Get started

BlueXP classification

NetApp
July 25, 2024

Table of Contents

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
- 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.

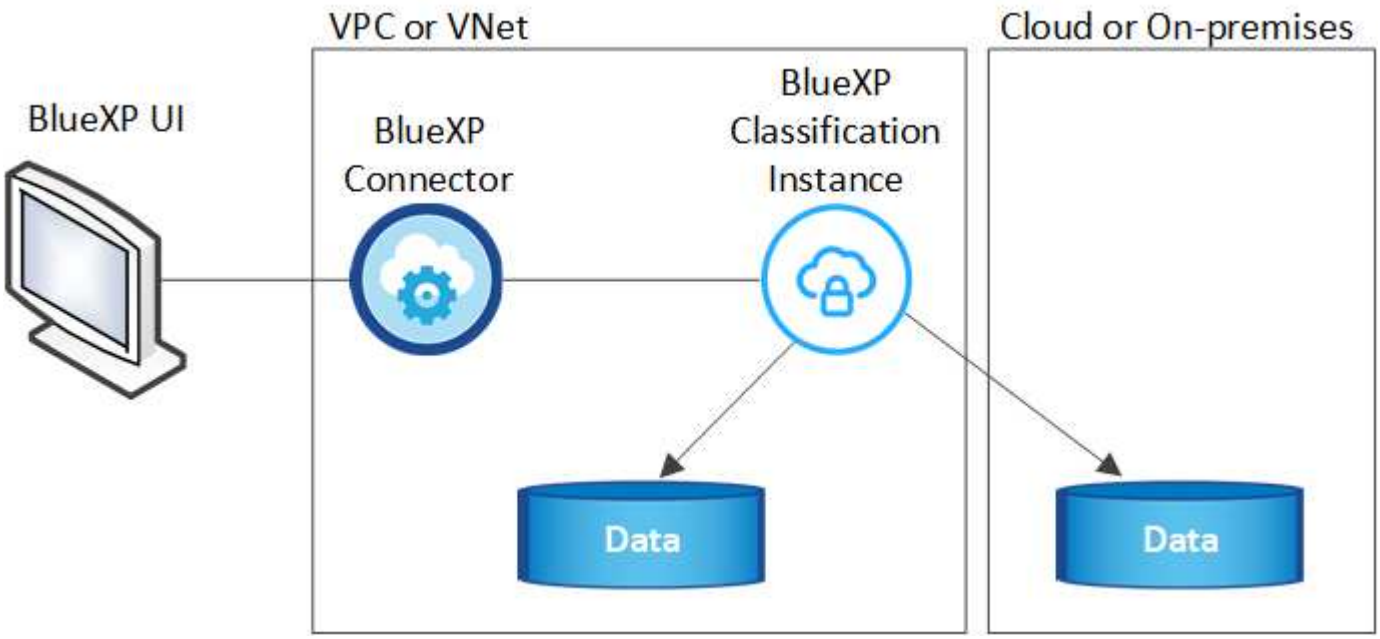
Data transfer costs

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: *CloudCompliance-16bb6564-38ad-4080-9a92-36f5fd2f71c7*
- 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

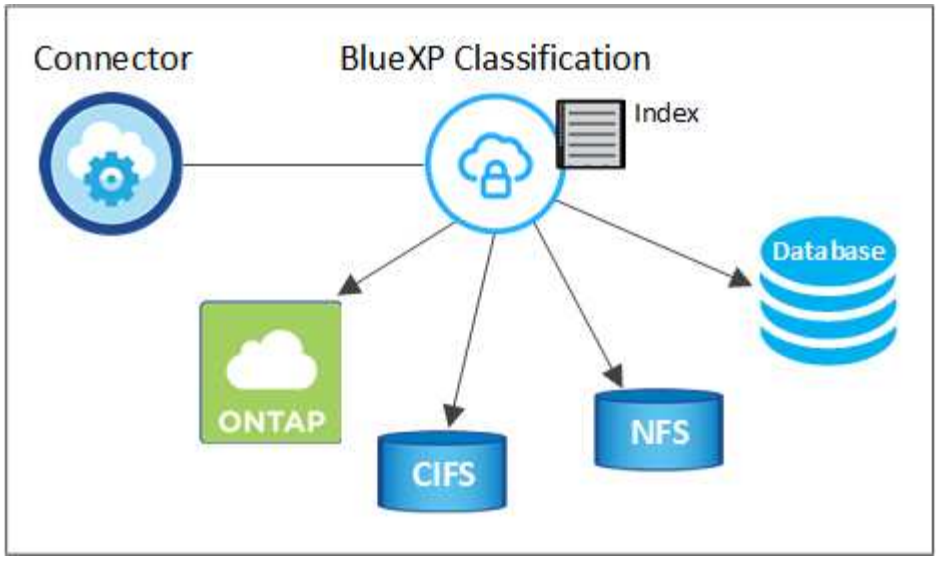
At a high-level, BlueXP classification works like this:

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

Feature	Map & Classify	Map
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 indexes

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

- **Standard metadata** BlueXP classification collects 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 access to compliance information

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.

3 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 you 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://dseasb33srnrn.cloudfront.net/ https://production.cloudflare.docker.com/	Provides access to software images, manifests, and templates.
https://kinesis.us-east-1.amazonaws.com	Enables NetApp to stream data from audit records.
https://cognito-idp.us-east-1.amazonaws.com https://cognito-identity.us-east-1.amazonaws.com https://user-feedback-store-prod.s3.us-west-2.amazonaws.com https://customer-data-production.s3.us-west-2.amazonaws.com	Enables 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://dseasb33srnrn.cloudfront.net/ https://production.cloudflare.docker.com/	Provides access to software images, manifests, templates, and to send logs and metrics.
https://support.compliance.api.bluexp.netapp.com/	Enables NetApp to stream data from audit records.

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://dseasb33srnrn.cloudfront.net/ https://production.cloudflare.docker.com/	Provides access to software images, manifests, templates, and to send logs and metrics.
https://support.compliance.api.bluexp.netapp.com/	Enables NetApp to stream data from audit records.

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

Note that you can deploy BlueXP classification on an instance in AWS cloud environments with fewer CPUs and less RAM, but there are limitations when using these systems. See Using a smaller instance type for details.

Deploy BlueXP classification in the cloud

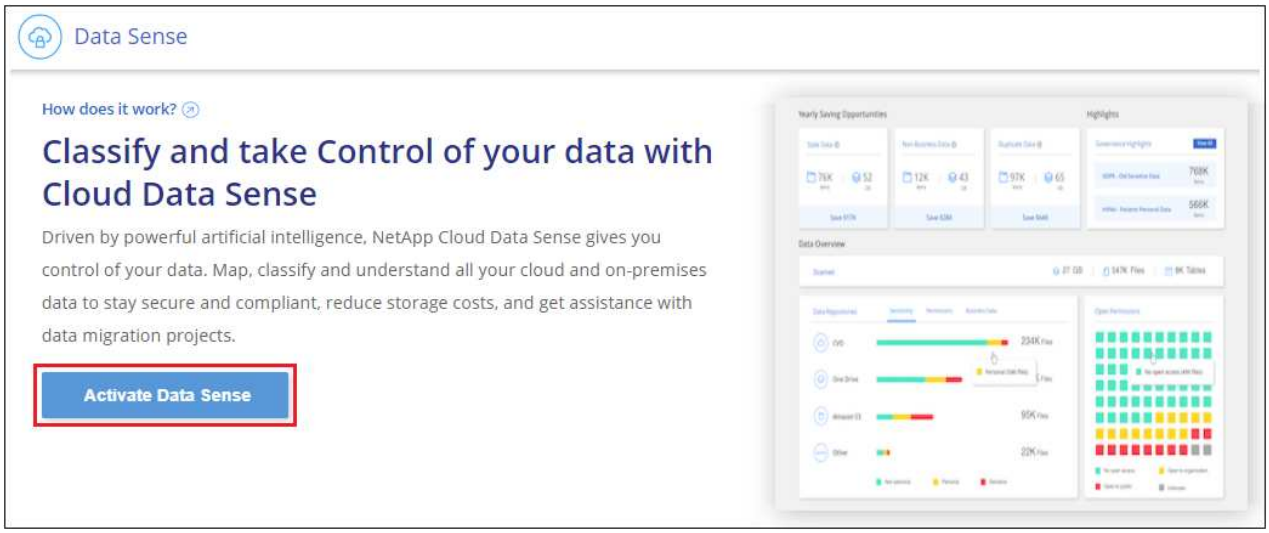
Follow these steps to deploy an instance of BlueXP classification in the cloud. The Connector will deploy the instance in the cloud, and then install BlueXP classification software on that instance.

Note that when deploying BlueXP classification from a BlueXP Connector in an AWS environment, you can select the default instance size or you can select from two smaller instance types. See the available instance types and limitations. 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**.



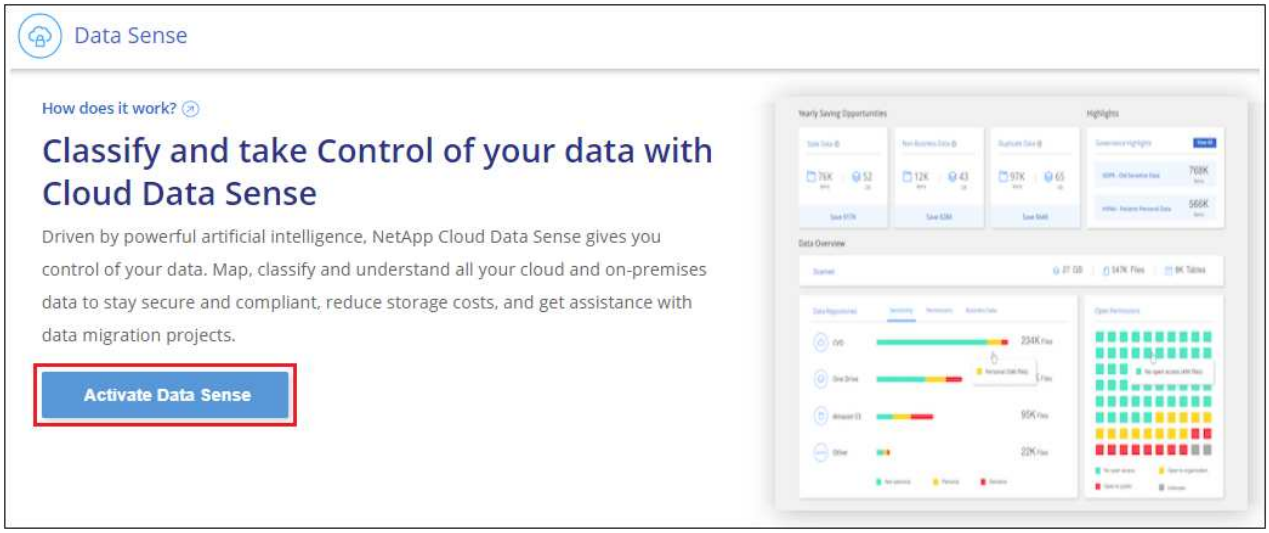
2. Click **Activate Data Sense**.
3. From the *Installation* page, click **Deploy > Deploy** to use the "Large" instance size and start the cloud deployment wizard.
4. The wizard displays progress as it goes through the deployment steps. It will stop and prompt for input if it runs into any issues.

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

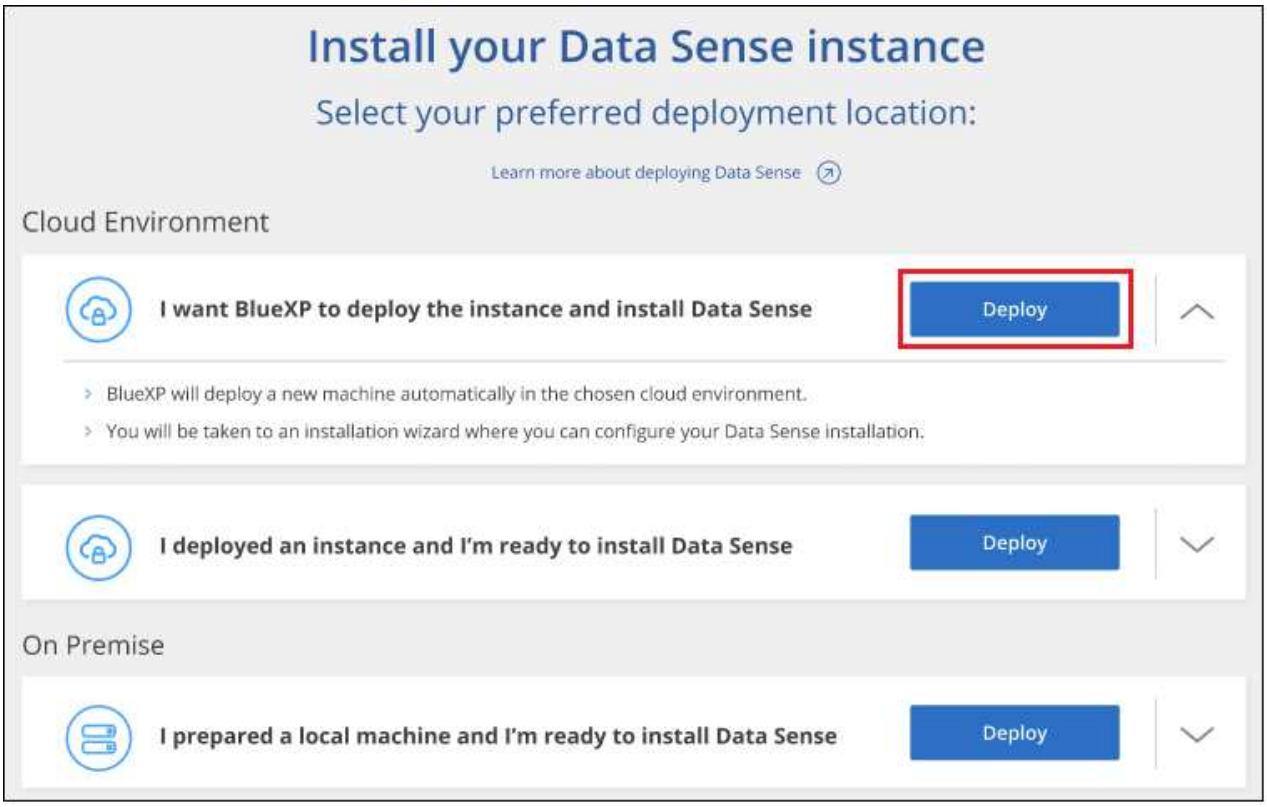
Deploy in Azure

Steps

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



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



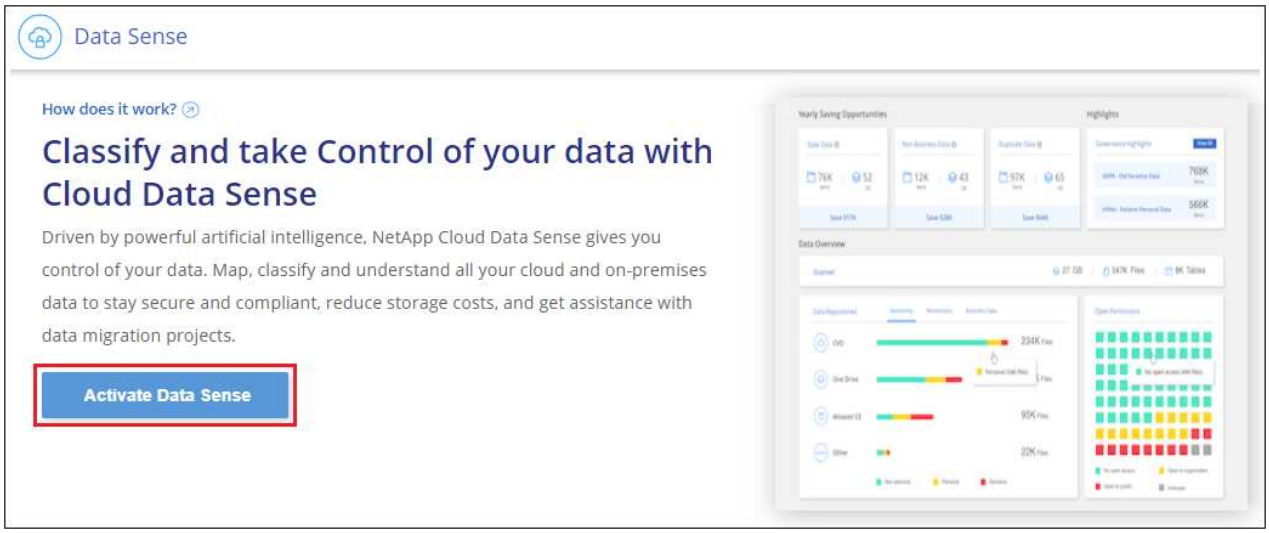
4. The wizard displays progress as it goes through the deployment steps. It will stop and prompt for input if it runs into any issues.

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

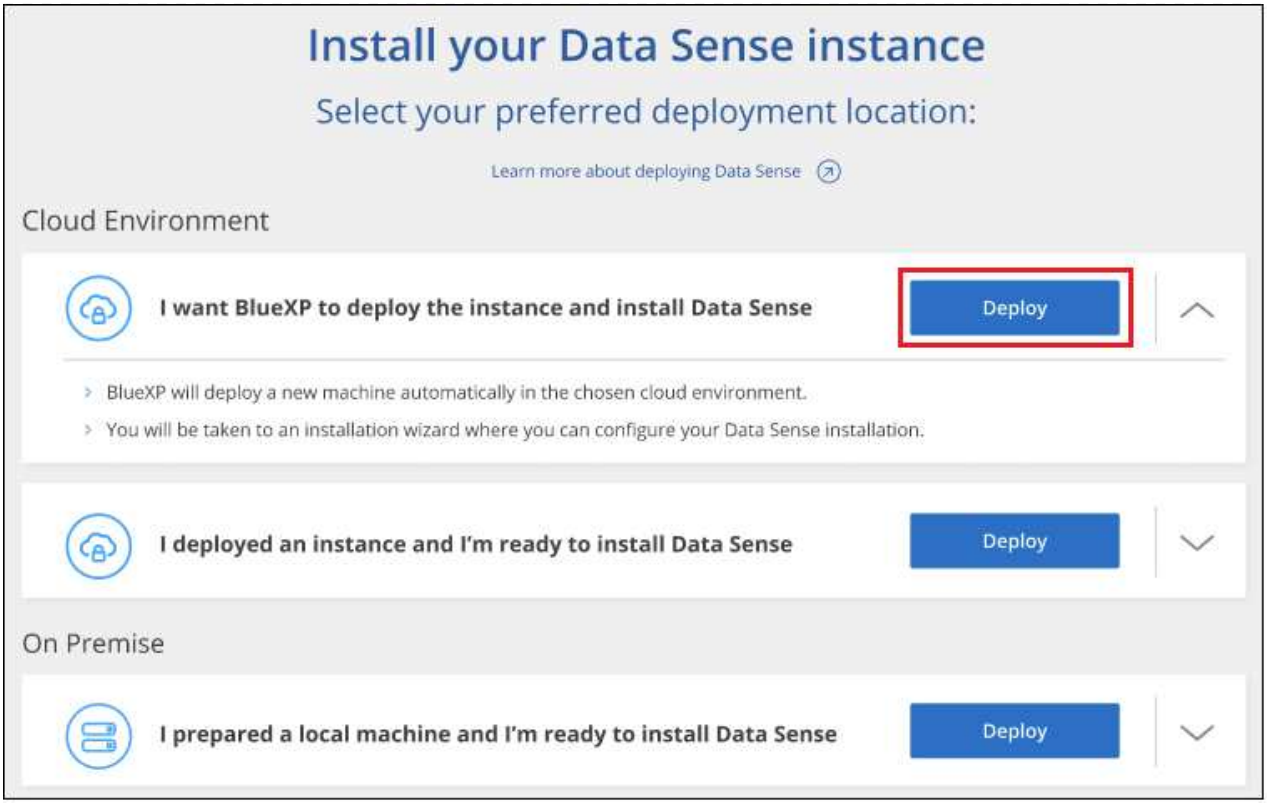
Deploy in Google Cloud

Steps

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



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



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

5. When the instance is deployed and 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 you 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
    - CentOS version 7.8 and 7.9
    - Ubuntu 22.04 (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://dseasb33srnrn.cloudfront.net/ https://production.cloudflare.docker.com/	Provides access to software images, manifests, templates, and to send logs and metrics.
https://support.compliance.api.bluexp.netapp.com/	Enables NetApp to stream data from audit records.
https://github.com/docker https://download.docker.com	Provides prerequisite packages for docker installation.
http://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

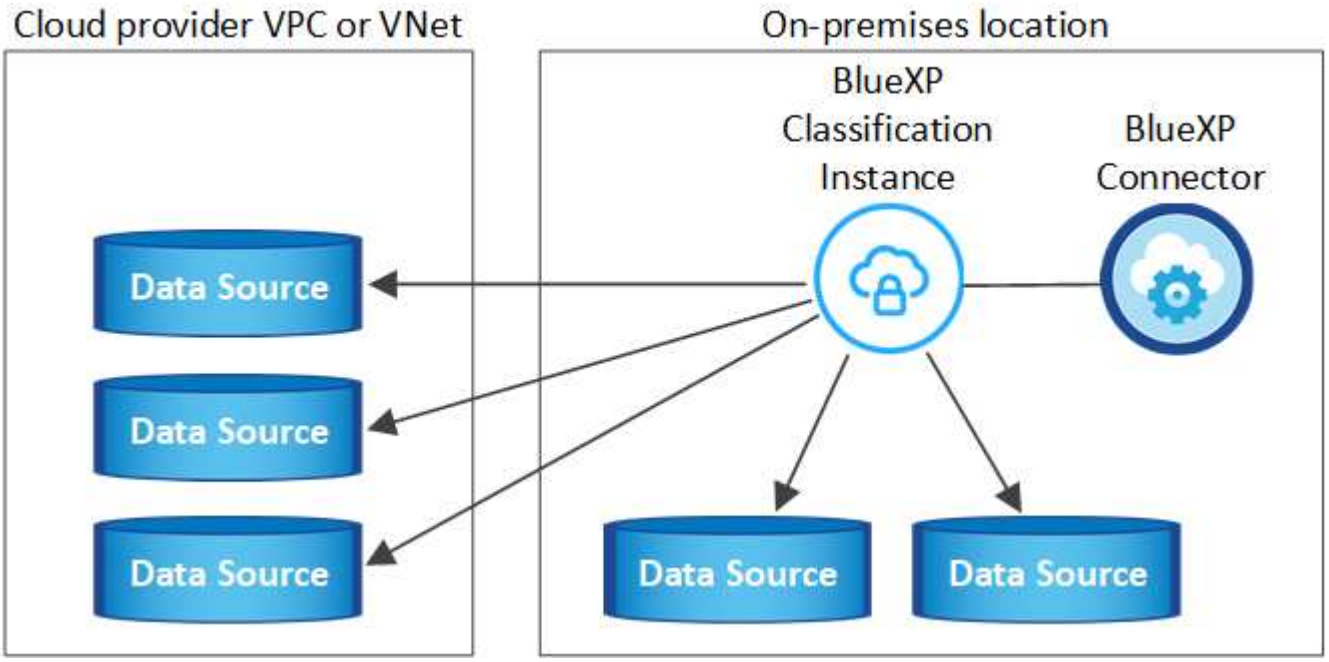
You must ensure that all required ports are open for communication between the Connector, BlueXP classification, Active Directory, and your data sources.

Connection Type	Ports	Description
Connector <> BlueXP classification	8080 (TCP), 443 (TCP), and 80	The firewall or routing rules for the Connector must allow inbound and outbound traffic over port 443 to and from the BlueXP classification instance. Make sure port 8080 is open so you can see the installation progress in BlueXP.

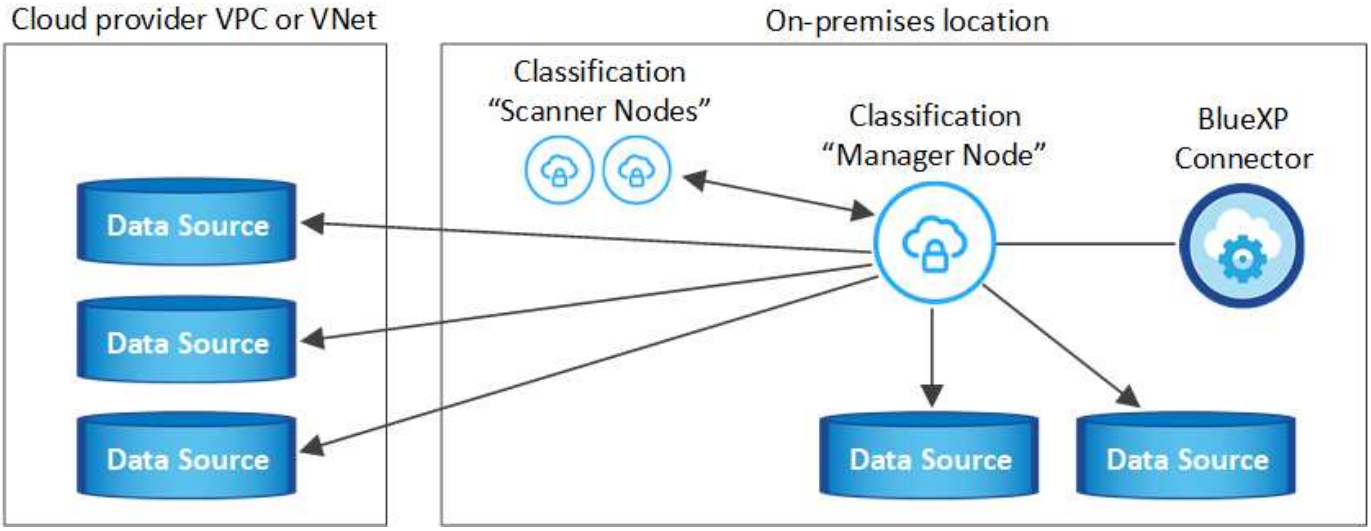
Connection Type	Ports	Description
Connector <> ONTAP cluster (NAS)	443 (TCP)	BlueXP discovers ONTAP clusters using HTTPS. If you use custom firewall policies, they must meet the following requirements: • The Connector host must allow outbound HTTPS access through port 443. If the Connector is in the 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	• For NFS - 111 (TCP\UDP) and 2049 (TCP\UDP) • For CIFS - 139 (TCP\UDP) and 445 (TCP\UDP)	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. Make sure these ports are open to the BlueXP classification instance: • For NFS - 111 and 2049 • For CIFS - 139 and 445 NFS volume export policies must allow access from the BlueXP classification instance.
BlueXP classification <> Active Directory	389 (TCP & UDP), 636 (TCP), 3268 (TCP), and 3269 (TCP)	You must have an Active Directory already set up for the users in your company. Additionally, BlueXP classification needs Active Directory credentials to scan CIFS volumes. You must have the information for the Active Directory: • DNS Server IP Address, or multiple IP Addresses • User Name and Password for the server • Domain Name (Active Directory Name) • Whether you are using secure LDAP (LDAPS) or not • LDAP Server Port (typically 389 for LDAP, and 636 for secure LDAP)

Install 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

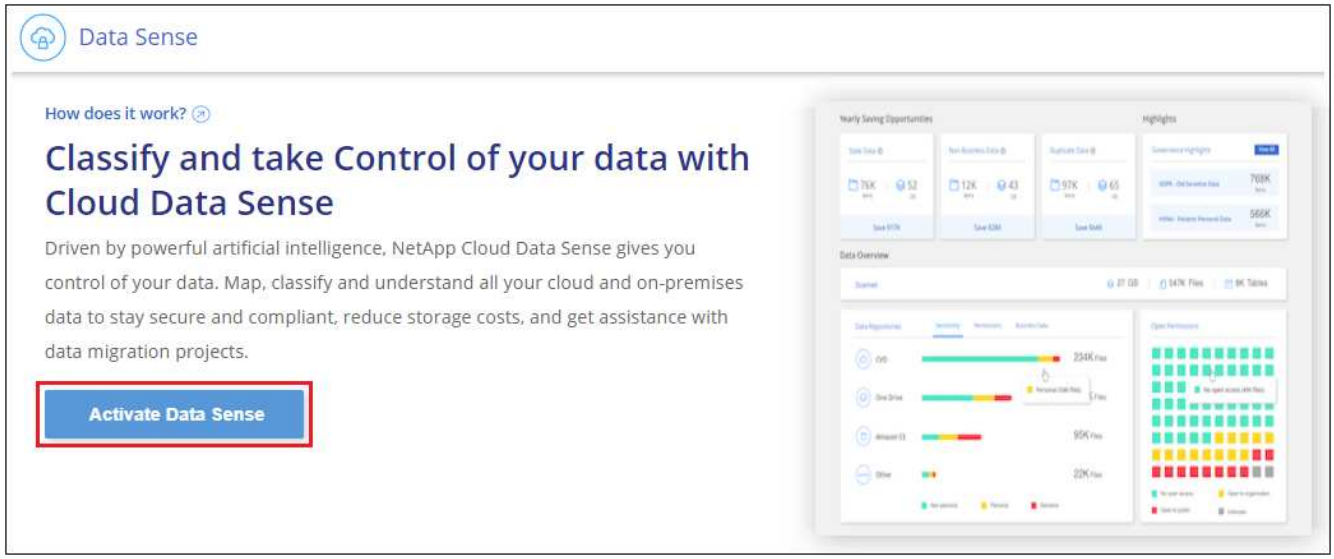
- Verify that your Linux system meets the host requirements.
- Verify that the system has the two prerequisite software packages installed (Docker Engine or Podman, and Python 3).
- Make sure you have root privileges on the Linux system.
- If you're using a proxy for access to the internet:
  - You'll 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

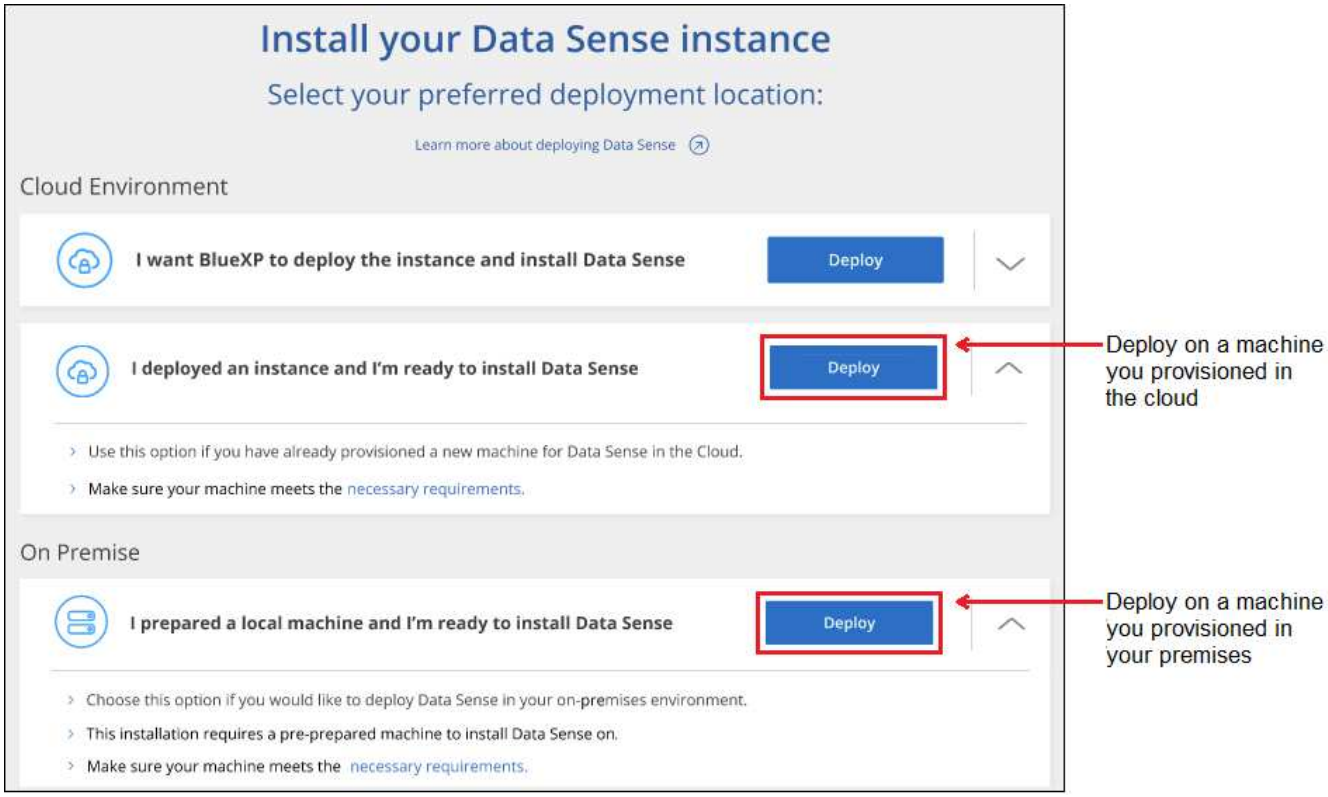
1. Download the BlueXP classification software from the NetApp Support Site. The file you should select is named **DATASENSE-INSTALLER-<version>.tar.gz**.
2. Copy the installer file to the Linux host you plan to use (using `scp` or some other method).
3. Unzip the installer file on the host machine,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

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

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



6. 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.



7. 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.

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

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

Enter parameters as prompted:	Enter the full command:
1. Paste the command you copied from step 7: `sudo ./install.sh -a <account_id> -c <client_id> -t <user_token>` If you are installing on a cloud instance (not on your premises), add `--manual-cloud -install <cloud_provider>`. 2. Enter the IP address or host name of the 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.	Alternatively, you can create the whole command in advance, providing the necessary host and proxy parameters: `sudo ./install.sh -a <account_id> -c <client_id> -t <user_token> --host <ds_host> --manager-host <cm_host> --manual-cloud-install <cloud_provider> --proxy-host <proxy_host> --proxy-port <proxy_port> --proxy-scheme <proxy_scheme> --proxy -user <proxy_user> --proxy-password <proxy_password> --cacert-folder-path <ca_cert_dir>`

Variable values:

- *account_id* = NetApp Account ID
- *client_id* = 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
    - CentOS version 7.8 and 7.9
    - Ubuntu 22.04 (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.

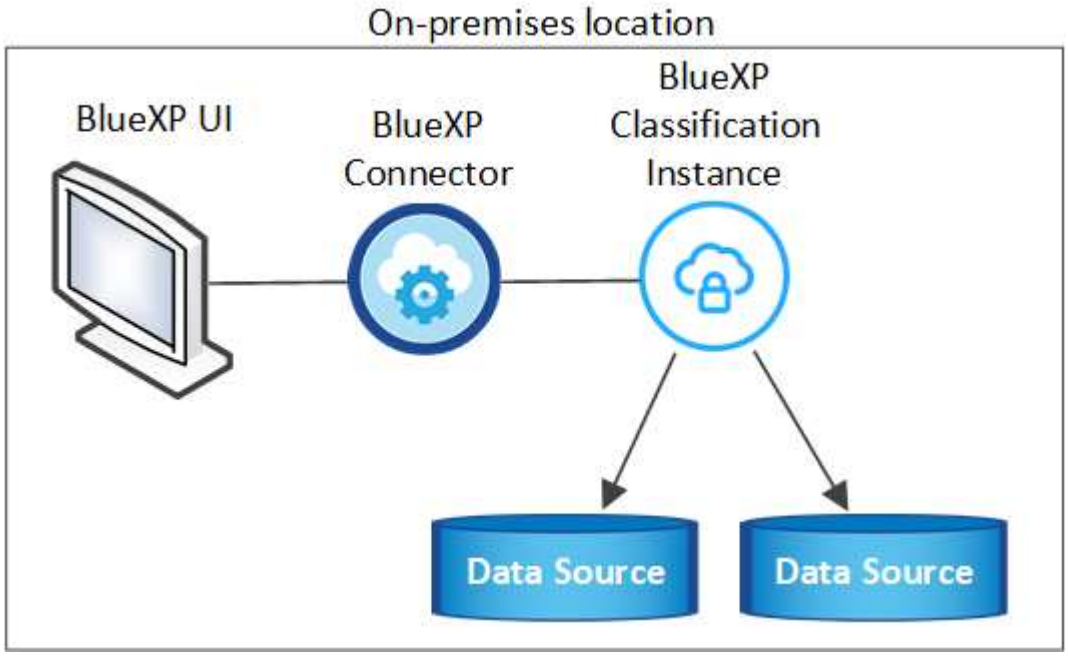
Connection Type	Ports	Description
Connector <> BlueXP classification	8080 (TCP), 6000 (TCP), 443 (TCP), and 80	The security group for the Connector must allow inbound and outbound traffic over ports 6000 and 443 to and from the BlueXP classification instance. • Port 6000 is required so that the BlueXP classification BYOL license works in a dark site. • Port 8080 should be open so you can see the installation progress in BlueXP.

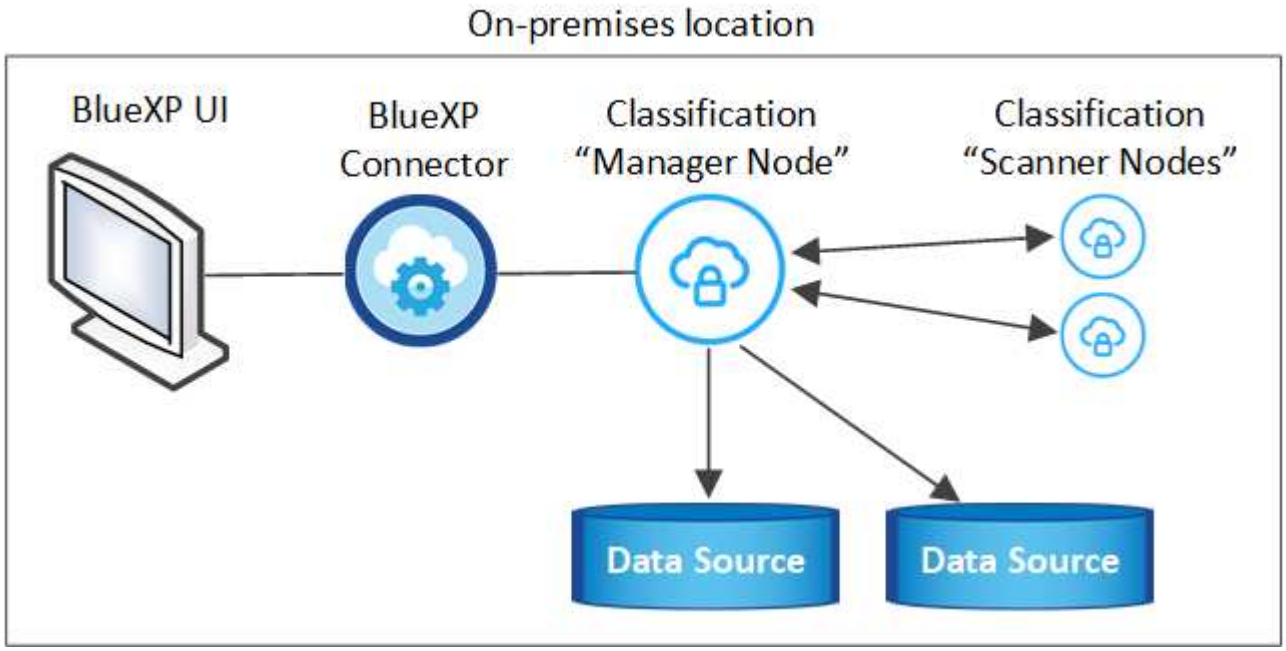
Connection Type	Ports	Description
Connector <> ONTAP cluster (NAS)	443 (TCP)	BlueXP discovers ONTAP clusters using HTTPS. If you use custom firewall policies, they must meet the following requirements: • The Connector host must allow outbound HTTPS access through port 443. If the Connector is in the 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	• For NFS - 111 (TCP\UDP) and 2049 (TCP\UDP) • For CIFS - 139 (TCP\UDP) and 445 (TCP\UDP)	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. Make sure these ports are open to the BlueXP classification instance: • For NFS - 111 and 2049 • For CIFS - 139 and 445 NFS volume export policies must allow access from the BlueXP classification instance.
BlueXP classification <> Active Directory	389 (TCP & UDP), 636 (TCP), 3268 (TCP), and 3269 (TCP)	You must have an Active Directory already set up for the users in your company. Additionally, BlueXP classification needs Active Directory credentials to scan CIFS volumes. You must have the information for the Active Directory: • DNS Server IP Address, or multiple IP Addresses • User Name and Password for the server • Domain Name (Active Directory Name) • Whether you are using secure LDAP (LDAPS) or not • LDAP Server Port (typically 389 for LDAP, and 636 for secure LDAP)

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.





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:

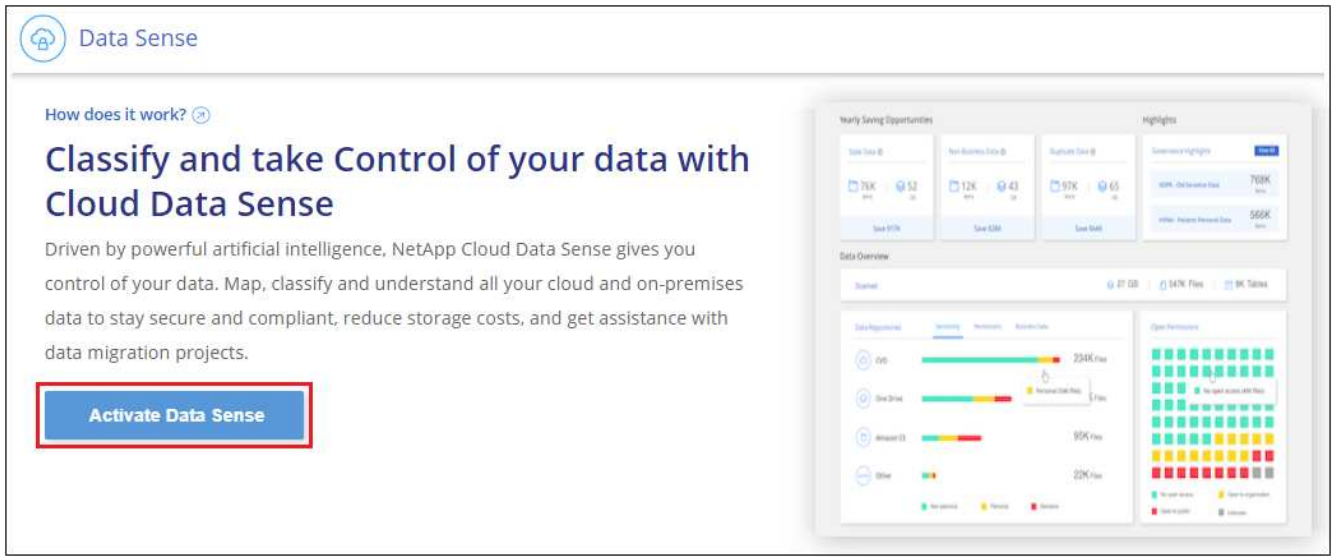
```
tar -xzf DataSense-offline-bundle-v1.25.0.tar.gz
```

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

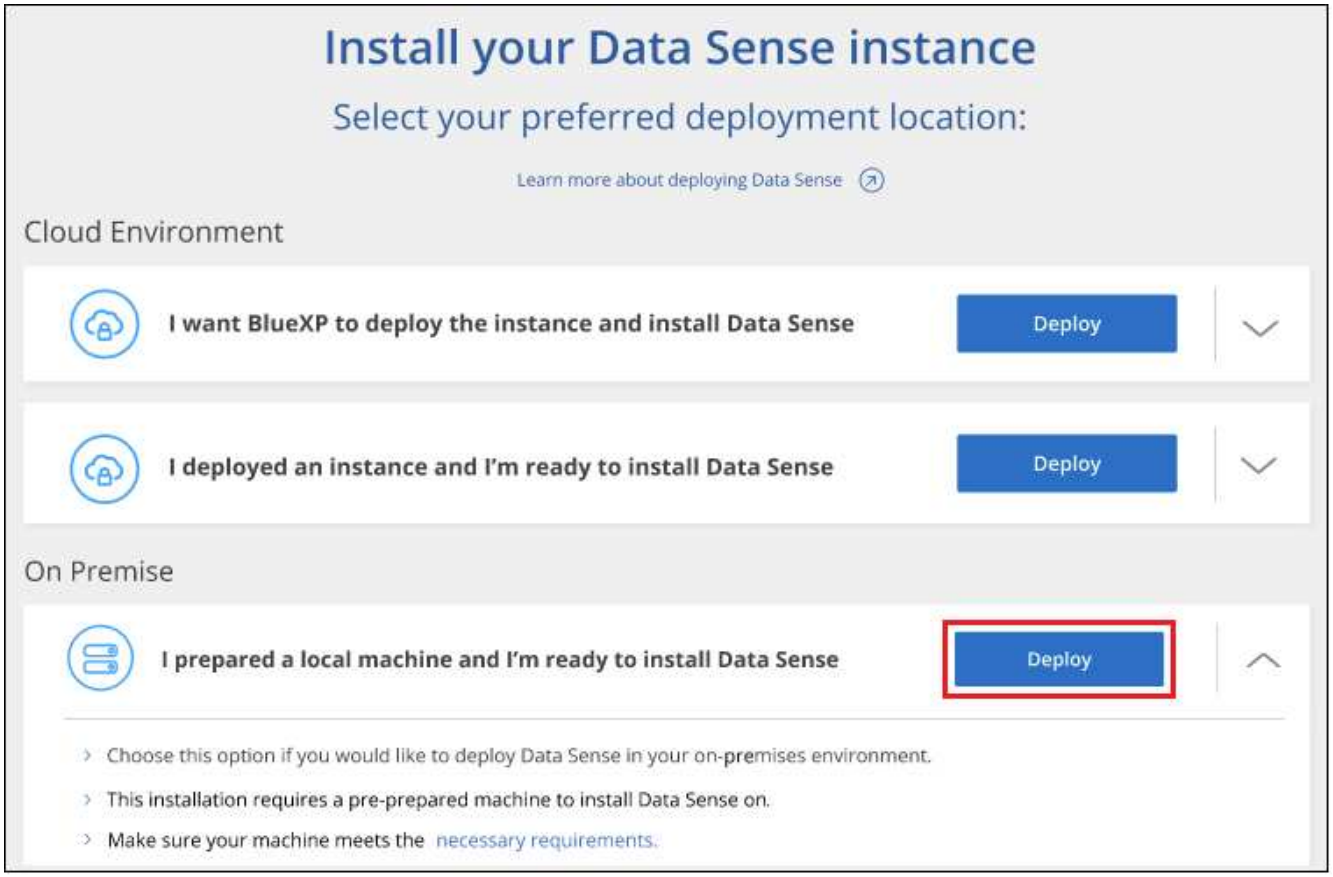
4. Unzip the installation file on the host machine, for example:

```
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:
1. Paste the information you copied from step 8: `sudo ./install.sh -a <account_id> -c <client_id> -t <user_token> --darksite` 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: `sudo ./install.sh -a <account_id> -c <client_id> -t <user_token> --host <ds_host> --manager-host <cm_host> --no-proxy --darksite`

   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
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

        - CentOS version 7.8 and 7.9
        - Ubuntu 22.04 (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
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://dseasb33srnrn.cloudfront.net/ https://production.cloudflare.docker.com/	Provides access to software images, manifests, templates, and to send logs and metrics.
https://support.compliance.api.bluexp.netapp.com/	Enables NetApp to stream data from audit records.
https://github.com/docker https://download.docker.com	Provides prerequisite packages for docker installation.
http://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	The firewall or routing rules for the Connector must allow inbound and outbound traffic over port 443 to and from the BlueXP classification instance. Make sure port 8080 is open so you can see the installation progress in BlueXP.

Connection Type	Ports	Description
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

Getting started with BlueXP classification for Cloud Volumes ONTAP and on-premises ONTAP

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.

Discovering 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.

Deploying 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.

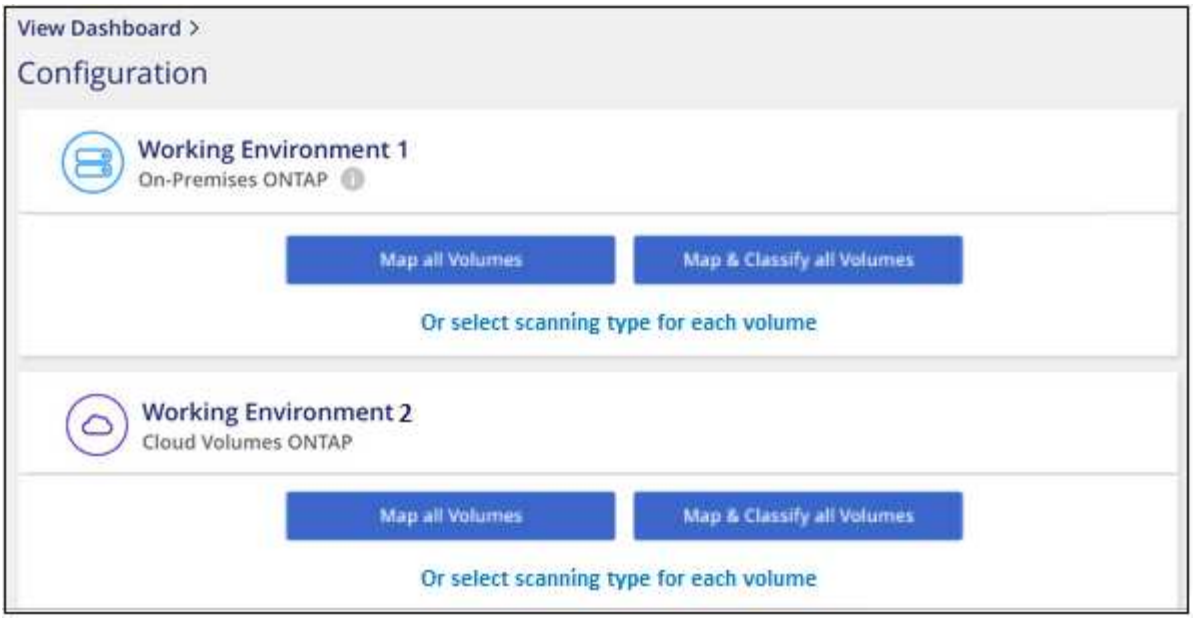
If you are scanning on-premises ONTAP systems that have been installed in a dark site that has no internet access, you need to deploy BlueXP classification in the same on-premises location that has no internet access. This also requires that the BlueXP Connector is deployed in that same on-premises location.

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

Enabling 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 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.

Verifying that BlueXP classification has access to volumes

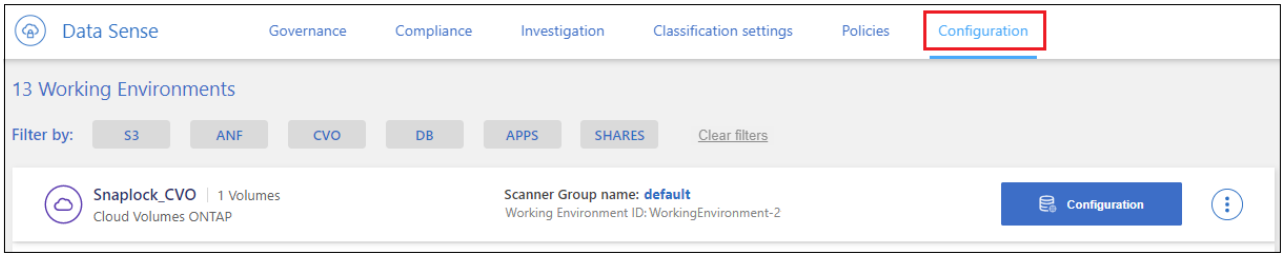
Make sure that BlueXP classification can access volumes by checking your networking, security groups, and export policies. You'll need to provide BlueXP classification with CIFS credentials so it can access CIFS volumes.

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.

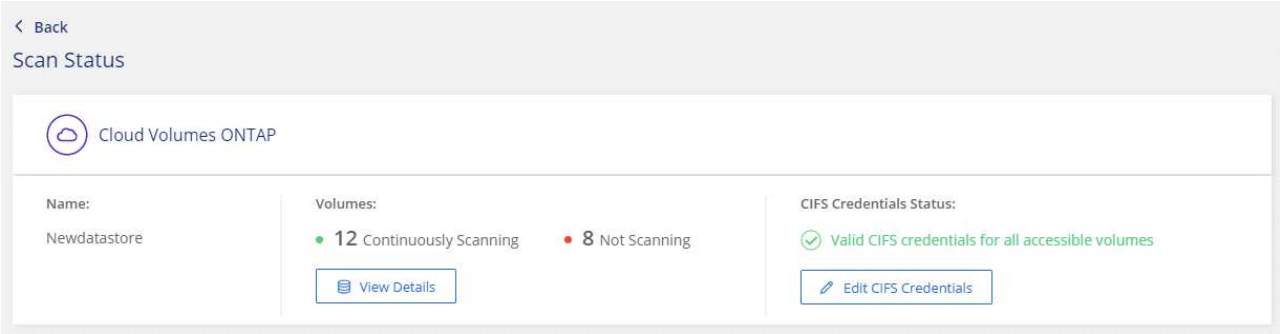


   b. For each working environment, click **Edit CIFS Credentials** and enter the user name and password that BlueXP classification needs to access CIFS volumes on the system.

      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.

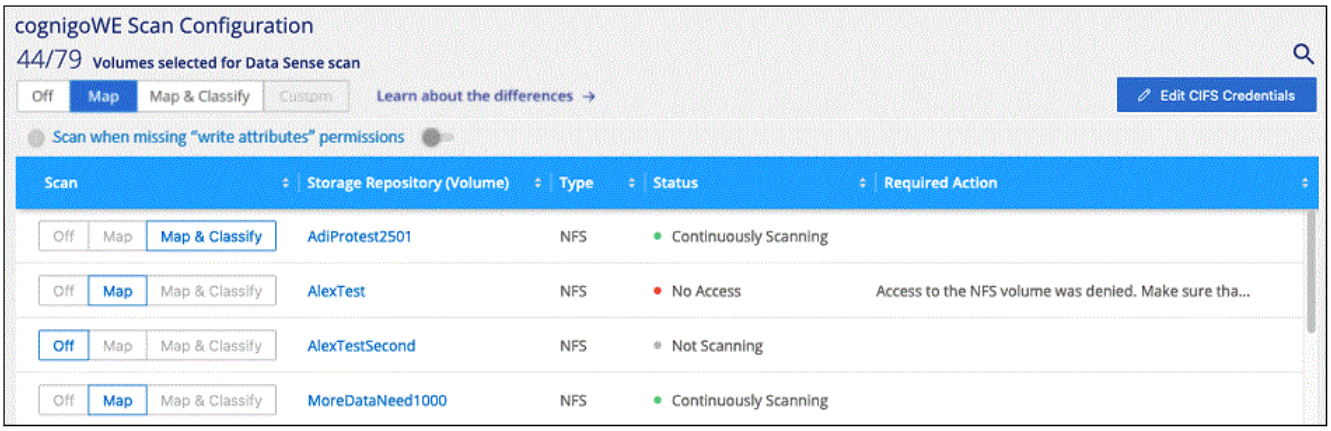
      If you want to make sure your files "last accessed times" are unchanged by BlueXP classification scans, we recommend that the user has Write Attributes permissions in CIFS or write permissions in NFS. If possible, we recommend making the Active Directory configured user part of a parent group in the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

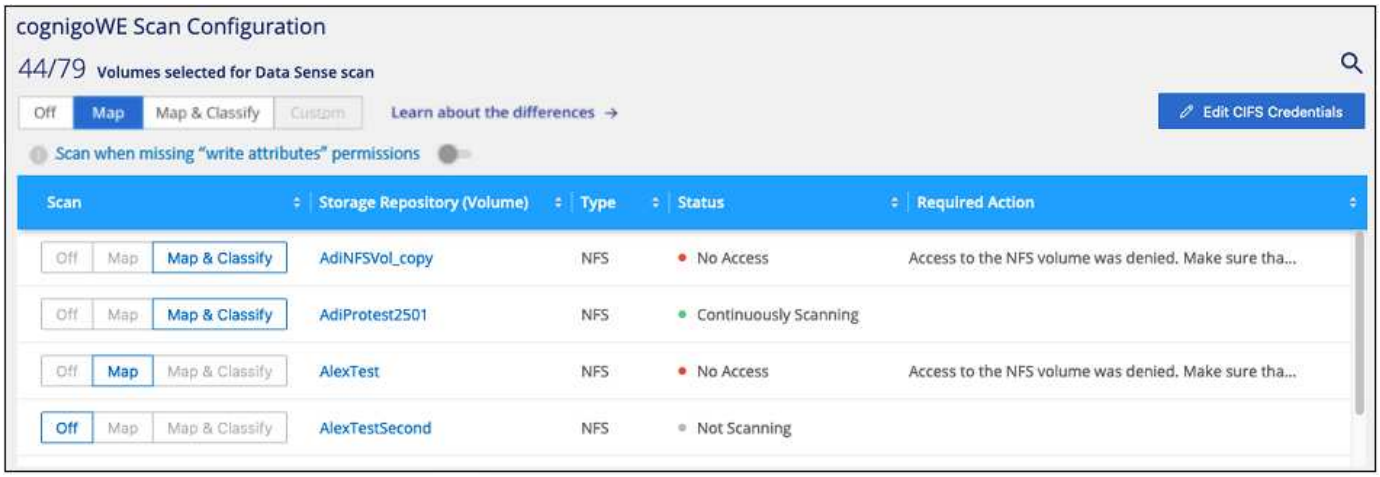
   For example, the following image shows four volumes; one of which BlueXP classification can't scan due to network connectivity issues between the BlueXP classification instance and the volume.



Enabling and disabling 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.



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**

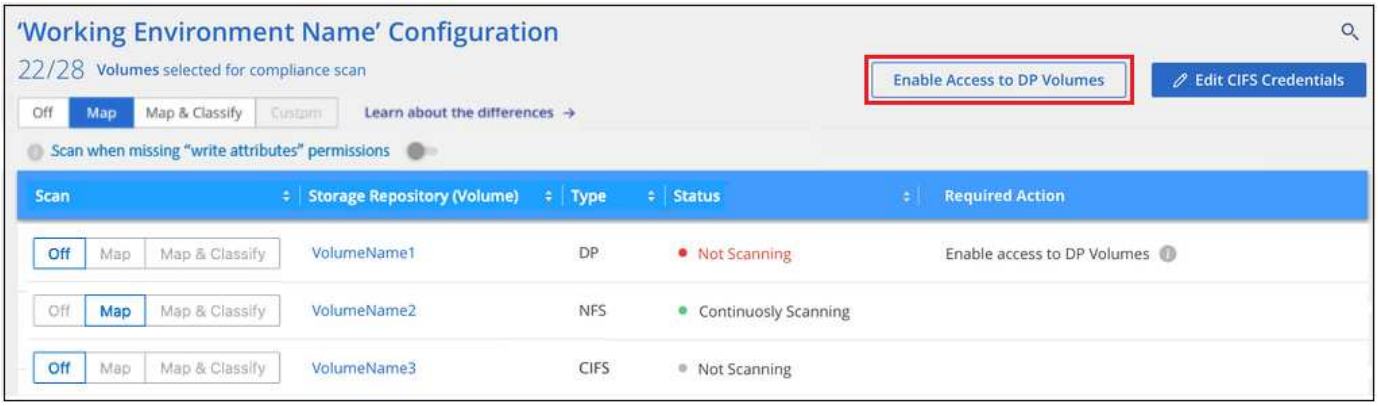
To:	Do this:
Enable mapping-only scans on all volumes	In the heading area, click **Map**
Enable full scanning on all volumes	In the heading area, click **Map & Classify**
Disable scanning on all volumes	In the heading area, click **Off**

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.

Scanning 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.

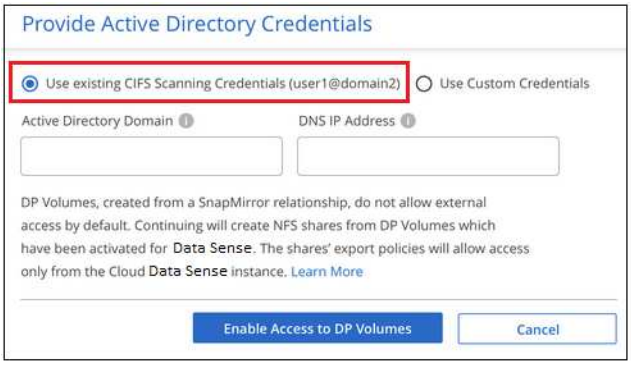
Initially, the volume list identifies these volumes as *Type* **DP** with the *Status* **Not Scanning** and the *Required Action* **Enable Access to DP volumes**.

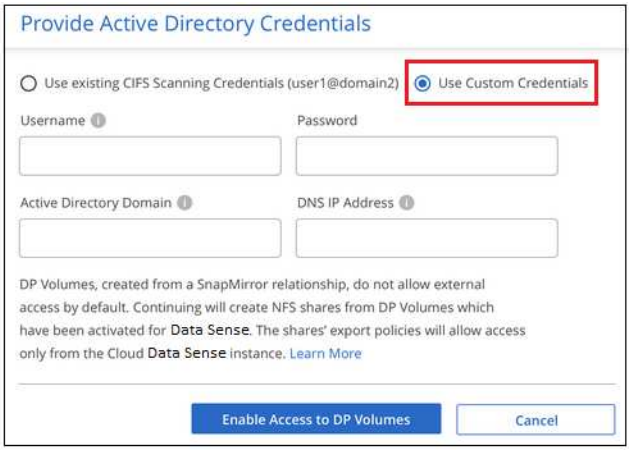


Steps

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.





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.

Getting started with BlueXP classification for Azure NetApp Files

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.

Discovering 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.

Deploying 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.

Enabling 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 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.

Verifying 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.

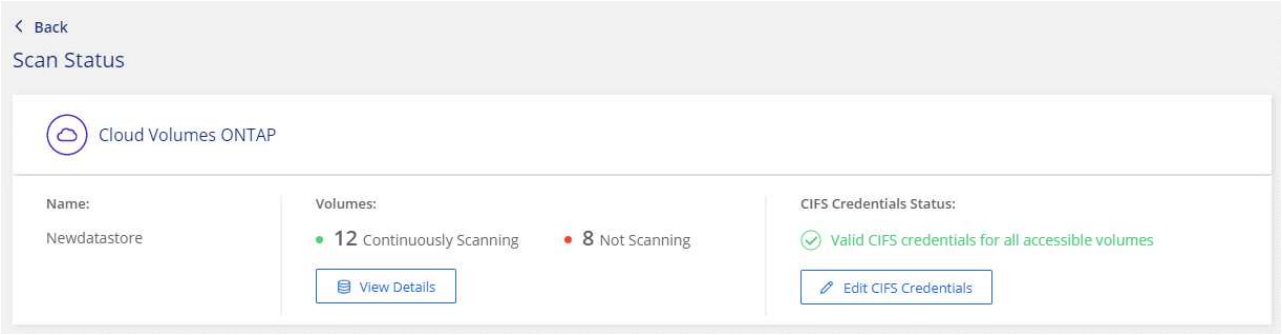
3. Ensure that NFS volume export policies include the IP address of the BlueXP classification instance so it can access the data on each volume.
4. If you use CIFS, provide BlueXP classification with Active Directory credentials so it can scan CIFS volumes.
   a. From the BlueXP left navigation menu, 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.

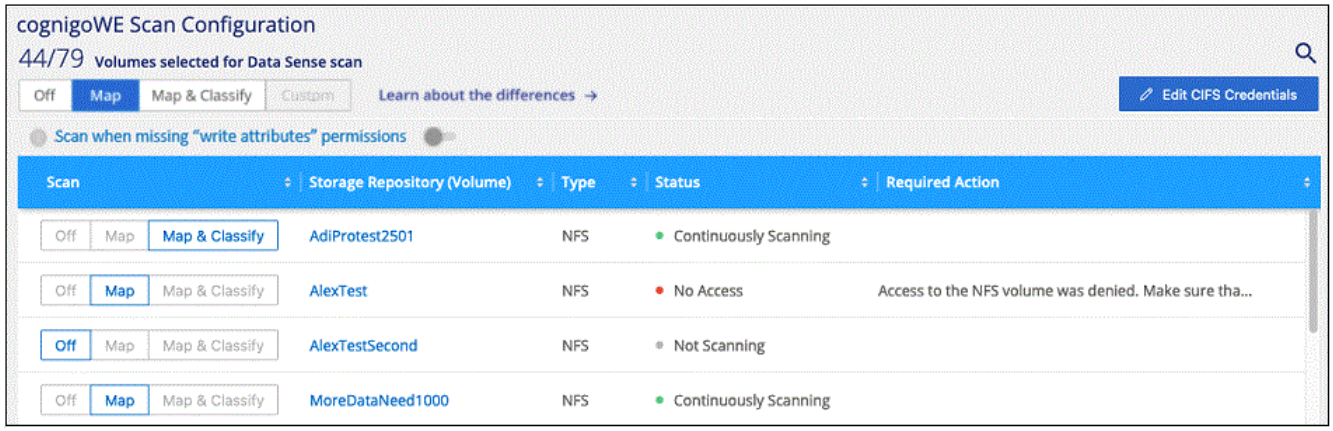
      If you want to make sure your files "last accessed times" are unchanged by BlueXP classification scans, we recommend that the user has Write Attributes permissions in CIFS or write permissions in NFS. If possible, we recommend making the Active Directory configured user part of a parent group in the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

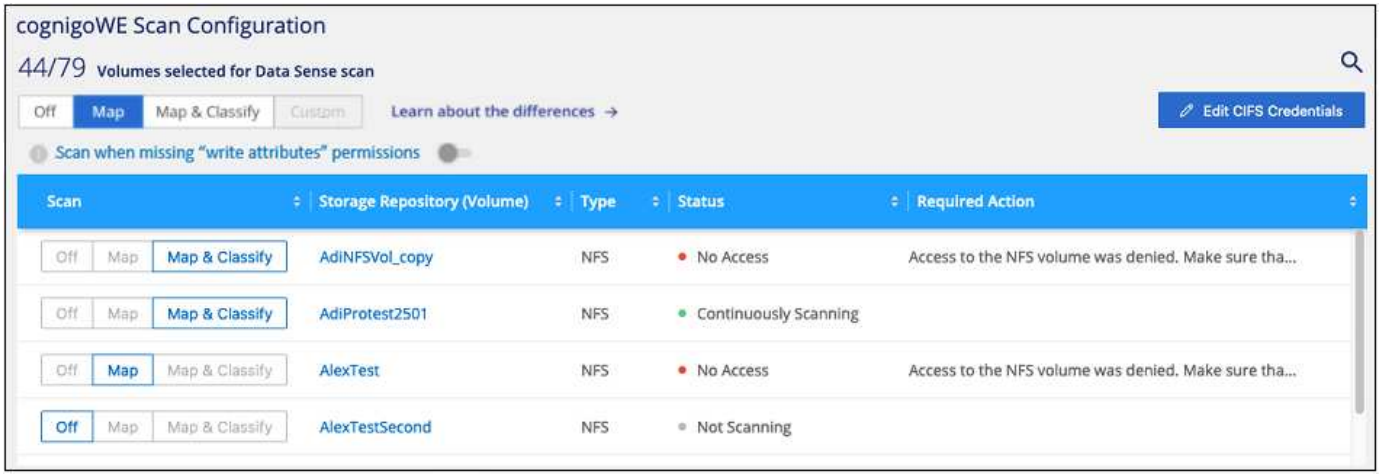
   For example, the following image shows four volumes; one of which BlueXP classification can't scan due to network connectivity issues between the BlueXP classification instance and the volume.



Enabling and disabling 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.



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.

Get started with BlueXP classification for Amazon FSx for ONTAP

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.

Discovering 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.

Deploying 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.

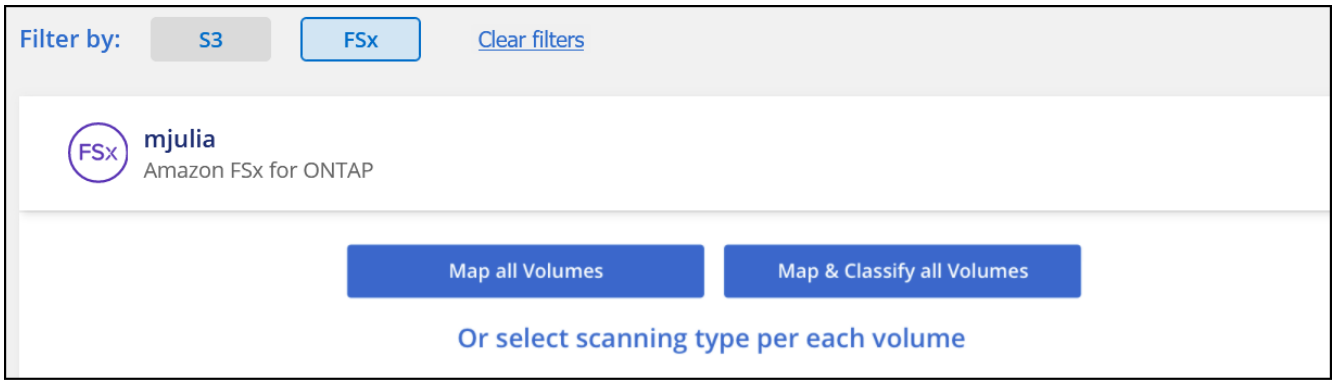
Note: Deploying BlueXP classification in an on-premises location is not currently supported when scanning FSx volumes.

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

Enabling 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.

Verifying 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
Off Map 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.

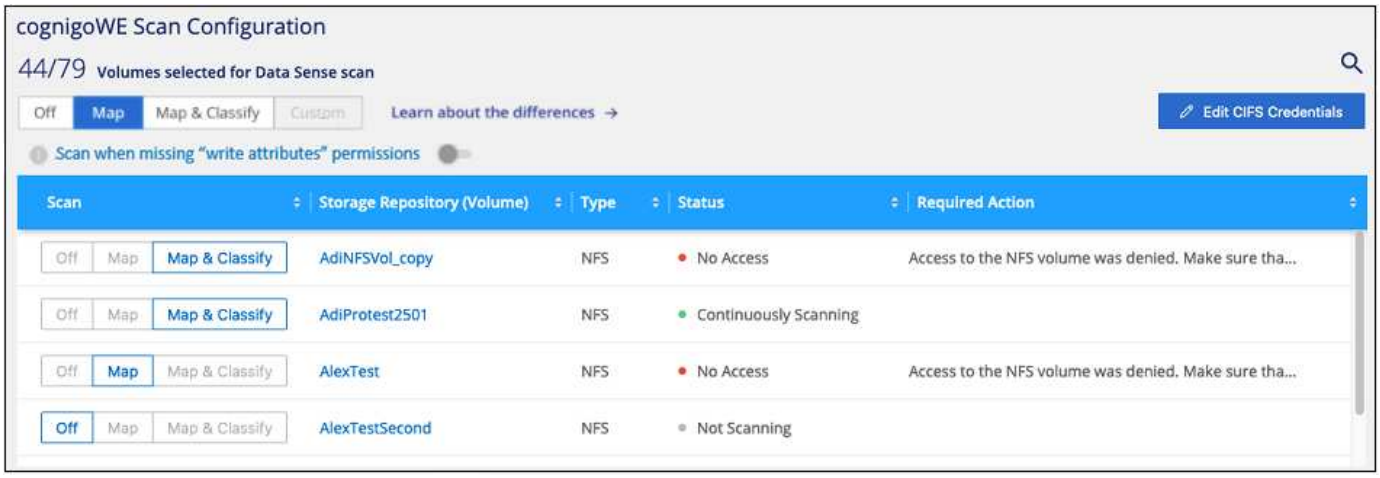
   If you want to make sure your files "last accessed times" are unchanged by BlueXP classification scans, we recommend that the user has Write Attributes permissions in CIFS or write permissions in NFS. If possible, we recommend making the Active Directory configured user part of a parent group in the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

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

Enabling and disabling 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.



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**

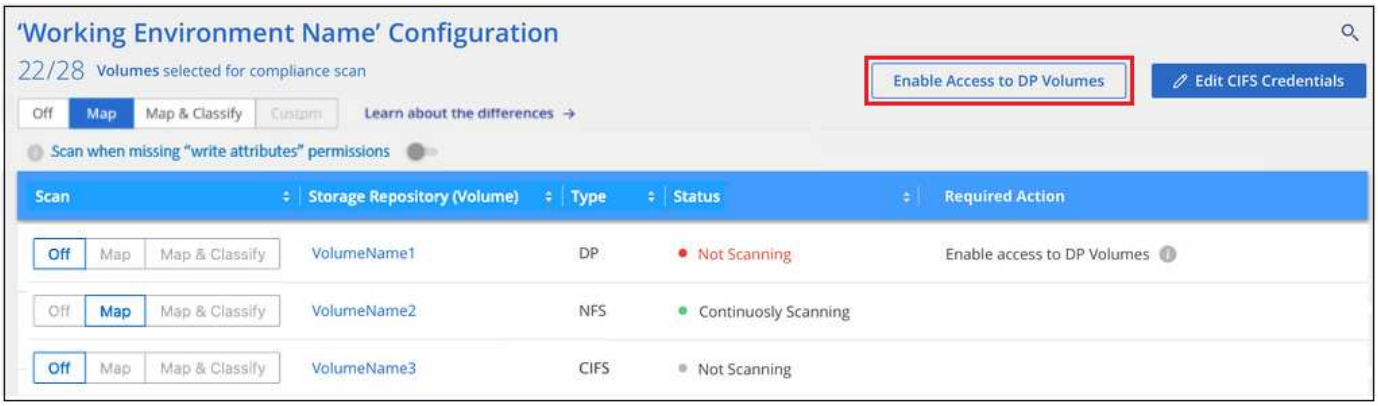
To:	Do this:
Enable full scanning on all volumes	In the heading area, click **Map & Classify**
Disable scanning on all volumes	In the heading area, click **Off**

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.

Scanning 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.

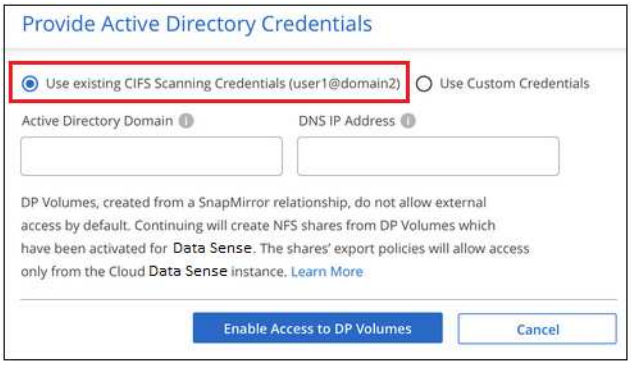
Initially, the volume list identifies these volumes as *Type* **DP** with the *Status* **Not Scanning** and the *Required Action* **Enable Access to DP volumes**.

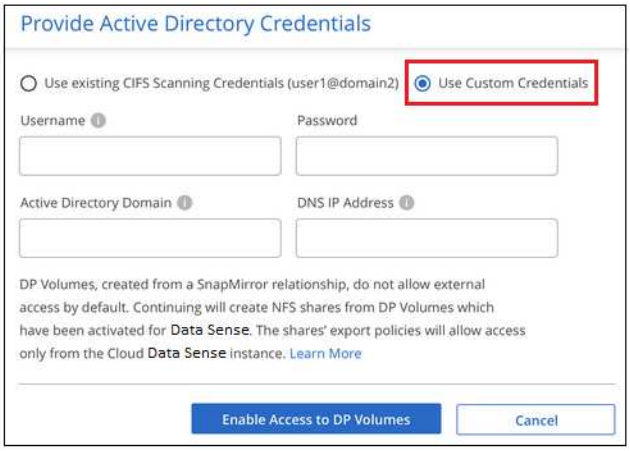


Steps

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.





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

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.

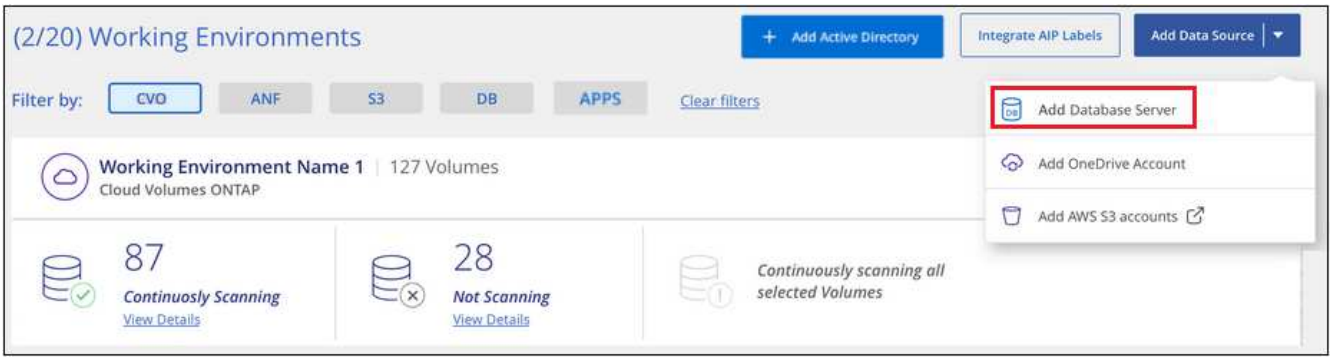
If you are scanning database schemas that have been installed in a dark site that has no internet access, you need to deploy BlueXP classification in the same on-premises location that has no internet access. This also requires that the BlueXP Connector is deployed in that same on-premises location.

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

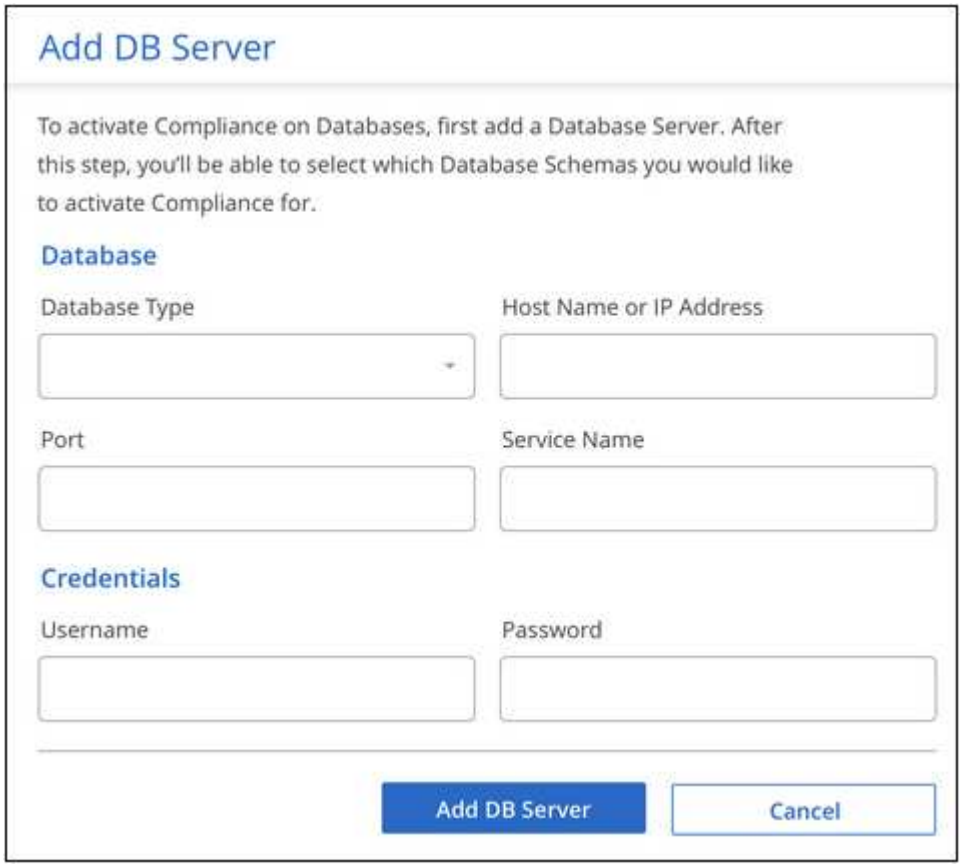
Add the database server

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 database is added to the list of working environments.

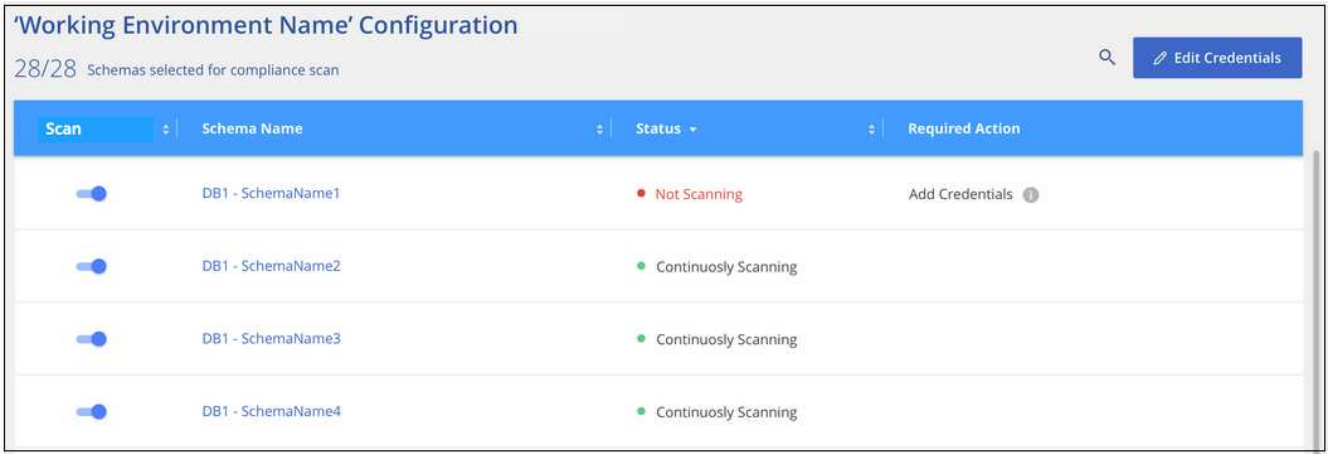
Enable and disable compliance scans on database schemas

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.

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



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.

Scanning file shares

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.

Reviewing 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.

Deploying the BlueXP classification instance

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

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

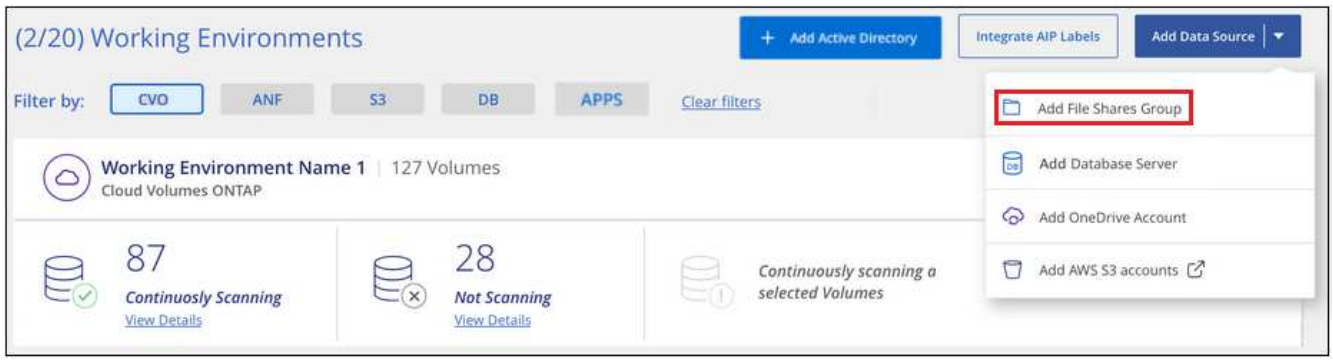
Creating the group for the file shares

You must add a files 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.

Adding 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>`.

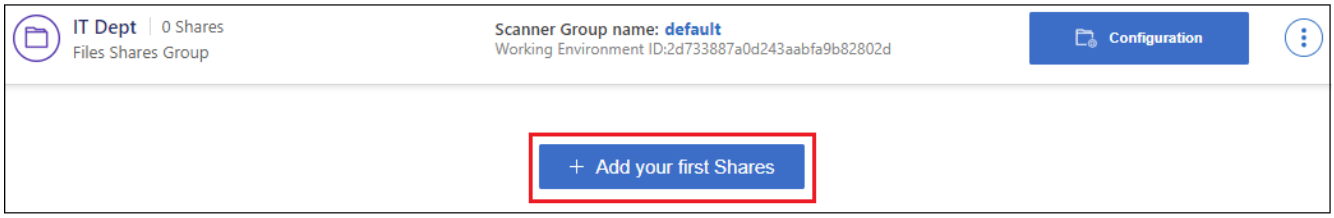
You can add individual file shares, or you can supply a line-separated list of the file shares you want to scan. You can add up to 100 shares at a time.

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.

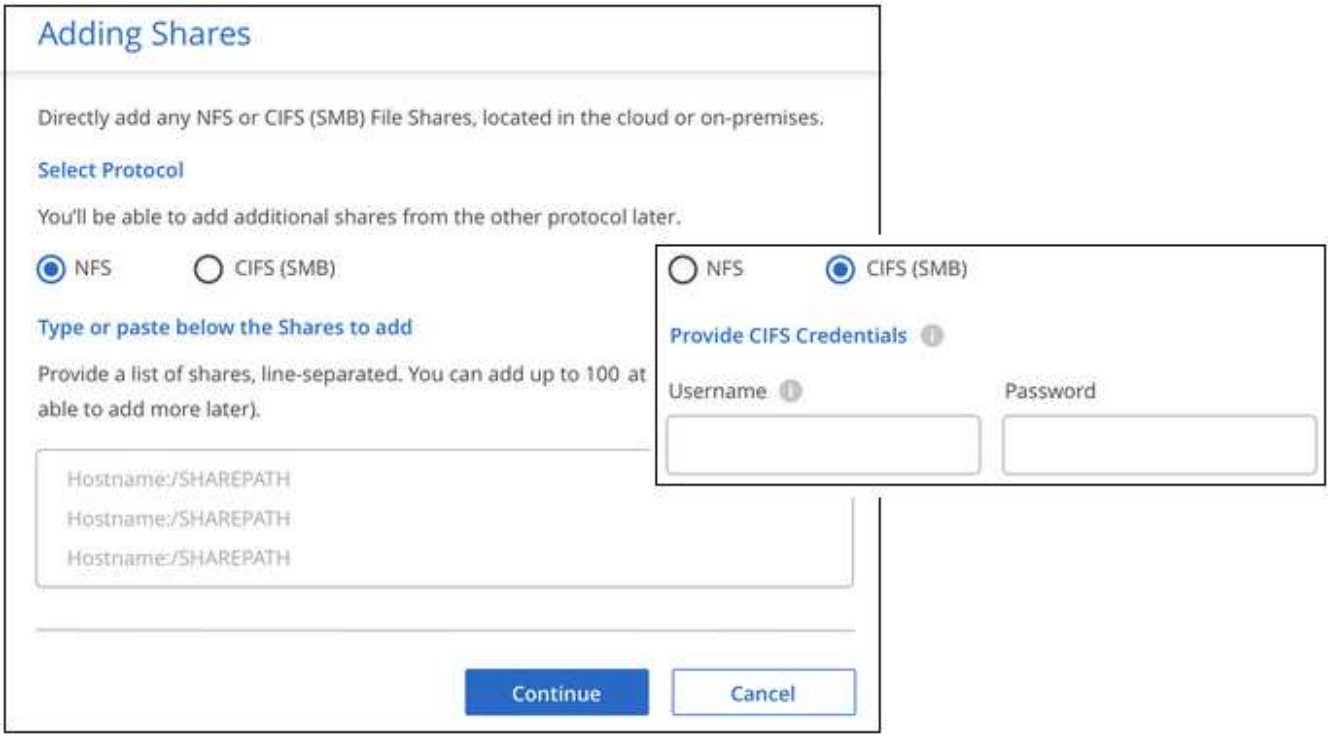
2. If this is the first time adding file shares for this File Shares Group, click **Add your first Shares**.



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

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.



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**

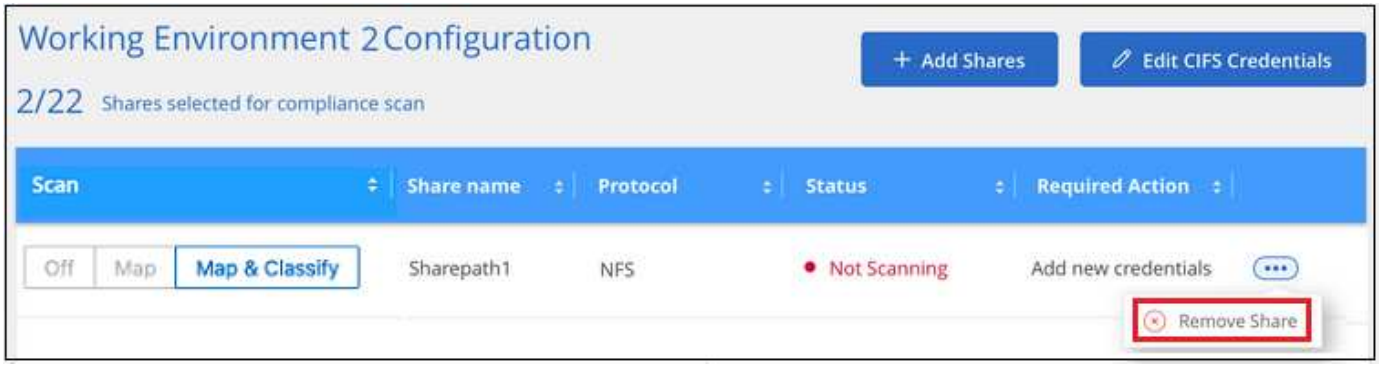
The switch at the top of the page for **Scan when missing "write attributes" permissions** is disabled by default. This means that if BlueXP classification doesn't have write attributes permissions in CIFS, or write permissions in NFS, that the system won't scan the files because BlueXP classification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

Removing 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.



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.

Integration with Active Directory is not supported in the BlueXP classification core version.

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 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 addresses

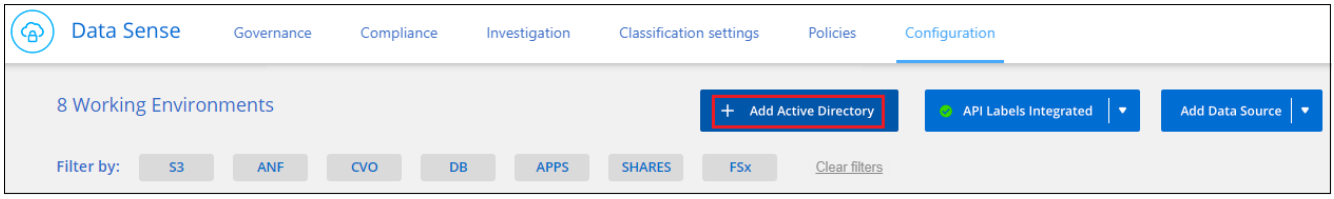
    or

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

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

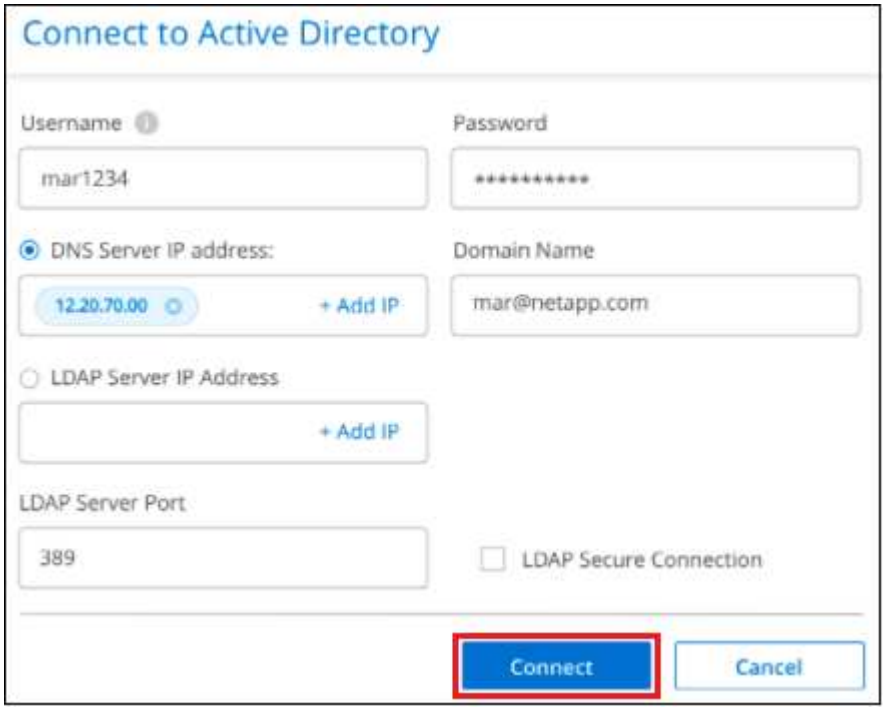
Steps

1. From the BlueXP classification Configuration page, click **Add Active Directory**.

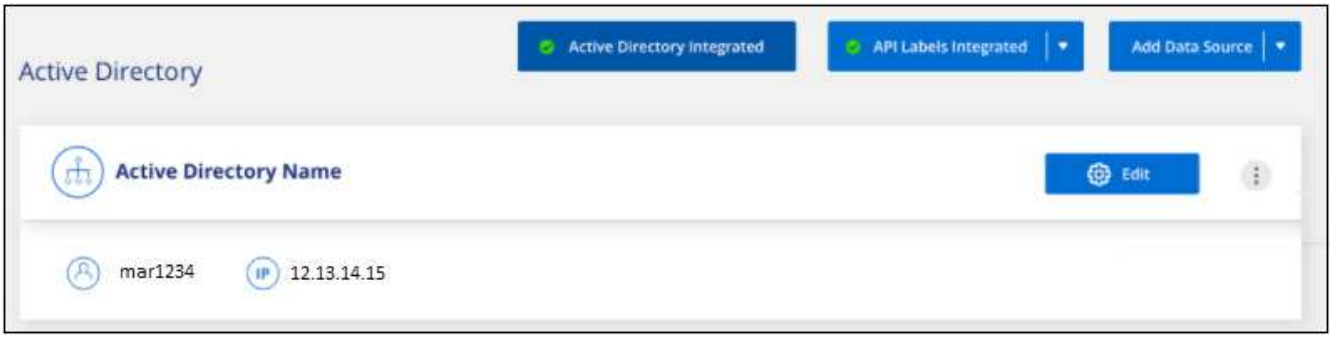


2. In the Connect to Active Directory dialog, enter the Active Directory details and click **Connect**.

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



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.

Note that you can deploy BlueXP classification on a system with fewer CPUs and less RAM, but there are limitations when using these systems. See Using a smaller instance type for details.

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.