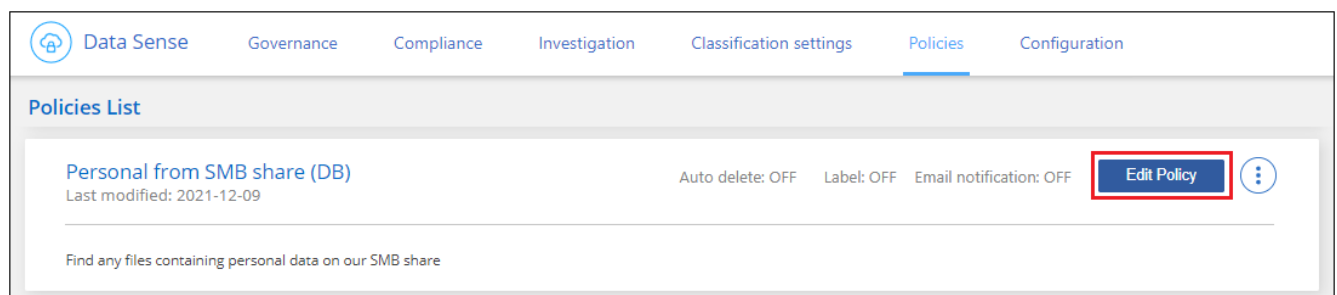
Edit Policies

You can modify any criteria for an existing policy that you previously created. This can be especially useful if you want to change the query (the items you defined using Filters) to add or remove certain parameters.

For Predefined Policies, you can only modify whether email notifications are sent and whether AIP labels are added. No other values can be changed.

Steps

1. From the Policies List page, click **Edit** for the Policy that you want to change.



Data Sense Governance Compliance Investigation Classification settings **Policies** Configuration

Policies List

Personal from SMB share (DB) Last modified: 2021-12-09	Auto delete: OFF Label: OFF Email notification: OFF	Edit Policy ⓘ
Find any files containing personal data on our SMB share		

2. If you just want to change the items on this page (the Name, Description, whether email notifications are sent, and whether AIP labels are added), make the change and click **Save Policy**.

If you want to change the filters for the saved query, click **Edit Query**.

Edit Policy

Edit Query

Name this Policy

Personal from SMB share (DB)

Give it a detailed description that explains what it searches for

Find any files containing personal data on our SMB share

☐ Automatically delete files that match this policy (Every Day)

Email updates about this Policy:

☐ Email all the users in this account Every Day

☐ Send Email Every Day to:

Label:

☐ Automatically label this Policy's matches with:

Cancel

Save Policy

3. In the Investigation page that defines that query, edit the query by adding, removing, or customizing the filters, and click **Save Changes** .

Data Investigation

Unstructured (16 Files)

Directories (0 Folders)

Structured (0 Tables)

Search by File, Table or Location

Download

FILTERS:

Clear All

16 items | 250.2 MB

Tags

Assign to

Label

Move

Copy

Delete

Policies 1

+

Open Permissions

+

User / Group Permissions

+

File Owner

+

Label

+

Working Environment Type

+

Working Environment

+

Save Changes

Cancel Edit Query

File Name

Personal

Sensitive Personal

Data Subjects

File Type

☐ cifs2.json

SHARES

1

0

0

JSON

▼

☐ cifs12.json

SHARES

1

0

0

JSON

▼

☐ TableTextServiceYi.txt

SHARES

1

0

0

TXT

▼

☐ testpass.json

SHARES

1

0

0

JSON

▼

☐ urlp.txt

SHARES

1

0

0

TXT

▼

☐ License.sharpen.txt

SHARES

1

0

1

TXT

▼

☐ TableTextServiceYi.txt

SHARES

1

0

0

TXT

▼

☐ Notice.txt

SHARES

1

0

0

TXT

▼

☐ urlp.txt

SHARES

1

0

0

TXT

▼

☐ Notice.txt

SHARES

1

0

0

TXT

▼

1-16 of 16

Result

The policy is changed immediately. Any actions defined for that policy to send an email, add AIP labels, or delete files will occur at the next internal.

Delete Policies

You can delete any custom policy that you created if you no longer need it. You can't delete any of the predefined policies.

To delete a policy, click the  button for a specific Policy, click **Delete Policy**, and then click **Delete Policy** again in the confirmation dialog.

List of predefined policies

BlueXP classification provides the following system-defined policies:

Name	Description	Logic
Private data - Stale over 7 years	Files containing personal or sensitive personal information, last modified over 7 years ago.	Files containing personal or sensitive personal information, last modified over 7 years ago
Data Subject names - High risk	Files with over 50 Data Subject names.	Files with over 50 Data Subject names
Email Addresses - High risk	Files with over 50 Email Addresses, or DB Columns with over 50% of their rows containing Email Addresses	Files with over 50 Email Addresses, or DB Columns with over 50% of their rows containing Email Addresses

Name	Description	Logic
Personal data - High risk	Files with over 20 Personal data identifiers, or DB Columns with over 50% of their rows containing Personal data identifiers.	Files with over 20 personal, or DB Columns with over 50% of their rows containing personal
Sensitive Personal data - High risk	Files with over 20 Sensitive Personal data identifiers, or DB Columns with over 50% of their rows containing Sensitive Personal data.	Files with over 20 sensitive personal, or DB Columns with over 50% of their rows containing sensitive personal

View compliance reports

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

By default, the BlueXP classification dashboards display compliance and governance data for all working environments, databases, and data sources. If you want to view reports that contain data for only some of the working environments, [select those working environments](#).



- The reports described in this section are available only if you have chosen to 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 can't guarantee 100% accuracy of the personal data and sensitive personal data that BlueXP classification identifies. You should always validate the information by reviewing the data.

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

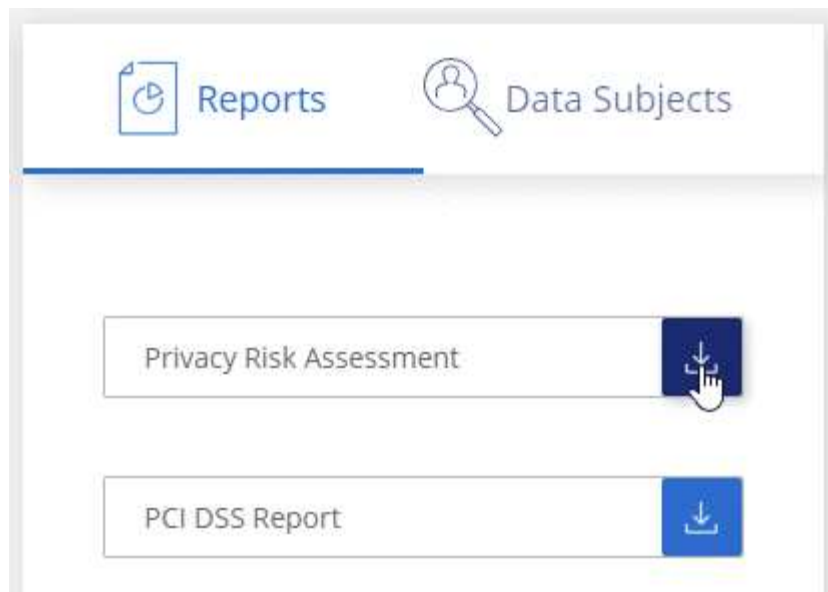
The number of people, by location, for which national identifiers were found.

Generate the Privacy Risk Assessment Report

Go to the Compliance tab to generate the report.

Steps

1. From the BlueXP menu, click **Governance > Classification**.
2. Click **Compliance**, and then click the download icon next to **Privacy Risk Assessment** under **Reports**.



Result

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

Severity score

BlueXP 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%

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 working environments.

Encryption

The percentage of files containing credit card information that are on encrypted or unencrypted working environments. This information is specific to Cloud Volumes ONTAP.

Ransomware Protection

The percentage of files containing credit card information that are on working environments 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

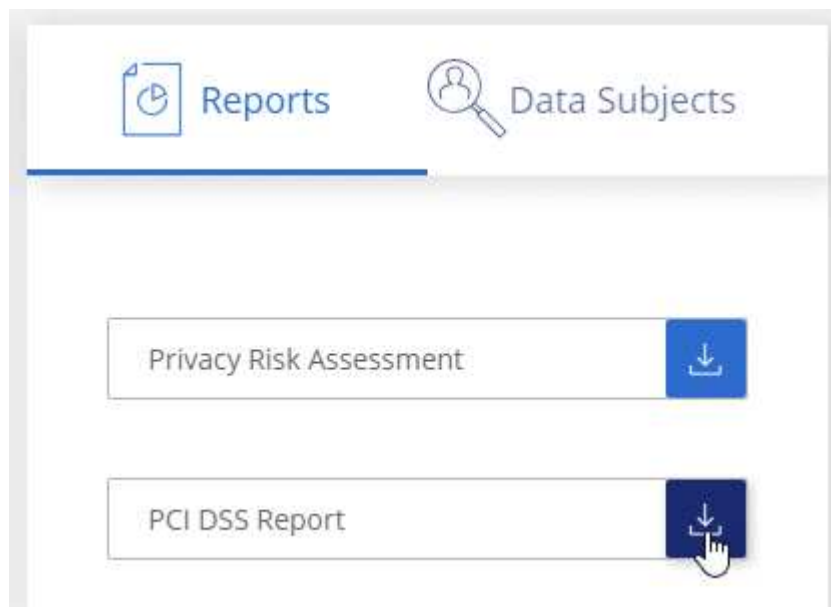
The working environments where the credit card information was found and whether encryption and ransomware protection are enabled.

Generate the PCI DSS Report

Go to the Compliance tab to generate the report.

Steps

1. From the BlueXP menu, click **Governance > Classification**.
2. Click **Compliance**, and then click the download icon next to **PCI DSS Report** under **Reports**.



Result

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

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 BlueXP 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 working environments.

Encryption

The percentage of files containing health information that are on encrypted or unencrypted working environments. This information is specific to Cloud Volumes ONTAP.

Ransomware Protection

The percentage of files containing health information that are on working environments 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

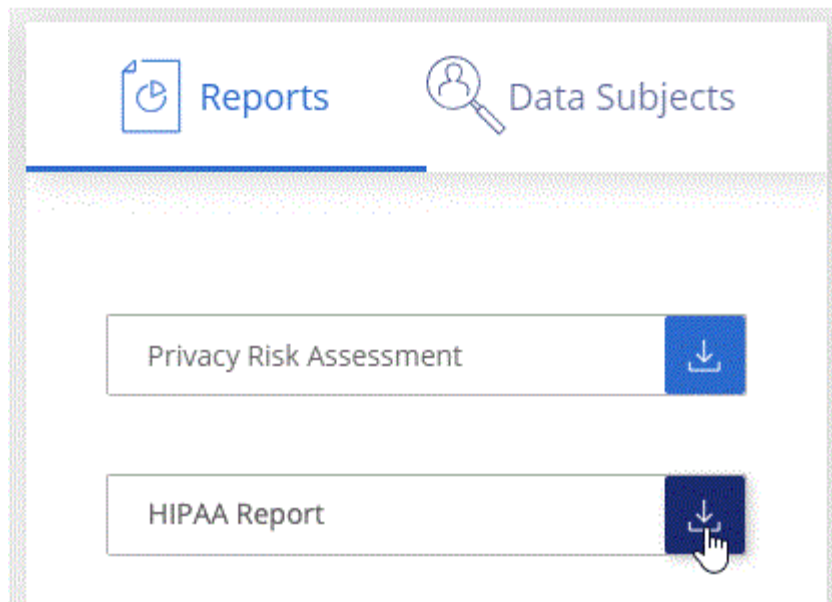
The working environments where the health information was found and whether encryption and ransomware protection are enabled.

Generate the HIPAA Report

Go to the Compliance tab to generate the report.

Steps

1. From the BlueXP menu, click **Governance > Classification**.
2. Click **Compliance**, and then click the download icon next to **HIPAA Report** under **Reports**.



Result

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

What is a Data Subject Access Request?

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 BlueXP classification help you respond to a DSAR?

When you perform a data subject search, BlueXP classification finds all of the files, buckets, OneDrive, and SharePoint accounts that have that person's name or identifier in it. BlueXP 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.



Data subject search is not supported within databases at this time.

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](#).

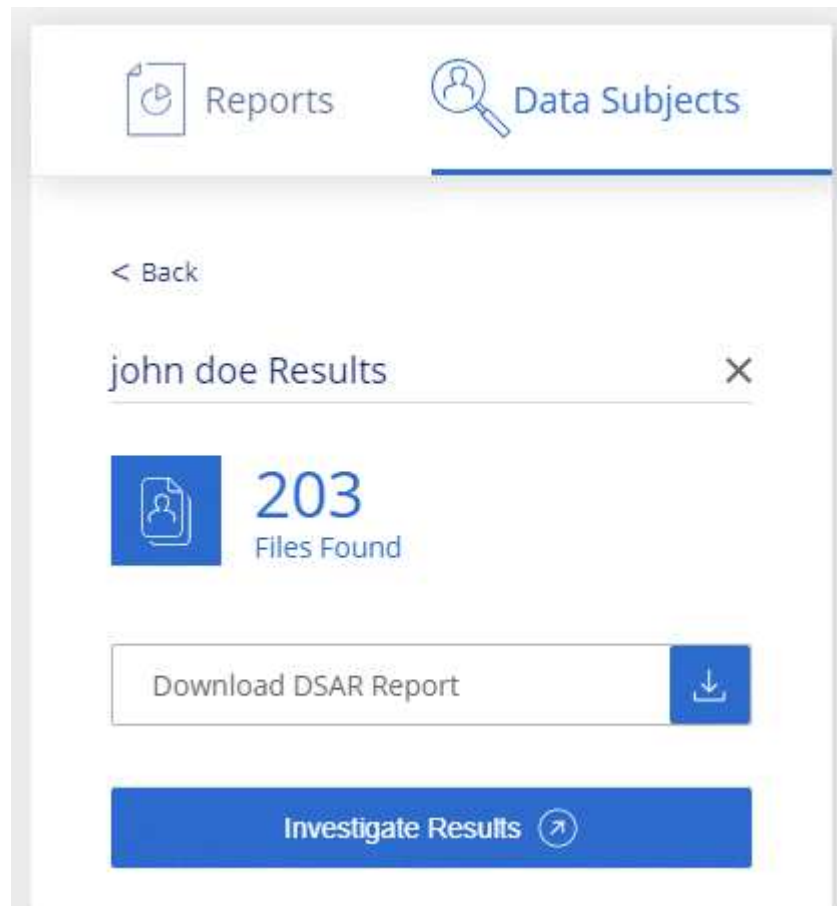


English, German, Japanese, and Spanish are supported when searching for the names of data subjects. Support for more languages will be added later.

Steps

1. From the BlueXP menu, click **Governance > Classification**.
2. Click **Data Subjects**.
3. Search for the data subject's full name or known identifier.

Here's an example that shows a search for the name *john doe*:



4. Choose one of the available options:
 - **Download DSAR Report:** A formal response to the access request that you can send to the data subject. This report contains automatically-generated information based on data that BlueXP classification found on the data subject and is designed to be used as a template. You should complete the form and review it internally before sending it to the data subject.
 - **Investigate Results:** A page that enables you to investigate the data by searching, sorting, expanding details for a specific file, and by downloading the file list.



If there are more than 10,000 results, only the top 10,000 appear in the file list.

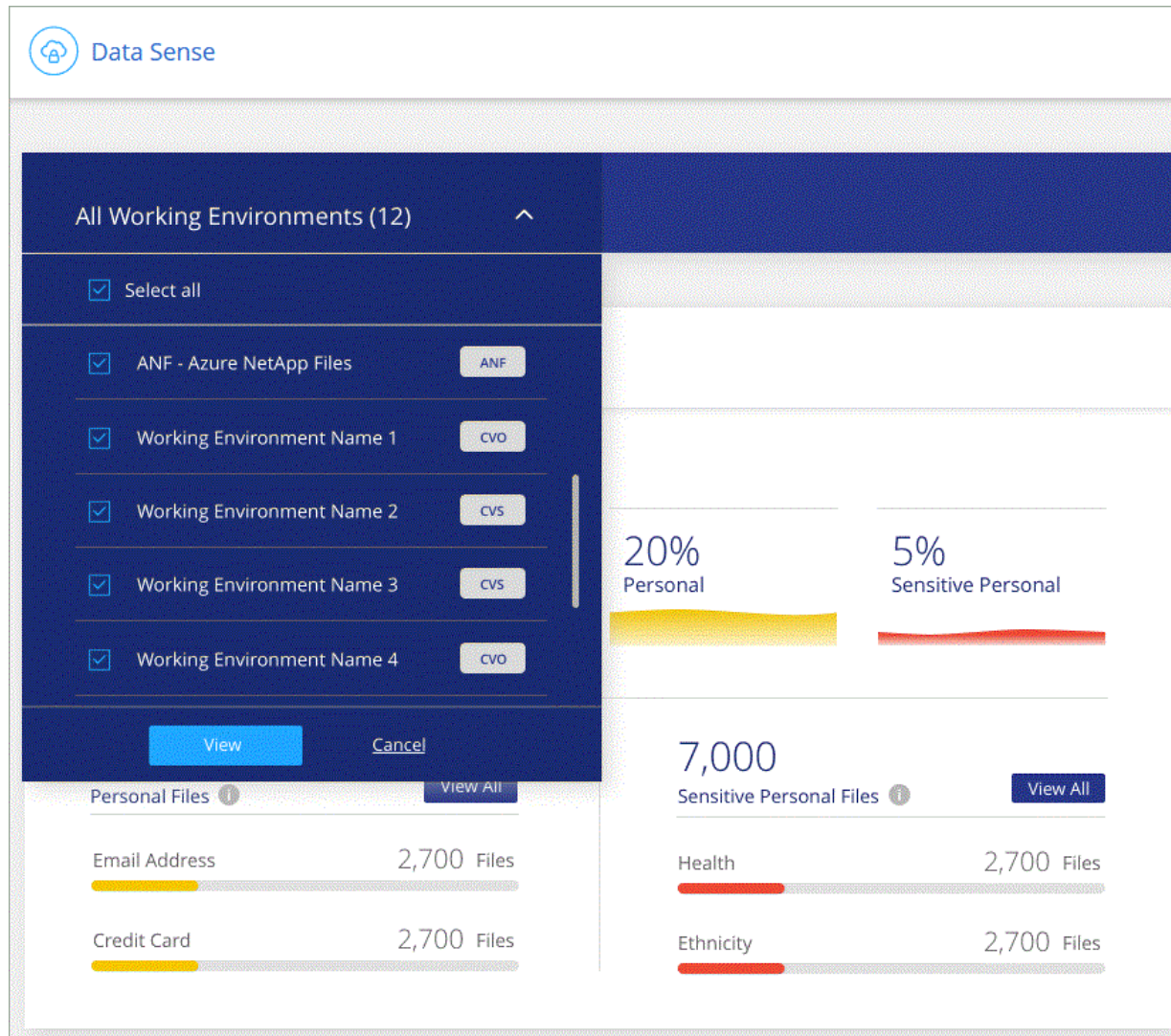
Select the working environments for reports

You can filter the contents of the BlueXP classification Compliance dashboard to see compliance data for all working environments and databases, or for just specific working environments.

When you filter the dashboard, BlueXP classification scopes the compliance data and reports to just those working environments that you selected.

Steps

1. Click the filter drop-down, select the working environments that you'd like to view data for, and click **View**.



Manage BlueXP classification

Exclude specific directories from BlueXP classification scans

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. After you apply this change, the BlueXP classification engine will exclude scanning data in those directories.

Note that BlueXP classification is configured by default to exclude scanning volume snapshot data because that content is identical to the content in the volume.

This functionality is available in BlueXP classification version 1.29 and greater (starting in March 2024).

Supported data sources

Excluding specific directories from BlueXP 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 BlueXP classification system so you can edit a configuration file and run a script. See how to [log in to the BlueXP classification system](#) depending on whether you manually installed the software on a Linux machine or if you deployed the instance in the cloud.



- You can exclude a maximum of 50 directory paths per BlueXP classification system.
- Excluding directory paths may affect scanning times.

Steps

1. On the BlueXP classification system, go to `/opt/netapp/config/custom_configuration` and 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 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 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 BlueXP 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

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 group IDs (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 BlueXP 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 BlueXP classification system so you can edit a configuration file and run a script. See how to [log in to the BlueXP 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 BlueXP 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 BlueXP classification, run the following command from `/opt/netapp/Datasense/tools/customer_configuration/data_providers`:

```
get_data_providers_configuration.sh
```

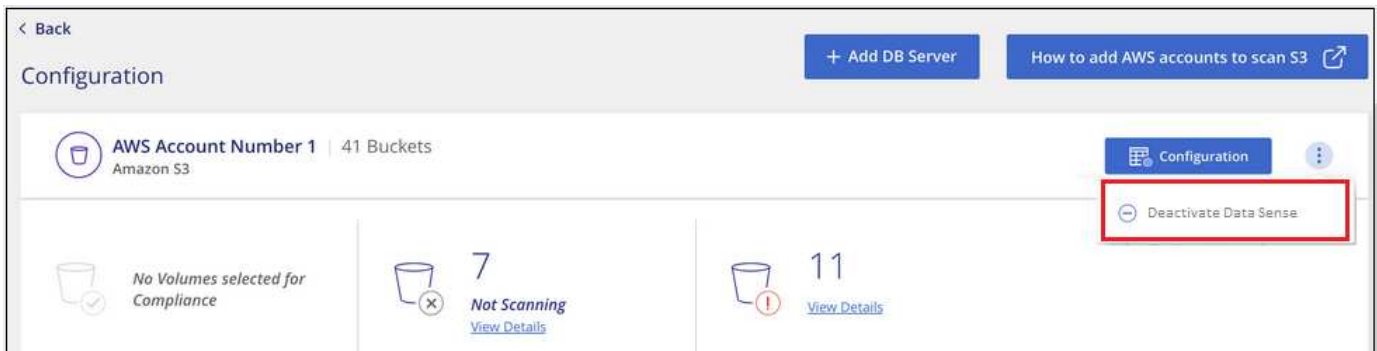
Remove data sources from BlueXP classification

If you need to, you can stop BlueXP classification from scanning one or more working environments, databases, or file share groups.

Deactivate compliance scans for a working environment

When you deactivate scans, BlueXP classification no longer scans the data on the working environment and it removes the indexed compliance insights from the BlueXP classification instance (the data from the working environment itself isn't deleted).


1. From the *Configuration* page, click the  button in the row for the working environment, and then click **Deactivate Data Sense**.



You can also disable compliance scans for a working environment from the Services panel when you select the working environment.

Remove a database from BlueXP classification

If you no longer want to scan a certain database, you can delete it from the BlueXP classification interface and stop all scans.

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


Server.



Remove a group of file shares from BlueXP classification

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

Steps

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



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


Uninstalling BlueXP classification

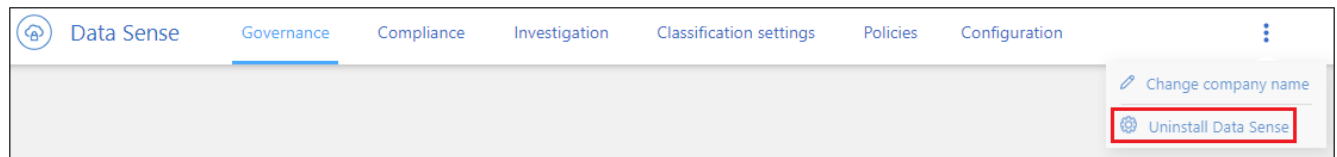
You can uninstall the BlueXP classification software to troubleshoot issues or to permanently remove the software from the host. Deleting the instance also deletes the associated disks where the indexed data resides - all the information BlueXP classification has scanned will be permanently deleted.

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

Uninstall BlueXP classification from a cloud deployment

You can uninstall and delete the BlueXP classification instance from the cloud provider environment if you no longer want to use BlueXP classification.

1. At the top of the BlueXP classification page, click  and then click **Uninstall Data Sense**.




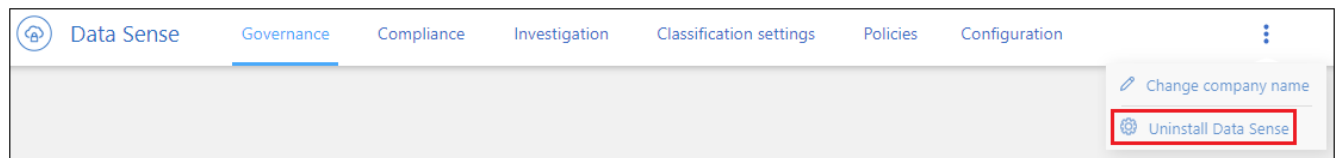
2. In the *Uninstall Data Sense* dialog, type **uninstall** to confirm that you want to disconnect the BlueXP classification instance from the BlueXP Connector, and then click **Uninstall**.
3. Go to your cloud provider's console and delete the BlueXP classification instance. The instance is named *CloudCompliance* with a generated hash (UUID) concatenated to it. For example: *CloudCompliance-16bb6564-38ad-4080-9a92-36f5fd2f71c7*

This deletes the instance and all associated data that had been collected by BlueXP classification.

Uninstall BlueXP classification from an on-premises deployment

You can uninstall BlueXP classification from a host if you no longer want to use BlueXP classification, or if you had an issue that requires reinstallation.

1. At the top of the BlueXP classification page, click  and then click **Uninstall Data Sense**.



2. In the *Uninstall Data Sense* dialog, type **uninstall** to confirm that you want to disconnect the BlueXP classification instance from the BlueXP Connector, and then click **Uninstall**.
3. To uninstall the software from the host, run the `cleanup.sh` script on the host machine, for example:

```
cleanup.sh
```

See how to [log in to the BlueXP classification host machine](#).

Deprecated features

BlueXP classification deprecated features

BlueXP classification is available as a core capability within BlueXP at no additional charge. By including BlueXP classification as a core BlueXP capability available to all customers, NetApp is enabling you to access tailored data management with core features.

There are some features and functionality that are deprecated in the BlueXP core version starting with version 1.31 and later and are still supported in legacy versions 1.30 and earlier.

Supported data sources

Data source	Legacy versions 1.30 and earlier	BlueXP core versions 1.31 and later
Cloud Volumes ONTAP (deployed in AWS, Azure, or GCP)	Yes	Yes
On-premises ONTAP clusters	Yes	Yes
StorageGRID	Yes	Yes
Azure NetApp Files	Yes	Yes
Amazon FSx for ONTAP	Yes	Yes
Google Cloud NetApp Volumes	Yes	Yes
Cloud Volumes Service for Google Cloud	Yes	Yes
Databases	Yes	Yes
Amazon S3	Yes	No
Google Cloud Storage	Yes	No
OneDrive	Yes	No
SharePoint Online	Yes	No
SharePoint On-premises (SharePoint Server)	Yes	No
Google Drive	Yes	No

Compliance features

Feature	Legacy versions 1.30 and earlier	BlueXP core versions 1.31 and later
Identify Personal Identifiable Information (PII)	Yes	Yes
Identify sensitive personal information	Yes	Yes
Respond to Data Subject Access Requests (DSAR)	Yes	Yes

Feature	Legacy versions 1.30 and earlier	BlueXP core versions 1.31 and later
Create a custom list of "personal data" that is identified	Yes	No
Notify users through email when files contain certain PII. (You define this criteria using Policies .)	Yes	No
Use directory-level filters	Yes	Yes
Use directory-level PII analysis	Yes	No

Features to manage your data

Feature	Legacy versions 1.30 and earlier	BlueXP core versions 1.31 and later
Move, copy, and delete source files	Yes	No
Categorize data using Status tags	Yes	No
Categorize data using AIP labels	Yes	No
Assign files to users	Yes	No
Rescan data on demand	Yes	No
Create custom classifiers	Yes	No
Exclude directories from scanning	Yes	Yes
Search for names within files	Yes	Yes
Export data to NFS from investigation	Yes	No
Export data to CSV from investigation	Yes	Yes
Support multiple scanners	Yes	No
Integrate Active Directory	Yes	Yes
Use permission analysis and filters	Yes	Yes
Use the file card	Yes	Yes
Use the heatmap	Yes	Yes
Use actions on Dashboard and file card	Yes	No
Use file access audit logging	Yes	No
Enable file access from the Configuration page	Yes	No
Use certain predefined policies	Yes	No

Deploy BlueXP classification deprecations

Install BlueXP classification on multiple hosts for large configurations with no internet access

Complete a few steps to install BlueXP classification on multiple hosts in an on-premises site that doesn't have internet access - also known as *private mode*. This type of installation is perfect for your secure sites.

For very large configurations where you'll be scanning petabytes of data in sites without internet access, you can include multiple hosts to provide additional processing power. When using multiple host systems, the primary system is called the *Manager node* and the additional systems that provide extra processing power are called *Scanner nodes*.

Follow these steps when installing BlueXP classification software on multiple on-premises hosts in an offline environment.

NOTE This information is relevant only for BlueXP classification legacy versions 1.30 and earlier.

What you'll need

- Verify that all your Linux systems for the Manager and Scanner nodes meet the host requirements.
- Verify that you have installed the two prerequisite software packages (Docker Engine or Podman, and Python 3).
- Make sure you have root privileges on the Linux systems.
- Verify that your offline environment meets the required permissions and connectivity.
- You must have the IP addresses of the scanner node hosts that you plan to use.
- The following ports and protocols must be enabled on all hosts:

Port	Protocols	Description
2377	TCP	Cluster management communications
7946	TCP, UDP	Inter-node communication
4789	UDP	Overlay network traffic
50	ESP	Encrypted IPsec overlay network (ESP) traffic
111	TCP, UDP	NFS Server for sharing files between the hosts (needed from each scanner node to manager node)
2049	TCP, UDP	NFS Server for sharing files between the hosts (needed from each scanner node to manager node)

Steps

1. Follow steps 1 through 8 from the [Single-host installation](#) on the manager node.
2. As shown in step 9, when prompted by the installer, you can enter the required values in a series of prompts, or you can provide the required parameters as command line arguments to the installer.

In addition to the variables available for a single-host installation, a new option **-n <node_ip>** is used to specify the IP addresses of the scanner nodes. Multiple node IPs are separated by a comma.

For example, this command adds 3 scanner nodes:

```
sudo ./install.sh -a <account_id> -c <client_id> -t <user_token> --host  
<ds_host> --manager-host <cm_host> -n <node_ip1>,<node_ip2>,<node_ip3> --no  
-proxy --darksite
```

3. Before the manager node installation completes, a dialog displays the installation command needed for the scanner nodes. Copy the command (for example: `sudo ./node_install.sh -m 10.11.12.13 -t ABCDEF-1-3u69m1-1s35212`) and save it in a text file.
4. On **each** scanner node host:
 - a. Copy the Data Sense installer file (**cc_onprem_installer.tar.gz**) to the host machine.
 - b. Unzip the installer file.
 - c. Paste and run the command that you copied in step 3.

When the installation finishes on all scanner nodes and they have been joined to the manager node, the manager node installation finishes as well.

Result

The BlueXP classification installer finishes installing packages, and registers the installation. Installation can take 15 to 25 minutes.

What's Next

From the Configuration page you can select the local [on-prem ONTAP clusters](#) and local [databases](#) that you want to scan.

Add scanner nodes to an existing deployment

You can add scanner nodes to an existing deployment on a Linux host with internet access.

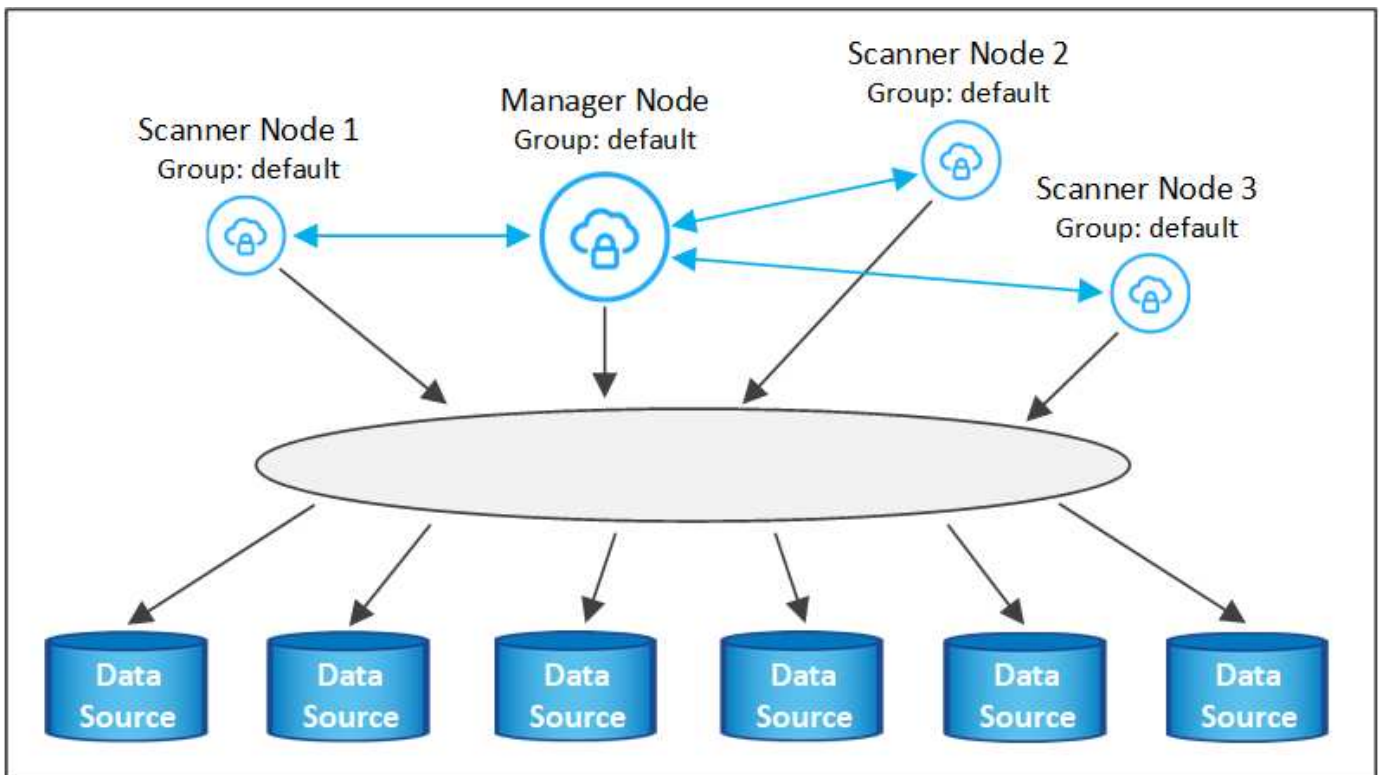
You can add more scanner nodes if you find that you need more scanning processing power to scan your data sources. You can add the scanner nodes immediately after installing the manager node, or you can add a scanner node later. For example, if you realize that the amount of data in one of your data sources has doubled or tripled in size after 6 months, you can add a new scanner node to assist with data scanning.

NOTE This information is relevant only for BlueXP classification legacy versions 1.30 and earlier.

There are two ways in which you can add additional scanner nodes:

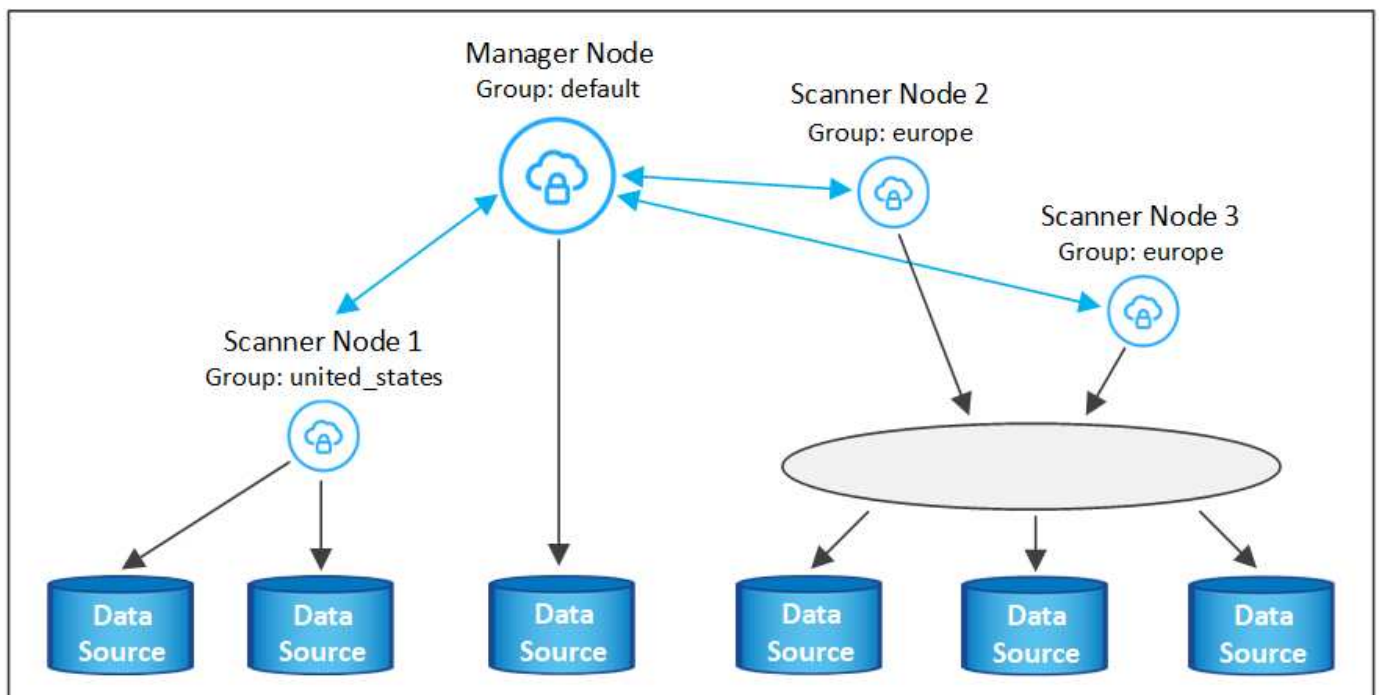
- add a node to assist with scanning all data sources
- add a node to assist with scanning a specific data source, or a specific group of data sources (typically based on location)

By default, any new scanner nodes you add are added to the general pool of scanning resources. This is called the "default scanner group". In the image below, there is 1 Manager node and 3 Scanner nodes in the "default" group that are all scanning data from all 6 data sources.



If you have certain data sources that you want to be scanned by scanner nodes that are physically closer to the data sources, you can define a scanner node, or group of scanner nodes, to scan a specific data source, or group of data sources. In the image below, there is 1 Manager node and 3 Scanner nodes.

- The Manager node is in the "default" group, and it is scanning 1 data source
- Scanner node 1 is in the "united_states" group, and it is scanning 2 data sources
- Scanner nodes 2 and 3 are in the "europe" group, and they share the scanning tasks for 3 data sources



BlueXP classification scanner groups can be defined as separate geographic areas where your data is stored.

You can deploy multiple BlueXP classification scanner nodes around the world and choose a scanner group for each node. In that way, each scanner node will scan the data that is the closest to it. The closer the scanner node is to the data, the better, because it reduces network latency as much as possible while scanning data.

You can choose which scanner groups to add to BlueXP classification and you can choose their names. BlueXP classification does not enforce that a node mapped to a scanner group named "europe" will be deployed in Europe.

You'll follow these steps to install additional BlueXP classification scanner nodes:

1. Prepare the Linux host systems that will act as the Scanner nodes
2. Download the Data Sense software to these Linux systems
3. Run a command on the Manager node to identify the Scanner nodes
4. Follow the steps to deploy the software on the Scanner nodes (and to optionally define a "scanner group" for certain Scanner nodes)
5. If you defined a scanner group, on the Manager node:
 - a. Open the file "working_environment_to_scanner_group_config.yml" and define the working environments that will be scanned by each scanner group
 - b. Run the following script to register this mapping information with all Scanner nodes:
`update_we_scanner_group_from_config_file.sh`

What you'll need

- Verify that all your Linux systems for Scanner nodes meet the host requirements.
- Verify that the systems have the two prerequisite software packages installed (Docker Engine or Podman, and Python 3).
- Make sure you have root privileges on the Linux systems.
- Verify that your environment meets the required permissions and connectivity.
- You must have the IP addresses of the Scanner node hosts that you are adding.
- You must have the IP address of the BlueXP classification Manager node host system
- You must have the IP address or host name of the Connector system, your NetApp Account ID, Connector Client ID, and user access token. If you're planning to use scanner groups, you'll need to know the Working Environment ID for each data source in your account. See **Prerequisite steps** below to get this information.
- The following ports and protocols must be enabled on all hosts:

Port	Protocols	Description
2377	TCP	Cluster management communications
7946	TCP, UDP	Inter-node communication
4789	UDP	Overlay network traffic
50	ESP	Encrypted IPsec overlay network (ESP) traffic
111	TCP, UDP	NFS Server for sharing files between the hosts (needed from each scanner node to manager node)
2049	TCP, UDP	NFS Server for sharing files between the hosts (needed from each scanner node to manager node)

- If you are using `firewalld` on your BlueXP classification machines, we recommend that you enable it before installing BlueXP classification. Run the following commands to configure `firewalld` so that it is compatible with BlueXP classification:

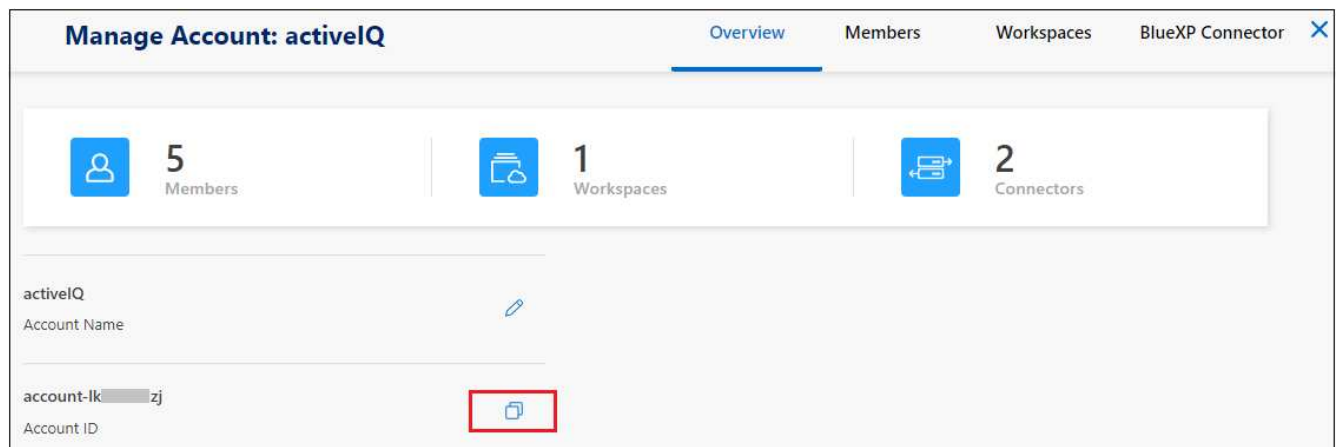
```
firewall-cmd --permanent --add-service=http
firewall-cmd --permanent --add-service=https
firewall-cmd --permanent --add-port=80/tcp
firewall-cmd --permanent --add-port=8080/tcp
firewall-cmd --permanent --add-port=443/tcp
firewall-cmd --permanent --add-port=2377/tcp
firewall-cmd --permanent --add-port=7946/udp
firewall-cmd --permanent --add-port=7946/tcp
firewall-cmd --permanent --add-port=4789/udp
firewall-cmd --reload
```

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

Prerequisite steps

Follow these steps to get the NetApp Account ID, Connector Client ID, Connector Server Name, and user access token that are required to add scanner nodes.

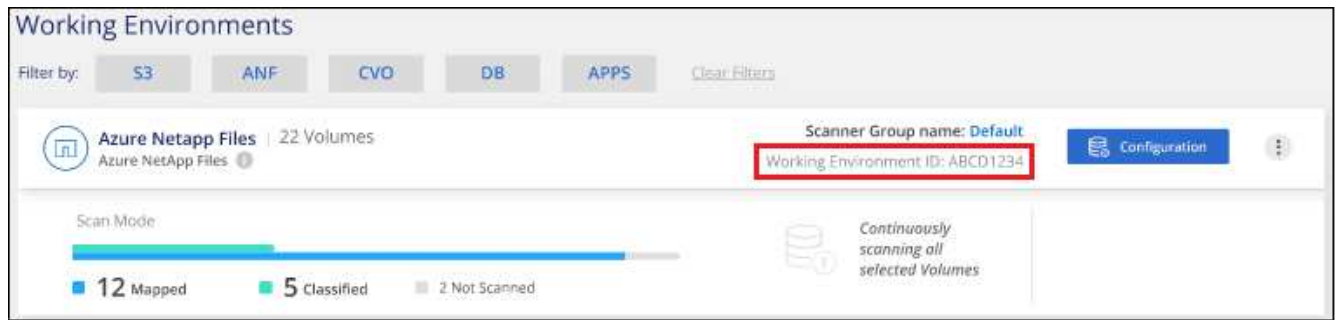
1. From the BlueXP menu bar, click **Account > Manage Accounts**.



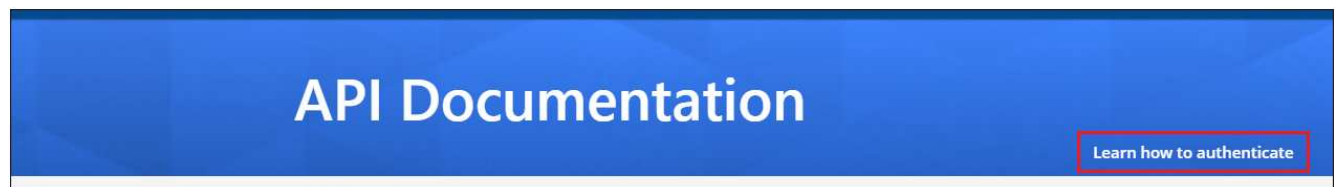
2. Copy the *Account ID*.
3. From the BlueXP menu bar, click **Help > Support > BlueXP Connector**.



4. Copy the connector *Client ID* and the *Server Name*.
5. If you're planning to use scanner groups, from the BlueXP classification Configuration tab, copy the Working Environment ID for each working environment that you plan to add to a scanner group.



6. Go to the [API Documentation Developer Hub](#) and click **Learn how to authenticate**.



7. Follow the authentication instructions, using the username and password of the account admin in the "username" and "password" parameters.
8. Then copy the *access token* from the response.

Steps

1. On the BlueXP classification Manager node, run the script "add_scanner_node.sh". For example, this command adds 2 scanner nodes:

```
sudo ./add_scanner_node.sh -a <account_id> -c <client_id> -m <cm_host> -h
<ds_manager_ip> -n <node_private_ip_1,node_private_ip_2> -t <user_token>
```

Variable values:

- *account_id* = NetApp Account ID
 - *client_id* = Connector Client ID (add the suffix "clients" to the client ID that you copied in the Prerequisite steps)
 - *cm_host* = IP address or host name of the Connector system
 - *ds_manager_ip* = Private IP address of the BlueXP classification Manager node system
 - *node_private_ip* = IP addresses of the BlueXP classification Scanner node systems (multiple scanner node IPs are separated by a comma)
 - *user_token* = JWT user access token
2. Before the add_scanner_node script completes, a dialog displays the installation command needed for the scanner nodes. Copy the command (for example: `sudo ./node_install.sh -m 10.11.12.13 -t ABCDEF1s35212 -u red95467j`) and save it in a text file.
 3. On **each** scanner node host:
 - a. Copy the Data Sense installer file (**DATASENSE-INSTALLER-<version>.tar.gz**) to the host machine (using `scp` or some other method).

- b. Unzip the installer file.
- c. Paste and execute the command that you copied in step 2.
- d. If you want to add a scanner node into a "scanner group", add the parameter **-r <scanner_group_name>** to the command. Otherwise, the scanner node is added to the "default" group.

When the installation finishes on all scanner nodes and they have been joined to the manager node, the "add_scanner_node.sh" script finishes as well. The installation can take 10 to 20 minutes.

4. If you added any scanner nodes into a scanner group, return to the Manager node and perform the following 2 tasks:

- a. Open the file `/opt/netapp/config/custom_configuration/working_environment_to_scanner_group_config.yml` and enter the mapping for which scanner groups will scan specific working environments. You'll need to have the *Working Environment ID* for each data source. For example, the following entries add 2 working environments to the "europe" scanner group and 2 to the "united_states" scanner group:

```
scanner_groups:
  europe:
    working_environments:
      - "working_environment_id1"
      - "working_environment_id2"
  united_states:
    working_environments:
      - "working_environment_id3"
      - "working_environment_id4"
```

Any working environment that is not added to the list is scanned by the "default" group - you must have at least one manager or scanner node in the "default" group.

- b. Run the following script to register this mapping information with all Scanner nodes:
`/opt/netapp/Datasense/tools/update_we_scanner_group_from_config_file.sh`

Result

BlueXP classification is set up with Manager and Scanner nodes to scan all your data sources.

What's Next

From the Configuration page you can select the data sources that you want to scan - if you haven't already done that. If you created scanner groups, each data source is scanned by the Scanner nodes in the respective group.

You can see the Scanner Group name for each working environment in the Configuration page.

Working Environments

Filter by: S3 ANF CVO DB APPS [Clear Filters](#)

Azure Netapp Files | 22 Volumes
 Azure NetApp Files ⓘ

Scanner Group name: **Default**
 Working Environment ID: **ABCD1234** [Configuration](#)

Scan Mode

12 Mapped 5 Classified 2 Not Scanned

Continuously scanning all selected Volumes

You can also see the list of all scanner groups along with the IP address and status for each scanner node in the group in the bottom of the Configuration page.

Scanner Groups

Scanner Group: **Default** ✓ Scanner nodes

2 Scanner nodes

Scanner node host name	IP	Last active time	Status	Error
ip-172-...us-west-2.compute	172-...	23/09/2022 14:32	Active	
ip-172-...us-west-2.compute	172-...	23/09/2022 14:32	Active	

Scanner Group: **United_States** ✓ Scanner nodes

2 Scanner nodes

Scanner node host name	IP	Last active time	Status	Error
ip-172-...us-west-2.compute	172-...	23/09/2022 14:32	Active	
ip-172-...us-west-2.compute	172-...	23/09/2022 14:32	Active	

Scanner Group: **Europe** ✓ Scanner nodes

Scan data deprecations

Scan Amazon S3 buckets

BlueXP classification can scan your Amazon S3 buckets to identify the personal and sensitive data that resides in S3 object storage. BlueXP classification can scan any bucket in the account, regardless if it was created for a NetApp solution.

NOTE This information is relevant only for BlueXP classification legacy versions 1.30 and earlier.

Quick start

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

1

Set up the S3 requirements in your cloud environment

Ensure that your cloud environment can meet the requirements for BlueXP classification, including preparing an IAM role and setting up connectivity from BlueXP classification to S3. [See the complete list.](#)

2

Deploy the BlueXP classification instance

[Deploy BlueXP classification](#) if there isn't already an instance deployed.

3

Activate BlueXP classification on your S3 working environment

Select the Amazon S3 working environment, click **Enable**, and select an IAM role that includes the required permissions.

4

Select the buckets to scan

Select the buckets that you'd like to scan and BlueXP classification will start scanning them.

Reviewing S3 prerequisites

The following requirements are specific to scanning S3 buckets.

Set up an IAM role for the BlueXP classification instance

BlueXP classification needs permissions to connect to the S3 buckets in your account and to scan them. Set up an IAM role that includes the permissions listed below. BlueXP prompts you to select an IAM role when you enable BlueXP classification on the Amazon S3 working environment.

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "s3:Get*",
        "s3:List*",
        "s3:PutObject"
      ],
      "Resource": "*"
    },
    {
      "Effect": "Allow",
      "Action": [
        "iam:GetPolicyVersion",
        "iam:GetPolicy",
        "iam:ListAttachedRolePolicies"
      ],
      "Resource": [
        "arn:aws:iam::*:policy/*",
        "arn:aws:iam::*:role/*"
      ]
    }
  ]
}

```

Provide connectivity from BlueXP classification to Amazon S3

BlueXP classification needs a connection to Amazon S3. The best way to provide that connection is through a VPC Endpoint to the S3 service. For instructions, see [AWS Documentation: Creating a Gateway Endpoint](#).

When you create the VPC Endpoint, be sure to select the region, VPC, and route table that corresponds to the BlueXP classification instance. You must also modify the security group to add an outbound HTTPS rule that enables traffic to the S3 endpoint. Otherwise, BlueXP classification can't connect to the S3 service.

If you experience any issues, see [AWS Support Knowledge Center: Why can't I connect to an S3 bucket using a gateway VPC endpoint?](#)

An alternative is to provide the connection by using a NAT Gateway.



You can't use a proxy to get to S3 over the internet.

Deploying the BlueXP classification instance

[Deploy BlueXP classification in BlueXP](#) if there isn't already an instance deployed.

You need to deploy the instance using a Connector deployed in AWS so that BlueXP automatically discovers

the S3 buckets in this AWS account and displays them in an Amazon S3 working environment.

Note: Deploying BlueXP classification in an on-premises location is not currently supported when scanning S3 buckets.

Upgrades to BlueXP classification software are automated as long as the instance has internet connectivity.

Activating BlueXP classification on your S3 working environment

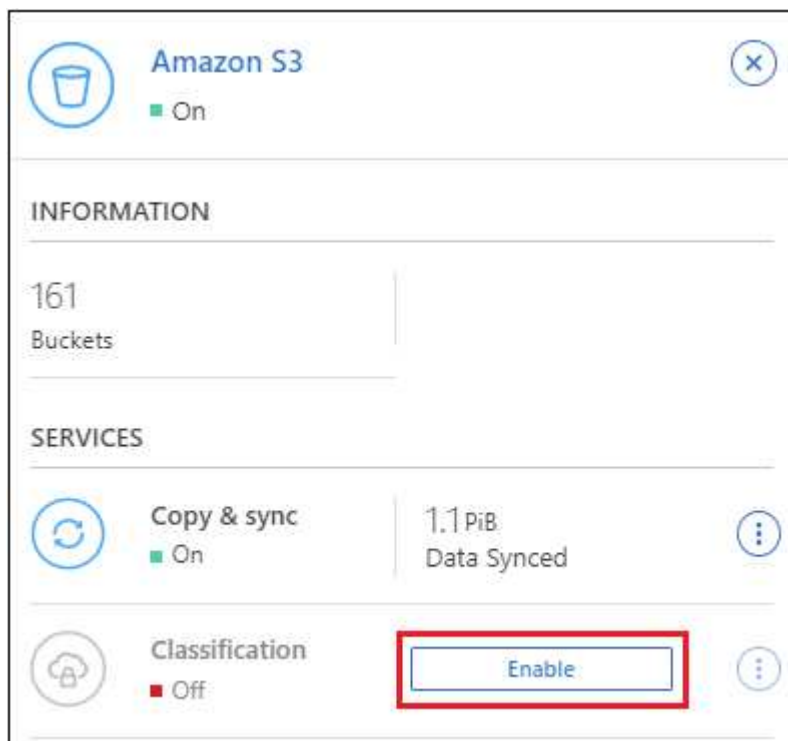
Enable BlueXP classification on Amazon S3 after you verify the prerequisites.

Steps

1. From the BlueXP left navigation menu, click **Storage > Canvas**.
2. Select the Amazon S3 working environment.



3. In the Services pane on the right, click **Enable** next to **Classification**.



4. When prompted, assign an IAM role to the BlueXP classification instance that has [the required permissions](#).

Assign an AWS IAM Role for Data Sense & Compliance

To enable Data Sense & Compliance on Amazon S3 buckets, select an existing IAM Role. Make sure that your AWS IAM Role has the permission defined in the [Policy Requirements](#).

Select IAM Role

Select a Role

▼

VPC Endpoint for Amazon S3 Required

A VPC endpoint to the Amazon S3 service is required so Data Sense & Compliance can securely scan the data.

Alternatively, ensure that the Data Sense & Compliance instance has direct access to the internet via a NAT Gateway or Internet Gateway.

Free for the 1st TB


Over 1 TB you pay only for what you use. [Learn more about pricing.](#)

Enable

Cancel

5. Click **Enable**.



You can also enable compliance scans for a working environment from the Configuration page by clicking the  button and selecting **Activate BlueXP classification**.

Result

BlueXP assigns the IAM role to the instance.

Enabling and disabling compliance scans on S3 buckets

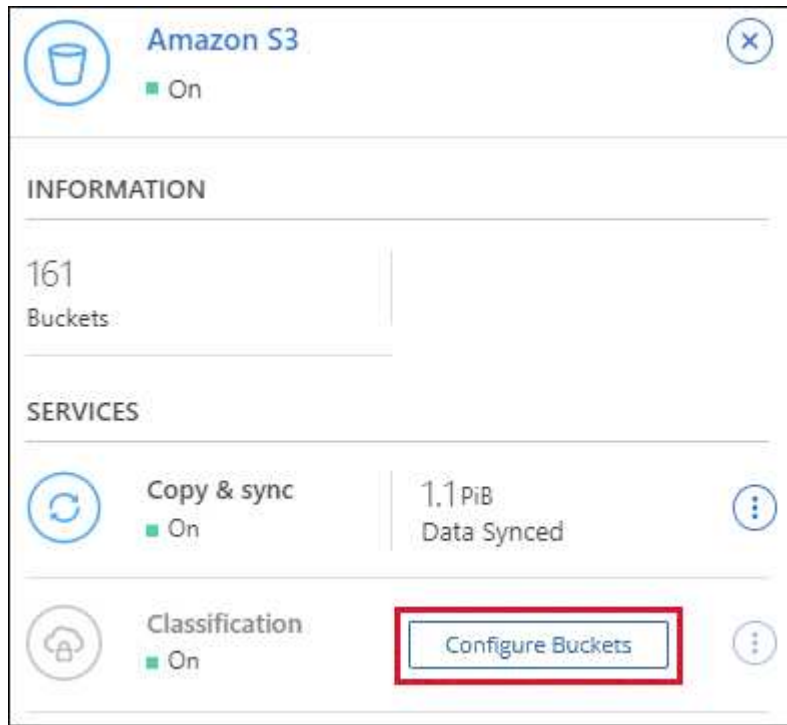
After BlueXP enables BlueXP classification on Amazon S3, the next step is to configure the buckets that you want to scan.

When BlueXP is running in the AWS account that has the S3 buckets you want to scan, it discovers those buckets and displays them in an Amazon S3 working environment.

BlueXP classification can also [scan S3 buckets that are in different AWS accounts](#).

Steps

1. Select the Amazon S3 working environment.
2. In the Services pane on the right, click **Configure Buckets**.



3. Enable mapping-only scans, or mapping and classification scans, on your buckets.

Amazon S3 Configuration 🔍			
15/28 Buckets in Scan Scope.			
Scan	Bucket Name	Status	Required Action
Off Map Map & Classify	BucketName1	● Not Scanning	Add Credentials
Off Map Map & Classify	BucketName2	● Continuosly Scanning	
Off Map Map & Classify	BucketName3	● Not Scanning	

To:	Do this:
Enable mapping-only scans on a bucket	Click Map
Enable full scans on a bucket	Click Map & Classify
Disable scanning on a bucket	Click Off

Result

BlueXP classification starts scanning the S3 buckets that you enabled. If there are any errors, they'll appear in the Status column, alongside the required action to fix the error.

Scanning buckets from additional AWS accounts

You can scan S3 buckets that are under a different AWS account by assigning a role from that account to access the existing BlueXP classification instance.





Steps

1. Go to the target AWS account where you want to scan S3 buckets and create an IAM role by selecting **Another AWS account**.

Create role




Select type of trusted entity

 AWS service EC2, Lambda and others	 Another AWS account Belonging to you or 3rd party	 Web identity Cognito or any OpenID provider	 SAML 2.0 federation Your corporate directory
--	---	---	--

Allows entities in other accounts to perform actions in this account. [Learn more](#)

Specify accounts that can use this role

Account ID*

- Options
- ☐ Require external ID (Best practice when a third party will assume this role)
 - ☐ Require MFA 

Be sure to do the following:

- Enter the ID of the account where the BlueXP classification instance resides.
- Change the **Maximum CLI/API session duration** from 1 hour to 12 hours and save that change.
- Attach the BlueXP classification IAM policy. Make sure it has the required permissions.

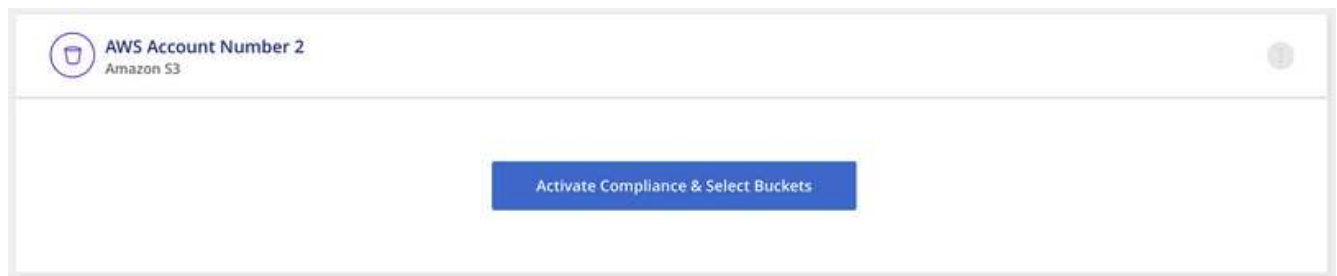
```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "s3:Get*",
        "s3:List*",
        "s3:PutObject"
      ],
      "Resource": "*"
    }
  ]
}
```

2. Go to the source AWS account where the BlueXP classification instance resides and select the IAM role that is attached to the instance.
 - a. Change the **Maximum CLI/API session duration** from 1 hour to 12 hours and save that change.
 - b. Click **Attach policies** and then click **Create policy**.
 - c. Create a policy that includes the "sts:AssumeRole" action and specify the ARN of the role that you created in the target account.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "sts:AssumeRole",
      "Resource": "arn:aws:iam::<ADDITIONAL-ACCOUNT-ID>:role/<ADDITIONAL_ROLE_NAME>"
    },
    {
      "Effect": "Allow",
      "Action": [
        "iam:GetPolicyVersion",
        "iam:GetPolicy",
        "iam:ListAttachedRolePolicies"
      ],
      "Resource": [
        "arn:aws:iam::*:policy/*",
        "arn:aws:iam::*:role/*"
      ]
    }
  ]
}
```

The BlueXP classification instance profile account now has access to the additional AWS account.

3. Go to the **Amazon S3 Configuration** page and the new AWS account is displayed. Note that it can take a few minutes for BlueXP classification to sync the new account's working environment and show this information.



4. Click **Activate BlueXP classification & Select Buckets** and select the buckets you want to scan.

Result

BlueXP classification starts scanning the new S3 buckets that you enabled.

Scan OneDrive accounts

Complete a few steps to start scanning files in your user's OneDrive folders with BlueXP classification.

NOTE This information is relevant only for BlueXP classification legacy versions 1.30 and earlier.

Quick start

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

1

Review OneDrive prerequisites

Ensure that you have the Admin credentials to log into the OneDrive account.

2

Deploy the BlueXP classification instance

[Deploy BlueXP classification](#) if there isn't already an instance deployed.

3

Add the OneDrive account

Using Admin user credentials, log into the OneDrive account that you want to access so that it is added as a new working environment.

4

Add the users and select the type of scanning

Add the list of users from the OneDrive account that you want to scan and select the type of scanning. You can add up to 100 users at time.

Reviewing OneDrive requirements

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

- You must have the Admin login credentials for the OneDrive for Business account that provides read access to the user's files.
- You'll need a line-separated list of the email addresses for all the users whose OneDrive folders you want to scan.

Deploying the BlueXP classification instance

Deploy BlueXP classification if there isn't already an instance deployed.

BlueXP classification can be [deployed in the cloud](#) or [in an on-premises location that has internet access](#).

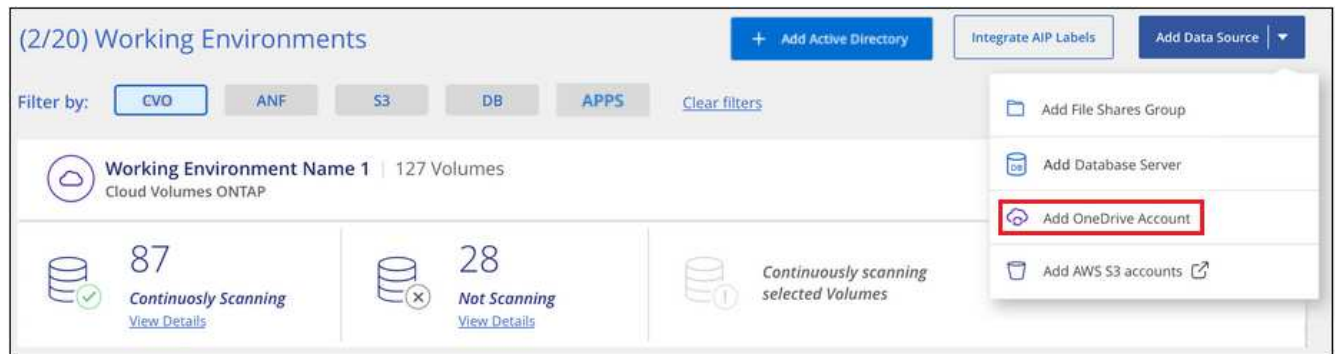
Upgrades to BlueXP classification software is automated as long as the instance has internet connectivity.

Adding the OneDrive account

Add the OneDrive account where the user files reside.

Steps

1. From the Working Environments Configuration page, click **Add Data Source > Add OneDrive Account**.



2. In the Add a OneDrive account dialog, click **Sign in to OneDrive**.
3. In the Microsoft page that appears, select the OneDrive account and enter the required Admin user and password, then click **Accept** to allow BlueXP classification to read data from this account.

The OneDrive account is added to the list of working environments.

Adding OneDrive users to compliance scans

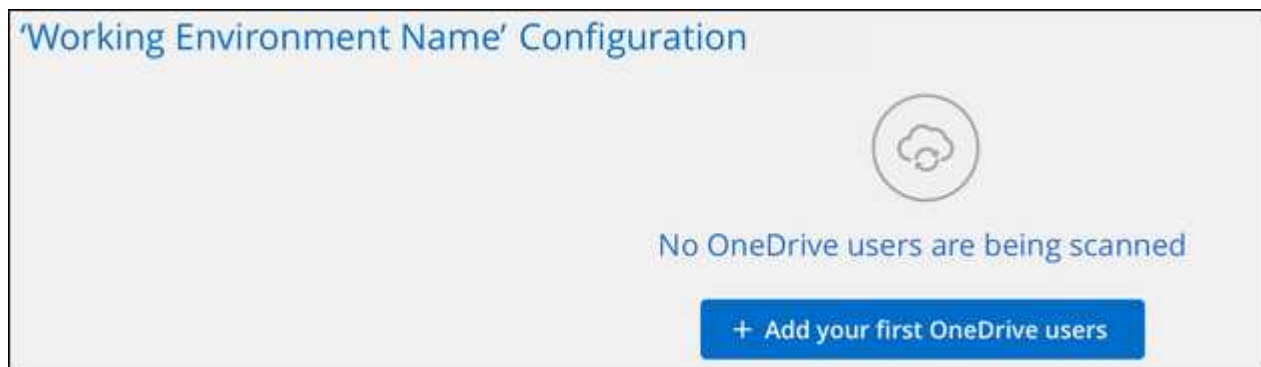
You can add individual OneDrive users, or all of your OneDrive users, so that their files will be scanned by BlueXP classification.

Steps

1. From the *Configuration* page, click the **Configuration** button for the OneDrive account.



2. If this is the first time adding users for this OneDrive account, click **Add your first OneDrive users**.



If you are adding additional users from a OneDrive account, click **Add OneDrive users**.

Working Environment 4 Configuration

+ Add OneDrive users

24 users are being scanned for compliance

Scan	Username	Status	Required Action
<div>Off</div> <div>Map</div> <div>Map & Classify</div>	user2@example.com	<div></div> <div>Continuously Scanning</div>	...
<div>Off</div> <div>Map</div> <div>Map & Classify</div>	user3@example.com	<div></div> <div>Continuously Scanning</div>	...

3. Add the email addresses for the users whose files you want to scan - one email address per line (up to 100 maximum per session) - and click **Add Users**.

Add OneDrive users

Provide a list of OneDrive users for Cloud Data Sense to scan their data, line-separated. You can add up to 100 users at a time.

Type or paste below the OneDrive user accounts to add

User Accounts

user@example.com

user@example.com

user@example.com

user@example.com

user@example.com

user@example.com

user@example.com

Add Users

Cancel

A confirmation dialog displays the number of users who were added.

If the dialog lists any users who could not be added, capture this information so that you can resolve the issue. In some cases you can re-add the user with a corrected email address.

4. Enable mapping-only scans, or mapping and classification scans, on user files.

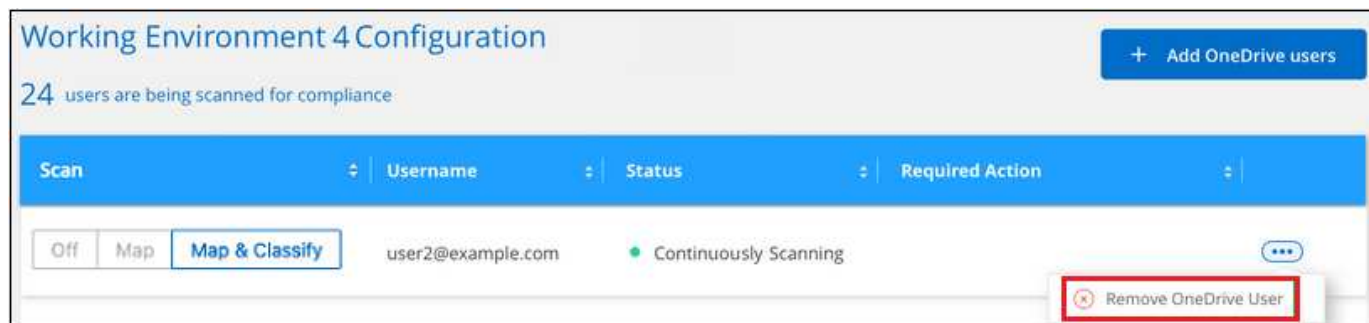
To:	Do this:
Enable mapping-only scans on user files	Click Map
Enable full scans on user files	Click Map & Classify
Disable scanning on user files	Click Off

Result

BlueXP classification starts scanning the files for the users you added, and the results are displayed in the Dashboard and in other locations.

Removing a OneDrive user from compliance scans

If users leave the company or if their email address changes, you can remove individual OneDrive users from having their files scanned at any time. Just click **Remove OneDrive User** from the Configuration page.



Scan SharePoint accounts

Complete a few steps to start scanning files in your SharePoint Online and SharePoint On-Premise accounts with BlueXP classification.

NOTE This information is relevant only for BlueXP classification legacy versions 1.30 and earlier.

Quick start

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

1

Review SharePoint prerequisites

Ensure that you have qualified credentials to log into the SharePoint account, and that you have the URLs for the SharePoint sites that you want to scan.

2

Deploy the BlueXP classification instance

[Deploy BlueXP classification](#) if there isn't already an instance deployed.

3

Log into the SharePoint account

Using qualified user credentials, log into the SharePoint account that you want to access so that it is added as a new data source/working environment.

4

Add the SharePoint site URLs to scan

Add the list of SharePoint site URLs that you want to scan in the SharePoint account, and select the type of scanning. You can add up to 100 URLs at time - and up to 1,000 sites total for each account.

Review SharePoint requirements

Review the following prerequisites to make sure you are ready to activate BlueXP classification on a SharePoint account.

- You must have the Admin user login credentials for the SharePoint account that provides read access to all SharePoint sites.
 - For SharePoint Online you can use a non-Admin account, but that user must have permission to access all the SharePoint sites that you want to scan.
- For SharePoint On-Premise, you'll also need the URL of the SharePoint Server.
- You will need a line-separated list of the SharePoint site URLs for all the data you want to scan.

Deploy the BlueXP classification instance

Deploy BlueXP classification if there isn't already an instance deployed.

- For SharePoint Online, BlueXP classification can be [deployed in the cloud](#).
- For SharePoint On-Premises, BlueXP classification can be installed [in an on-premises location that has internet access](#) or [in an on-premises location that does not have internet access](#).

When BlueXP classification is installed in a site without internet access, the BlueXP Connector also must be installed in that same site without internet access. [Learn more](#).

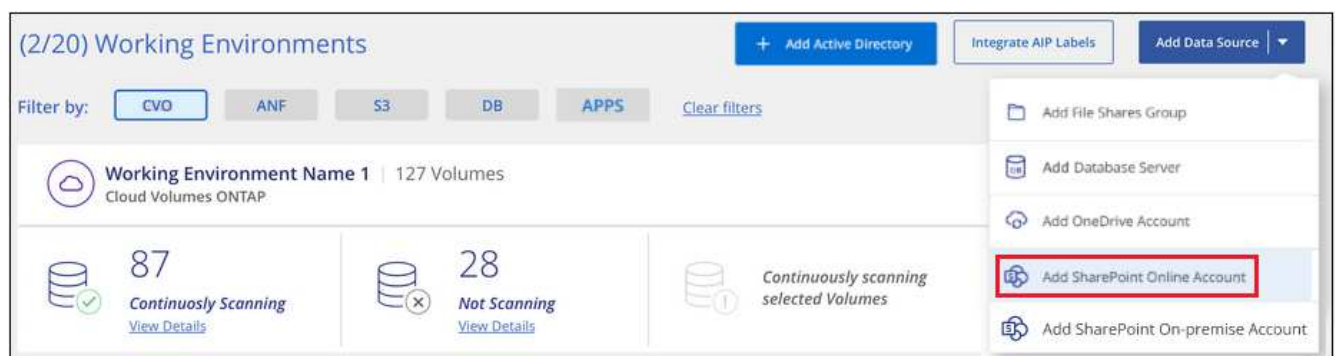
Upgrades to BlueXP classification software is automated as long as the instance has internet connectivity.

Add a SharePoint Online account

Add the SharePoint Online account where the user files reside.

Steps

1. From the Working Environments Configuration page, click **Add Data Source > Add SharePoint Online Account**.



2. In the Add a SharePoint Online Account dialog, click **Sign in to SharePoint**.
3. In the Microsoft page that appears, select the SharePoint account and enter the user and password (Admin user or other user with access to the SharePoint sites), then click **Accept** to allow BlueXP classification to read data from this account.

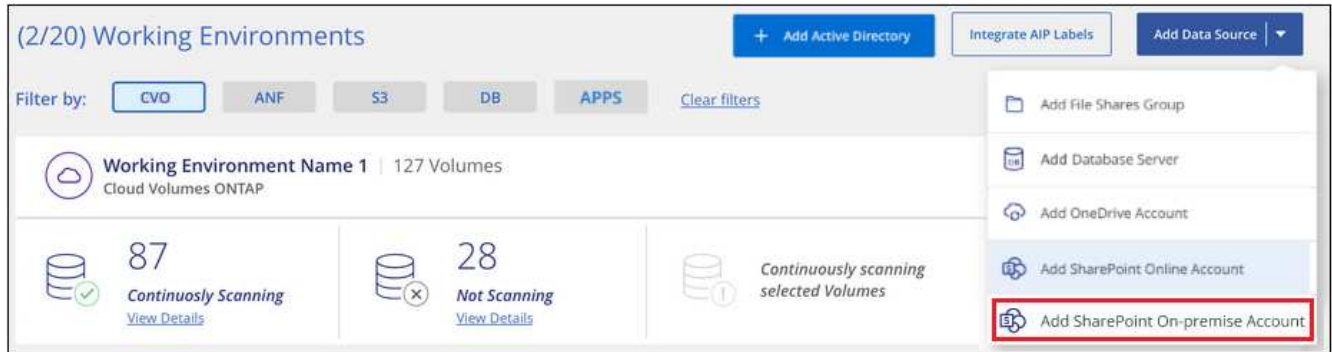
The SharePoint Online account is added to the list of working environments.

Add a SharePoint On-premise account

Add the SharePoint On-premise account where the user files reside.

Steps

1. From the Working Environments Configuration page, click **Add Data Source > Add SharePoint On-premise Account**.



2. In the Log into the SharePoint On-Premise Server dialog, enter the following information:
 - Admin user in the format "domain/user" or "user@domain", and admin password
 - URL of the SharePoint Server

The screenshot shows a dialog box titled 'Log into the SharePoint On-Premises Server'. It contains the instruction: 'To activate Data Sense on your SharePoint business account, sign in to SharePoint with an Admin user.' Below this, there are three input fields: 'Username' (with placeholder text 'domain/user or user@domain'), 'Password' (with placeholder text 'Password'), and 'URL' (with placeholder text 'http://10.0.0.1'). At the bottom right, there are two buttons: 'Connect' and 'Cancel'.

3. Click **Connect**.

The SharePoint On-premise account is added to the list of working environments.

Add SharePoint sites to compliance scans

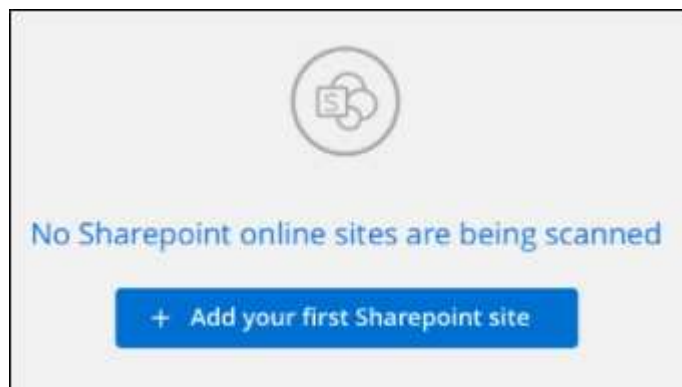
You can add individual SharePoint sites, or up to 1,000 SharePoint sites in the account, so that the associated files will be scanned by BlueXP classification. The steps are the same whether you are adding SharePoint Online or SharePoint On-premise sites.

Steps

1. From the *Configuration* page, click the **Configuration** button for the SharePoint account.



2. If this is the first time adding sites for this SharePoint account, click **Add your first SharePoint site**.



If you are adding additional users from a SharePoint account, click **Add SharePoint Sites**.



3. Add the URLs for the sites whose files you want to scan - one URL per line (up to 100 maximum per session) - and click **Add Sites**.

A confirmation dialog displays the number of sites that were added.

If the dialog lists any sites that could not be added, capture this information so that you can resolve the issue. In some cases you can re-add the site with a corrected URL.

4. If you need to add more than 100 sites for this account, just click **Add SharePoint Sites** again until you have added all your sites for this account (up to 1,000 sites total for each account).
5. Enable mapping-only scans, or mapping and classification scans, on the files in the SharePoint sites.

To:	Do this:
Enable mapping-only scans on files	Click Map
Enable full scans on files	Click Map & Classify
Disable scanning on files	Click Off

Result

BlueXP classification starts scanning the files in the SharePoint sites you added, and the results are displayed in the Dashboard and in other locations.

Remove a SharePoint site from compliance scans

If you remove a SharePoint site in the future, or decide not to scan files in a SharePoint site, you can remove individual SharePoint sites from having their files scanned at any time. Just click **Remove SharePoint Site** from the Configuration page.

Scan	Site URL	Status	Required Action
Off Map Map & Classify	Site URL	Continuously Scanning	...
Off Map Map & Classify	Site URL	Continuously Scanning	Remove SharePoint Site

Note that you can [delete the entire SharePoint account from BlueXP classification](#) if you no longer want to scan any user data from the SharePoint account.

Scan Google Drive accounts

Complete a few steps to start scanning user files in your Google Drive accounts with BlueXP classification.

NOTE This information is relevant only for BlueXP classification legacy versions 1.30 and earlier.

Quick start

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

1

Review Google Drive prerequisites

Ensure that you have the Admin credentials to log into the Google Drive account.

2

Deploy BlueXP classification

[Deploy BlueXP classification](#) if there isn't already an instance deployed.

3

Log into the Google Drive account

Using Admin user credentials, log into the Google Drive account that you want to access so that it is added as a new data source.

4

Select the type of scanning for the user files

Select the type of scanning you want to perform on the user files; mapping or mapping and classifying.

Review Google Drive requirements

Review the following prerequisites to make sure you are ready to enable BlueXP classification on a Google Drive account.

- You must have the Admin login credentials for the Google Drive account that provides read access to the user's files

Current restrictions

The following BlueXP classification features are not currently supported with Google Drive files:

- When viewing files in the Data Investigation page, the actions in the button bar aren't active. You can't copy, move, delete, etc. any files.
- Permissions can't be identified within files in Google Drive, so no permission information is displayed in the Investigation page.

Deploy BlueXP classification

Deploy BlueXP classification if there isn't already an instance deployed.

BlueXP classification can be [deployed in the cloud](#) or [in an on-premises location that has internet access](#).

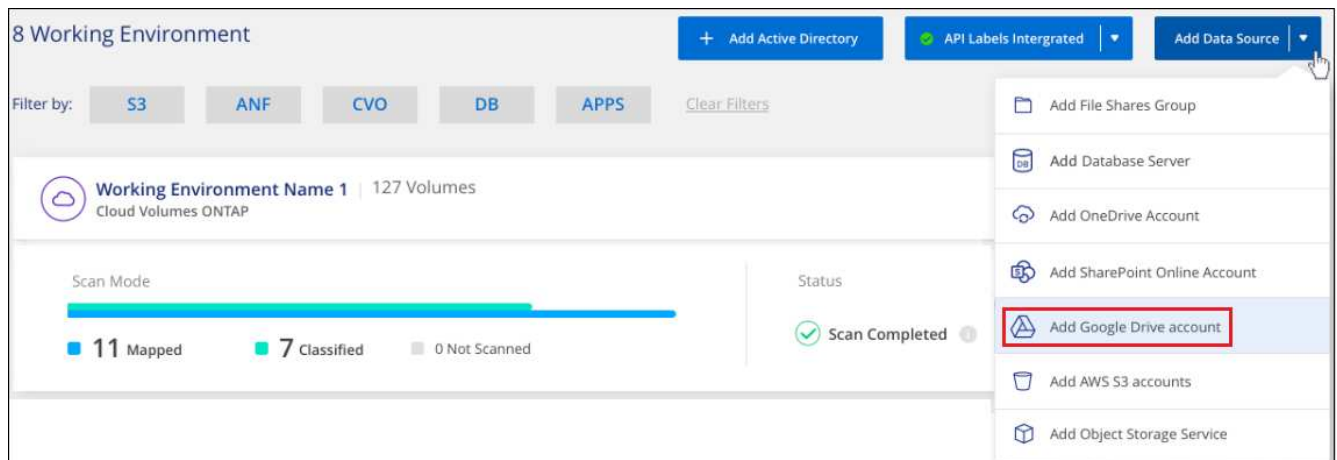
Upgrades to BlueXP classification software is automated as long as the instance has internet connectivity.

Add the Google Drive account

Add the Google Drive account where the user files reside. If you want to scan files from multiple users, you'll need to run through this step for each user.

Steps

1. From the Working Environments Configuration page, click **Add Data Source > Add Google Drive Account**.



2. In the Add a Google Drive Account dialog, click **Sign in to Google Drive**.
3. In the Google page that appears, select the Google Drive account and enter the required Admin user and password, then click **Accept** to allow BlueXP classification to read data from this account.

The Google Drive account is added to the list of working environments.

Select the type of scanning for user data

Select the type of scanning that BlueXP classification will perform on the user's data.

Steps

1. From the *Configuration* page, click the **Configuration** button for the Google Drive account.



2. Enable mapping-only scans, or mapping and classification scans, on the files in the Google Drive account.



To:	Do this:
Enable mapping-only scans on files	Click Map
Enable full scans on files	Click Map & Classify
Disable scanning on files	Click Off

Result

BlueXP classification starts scanning the files in the Google Drive account you added, and the results are displayed in the Dashboard and in other locations.

Remove a Google Drive account from compliance scans

Since only a single user's Google Drive files are part of a single Google Drive account, if you want to stop scanning files from a user's Google Drive account, then you should [delete the Google Drive account from BlueXP classification](#).

Scan StorageGRID data

Complete a few steps to start scanning data within object storage directly with BlueXP classification. BlueXP classification can scan data from any Object Storage service which uses the Simple Storage Service (S3) protocol. This includes NetApp StorageGRID, IBM Cloud Object Store, Linode, B2 Cloud Storage, Amazon S3, and more.

NOTE Using the BlueXP classification that is part of the core BlueXP, you can now scan StorageGRID data. see [Scan StorageGRID data](#). The remaining information here is relevant only for BlueXP classification legacy versions 1.30 and earlier.

Quick start

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



Review object storage prerequisites

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 the object storage provider so that BlueXP classification can access the buckets.

2

Deploy the BlueXP classification instance

[Deploy BlueXP classification](#) if there isn't already an instance deployed.

3

Add the Object Storage Service

Add the object storage service to BlueXP classification.

4

Select the buckets to scan

Select the buckets that you'd like to scan and BlueXP classification will start scanning them.

Reviewing object storage requirements

Review the following prerequisites to make sure that you have a supported configuration before you enable BlueXP 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 the object storage provider so that BlueXP classification can access the buckets.

Deploying the BlueXP classification instance

Deploy BlueXP classification if there isn't already an instance deployed.

If you are scanning data from S3 object storage that is accessible over the internet, you can [deploy BlueXP classification in the cloud](#) or [deploy BlueXP classification in an on-premises location that has internet access](#).

If you are scanning data from S3 object storage that has been installed in a dark site that has no internet access, you need to [deploy BlueXP classification in the same on-premises location that has no internet access](#). This also requires that the BlueXP Connector is deployed in that same on-premises location.

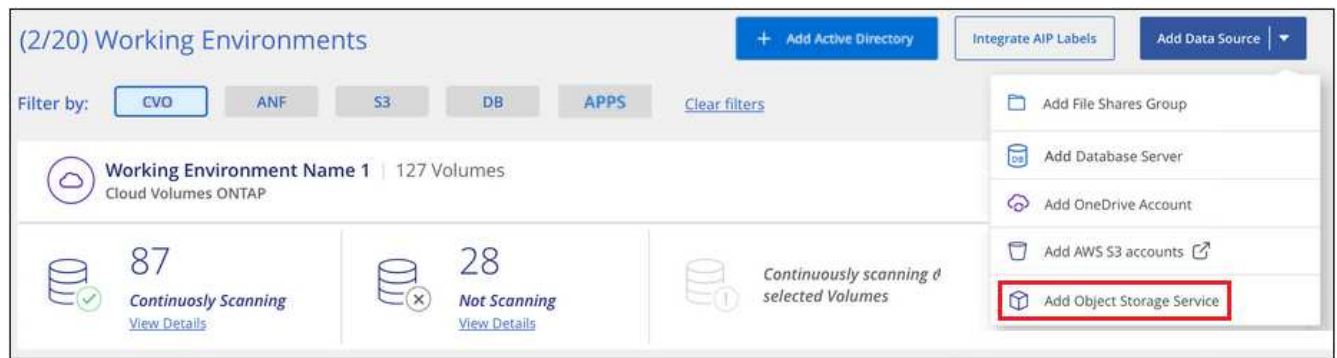
Upgrades to BlueXP classification software is automated as long as the instance has internet connectivity.

Adding the object storage service to BlueXP classification

Add the object storage service.

Steps

1. From the Working Environments Configuration page, click **Add Data Source > Add Object Storage Service**.



2. In the Add Object Storage Service dialog, enter the details for the object storage service and click **Continue**.
 - a. Enter the name you want to use for the Working Environment. This name should reflect the name of the object storage 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 BlueXP classification can access the buckets in the object storage.

Add Object Storage Service

Cloud Data Sense can scan data from any Object Storage service which uses the S3 protocol. This includes NetApp StorageGRID, IBM Object Store, and more.

To continue, enter the following information. In the next steps you'll need to select the buckets you want to scan.

Name the Working Environment	Endpoint URL
<input type="text" value="object_myIBM"/>	<input type="text" value="http://my.endpoint.com"/>
Access Key	Secret Key
<input type="text" value="AJUKDO574NDJG86795"/>	<input type="password" value="....."/>

Result

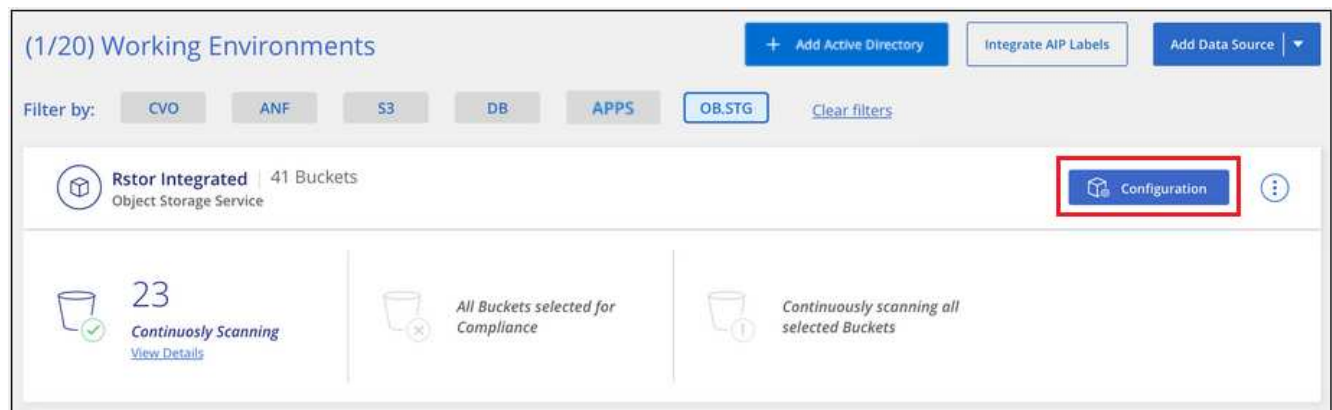
The new Object Storage Service is added to the list of working environments.

Enabling and disabling compliance scans on object storage buckets

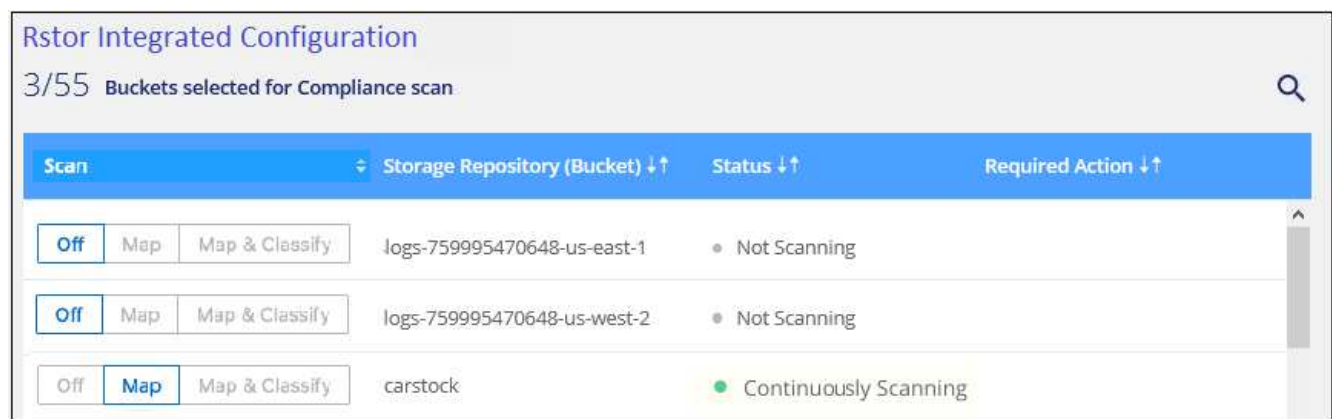
After you enable BlueXP classification on your Object Storage Service, the next step is to configure the buckets that you want to scan. BlueXP classification discovers those buckets and displays them in the working environment you created.

Steps

1. In the Configuration page, click **Configuration** from the Object Storage Service working environment.



2. Enable mapping-only scans, or mapping and classification scans, on your buckets.



To:	Do this:
Enable mapping-only scans on a bucket	Click Map
Enable full scans on a bucket	Click Map & Classify
Disable scanning on a bucket	Click Off

Result

BlueXP classification starts scanning the buckets that you enabled. If there are any errors, they'll appear in the Status column, alongside the required action to fix the error.

Manage data deprecations

View governance details about your data using the Governance dashboard

Gain control of the costs related to the data on your organizations' storage resources. BlueXP classification identifies the amount of stale data, non-business 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.

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.

NOTE This information is relevant only for BlueXP classification legacy versions 1.30 and earlier.

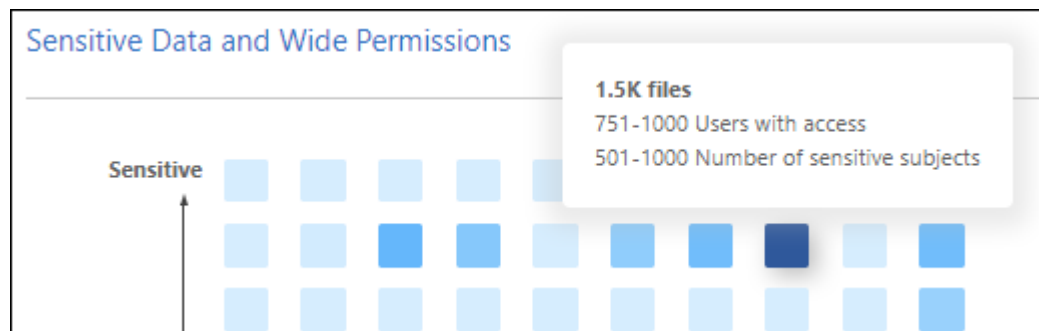
Data listed by sensitivity and wide permissions on the Governance dashboard

The *Sensitive Data and Wide Permissions* area on the Governance dashboard 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.



This applies to BlueXP classification versions 1.30 and earlier.

Files are rated based on the number of users with permission to access the files on the X axis (lowest to highest), and the number of sensitive identifiers within the files on the Y axis (lowest to highest). The blocks represent the number of files that match the items from the X and Y axes. The lighter colored blocks are good; with fewer users able to access the files, and with fewer sensitive identifiers per file. The darker blocks are the items you may want to investigate. For example, the screen below shows the tooltip text for the dark blue block. It shows that you have 1,500 files where 751-1000 users have access, and where there are 501-1000 sensitive identifiers per file.



You can click the block you are interested in to view the filtered results of the affected files in the Investigation page so that you can investigate further.

No data is displayed in this panel if you haven't integrated an identity service with BlueXP classification. [See how to integrate your Active Directory service with BlueXP classification.](#)



This panel supports files in CIFS shares, OneDrive, and SharePoint data sources. There is currently no support for databases, Google Drive, Amazon S3, and generic object storage.

Classification area on the dashboard showing AIP labels

The *Classification* area on the dashboard provides a list of the most identified Azure Information Protection (AIP) Labels in your scanned data.

If you have subscribed to Azure Information Protection (AIP), you can classify and protect documents and files by applying labels to content. Reviewing the most used AIP labels that are assigned to files enables you to see which labels are most used in your files.

See [AIP Labels](#) for more information.

Organize your private data

BlueXP classification provides many ways for you to manage and organize your private data. This makes it easier to see the data that is most important to you.

NOTE This information is relevant only for BlueXP classification legacy versions 1.30 and earlier. The December 2023 (v1.26.6) release removed the option to integrate data using Azure Information Protection (AIP) labels.

- If you are subscribed to [Azure Information Protection \(AIP\)](#) to classify and protect your files, you can use BlueXP classification to manage those AIP labels.
- You can add Tags to files that you want to mark for organization or for some type of follow-up.
- You can assign a BlueXP user to a specific file, or to multiple files, so that person can be responsible for managing the file.
- Using the "Policy" functionality you can create your own custom search queries so that you can easily see the results by clicking one button.
- You can send email alerts to BlueXP users, or any other email address, when certain critical Policies return results.



The capabilities described in this section are available only if you have chosen to perform a full classification scan on your data sources. Data sources that have had a mapping-only scan do not show file-level details.

Should I use tags or labels?

Below is a comparison of BlueXP classification tagging and Azure Information Protection labeling.

Tags	Labels
File tags are an integrated part of BlueXP classification.	Requires that you have subscribed to Azure Information Protection (AIP).
The tag is only kept in the BlueXP classification database - it is not written to the file. It does not change the file, or the file accessed or modified times.	The label is part of the file and when the label changes, the file changes. This change also changes the file accessed and modified times.
You can have multiple tags on a single file.	You can have one label on a single file.
The tag can be used for internal BlueXP classification action, such as copy, move, delete, run a policy, etc.	Other systems that can read the file can see the label - which can be used for additional automation.
Only a single API call is used to see if a file has a tag.	

Categorize your data using AIP labels

You can manage AIP labels in the files that BlueXP classification is scanning if you have subscribed to [Azure Information Protection \(AIP\)](#). AIP enables you to classify and protect documents and files by applying labels to content. BlueXP classification enables you to view the labels that are already assigned to files, add labels to files, and change labels when a label already exists.

BlueXP classification supports AIP labels within the following file types: .DOC, .DOCX, .PDF, .PPTX, .XLS, .XLSX.



- You can't currently change labels in files larger than 30 MB. For OneDrive, SharePoint, and Google Drive accounts the maximum file size is 4 MB.
- If a file has a label which doesn't exist anymore in AIP, BlueXP classification considers it as a file without a label.
- If you've deployed BlueXP classification in a Government region, or in an on-prem location that has no internet access (also known as a dark site), then the AIP label functionality is unavailable.

Integrate AIP labels in your workspace

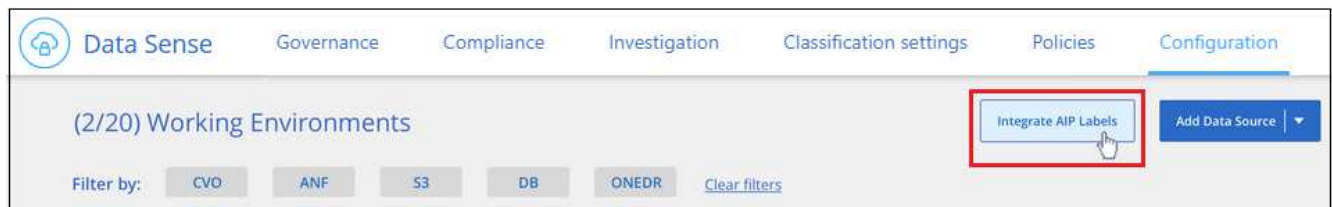
Before you can manage AIP labels, you need to integrate the AIP label functionality into BlueXP classification by signing into your existing Azure account. Once enabled, you can manage AIP labels within files for all [data sources](#) in your BlueXP workspace.

Requirements

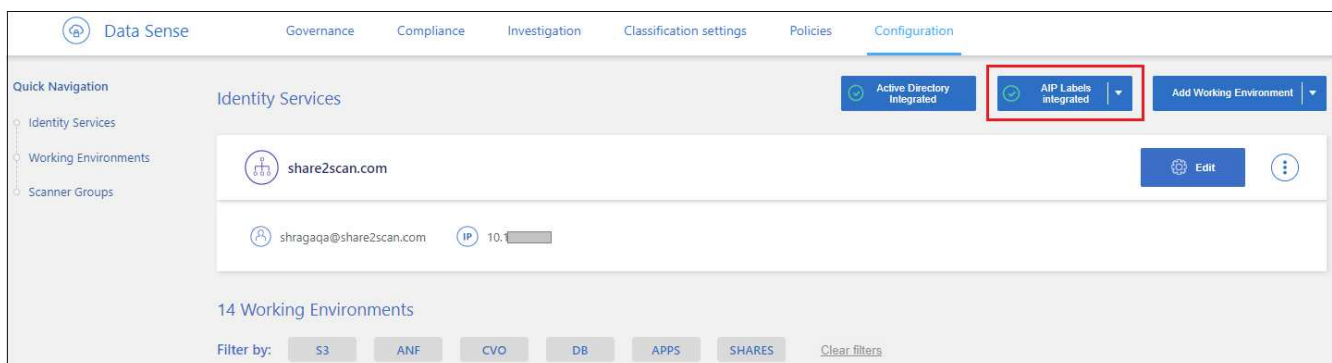
- You must have an account and an Azure Information Protection license.
- You must have the login credentials for the Azure account.
- If you plan to change labels in files that reside in Amazon S3 buckets, ensure that the permission `s3:PutObject` is included in the IAM role. See [setting up the IAM role](#).

Steps

1. From the BlueXP classification Configuration page, click **Integrate AIP Labels**.



2. In the Integrate AIP Labels dialog, click **Sign in to Azure**.
3. In the Microsoft page that appears, select the account and enter the required credentials.
4. Return to the BlueXP classification tab and you'll see the message "AIP Labels were integrated successfully with the account <account_name>".
5. Click **Close** and you'll see the text *AIP Labels integrated* at the top of the page.



Result

You can view and assign AIP labels from the results pane of the Investigation page. You can also assign AIP labels to files using Policies.

View AIP labels in your files

You can view the current AIP label that is assigned to a file.

In the Data Investigation results pane, click  for the file to expand the file metadata details.



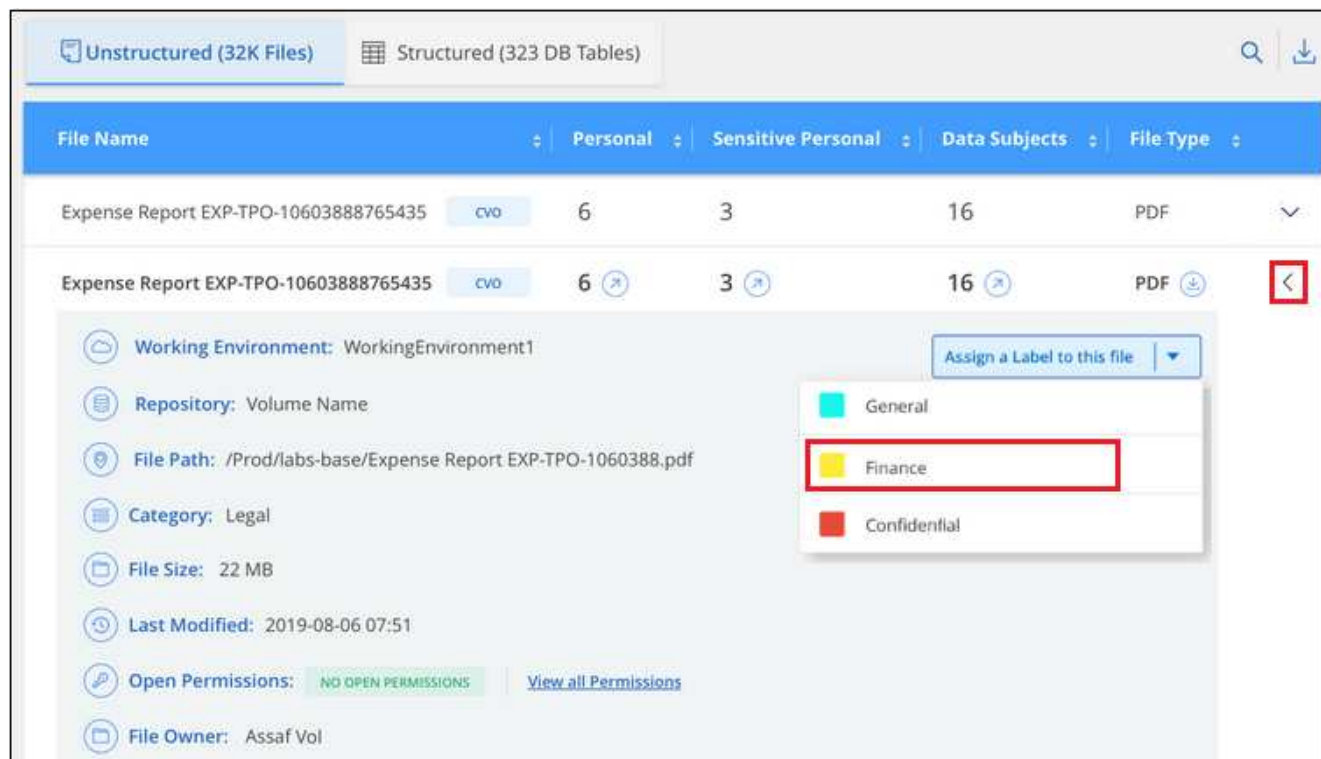
Assign AIP labels manually

You can add, change, and remove AIP labels from your files using BlueXP classification.

Follow these steps to assign an AIP label to a single file.

Steps

1. In the Data Investigation results pane, click  for the file to expand the file metadata details.



2. Click **Assign a Label to this file** and then select the label.

The label appears in the file metadata.

Follow these steps to assign an AIP label to multiple files. Note that you can assign an AIP label to a maximum of 20 files at a time (one page in the UI).

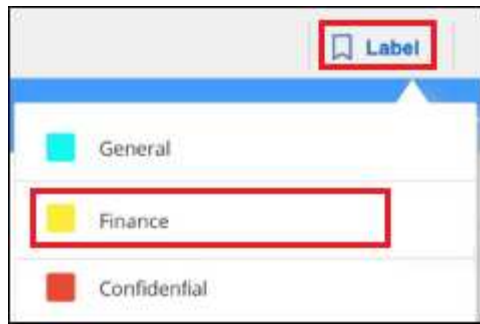
Steps

1. In the Data Investigation results pane, select the file, or files, that you want to label.



- To select individual files, check the box for each file (☒ Volume_1).
- To select all files on the current page, check the box in the title row (☒ File Name).

2. From the button bar, click **Label** and select the AIP label:



The AIP label is added to the metadata for all selected files.

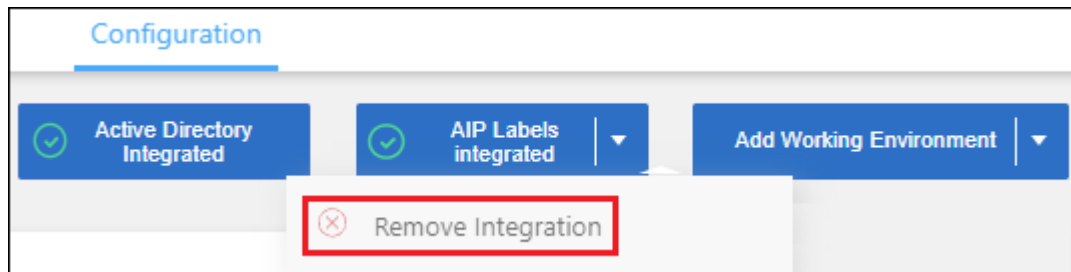
Remove the AIP integration

If you no longer want the ability to manage AIP labels in files, you can remove the AIP account from the BlueXP classification interface.

Note that no changes are made to the labels you have added using BlueXP classification. The labels that exist in files will stay as they currently exist.

Steps

1. From the *Configuration* page, click **AIP Labels integrated > Remove Integration**.



2. Click **Remove Integration** from the confirmation dialog.

Apply tags to manage your scanned files

You can add a tag to files that you want to mark for some type of follow-up. For example, you may have found some duplicate files and you want to delete one of them, but you need to check to see which one should be deleted. You could add a tag of "Check to delete" to the file so you know this file requires some research and some type of future action.

BlueXP classification enables you to view the tags that are assigned to files, add or remove tags from files, and change the name or delete an existing tag.

Note that the tag is not added to the file in the same way as AIP Labels are part of the file metadata. The tag is just seen by BlueXP users using BlueXP classification so you can see if a file needs to be deleted or checked for some type of follow-up.

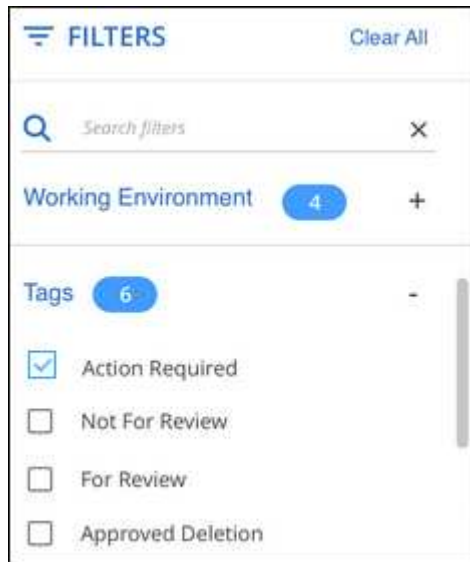


Tags assigned to files in BlueXP classification are not related to the tags you can add to resources, such as volumes or virtual machine instances. BlueXP classification tags are applied at the file level.

View files that have certain tags applied

You can view all the files that have specific tags assigned.

1. Click the **Investigation** tab from BlueXP classification.
2. In the Data Investigation page, click **Tags** in the Filters pane and then select the required tags.




The Investigation Results pane displays all the files that have those tags assigned.

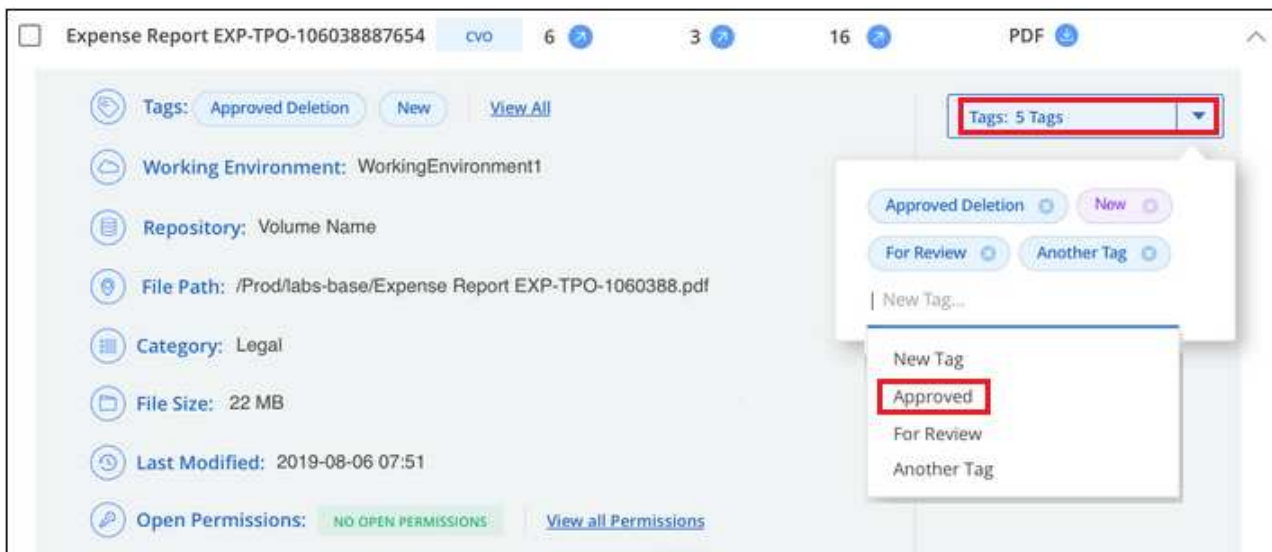
Assign tags to files

You can add tags to a single file or to a group of files.

To add a tag to a single file:

Steps

1. In the Data Investigation results pane, click  for the file to expand the file metadata details.
2. Click the **Tags** field and the currently assigned tags are displayed.
3. Add the tag or tags:
 - To assign an existing tag, click in the **New Tag...** field and start typing the name of the tag. When the tag you are looking for appears, select it and press **Enter**.
 - To create a new tag and assign it to the file, click in the **New Tag...** field, enter the name of the new tag, and press **Enter**.



The tag appears in the file metadata.

To add a tag to multiple files:

Steps

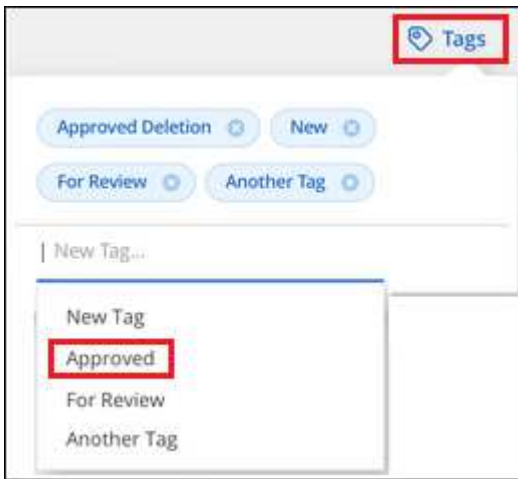
1. In the Data Investigation results pane, select the file, or files, that you want to tag.

255 items 1.2 GB 2 Selected 3 MB							Tags	Assign to	Label	Copy	Move	Delete
<input type="checkbox"/>	File Name		Personal	Sensitive Personal	Data Subjects	File Type						
<input checked="" type="checkbox"/>	Expense Report EXP-TPO-106038887654	cvo	6	3	16	PDF						
<input checked="" type="checkbox"/>	Expense Report EXP-TPO-106038887654	cvo	6	3	6	PDF						
<input type="checkbox"/>	Expense Report EXP-TPO-106038887654	cvo	6	3	6	PDF						
<input type="checkbox"/>	Expense Report EXP-TPO-106038887654	cvo	6	3	6	PDF						

- To select individual files, check the box for each file (☒ Volume_1).
- To select all files on the current page, check the box in the title row (☒ File Name).
- To select all files on all pages, check the box in the title row (☒ File Name), and then in the pop-up message **All 20 items on this page selected** [Select all items in list \(63K items\)](#), click **Select all items in list (xxx items)**.

You can apply tags to a maximum of 100,000 files at a time.

2. From the button bar, click **Tags** and the currently assigned tags are displayed.
3. Add the tag or tags:
 - To assign an existing tag, click in the **New Tag...** field and start typing the name of the tag. When the tag you are looking for appears, select it and press **Enter**.
 - To create a new tag and assign it to the file, click in the **New Tag...** field, enter the name of the new tag, and press **Enter**.



4. Approve adding the tags in the confirmation dialog and the tags are added to the metadata for all selected files.

Delete tags from files

You can delete a tag if you don't need to use it anymore.

Just click the **x** for an existing tag.



If you had selected multiple files, the tag is removed from all the files.

Assign users to manage certain files

You can assign a BlueXP user to a specific file, or to multiple files, so that person can be responsible for any follow-up actions that need to be done on the file. This capability is often used with the feature to add custom Status tags to a file.


For example, you might have a file that contains certain personal data that allows too many users read and write access (open permissions). So you could assign the Status tag "Change permissions" and assign this file to user "Joan Smith" so they can decide how to fix the issue. When they have fixed the issue they could change the Status tag to "Completed".

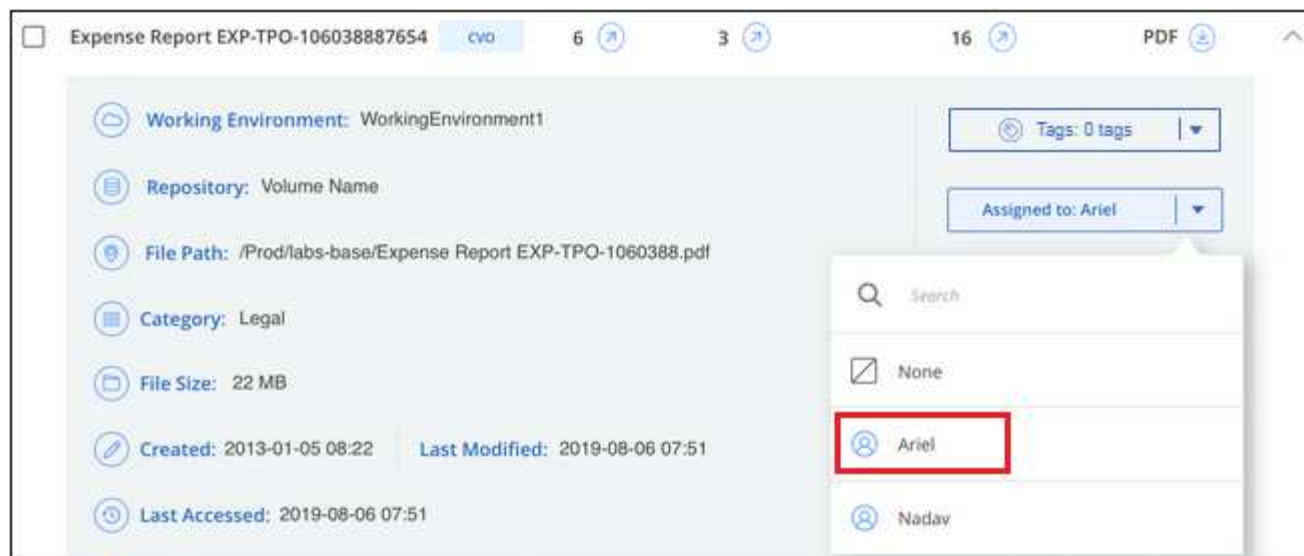
Note that the user name is not added to the file as part of the file metadata - it is just seen by BlueXP users when using BlueXP classification.

A new Filter in the Investigation page enables you to easily view all files that have the same person in the "Assigned To" field.

Follow these steps to assign a user to a single file.

Steps

1. In the Data Investigation results pane, click  for the file to expand the file metadata details.
2. Click the **Assigned to** field and select the user name.



The User name appears in the file metadata.

Follow these steps to assign a user to multiple files. Note that you can assign a user to a maximum of 20 files at a time (one page in the UI).

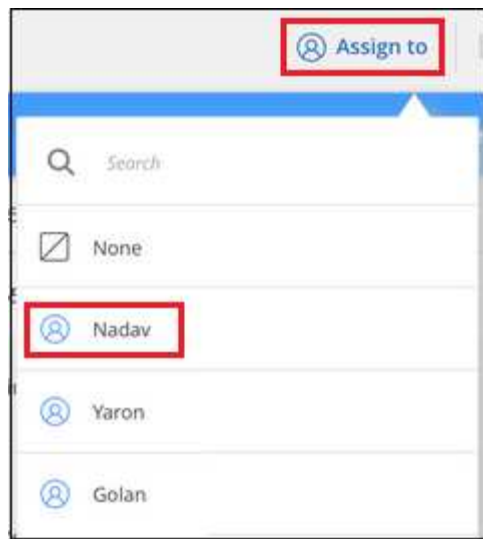
Steps

1. In the Data Investigation results pane, select the file, or files, that you want to assign to a user.

255 items 1.2 GB 2 Selected 3 MB							Tags	Assign to	Label	Copy	Move	Delete
<input type="checkbox"/>	File Name		Personal	Sensitive Personal	Data Subjects	File Type						
<input checked="" type="checkbox"/>	Expense Report EXP-TPO-106038887654	cvo	6	3	16	PDF						
<input checked="" type="checkbox"/>	Expense Report EXP-TPO-106038887654	cvo	6	3	6	PDF						
<input type="checkbox"/>	Expense Report EXP-TPO-106038887654	cvo	6	3	6	PDF						
<input type="checkbox"/>	Expense Report EXP-TPO-106038887654	cvo	6	3	6	PDF						

- To select individual files, check the box for each file (☒ Volume_1).
- To select all files on the current page, check the box in the title row (☒ File Name).

2. From the button bar, click **Assign to** and select the user name:



The user is added to the metadata for all selected files.

Manage your private data

BlueXP classification provides many ways for you to manage your private data. Some functionality makes it easier to prepare for migrating your data, while other functionality allows you to make changes to the data.

NOTE This information is relevant only for BlueXP classification legacy versions 1.30 and earlier.

- You can copy files to a destination NFS share if you want to make a copy of certain data and move it to a different NFS location.
- You can clone an ONTAP volume to a new volume, while including only selected files from the source volume in the new cloned volume. This is useful for situations where you're migrating data and you want to exclude certain files from the original volume.
- You can copy and synchronize files from a source repository to a directory in a specific destination location. This is useful for situations where you're migrating data from one source system to another while there is still some final activity on the source files.
- You can move source files that BlueXP classification is scanning to any NFS share.
- You can delete files that seem insecure or too risky to leave in your storage system, or that you have identified as duplicate.



- The capabilities described in this section are available only if you have chosen to 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 from Google Drive accounts can't use any of these capabilities at this time.

Copy source files

You can copy any source files that BlueXP classification is scanning. There are three types of copy operations depending on what you're trying to accomplish:

- **Copy files** from the same, or different, volumes or data sources to a destination NFS share.

This is useful if you want to make a copy of certain data and move it to a different NFS location.

- **Clone an ONTAP volume** to a new volume in the same aggregate, but include only selected files from the source volume in the new cloned volume.

This is useful for situations where you're migrating data and you want to exclude certain files from the original volume. This action uses the [NetApp FlexClone](#) functionality to quickly duplicate the volume and then remove the files that you **didn't** select.

- **Copy and synchronize files** from a single source repository (ONTAP volume, S3 bucket, NFS share, etc.) to a directory in a specific destination (target) location.

This is useful for situations where you're migrating data from one source system to another. After the initial copy, the service syncs any changed data based on the schedule that you set. This action uses the [NetApp BlueXP copy and sync](#) functionality to copy and sync data from a source to a target.

Copy source files to an NFS share

You can copy source files that BlueXP classification is scanning to any NFS share. The NFS share doesn't need to be integrated with BlueXP classification, you just need to know the name of the NFS share where all selected files will be copied in the format `<host_name>:/<share_path>`.



You can't copy files that reside in databases.

Requirements

- You must have the Account Admin or Workspace Admin role to copy files.
- Copying files requires that the destination NFS share allows access from the BlueXP classification instance.
- You can copy between 1 and 100,000 files at a time.

Steps

1. In the Data Investigation results pane, select the file, or files, that you want to copy, and click **Copy**.

<input type="checkbox"/>	File Name	Personal	Sensitive Personal	Data Subjects	File Type	
<input checked="" type="checkbox"/>	Expense Report EXP-TPO-106038887654	cvo	6	3	16	PDF
<input checked="" type="checkbox"/>	Expense Report EXP-TPO-106038887654	cvo	6	3	6	PDF
<input type="checkbox"/>	Expense Report EXP-TPO-106038887654	cvo	6	3	6	PDF
<input type="checkbox"/>	Expense Report EXP-TPO-106038887654	cvo	6	3	6	PDF

- To select individual files, check the box for each file (☒ Volume_1).
- To select all files on the current page, check the box in the title row (☒ File Name).
-

To select all files on all pages, check the box in the title row (☒ **File Name**), and then in the pop-up message **All 20 Items on this page selected** [Select all Items in list \(63K Items\)](#), click **Select all items in list (xxx items)**.

2. In the *Copy Files* dialog, select the **Regular Copy** tab.



3. Enter the name of the NFS share where all selected files will be copied in the format `<host_name>:/<share_path>`, and click **Copy**.

A dialog appears with the status of the copy operation.

You can view the progress of the copy operation in the [Actions Status](#) pane.

Note that you can also copy an individual file when viewing the metadata details for a file. Just click **Copy File**.



Clone volume data to a new volume

You can clone an existing ONTAP volume that BlueXP classification is scanning using NetApp *FlexClone* functionality. This allows you to quickly duplicate the volume while including only those files you selected. This is useful if you're migrating data and you want to exclude certain files from the original volume, or if you want to create a copy of a volume for testing.

The new volume is created in the same aggregate as the source volume. Ensure that you have enough space for this new volume in the aggregate before you start this task. Contact your storage administrator if necessary.

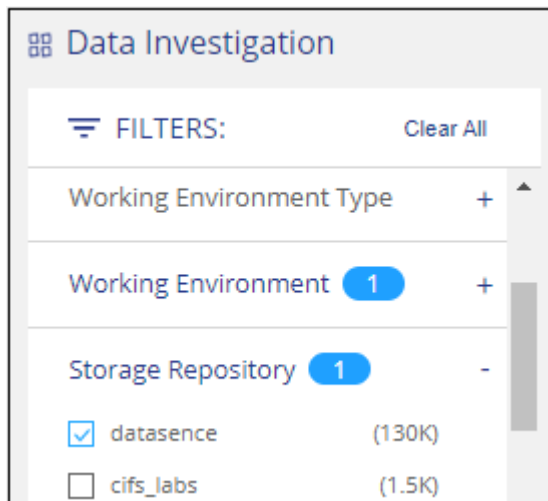
Note: FlexGroup volumes can't be cloned because they're not supported by FlexClone.

Requirements

- You must have the Account Admin or Workspace Admin role to copy files.
- You must select a minimum of 20 files.
- All selected files must be from the same volume, and the volume must be online.
- The volume must be from a Cloud Volumes ONTAP or on-premises ONTAP system. No other data sources are currently supported.
- The FlexClone license must be installed on the cluster. This license is installed by default on Cloud Volumes ONTAP systems.

Steps

1. In the Data Investigation pane, create a filter by selecting a single **Working Environment** and a single **Storage Repository** to make sure all the files are from the same ONTAP volume.



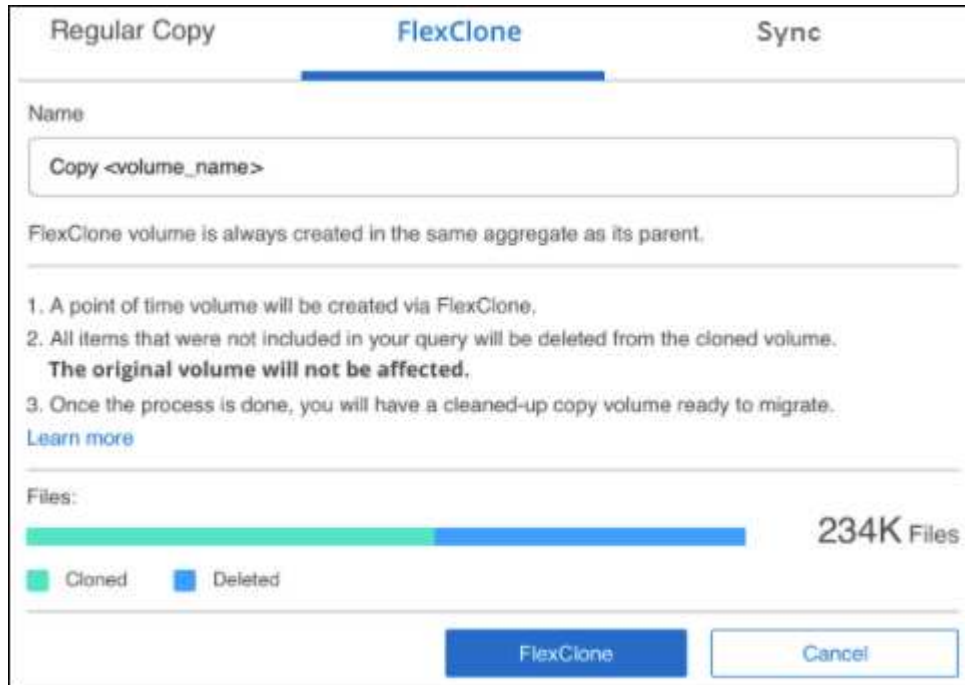
Apply any other filters so that you're seeing only the files that you want to clone to the new volume.

2. In the Investigation results pane, select the files that you want to clone and click **Copy**.



- To select individual files, check the box for each file (☒ Volume_1).
- To select all files on the current page, check the box in the title row (☒ File Name).
- To select all files on all pages, check the box in the title row (☒ File Name), and then in the pop-up message **All 20 Items on this page selected** [Select all Items in list \(63K Items\)](#), click **Select all items in list (xxx items)**.

3. In the *Copy Files* dialog, select the **FlexClone** tab. This page shows the total number of files that will be cloned from the volume (the files you selected), and the number of files that are not included/deleted (the files you didn't select) from the cloned volume.



The screenshot shows a dialog box with three tabs: "Regular Copy", "FlexClone" (selected), and "Sync". Under the "FlexClone" tab, there is a "Name" field with the placeholder text "Copy <volume_name>". Below this, a note states: "FlexClone volume is always created in the same aggregate as its parent." A list of three points follows: 1. A point of time volume will be created via FlexClone. 2. All items that were not included in your query will be deleted from the cloned volume. **The original volume will not be affected.** 3. Once the process is done, you will have a cleaned-up copy volume ready to migrate. A "Learn more" link is provided. Below the list, a "Files:" section shows a progress bar with a green segment (Cloned) and a blue segment (Deleted). To the right of the bar, it says "234K Files". At the bottom, there are two buttons: "FlexClone" and "Cancel".

4. Enter the name of the new volume, and click **FlexClone**.

A dialog appears with the status of the clone operation.

Result

The new, cloned volume is created in the same aggregate as the source volume.

You can view the progress of the clone operation in the [Actions Status pane](#).

If you initially selected **Map all volumes** or **Map & Classify all volumes** when you enabled BlueXP classification for the working environment where the source volume resides, then BlueXP classification will scan the new cloned volume automatically. If you didn't use either of these selections initially, then if you want to scan this new volume, you'll need to [enable scanning on the volume manually](#).

Copy and synchronize source files to a target system

You can copy source files that BlueXP classification is scanning from any supported unstructured data source to a directory in a specific target destination location ([target locations that are supported by BlueXP copy and sync](#)). After the initial copy, any data changed in the files are synchronized based on the schedule that you configure.

This is useful for situations where you're migrating data from one source system to another. This action uses the [NetApp BlueXP copy and sync](#) functionality to copy and sync data from a source to a target.



You can't copy and sync files that reside in databases, OneDrive accounts, or SharePoint accounts.

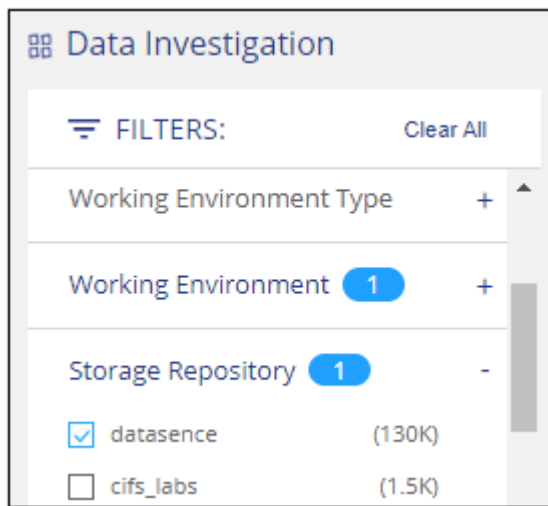
Requirements

- You must have the Account Admin or Workspace Admin role to copy and sync files.
- You must select a minimum of 20 files.
- All selected files must be from the same source repository (ONTAP volume, S3 bucket, NFS or CIFS share, etc.).
- You'll need to activate the BlueXP copy and sync service and configure a minimum of one data broker that can be used to transfer files between the source and target systems. Review the BlueXP copy and sync requirements beginning with the [Quick Start description](#).

Note that the BlueXP copy and sync service has separate service charges for your sync relationships, and will incur resource charges if you deploy the data broker in the cloud.

Steps

1. In the Data Investigation pane, create a filter by selecting a single **Working Environment** and a single **Storage Repository** to make sure all the files are from the same repository.



Apply any other filters so that you're seeing only the files that you want to copy and sync to the destination system.

2. In the Investigation results pane, select all files on all pages by checking the box in the title row

(☒ **File Name**), then in the pop-up message **All 20 Items on this page selected** [Select all Items in list \(63K Items\)](#) click **Select all items in list (xxx items)**, and then click **Copy**.

238.1 Items | 244.2 GB

Tags | Assign to | Label | Move | Copy | Delete

☒ File Name 1

Personal | Sensitive Personal | Data Subjects | File Type

All 20 Items on this page selected | 24 MB

Select all items in list (238k items | 244GB) 2

File Name	CVO	652	0	1	TXT	
<input checked="" type="checkbox"/> CRM_Customers.txt	CVO	652	0	1	TXT	▼
<input checked="" type="checkbox"/> truepositive.txt	CVO	0	61	11	TXT	▼
<input checked="" type="checkbox"/> test_file.txt	CVO	6	611	111	TXT	▼
<input checked="" type="checkbox"/> test_positive.txt	CVO	0	65	51	TXT	▼

3. In the *Copy Files* dialog, select the **Sync** tab.

Regular Copy | FlexClone | **Sync**

An easy to use replication service for transferring data between any file or object store, on prem or in the cloud.

[Learn More](#)

32K items will be synced using Cloud Sync.

Source ↔ Target

Data Sense

Data Broker

OK Cancel

4. If you are sure that you want to sync the selected files to a destination location, click **OK**.

The BlueXP copy and sync UI is opened in BlueXP.

You are prompted to define the sync relationship. The Source system is pre-populated based on the repository and files you already selected in BlueXP classification.

5. You'll need to select the Target system and then select (or create) the Data Broker you plan to use. Review the BlueXP copy and sync requirements beginning with the [Quick Start description](#).

Result

The files are copied to the target system and they'll be synchronized based on the schedule you define. If you select a one-time sync then the files are copied and synchronized one time only. If you choose a periodic sync, then the files are synchronized based on the schedule. Note that if the source system adds new files that match the query you created using filters, those *new* files will be copied to the destination and synchronized in the future.

Note that some of the usual BlueXP copy and sync operations are disabled when it is invoked from BlueXP classification:

- You can't use the **Delete Files on Source** or **Delete Files on Target** buttons.
- Running a report is disabled.

Move source files to an NFS share

You can move source files that BlueXP classification is scanning to any NFS share. The NFS share doesn't need to be integrated with BlueXP classification.

Optionally, you can leave a breadcrumb file in the location of the moved file. A breadcrumb file helps your users understand why a file was moved from its original location. For each moved file, the system creates a breadcrumb file in the source location named `<filename>-breadcrumb-<date>.txt`. You can add text in the dialog box that will be added to the breadcrumb file to indicate the location where the file was moved and the user who moved the file.

Note that the subdirectory structure from the source file is recreated on the destination share when the file is moved so it is easier to understand where the file was moved from. If a file with the same name exists in the destination location, the file will not be moved.



You can't move files that reside in databases.

Requirements

- You must have the Account Admin or Workspace Admin role to move files.
- The source files can be located in the following data sources: On-premises ONTAP, Cloud Volumes ONTAP, Azure NetApp Files, File Shares, and SharePoint Online.
- You can move a maximum of 15 million files at a time.
- Only files which are 50 MB or smaller are moved.
- The destination NFS share must allow access from the BlueXP classification instance IP address.

Steps

1. In the Data Investigation results pane, select the file, or files, that you want to move.

255 items 1.2 GB 2 Selected 3 MB							Tags	Assign to	Label	Copy	Move	Delete
<input type="checkbox"/>	File Name		Personal	Sensitive Personal	Data Subjects	File Type						
<input checked="" type="checkbox"/>	Expense Report EXP-TPO-106038887654	cvo	6	3	16	PDF						
<input checked="" type="checkbox"/>	Expense Report EXP-TPO-106038887654	cvo	6	3	6	PDF						
<input type="checkbox"/>	Expense Report EXP-TPO-106038887654	cvo	6	3	6	PDF						
<input type="checkbox"/>	Expense Report EXP-TPO-106038887654	cvo	6	3	6	PDF						

- To select individual files, check the box for each file (☒ Volume_1).
- To select all files on the current page, check the box in the title row (☒ File Name).
-

To select all files on all pages, check the box in the title row (☒ **File Name**), and then in the pop-up message **All 20 Items on this page selected** [Select all Items in list \(63K Items\)](#), click **Select all items in list (xxx items)**.

- From the button bar, click **Move**.

Move Files (63)

The files will be moved to the destination folder you provide and will no longer be available at their current location.

Moving files is supported only to destination folders in NFS Shares. Any NFS Share is supported, no matter where it is hosted, as long as the share's export policy allows access from the data connector instance IP address.

The status of this action will appear in the Action Status.

Enter the NFS destination folder path to continue

Hostname:/SHAREPATH

☒ **Leave breadcrumb**

A breadcrumb file helps your users understand why a file was moved from its original location. For each moved file, the system creates a breadcrumb file in the source location named **<filename>-breadcrumb-<date>.txt**.

Enter the content of your breadcrumb

Max length should be maximum 400 characters

Move Files **Cancel**

- In the *Move Files* dialog, enter the name of the NFS share where all selected files will be moved in the format **<host_name>:/<share_path>**.
- If you want to leave a breadcrumb file, check the *Leave breadcrumb* box. You can enter text in the dialog box to indicate the location where the file was moved and the user who moved the file, and any other information, such as the reason the file was moved.
- Click **Move Files**.

Note that you can also move an individual file when viewing the metadata details for a file. Just click **Move File**.



Delete source files

You can permanently remove source files that seem insecure or too risky to leave in your storage system, or that you've identified as a duplicate. This action is permanent and there is no undo or restore.

You can delete files manually from the Investigation pane, or [automatically using Policies](#).



You can't delete files that reside in databases. All other data sources are supported.

Deleting files requires the following permissions:

- For NFS data - the export policy needs to be defined with write permissions.
- For CIFS data - the CIFS credentials need to have write permissions.
- For S3 data - the IAM role must include the following permission: `s3:DeleteObject`.

Delete source files manually

Requirements

- You must have the Account Admin or Workspace Admin role to delete files.
- You can delete a maximum of 100,000 files at a time.

Steps

1. In the Data Investigation results pane, select the file, or files, that you want to delete.



- To select individual files, check the box for each file (☒ Volume_1).

- To select all files on the current page, check the box in the title row (☒ File Name).
 - To select all files on all pages, check the box in the title row (☒ File Name), and then in the pop-up message **All 20 Items on this page selected Select all Items in list (63K Items)**, click **Select all items in list (xxx items)**.
2. From the button bar, click **Delete**.
 3. Because the delete operation is permanent, you must type "**permanently delete**" in the subsequent *Delete File* dialog and click **Delete File**.

You can view the progress of the delete operation in the [Actions Status pane](#).

Note that you can also delete an individual file when viewing the metadata details for a file. Just click **Delete file**.



Add personal data identifiers to your BlueXP classification scans

BlueXP classification provides many ways for you to add a custom list of "personal data" that BlueXP classification will identify in future scans, giving you the full picture about where potentially sensitive data resides in *all* your organizations' files.

NOTE This information is relevant only for BlueXP classification legacy versions 1.30 and earlier.

- You can add unique identifiers based on specific columns in databases you are scanning.
- You can add custom keywords from a text file — these words are identified within your data.
- You can add a personal pattern using a regular expression (regex) — the regex is added to the existing predefined patterns.
- You can add custom categories to identify where specific categories of information are found in your data.

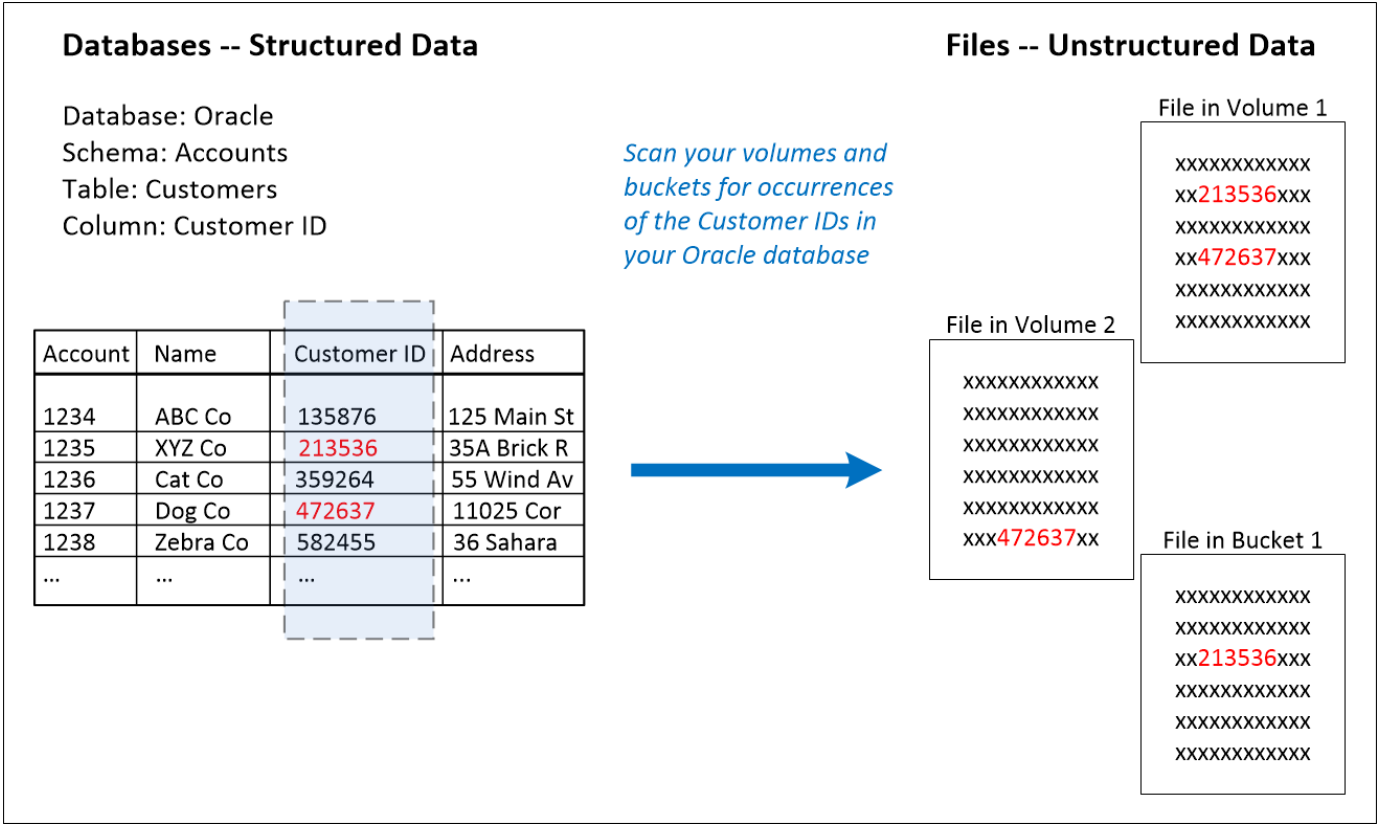
All of these mechanisms to add custom scanning criteria are supported in all languages.



The capabilities described in this section are available only if you have chosen to perform a full classification scan on your data sources. Data sources that have had a mapping-only scan do not show file-level details.

Add custom personal data identifiers from your databases

A feature we call *Data Fusion* allows you to scan your organizations' data to identify whether unique identifiers from your databases are found in any of your other data sources. You can choose the additional identifiers that BlueXP classification will look for in its' scans by selecting a specific column, or columns, in a database table. For example, the diagram below shows how data fusion is used to scan your volumes, buckets, and databases for occurrences of all your Customer IDs from your Oracle database.



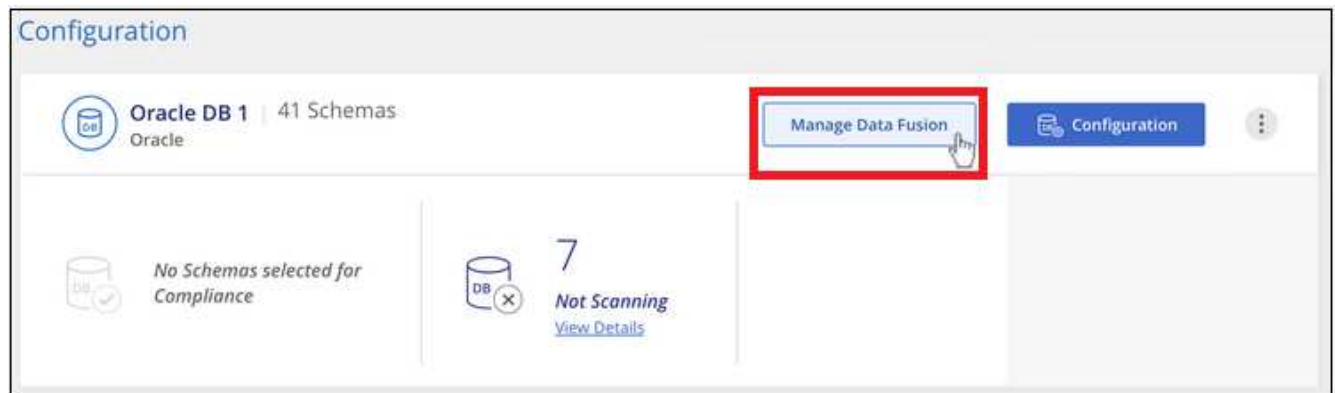
As you can see, two unique Customer IDs have been found in two volumes and in one S3 bucket. Any matches in database tables will also be identified.

Note that since you're scanning your own databases, whatever language your data is stored in will be used to identify data in future BlueXP classification scans.

Steps

You must have [added at least one database server](#) to BlueXP classification before you can add data fusion sources.

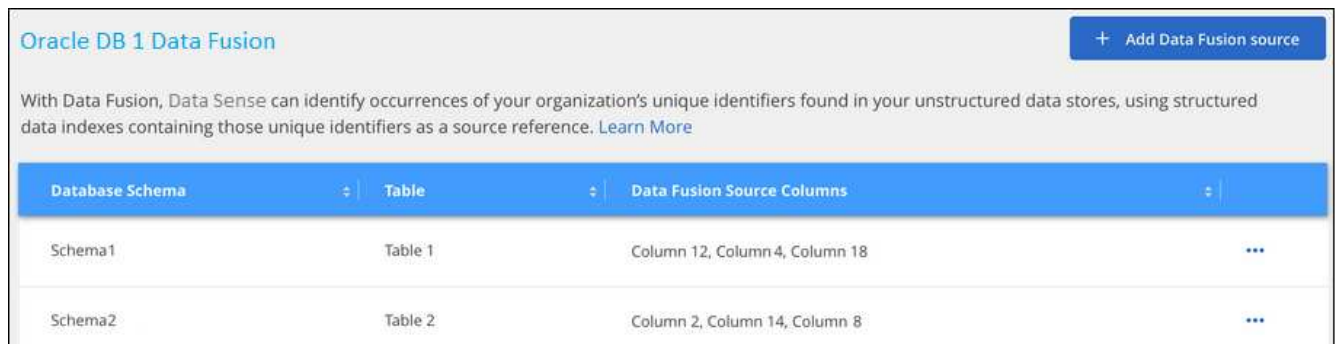
- 1. In the Configuration page, click **Manage Data Fusion** in the database where the source data resides.



2. Click **Add Data Fusion source** on the next page.
3. In the *Add Data Fusion Source* page:
 - a. Select the Database Schema from the drop-down menu.
 - b. Enter the Table name in that schema.
 - c. Enter the Column, or Columns, that contain the unique identifiers you want to use.

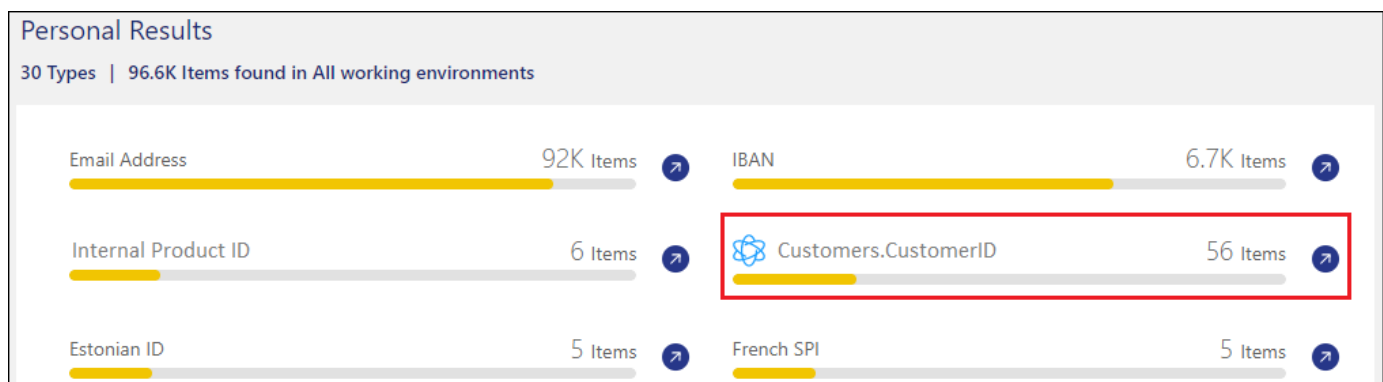
When adding multiple columns, enter each column name, or table view name, on a separate line.

4. Click **Add Data Fusion Source**.



Results

After the next scan, the results will include this new information in the Compliance Dashboard under the "Personal Results" section, and in the Investigation page in the "Personal Data" filter. The name you used for the classifier appears in the filter list, for example `Customers.CustomerID`.



Delete a Data Fusion source

If at some point you decide not to scan your files using a certain Data Fusion source, you can select the source row from the Data Fusion inventory page and click **Delete Data Fusion Source**.



Add custom keywords from a list of words

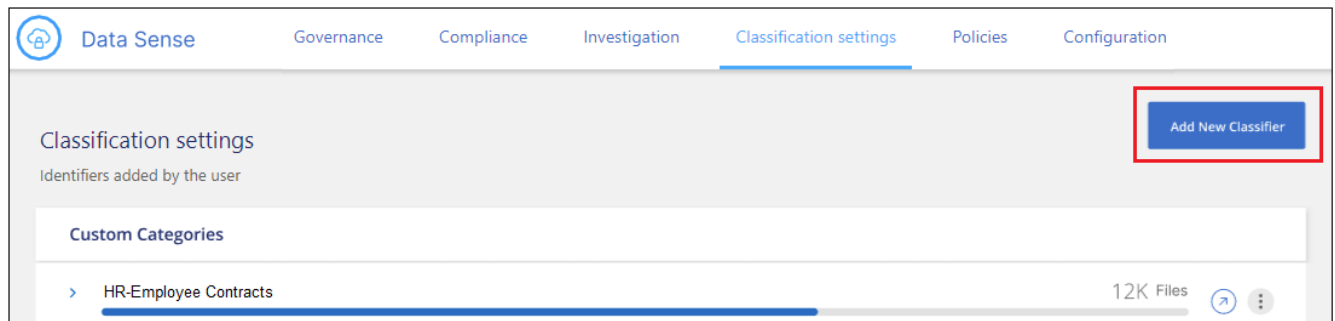
You can add custom keywords to BlueXP classification so that it will identify where that information is found in your data. You add the keywords just by entering each word you want BlueXP classification to recognize. The keywords are added to the existing predefined keywords that BlueXP classification already uses, and the results will be visible under the personal patterns section.

For example, you may want to see where internal Product Names are mentioned in all of your files to make sure these names are not accessible in locations that are not secure.

After updating the custom keywords, BlueXP classification will restart scanning all data sources. After the scan has completed, the new results will appear in the BlueXP classification Compliance Dashboard under the "Personal Results" section, and in the Investigation page in the "Personal Data" filter.

Steps

1. From the *Classification settings* tab, click **Add New Classifier** to launch the *Add Custom Classifier* wizard.



2. In the *Select type* page, enter the name of the classifier, provide a brief description, select **Personal identifier**, and then click **Next**.

The name you enter will appear in the BlueXP classification UI as the heading for scanned files that match the classifier requirements, and as the name of the filter in the Investigation page.

You can also check the box to "Mask detected results in the system" so the full result won't appear in the UI. For example, you may want to do this to hide full credit card numbers or similar personal data (the mask would appear in the UI like this: "**** * 3434").

1 Select type

2 Select tool

3 Create Logic

Select type

Select the type of classifier that you want to add to the system, and provide the name and description. Data Sense re-scans all your data sources after you add a new classifier. When the scan is complete, all matching results are displayed in the "Classification Settings" dashboard and in other Data Sense pages.

Classifier name

Internal Product Names

Description

Identify internal product names found in all files

☒ **Personal identifier**

The classifier will be added to the system as a new personal identifier. Any matches are considered "personal data", and they are added to the results that are displayed in the Personal Results page and in the Investigation page. [See the list of personal data that Data Sense identifies by default.](#)

☐ Mask detected results in the system

☐ **Category**

The classifier will be added to the system as a new Category. Any matches are added to the results that are displayed in the Categories page and in the Investigation page. [See the list of categories that Data Sense identifies by default.](#)

Previous

Next

- In the *Select Data Analysis Tool* page, select **Custom keywords** as the method you want to use to define the classifier, and then click **Next**.

Select Data Analysis Tool

Select the tool that will be used to build the list of words, or patterns, that Data Sense will attempt to match in your data sources.

☒

Custom keywords ⓘ
Create a custom personal pattern based on a list of keywords that you provide.

☐

Custom regular expression ⓘ
Create a custom personal pattern based on a regular expression that you define.

☐

DB fusion ⓘ
Create a custom personal pattern based on the values found in specific columns in your scanned database tables. This allows you to identify whether unique identifiers from your databases are found in any of your other data sources.

Previous

Next

4. In the *Create Logic* page, enter the keywords you want to recognize - each word on a separate line - and click **Validate**.

The screenshot below shows internal Product Names (different types of owls). The BlueXP classification search for these items is not case sensitive.

Create Logic

Create logic for the new identifier, based on regular expression and keywords that should be detected. You will be able to change the logic in the future, by clicking on "edit" from the custom classification dashboard.

Custom keywords list ¹

- Maximum of 100,000 words.
- Separate between keywords with a new line
- The keywords are not case sensitive
- Each word must be at least 3 characters long. Shorter words are ignored.
- Duplicate words are only added once.

barred
barn
horned
snowy
screech

Validate

✔ Keywords list is valid.

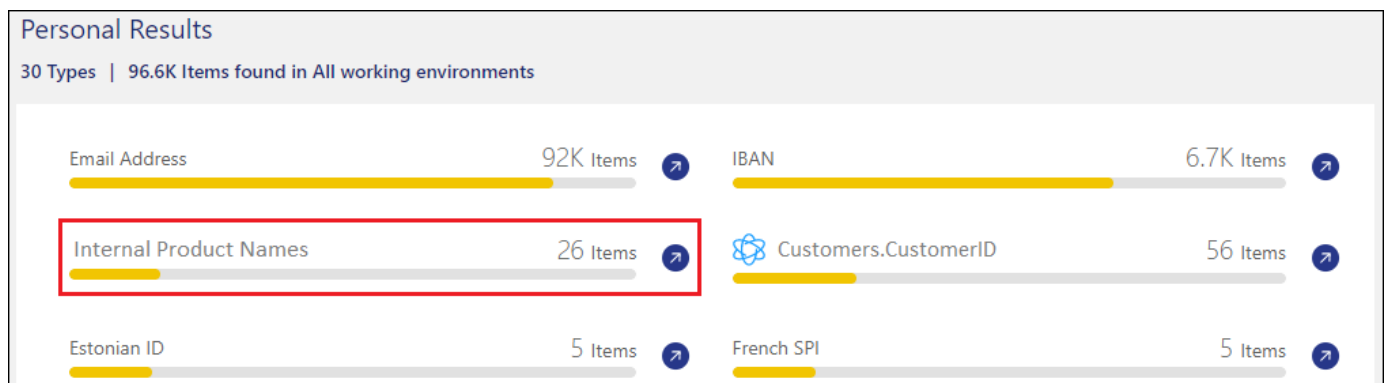
Previous

Done

5. Click **Done** and BlueXP classification starts to rescan your data.

Results

After the scan is complete, the results will include this new information in the Compliance Dashboard under the "Personal Results" section, and in the Investigation page in the "Personal Data" filter.



As you can see, the name of the classifier is used as the name in the Personal Results panel. In this manner you can activate many different groups of keywords and see the results for each group.

Add custom personal data identifiers using a regex

You can add a personal pattern to identify specific information in your data using a custom regular expression (regex). This allows you to create a new custom regex to identify new personal information elements that don't yet exist in the system. The regex is added to the existing predefined patterns that BlueXP classification

already uses, and the results will be visible under the personal patterns section.

For example, you may want to see where your internal Product IDs are mentioned in all of your files. If the Product ID has a clear structure, for example, it is a 12-digit number that starts with 201, you can use the custom regex feature to search for it in your files. The regular expression for this example is `\b201\d{9}\b`.

After adding the regex, BlueXP classification will restart scanning all data sources. After the scan has completed, the new results will appear in the BlueXP classification Compliance Dashboard under the "Personal Results" section, and in the Investigation page in the "Personal Data" filter.

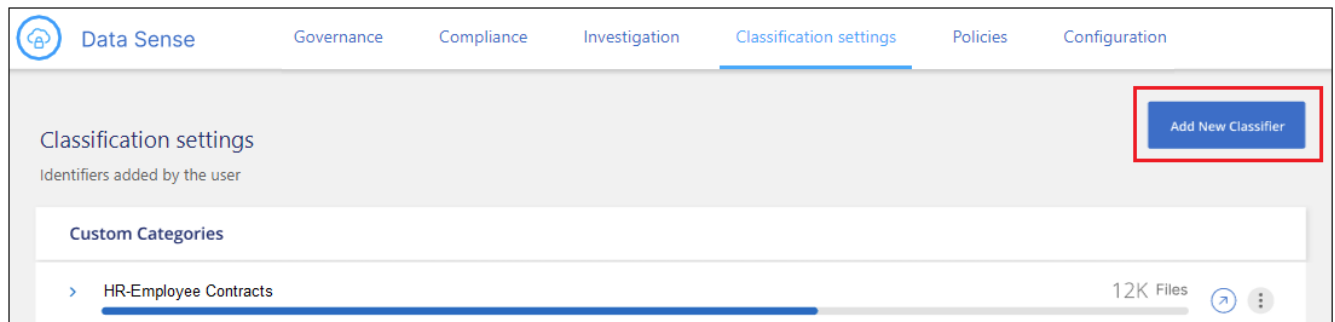
If you need assistance in building the regular expression, refer to [Regular expressions 101](#). Choose **Python** for the Flavor to see the types of results BlueXP classification will match from the regular expression. The [Python Regex Tester page](#) is also useful by displaying a graphical representation of your patterns.



Currently we do not allow the use of pattern flags when creating a regex - this means you should not use `/`.

Steps

1. From the *Classification settings* tab, click **Add New Classifier** to launch the *Add Custom Classifier* wizard.



2. In the *Select type* page, enter the name of the classifier, provide a brief description, select **Personal identifier**, and then click **Next**.

The name you enter will appear in the BlueXP classification UI as the heading for scanned files that match the classifier requirements, and as the name of the filter in the Investigation page. You can also check the box to "Mask detected results in the system" so the full result won't appear in the UI. For example, you may want to do this to hide full credit card numbers or similar personal data.

1 Select type

2 Select tool

3 Create Logic

Select type

Select the type of classifier that you want to add to the system, and provide the name and description. Data Sense re-scans all your data sources after you add a new classifier. When the scan is complete, all matching results are displayed in the "Classification Settings" dashboard and in other Data Sense pages.

Classifier name

Internal Product ID

Description

Identify internal product IDs found in all files

☒ **Personal identifier**

The classifier will be added to the system as a new personal identifier. Any matches are considered "personal data", and they are added to the results that are displayed in the Personal Results page and in the Investigation page. [See the list of personal data that Data Sense identifies by default.](#)

☐ Mask detected results in the system

☐ **Category**

The classifier will be added to the system as a new Category. Any matches are added to the results that are displayed in the Categories page and in the Investigation page. [See the list of categories that Data Sense identifies by default.](#)

Previous

Next

- In the *Select Data Analysis Tool* page, select **Custom regular expression** as the method you want to use to define the classifier, and then click **Next**.

Select Data Analysis Tool

Select the tool that will be used to build the list of words, or patterns, that Data Sense will attempt to match in your data sources.

☐

Custom keywords ⓘ
Create a custom personal pattern based on a list of keywords that you provide.

☒

Custom regular expression ⓘ
Create a custom personal pattern based on a regular expression that you define.

☐

DB fusion ⓘ
Create a custom personal pattern based on the values found in specific columns in your scanned database tables. This allows you to identify whether unique identifiers from your databases are found in any of your other data sources.

Previous

Next

4. In the *Create Logic* page, enter the regular expression and any proximity words, and click **Done**.
 - a. You can enter any legal regular expression. Click the **Validate** button to have BlueXP classification verify that the regular expression is valid, and that it is not too broad — meaning it will return too many results.
 - b. Optionally, you can enter some proximity words to help refine the accuracy of the results. These are words that will typically be found within 300 characters of the pattern you are searching for (either before or after the found pattern). Enter each word, or phrase, on a separate line.

Create Logic

Create logic for the new identifier, based on regular expression and keywords that should be detected.

Regular expression ⓘ

Add the pattern that should be detected to identify specific information in your data, using a custom regular expression.

✓ **Success:** Regular expression is valid.

☒ **Proximity words** - To improve the detection accuracy, insert phrases that must appear near by the regular expression's match.

Results

The classifier is added and BlueXP classification starts to rescan all your data sources. You are returned to the Custom Classifiers page where you can view the number of files that have matched your new classifier. Results from scanning all of your data sources will take some time depending on the number of files that need to be scanned.

Data Sense	Governance	Compliance	Investigation	Classification settings	Policies	Configuration
Classification settings						
Identifiers added by the user						
Custom Categories						
HR - Employee Contracts 7.5K Files						
Personal information						
Internal Product ID 12K Files						

Add custom categories

BlueXP classification takes the data that it scans and divides it into different types of categories. Categories are topics based on artificial intelligence analysis of the content and metadata of each file. [See the list of](#)

predefined categories.

Categories can help you understand what's happening with your data by showing you the types of information that you have. For example, a category like *resumes* or *employee contracts* may include sensitive data. When you investigate the results, you might find that employee contracts are stored in an insecure location. You can then correct that issue.

You can add custom categories to BlueXP classification so you can identify where categories of information that are unique for your data estate are found in your data. You add each category by creating "training" files that contain the categories of data that you want to identify, and then have BlueXP classification scan those files to "learn" through AI so that it can identify that data in your data sources. The categories are added to the existing predefined categories that BlueXP classification already identifies, and the results are visible under the Categories section.

For example, you may want to see where compressed installation files in .gz format are located in your files so that you can remove them, if necessary.

After updating the custom categories, BlueXP classification will restart scanning all data sources. After the scan has completed, the new results will appear in the BlueXP classification Compliance Dashboard under the "Categories" section, and in the Investigation page in the "Category" filter. [See how to view files by categories.](#)

What you'll need

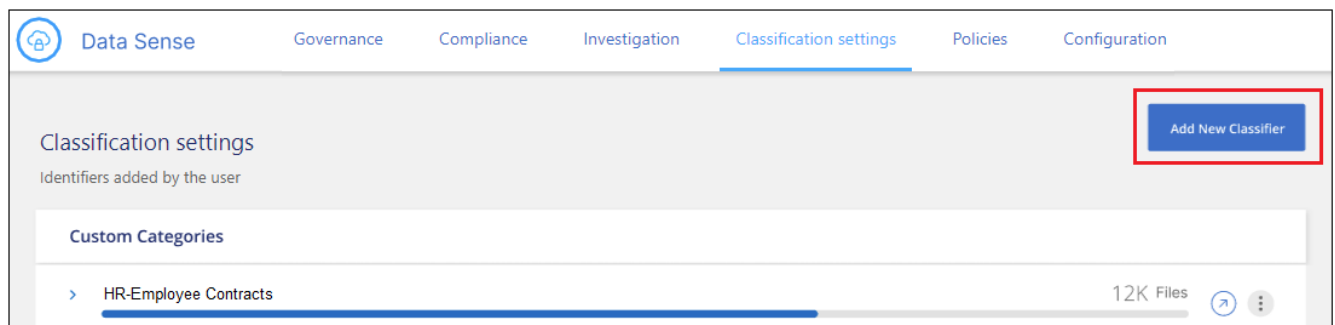
You'll need to create a minimum of 25 training files that contain samples of the categories of data that you want BlueXP classification to recognize. The following file types are supported:

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

The files must be a minimum of 100 bytes, and they must be located in a folder that is accessible by BlueXP classification.

Steps

1. From the *Classification settings* tab, click **Add New Classifier** to launch the *Add Custom Classifier* wizard.



2. In the *Select type* page, enter the name of the classifier, provide a brief description, select **Category**, and then click **Next**.

The name you enter will appear in the BlueXP classification UI as the heading for scanned files that match the category of data you are defining, and as the name of the filter in the Investigation page.

1 Select type
2 Select tool
3 Create Logic

Select type

Select the type of classifier that you want to add to the system, and provide the name and description. Data Sense re-scans all your data sources after you add a new classifier. When the scan is complete, all matching results are displayed in the "Classification Settings" dashboard and in other Data Sense pages.

Classifier name

Description

☐ **Personal identifier**
The classifier will be added to the system as a new personal identifier. Any matches are considered "personal data", and they are added to the results that are displayed in the Personal Results page and in the Investigation page. [See the list of personal data that Data Sense identifies by default.](#)
☐ Mask detected results in the system

☒ **Category**
The classifier will be added to the system as a new Category. Any matches are added to the results that are displayed in the Categories page and in the Investigation page. [See the list of categories that Data Sense identifies by default.](#)

Previous
Next

- In the *Create Logic* page, make sure you have the learning files prepared, and then click **Select files**.

Create Logic

AI-based similarity training

- Insert NFS folder path
- The folder should contain minimum 25 files and maximum 1000 files that will be used for the AI training.
- Supported file types: pdf, docx, doc, pptx, xls, xlsx, csv, txt, gz, rtf, docs, sheets, slides, json
- The keywords are not case sensitive
- Minimum file size: 100B

Compressed Installer files

- Enter the IP address of the volume, and the path where the training files are located, and click **Add**.

Insert folder path that contains at least 25 files for the training

Enter the IP address and volume name, along with the path to the location of the training files.

IP

Training Data - Folder path

XXX.XXX.XXX.XXX:/VolumeName

folder/path/

Add

Cancel

- Verify that the training files were recognized by BlueXP classification. Click the **x** to remove any training files that do not meet the requirements. Then click **Done**.

Create Logic

AI-based similarity training

- Insert NFS folder path
- The folder should contain minimum 25 files and maximum 1000 files that will be used for the AI training.
- Supported file types: pdf, docx, doc, pptx, xls, xlsx, csv, txt, gz, rtf, docs, sheets, slides, json
- The keywords are not case sensitive
- Minimum file size: 100B

Select Files

Compressed Installer files

Total uploaded files: 54

File name	File Size	File Type	Reliability	included in training
File1	56	File type	Sufficient	x
File2	22	File type	Sufficient	x
File3	43	File type	Sufficient	x
File4	11	File type	Sufficient	x

Previous

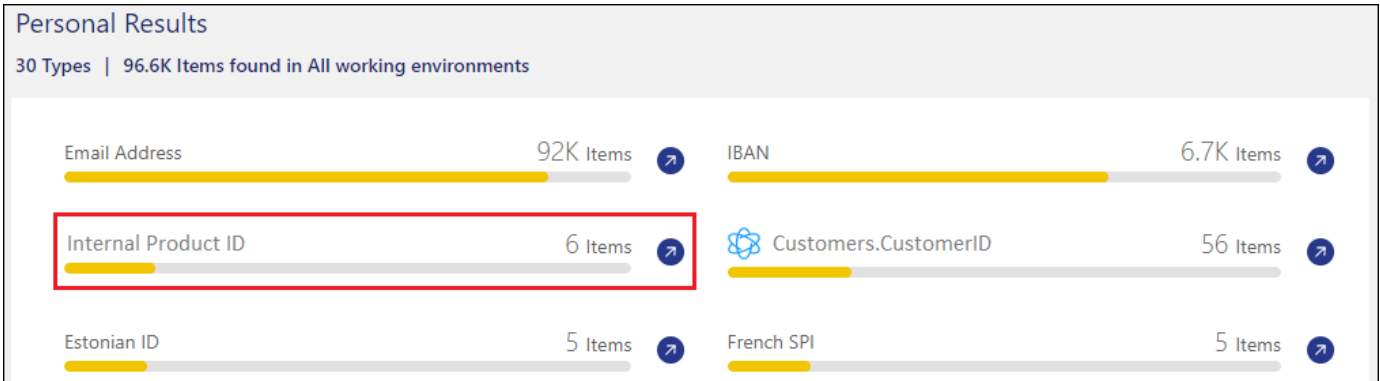
Done

Results

The new category is created as defined by the training files and added to BlueXP classification. Then BlueXP classification starts to rescan all your data sources to identify files that fit into this new category. You are returned to the Custom Classifiers page where you can view the number of files that have matched your new category. Results from scanning all of your data sources will take some time depending on the number of files that need to be scanned.

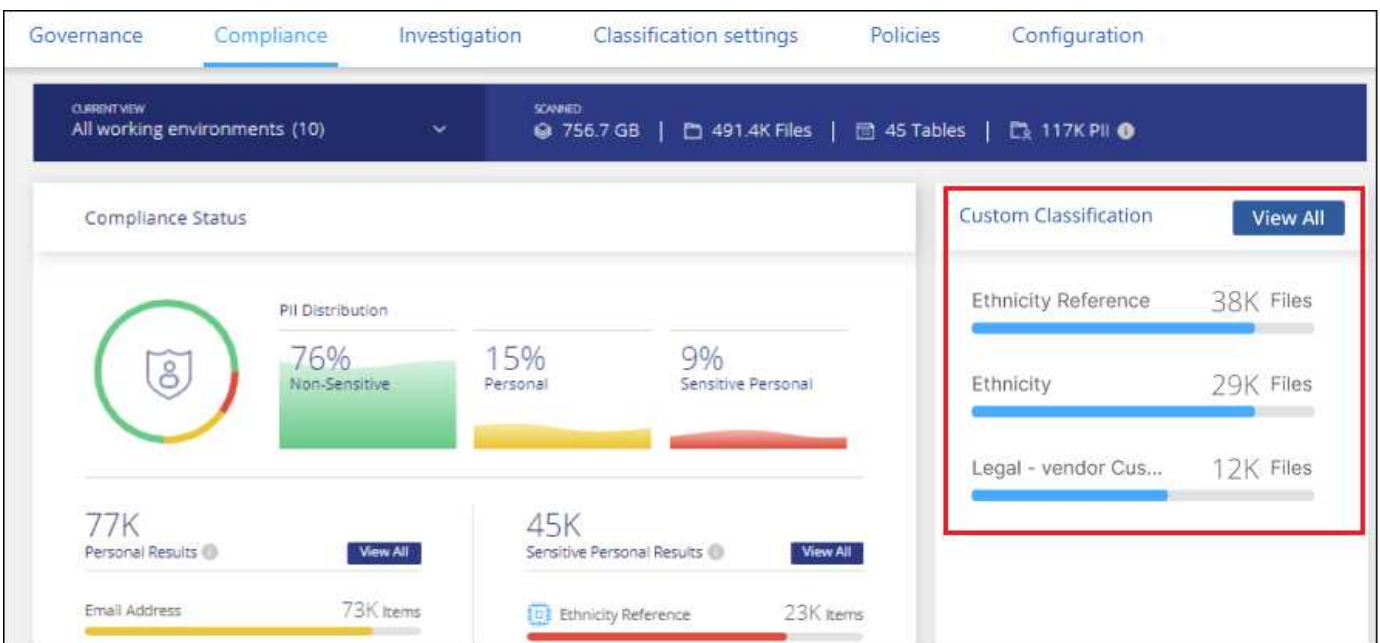
View results from your custom classifiers

You can view the results from any of your custom classifiers in the Compliance Dashboard and in the Investigation page. For example, this screenshot shows the matched information in the Compliance Dashboard under the "Personal Results" section.



Click the button to see the detailed results in the Investigation page.

Additionally, all of your custom classifier results appear in the Custom Classifiers tab, and the top 6 custom classifier results are displayed in the Compliance Dashboard, as shown below.



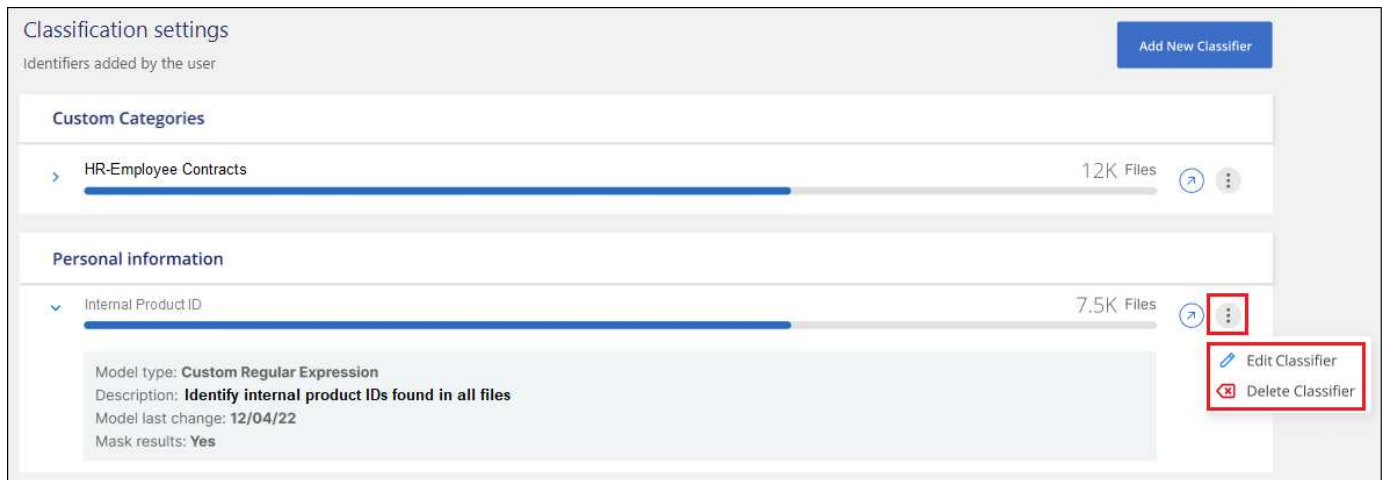
Manage custom classifiers

You can change any of the custom classifiers that you have created by using the **Edit Classifier** button.



You can't edit Data Fusion classifiers at this time.

And if you decide at some later point that you don't need BlueXP classification to identify the custom patterns that you added, you can use the **Delete Classifier** button to remove each item.



Viewing the status of your compliance actions

When you run an asynchronous action from the Investigation Results pane across many files, for example, moving or deleting 100 files, the process can take some time. You can monitor the status of these actions in the *Action Status* pane so you'll know when it has been applied to all files.

This allows you to see the actions that have completed successfully, those currently in progress, and those that have failed so you can diagnose and fix any problems. Note that short operations that complete quickly, such as moving a single file, do not appear in the Actions Status pane.

NOTE This information is relevant only for BlueXP classification legacy versions 1.30 and earlier.

The status can be:

- Success - A BlueXP classification action is finished and all items succeeded.
- Partial Success - A BlueXP classification action is finished and some items failed and some succeeded.
- In Progress - The action is still in progress.
- Queued - The action has not started.
- Canceled - The action has been canceled.
- Failed - The action failed.

Note that you can Cancel any actions that have the "Queued" or "In Progress" status.

Steps

1.

In the bottom-right of the BlueXP classification UI you can see the **Actions Status** button



2. Click this button and the most recent 20 actions are listed.

You can click the name of an action to view details corresponding to that operation.

Audit the history of BlueXP classification actions

BlueXP classification logs management activities that have been performed on files from all the working environments and data sources that BlueXP classification is scanning. BlueXP classification also logs the activities when deploying the BlueXP classification instance.

You can view the contents of the BlueXP classification audit log files, or download them, to see what file changes have occurred, and when. For example, you can see what request was issued, the time of the request, and details such as source location in case a file was deleted, or source and destination location in case a file was moved.

NOTE This information is relevant only for BlueXP classification legacy versions 1.30 and earlier.

Log file contents

Each line in the audit log contains information in this format:

```
<full date> | <status> | ds_audit_logger | <module> | 0 | 0 | File <full file path> deleted from device <device path> - <result>
```

- Date and time - full timestamp for the event
- Status - INFO, WARNING
- Action type (delete, copy, move, create policy, update policy, rescan files, download JSON report, etc.)
- File name (if the action is relevant to a file)
- Details for the action - what was done: depends on the action
 - Policy name
 - For move - Source and destination
 - For copy - Source and destination
 - For tag - tag name
 - For assign to - user name
 - For email alert - email address / account

For example, the following lines from the log file show a successful copy operation and a failed copy operation.

```
2022-06-06 15:23:08,910 | INFO | ds_audit_logger | es_scanned_file | 237 | 49 | Copy file /CIFS_share/data/dopl/random_positives.tsv from device 10.31.133.183 (type: SMB_SHARE) to device 10.31.130.133:/export_reports (NFS_SHARE) - SUCCESS
2022-06-06 15:23:08,968 | WARNING | ds_audit_logger | es_scanned_file | 239 | 153 | Copy file /CIFS_share/data/compliance-netapp.tar.gz from device 10.31.133.183 (type: SMB_SHARE) to device 10.31.130.133:/export_reports (NFS_SHARE) - FAILURE
```


Log file locations

The management audit log files are located on the BlueXP classification machine in:
`/opt/netapp/audit_logs/`

The installation audit log files are written to `/opt/netapp/install_logs/`

Each log file can be a maximum of 10 MB in size. When that limit is reached, a new log file is started. The log files are named "DataSense_audit.log", "DataSense_audit.log.1", "DataSense_audit.log.2", and so on. A maximum of 100 log files are retained on the system - older log files are deleted automatically after the maximum has been reached.

Access the log files

You'll need to log into the BlueXP classification system to access the log files. See how to [log in to the BlueXP classification system](#) depending on whether you manually installed the software on a Linux machine or if you deployed the instance in the cloud.

Reducing the BlueXP classification scan speed

Data scans have a negligible impact on your storage systems and on your data. However, if you are concerned with even a very small impact, you can configure BlueXP classification to perform "slow" scans.

When enabled, slow scanning is used on all data sources - you can't configure slow scanning for a single working environment or data source.

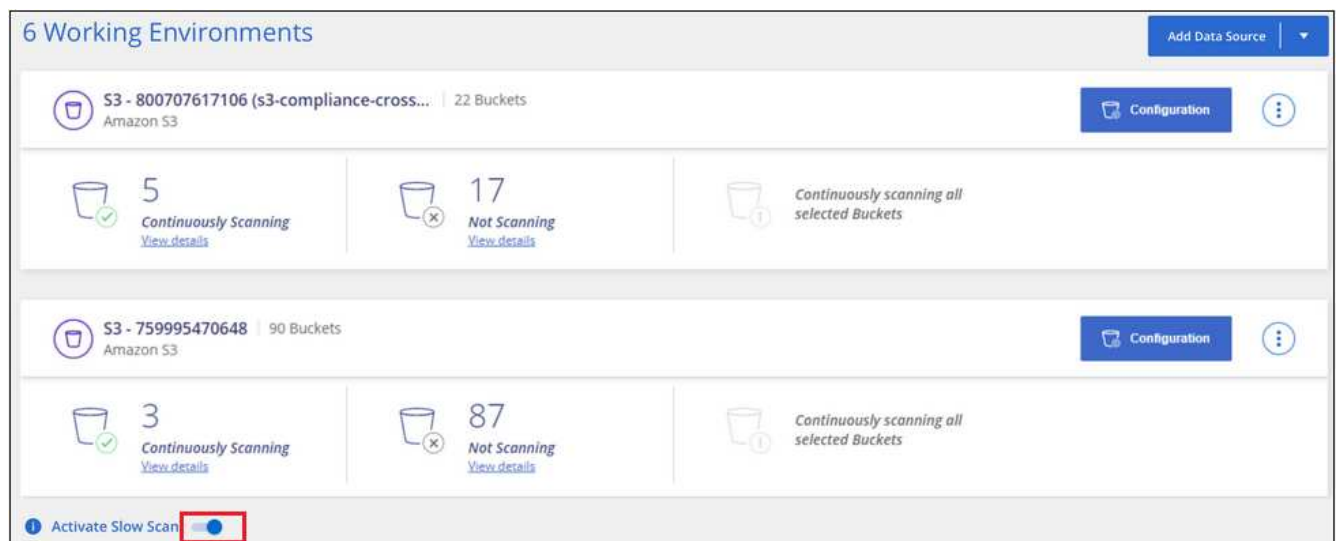


The scan speed can't be reduced when scanning databases.

NOTE This information is relevant only for BlueXP classification legacy versions 1.30 and earlier.

Steps

1. From the bottom of the *Configuration* page, move the slider to the right to activate slow scanning.



The top of the Configuration page indicates that slow scanning is enabled.




2. You can disable slow scanning by clicking **Disable** from this message.

Remove a OneDrive, SharePoint, or Google Drive account from BlueXP classification

If you no longer want to scan user files from a certain OneDrive account, from a specific SharePoint account, or from a Google Drive account, you can delete the account from the BlueXP classification interface and stop all scans.

Steps

1. From the *Configuration* page, click the  button in the row for the OneDrive, SharePoint, or Google Drive account, and then click **Remove OneDrive Account**, **Remove SharePoint Account**, or **Remove Google Drive account**.



2. Click **Delete Account** from the confirmation dialog.

Reference

Supported BlueXP classification instance types

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

You can deploy BlueXP 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 BlueXP 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

BlueXP classification collects certain metadata when performing classification scans on the data from your data sources and working environments. BlueXP 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	<i>Creation time</i>	Available	Not available (Unsupported in Linux)
	<i>Last access time</i>	Available	Available
	<i>Last modify time</i>	Available	Available
Permissions	<i>Open permissions</i>	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"
	<i>Users/group access</i>	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)



- BlueXP 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, BlueXP 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 BlueXP 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, BlueXP classification attempts to revert the last access time to the original timestamp. If BlueXP 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 BlueXP classification doesn't have these permissions, the system won't scan those 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 click the **Scan when missing "write attributes" permissions** switch at the bottom of the Configuration page so that BlueXP classification will scan the volumes regardless of permissions.

SMB_Shares Scan Configuration

2 Shares selected for Data Sense scan

Scan when missing "write" permissions

Add Shares

Edit CIFS Credentials

Scan	Storage Repository (Share)	Protocol	Access	Scan Status	Required Action
<div>Map</div> <div>Map & Classify</div>	\\10.1.7.16\CIFS_LABS_SHARE6	CIFS	<div></div> Continuously Scanning	<div></div> <div><div>Mapped: 5.8K</div><div>Classified: 5.8K</div></div>	<div></div>
<div>Map</div> <div>Map & Classify</div>	\\10.1.7.16\CIFS_LABS_SHARE7	CIFS	<div></div> Continuously Scanning	<div></div> <div><div>Mapped: 5.8K</div><div>Classified: 5.8K</div></div>	<div></div>

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

Note that there is a filter in the Investigation page called *Scan Analysis Event* that enables you to display either the files that were not classified because BlueXP classification couldn't revert the last accessed time, or the files that were classified even though BlueXP 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 BlueXP 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 the Data Investigation Report.](#)

Log in to the BlueXP classification system

At times you may need to log into the BlueXP classification system so you can access log files or edit configuration files.

When BlueXP 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 BlueXP classification is deployed in the cloud, you need to SSH to the BlueXP classification instance. You SSH to the system by entering the user and password, or by using the SSH key you provided during the BlueXP Connector 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 BlueXP instance
 - For GCP: use the user created for the BlueXP instance
- `<datasense_ip>` = IP address of the virtual machine instance

Note that you'll 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](#)

BlueXP classification APIs

The BlueXP classification capabilities that are available through the web UI are also available through the Swagger API.

There are four categories defined within BlueXP 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 create actions like copy, move, and more.

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: https://docs.netapp.com/us-en/bluexp-automation/platform/get_identifiers.html

Accessing the Swagger API reference

To get into Swagger you'll need the IP address of the your BlueXP 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 working environment 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": [
    {
```

```

    "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": "Working Environment Type",
    "operators": [
        "IN",
        "NOT_IN"
    ],
    "server_data": true,
    "type": "SELECT"
},

```

```

{
  "active_directory_affected": false,
  "data_mode": "ALL_EXTRACTABLE",
  "field": "ENVIRONMENT",
  "name": "Working Environment",
  "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, copy, assign to, FlexClone, export data, rescan, and label.

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 support

Support registration is required to receive technical support specific to BlueXP and its storage solutions and 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 BlueXP documentation for that product.

- [Amazon FSx for ONTAP](#)
- [Azure NetApp Files](#)
- [Cloud Volumes Service for Google Cloud](#)

Support registration overview

There are two forms of registration to activate support entitlement:

- Registering your BlueXP account ID support subscription (your 20 digit 960xxxxxxx serial number located on the Support Resources page in BlueXP).

This serves as your single support subscription ID for any service within BlueXP. Each BlueXP account-level support subscription must be registered.

- Registering the Cloud Volumes ONTAP serial numbers associated with a subscription in your cloud provider's marketplace (these are 20 digit 909201xxxxxxx serial numbers).

These serial numbers are commonly referred to as *PAYGO serial numbers* and get generated by BlueXP 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 BlueXP as described below.

Register your BlueXP account for NetApp support

To register for support and activate support entitlement, one user in your BlueXP account must associate a NetApp Support Site account with their BlueXP 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

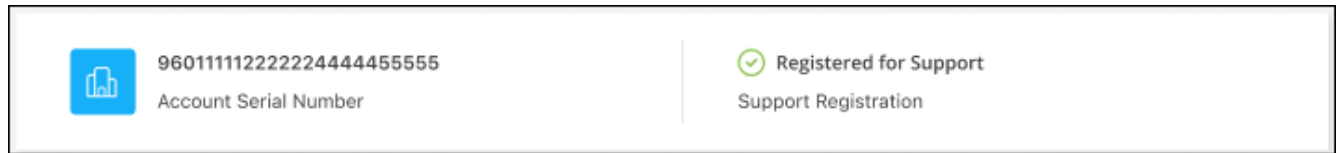
If you're a NetApp customer with an NSS account, you simply need to register for support through BlueXP.

Steps

1. In the upper right of the BlueXP console, select the Settings icon, and select **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 account is registered for support.



Note that other BlueXP users will not see this same support registration status if they have not associated a NetApp Support Site account with their BlueXP login. However, that doesn't mean that your BlueXP account is not registered for support. As long as one user in the account 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 BlueXP login.

Steps

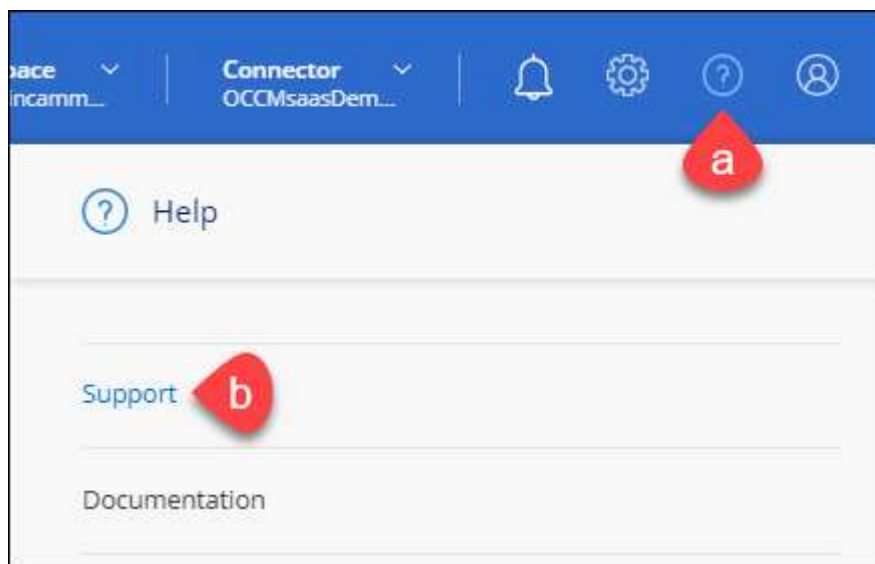
1. Create a NetApp Support Site account by completing the [NetApp Support Site User Registration form](#)
 - a. Be sure to select the appropriate User Level, which is typically **NetApp Customer/End User**.
 - b. Be sure to copy the BlueXP account serial number (960xxxx) used above for the serial number field. This will speed up the account processing.
2. Associate your new NSS account with your BlueXP 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 BlueXP console, select the Help icon, and select **Support**.



2. Locate your account ID serial number from the Support Registration page.

 96015585434285107893 Account serial number	 Not Registered Add your NetApp Support Site (NSS) credentials to BlueXP Follow these instructions to register for support in case you don't have an NSS account yet.
--	--

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 the account 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 BlueXP 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 BlueXP 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 BlueXP 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 BlueXP account is different than the NSS account that is associated with a BlueXP user login.

These NSS credentials are associated with your specific BlueXP account ID. Users who belong to the BlueXP account 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 BlueXP 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 BlueXP 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

NetApp provides support for BlueXP and its cloud services in a variety of ways. Extensive free self-support options are available 24x7, such as knowledgebase (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 "Getting help" in the BlueXP documentation for that product.

- [Amazon FSx for ONTAP](#)
- [Azure NetApp Files](#)
- [Cloud Volumes Service for Google Cloud](#)

To receive technical support specific to BlueXP and its storage solutions and 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 BlueXP documentation that you're currently viewing.

- [Knowledge base](#)

Search through the BlueXP knowledge base to find helpful articles to troubleshoot issues.

- [Communities](#)

Join the BlueXP 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 BlueXP login. [Learn how to manage credentials associated with your BlueXP 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 BlueXP, 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, BlueXP when specific to a technical support issue with workflows or functionality within the service.
 - **Working Environment:** If applicable to storage, select **Cloud Volumes ONTAP** or **On-Prem** and then the associated working environment.


The list of working environments are within scope of the BlueXP account, workspace, and Connector you have selected in the top banner of the service.

- **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.

ntapitdemo 


NetApp Support Site Account

Service

Select ▼

Working Enviroment


Select ▼

Case Priority 

Low - General guidance ▼

Issue Description



Provide detailed description of problem, applicable error messages and troubleshooting steps taken.



Additional Email Addresses (Optional) 

Type here

Attachment (Optional)

No files selected

 Upload 

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 BlueXP account serial number (ie. 960xxxx) or the working environment serial number. You can seek assistance using one of the following options:

- Use the in-product chat
- Submit a non-technical case at <https://mysupport.netapp.com/site/help>

Manage your support cases (Preview)

You can view and manage active and resolved support cases directly from BlueXP. You can manage the cases associated with your NSS account and with your company.

Case management is available as a Preview. We plan to refine this experience and add enhancements in upcoming releases. Please send us feedback by using the in-product chat.

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 BlueXP, select **Help > Support**.
2. Select **Case Management** and if you're prompted, add your NSS account to BlueXP.

The **Case management** page shows open cases related to the NSS account that is associated with your BlueXP 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.

Search icon | Cases opened on the last 3 months | Create a case

Date created	Last updated		Status (5)	
December 22, 2022	December 29, 2022	Last 7 days	Assigned	...
December 21, 2022	December 28, 2022	Last 30 days	Active	...
December 15, 2022	December 27, 2022	Last 3 months	Pending customer	...
December 14, 2022	December 26, 2022	Medium (P3)	Solution proposed	...
		Low (P4)		


Apply | Reset

- Filter the contents of the columns.

Search icon | Cases opened on the last 3 months | Create a case

Last updated	Priority	Status (5)	
December 29, 2022	Critical (P1)	Active	...
December 28, 2022	High (P2)	Pending customer	...
December 27, 2022	Medium (P3)	Solution proposed	...
December 26, 2022	Low (P4)	Pending closed	...
		Closed	...

Apply | Reset

- Change the columns that appear in the table by selecting  and then choosing the columns that you'd like to display.

Search icon | Cases opened on the last 3 months | Create a case

Last updated	Priority	Status (5)	
December 29, 2022	Critical (P1)	Last updated	...
December 28, 2022	High (P2)	Priority	...
December 27, 2022	Medium (P3)	Cluster name	...
December 26, 2022	Low (P4)	Case owner	...
		Opened by	...

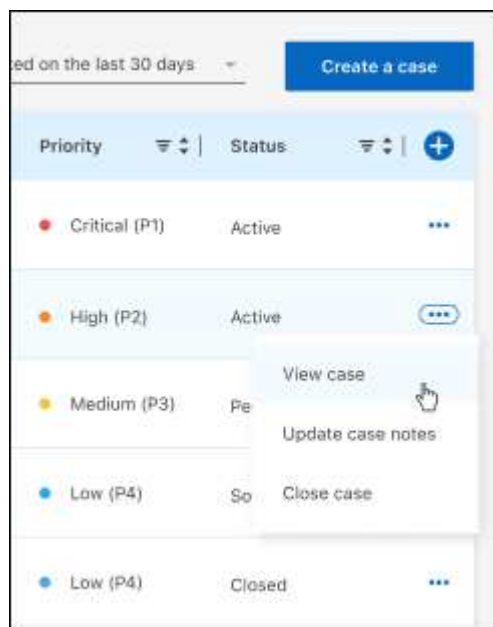
Apply | Reset

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**.



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 BlueXP](#)
- [Notice for BlueXP classification](#)

Copyright information

Copyright © 2024 NetApp, Inc. All Rights Reserved. Printed in the U.S. No part of this document covered by copyright may be reproduced in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or storage in an electronic retrieval system—without prior written permission of the copyright owner.

Software derived from copyrighted NetApp material is subject to the following license and disclaimer:

THIS SOFTWARE IS PROVIDED BY NETAPP “AS IS” AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, WHICH ARE HEREBY DISCLAIMED. IN NO EVENT SHALL NETAPP BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

NetApp reserves the right to change any products described herein at any time, and without notice. NetApp assumes no responsibility or liability arising from the use of products described herein, except as expressly agreed to in writing by NetApp. The use or purchase of this product does not convey a license under any patent rights, trademark rights, or any other intellectual property rights of NetApp.

The product described in this manual may be protected by one or more U.S. patents, foreign patents, or pending applications.

LIMITED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (b)(3) of the Rights in Technical Data -Noncommercial Items at DFARS 252.227-7013 (FEB 2014) and FAR 52.227-19 (DEC 2007).

Data contained herein pertains to a commercial product and/or commercial service (as defined in FAR 2.101) and is proprietary to NetApp, Inc. All NetApp technical data and computer software provided under this Agreement is commercial in nature and developed solely at private expense. The U.S. Government has a non-exclusive, non-transferrable, nonsublicensable, worldwide, limited irrevocable license to use the Data only in connection with and in support of the U.S. Government contract under which the Data was delivered. Except as provided herein, the Data may not be used, disclosed, reproduced, modified, performed, or displayed without the prior written approval of NetApp, Inc. United States Government license rights for the Department of Defense are limited to those rights identified in DFARS clause 252.227-7015(b) (FEB 2014).

Trademark information

NETAPP, the NETAPP logo, and the marks listed at <http://www.netapp.com/TM> are trademarks of NetApp, Inc. Other company and product names may be trademarks of their respective owners.