Get started

BlueXP classification

NetApp
March 14, 2024

Table of Contents

Get started

Learn about BlueXP classification

BlueXP classification 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.

Learn about the use cases for BlueXP classification.

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.

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.
- Identify whether unique identifiers from your databases are found in files in other repositories - basically making your own list of "personal data" that is identified in BlueXP classification scans.
- Notify certain users through email when files contain certain PII (you define this criteria using Policies) so you can decide on an action plan.

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 notify security staff of new security issues so they can take action immediately.
- Add custom tags to files (for example, "needs to be moved") and assign a BlueXP user so that person can own updates to the files.
- View and modify Azure Information Protection (AIP) labels in your files.

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. You can use this information to decide whether you want to move or delete certain files.
- Delete files that seem insecure or too risky to leave in your storage system, or that you have identified as duplicate. You can use *Policies* to automatically delete files that match certain criteria.
- 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.

Accelerate data migration

BlueXP classification can be used to scan your on-premises data before migrating it to the public or private cloud. You can use BlueXP classification to:

- View the size of data and whether any of the data contains sensitive information prior to moving it.
- Filter the source data (based on over 25 types of criteria) so you can move only the required files to the destination - unnecessary data is not moved.
- Automatically and continually move, copy, or sync only the required data into the cloud repository.

Supported data sources

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

NetApp:

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

Non-NetApp:

- Dell EMC Isilon
- Pure Storage
- Nutanix
- Any other storage vendor

Cloud:

- Amazon S3
- Google Cloud Storage

- OneDrive
- SharePoint Online
- SharePoint On-premises (SharePoint Server)
- Google Drive

Databases:

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

BlueXP classification supports NFS versions 3.x, 4.0, and 4.1, and CIFS versions 1.x, 2.0, 2.1, and 3.0.

Cost

- The cost to use BlueXP classification depends on the amount of data that you're scanning. The first 1 TB of data that BlueXP classification scans in a BlueXP workspace is free for 30 days. This includes all data from all working environments and data sources. A subscription to the AWS, Azure, or GCP Marketplace, or a BYOL license from NetApp, is required to continue scanning data after that point. See pricing for details.

  Learn how to license BlueXP classification.

- 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.

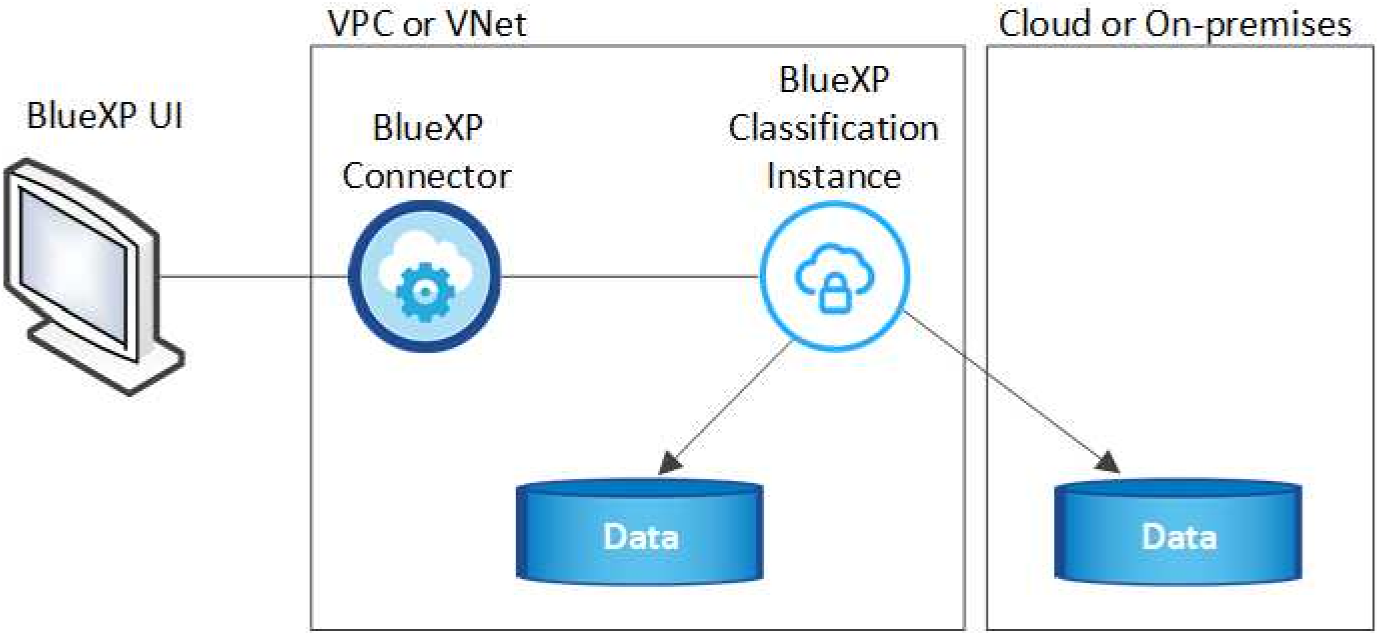
Data transfer costs

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 or S3 Bucket, 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 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.

Using a smaller instance type

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

System size	Specs	Limitations
Extra Large	32 CPUs, 128 GB RAM, 1 TiB SSD	Can scan up to 500 million files.

System size	Specs	Limitations
Large (default)	16 CPUs, 64 GB RAM, 500 GiB SSD	Can scan up to 250 million files.
Medium	8 CPUs, 32 GB RAM, 200 GiB SSD	Slower scanning, and can only scan up to 1 million files.
Small	8 CPUs, 16 GB RAM, 100 GiB SSD	Same limitations as "Medium", plus the ability to identify data subject names inside files is disabled.

When deploying BlueXP classification in the cloud on AWS you can choose a large/medium/small instance. When deploying BlueXP classification in Azure or GCP, email ng-contact-data-sense@netapp.com for assistance if you want to use one of these alternate systems. We'll need to work with you to deploy these other cloud configurations.

When deploying BlueXP classification on-premises, just use a Linux host with the alternate specifications. You do not need to contact NetApp for assistance.

How BlueXP classification works

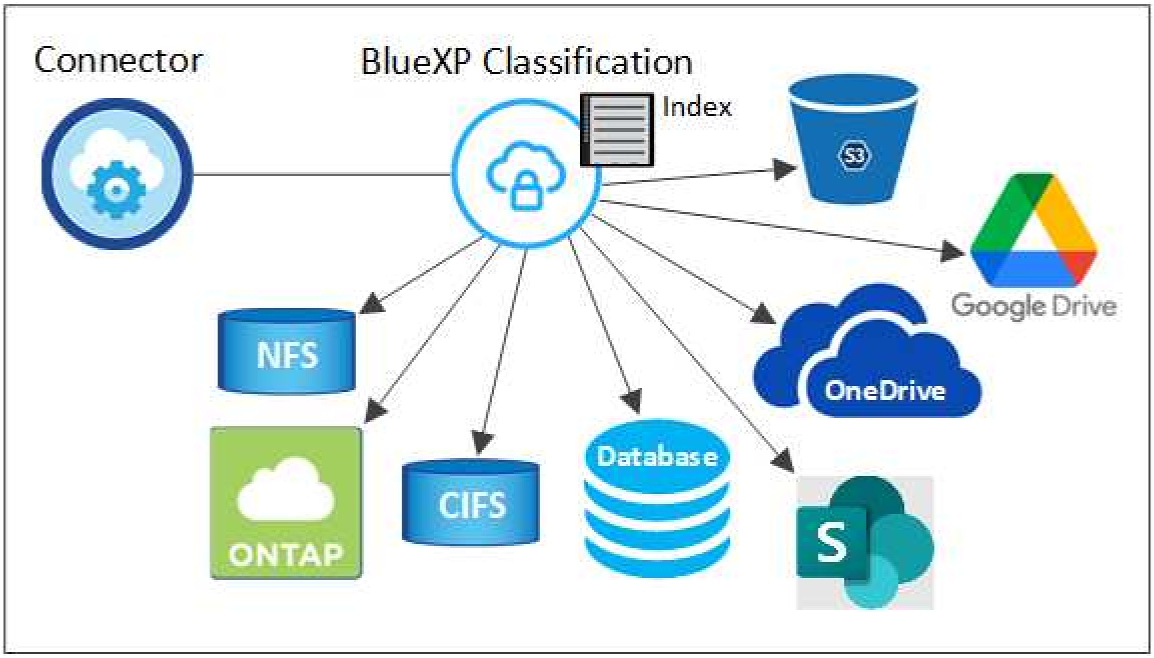
At a high-level, BlueXP classification works like this:

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, buckets, database schemas, or OneDrive or SharePoint 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, at the bucket level, at the database schema level, at the OneDrive user level, and at the SharePoint site 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
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
Categorize data using AIP labels and Status tags	Yes	No

Feature	Classification	Mapping
Copy, delete, and move source files	Yes	No
Ability to run other reports	Yes	No

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

Learn more about deploying multiple scanner nodes to scan your data.

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 such as email addresses, identification numbers, or credit card numbers. Learn more about personal data.

Sensitive personal data

Special types of sensitive information, 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 natural persons' 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. These users can’t copy, move, or delete files either.

Learn more about BlueXP roles and how to add users with specific roles.

Deploy 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.

4

Subscribe to the BlueXP classification service

The first 1 TB of data that BlueXP classification scans in BlueXP is free for 30 days. A BlueXP subscription through your cloud provider Marketplace, or a BYOL license from NetApp, is required to continue scanning data after that point.

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, Amazon FSx for ONTAP, or in AWS S3 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, non-NetApp file shares, generic S3 Object storage, databases, OneDrive folders, SharePoint accounts, and Google Drive accounts 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:

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

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

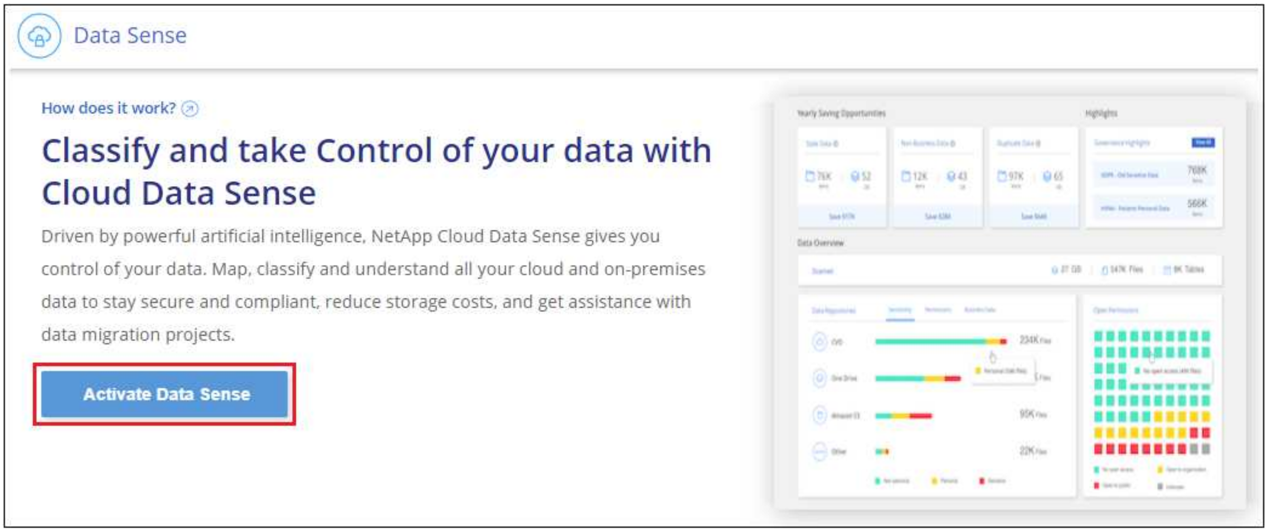
Follow these steps to deploy an instance of BlueXP classification in the cloud. The Connector will deploy the instance in the cloud, and then install BlueXP classification software on that instance.

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**.



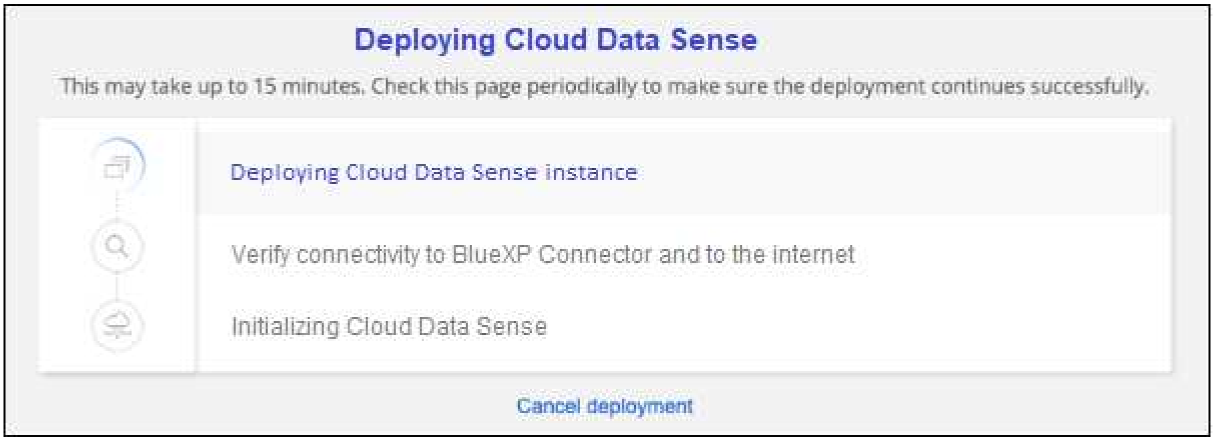
Data Sense

How does it work?

Classify and take Control of your data with Cloud Data Sense

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

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.



Deploying Cloud Data Sense

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

Deploying Cloud Data Sense instance

Verify connectivity to BlueXP Connector and to the internet

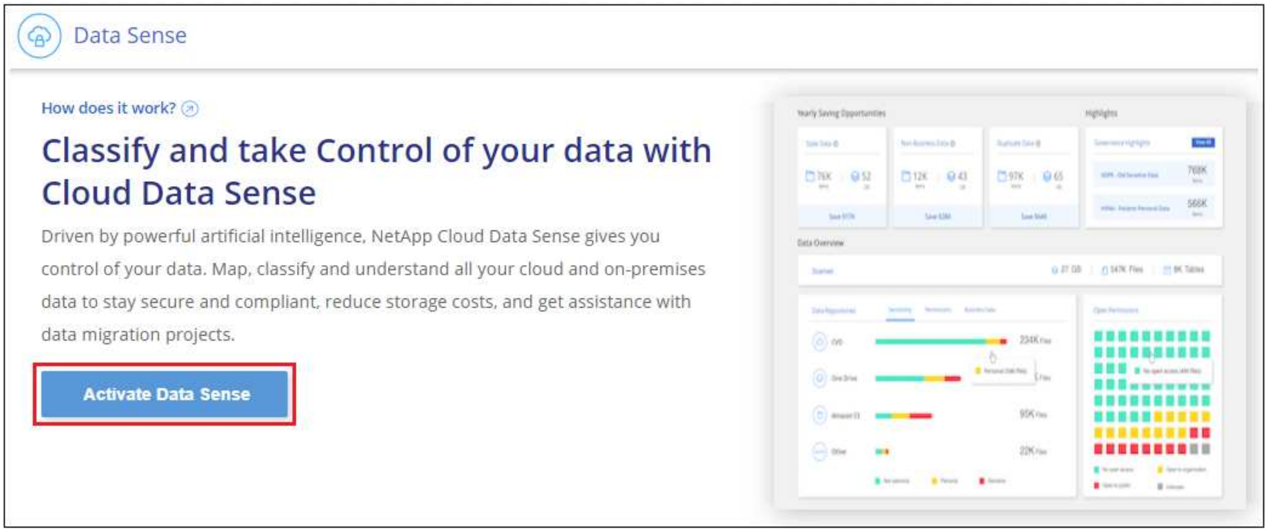
Initializing Cloud Data Sense

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**.



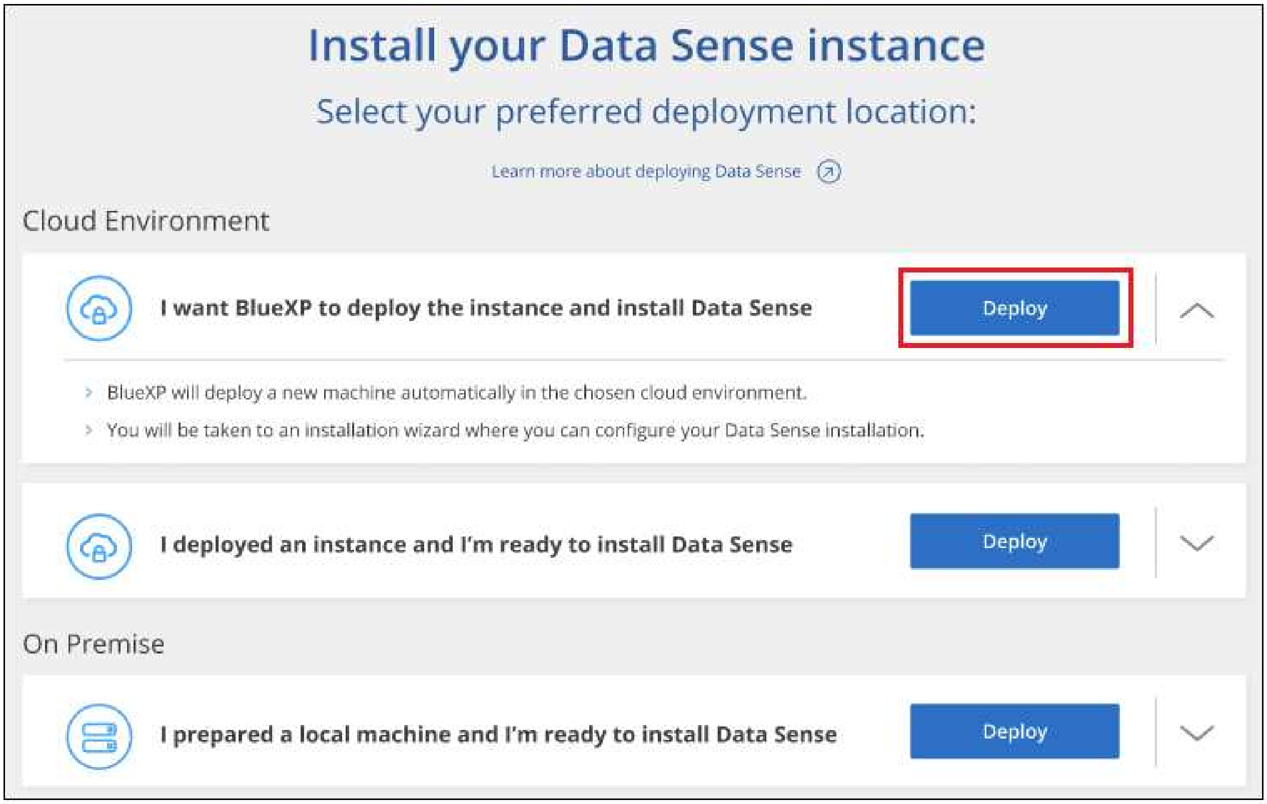
Data Sense

How does it work?

Classify and take Control of your data with Cloud Data Sense

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

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



Install your Data Sense instance

Select your preferred deployment location:

Learn more about deploying Data Sense

Cloud Environment

I want BlueXP to deploy the instance and install Data Sense

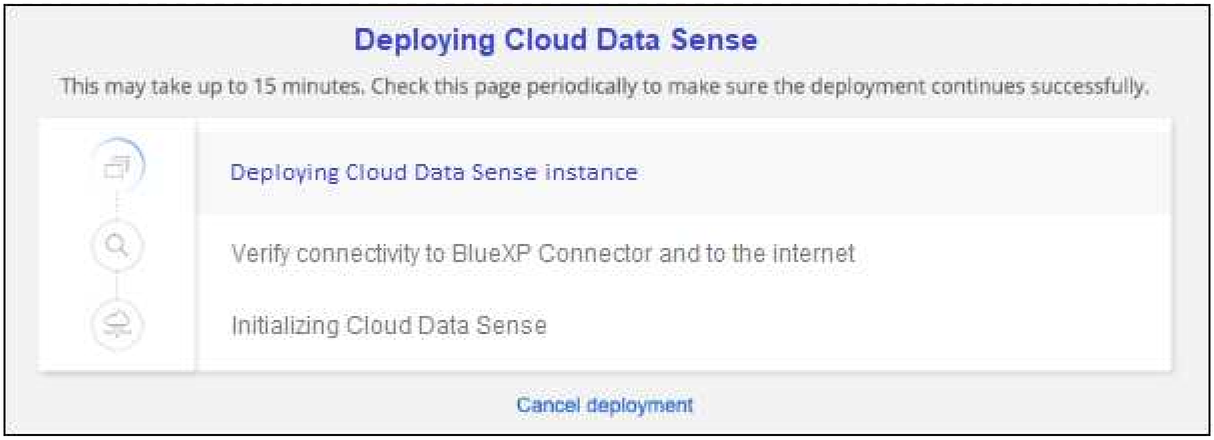
- BlueXP will deploy a new machine automatically in the chosen cloud environment.
- You will be taken to an installation wizard where you can configure your Data Sense installation.

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

On Premise

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

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



Deploying Cloud Data Sense

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

Deploying Cloud Data Sense instance

Verify connectivity to BlueXP Connector and to the internet

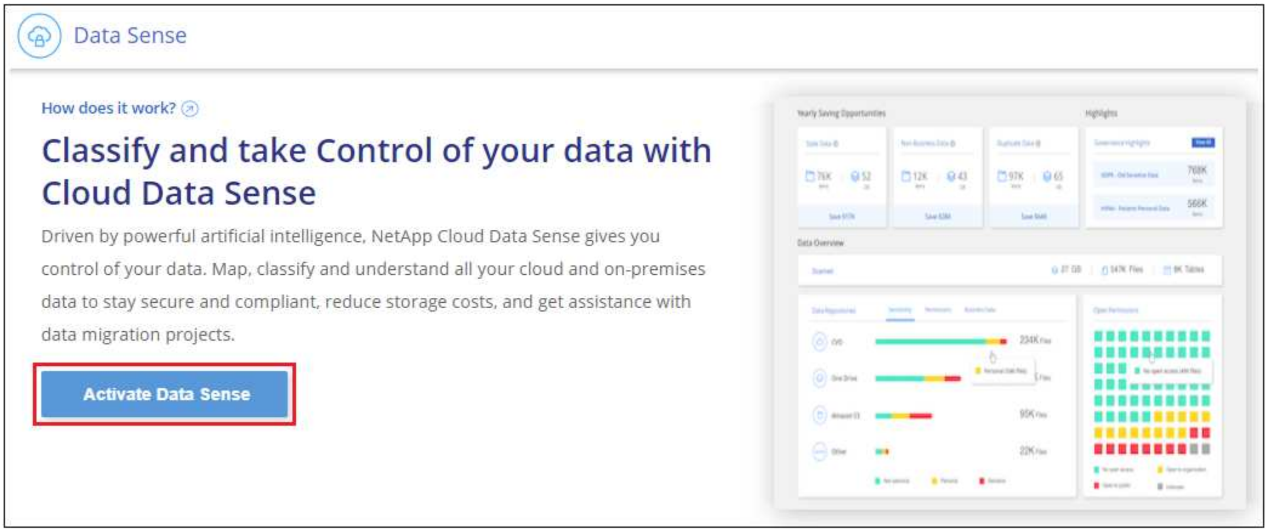
Initializing Cloud Data Sense

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**.



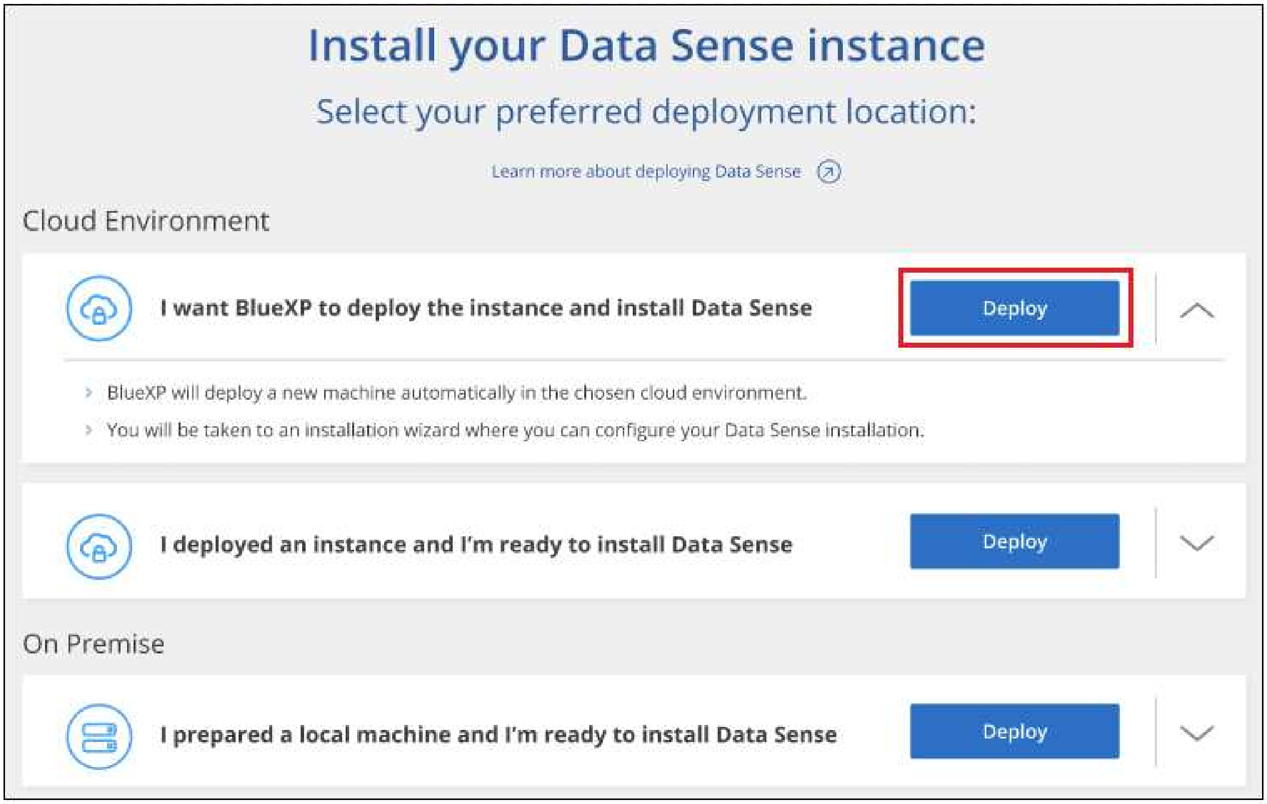
Data Sense

How does it work?

Classify and take Control of your data with Cloud Data Sense

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

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



Install your Data Sense instance

Select your preferred deployment location:

Learn more about deploying Data Sense

Cloud Environment

I want BlueXP to deploy the instance and install Data Sense

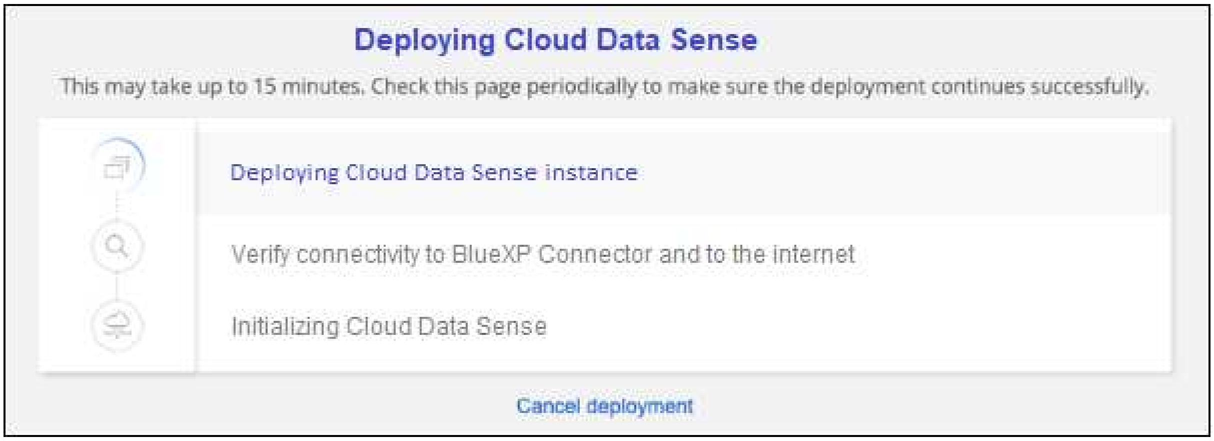
- BlueXP will deploy a new machine automatically in the chosen cloud environment.
- You will be taken to an installation wizard where you can configure your Data Sense installation.

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

On Premise

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

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



Deploying Cloud Data Sense

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

Deploying Cloud Data Sense instance

Verify connectivity to BlueXP Connector and to the internet

Initializing Cloud Data Sense

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.

You can also set up licensing for BlueXP classification at this time. You will not be charged until your 30-day free trial ends.

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, 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*.

Note that you can also install BlueXP classification in an on-premises site that doesn't have internet access for completely secure sites.

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.

4

Subscribe to the BlueXP classification service

The first 1 TB of data that BlueXP classification scans in BlueXP is free for 30 days. A subscription to your cloud provider Marketplace, or a BYOL license from NetApp, is required to continue scanning data after that point.

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, Amazon FSx for ONTAP, or in AWS S3 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, non-NetApp file shares, generic S3 Object storage, databases, OneDrive folders, SharePoint accounts, and Google Drive 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.

As you can see, there may be some situations where you need to use multiple Connectors.

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 three system sizes depending on the size of the data set 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 - 5 GiB on /tmp
Medium	8 CPUs	32 GB RAM	200 GiB SSD on /, or - 50 GiB available on /opt - 145 GiB available on /var/lib/docker - 5 GiB on /tmp
Small	8 CPUs	16 GB RAM	100 GiB SSD on /, or - 50 GiB available on /opt - 45 GiB available on /var/lib/docker - 5 GiB on /tmp

Note that there are limitations when using the smaller systems. See Using a smaller instance type for details.

- When deploying a compute instance in the cloud for your BlueXP classification installation, we recommend a system that meets the "Large" system requirements above:
    - **AWS 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 1.26 or greater:
        - Red Hat Enterprise Linux version 9.0, 9.1, and 9.2

          Note that the following features are not currently supported when using 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, update your system packages (`sudo yum update -y`), and then install Podman (`sudo yum install podman -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.

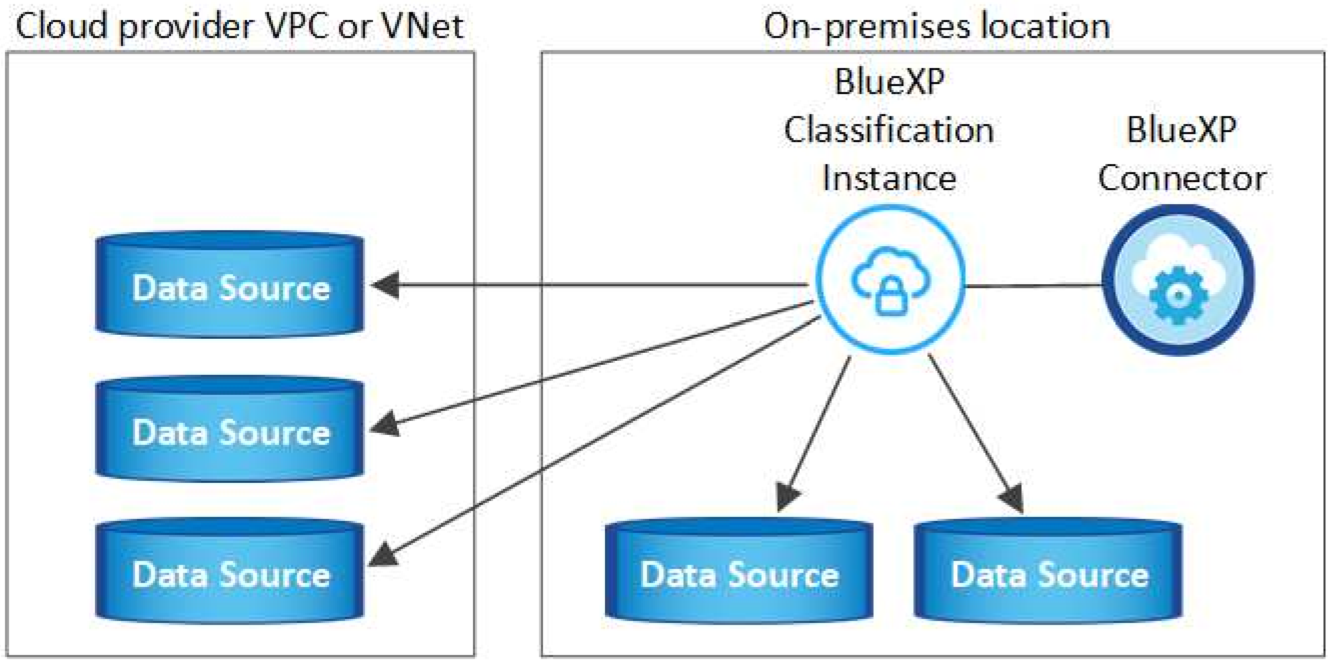
Connection Type	Ports	Description
Connector <> BlueXP classification	8080 (TCP), 443 (TCP), and 80	The firewall or routing rules for the Connector must allow inbound and outbound traffic over port 443 to and from the BlueXP classification instance. Make sure port 8080 is open so you can see the installation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

Connection Type	Ports	Description
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)

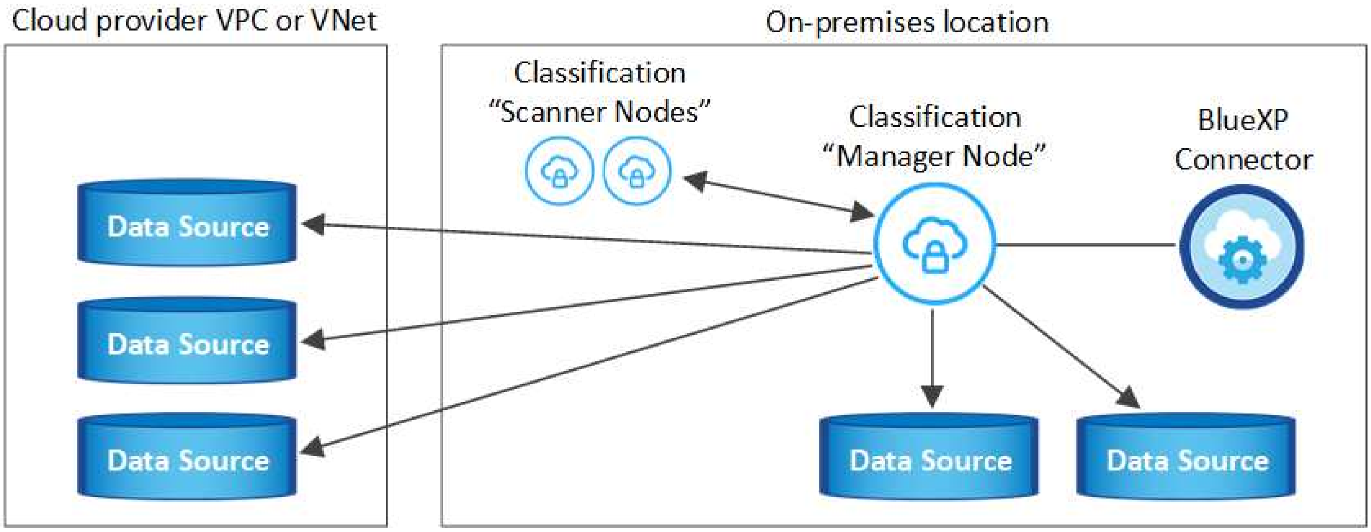
If you are using multiple BlueXP classification hosts to provide additional processing power to scan your data sources, you'll need to enable additional ports/protocols. See the additional port requirements.

Install BlueXP classification on the 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. See those steps here.



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.

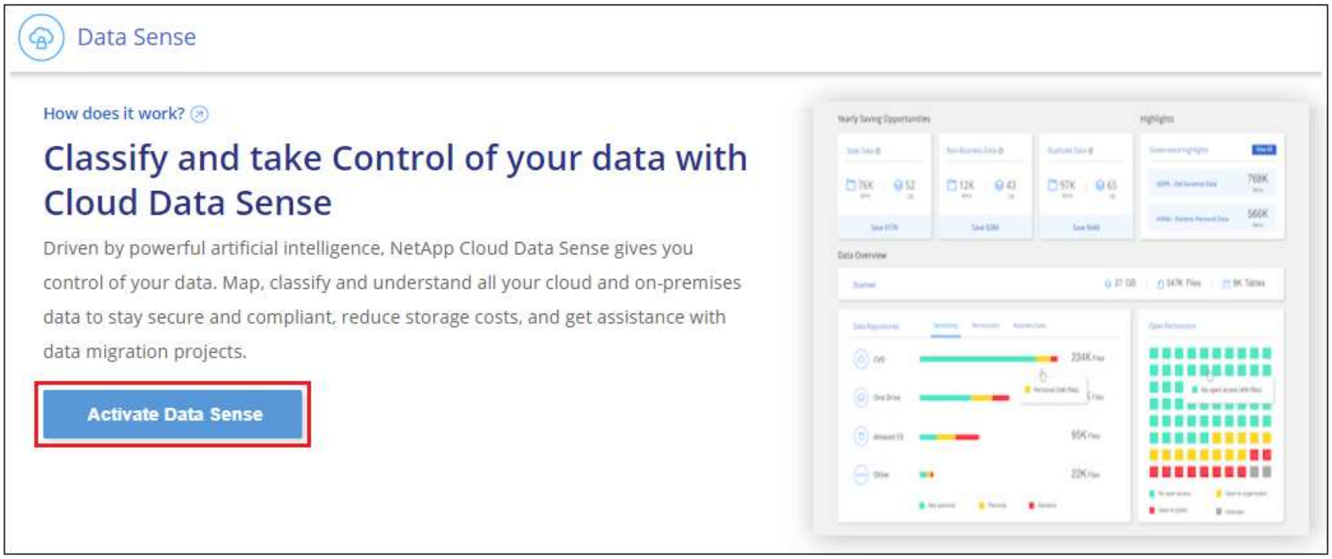
- Verify that your offline environment meets the required permissions and connectivity.

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**.



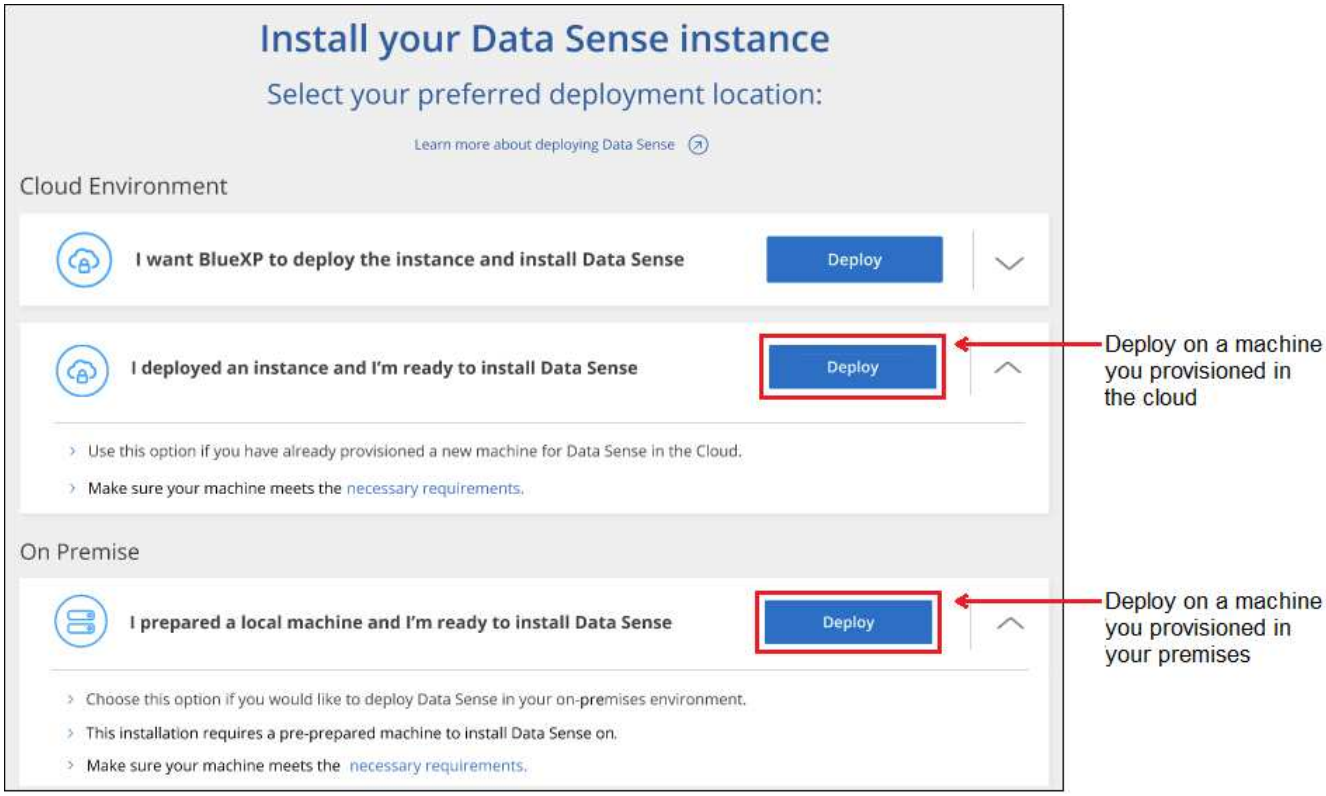
Data Sense

How does it work?

Classify and take Control of your data with Cloud Data Sense

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

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.



Install your Data Sense instance

Select your preferred deployment location:

Learn more about deploying Data Sense

Cloud Environment

I want BlueXP to deploy the instance and install Data Sense

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

- Use this option if you have already provisioned a new machine for Data Sense in the Cloud.
- Make sure your machine meets the necessary requirements.

On Premise

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

- Choose this option if you would like to deploy Data Sense in your on-premises environment.
- This installation requires a pre-prepared machine to install Data Sense on.
- Make sure your machine meets the necessary requirements.

Deploy
Deploy
Deploy
Deploy on a machine you provisioned in the cloud
Deploy on a machine you provisioned in your premises

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.

You can also set up licensing for BlueXP classification at this time. You will not be charged until your 30-day free trial ends.

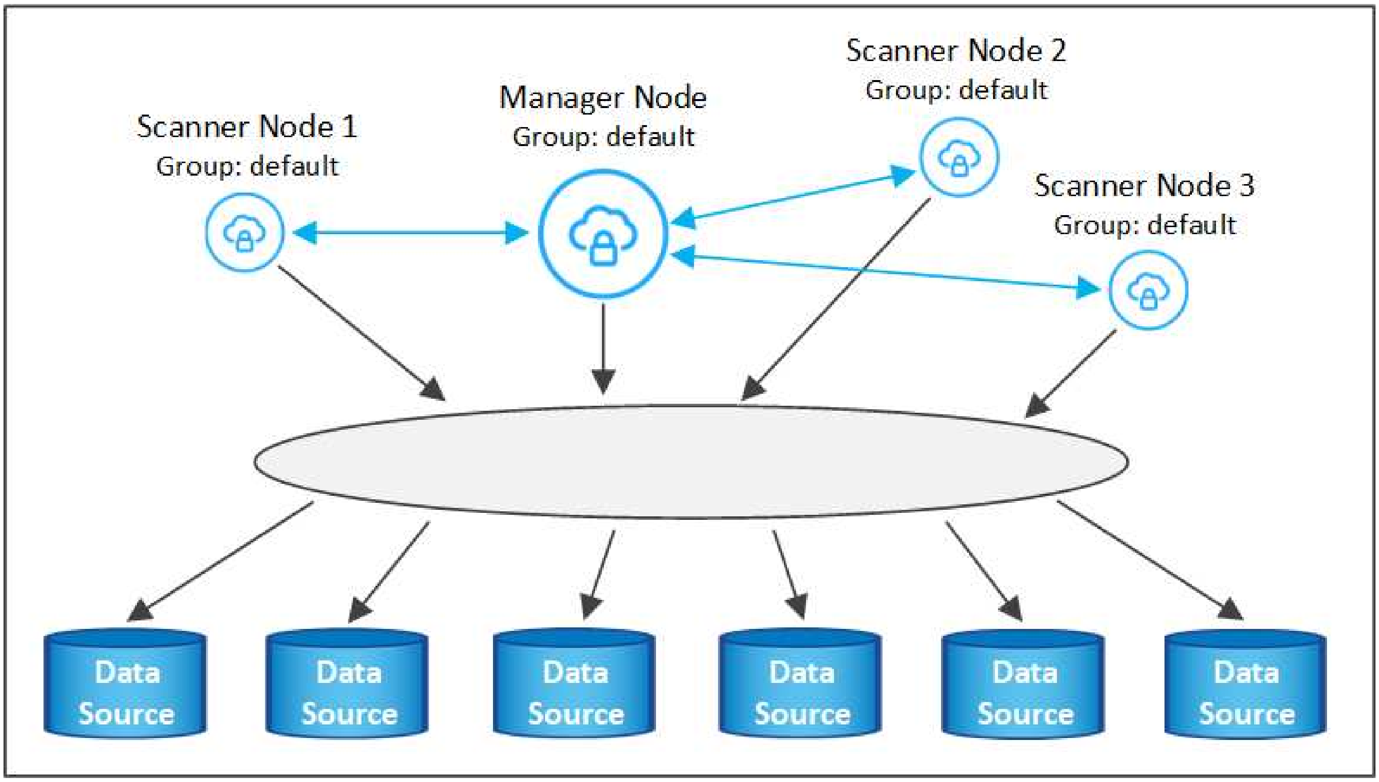
Add scanner nodes to an existing deployment

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

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

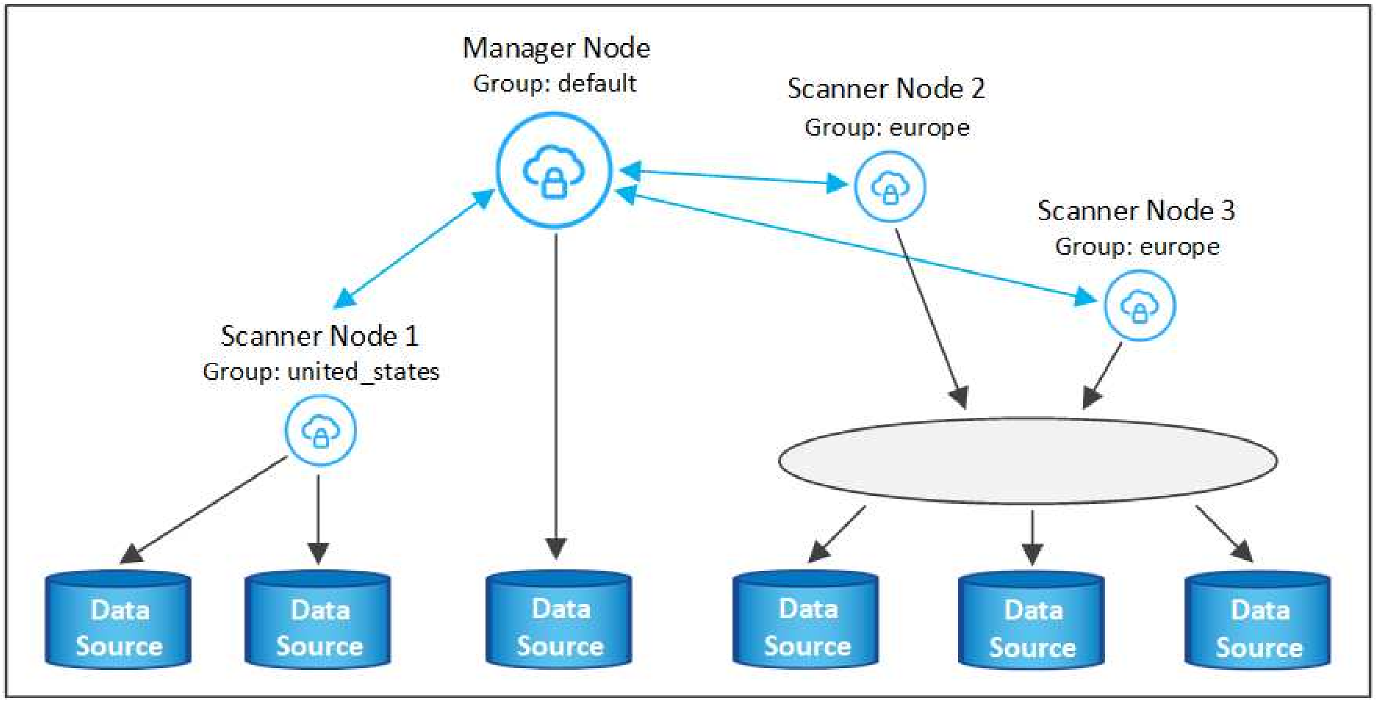
- add a node to assist with scanning all data sources
- add a node to assist with scanning a specific data source, or a specific group of data sources (typically based on location)

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



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

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



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

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

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

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

What you'll need

- Verify that all your Linux systems for Scanner nodes meet the host requirements.
- Verify that the systems have the two prerequisite software packages installed (Docker Engine or Podman, and Python 3).
- Make sure you have root privileges on the Linux systems.
- Verify that your environment meets the required permissions and connectivity.

- You must have the IP addresses of the Scanner node hosts that you are adding.
- You must have the IP address of the BlueXP classification Manager node host system
- You must have the IP address or host name of the Connector system, your NetApp Account ID, Connector Client ID, and user access token. If you're planning to use scanner groups, you'll need to know the Working Environment ID for each data source in your account. See ***Prerequisite steps*** below to get this information.
- The following ports and protocols must be enabled on all hosts:

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

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

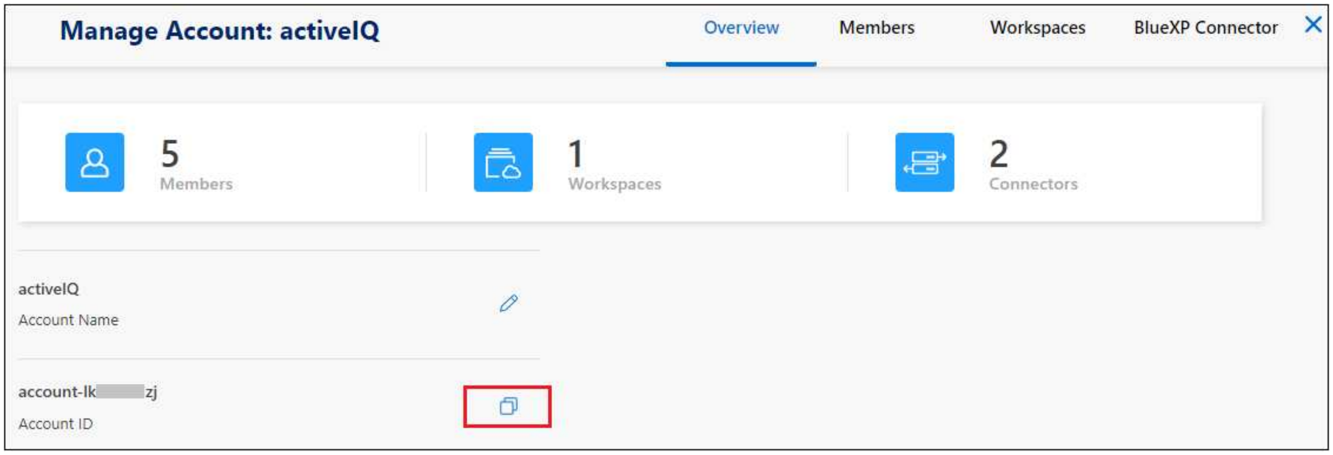
```
firewall-cmd --permanent --add-service=http
firewall-cmd --permanent --add-service=https
firewall-cmd --permanent --add-port=80/tcp
firewall-cmd --permanent --add-port=8080/tcp
firewall-cmd --permanent --add-port=443/tcp
firewall-cmd --permanent --add-port=2377/tcp
firewall-cmd --permanent --add-port=7946/udp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

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

Prerequisite steps

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

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



5 Members

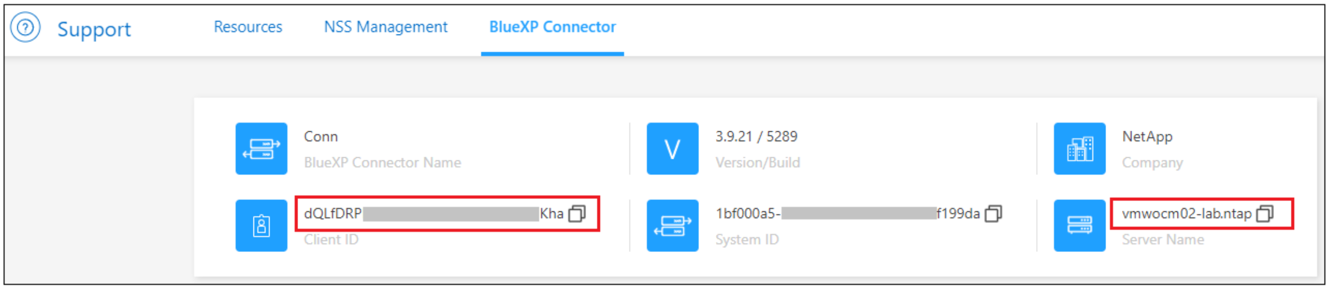
1 Workspaces

2 Connectors

activeIQ
Account Name

account-lk[redacted]zj
Account ID

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



Conn
BlueXP Connector Name

3.9.21 / 5289
Version/Build

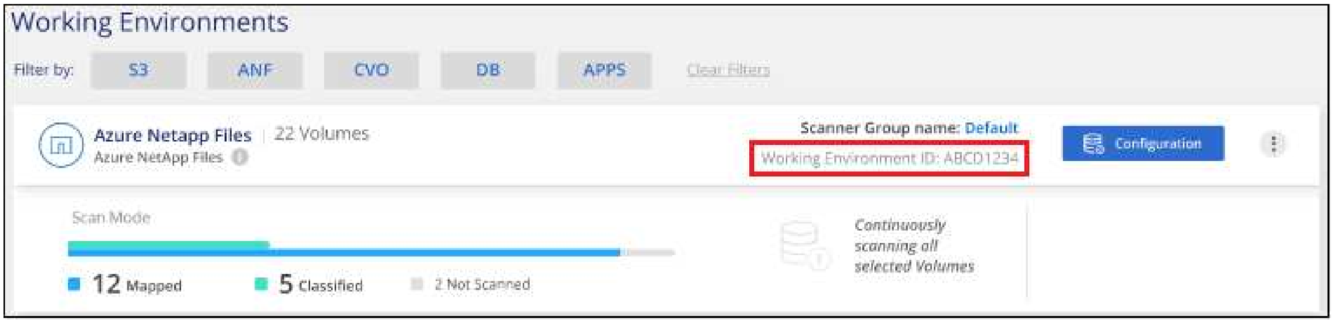
NetApp
Company

dQLfDRP[redacted]Kha
Client ID

1bf000a5-[redacted]f199da
System ID

vmwocm02-lab.ntap
Server Name

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



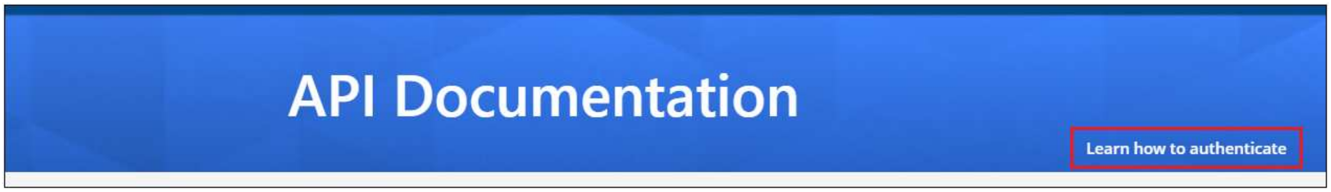
Azure Netapp Files | 22 Volumes
Azure NetApp Files

Scanner Group name: Default
Working Environment ID: ABCD1234

12 Mapped
5 Classified
2 Not Scanned

Continuously scanning all selected Volumes

6. Go to the API Documentation Developer Hub and click **Learn how to authenticate**.



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

Steps

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

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

   Variable values:

   - *account_id* = NetApp Account ID
   - *client_id* = Connector Client ID (add the suffix "clients" to the client ID that you copied in the Prerequisite steps)
   - *cm_host* = IP address or host name of the Connector system
   - *ds_manager_ip* = Private IP address of the BlueXP classification Manager node system
   - *node_private_ip* = IP addresses of the BlueXP classification Scanner node systems (multiple scanner node IPs are separated by a comma)
   - *user_token* = JWT user access token

2. Before the add_scanner_node script completes, a dialog displays the installation command needed for the scanner nodes. Copy the command (for example: `sudo ./node_install.sh -m 10.11.12.13 -t ABCDEF1s35212 -u red95467j`) and save it in a text file.

3. On **each** scanner node host:

   a. Copy the Data Sense installer file (**DATASENSE-INSTALLER-<version>.tar.gz**) to the host machine (using `scp` or some other method).

   b. Unzip the installer file.

   c. Paste and execute the command that you copied in step 2.

   d. If you want to add a scanner node into a "scanner group", add the parameter **-r <scanner_group_name>** to the command. Otherwise, the scanner node is added to the "default" group.

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

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

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

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

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

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

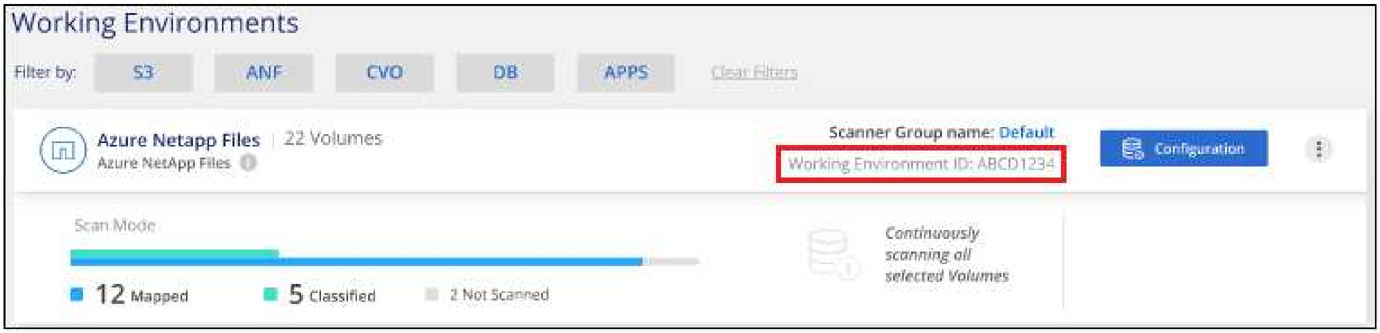
Result

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

What’s Next

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

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



Working Environments

Azure Netapp Files | 22 Volumes

Azure NetApp Files

Scanner Group name: Default

Working Environment ID: ABCD1234

Scan Mode

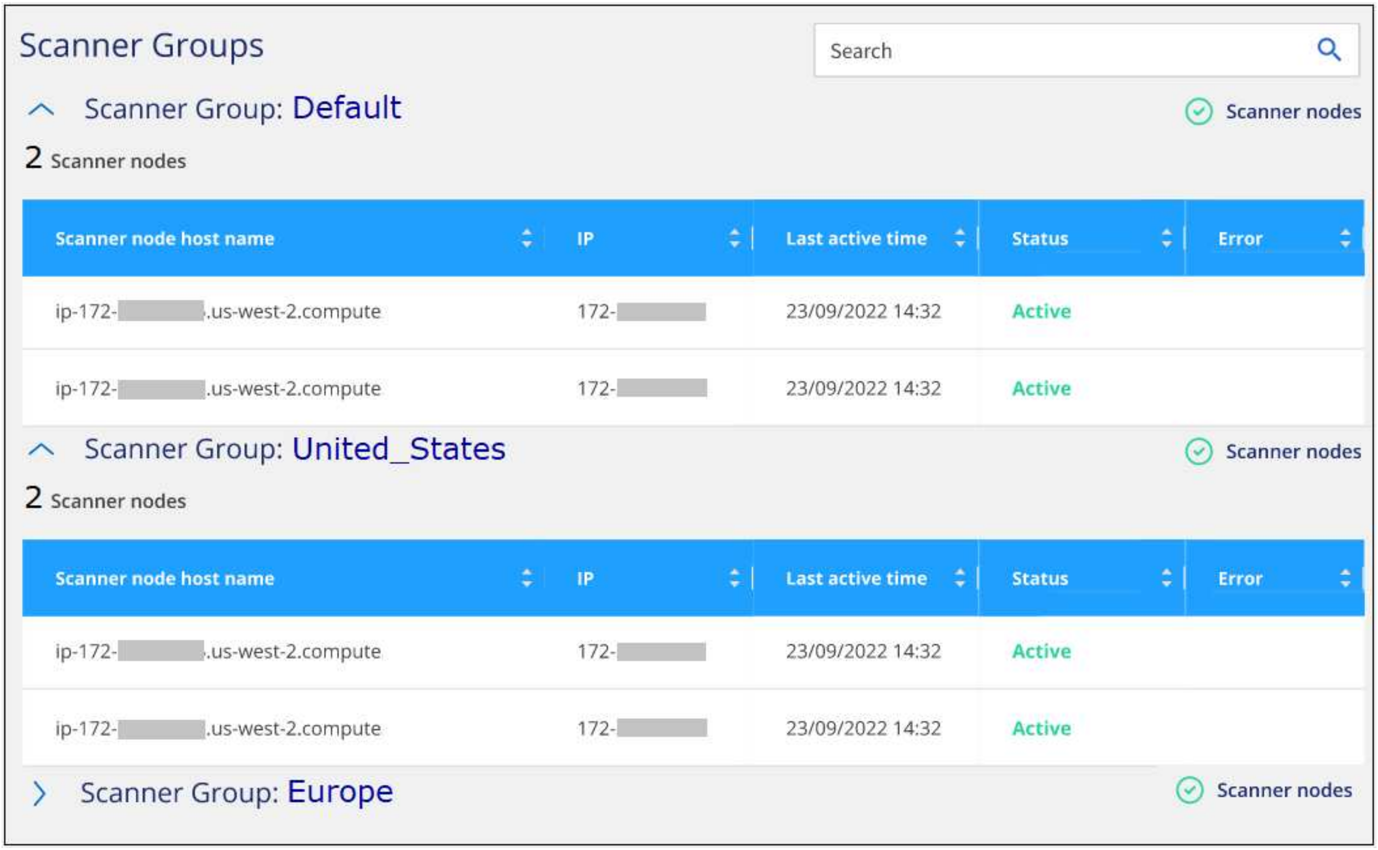
12 Mapped

5 Classified

2 Not Scanned

Continuously scanning all selected Volumes

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


Scanner Groups

Search

Scanner Group: Default

Scanner nodes

2 Scanner nodes

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

Scanner Group: United_States

Scanner nodes

2 Scanner nodes

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

Scanner Group: Europe

Scanner nodes

You can set up licensing for BlueXP classification at this time. You will not be charged until your 30-day free trial ends.

Multi-host installation for large configurations

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

Follow these steps when installing BlueXP classification software on multiple on-premises hosts at the same time. Note that you can't use "scanner groups" when deploying multiple hosts in this fashion.

What you'll need

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

Port	Protocols	Description
2377	TCP	Cluster management communications
7946	TCP, UDP	Inter-node communication

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

Steps

1. Follow steps 1 through 7 from the Single-host installation on the manager node.
2. As shown in step 8, when prompted by the installer, you can enter the required values in a series of prompts, or you can provide the required parameters as command line arguments to the installer.

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

   For example, this command adds 3 scanner nodes:
   ```
   sudo ./install.sh -a <account_id> -c <client_id> -t <user_token> --host
   <ds_host> --manager-host <cm_host> -n <node_ip1>,<node_ip2>,<node_ip3> --proxy
   -host <proxy_host> --proxy-port <proxy_port> --proxy-scheme <proxy_scheme>
   --proxy-user <proxy_user> --proxy-password <proxy_password>
   ```
3. Before the manager node installation completes, a dialog displays the installation command needed for the scanner nodes. Copy the command (for example, `sudo ./node_install.sh -m 10.11.12.13 -t ABCDEF-1-3u69m1-1s35212`) and save it in a text file.
4. On **each** scanner node host:
   a. Copy the Data Sense installer file (**DATASENSE-INSTALLER-<version>.tar.gz**) to the host machine (using `scp` or some other method).
   b. Unzip the installer file.
   c. Paste and execute the command that you copied in step 3.

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

Result

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

What's Next

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

You can also set up licensing for BlueXP classification at this time. You will not be charged until your 30-day free trial ends.

Install BlueXP classification on a 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.

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
- SharePoint On-Premises accounts (SharePoint Server)
- Non-NetApp NFS or CIFS file shares
- Object Storage that uses the Simple Storage Service (S3) protocol

There is no support currently for scanning Cloud Volumes ONTAP, Azure NetApp Files, FSx for ONTAP, AWS S3, or Google Drive, OneDrive, or SharePoint Online 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:

- Managing Microsoft Azure Information Protection (AIP) labels
- Sending email alerts to BlueXP users when certain critical Policies return results
- Setting BlueXP roles for different users (for example, Account Admin or Compliance Viewer)
- Copying and synchronizing source files using BlueXP copy and sync
- Receiving user feedback
- 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.

4

Subscribe to the BlueXP classification service

The first 1 TB of data that BlueXP classification scans in BlueXP is free for 30 days. A BYOL license from NetApp is required to continue scanning data after that point.

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 three system sizes depending on the size of the data set that you plan to have BlueXP classification scan.

System size	CPU	RAM (swap memory must be disabled)	Disk
Extra Large	32 CPUs	128 GB RAM	1 TiB SSD on /, or - 100 GiB available on /opt - 895 GiB available on /var/lib/docker - 5 GiB on /tmp

System size	CPU	RAM (swap memory must be disabled)	Disk
Large	16 CPUs	64 GB RAM	500 GiB SSD on /, or - 100 GiB available on /opt - 395 GiB available on /var/lib/docker - 5 GiB on /tmp
Medium	8 CPUs	32 GB RAM	200 GiB SSD on /, or - 50 GiB available on /opt - 145 GiB available on /var/lib/docker - 5 GiB on /tmp
Small	8 CPUs	16 GB RAM	100 GiB SSD on /, or - 50 GiB available on /opt - 45 GiB available on /var/lib/docker - 5 GiB on /tmp

Note that there are limitations when using the smaller systems. See Using a smaller instance type for details.

- When deploying a compute instance in the cloud for your BlueXP classification installation, we recommend a system that meets the "Large" system requirements above:
  - **AWS 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 1.26 or greater:
    - Red Hat Enterprise Linux version 9.0, 9.1, and 9.2

      Note that the following features are not currently supported when using 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, update your system packages (`sudo yum update -y`), and then install Podman (`sudo yum install podman -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.

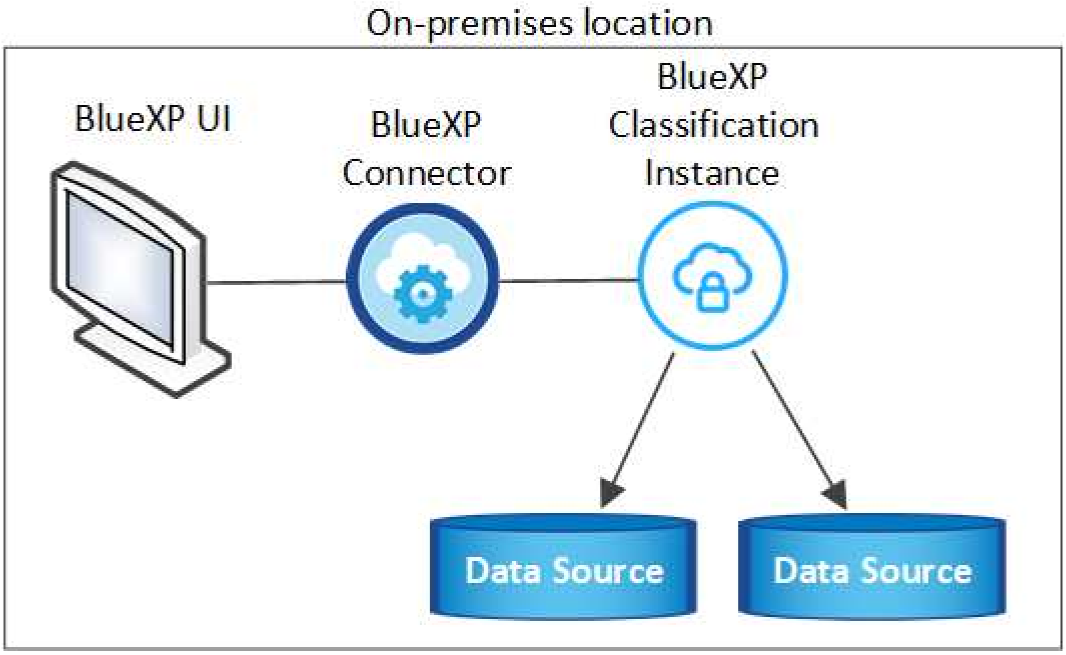
Connection Type	Ports	Description
Connector <> BlueXP classification	8080 (TCP), 6000 (TCP), 443 (TCP), and 80	The security group for the Connector must allow inbound and outbound traffic over ports 6000 and 443 to and from the BlueXP classification instance. • Port 6000 is required so that the BlueXP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

Connection Type	Ports	Description
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)

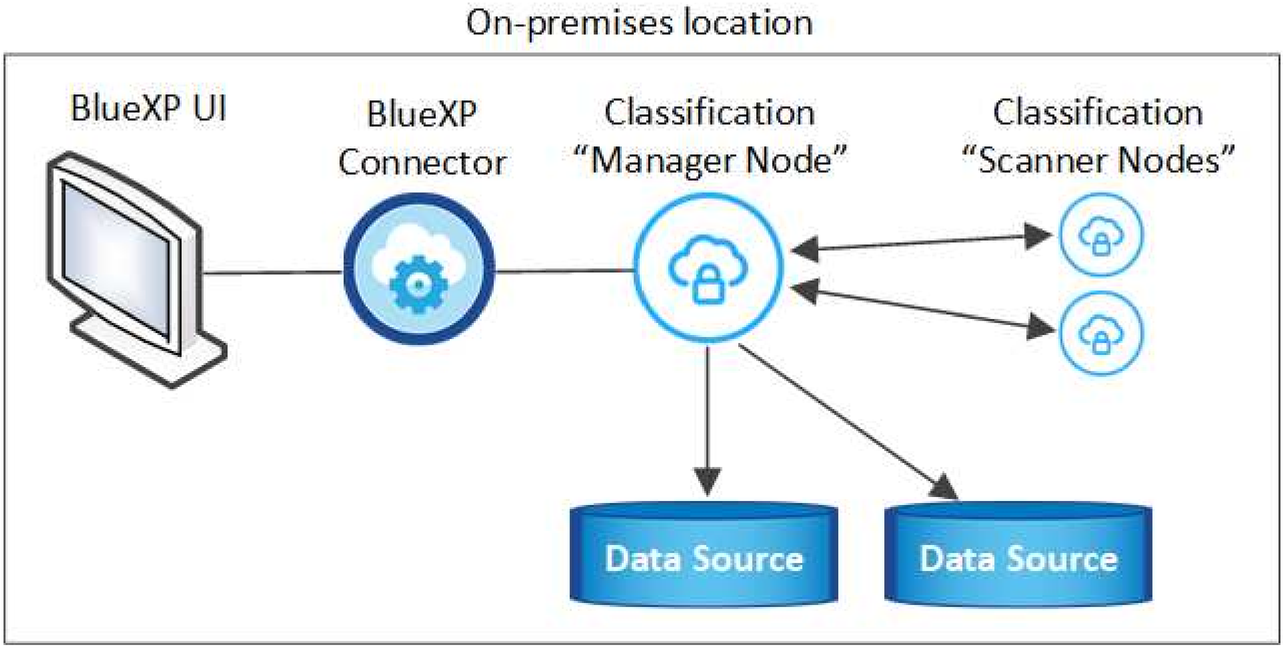
If you are using multiple BlueXP classification hosts to provide additional processing power to scan your data sources, you’ll need to enable additional ports/protocols. See the additional port requirements.

Install BlueXP classification on the on-premises Linux host

For typical configurations you’ll install the software on a single host system. 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. See those steps here.



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:

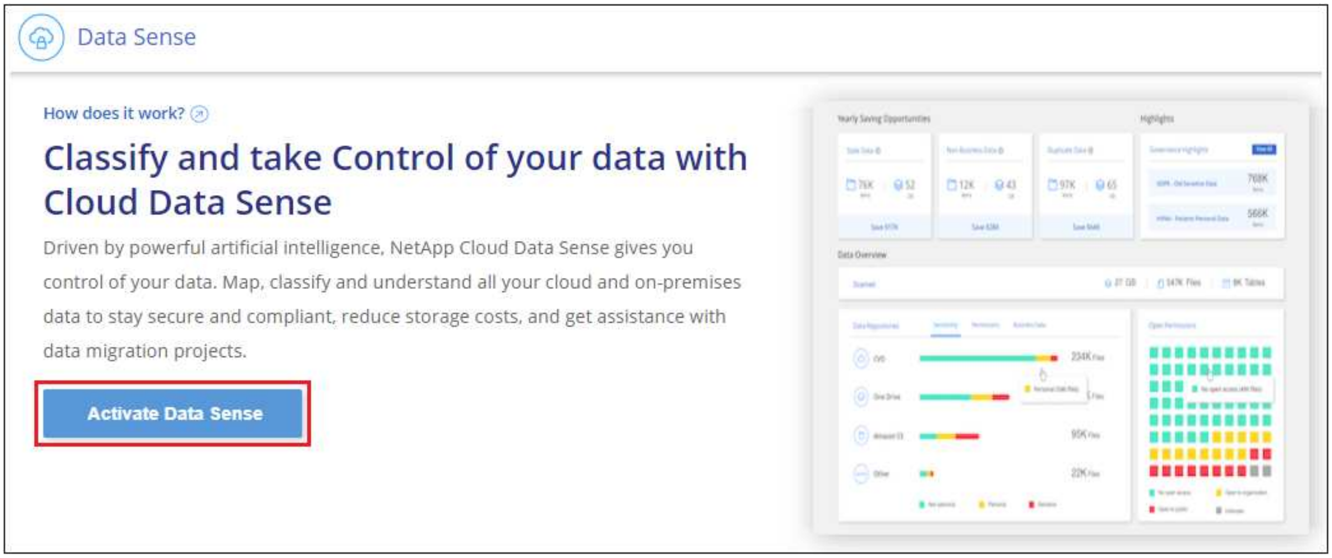
   ```
   tar -xzf DataSense-offline-bundle-v1.25.0.tar.gz
   ```

   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**.



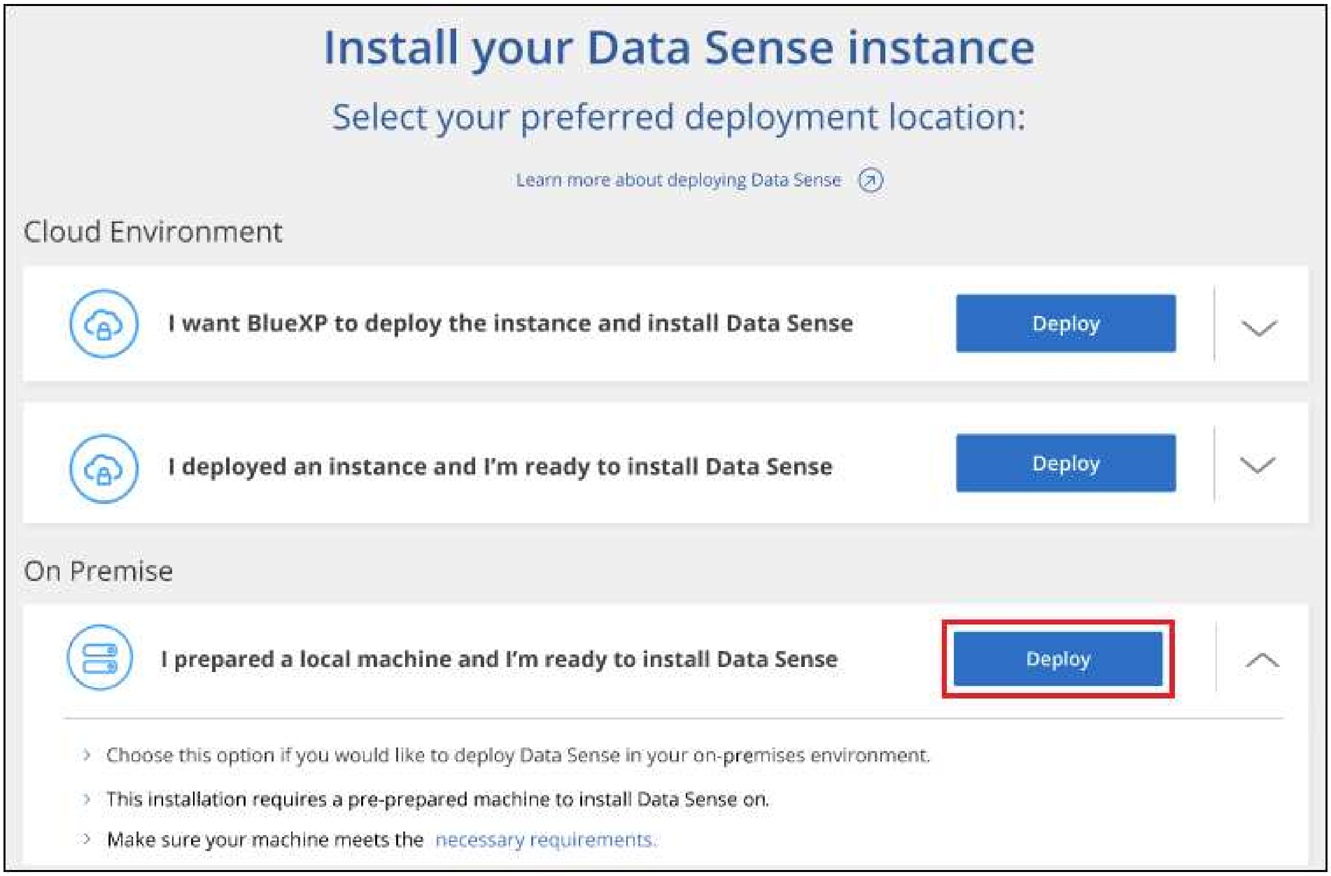
Data Sense

How does it work?

Classify and take Control of your data with Cloud Data Sense

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

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



Install your Data Sense instance

Select your preferred deployment location:

Learn more about deploying Data Sense

Cloud Environment

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

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

On Premise

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

- Choose this option if you would like to deploy Data Sense in your on-premises environment.
- This installation requires a pre-prepared machine to install Data Sense on.
- Make sure your machine meets the necessary requirements.

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.

You can also set up BYOL licensing for BlueXP classification from the BlueXP digital wallet page at this time. You will not be charged until your 30-day free trial ends.

Multi-host installation for large configurations

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

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

What you'll need

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

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

Steps

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

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

   For example, this command adds 3 scanner nodes:
   `sudo ./install.sh -a <account_id> -c <client_id> -t <user_token> --host <ds_host> --manager-host <cm_host>` **`-n <node_ip1>,<node_ip2>,<node_ip3>`** `--no-proxy --darksite`

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

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

Result

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

What's Next

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

You can also set up BYOL licensing for BlueXP classification from the BlueXP digital wallet page at this time. You will not be charged until your 30-day free trial ends.

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.

Note that no upgrade is required on scanner nodes if you have deployed BlueXP classification on multiple hosts systems for scanning very large configurations.

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 three system sizes depending on the size of the data set 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 - 5 GiB on /tmp

System size	CPU	RAM (swap memory must be disabled)	Disk
Medium	8 CPUs	32 GB RAM	200 GiB SSD on /, or - 50 GiB available on /opt - 145 GiB available on /var/lib/docker - 5 GiB on /tmp
Small	8 CPUs	16 GB RAM	100 GiB SSD on /, or - 50 GiB available on /opt - 45 GiB available on /var/lib/docker - 5 GiB on /tmp

Note that there are limitations when using the smaller systems. See Using a smaller instance type for details.

- When deploying a compute instance in the cloud for your BlueXP classification installation, we recommend a system that meets the "Large" system requirements above:
  - **AWS 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 1.26 or greater:
    - Red Hat Enterprise Linux version 9.0, 9.1, and 9.2

      Note that the following features are not currently supported when using 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, update your system packages (`sudo yum update -y`), and then install Podman (`sudo yum install podman -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.
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.

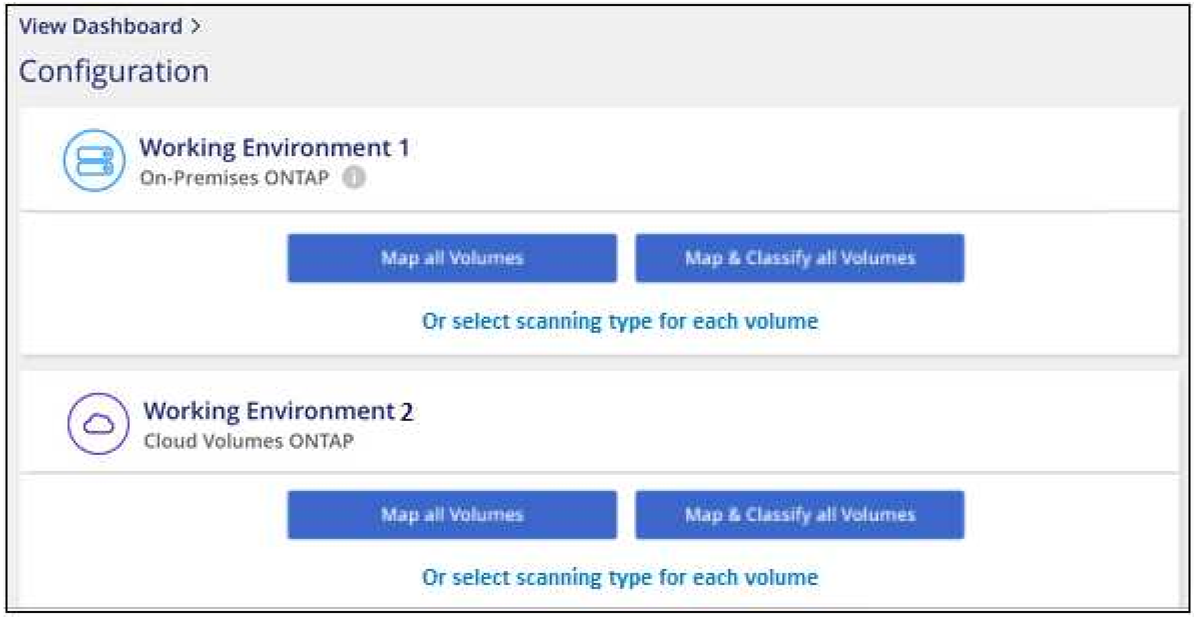
If you are scanning on-premises ONTAP systems that have been installed in a dark site that has no internet access, you need to deploy BlueXP classification in the same on-premises location that has no internet access. This also requires that the BlueXP Connector is deployed in that same on-premises location.

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

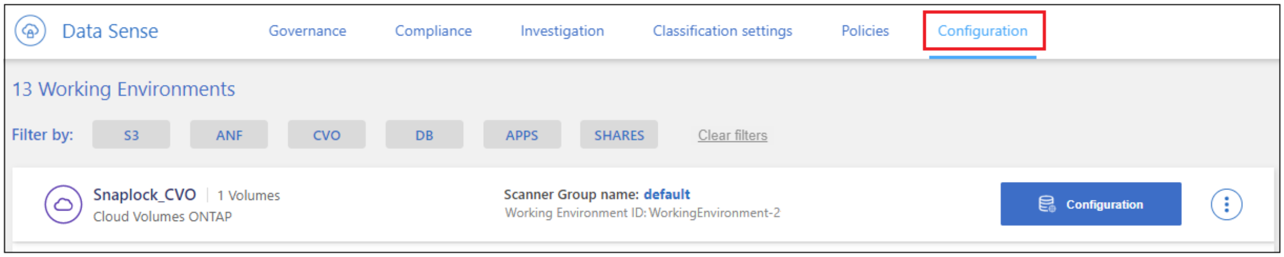
Make sure that BlueXP classification can access volumes by checking your networking, security groups, and export policies. You'll need to provide BlueXP classification with CIFS credentials so it can access CIFS volumes.

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.



13 Working Environments

Snaplock_CVO | 1 Volumes
Cloud Volumes ONTAP

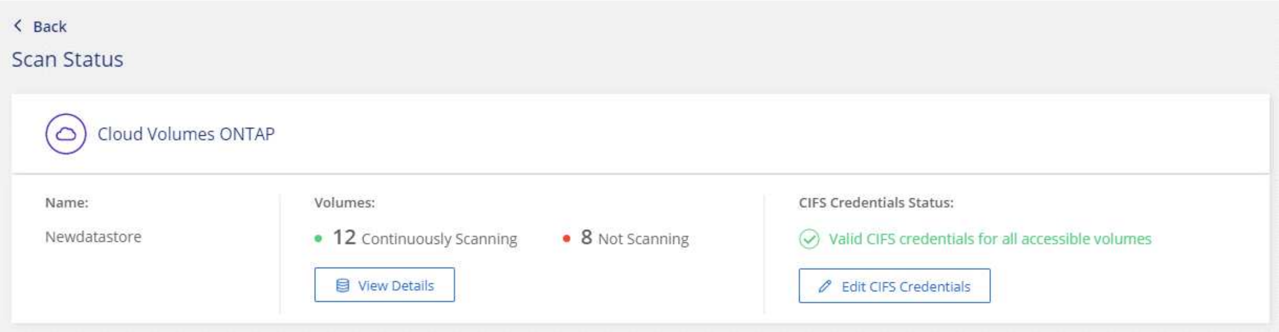
Scanner Group name: default
Working Environment ID: WorkingEnvironment-2

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.



Scan Status

Cloud Volumes ONTAP

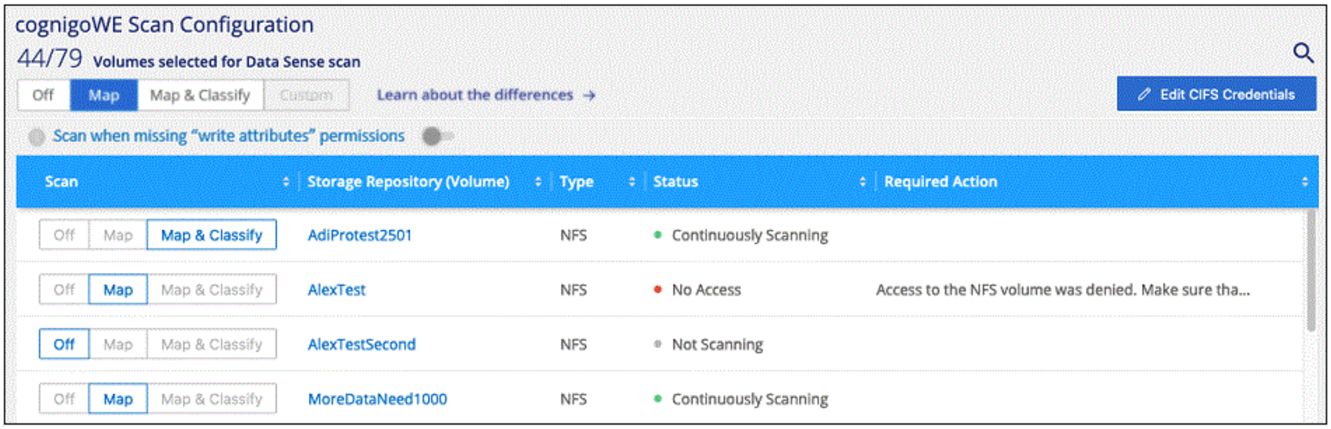
Name: Newdatastore

Volumes: ● 12 Continuously Scanning ● 8 Not Scanning

CIFS Credentials Status: ✓ Valid CIFS credentials for all accessible volumes

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.


cognigoWE Scan Configuration

44/79 Volumes selected for Data Sense scan

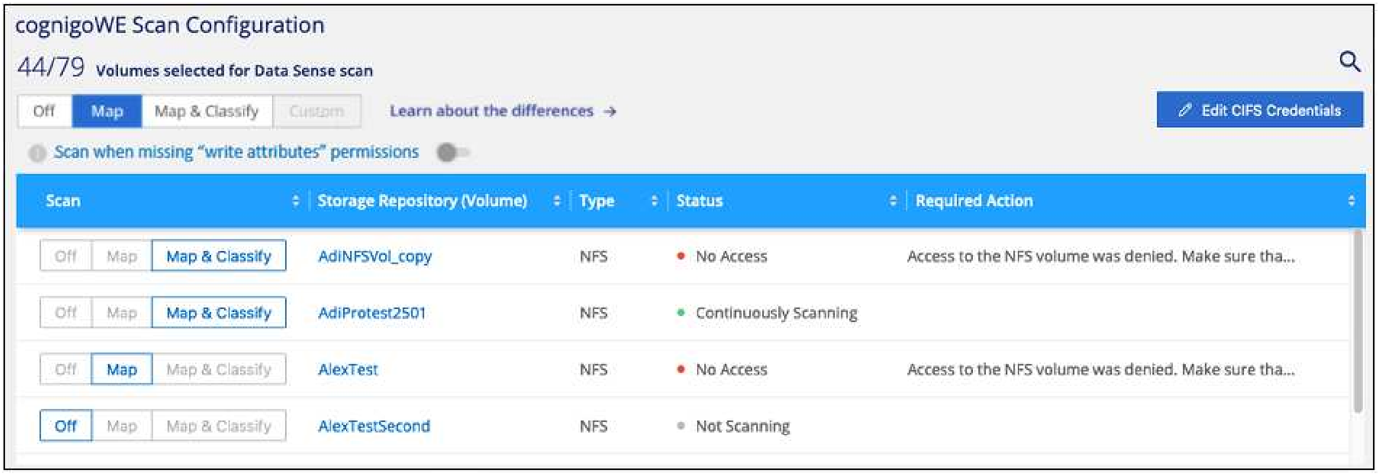
Scan when missing "write attributes" permissions

Scan	Storage Repository (Volume)	Type	Status	Required Action
Off / Map / **Map & Classify**	AdiProtest2501	NFS	● Continuously Scanning	
Off / **Map** / Map & Classify	AlexTest	NFS	● No Access	Access to the NFS volume was denied. Make sure tha...
Off / Map / Map & Classify	AlexTestSecond	NFS	● Not Scanning	
Off / **Map** / Map & Classify	MoreDataNeed1000	NFS	● Continuously Scanning	

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.



cognigoWE Scan Configuration

44/79 Volumes selected for Data Sense scan

Off [✓ Map] Map & Classify Custom — Learn about the differences → — Edit CIFS Credentials

Scan when missing "write attributes" permissions

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

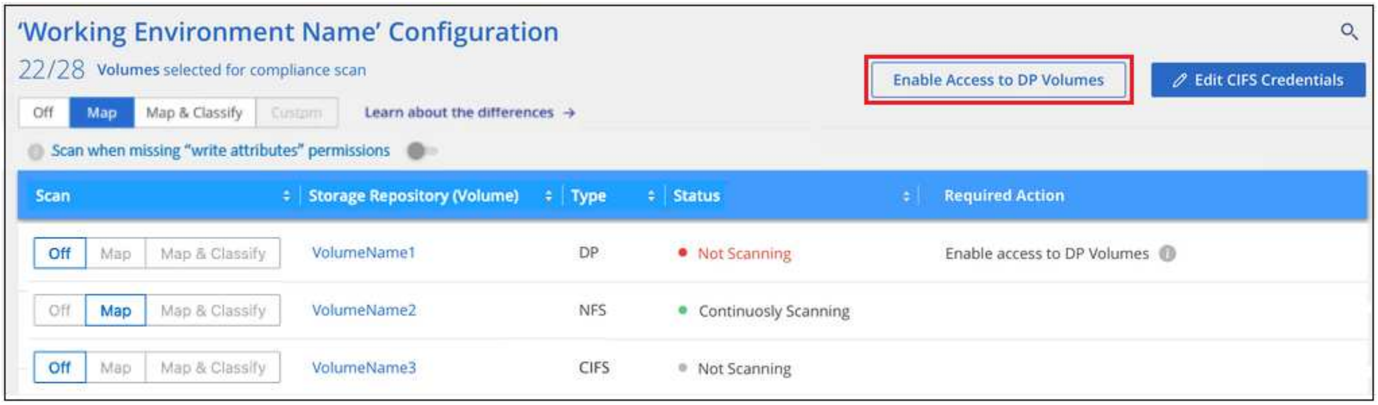
To:	Do this:
Enable mapping-only scans on a volume	In the volume area, click **Map**
Enable full scanning on a volume	In the volume area, click **Map & Classify**
Disable scanning on a volume	In the volume area, click **Off**
Enable mapping-only scans on all volumes	In the heading area, click **Map**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

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

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**.



'Working Environment Name' Configuration

22/28 Volumes selected for compliance scan

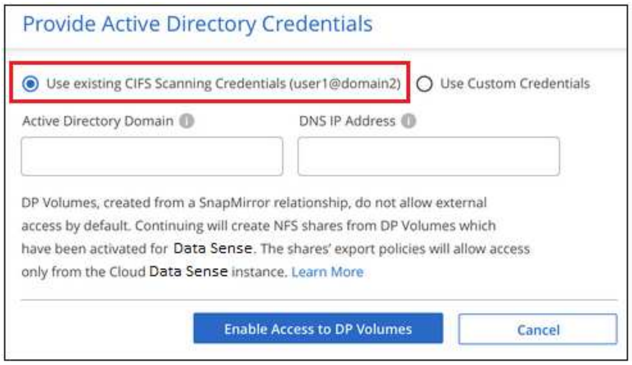
Scan when missing "write attributes" permissions

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

Steps

If you want to scan these data protection volumes:

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



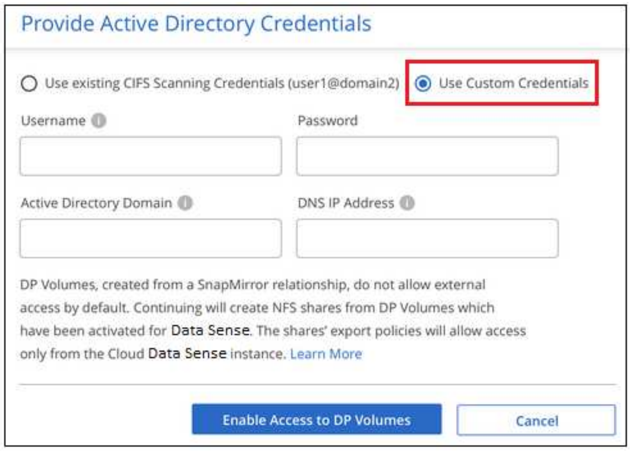
Provide Active Directory Credentials

● Use existing CIFS Scanning Credentials (user1@domain2) ○ Use Custom Credentials

Active Directory Domain

DNS IP Address

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



Provide Active Directory Credentials

○ Use existing CIFS Scanning Credentials (user1@domain2) ● Use Custom Credentials

Username

Password

Active Directory Domain

DNS IP Address

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

3. Activate each DP volume that you want to scan the same way you enabled other volumes.

Result

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

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

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

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.

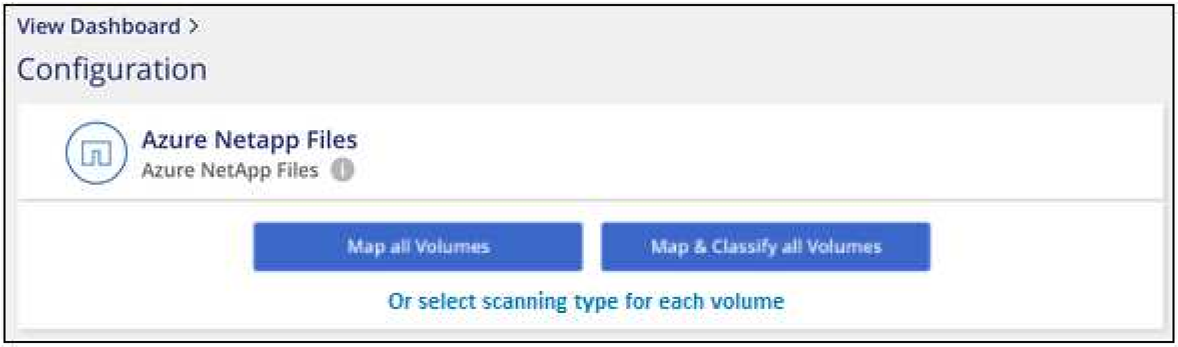
Note: Deploying BlueXP classification in an on-premises location is not currently supported when scanning Azure NetApp Files volumes.

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

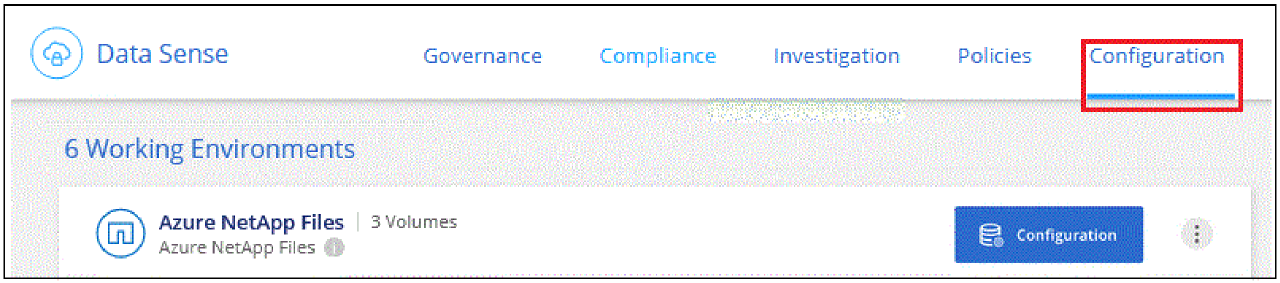
Make sure that BlueXP classification can access volumes by checking your networking, security groups, and export policies. You'll need to provide BlueXP classification with CIFS credentials so it can access CIFS volumes.

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.

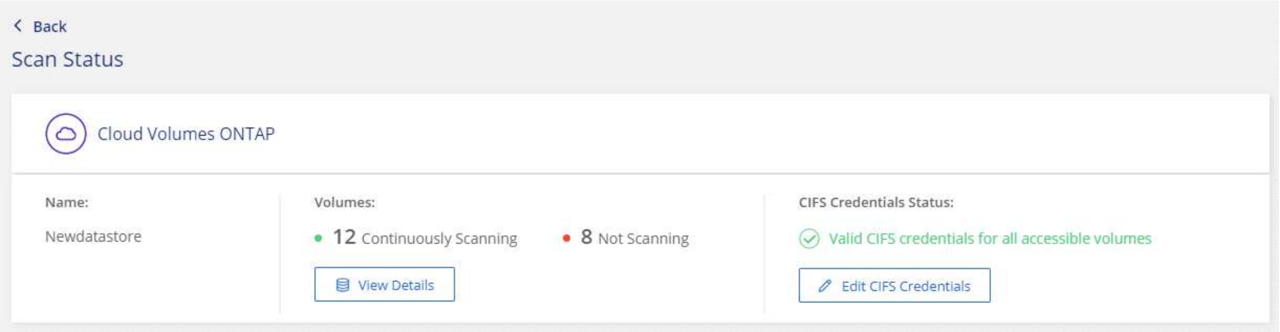


   b. For each working environment, click **Edit CIFS Credentials** and enter the user name and password that BlueXP classification needs to access CIFS volumes on the system.

   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.



< Back

Scan Status

Cloud Volumes ONTAP

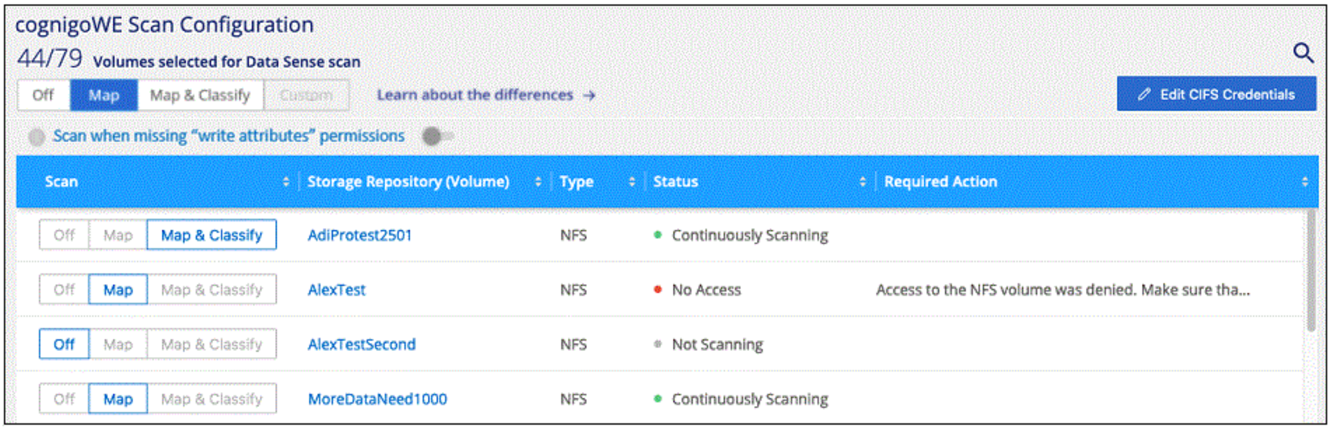
Name: Newdatastore

Volumes: ● 12 Continuously Scanning ● 8 Not Scanning

CIFS Credentials Status: ✓ Valid CIFS credentials for all accessible volumes

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.



cognigoWE Scan Configuration

44/79 Volumes selected for Data Sense scan

Off [✓ Map] Map & Classify Custom

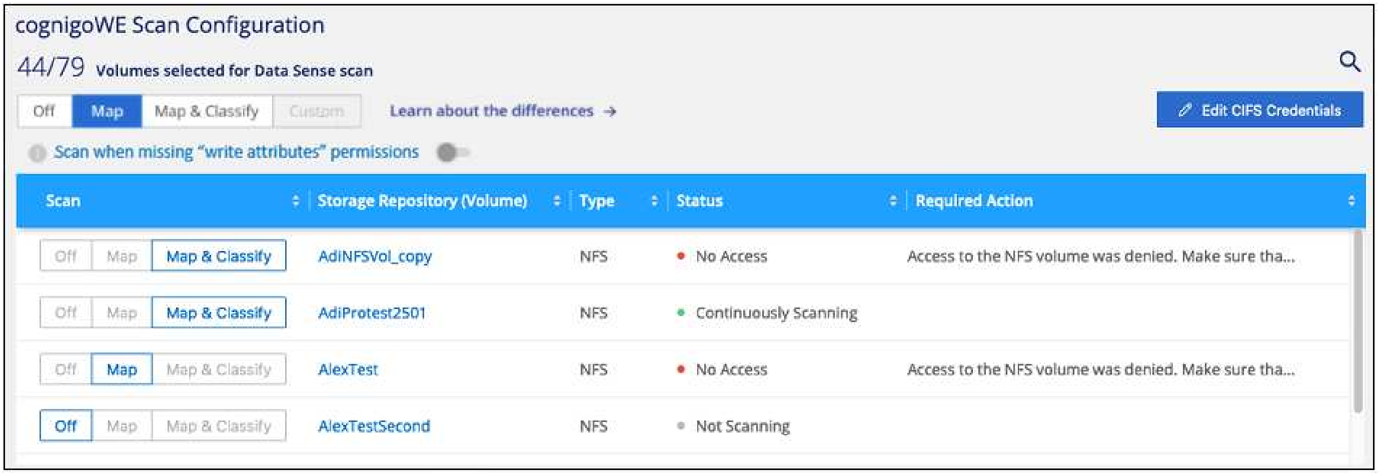
Scan when missing "write attributes" permissions ○

Scan	Storage Repository (Volume)	Type	Status	Required Action
Off Map [✓ Map & Classify]	AdiProtest2501	NFS	● Continuously Scanning	
Off [✓ Map] Map & Classify	AlexTest	NFS	● No Access	Access to the NFS volume was denied. Make sure tha...
[✓ Off] Map Map & Classify	AlexTestSecond	NFS	● Not Scanning	
Off [✓ Map] Map & Classify	MoreDataNeed1000	NFS	● Continuously Scanning	

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.



cognigoWE Scan Configuration

44/79 Volumes selected for Data Sense scan

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

○ Scan when missing "write attributes" permissions

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

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

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

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.

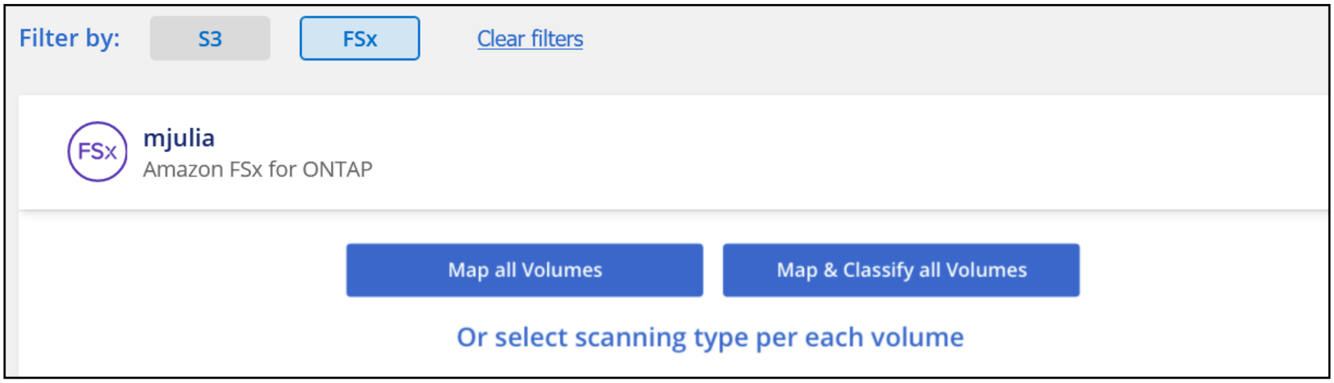
Note: Deploying BlueXP classification in an on-premises location is not currently supported when scanning FSx volumes.

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

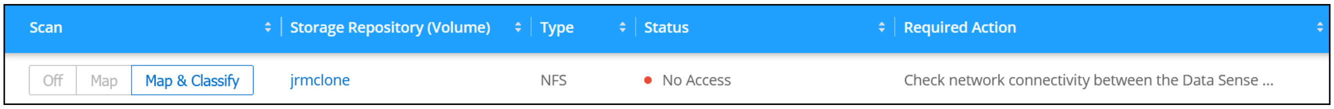
Make sure BlueXP classification can access volumes by checking your networking, security groups, and export policies.

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.

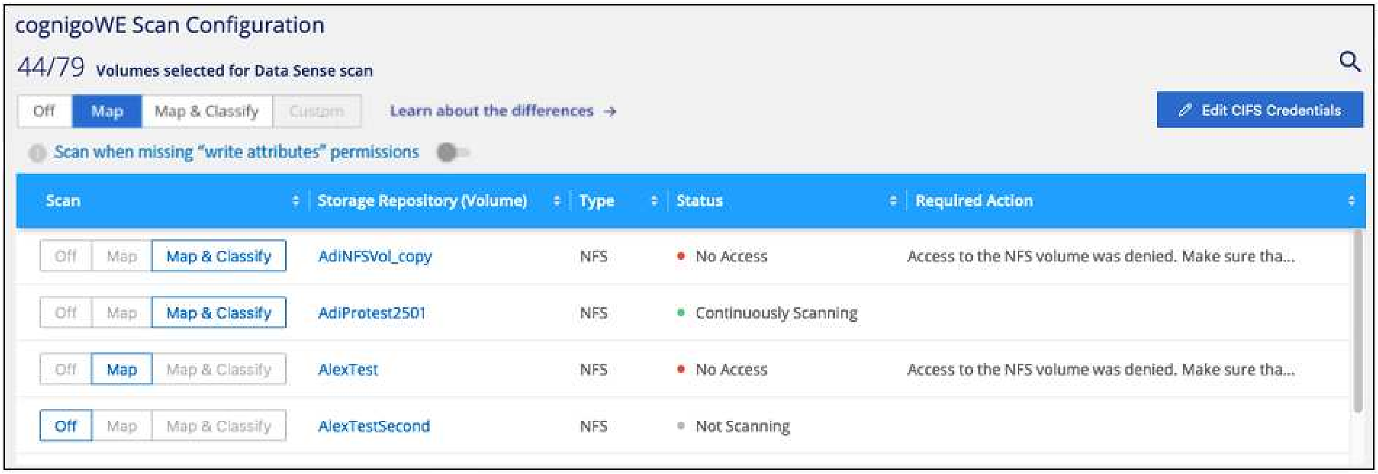
      If you want to make sure your files "last accessed times" are unchanged by BlueXP classification scans, we recommend that the user has Write Attributes permissions in CIFS or write permissions in NFS. If possible, we recommend making the Active Directory configured user part of a parent group in the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



cognigoWE Scan Configuration

44/79 Volumes selected for Data Sense scan

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

Scan when missing "write attributes" permissions

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

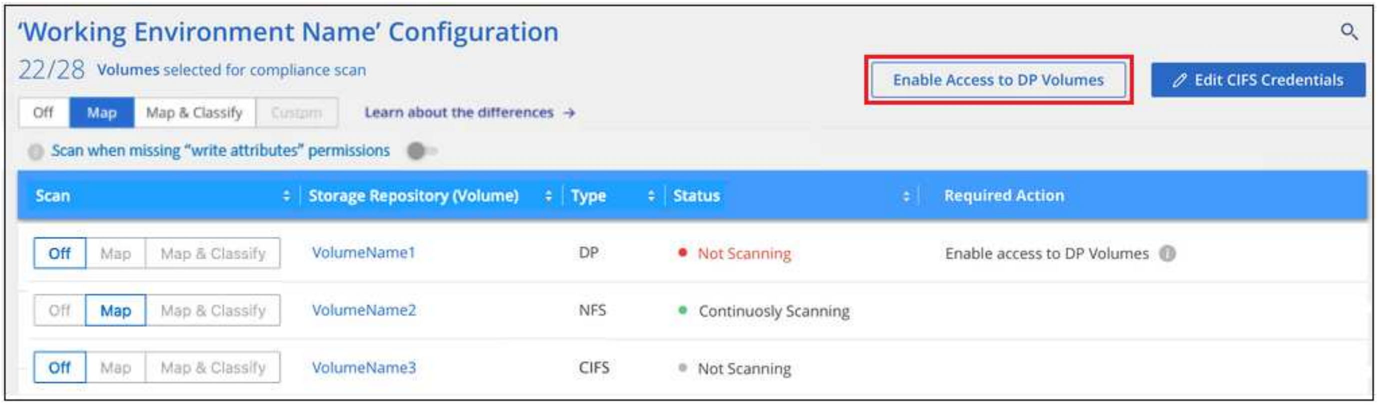
To:	Do this:
Enable mapping-only scans on a volume	In the volume area, click **Map**
Enable full scanning on a volume	In the volume area, click **Map & Classify**
Disable scanning on a volume	In the volume area, click **Off**
Enable mapping-only scans on all volumes	In the heading area, click **Map**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

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

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**.



'Working Environment Name' Configuration

22/28 Volumes selected for compliance scan

Off | Map | Map & Classify | Custom
Learn about the differences →
Enable Access to DP Volumes
Edit CIFS Credentials

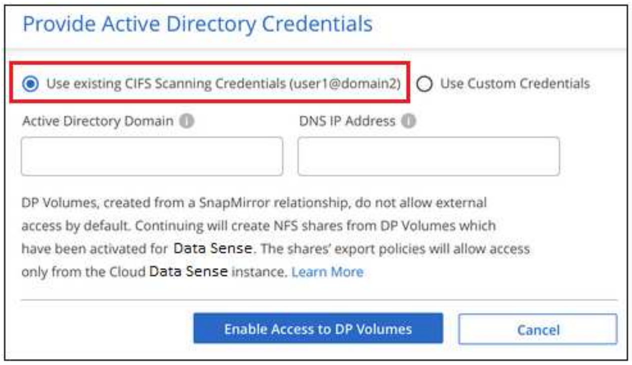
Scan when missing "write attributes" permissions

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

Steps

If you want to scan these data protection volumes:

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



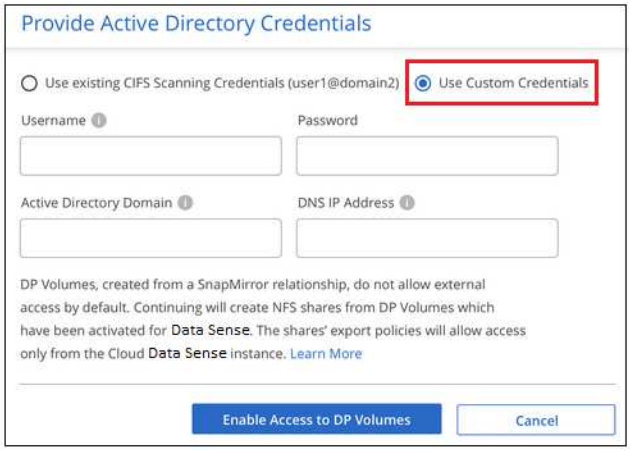
Provide Active Directory Credentials

● Use existing CIFS Scanning Credentials (user1@domain2) ○ Use Custom Credentials

Active Directory Domain

DNS IP Address

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



Provide Active Directory Credentials

○ Use existing CIFS Scanning Credentials (user1@domain2) ● Use Custom Credentials

Username

Password

Active Directory Domain

DNS IP Address

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

3. Activate each DP volume that you want to scan the same way you enabled other volumes.

Result

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

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

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

Getting started with BlueXP classification for Amazon S3

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

Quick start

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

❶ **Set up the S3 requirements in your cloud environment**

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

❷ **Deploy the BlueXP classification instance**

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

❸ **Activate BlueXP classification on your S3 working environment**

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

❹ **Select the buckets to scan**

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

Reviewing S3 prerequisites

The following requirements are specific to scanning S3 buckets.

Set up an IAM role for the BlueXP classification instance

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

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

Provide connectivity from BlueXP classification to Amazon S3

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

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

If you experience any issues, see AWS Support Knowledge Center: Why can’t I connect to an S3 bucket using a gateway VPC endpoint?

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

> ⓘ You can’t use a proxy to get to S3 over the internet.

Deploying the BlueXP classification instance

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

You need to deploy the instance using a Connector deployed in AWS so that BlueXP automatically discovers the S3 buckets in this AWS account and displays them in an Amazon S3 working environment.

Note: Deploying BlueXP classification in an on-premises location is not currently supported when scanning S3 buckets.

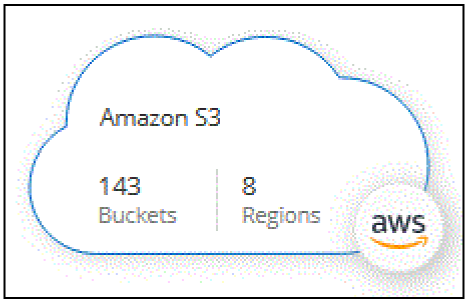
Upgrades to BlueXP classification software is automated as long as the instance has internet connectivity.

Activating BlueXP classification on your S3 working environment

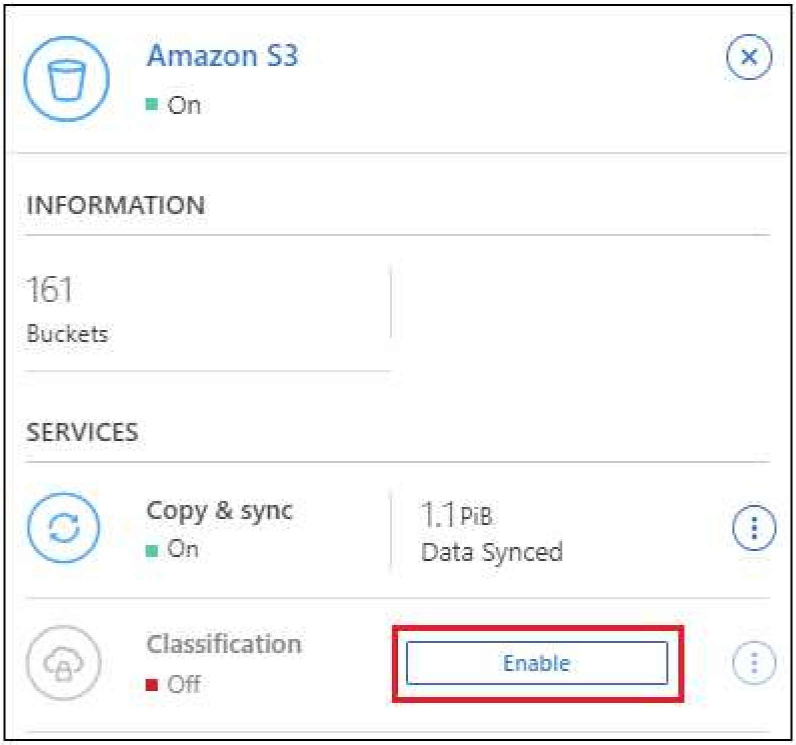
Enable BlueXP classification on Amazon S3 after you verify the prerequisites.

Steps

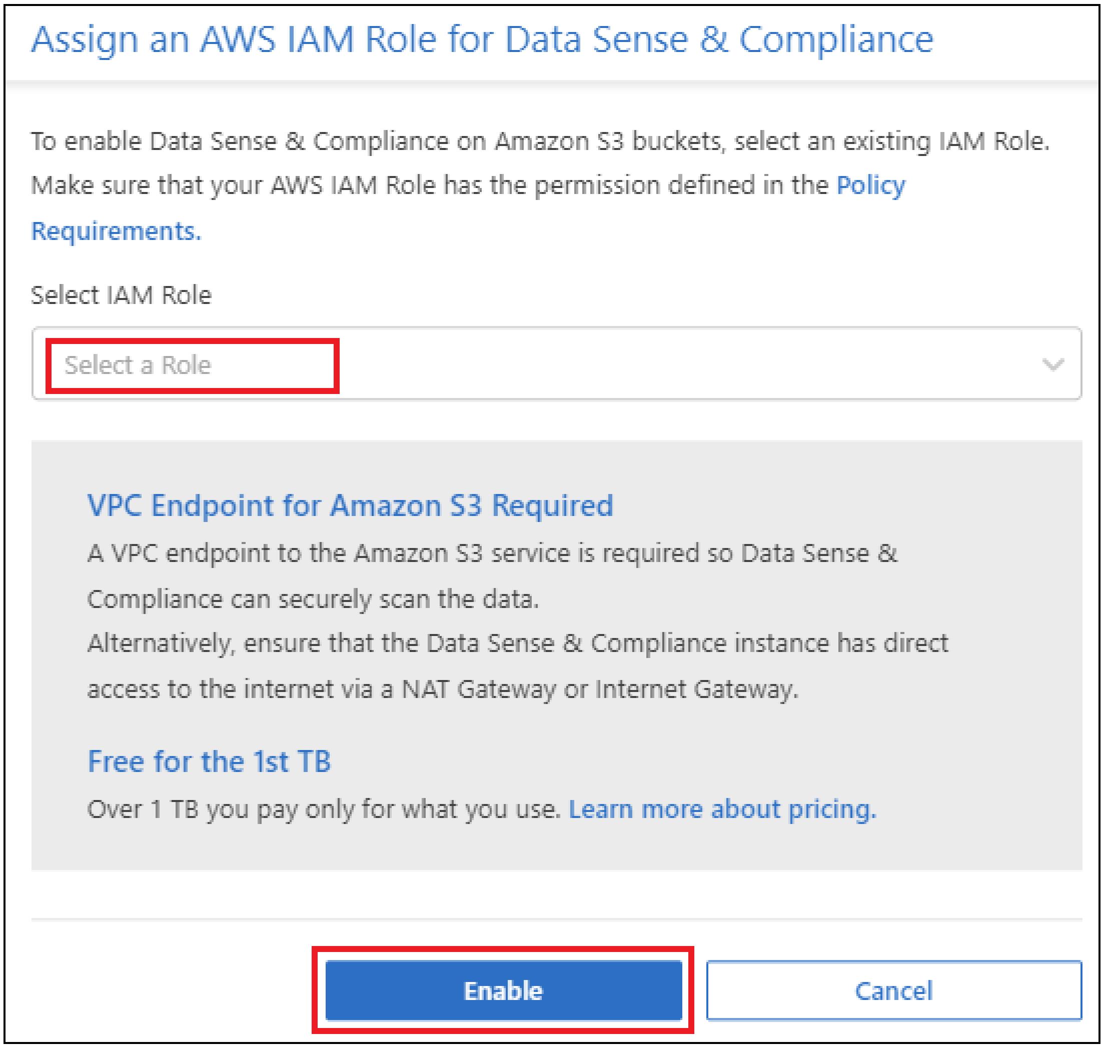
1. From the BlueXP left navigation menu, click **Storage > Canvas**.
2. Select the Amazon S3 working environment.



3. In the Services pane on the right, click **Enable** next to **Classification**.



4. When prompted, assign an IAM role to the BlueXP classification instance that has the required permissions.



Assign an AWS IAM Role for Data Sense & Compliance

To enable Data Sense & Compliance on Amazon S3 buckets, select an existing IAM Role. Make sure that your AWS IAM Role has the permission defined in the Policy Requirements.

Select IAM Role

Select a Role

VPC Endpoint for Amazon S3 Required

A VPC endpoint to the Amazon S3 service is required so Data Sense & Compliance can securely scan the data.
Alternatively, ensure that the Data Sense & Compliance instance has direct access to the internet via a NAT Gateway or Internet Gateway.

Free for the 1st TB

Over 1 TB you pay only for what you use. Learn more about pricing.

5. Click **Enable**.

> You can also enable compliance scans for a working environment from the Configuration page by clicking the ⋮ button and selecting **Activate BlueXP classification**.

Result

BlueXP assigns the IAM role to the instance.

Enabling and disabling compliance scans on S3 buckets

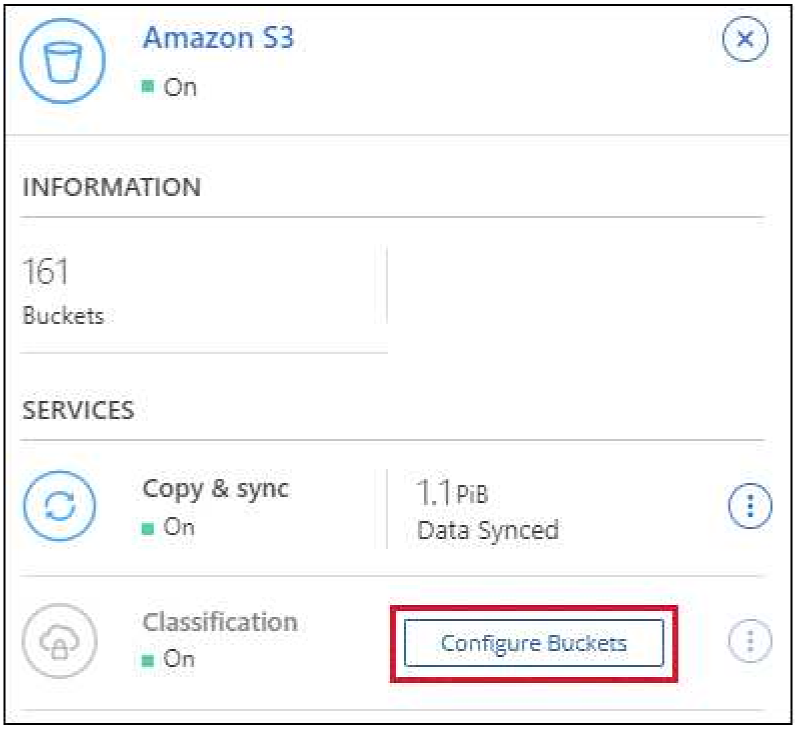
After BlueXP enables BlueXP classification on Amazon S3, the next step is to configure the buckets that you want to scan.

When BlueXP is running in the AWS account that has the S3 buckets you want to scan, it discovers those buckets and displays them in an Amazon S3 working environment.

BlueXP classification can also scan S3 buckets that are in different AWS accounts.

Steps

1. Select the Amazon S3 working environment.
2. In the Services pane on the right, click **Configure Buckets**.



Amazon S3

🟩 On

INFORMATION

161
Buckets

SERVICES

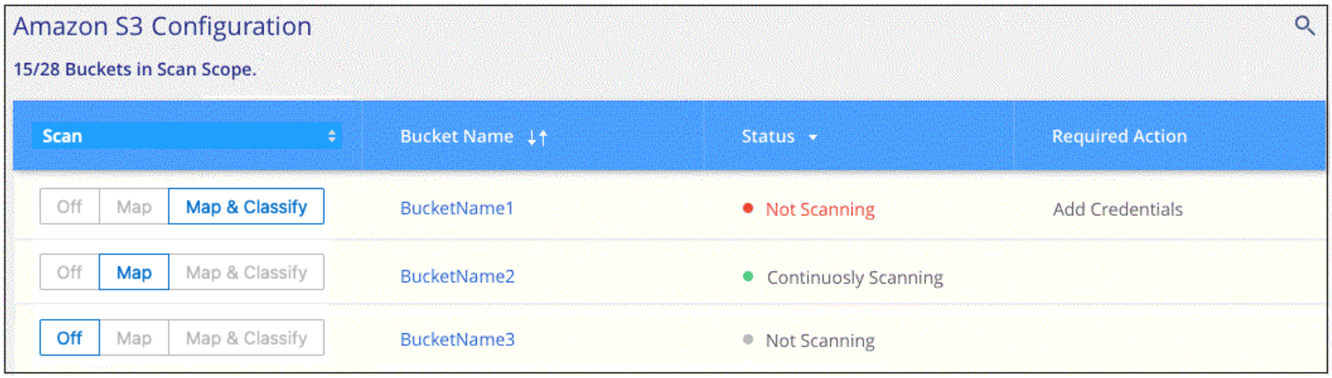
Copy & sync
🟩 On

1.1 PiB
Data Synced

Classification
🟩 On

Configure Buckets

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



Amazon S3 Configuration

15/28 Buckets in Scan Scope.

Scan	Bucket Name	Status	Required Action
Off / Map / **Map & Classify**	BucketName1	● Not Scanning	Add Credentials
Off / **Map** / Map & Classify	BucketName2	● Continuosly Scanning	
Off / Map / Map & Classify	BucketName3	● Not Scanning	

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

Result

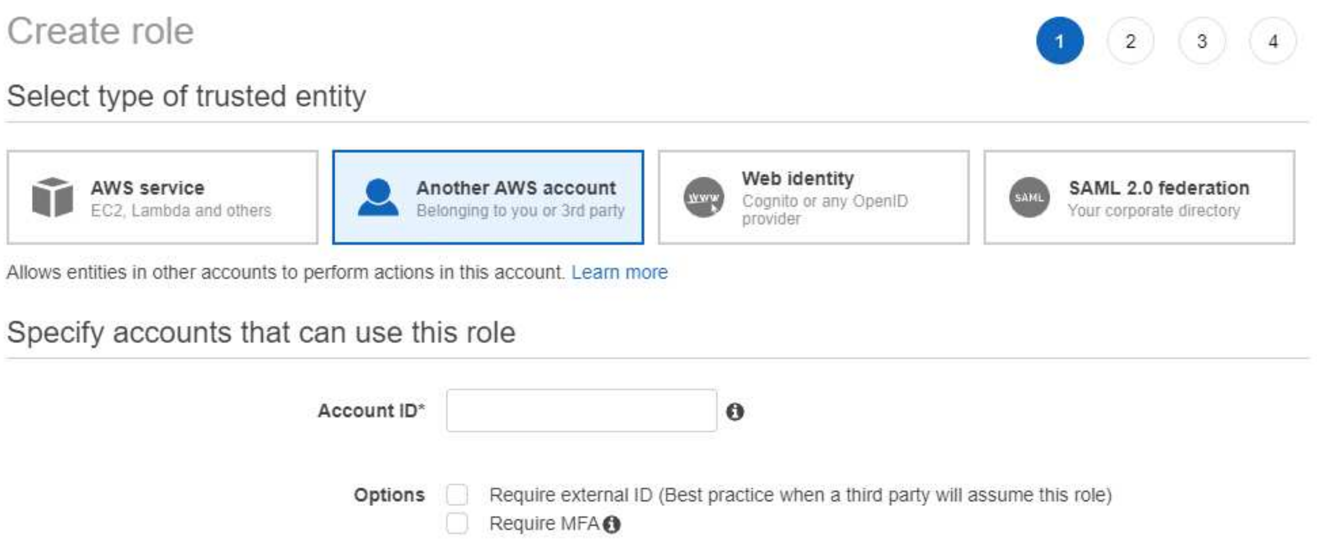
BlueXP classification starts scanning the S3 buckets that you enabled. If there are any errors, they'll appear in the Status column, alongside the required action to fix the error.

Scanning buckets from additional AWS accounts

You can scan S3 buckets that are under a different AWS account by assigning a role from that account to access the existing BlueXP classification instance.

Steps

1. Go to the target AWS account where you want to scan S3 buckets and create an IAM role by selecting **Another AWS account**.



Allows entities in other accounts to perform actions in this account. Learn more

Specify accounts that can use this role

Account ID*

Options
□ Require external ID (Best practice when a third party will assume this role)
□ Require MFA

Be sure to do the following:

- Enter the ID of the account where the BlueXP classification instance resides.
- Change the **Maximum CLI/API session duration** from 1 hour to 12 hours and save that change.
- Attach the BlueXP classification IAM policy. Make sure it has the required permissions.

```
{
  "Version": "2012-10-17",
  "Statement": [
      {
          "Effect": "Allow",
          "Action": [
              "s3:Get*",
              "s3:List*",
              "s3:PutObject"
          ],
          "Resource": "*"
      },
  ]
}
```

2. Go to the source AWS account where the BlueXP classification instance resides and select the IAM role that is attached to the instance.

   a. Change the **Maximum CLI/API session duration** from 1 hour to 12 hours and save that change.

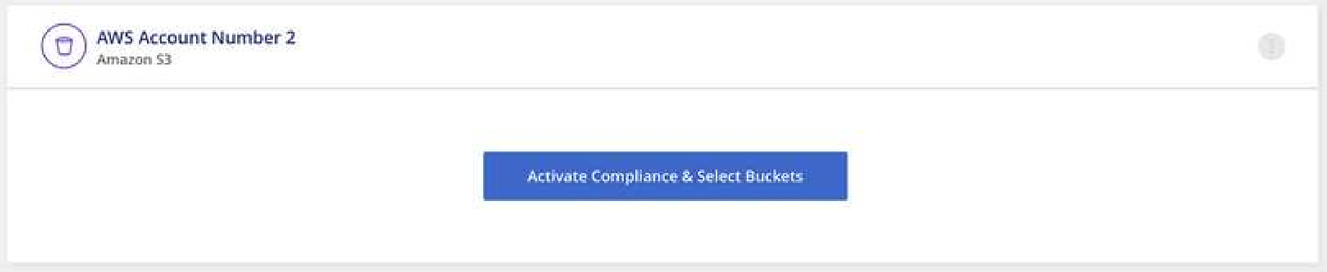
   b. Click **Attach policies** and then click **Create policy**.

   c. Create a policy that includes the "sts:AssumeRole" action and specify the ARN of the role that you created in the target account.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "sts:AssumeRole",
            "Resource": "arn:aws:iam::<ADDITIONAL-ACCOUNT-
ID>:role/<ADDITIONAL_ROLE_NAME>"
        },
        {
            "Effect": "Allow",
            "Action": [
                "iam:GetPolicyVersion",
                "iam:GetPolicy",
                "iam:ListAttachedRolePolicies"
            ],
            "Resource": [
                "arn:aws:iam::*:policy/*",
                "arn:aws:iam::*:role/*"
            ]
        }
    ]
}
```

The BlueXP classification instance profile account now has access to the additional AWS account.

3. Go to the **Amazon S3 Configuration** page and the new AWS account is displayed. Note that it can take a few minutes for BlueXP classification to sync the new account's working environment and show this information.



4. Click **Activate BlueXP classification & Select Buckets** and select the buckets you want to scan.

Result

BlueXP classification starts scanning the new S3 buckets that you enabled.

Scan database schemas

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

Note that after you have enabled database scanning that you can add unique identifiers that BlueXP classification will identify in all your data sources based on specific columns in your databases. This is called the *Data Fusion* feature. Learn how to add custom personal data identifiers from your databases.

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.

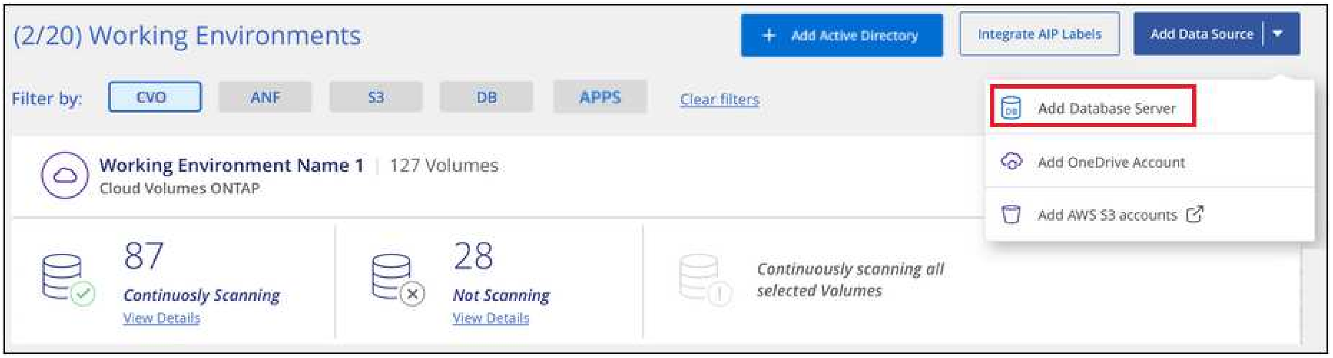
If you are scanning database schemas that have been installed in a dark site that has no internet access, you need to deploy BlueXP classification in the same on-premises location that has no internet access. This also requires that the BlueXP Connector is deployed in that same on-premises location.

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

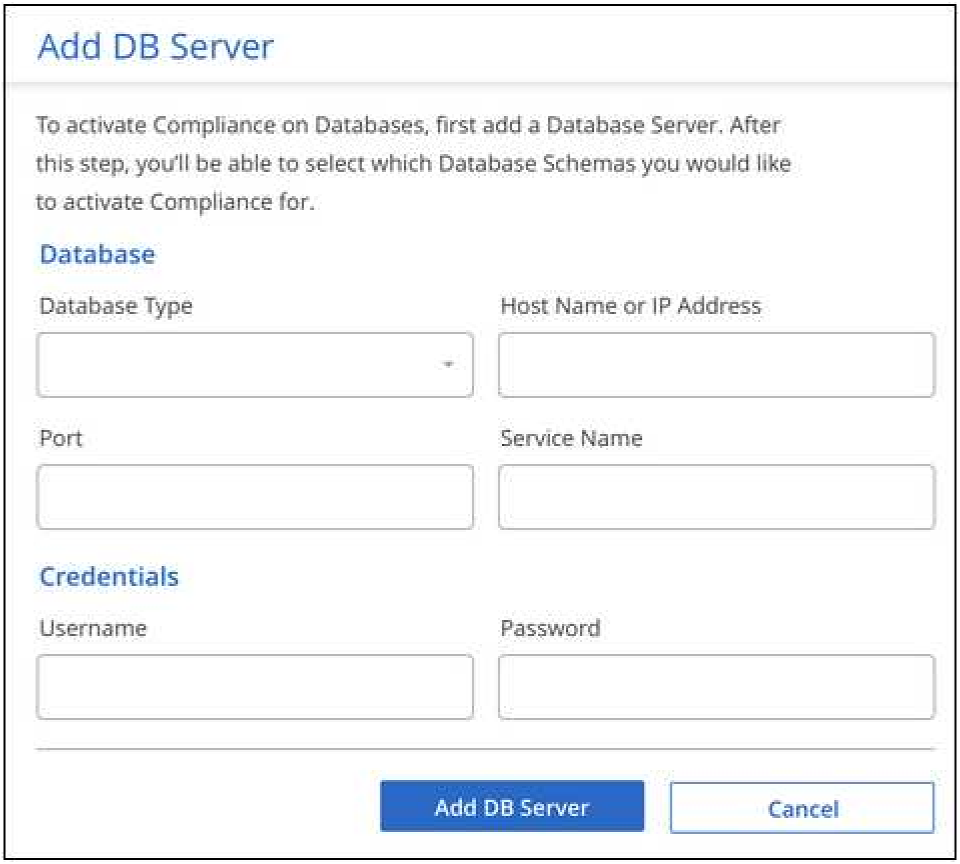
Add the database server

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**.



Add DB Server

To activate Compliance on Databases, first add a Database Server. After this step, you'll be able to select which Database Schemas you would like to activate Compliance for.

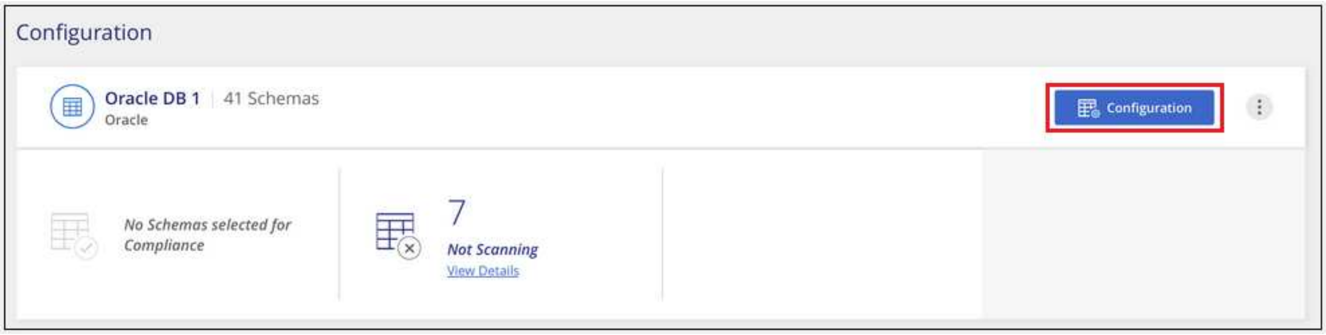
The database is added to the list of working environments.

Enable and disable compliance scans on database schemas

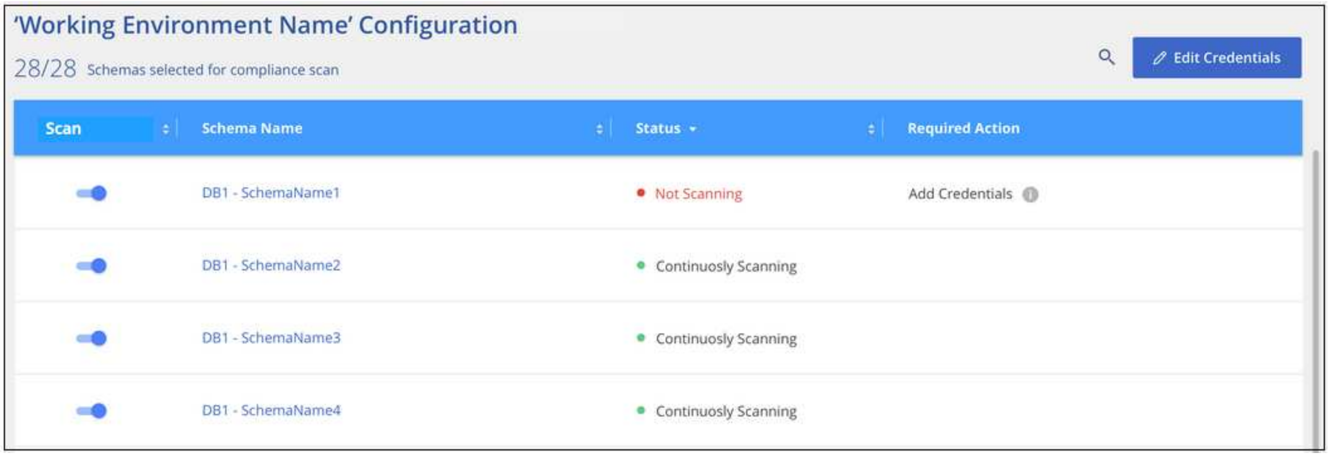
You can stop or start full scanning of your schemas at any time.

> 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.



'Working Environment Name' Configuration

28/28 Schemas selected for compliance scan

Scan	Schema Name	Status	Required Action
●	DB1 - SchemaName1	● Not Scanning	Add Credentials ⓘ
●	DB1 - SchemaName2	● Continuosly Scanning	
●	DB1 - SchemaName3	● Continuosly Scanning	
●	DB1 - SchemaName4	● Continuosly Scanning	

Result

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

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

Scanning OneDrive accounts

Complete a few steps to start scanning files in your user's OneDrive folders with BlueXP classification.

Quick start

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

1 **Review OneDrive prerequisites**

Ensure that you have the Admin credentials to log into the OneDrive account.

2 **Deploy the BlueXP classification instance**

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

3 **Add the OneDrive account**

Using Admin user credentials, log into the OneDrive account that you want to access so that it is added as a new working environment.

4 **Add the users and select the type of scanning**

Add the list of users from the OneDrive account that you want to scan and select the type of scanning. You can add up to 100 users at time.

Reviewing OneDrive requirements

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

- You must have the Admin login credentials for the OneDrive for Business account that provides read access to the user’s files.
- You’ll need a line-separated list of the email addresses for all the users whose OneDrive folders you want to scan.

Deploying the BlueXP classification instance

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

BlueXP classification can be deployed in the cloud or in an on-premises location that has internet access.

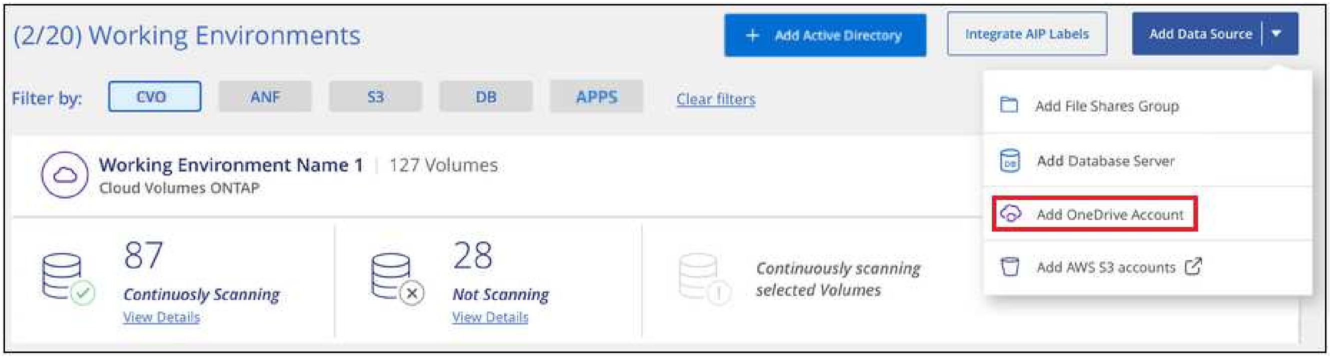
Upgrades to BlueXP classification software is automated as long as the instance has internet connectivity.

Adding the OneDrive account

Add the OneDrive account where the user files reside.

Steps

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



2. In the Add a OneDrive account dialog, click **Sign in to OneDrive**.
3. In the Microsoft page that appears, select the OneDrive account and enter the required Admin user and password, then click **Accept** to allow BlueXP classification to read data from this account.

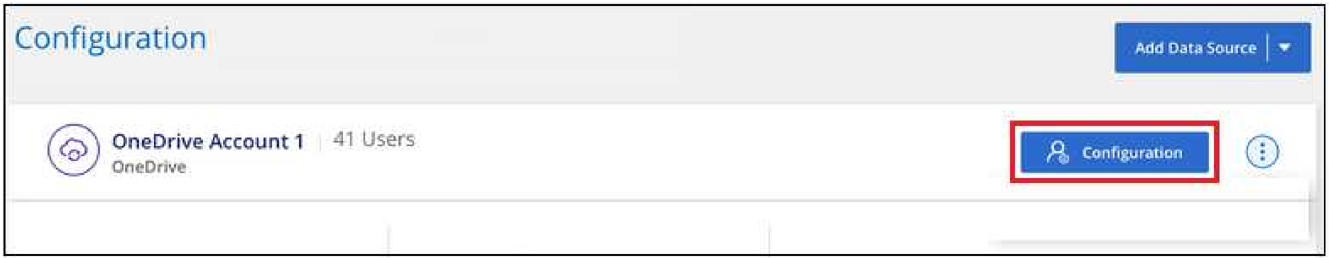
The OneDrive account is added to the list of working environments.

Adding OneDrive users to compliance scans

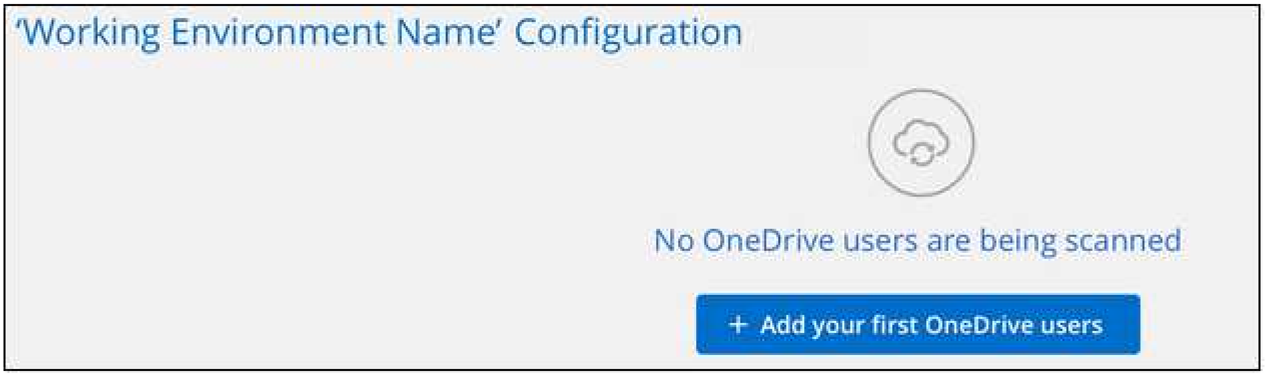
You can add individual OneDrive users, or all of your OneDrive users, so that their files will be scanned by BlueXP classification.

Steps

1. From the *Configuration* page, click the **Configuration** button for the OneDrive account.

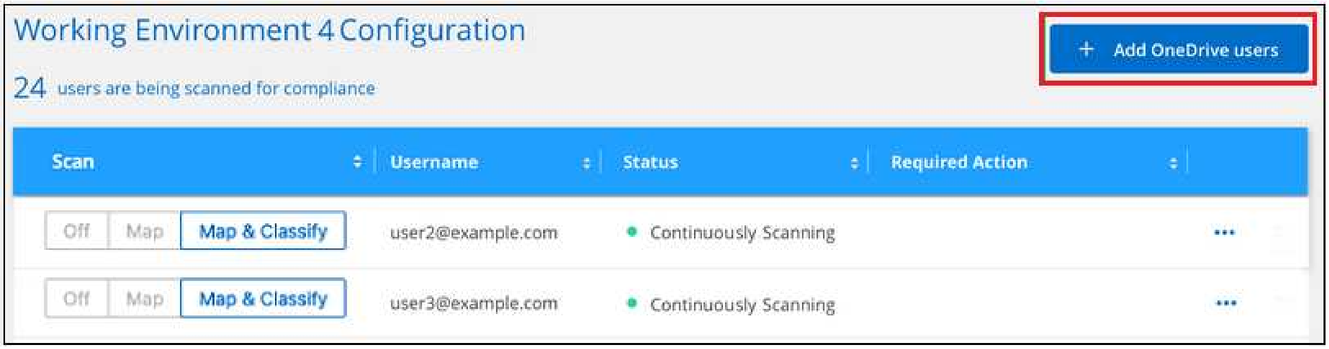


2. If this is the first time adding users for this OneDrive account, click **Add your first OneDrive users**.


'Working Environment Name' Configuration
+ Add your first OneDrive users

No OneDrive users are being scanned

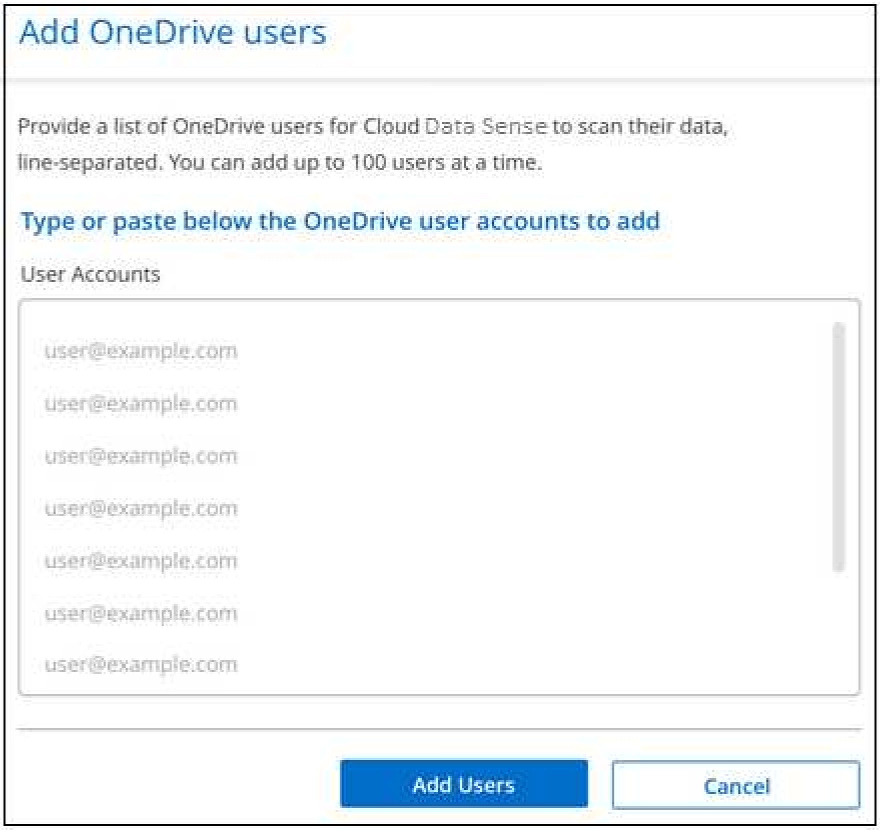
If you are adding additional users from a OneDrive account, click **Add OneDrive users**.



24 users are being scanned for compliance

Scan	Username	Status	Required Action	
Off / Map / Map & Classify	user2@example.com	● Continuously Scanning		•••
Off / Map / Map & Classify	user3@example.com	● Continuously Scanning		•••

3. Add the email addresses for the users whose files you want to scan - one email address per line (up to 100 maximum per session) - and click **Add Users**.



Add OneDrive users

Provide a list of OneDrive users for Cloud Data Sense to scan their data, line-separated. You can add up to 100 users at a time.

Type or paste below the OneDrive user accounts to add

User Accounts

user@example.com
user@example.com
user@example.com
user@example.com
user@example.com
user@example.com
user@example.com

A confirmation dialog displays the number of users who were added.

If the dialog lists any users who could not be added, capture this information so that you can resolve the issue. In some cases you can re-add the user with a corrected email address.

4. Enable mapping-only scans, or mapping and classification scans, on user files.

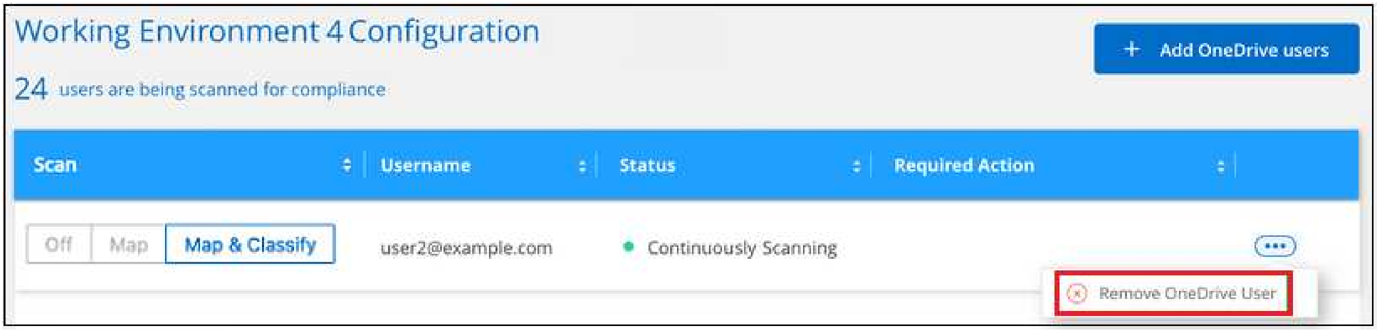
To:	Do this:
Enable mapping-only scans on user files	Click **Map**
Enable full scans on user files	Click **Map & Classify**
Disable scanning on user files	Click **Off**

Result

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

Removing a OneDrive user from compliance scans

If users leave the company or if their email address changes, you can remove individual OneDrive users from having their files scanned at any time. Just click **Remove OneDrive User** from the Configuration page.



Working Environment 4 Configuration

24 users are being scanned for compliance

Scan	Username	Status	Required Action
Off Map [✓ Map & Classify]	user2@example.com	Continuously Scanning	

Note that you can delete the entire OneDrive account from BlueXP classification if you no longer want to scan any user data from the OneDrive account.

Scanning SharePoint accounts

Complete a few steps to start scanning files in your SharePoint Online and SharePoint On-Premise accounts with BlueXP classification.

Quick start

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

1 **Review SharePoint prerequisites**

Ensure that you have qualified credentials to log into the SharePoint account, and that you have the URLs for the SharePoint sites that you want to scan.

2 **Deploy the BlueXP classification instance**

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

3 **Log into the SharePoint account**

Using qualified user credentials, log into the SharePoint account that you want to access so that it is added as a new data source/working environment.

4 **Add the SharePoint site URLs to scan**

Add the list of SharePoint site URLs that you want to scan in the SharePoint account, and select the type of scanning. You can add up to 100 URLs at time - and up to 1,000 sites total for each account.

Reviewing SharePoint requirements

Review the following prerequisites to make sure you are ready to activate BlueXP classification on a SharePoint account.

- You must have the Admin user login credentials for the SharePoint account that provides read access to all SharePoint sites.
    - For SharePoint Online you can use a non-Admin account, but that user must have permission to

access all the SharePoint sites that you want to scan.

- For SharePoint On-Premise, you'll also need the URL of the SharePoint Server.
- You will need a line-separated list of the SharePoint site URLs for all the data you want to scan.

Deploying the BlueXP classification instance

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

- For SharePoint Online, BlueXP classification can be deployed in the cloud.
- For SharePoint On-Premises, BlueXP classification can be installed in an on-premises location that has internet access or in an on-premises location that does not have internet access.

When BlueXP classification is installed in a site without internet access, the BlueXP Connector also must be installed in that same site without internet access. Learn more.

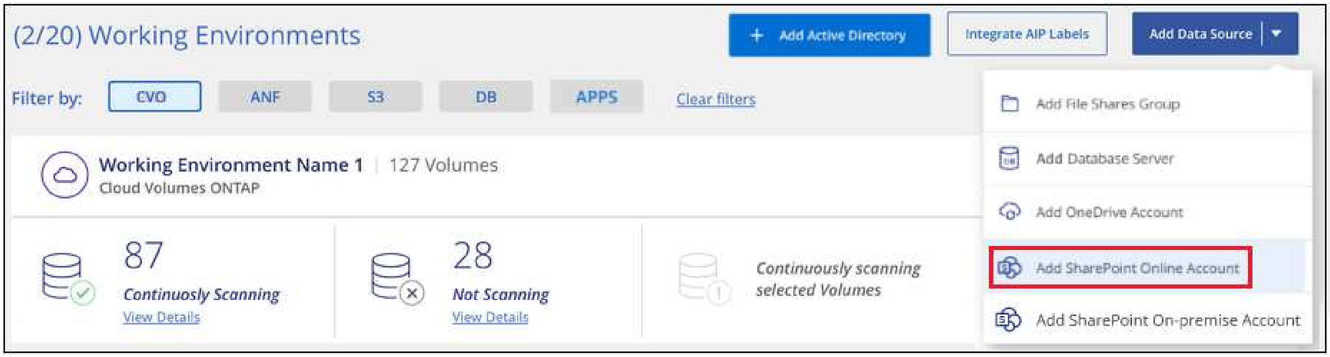
Upgrades to BlueXP classification software is automated as long as the instance has internet connectivity.

Adding a SharePoint Online account

Add the SharePoint Online account where the user files reside.

Steps

1. From the Working Environments Configuration page, click **Add Data Source** > **Add SharePoint Online Account**.



2. In the Add a SharePoint Online Account dialog, click **Sign in to SharePoint**.
3. In the Microsoft page that appears, select the SharePoint account and enter the user and password (Admin user or other user with access to the SharePoint sites), then click **Accept** to allow BlueXP classification to read data from this account.

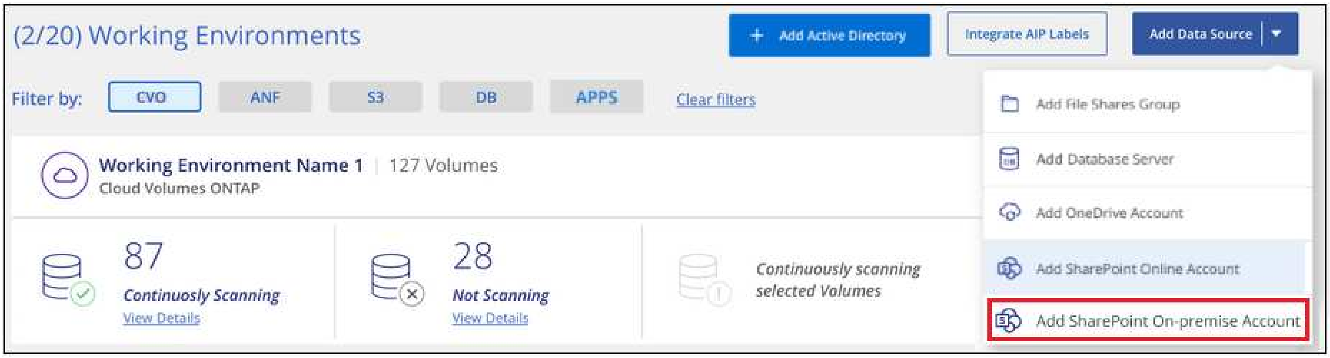
The SharePoint Online account is added to the list of working environments.

Adding a SharePoint On-premise account

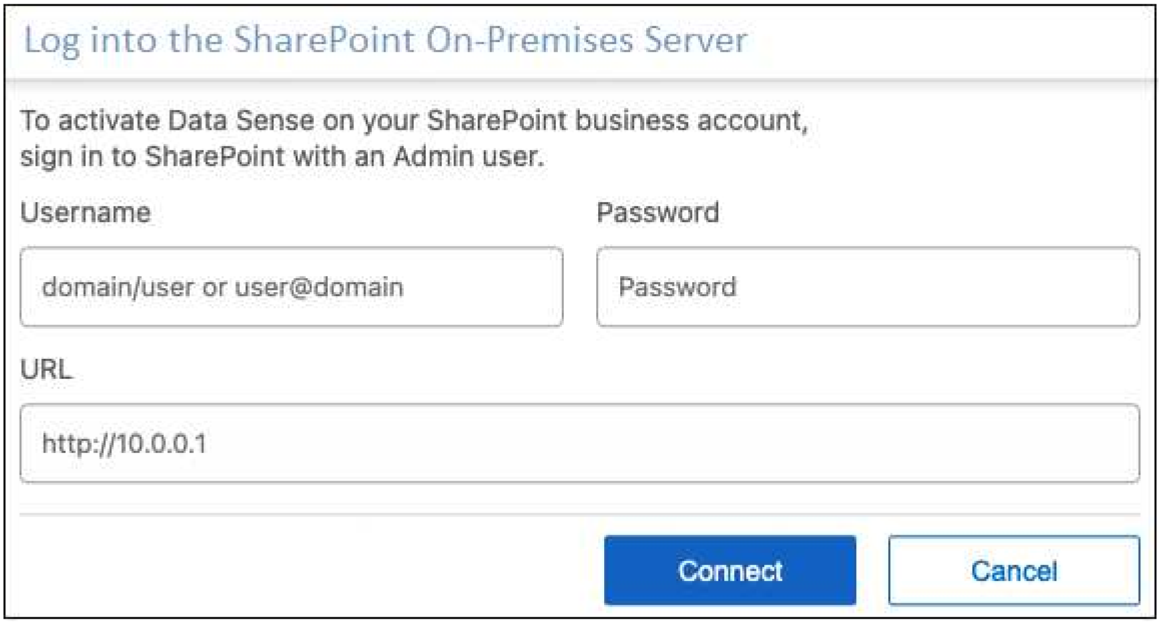
Add the SharePoint On-premise account where the user files reside.

Steps

1. From the Working Environments Configuration page, click **Add Data Source** > **Add SharePoint On-premise Account**.



2. In the Log into the SharePoint On-Premise Server dialog, enter the following information:
   - Admin user in the format "domain/user" or "user@domain", and admin password
   - URL of the SharePoint Server



Log into the SharePoint On-Premises Server

To activate Data Sense on your SharePoint business account, sign in to SharePoint with an Admin user.

Username: domain/user or user@domain

Password: Password

URL: http://10.0.0.1

3. Click **Connect**.

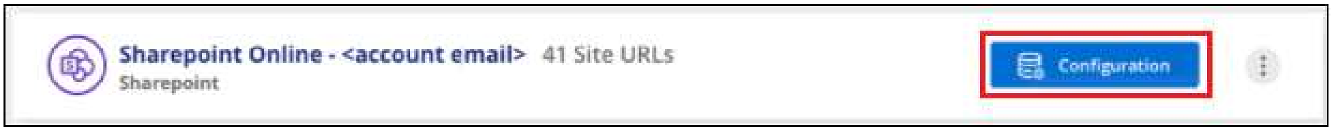
The SharePoint On-premise account is added to the list of working environments.

Adding SharePoint sites to compliance scans

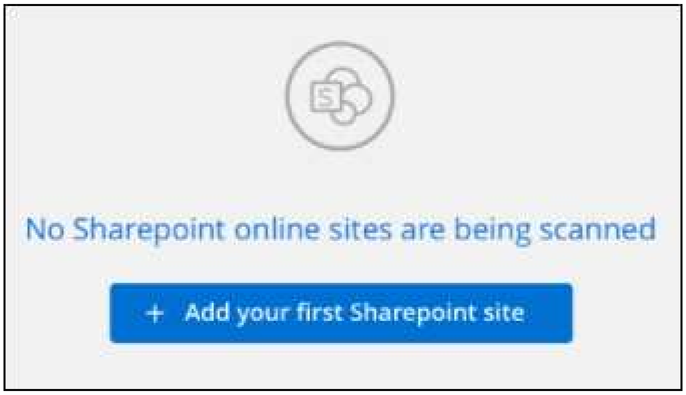
You can add individual SharePoint sites, or up to 1,000 SharePoint sites in the account, so that the associated files will be scanned by BlueXP classification. The steps are the same whether you are adding SharePoint Online or SharePoint On-premise sites.

Steps

1. From the *Configuration* page, click the **Configuration** button for the SharePoint account.

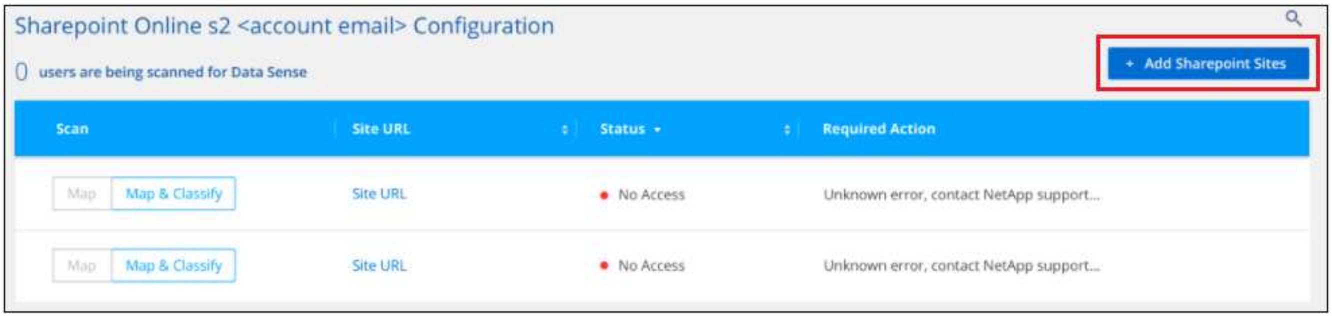


2. If this is the first time adding sites for this SharePoint account, click **Add your first SharePoint site**.



No Sharepoint online sites are being scanned

If you are adding additional users from a SharePoint account, click **Add SharePoint Sites**.

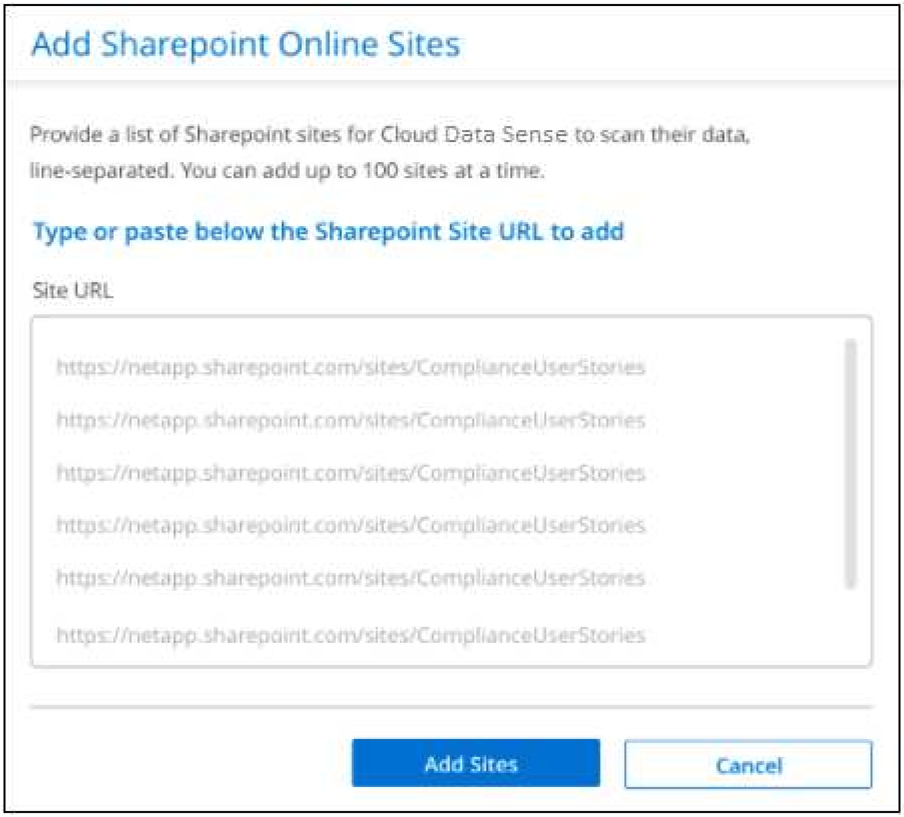


Sharepoint Online s2 <account email> Configuration

0 users are being scanned for Data Sense

Scan	Site URL	Status	Required Action
Map / Map & Classify	Site URL	● No Access	Unknown error, contact NetApp support...
Map / Map & Classify	Site URL	● No Access	Unknown error, contact NetApp support...

3. Add the URLs for the sites whose files you want to scan - one URL per line (up to 100 maximum per session) - and click **Add Sites**.



Add Sharepoint Online Sites

Provide a list of Sharepoint sites for Cloud Data Sense to scan their data, line-separated. You can add up to 100 sites at a time.

Type or paste below the Sharepoint Site URL to add

Site URL

```
https://netapp.sharepoint.com/sites/ComplianceUserStories
https://netapp.sharepoint.com/sites/ComplianceUserStories
https://netapp.sharepoint.com/sites/ComplianceUserStories
https://netapp.sharepoint.com/sites/ComplianceUserStories
https://netapp.sharepoint.com/sites/ComplianceUserStories
https://netapp.sharepoint.com/sites/ComplianceUserStories
```

A confirmation dialog displays the number of sites that were added.

If the dialog lists any sites that could not be added, capture this information so that you can resolve the issue. In some cases you can re-add the site with a corrected URL.

4. If you need to add more than 100 sites for this account, just click **Add SharePoint Sites** again until you have added all your sites for this account (up to 1,000 sites total for each account).
5. Enable mapping-only scans, or mapping and classification scans, on the files in the SharePoint sites.

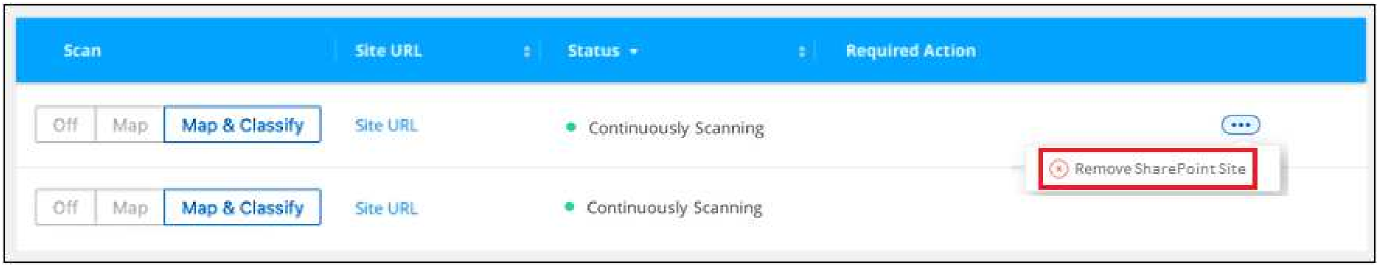
To:	Do this:
Enable mapping-only scans on files	Click **Map**
Enable full scans on files	Click **Map & Classify**
Disable scanning on files	Click **Off**

Result

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

Removing a SharePoint site from compliance scans

If you remove a SharePoint site in the future, or decide not to scan files in a SharePoint site, you can remove individual SharePoint sites from having their files scanned at any time. Just click **Remove SharePoint Site** from the Configuration page.



Scan	Site URL	Status	Required Action
Off Map Map & Classify	Site URL	● Continuously Scanning	•••
Off Map Map & Classify	Site URL	● Continuously Scanning	

Note that you can delete the entire SharePoint account from BlueXP classification if you no longer want to scan any user data from the SharePoint account.

Scanning Google Drive accounts

Complete a few steps to start scanning user files in your Google Drive accounts with BlueXP classification.

Quick start

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

1

Review Google Drive prerequisites

Ensure that you have the Admin credentials to log into the Google Drive account.

2

Deploy BlueXP classification

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

3 Log into the Google Drive account

Using Admin user credentials, log into the Google Drive account that you want to access so that it is added as a new data source.

4 Select the type of scanning for the user files

Select the type of scanning you want to perform on the user files; mapping or mapping and classifying.

Reviewing Google Drive requirements

Review the following prerequisites to make sure you are ready to enable BlueXP classification on a Google Drive account.

- You must have the Admin login credentials for the Google Drive account that provides read access to the user’s files

Current restrictions

The following BlueXP classification features are not currently supported with Google Drive files:

- When viewing files in the Data Investigation page, the actions in the button bar aren’t active. You can’t copy, move, delete, etc. any files.
- Permissions can’t be identified within files in Google Drive, so no permission information is displayed in the Investigation page.

Deploying BlueXP classification

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

BlueXP classification can be deployed in the cloud or in an on-premises location that has internet access.

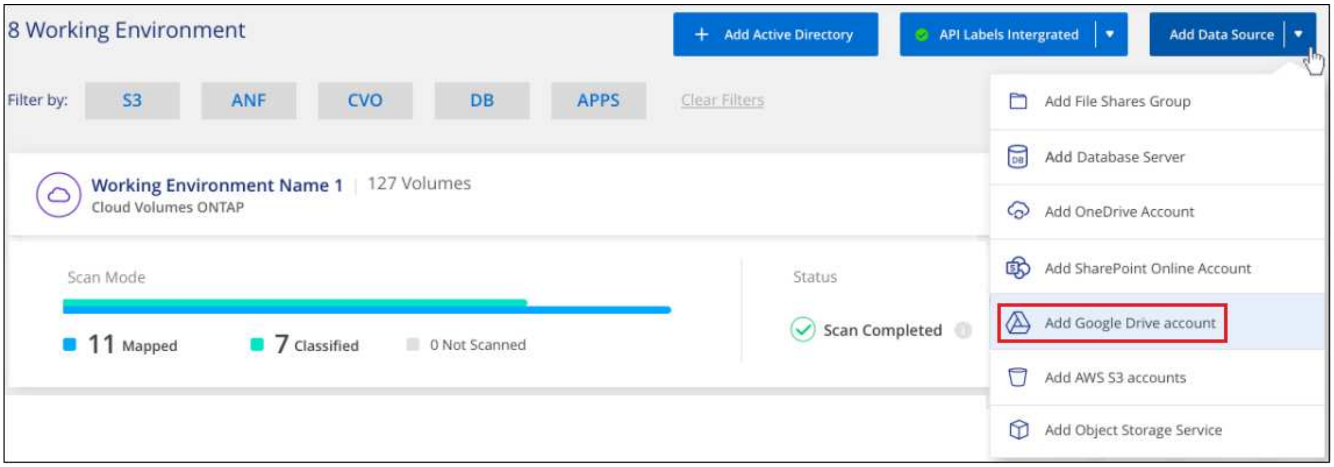
Upgrades to BlueXP classification software is automated as long as the instance has internet connectivity.

Adding the Google Drive account

Add the Google Drive account where the user files reside. If you want to scan files from multiple users, you’ll need to run through this step for each user.

Steps

1. From the Working Environments Configuration page, click **Add Data Source** > **Add Google Drive Account**.



2. In the Add a Google Drive Account dialog, click **Sign in to Google Drive**.
3. In the Google page that appears, select the Google Drive account and enter the required Admin user and password, then click **Accept** to allow BlueXP classification to read data from this account.

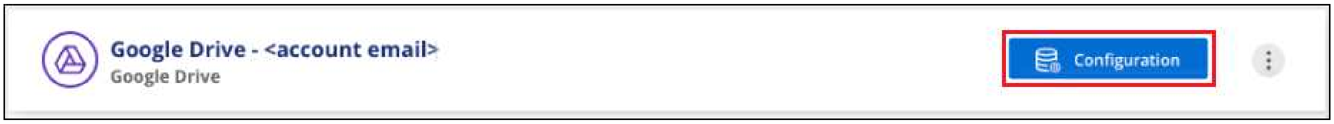
The Google Drive account is added to the list of working environments.

Selecting the type of scanning for user data

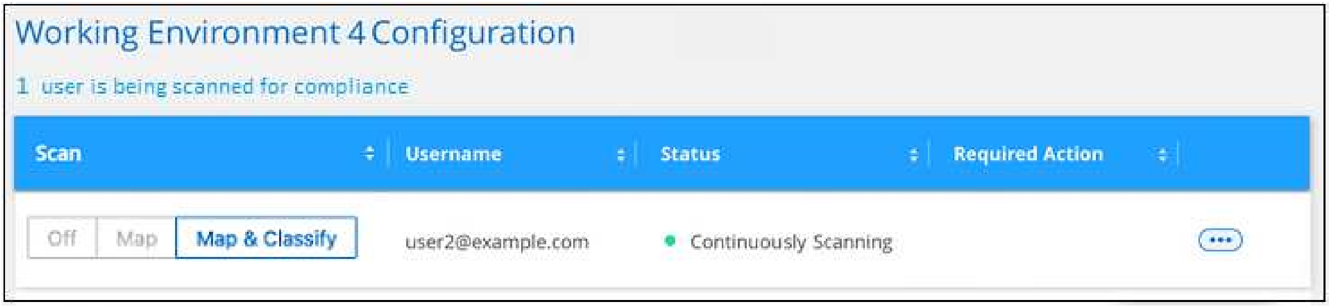
Select the type of scanning that BlueXP classification will perform on the user's data.

Steps

1. From the *Configuration* page, click the **Configuration** button for the Google Drive account.



2. Enable mapping-only scans, or mapping and classification scans, on the files in the Google Drive account.


Working Environment 4 Configuration

1 user is being scanned for compliance

Scan	Username	Status	Required Action
Off Map Map & Classify	user2@example.com	● Continuously Scanning	•••

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

Result

BlueXP classification starts scanning the files in the Google Drive account you added, and the results are displayed in the Dashboard and in other locations.

Removing a Google Drive account from compliance scans

Since only a single user’s Google Drive files are part of a single Google Drive account, if you want to stop scanning files from a user’s Google Drive account, then you should delete the Google Drive account from BlueXP classification.

Scanning file shares

Complete a few steps to start scanning non-NetApp NFS or CIFS file shares directly with BlueXP classification. These file shares can reside on-premises or in the cloud.

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. In most cases these are file shares that reside on non-NetApp storage systems. However, 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.

If you are scanning non-NetApp NFS or CIFS file shares 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 non-NetApp NFS or CIFS file shares 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.

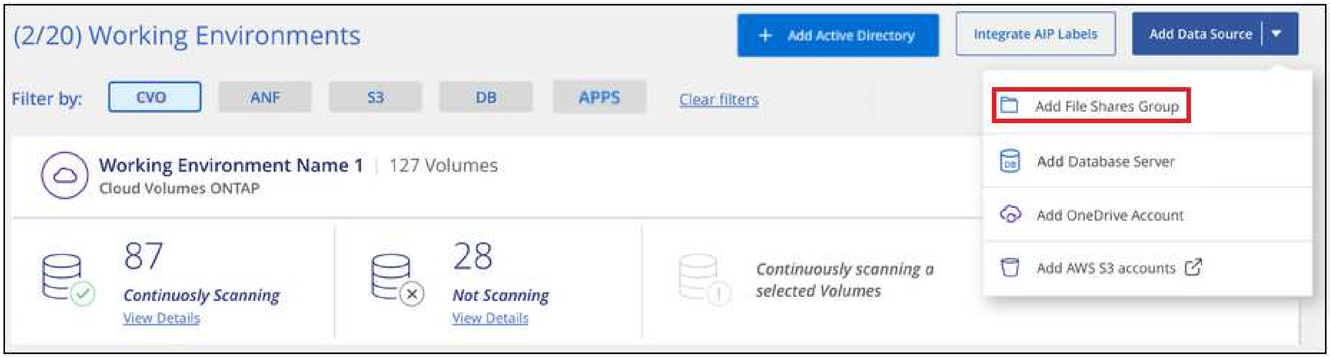
Creating the group for the file shares

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

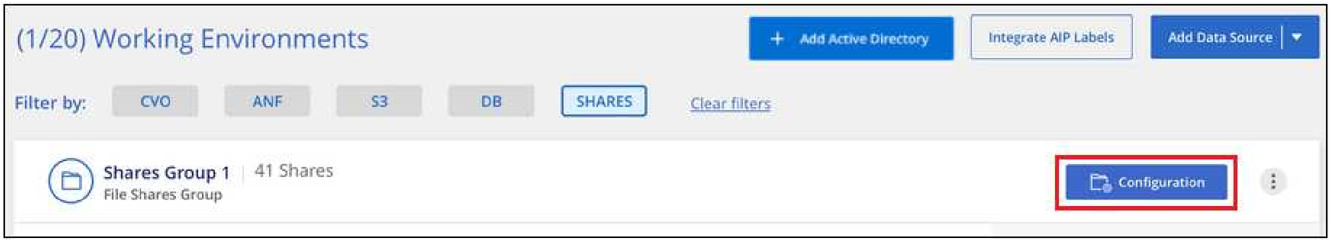
You add file shares to the File Shares Group so that the files in those shares will be scanned by BlueXP classification. You add the shares in the format `<host_name>:/<share_path>`.

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.

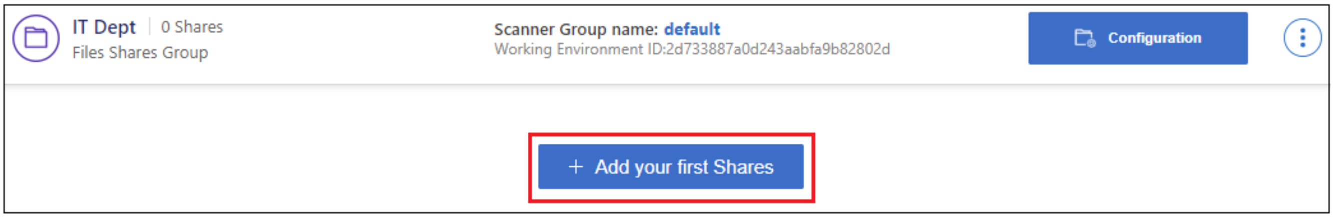
When adding both NFS and CIFS shares in a single group, you’ll need to run through the process twice - once adding NFS shares, and then again adding the CIFS shares.

Steps

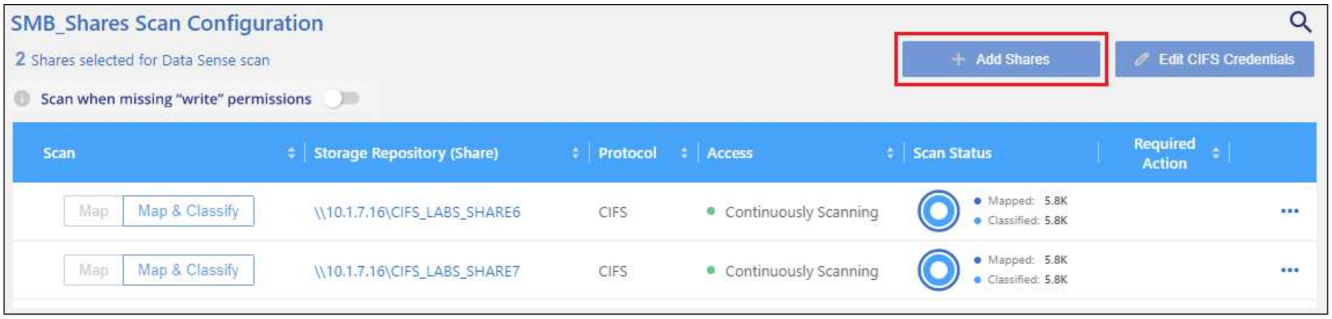
1. From the *Working Environments* page, click the **Configuration** button for the File Shares Group.



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


IT Dept
0 Shares
Files Shares Group
Scanner Group name: default
Working Environment ID:2d733887a0d243aabfa9b82802d
Configuration
+ Add your first Shares

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



SMB_Shares Scan Configuration

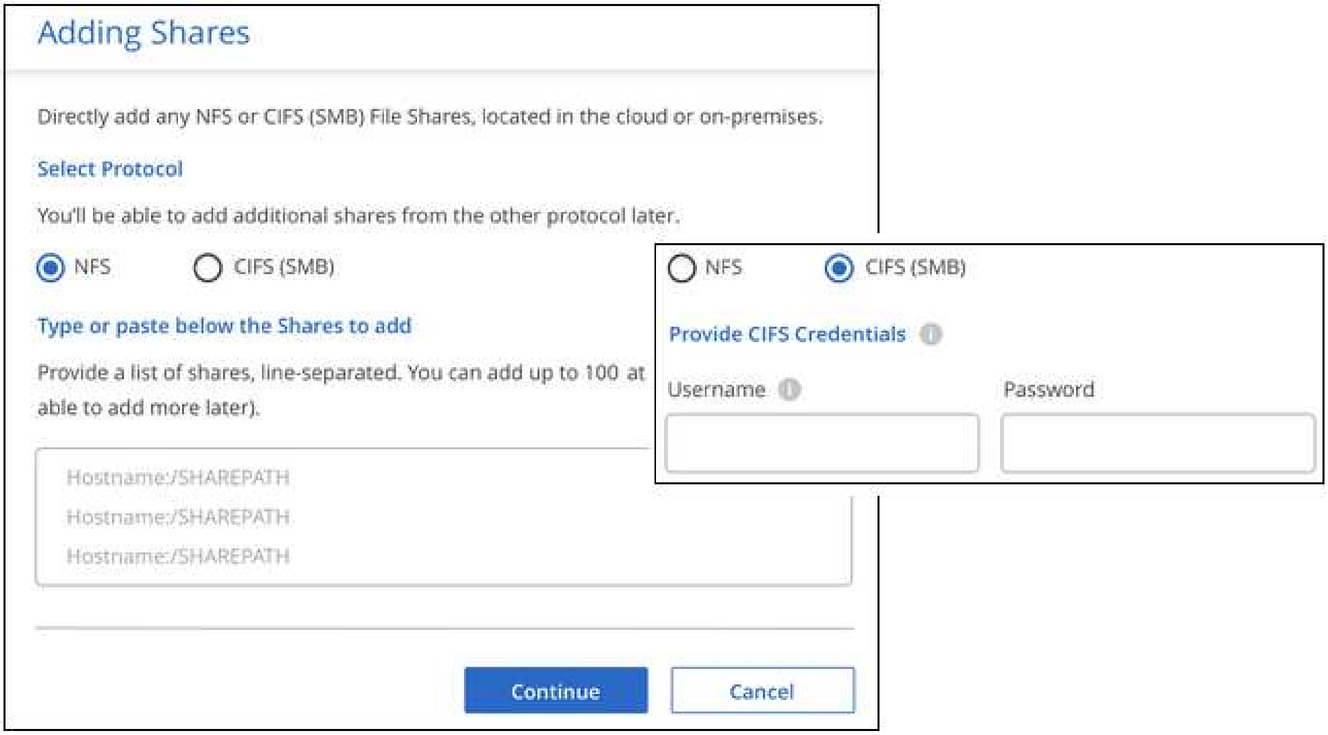
2 Shares selected for Data Sense scan

Scan when missing "write" permissions

Scan	Storage Repository (Share)	Protocol	Access	Scan Status	Required Action
Map / Map & Classify	\\10.1.7.16\CIFS_LABS_SHARE6	CIFS	● Continuously Scanning	● Mapped: 5.8K ● Classified: 5.8K	•••
Map / Map & Classify	\\10.1.7.16\CIFS_LABS_SHARE7	CIFS	● Continuously Scanning	● Mapped: 5.8K ● Classified: 5.8K	•••

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

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



Adding Shares

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

Select Protocol

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

● NFS ○ CIFS (SMB)

Type or paste below the Shares to add

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

Hostname:/SHAREPATH
Hostname:/SHAREPATH
Hostname:/SHAREPATH

○ NFS ● CIFS (SMB)

Provide CIFS Credentials

Username

Password

   A confirmation dialog displays the number of shares that were added.

   If the dialog lists any shares that could not be added, capture this information so that you can resolve the issue. In some cases you can re-add the share with a corrected host name or share name.

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

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

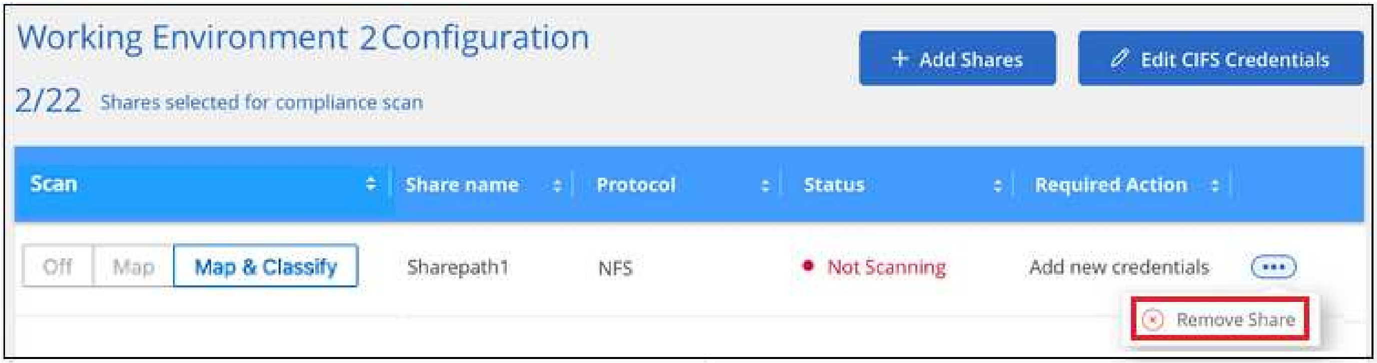
   The switch at the top of the page for **Scan when missing "write attributes" permissions** is disabled by default. This means that if BlueXP classification doesn’t have write attributes permissions in CIFS, or write permissions in NFS, that the system won’t scan the files because BlueXP classification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

Result

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

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.



Working Environment 2Configuration

2/22 Shares selected for compliance scan

Scan	Share name	Protocol	Status	Required Action
Off Map **Map & Classify**	Sharepath1	NFS	● Not Scanning	Add new credentials

Scanning object storage that uses S3 protocol

Complete a few steps to start scanning data within object storage directly with BlueXP classification. BlueXP classification can scan data from any Object Storage service which uses the Simple Storage Service (S3) protocol. This includes NetApp StorageGRID, IBM Cloud Object Store, Linode, B2 Cloud Storage, Amazon S3, and more.

Quick start

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

1

Review object storage prerequisites

You need to have the endpoint URL to connect with the object storage service.

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

2

Deploy the BlueXP classification instance

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

3

Add the Object Storage Service

Add the object storage service to BlueXP classification.

4

Select the buckets to scan

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

Reviewing object storage requirements

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

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

Deploying the BlueXP classification instance

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

If you are scanning data from S3 object storage that is accessible over the internet, you can deploy BlueXP classification in the cloud or deploy BlueXP classification in an on-premises location that has internet access.

If you are scanning data from S3 object storage that has been installed in a dark site that has no internet access, you need to deploy BlueXP classification in the same on-premises location that has no internet access. This also requires that the BlueXP Connector is deployed in that same on-premises location.

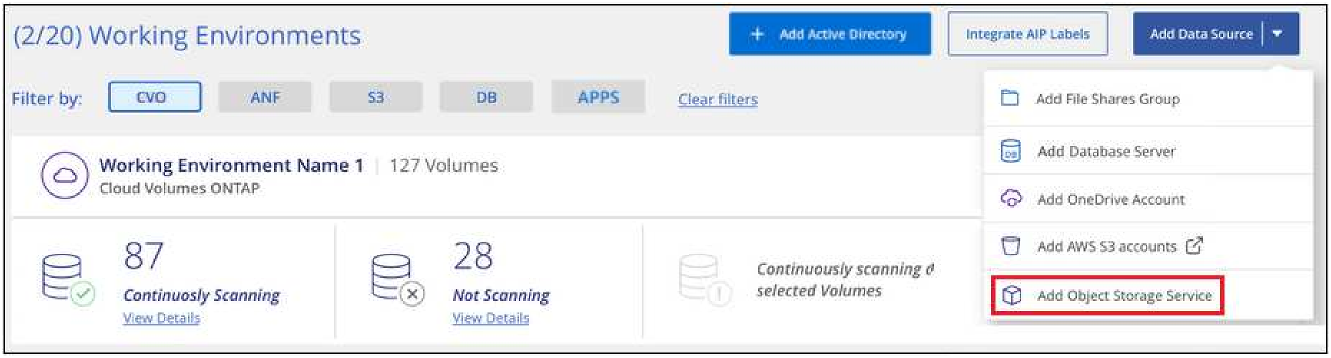
Upgrades to BlueXP classification software is automated as long as the instance has internet connectivity.

Adding the object storage service to BlueXP classification

Add the object storage service.

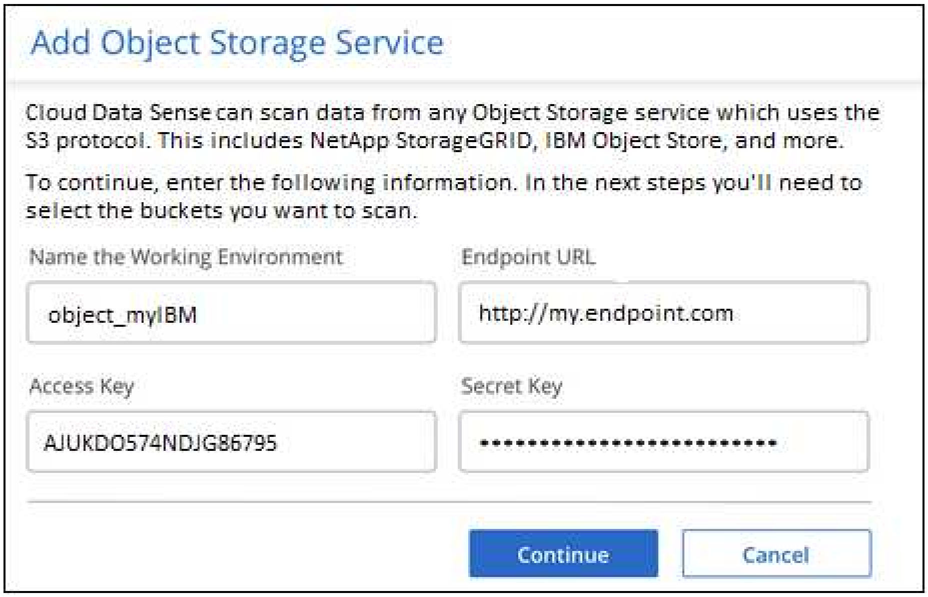
Steps

1. From the Working Environments Configuration page, click **Add Data Source** > **Add Object Storage Service**.



2. In the Add Object Storage Service dialog, enter the details for the object storage service and click **Continue**.
   a. Enter the name you want to use for the Working Environment. This name should reflect the name of the object storage service to which you are connecting.
   b. Enter the Endpoint URL to access the object storage service.
   c. Enter the Access Key and Secret Key so that BlueXP classification can access the buckets in the

object storage.



Add Object Storage Service

Cloud Data Sense can scan data from any Object Storage service which uses the S3 protocol. This includes NetApp StorageGRID, IBM Object Store, and more.

To continue, enter the following information. In the next steps you'll need to select the buckets you want to scan.

Name the Working Environment: object_myIBM

Endpoint URL: http://my.endpoint.com

Access Key: AJUKDO574NDJG86795

Secret Key: ••••••••••••••••••••••••••

Result

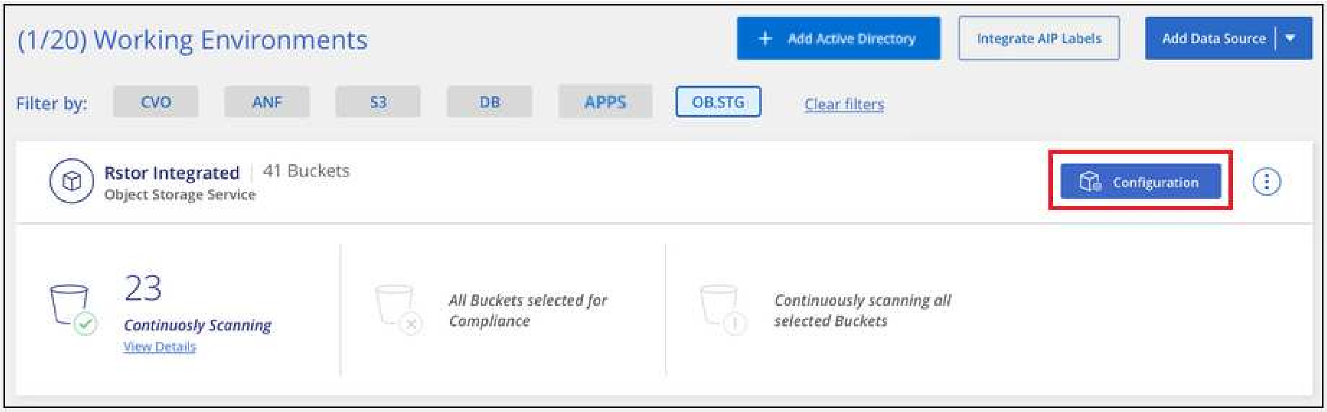
The new Object Storage Service is added to the list of working environments.

Enabling and disabling compliance scans on object storage buckets

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

Steps

1. In the Configuration page, click **Configuration** from the Object Storage Service working environment.



(1/20) Working Environments

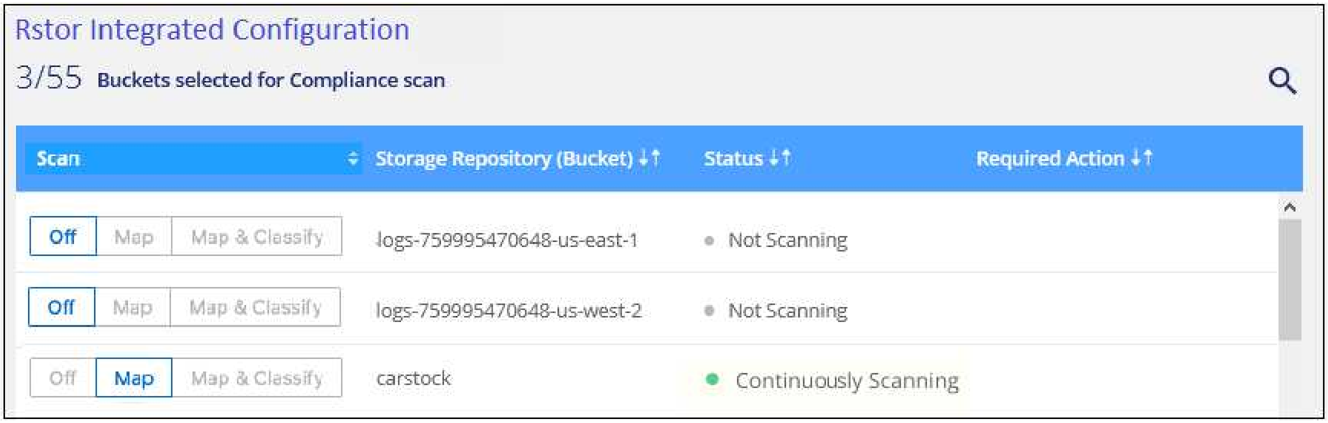
Rstor Integrated | 41 Buckets
Object Storage Service

23
Continuosly Scanning

All Buckets selected for Compliance

Continuously scanning all selected Buckets

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


Rstor Integrated Configuration

3/55 Buckets selected for Compliance scan

Scan	Storage Repository (Bucket)	Status	Required Action
Off / Map / Map & Classify	logs-759995470648-us-east-1	● Not Scanning	
Off / Map / Map & Classify	logs-759995470648-us-west-2	● Not Scanning	
Off / Map / Map & Classify	carstock	● Continuously Scanning	

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

Result

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

Integrate your Active Directory with BlueXP classification

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

When you set up certain data sources (listed below), you need to enter Active Directory credentials in order for BlueXP classification to scan CIFS volumes. This integration provides BlueXP classification with file owner and permissions details for the data that resides in those data sources. The Active Directory entered for those data sources may be different than 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
- Non-NetApp CIFS file shares (not NFS file shares)
- OneDrive accounts
- SharePoint accounts

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

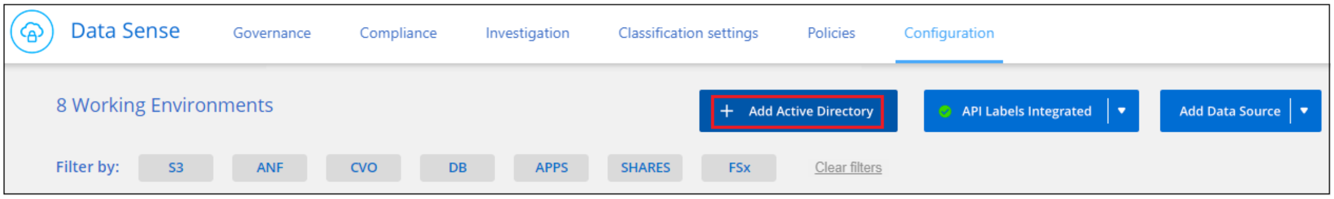
    LDAP Server IP address, or multiple IP addresses
  - User Name and Password to access the server
  - Domain Name (Active Directory Name)
  - Whether you are using secure LDAP (LDAPS) or not
  - LDAP Server Port (typically 389 for LDAP, and 636 for secure LDAP)
- The following ports must be open for outbound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

Protocol	Port	Destination	Purpose
TCP	3269	Active Directory	Global Catalog over SSL

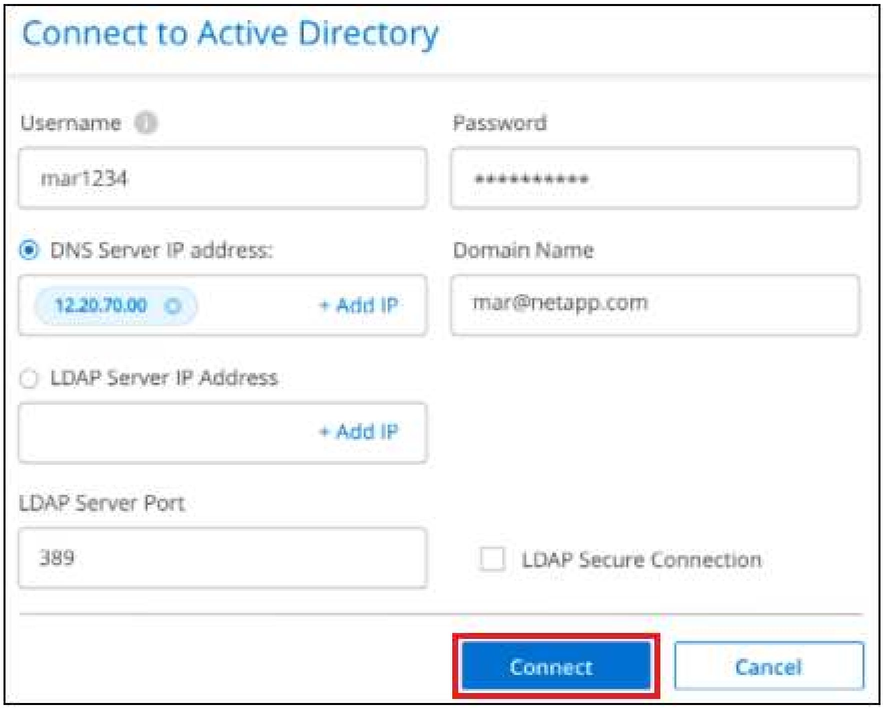
Steps

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**.


Connect to Active Directory

Username: mar1234

Password: **********

● DNS Server IP address: 12.20.70.00

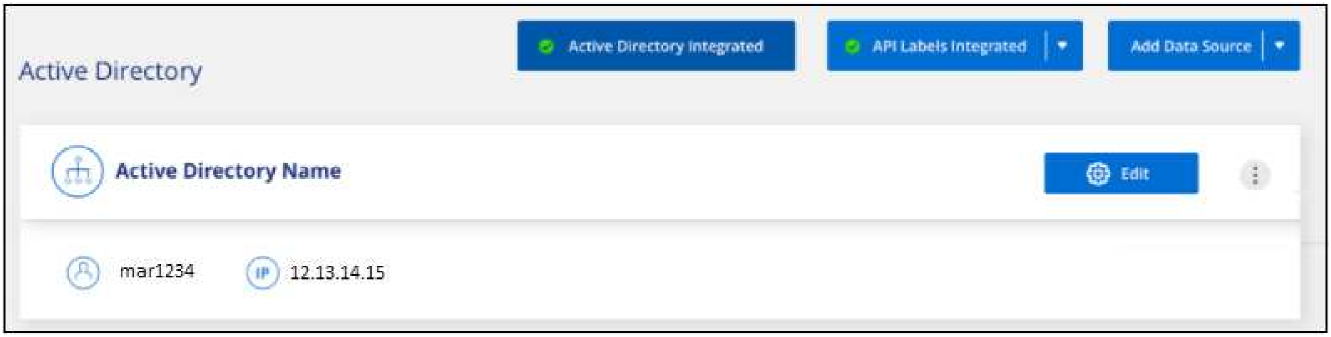
Domain Name: mar@netapp.com

○ LDAP Server IP Address

LDAP Server Port: 389

□ LDAP Secure Connection

   BlueXP classification integrates to the Active Directory, and a new section is added to the Configuration page.


Active Directory

Active Directory Name

mar1234 12.13.14.15

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**.

Set up licensing for BlueXP classification

The first 1 TB of data that BlueXP classification scans in a BlueXP workspace is free for 30 days. A BYOL license from NetApp, or a subscription from your cloud provider's marketplace, is required to continue scanning data after that point.

A few notes before you read any further:

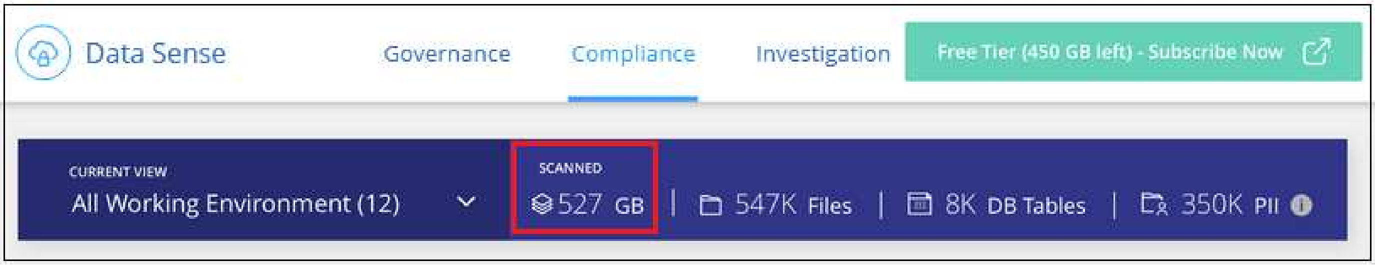
- If you've already subscribed to the BlueXP pay-as-you-go (PAYGO) subscription in your cloud provider's marketplace, then you're automatically subscribed to BlueXP classification as well. You won't need to subscribe again.
- The BlueXP classification (Data Sense) bring-your-own-license (BYOL) is a *floating* license that you can use across all the working environments and data sources in the workspace that you plan to scan. You'll see an active subscription in the BlueXP digital wallet.
- The amount of data being scanned is calculated based on logical file size; without any storage efficiencies.

Learn more about the licensing and costs related to BlueXP classification.

30-day free trial

A 30-day free trial is available for up to 1 TB of data that BlueXP classification scans in a BlueXP workspace. You'll need to purchase a BYOL license from NetApp, or a sign up for a subscription from your cloud provider's marketplace, to continue scanning data after that point.

You can subscribe at any time and you will not be charged until either the 30 day trial ends, or the amount of data exceeds 1 TB. You can always see the total amount of data that is being scanned from the BlueXP classification Governance Dashboard. And the *Subscribe Now* button makes it easy to subscribe when you are ready.



Use a BlueXP classification PAYGO subscription

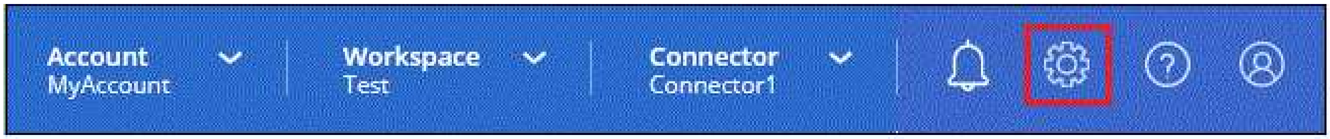
Pay-as-you-go subscriptions from your cloud provider's marketplace enable you to license the use of Cloud Volumes ONTAP systems and many BlueXP services, such as BlueXP classification. You'll pay your cloud provider for the amount of data BlueXP classification is scanning on an hourly basis in a single subscription.

Subscribing ensures that there’s no disruption of service after your free trial ends. When the trial ends, you’ll be charged hourly according to the amount of data that you are scanning. You won’t be charged from your subscription during your free trial.

Steps

These steps must be completed by a user who has the *Account Admin* role.

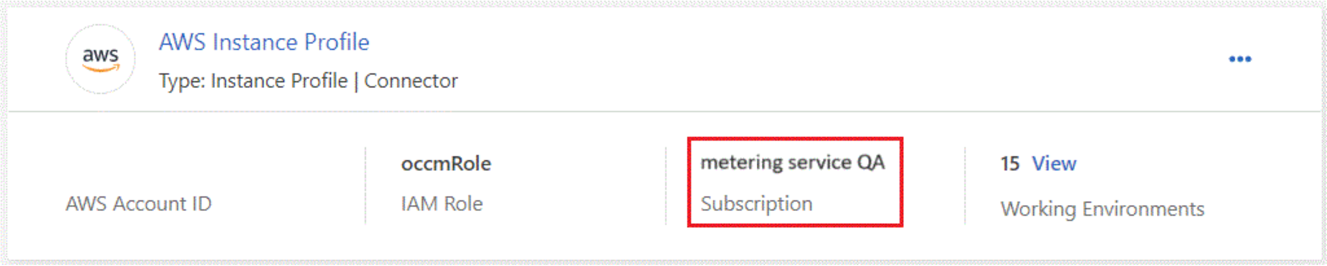
1. In the upper right of the BlueXP console, click the Settings icon, and select **Credentials**.



2. Click **Credentials** and then find the credentials for the AWS Instance Profile, Azure Managed Service Identity, or Google Project.

   The subscription must be added to the Instance Profile, Managed Service Identity, or Google Project. Charging won’t work otherwise.

   If you already have a BlueXP subscription (as shown below for AWS), then you’re all set—there’s nothing else that you need to do.



AWS Instance Profile
Type: Instance Profile | Connector

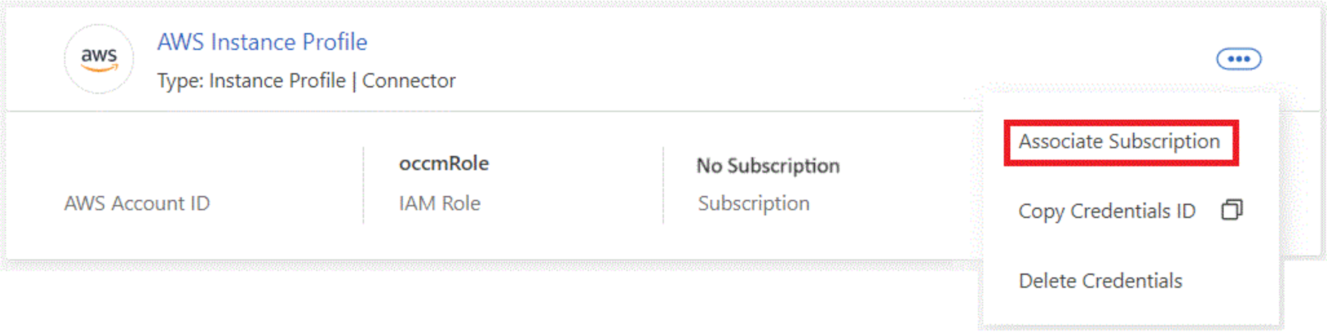
AWS Account ID

occmRole
IAM Role

metering service QA
Subscription

15 View
Working Environments

3. If you don’t have a subscription yet, click the action menu and click **Associate Subscription**.



AWS Instance Profile
Type: Instance Profile | Connector

AWS Account ID

occmRole
IAM Role

No Subscription
Subscription

4. Select an existing subscription and click **Associate**, or click **Add Subscription** and follow the steps.

   The following video shows how to associate an AWS Marketplace subscription to an AWS subscription:

   ► https://docs.netapp.com/us-en/bluexp-classification//media/video_subscribing_aws.mp4 *(video)*

   The following video shows how to associate an Azure Marketplace subscription to an Azure subscription:

   ► https://docs.netapp.com/us-en/bluexp-classification//media/video_subscribing_azure.mp4 *(video)*

The following video shows how to associate a Google Cloud Marketplace subscription to a GCP subscription:

► https://docs.netapp.com/us-en/bluexp-classification//media/video_subscribing_gcp.mp4 *(video)*

Use an annual contract

Pay for BlueXP classification annually by purchasing an annual contract. They're available in 1-, 2-, or 3-year terms.

If you have an annual contract from a marketplace, all BlueXP classification data scanning is charged against that contract. You can't mix and match an annual marketplace contract with a BYOL.

- AWS: Go to the BlueXP Marketplace offering for pricing details.
- Azure: Go to the BlueXP Marketplace offering for pricing details.
- Google Cloud: Contact your NetApp sales representative to purchase an annual contract. The contract is available as a private offer in the Google Cloud Marketplace. After NetApp shares the private offer with you, you can select the annual plan when you subscribe from the Google Cloud Marketplace during BlueXP classification activation.

Use a BlueXP classification BYOL license

Bring-your-own licenses from NetApp provide 1-, 2-, or 3-year terms. The BYOL BlueXP classification (Data Sense) license is a *floating* license where the total capacity is shared among **all** of your working environments and data sources, making initial licensing and renewal easy.

If you don't have a BlueXP classification license, contact us to purchase one:

- Send email to purchase a license.
- Click the chat icon in the lower-right of BlueXP to request a license.

Optionally, if you have an unassigned node-based license for Cloud Volumes ONTAP that you won't be using, you can convert it to a BlueXP classification license with the same dollar-equivalence and the same expiration date. Go here for details.

You use the BlueXP digital wallet to manage BlueXP classification BYOL licenses. You can add new licenses, update existing licenses, and view license status from the BlueXP digital wallet.

Obtain your BlueXP classification license file

After you've purchased your BlueXP classification (Data Sense) license, you activate the license in BlueXP by entering the BlueXP classification serial number and NetApp Support Site (NSS) account, or by uploading the NetApp License File (NLF). The steps below show how to get the NLF license file if you plan to use that method.

If you've deployed BlueXP classification on a host in an on-premises site that doesn't have internet access, meaning that you've deployed the BlueXP Connector in private mode, you'll need to obtain the license file from an internet-connected system. Activating the license using the serial number and NSS account is not available for private mode installations.

Before you begin

You'll need to have the following information before you start:

- BlueXP classification serial number

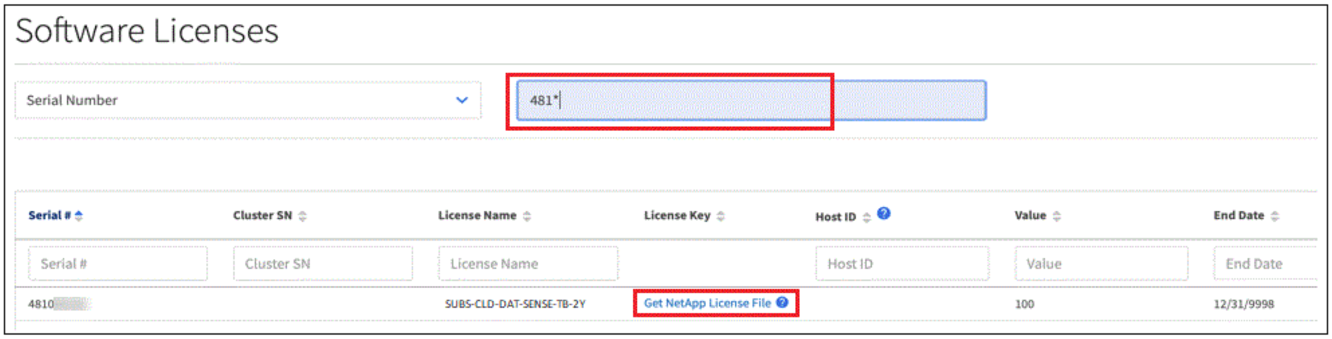
  Locate this number from your Sales Order, or contact the account team for this information.

- BlueXP Account ID

  You can find your BlueXP Account ID by selecting the **Account** drop-down from the top of BlueXP, and then clicking **Manage Account** next to your account. Your Account ID is in the Overview tab. For private mode sites without internet access, use **account-DARKSITE1**.

Steps

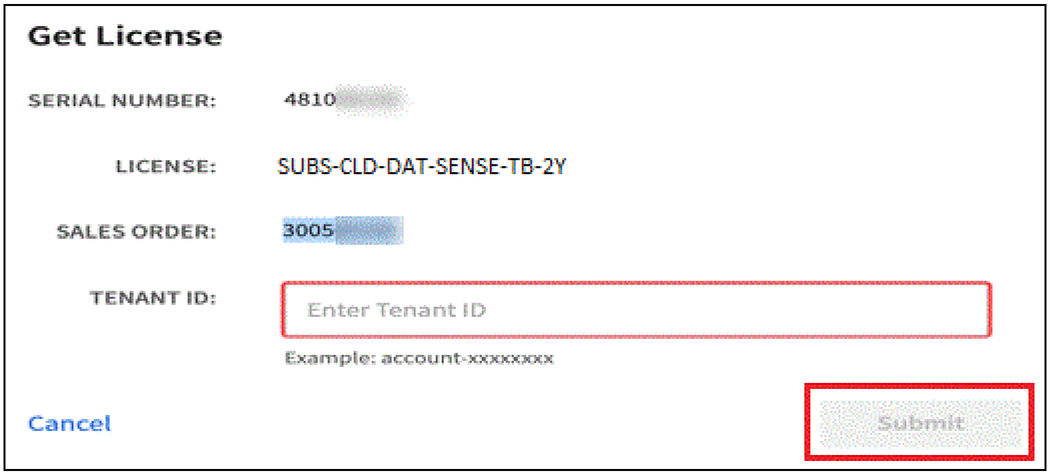
1. Sign in to the NetApp Support Site and click **Systems > Software Licenses**.
2. Enter your BlueXP classification license serial number.



Software Licenses

Serial #	Cluster SN	License Name	License Key	Host ID	Value	End Date
Serial #	Cluster SN	License Name		Host ID	Value	End Date
4810[redacted]		SUBS-CLD-DAT-SENSE-TB-2Y	Get NetApp License File		100	12/31/9998

3. Under the **License Key** column, click **Get NetApp License File**.
4. Enter your BlueXP Account ID (this is called a Tenant ID on the support site) and click **Submit** to download the license file.



Get License

SERIAL NUMBER:	4810[redacted]
LICENSE:	SUBS-CLD-DAT-SENSE-TB-2Y
SALES ORDER:	3005[redacted]
TENANT ID:	Enter Tenant ID Example: account-xxxxxxxx

Add BlueXP classification BYOL licenses to your account

After you purchase a BlueXP classification (Data Sense) license for your BlueXP account, you need to add the license to BlueXP to use the BlueXP classification service.

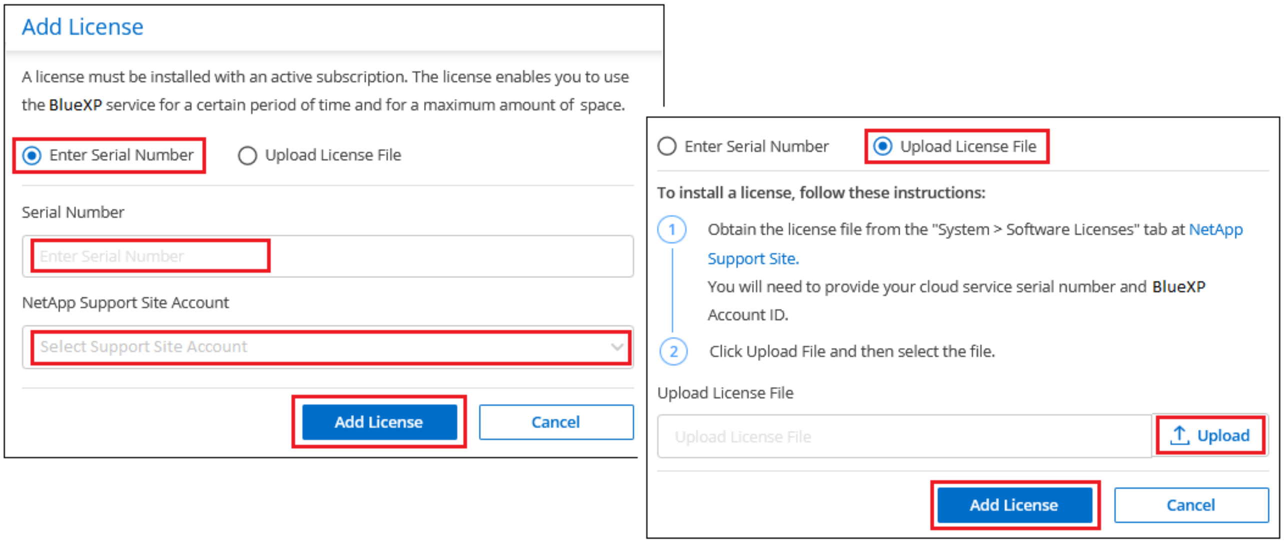
Steps

1. From the BlueXP menu, click **Governance > Digital wallet** and then select the **Data Services Licenses** tab.

2. Click **Add License**.
3. In the *Add License* dialog, enter the license information and click **Add License**:
   - If you have the BlueXP classification license serial number and know your NSS account, select the **Enter Serial Number** option and enter that information.

     If your NetApp Support Site account isn't available from the drop-down list, add the NSS account to BlueXP.

   - If you have the BlueXP classification license file (required when installed in a dark site), select the **Upload License File** option and follow the prompts to attach the file.



Add License

A license must be installed with an active subscription. The license enables you to use the **BlueXP** service for a certain period of time and for a maximum amount of space.

● Enter Serial Number ○ Upload License File

Serial Number

Enter Serial Number

NetApp Support Site Account

Select Support Site Account

Add License Cancel

○ Enter Serial Number ● Upload License File

To install a license, follow these instructions:

1. Obtain the license file from the "System > Software Licenses" tab at NetApp Support Site.
   You will need to provide your cloud service serial number and **BlueXP** Account ID.
2. Click Upload File and then select the file.

Upload License File

Upload License File

Upload

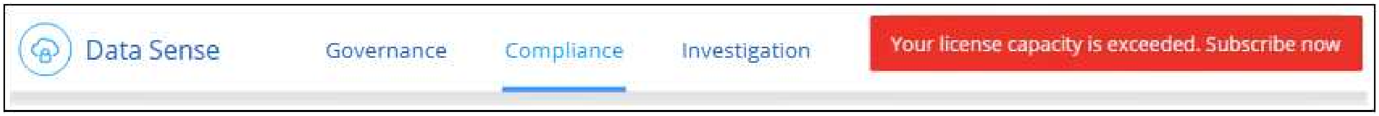
Add License Cancel

Result

BlueXP adds the license so that your BlueXP classification service is active.

Update a BlueXP classification BYOL license

If your licensed term is nearing the expiration date, or if your licensed capacity is reaching the limit, you'll be notified in the Classification UI.



Data Sense

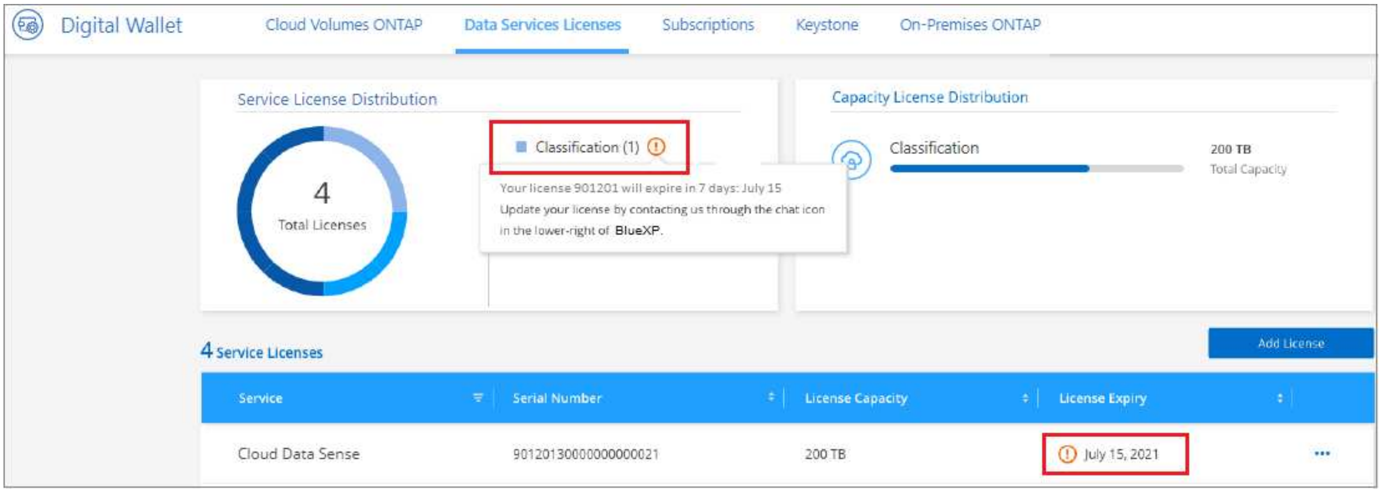
Governance

Compliance

Investigation

Your license capacity is exceeded. Subscribe now

This status also appears in the BlueXP digital wallet and in Notifications.



Your license 901201 will expire in 7 days: July 15
Update your license by contacting us through the chat icon in the lower-right of BlueXP.

4 Service Licenses

Service	Serial Number	License Capacity	License Expiry	
Cloud Data Sense	90120130000000000021	200 TB	July 15, 2021	•••

You can update your BlueXP classification license before it expires so that there is no interruption in your ability to access your scanned data.

Steps

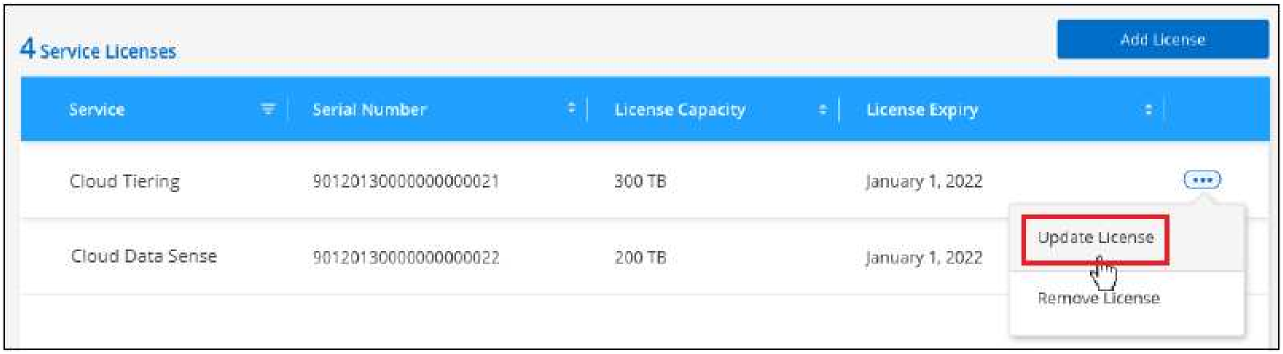
1. Click the chat icon in the lower-right of BlueXP to request an extension to your term or additional capacity to your Cloud Data Sense license for the particular serial number. You can also send an email to request an update to your license.

   After you pay for the license and it is registered with the NetApp Support Site, BlueXP automatically updates the license in the BlueXP digital wallet and the Data Services Licenses page will reflect the change in 5 to 10 minutes.

2. If BlueXP can’t automatically update the license (for example, when installed in a dark site), then you’ll need to manually upload the license file.

   a. You can obtain the license file from the NetApp Support Site.

   b. On the BlueXP digital wallet page in the *Data Services Licenses* tab, click ••• for the service serial number you are updating, and click **Update License**.

   4 Service Licenses

Service	Serial Number	License Capacity	License Expiry	
Cloud Tiering	90120130000000000021	300 TB	January 1, 2022	•••
Cloud Data Sense	90120130000000000022	200 TB	January 1, 2022	

   c. In the *Update License* page, upload the license file and click **Update License**.

Result

BlueXP updates the license so that your BlueXP classification service continues to be active.

BYOL license considerations

When using a BlueXP classification (Data Sense) BYOL license, BlueXP displays a warning in the BlueXP classification UI and in the BlueXP digital wallet UI when the size of all the data you are scanning is nearing the capacity limit or nearing the license expiration date. You receive these warnings:

- When the amount of data you are scanning has reached 80% of licensed capacity, and again when you have reached the limit
- 30 days before a license is due to expire, and again when the license expires

Use the chat icon in the lower right of the BlueXP interface to renew your license when you see these warnings.

If your license expires or you have reached the BYOL limit, BlueXP classification continues to run, but access to the Dashboards is blocked so that you can’t view information about any of your scanned data. Only the *Configuration* page is available in case you want to reduce the number of volumes being scanned to potentially bring your capacity usage under the license limit.

Once you renew your BYOL license, BlueXP automatically updates the license in the BlueXP digital wallet and provides full access to all Dashboards. If BlueXP can’t access the license file over the secure internet connection (for example, when installed in a dark site), you can obtain the file yourself and manually upload it to BlueXP. For instructions, see how to update a BlueXP classification license.

> ⓘ If the account you are using has both a BYOL license and a PAYGO subscription, BlueXP classification *will not* shift over to the PAYGO subscription when the BYOL license expires. You must renew the BYOL license.

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.

What are the common use cases for BlueXP classification?

- Identify Personal Identifiable Information (PII).
- Easily locate and report on specific data in response to data subjects, as required by GDPR, CCPA, HIPAA, and other data privacy regulations.
- Comply with new and upcoming data privacy regulations.
- Comply with data compliance and privacy regulations.
- Migrate data from legacy systems to the cloud.
- Comply with data retention policies.

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?

BlueXP supports REST API capabilities for its services. If BlueXP isn't the preferred point of management, services such as BlueXP classification can also be used via a REST API. Every user action has a REST API that can be integrated with third-party systems. See BlueXP classification APIs for details.

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.

What kind of analytics does BlueXP classification provide?

Data sources can be represented visually, and relationships defined and depicted graphically. For example, admins can see all stale, duplicate, and non-business-related data across data sources throughout the enterprise (on-premises systems, databases, file shares, S3 stores, OneDrive, etc.). They can then copy, move, delete, and manage data to optimize storage costs and reduce risk. Users can reduce risk by seeing what sensitive data might be exposed, and they can create jobs to manage permissions for strong data protection. BlueXP classification also classifies all the different types of data, so admins can investigate data by type and see what actions have been taken on the data, and when.

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, file share group, OneDrive account, or SharePoint account 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 out-of-the-box 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 snapshot copies 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.

Can BlueXP classification send notifications to my organization?

Yes. In conjunction with the Policies feature, you can send email alerts to BlueXP users (daily, weekly, or monthly), or any other email address, when a Policy returns results so you can get notifications to protect your data. Learn more about Policies.

You can also download status reports from the Governance page and Investigation page that you can share internally in your organization.

Can BlueXP classification work with the AIP labels I have embedded in my files?

Yes. You can manage AIP labels in the files that BlueXP classification is scanning if you have subscribed to Azure Information Protection (AIP). You can view the labels that are already assigned to files, add labels to files, and change existing labels.

Learn more.

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.

Working environments:

- Cloud Volumes ONTAP (deployed in AWS, Azure, or GCP)
- On-premises ONTAP clusters
- Azure NetApp Files
- Amazon FSx for ONTAP
- Amazon S3

Data sources:

- Non-NetApp file shares
- Object storage (that uses S3 protocol)
- Databases (Amazon RDS, MongoDB, MySQL, Oracle, PostgreSQL, SAP HANA, SQL Server)
- OneDrive accounts
- SharePoint Online and On-Premises accounts
- Google Drive accounts

BlueXP classification supports NFS versions 3.x, 4.0, and 4.1, and CIFS versions 1.x, 2.0, 2.1, and 3.0.

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:

- 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
- SharePoint On-Premises accounts (SharePoint Server)
- Non-NetApp NFS or CIFS file shares
- Object Storage that uses the Simple Storage Service (S3) protocol

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 data and sensitive personal data. 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 third 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 questions relate to licensing and costs to use BlueXP classification.

How much does BlueXP classification cost?

The cost to use BlueXP classification depends on the amount of data that you’re scanning. The first 1 TB of data that BlueXP classification scans in a BlueXP workspace is free for 30 days. After reaching either limit, you’ll need one of the following to continue scanning data:

- A subscription to the BlueXP Marketplace listing from your cloud provider, or
- A Bring-your-own-license (BYOL) from NetApp

See pricing for details.

What happens if I have reached the BYOL capacity limit?

If you reach a BYOL capacity limit, BlueXP classification continues to run, but access to the Dashboards is blocked so that you can’t view information about any of your scanned data. Only the Configuration page is available in case you want to reduce the number of volumes being scanned to potentially bring your capacity usage under the license limit. You must renew your BYOL license to regain full access to BlueXP classification.

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, Amazon FSx for ONTAP, or in AWS S3 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.
- When scanning data in Cloud Volumes ONTAP in GCP, you use a Connector in GCP.

- When scanning data in on-premises ONTAP systems, non-NetApp file shares, generic S3 Object storage, databases, OneDrive folders, SharePoint accounts, and Google Drive accounts, 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.

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.

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.

Copyright information

Trademark information