Deprecated features

BlueXP classification

NetApp
August 11, 2025

Table of Contents

Deprecated features

BlueXP classification deprecated features

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

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

Supported data sources

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

Compliance features

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

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

Features to manage your data

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

Deploy BlueXP classification deprecations

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

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

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

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

The following information is relevant only for BlueXP classification legacy versions 1.30 and earlier.

Before you begin

- 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.

Scan data deprecations

Scan Amazon S3 buckets with BlueXP classification

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

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

Quick start

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

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.

Deploy the BlueXP classification instance

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

3 Activate BlueXP classification on your S3 working environment

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

4 Select the buckets to scan

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

Reviewing S3 prerequisites

The following requirements are specific to scanning S3 buckets.

Set up an IAM role for the BlueXP classification instance

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

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

Provide connectivity from BlueXP classification to Amazon S3

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

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

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

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

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

Deploying the BlueXP classification instance

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

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

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

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

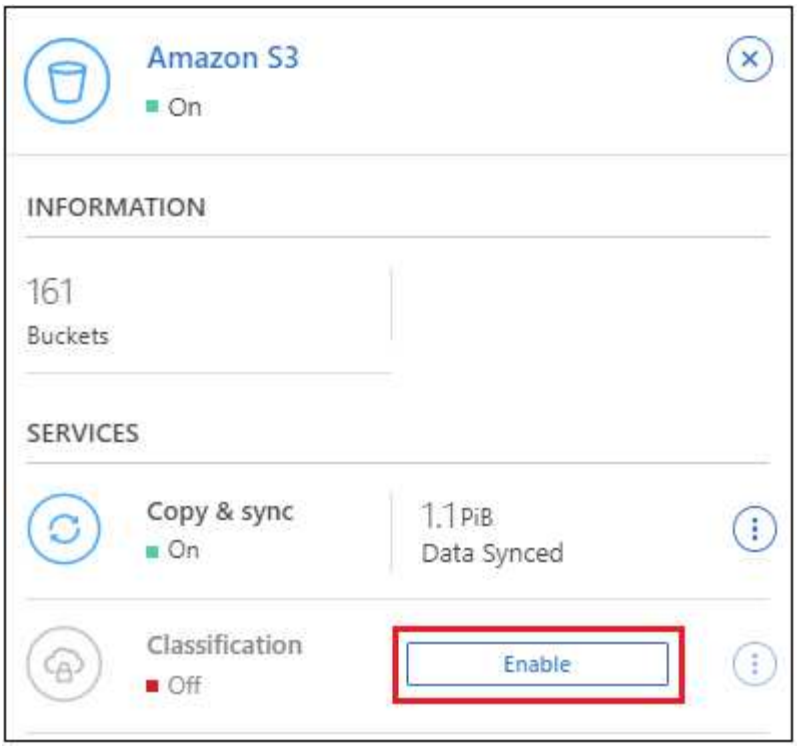
Activating BlueXP classification on your S3 working environment

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

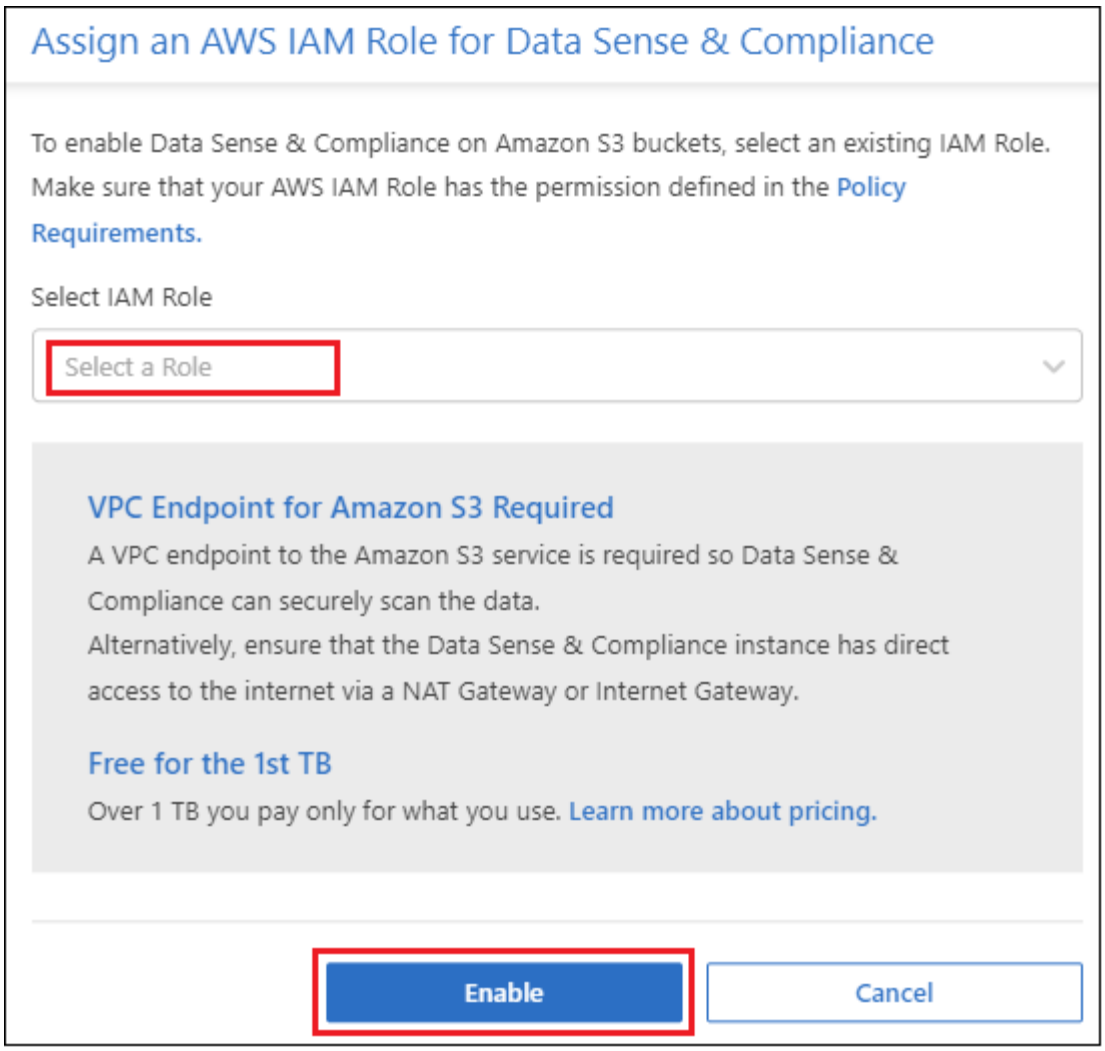
Steps

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

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



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



5. Select **Enable**.

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

Result

BlueXP assigns the IAM role to the instance.

Enabling and disabling compliance scans on S3 buckets

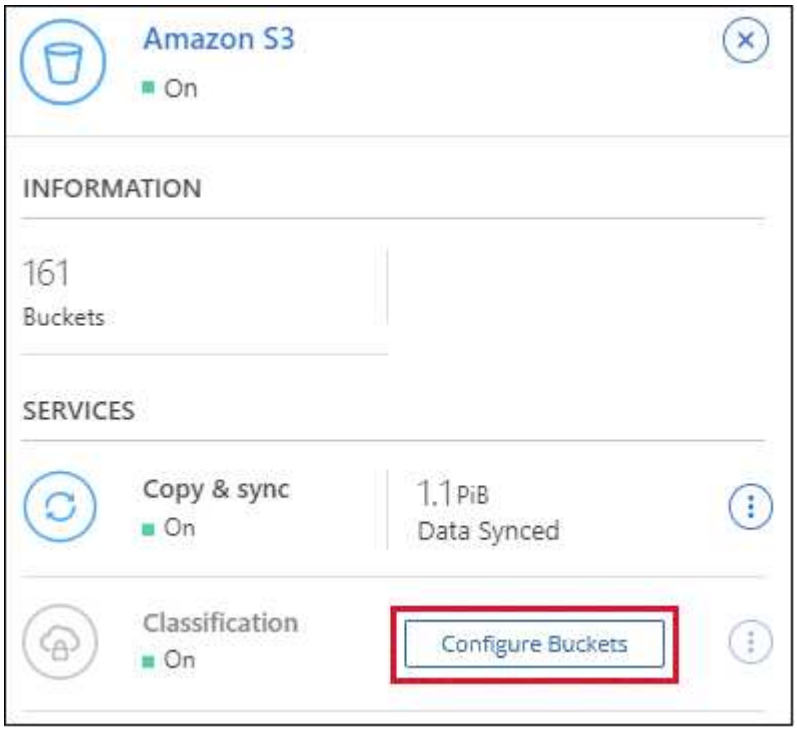
After BlueXP enables BlueXP classification on Amazon S3, the next step is to configure the buckets that you want to scan.

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

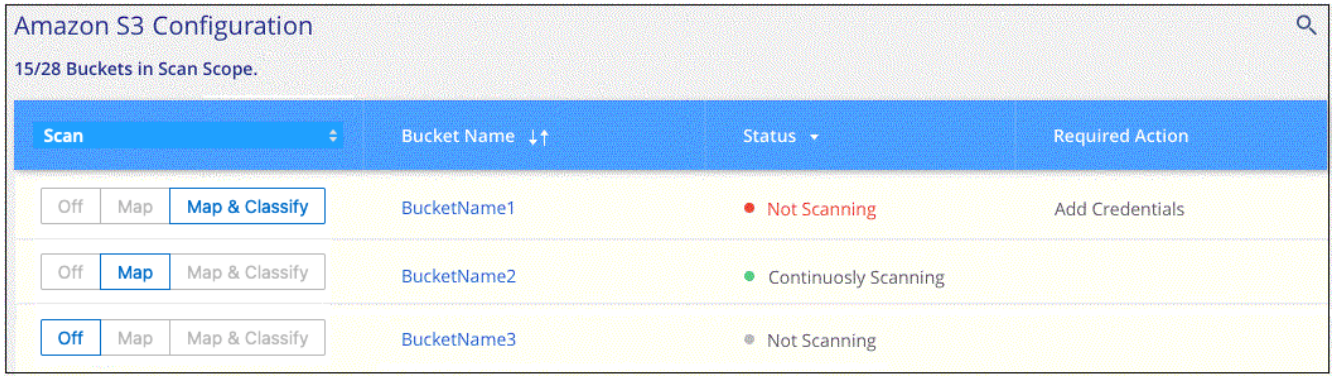
BlueXP classification can also scan S3 buckets that are in different AWS accounts.

Steps

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



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



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

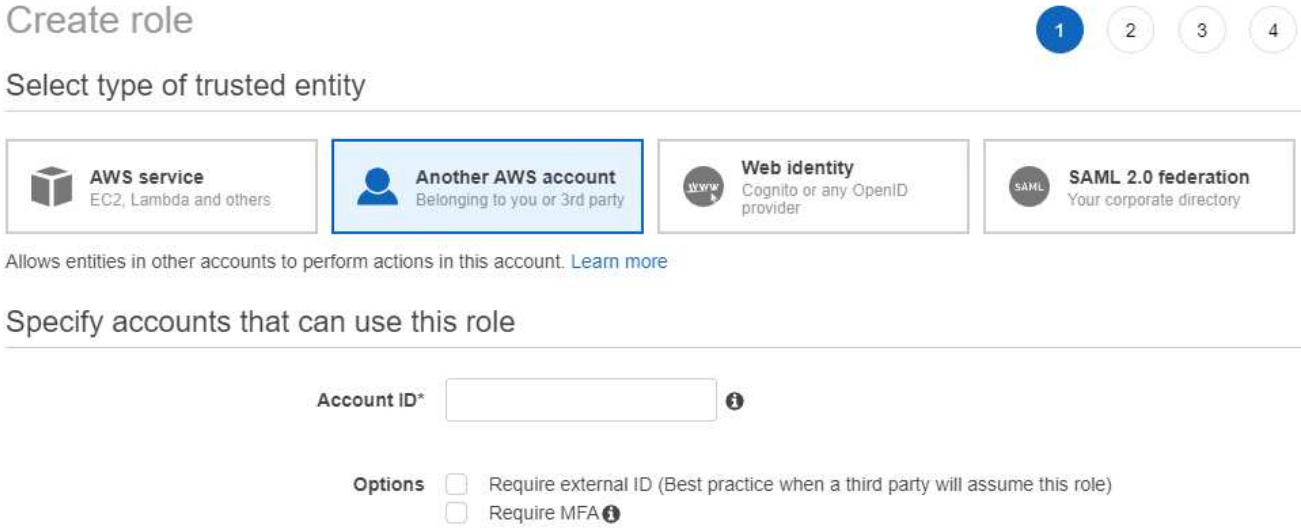
BlueXP classification starts scanning the S3 buckets that you enabled. If there are any errors, they'll appear in the Status column, alongside the required action to fix the error.

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**.



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.

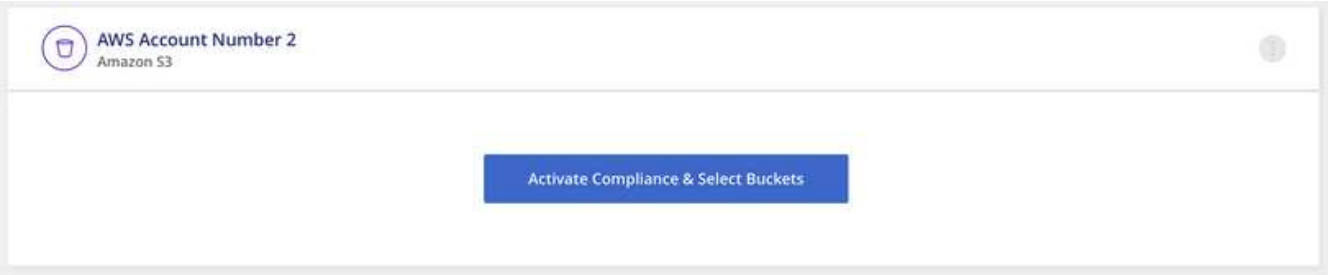
```
{
  "Version": "2012-10-17",
  "Statement": [
      {
          "Effect": "Allow",
          "Action": [
              "s3:Get*",
              "s3:List*",
              "s3:PutObject"
          ],
          "Resource": "*"
      },
  ]
}
```

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. Save the change.
   b. Select **Attach policies** then **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 receives access to the additional AWS account.

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



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

Result

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

Scan OneDrive accounts with BlueXP classification

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

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

Quick start

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

Review OneDrive prerequisites

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

Deploy the BlueXP classification instance

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

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.

Add the users and select the type of scanning

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

Reviewing OneDrive requirements

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

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

Deploying the BlueXP classification instance

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

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

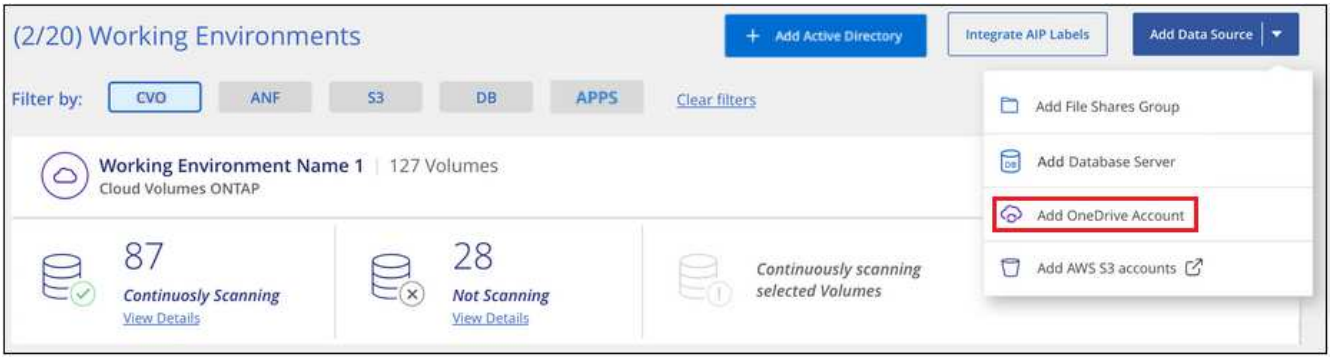
Upgrades to BlueXP classification software is automated as long as the instance has internet connectivity.

Adding the OneDrive account

Add the OneDrive account where the user files reside.

Steps

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



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

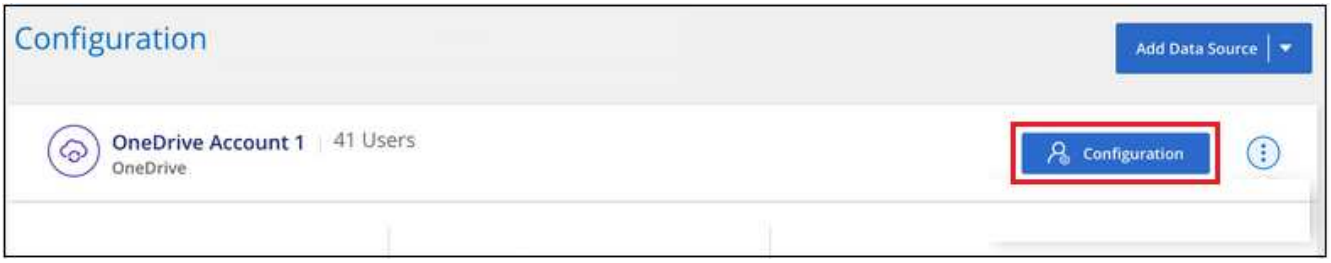
The OneDrive account is added to the list of working environments.

Adding OneDrive users to compliance scans

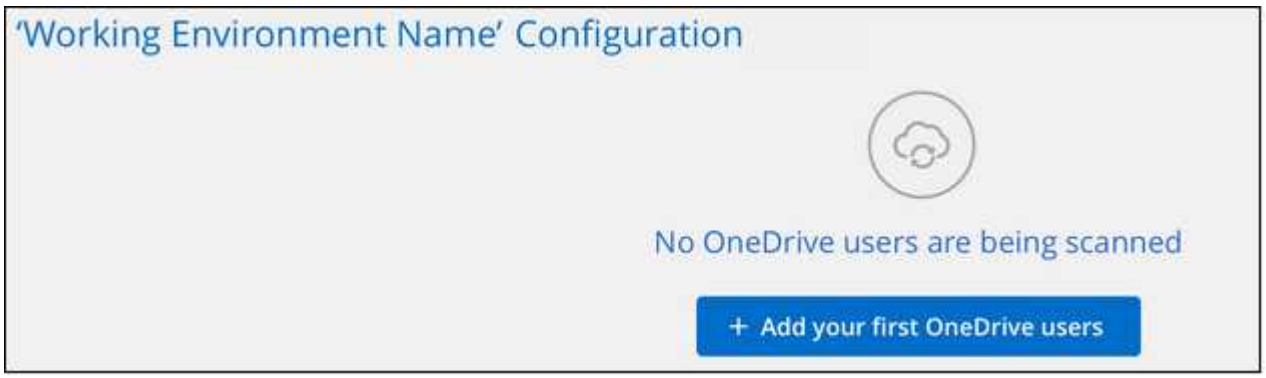
You can add individual OneDrive users, or all of your OneDrive users, so that their files will be scanned by BlueXP classification.

Steps

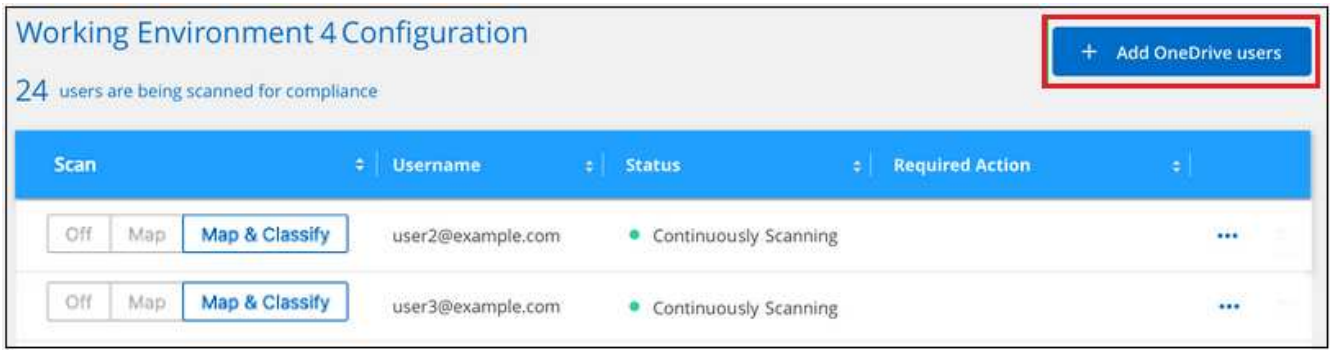
1. From the *Configuration* page, click the **Configuration** button for the OneDrive account.



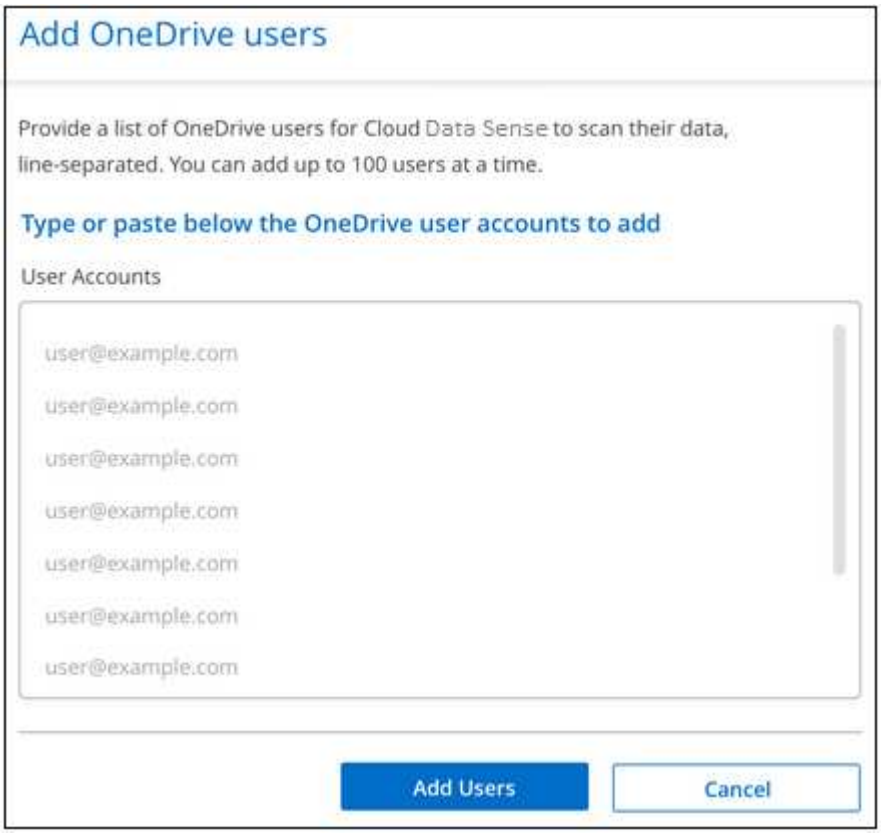
2. If this is the first time adding users for this OneDrive account, click **Add your first OneDrive users**.



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



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**.



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

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

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

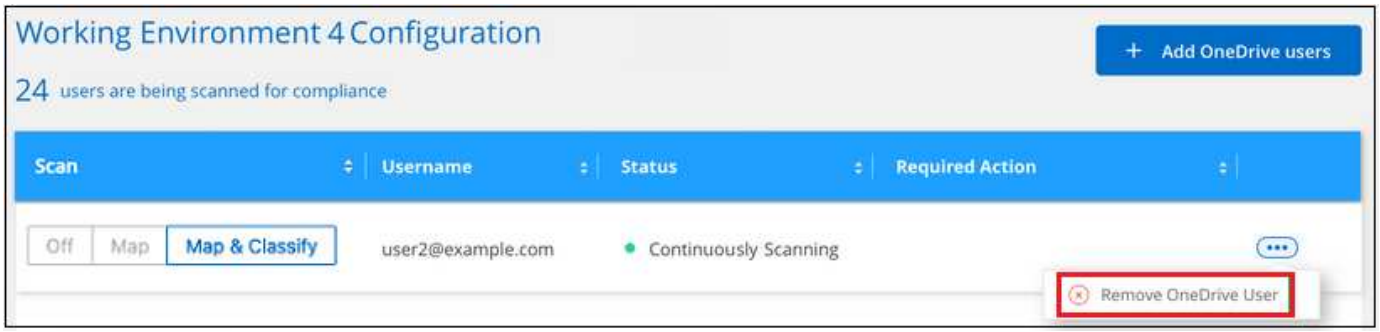
To:	Do this:
Enable mapping-only scans on user files	Click **Map**
Enable full scans on user files	Click **Map & Classify**
Disable scanning on user files	Click **Off**

Result

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

Removing a OneDrive user from compliance scans

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



Scan SharePoint accounts with BlueXP classification

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

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

Review SharePoint requirements

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

- You must have the Admin user login credentials for the SharePoint account that provides read access to all SharePoint sites.
  - For SharePoint Online you can use a non-Admin account, but that user must have permission to access all the SharePoint sites that you want to scan.
- For SharePoint On-Premise, you'll also need the URL of the SharePoint Server.
- You will need a line-separated list of the SharePoint site URLs for all the data you want to scan.

Deploy the BlueXP classification instance

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

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

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

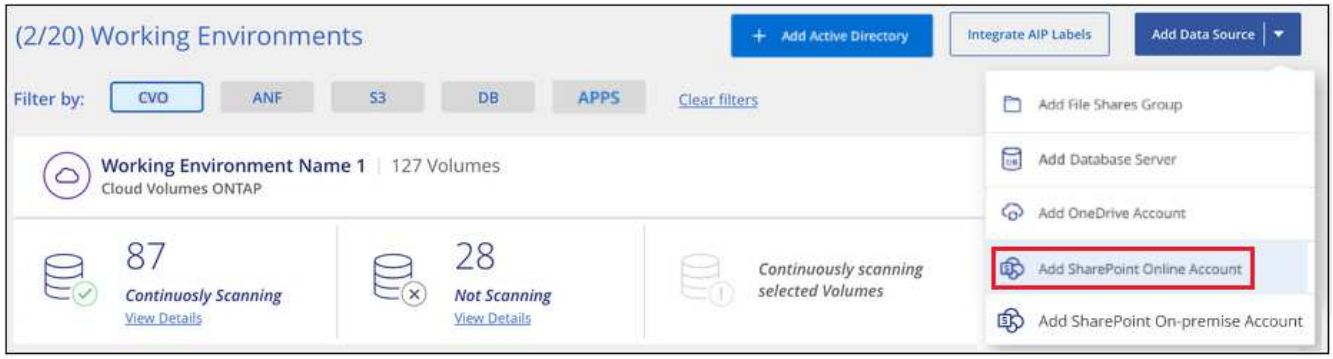
Upgrades to BlueXP classification software is automated as long as the instance has internet connectivity.

Add a SharePoint Online account

Add the SharePoint Online account where the user files reside.

Steps

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



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

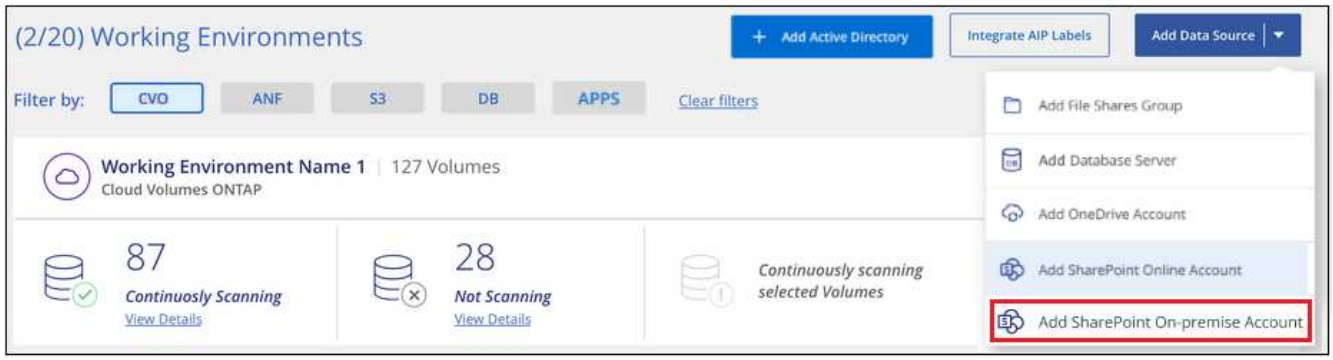
The SharePoint Online account is added to the list of working environments.

Add a SharePoint On-premise account

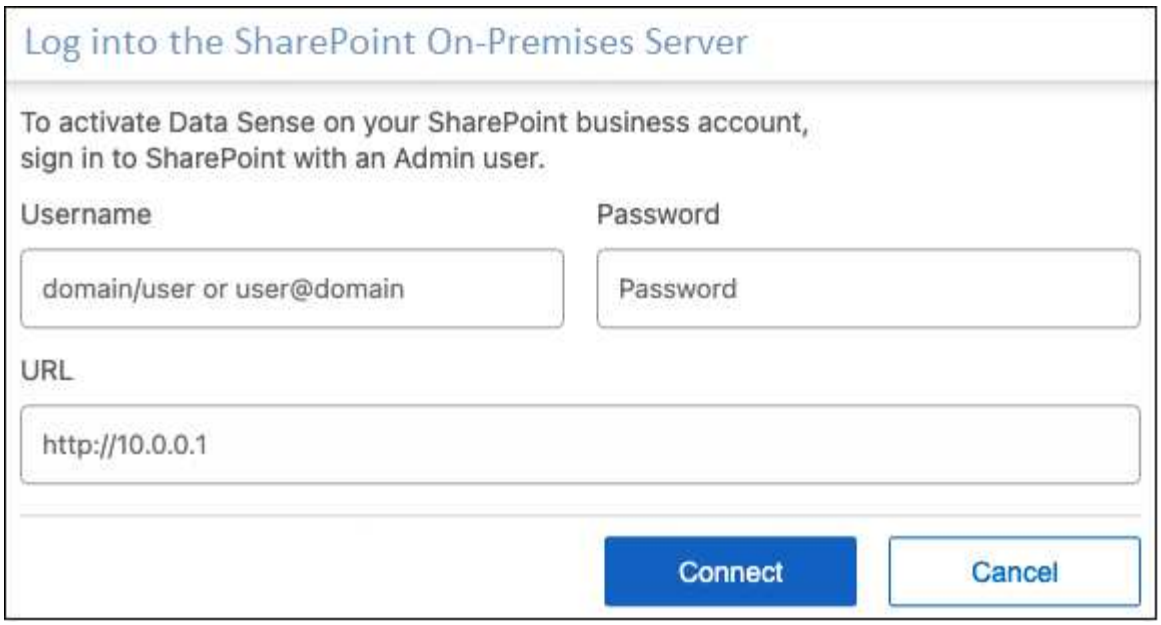
Add the SharePoint On-premise account where the user files reside.

Steps

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



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



3. Click **Connect**.

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

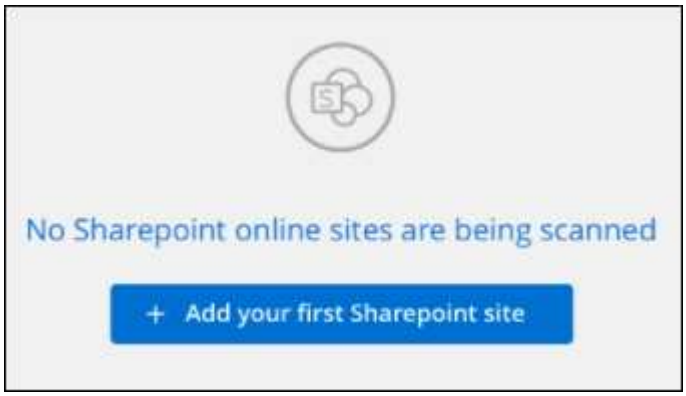
Add SharePoint sites to compliance scans

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

Steps

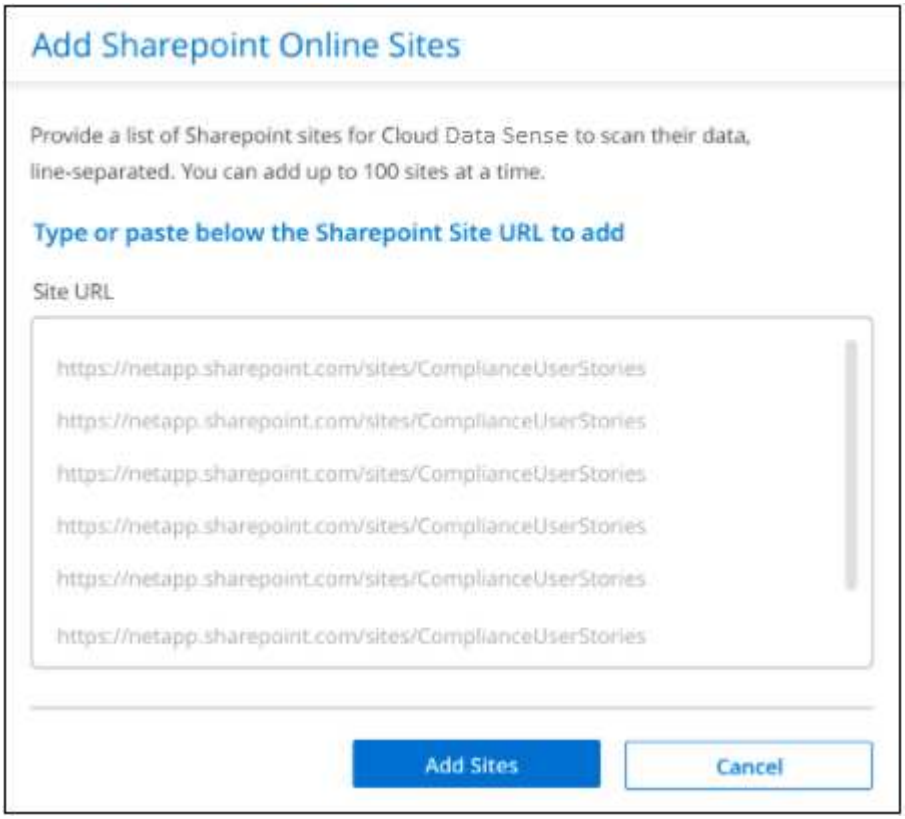
1. From the *Configuration* page, click the **Configuration** button for the SharePoint account.

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



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

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



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

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

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

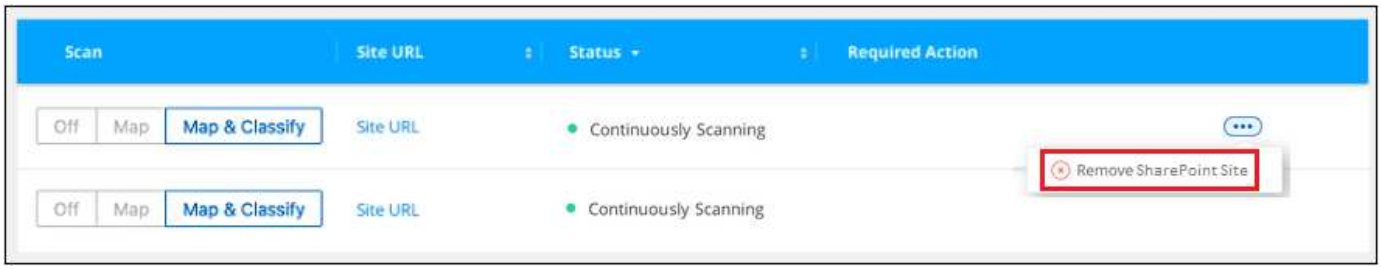
To:	Do this:
Enable mapping-only scans on files	Click **Map**
Enable full scans on files	Click **Map & Classify**
Disable scanning on files	Click **Off**

Result

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

Remove a SharePoint site from compliance scans

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



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

Scan Google Drive accounts with BlueXP classification

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

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

Quick start

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

1 **Review Google Drive prerequisites**

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

2 **Deploy BlueXP classification**

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

3 **Log into the Google Drive account**

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

4 **Select the type of scanning for the user files**

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

Review Google Drive requirements

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

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

Current restrictions

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

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

Deploy BlueXP classification

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

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

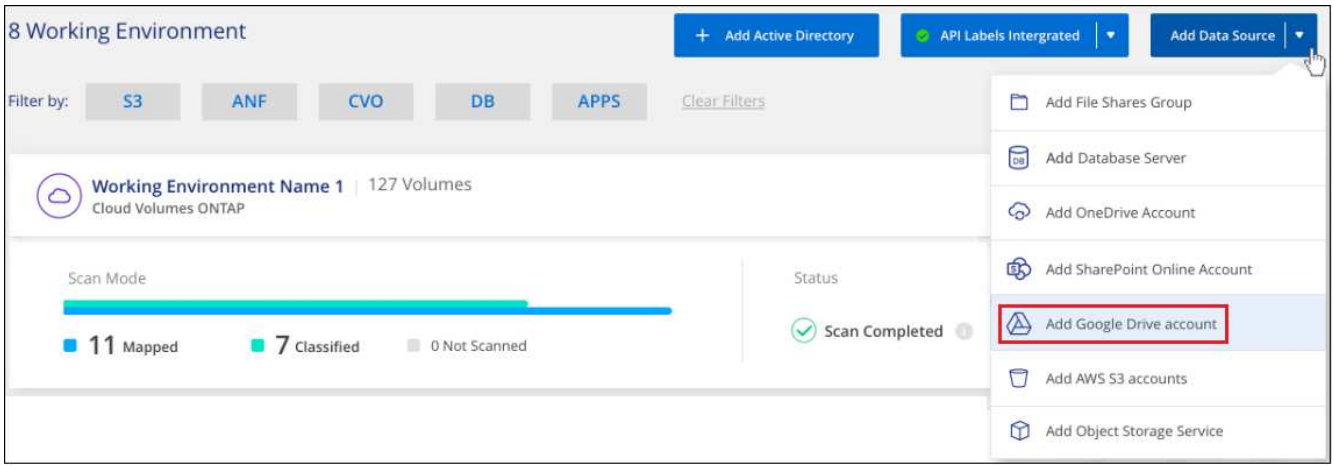
Upgrades to BlueXP classification software is automated as long as the instance has internet connectivity.

Add the Google Drive account

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

Steps

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



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

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

Select the type of scanning for user data

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

Steps

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

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

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

Result

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

Remove a Google Drive account from compliance scans

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

Scan StorageGRID data with BlueXP classification

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.

The following information is relevant only for BlueXP classification legacy versions 1.30 and earlier.
Beginning with version 1.31, BlueXP classification is part of the core BlueXP offering. For more information, see Scan StorageGRID data.

Quick start

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

Review object storage prerequisites

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

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

Deploy the BlueXP classification instance

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

Add the Object Storage Service

Add the object storage service to BlueXP classification.

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.

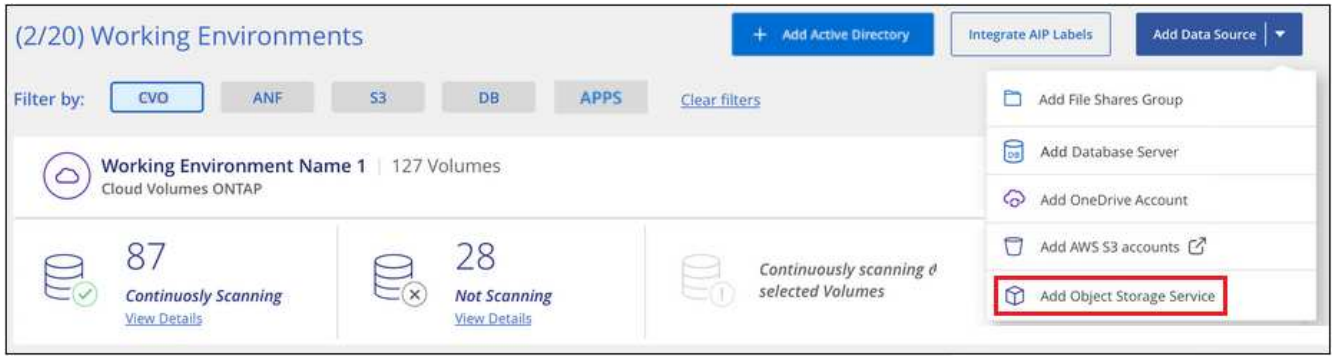
Upgrades to BlueXP classification software is automated as long as the instance has internet connectivity.

Adding the object storage service to BlueXP classification

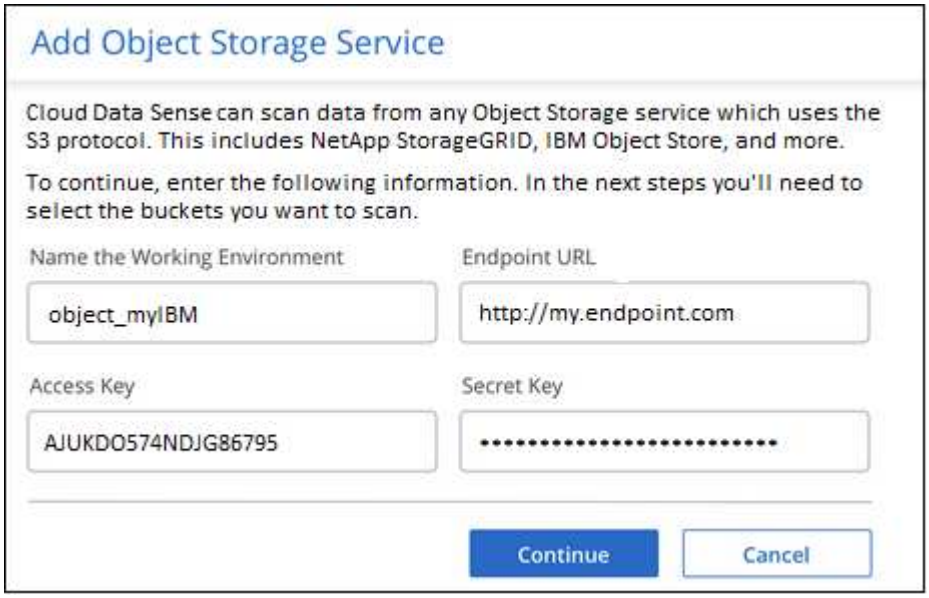
Add the object storage service.

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.



Result

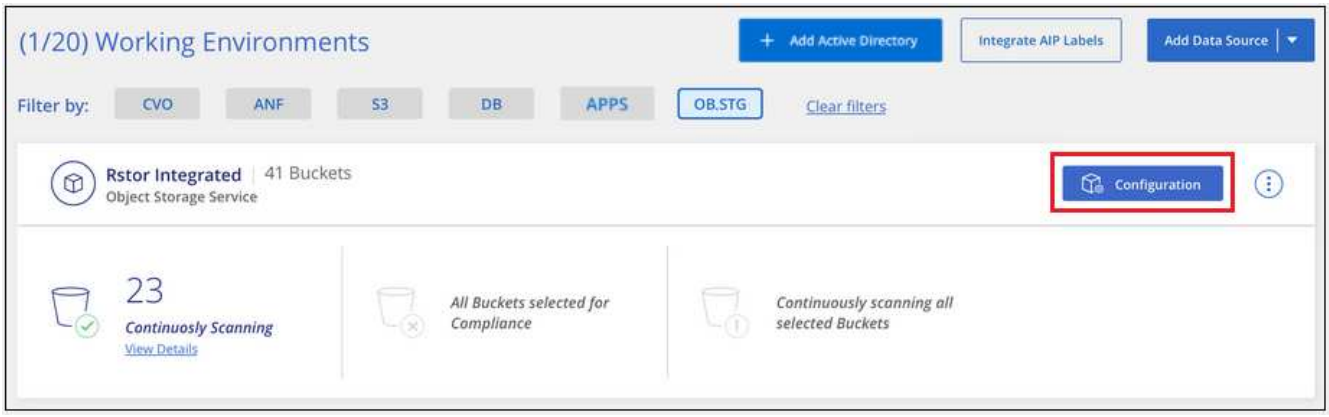
The new Object Storage Service is added to the list of working environments.

Enabling and disabling compliance scans on object storage buckets

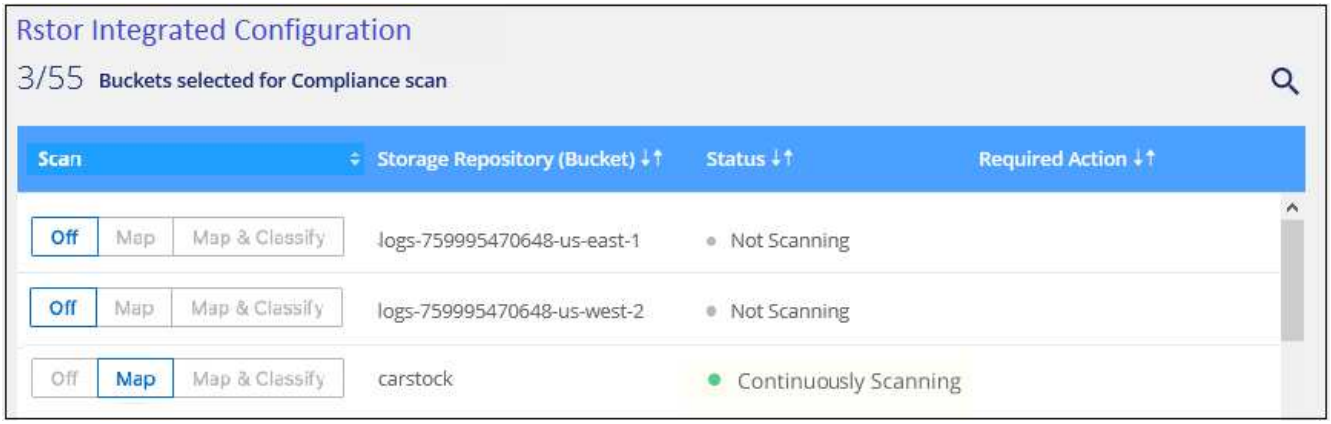
After you enable BlueXP classification on your Object Storage Service, the next step is to configure the buckets that you want to scan. BlueXP classification discovers those buckets and displays them in the working environment you created.

Steps

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



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



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

Result

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

Manage data deprecations

View governance details about your data using the BlueXP classification Governance dashboard

Gain control of the costs related to the data on your organizations' storage resources. BlueXP classification identifies the amount of stale data, non-business data, duplicate files, and very large files in your systems so you can decide whether you want to remove or tier some files to less expensive object storage.

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

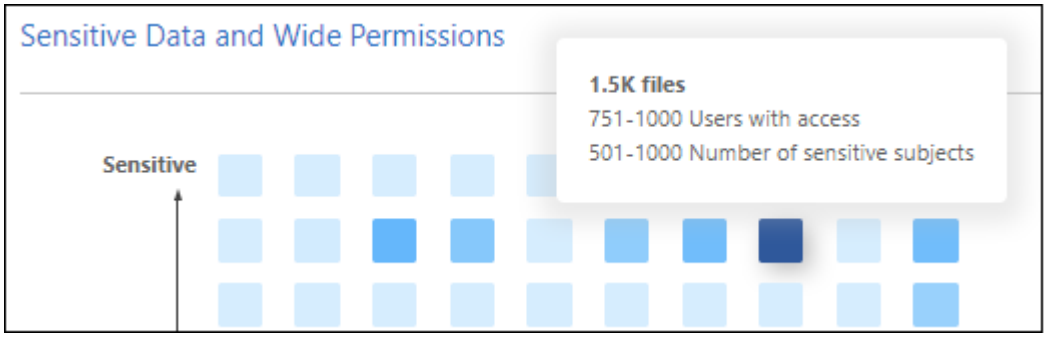
NOTE This information is relevant only for BlueXP classification legacy versions 1.30 and earlier.

Data listed by sensitivity and wide permissions on the Governance dashboard

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

This applies to BlueXP classification versions 1.30 and earlier.

Files are rated based on the number of users with permission to access the files on the X axis (lowest to highest), and the number of sensitive identifiers within the files on the Y axis (lowest to highest). The blocks represent the number of files that match the items from the X and Y axes. The lighter colored block are good; with fewer users able to access the files, and with fewer sensitive identifiers per file. The darker blocks are the items you may want to investigate. For example, the screen below shows the tooltip text for the dark blue block. It shows that you have 1,500 files where 751-100 users have access, and where there are 501-100 sensitive identifiers per file.



You can click the block you are interested in to view the filtered results of the affected files in the Investigation page so that you can investigate further.

No data is displayed in this panel if you haven't integrated an identity service with BlueXP classification. See how to integrate your Active Directory service with BlueXP classification.

This panel supports files in CIFS shares, OneDrive, and SharePoint data sources. There is currently no support for databases, Google Drive, Amazon S3, and generic object storage.

Classification area on the dashboard showing AIP labels

The *Classification* area on the dashboard provides a list of the most identified Azure Information Protection (AIP) Labels in your scanned data.

If you have subscribed to Azure Information Protection (AIP), you can classify and protect documents and files by applying labels to content. Reviewing the most used AIP labels that are assigned to files enables you to see which labels are most used in your files.

See AIP Labels for more information.

Organize your private data with BlueXP classification

BlueXP classification provides many ways for you to manage and organize your private data. This makes it easier to see the data that is most important to you.

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

- If you are subscribed to Azure Information Protection (AIP) to classify and protect your files, you can use BlueXP classification to manage AIP labels.
- You can add Tags to files that you want to mark for organization or for some type of follow-up.
- You can assign a BlueXP user to a specific file, or to multiple files, so that person can be responsible for managing the file.
- With the saved search functionality, you can create your own custom search queries so that you can easily see the results by clicking one button.
- You can send email alerts to BlueXP users, or any other email address, when certain critical Policies return results.

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

Should I use tags or labels?

Below is a comparison of BlueXP classification tagging and Azure Information Protection labeling.

Tags	Labels
File tags are an integrated part of BlueXP classification.	Requires that you have subscribed to Azure Information Protection (AIP).
The tag is only kept in the BlueXP classification database - it is not written to the file. It does not change the file, or the file accessed or modified times.	The label is part of the file and when the label changes, the file changes. This change also changes the file accessed and modified times.
You can have multiple tags on a single file.	You can have one label on a single file.
The tag can be used for internal BlueXP classification action, such as copy, move, delete, run a policy, etc.	Other systems that can read the file can see the label - which can be used for additional automation.
Only a single API call is used to see if a file has a tag.	

Categorize your data using AIP labels

You can manage AIP labels in the files that BlueXP classification is scanning if you have subscribed to Azure Information Protection (AIP). AIP enables you to classify and protect documents and files by applying labels to content. BlueXP classification enables you to view the labels that are already assigned to files, add labels to files, and change labels when a label already exists.

BlueXP classification supports AIP labels within the following file types: .DOC, .DOCX, .PDF, .PPTX, .XLS, .XLSX.

- You can't currently change labels in files larger than 30 MB. For OneDrive, SharePoint, and Google Drive accounts the maximum file size is 4 MB.
- If a file has a label which doesn't exist anymore in AIP, BlueXP classification considers it as a file without a label.
- If you've deployed BlueXP classification in a Government region, or in an on-prem location that has no internet access (also known as a dark site), then the AIP label functionality is unavailable.

Integrate AIP labels in your project or workspace

Before you can manage AIP labels, you need to integrate the AIP label functionality into BlueXP classification by signing into your existing Azure account. Once enabled, you can manage AIP labels within files for all data sources in your BlueXP project or workspace.

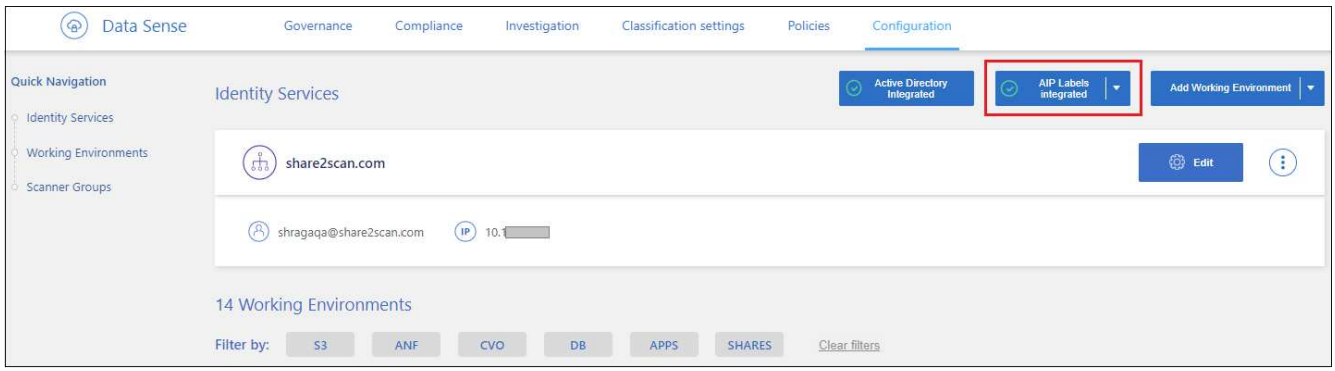
Requirements

- You must have an account and an Azure Information Protection license.
- You must have the login credentials for the Azure account.
- If you plan to change labels in files that reside in Amazon S3 buckets, ensure that the permission `s3:PutObject` is included in the IAM role. See setting up the IAM role.

Steps

1. From the BlueXP classification Configuration page, click **Integrate AIP Labels**.

2. In the Integrate AIP Labels dialog, click **Sign in to Azure**.
3. In the Microsoft page that appears, select the account and enter the required credentials.
4. Return to the BlueXP classification tab and you'll see the message "*AIP Labels were integrated successfully with the account <account_name>*".
5. Click **Close** and you'll see the text *AIP Labels integrated* at the top of the page.



Result

You can view and assign AIP labels from the results pane of the Investigation page. You can also assign AIP

labels to files using Policies.

View AIP labels in your files

You can view the current AIP label that is assigned to a file.

In the Data Investigation results pane, click ⌄ for the file to expand the file metadata details.

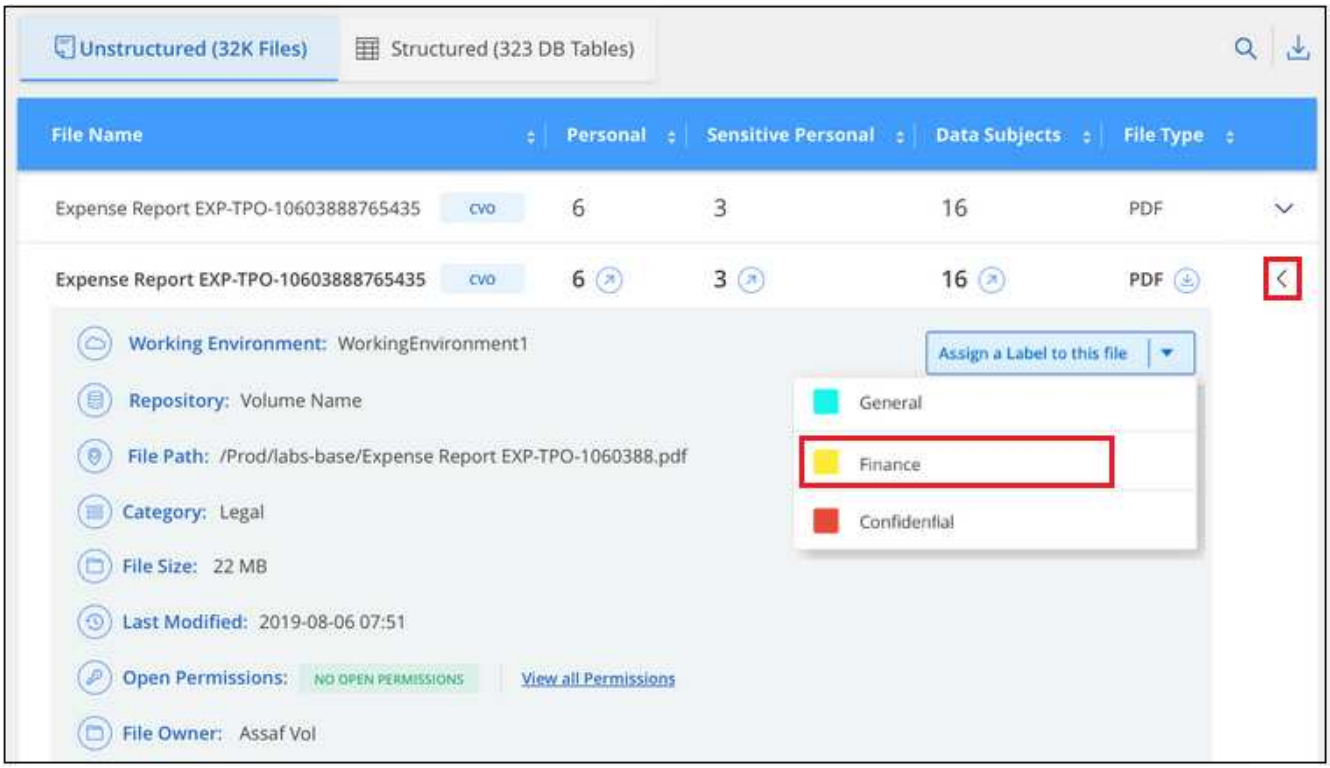
Assign AIP labels manually

You can add, change, and remove AIP labels from your files using BlueXP classification.

Follow these steps to assign an AIP label to a single file.

Steps

1. In the Data Investigation results pane, click ⌄ for the file to expand the file metadata details.



2. Click **Assign a Label to this file** and then select the label.

   The label appears in the file metadata.

Follow these steps to assign an AIP label to multiple files. Note that you can assign an AIP label to a maximum of 20 files at a time (one page in the UI).

Steps

1. In the Data Investigation results pane, select the file, or files, that you want to label.

   - To select individual files, check the box for each file ().
   - To select all files on the current page, check the box in the title row ().

2. From the button bar, click **Label** and select the AIP label:

   The AIP label is added to the metadata for all selected files.

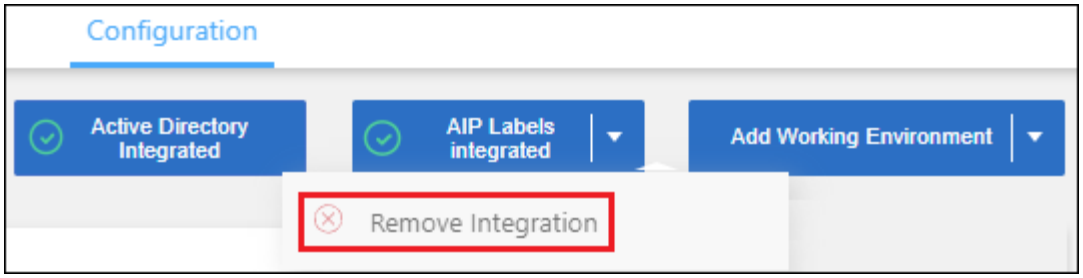
Remove the AIP integration

If you no longer want the ability to manage AIP labels in files, you can remove the AIP account from the BlueXP classification interface.

Note that no changes are made to the labels you have added using BlueXP classification. The labels that exist in files will stay as they currently exist.

Steps

1. From the *Configuration* page, click **AIP Labels integrated > Remove Integration**.



2. Click **Remove Integration** from the confirmation dialog.

Apply tags to manage your scanned files

You can add a tag to files that you want to mark for some type of follow-up. For example, you may have found some duplicate files and you want to delete one of them, but you need to check to see which one should be deleted. You could add a tag of "Check to delete" to the file so you know this file requires some research and some type of future action.

BlueXP classification enables you to view the tags that are assigned to files, add or remove tags from files, and change the name or delete an existing tag.

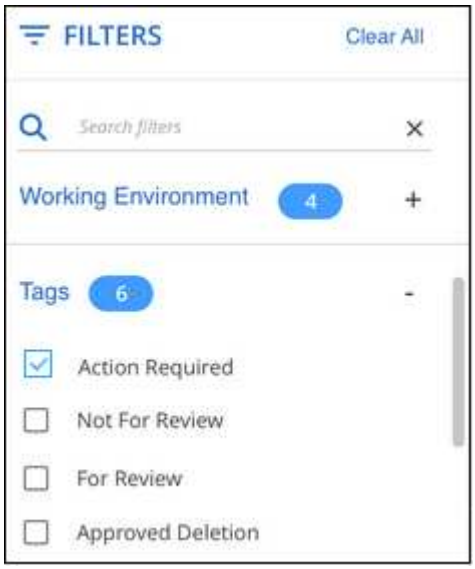
Note that the tag is not added to the file in the same way as AIP Labels are part of the file metadata. The tag is just seen by BlueXP users using BlueXP classification so you can see if a file needs to be deleted or checked for some type of follow-up.

Tags assigned to files in BlueXP classification are not related to the tags you can add to resources, such as volumes or virtual machine instances. BlueXP classification tags are applied at the file level.

View files that have certain tags applied

You can view all the files that have specific tags assigned.

1. Click the **Investigation** tab from BlueXP classification.
2. In the Data Investigation page, click **Tags** in the Filters pane and then select the required tags.



The Investigation Results pane displays all the files that have those tags assigned.

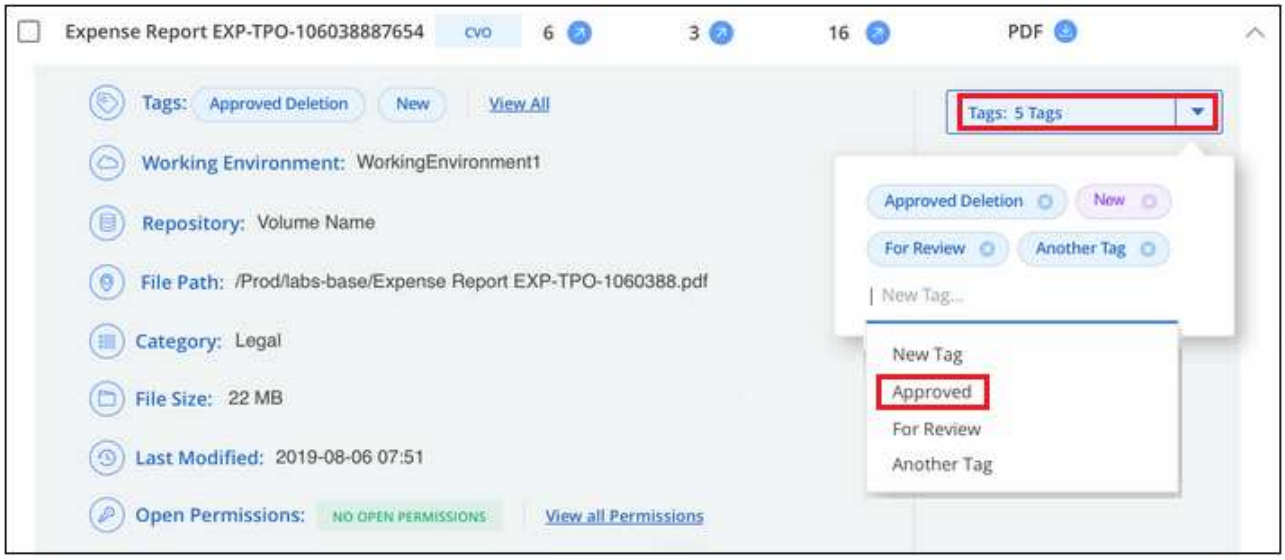
Assign tags to files

You can add tags to a single file or to a group of files.

To add a tag to a single file:

Steps

1. In the Data Investigation results pane, click ⌄ for the file to expand the file metadata details.
2. Click the **Tags** field and the currently assigned tags are displayed.
3. Add the tag or tags:
   - To assign an existing tag, click in the **New Tag…** field and start typing the name of the tag. When the tag you are looking for appears, select it and press **Enter**.
   - To create a new tag and assign it to the file, click in the **New Tag…** field, enter the name of the new tag, and press **Enter**.



The tag appears in the file metadata.

To add a tag to multiple files:

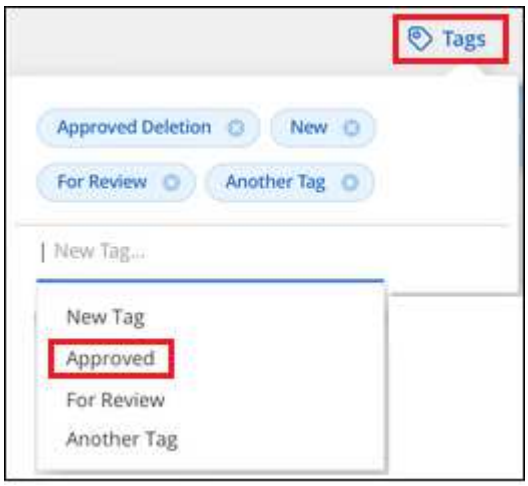
Steps

1. In the Data Investigation results pane, select the file, or files, that you want to tag.

- To select individual files, check the box for each file (Volume_1).
- To select all files on the current page, check the box in the title row (File Name).
- To select all files on all pages, check the box in the title row (File Name), and then in the pop-up message All **20** Items on this page selected Select all Items in list (63K Items), click **Select all items in list (xxx items)**.

  You can apply tags to a maximum of 100,000 files at a time.

2. From the button bar, click **Tags** and the currently assigned tags are displayed.
3. Add the tag or tags:
   - To assign an existing tag, click in the **New Tag…** field and start typing the name of the tag. When the tag you are looking for appears, select it and press **Enter**.
   - To create a new tag and assign it to the file, click in the **New Tag…** field, enter the name of the new tag, and press **Enter**.



4. Approve adding the tags in the confirmation dialog and the tags are added to the metadata for all selected files.

Delete tags from files

You can delete a tag if you don't need to use it anymore.

Just click the **x** for an existing tag.

If you had selected multiple files, the tag is removed from all the files.

Assign users to manage certain files

You can assign a BlueXP user to a specific file, or to multiple files, so that person can be responsible for any follow-up actions that need to be done on the file. This capability is often used with the feature to add custom Status tags to a file.

For example, you might have a file that contains certain personal data that allows too many users read and write access (open permissions). So you could assign the Status tag "Change permissions" and assign this file to user "Joan Smith" so they can decide how to fix the issue. When they have fixed the issue they could change the Status tag to "Completed".

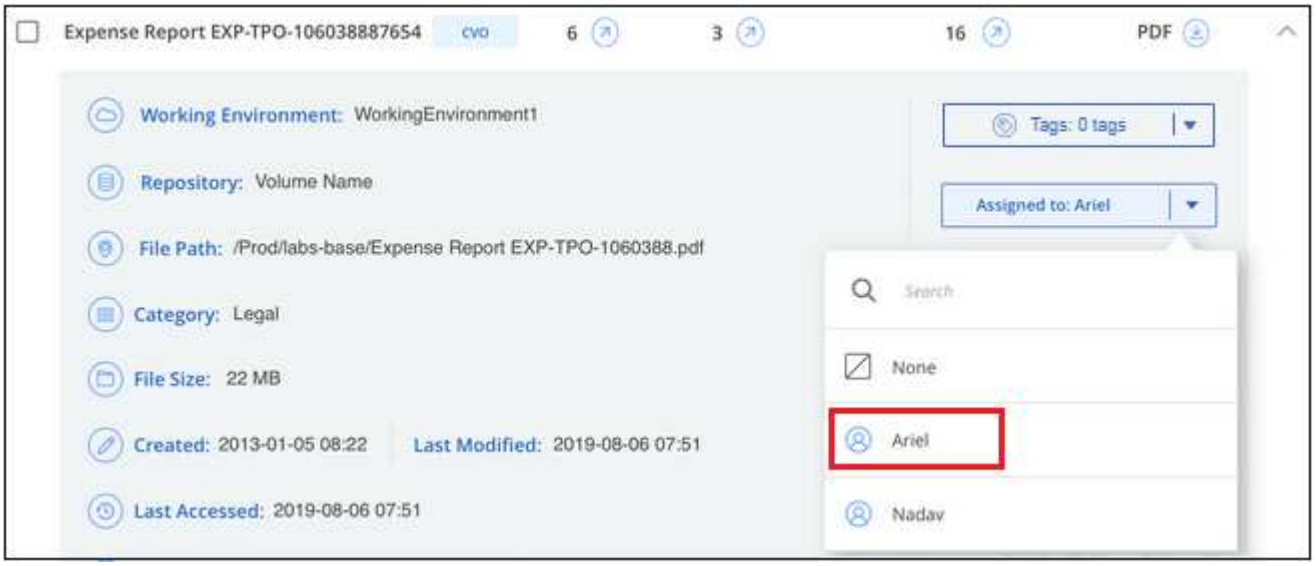
Note that the user name is not added to the file as part of the file metadata - it is just seen by BlueXP users when using BlueXP classification.

A new Filter in the Investigation page enables you to easily view all files that have the same person in the "Assigned To" field.

Follow these steps to assign a user to a single file.

Steps

1. In the Data Investigation results pane, click ⌄ for the file to expand the file metadata details.
2. Click the **Assigned to** field and select the user name.



The User name appears in the file metadata.

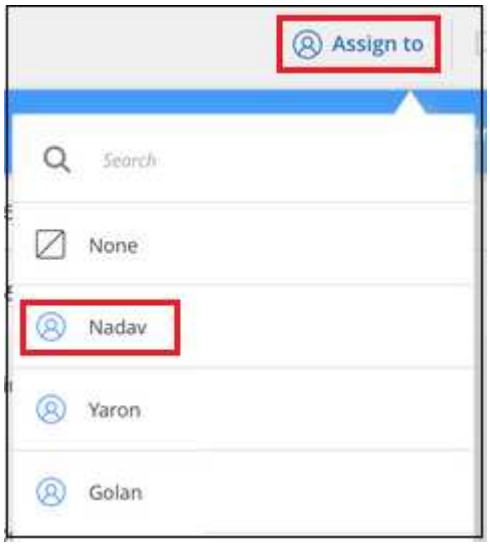
Follow these steps to assign a user to multiple files. Note that you can assign a user to a maximum of 20 files at a time (one page in the UI).

Steps

1. In the Data Investigation results pane, select the file, or files, that you want to assign to a user.

- To select individual files, check the box for each file (Volume_1).
- To select all files on the current page, check the box in the title row (File Name).

2. From the button bar, click **Assign to** and select the user name:



The user is added to the metadata for all selected files.

Manage your private data with BlueXP classification

BlueXP classification provides many ways for you to manage your private data. Some functionality makes it easier to prepare for migrating your data, while other functionality allows you to make changes to the data.

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

- You can copy files to a destination NFS share if you want to make a copy of certain data and move it to a different NFS location.
- You can clone an ONTAP volume to a new volume, while including only selected files from the source volume in the new cloned volume. This is useful for situations where you're migrating data and you want to exclude certain files from the original volume.
- You can copy and synchronize files from a source repository to a directory in a specific destination location. This is useful for situations where you're migrating data from one source system to another while there is

still some final activity on the source files.

- You can move source files that BlueXP classification is scanning to any NFS share.
- You can delete files that seem insecure or too risky to leave in your storage system, or that you have identified as duplicate.

- The capabilities described in this section are available only if you have chosen to perform a full classification scan on your data sources. Data sources that have had a mapping-only scan do not show file-level details.
- Data from Google Drive accounts can't use any of these capabilities at this time.

Copy source files

You can copy any source files that BlueXP classification is scanning. There are three types of copy operations depending on what you're trying to accomplish:

- **Copy files** from the same, or different, volumes or data sources to a destination NFS share.

  This is useful if you want to make a copy of certain data and move it to a different NFS location.

- **Clone an ONTAP volume** to a new volume in the same aggregate, but include only selected files from the source volume in the new cloned volume.

  This is useful for situations where you're migrating data and you want to exclude certain files from the original volume. This action uses the NetApp FlexClone functionality to quickly duplicate the volume and then remove the files that you **didn't** select.

- **Copy and synchronize files** from a single source repository (ONTAP volume, S3 bucket, NFS share, etc.) to a directory in a specific destination (target) location.

  This is useful for situations where you're migrating data from one source system to another. After the initial copy, the service syncs any changed data based on the schedule that you set. This action uses the NetApp BlueXP copy and sync functionality to copy and sync data from a source to a target.

Copy source files to an NFS share

You can copy source files that BlueXP classification is scanning to any NFS share. The NFS share doesn't need to be integrated with BlueXP classification, you just need to know the name of the NFS share where all selected files will be copied in the format `<host_name>:/<share_path>`.

You can't copy files that reside in databases.

Requirements

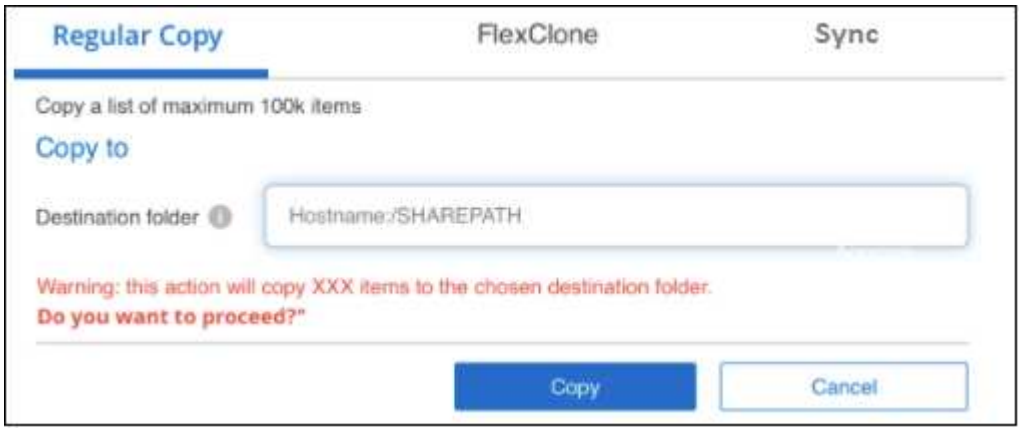
- You must have permissions to copy files. Learn about user access to compliance information.
- Copying files requires that the destination NFS share allows access from the BlueXP classification instance.
- You can copy between 1 and 100,000 files at a time.

Steps

1. In the Data Investigation results pane, select the file, or files, that you want to copy, and click **Copy**.

- To select individual files, check the box for each file (Volume_1).
- To select all files on the current page, check the box in the title row (File Name).
- To select all files on all pages, check the box in the title row (File Name), and then in the pop-up message All **20** Items on this page selected Select all Items in list (63K Items), click **Select all items in list (xxx items)**.

2. In the *Copy Files* dialog, select the **Regular Copy** tab.



3. Enter the name of the NFS share where all selected files will be copied in the format `<host_name>:/<share_path>`, and click **Copy**.

   A dialog appears with the status of the copy operation.

You can view the progress of the copy operation in the Actions Status pane.

Note that you can also copy an individual file when viewing the metadata details for a file. Just click **Copy File**.

Clone volume data to a new volume

You can clone an existing ONTAP volume that BlueXP classification is scanning using NetApp *FlexClone* functionality. This allows you to quickly duplicate the volume while including only those files you selected. This is useful if you're migrating data and you want to exclude certain files from the original volume, or if you want to create a copy of a volume for testing.

The new volume is created in the same aggregate as the source volume. Ensure that you have enough space for this new volume in the aggregate before you start this task. Contact your storage administrator if necessary.

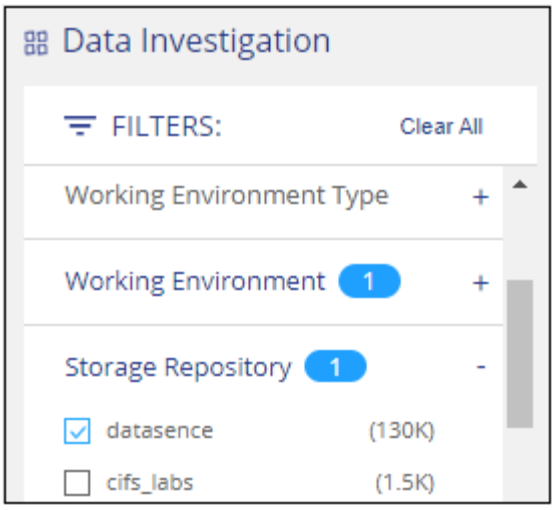
Note: FlexGroup volumes can't be cloned because they're not supported by FlexClone.

Requirements

- You must have permissions to copy files. Learn about user access to compliance information.
- You must select a minimum of 20 files.
- All selected files must be from the same volume, and the volume must be online.
- The volume must be from a Cloud Volumes ONTAP or on-premises ONTAP system. No other data sources are currently supported.
- The FlexClone license must be installed on the cluster. This license is installed by default on Cloud Volumes ONTAP systems.

Steps

1. In the Data Investigation pane, create a filter by selecting a single **Working Environment** and a single **Storage Repository** to make sure all the files are from the same ONTAP volume.

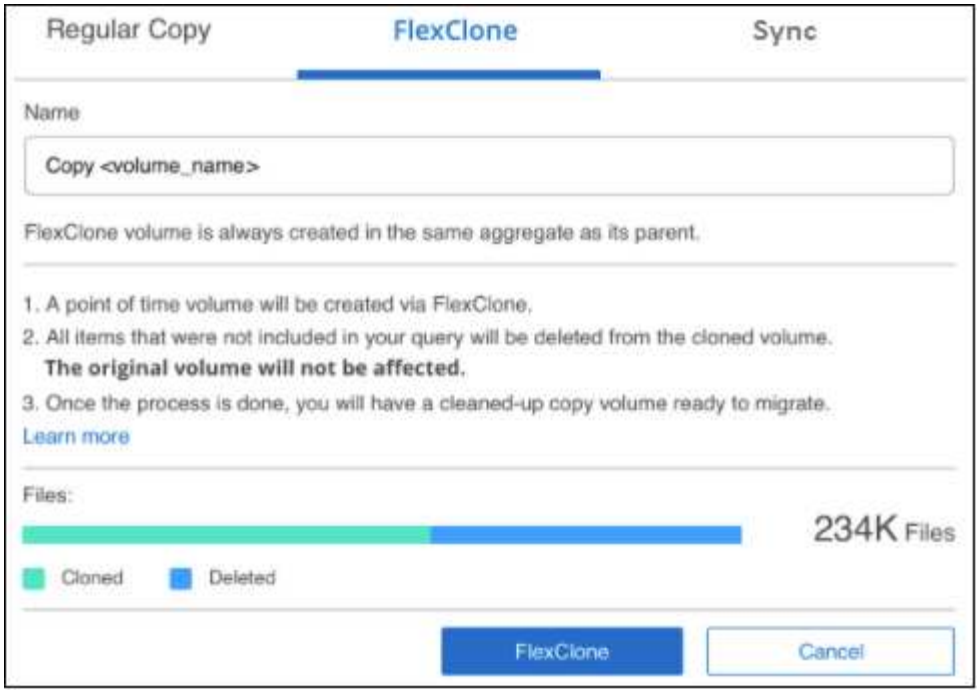


Apply any other filters so that you're seeing only the files that you want to clone to the new volume.

2. In the Investigation results pane, select the files that you want to clone and click **Copy**.

   - To select individual files, check the box for each file (☑ Volume_1).
   - To select all files on the current page, check the box in the title row (☑ File Name).
   - To select all files on all pages, check the box in the title row (☑ File Name), and then in the pop-up message All **20** Items on this page selected Select all Items in list (63K Items), click **Select all items in list (xxx items)**.

3. In the *Copy Files* dialog, select the **FlexClone** tab. This page shows the total number of files that will be cloned from the volume (the files you selected), and the number of files that are not included/deleted (the files you didn't select) from the cloned volume.



4. Enter the name of the new volume, and click **FlexClone**.

   A dialog appears with the status of the clone operation.

Result

The new, cloned volume is created in the same aggregate as the source volume.

You can view the progress of the clone operation in the Actions Status pane.

If you initially selected **Map all volumes** or **Map & Classify all volumes** when you enabled BlueXP classification for the working environment where the source volume resides, then BlueXP classification will scan the new cloned volume automatically. If you didn't use either of these selections initially, then if you want to scan this new volume, you'll need to enable scanning on the volume manually.

Copy and synchronize source files to a target system

You can copy source files that BlueXP classification is scanning from any supported unstructured data source to a directory in a specific target destination location (target locations that are supported by BlueXP copy and sync). After the initial copy, any data changed in the files are synchronized based on the schedule that you configure.

This is useful for situations where you're migrating data from one source system to another. This action uses the NetApp BlueXP copy and sync functionality to copy and sync data from a source to a target.

You can't copy and sync files that reside in databases, OneDrive accounts, or SharePoint accounts.

Requirements

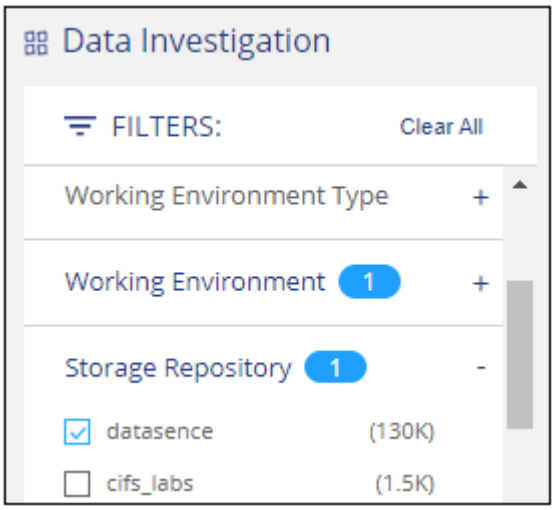
- You must have permissions to copy and sync files. Learn about user access to compliance information.
- You must select a minimum of 20 files.
- All selected files must be from the same source repository (ONTAP volume, S3 bucket, NFS or CIFS share, etc.).

- You'll need to activate the BlueXP copy and sync service and configure a minimum of one data broker that can be used to transfer files between the source and target systems. Review the BlueXP copy and sync requirements beginning with the Quick Start description.

  Note that the BlueXP copy and sync service has separate service charges for your sync relationships, and will incur resource charges if you deploy the data broker in the cloud.

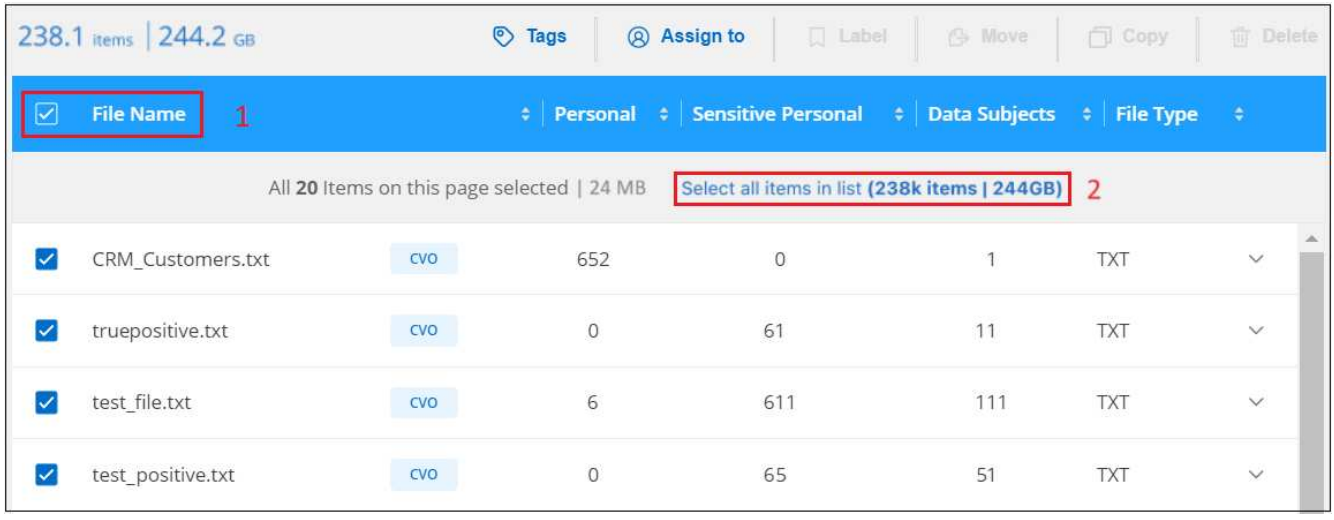
Steps

1. In the Data Investigation pane, create a filter by selecting a single **Working Environment** and a single **Storage Repository** to make sure all the files are from the same repository.

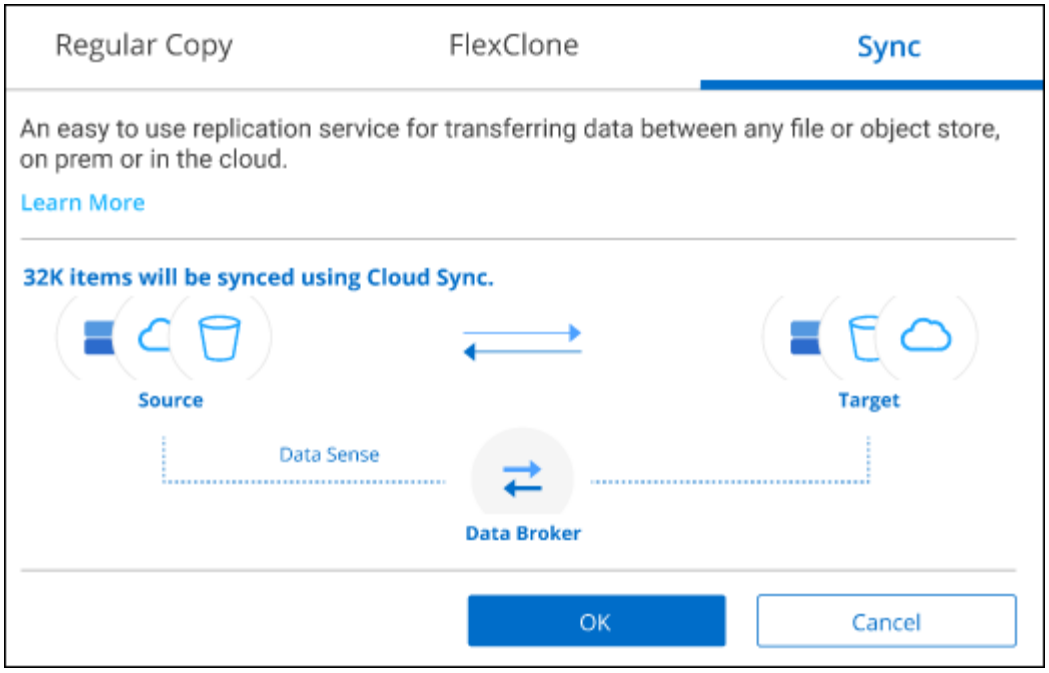


   Apply any other filters so that you're seeing only the files that you want to copy and sync to the destination system.

2. In the Investigation results pane, select all files on all pages by checking the box in the title row (), then in the pop-up message  click **Select all items in list (xxx items)**, and then click **Copy**.



3. In the *Copy Files* dialog, select the **Sync** tab.



4. If you are sure that you want to sync the selected files to a destination location, click **OK**.

   The BlueXP copy and sync UI is opened in BlueXP.

   You are prompted to define the sync relationship. The Source system is pre-populated based on the repository and files you already selected in BlueXP classification.

5. You'll need to select the Target system and then select (or create) the Data Broker you plan to use. Review the BlueXP copy and sync requirements beginning with the Quick Start description.

Result

The files are copied to the target system and they'll be synchronized based on the schedule you define. If you select a one-time sync then the files are copied and synchronized one time only. If you choose a periodic sync, then the files are synchronized based on the schedule. Note that if the source system adds new files that match the query you created using filters, those *new* files will be copied to the destination and synchronized in the future.

Note that some of the usual BlueXP copy and sync operations are disabled when it is invoked from BlueXP classification:

- You can't use the **Delete Files on Source** or **Delete Files on Target** buttons.
- Running a report is disabled.

Move source files to an NFS share

You can move source files that BlueXP classification is scanning to any NFS share. The NFS share doesn't need to be integrated with BlueXP classification.

Optionally, you can leave a breadcrumb file in the location of the moved file. A breadcrumb file helps your users understand why a file was moved from its original location. For each moved file, the system creates a breadcrumb file in the source location named `<filename>-breadcrumb-<date>.txt`. You can add text in the dialog box that will be added to the breadcrumb file to indicate the location where the file was moved and the user who moved the file.

Note that the subdirectory structure from the source file is recreated on the destination share when the file is moved so it is easier to understand where the file was moved from. If a file with the same name exists in the destination location, the file will not be moved.

You can't move files that reside in databases.

Requirements

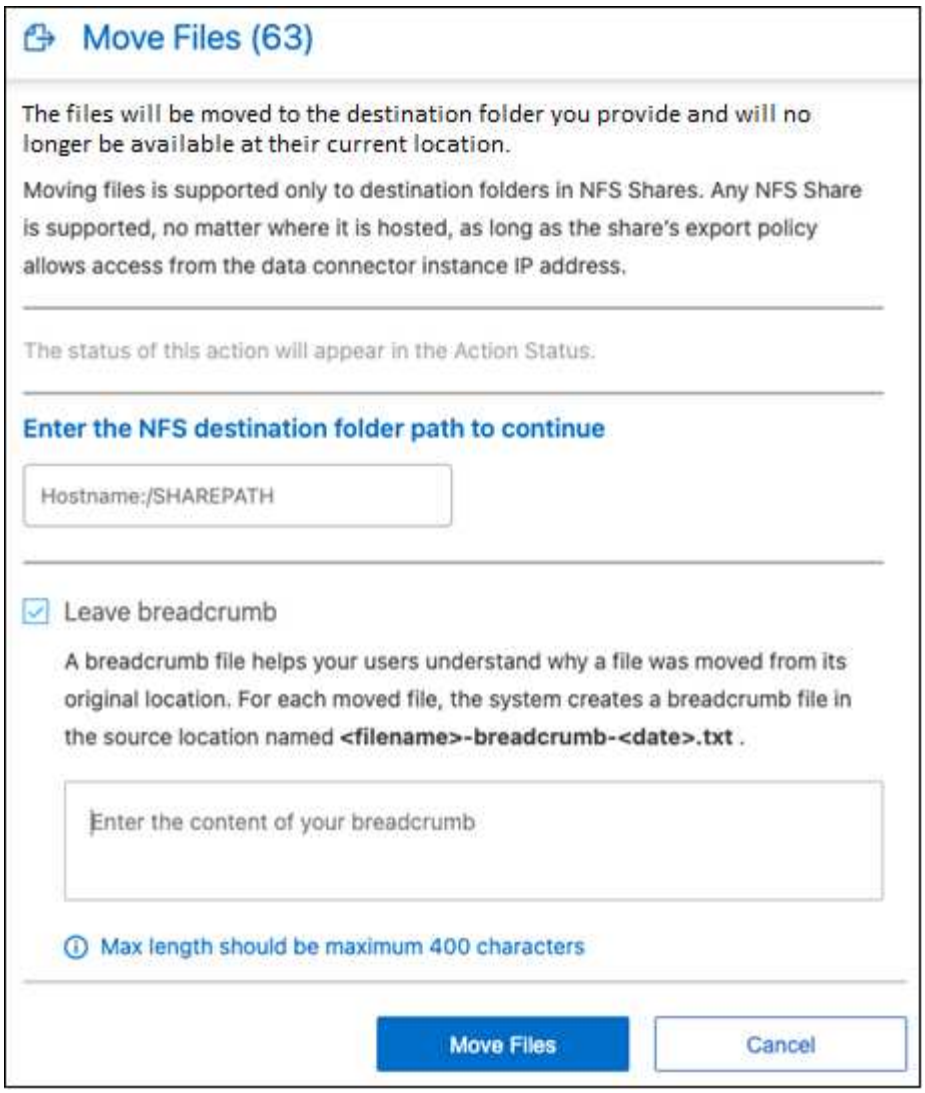
- You must have permissions to move files. Learn about user access to compliance information.
- The source files can be located in the following data sources: On-premises ONTAP, Cloud Volumes ONTAP, Azure NetApp Files, File Shares, and SharePoint Online.
- You can move a maximum of 15 million files at a time.
- Only files which are 50 MB or smaller are moved.
- The destination NFS share must allow access from the BlueXP classification instance IP address.

Steps

1. In the Data Investigation results pane, select the file, or files, that you want to move.

   - To select individual files, check the box for each file (Volume_1).
   - To select all files on the current page, check the box in the title row (File Name).
   - To select all files on all pages, check the box in the title row (File Name), and then in the pop-up message All 20 Items on this page selected Select all Items in list (63K Items), click **Select all items in list (xxx items)**.

2. From the button bar, click **Move**.



3. In the *Move Files* dialog, enter the name of the NFS share where all selected files will be moved in the format `<host_name>:/<share_path>`.
4. If you want to leave a breadcrumb file, check the *Leave breadcrumb* box. You can enter text in the dialog box to indicate the location where the file was moved and the user who moved the file, and any other information, such as the reason the file was moved.
5. Click **Move Files**.

Note that you can also move an individual file when viewing the metadata details for a file. Just click **Move File**.

Delete source files

You can permanently remove source files that seem insecure or too risky to leave in your storage system, or that you've identified as a duplicate. This action is permanent and there is no undo or restore.

You can't delete files that reside in databases. All other data sources are supported.

Deleting files requires the following permissions:

- For NFS data - the export policy needs to be defined with write permissions.
- For CIFS data - the CIFS credentials need to have write permissions.
- For S3 data - the IAM role must include the following permission: `s3:DeleteObject`.

Delete source files manually

Requirements

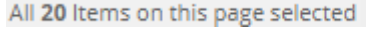
- You must have permissions to delete files. Learn about user access to compliance information.
- You can delete a maximum of 100,000 files at a time.

Steps

1. In the Data Investigation results pane, select the file, or files, that you want to delete.

   - To select individual files, check the box for each file (Volume_1).
   - To select all files on the current page, check the box in the title row (File Name).

- To select all files on all pages, check the box in the title row (

), and then in the pop-up message 

, click **Select all items in list (xxx items)**.

2. From the button bar, click **Delete**.
3. Because the delete operation is permanent, you must type "**permanently delete**" in the subsequent *Delete File* dialog and click **Delete File**.

You can view the progress of the delete operation in the Actions Status pane.

Note that you can also delete an individual file when viewing the metadata details for a file. Just click **Delete file**.

Add personal data identifiers to your BlueXP classification scans

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

To create a custom classification in version 1.43 and later, see Create a custom classification.

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

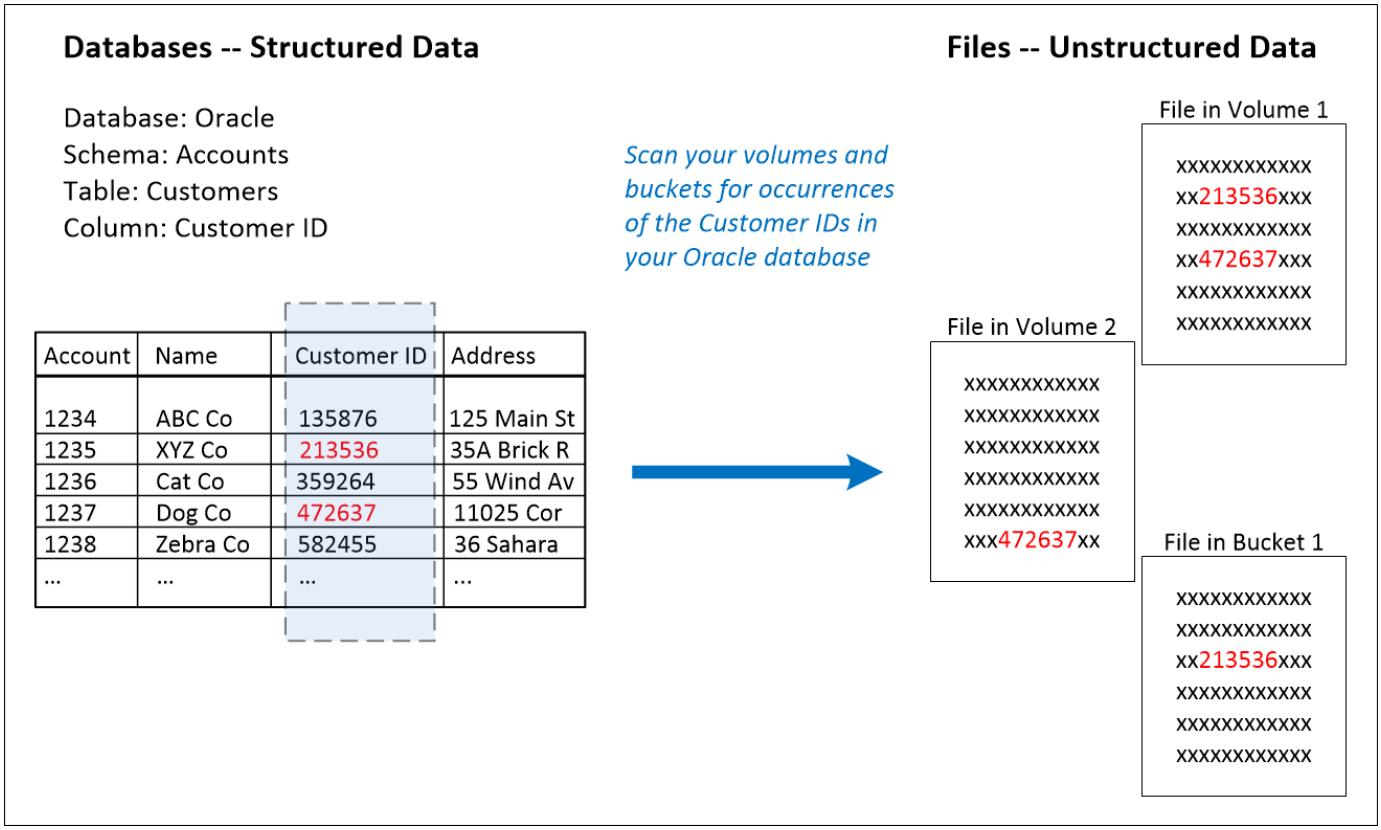
- You can add unique identifiers based on specific columns in databases you are scanning.
- You can add custom keywords from a text file — these words are identified within your data.
- You can add a personal pattern using a regular expression (regex) — the regex is added to the existing predefined patterns.
- You can add custom categories to identify where specific categories of information are found in your data.

All of these mechanisms to add custom scanning criteria are supported in all languages.

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

Add custom personal data identifiers from your databases

Data Fusion allows you to scan your organization's data to identify whether unique identifiers from your databases are found in any of your other data sources. You can choose the additional identifiers that BlueXP classification will look for in its scans by selecting a specific column or columns in a database table. For example, the diagram below shows how data fusion is used to scan your volumes, buckets, and databases for occurrences of all your Customer IDs from your Oracle database.



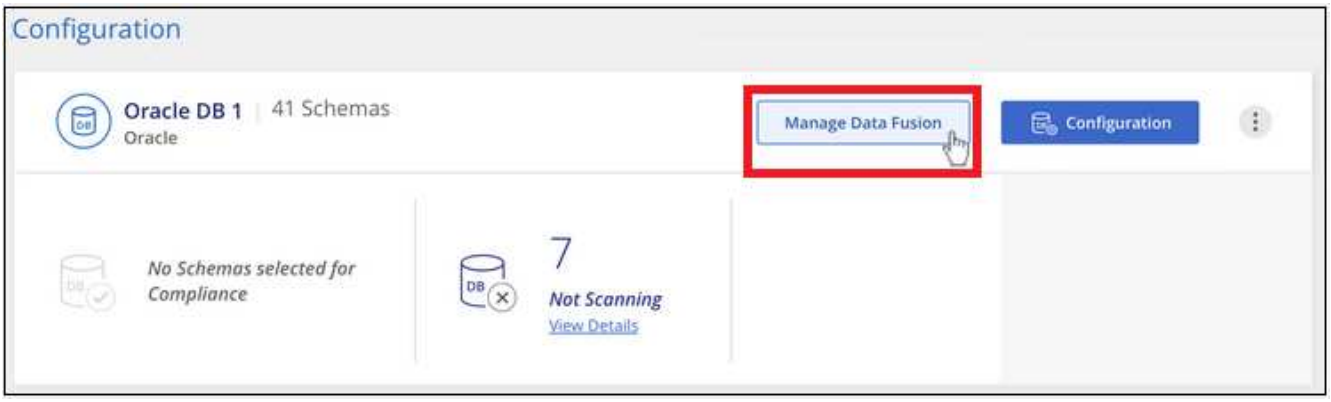
As you can see, two unique Customer IDs have been found in two volumes and in one S3 bucket. Any matches in database tables will also be identified.

Note that since you're scanning your own databases, whatever language your data is stored in will be used to identify data in future BlueXP classification scans.

Steps

You must have added at least one database server to BlueXP classification before you can add data fusion sources.

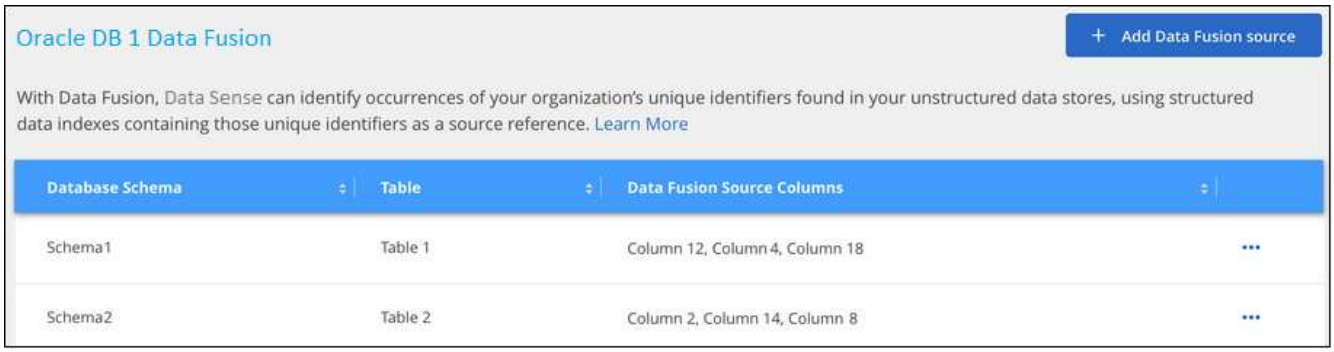
1. In the Configuration page, click **Manage Data Fusion** in the database where the source data resides.



2. Click **Add Data Fusion source** on the next page.
3. In the *Add Data Fusion Source* page:
   a. Select the Database Schema from the drop-down menu.
   b. Enter the Table name in that schema.
   c. Enter the Column, or Columns, that contain the unique identifiers you want to use.

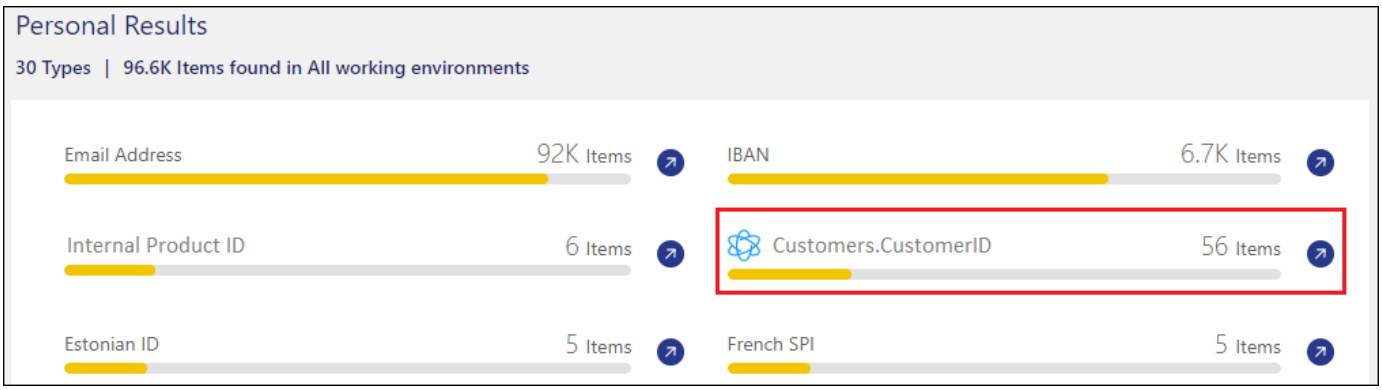
      When adding multiple columns, enter each column name, or table view name, on a separate line.

4. Click **Add Data Fusion Source**.



Results

After the next scan, the results will include this new information in the Compliance Dashboard under the "Personal Results" section, and in the Investigation page in the "Personal Data" filter. The name you used for the classifier appears in the filter list, for example `Customers.CustomerID`.



Delete a Data Fusion source

If at some point you decide not to scan your files using a certain Data Fusion source, you can select the source row from the Data Fusion inventory page and click **Delete Data Fusion Source**.

Add custom keywords from a list of words

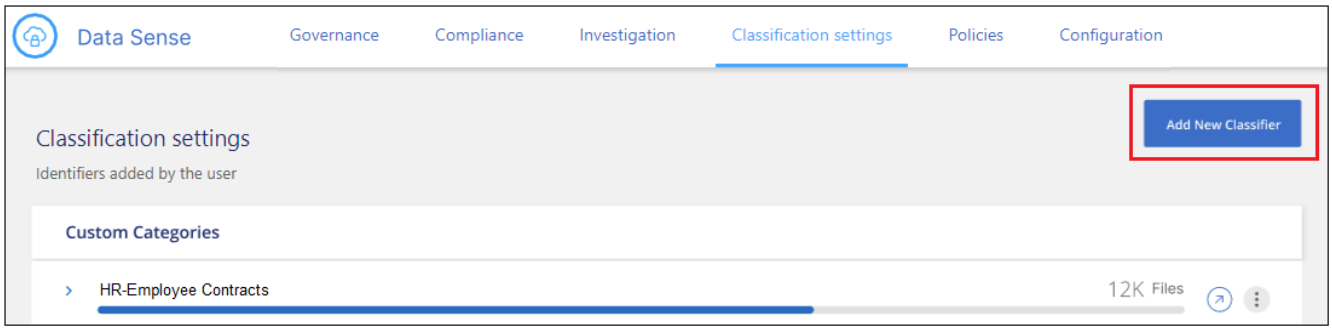
You can add custom keywords to BlueXP classification so that it will identify where that information is found in your data. You add the keywords just by entering each word you want BlueXP classification to recognize. The keywords are added to the existing predefined keywords that BlueXP classification already uses, and the results will be visible under the personal patterns section.

For example, you may want to see where internal Product Names are mentioned in all of your files to make sure these names are not accessible in locations that are not secure.

After updating the custom keywords, BlueXP classification will restart scanning all data sources. After the scan has completed, the new results will appear in the BlueXP classification Compliance Dashboard under the "Personal Results" section, and in the Investigation page in the "Personal Data" filter.

Steps

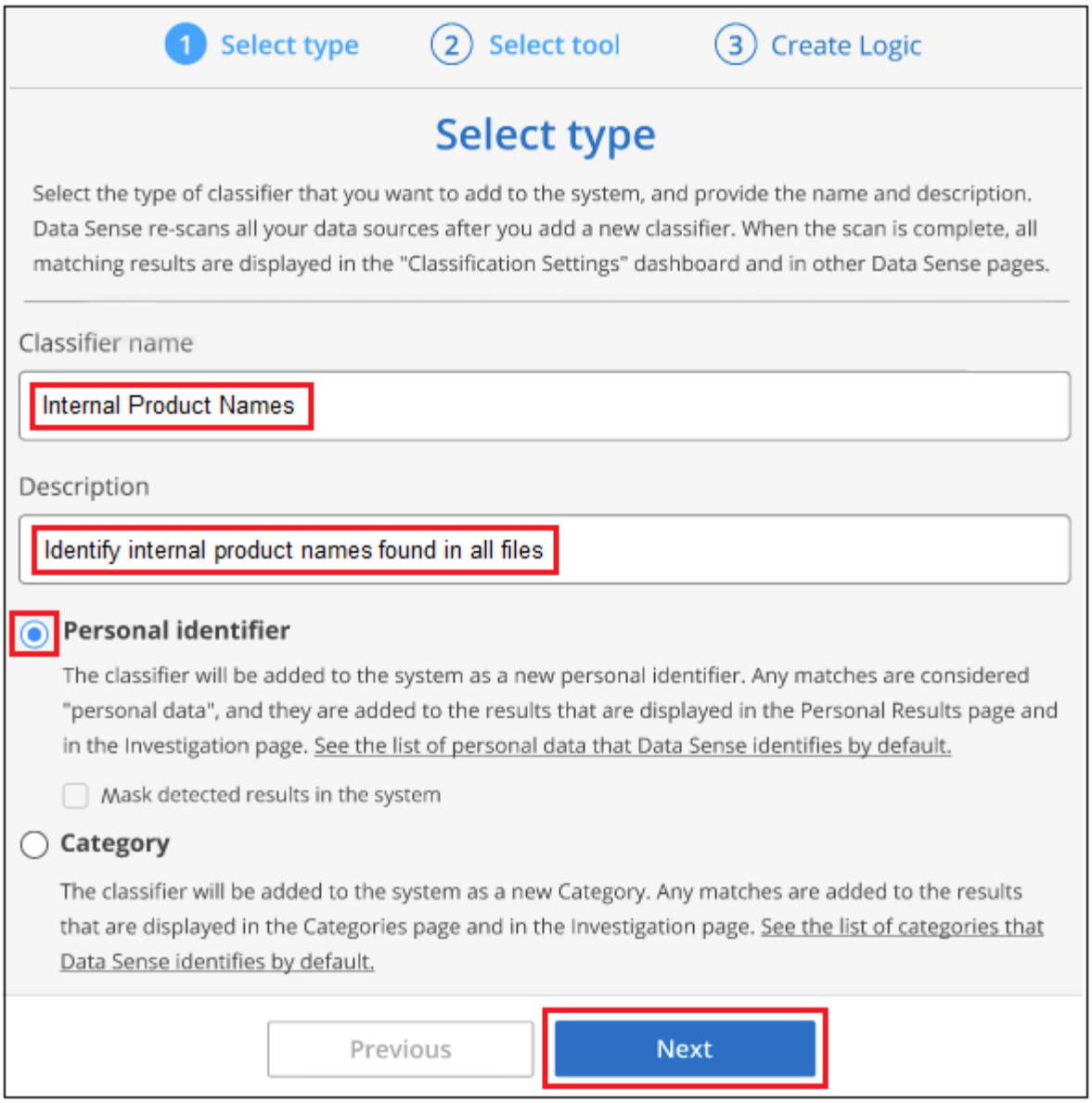
1. From the *Classification settings* tab, click **Add New Classifier** to launch the *Add Custom Classifier* wizard.



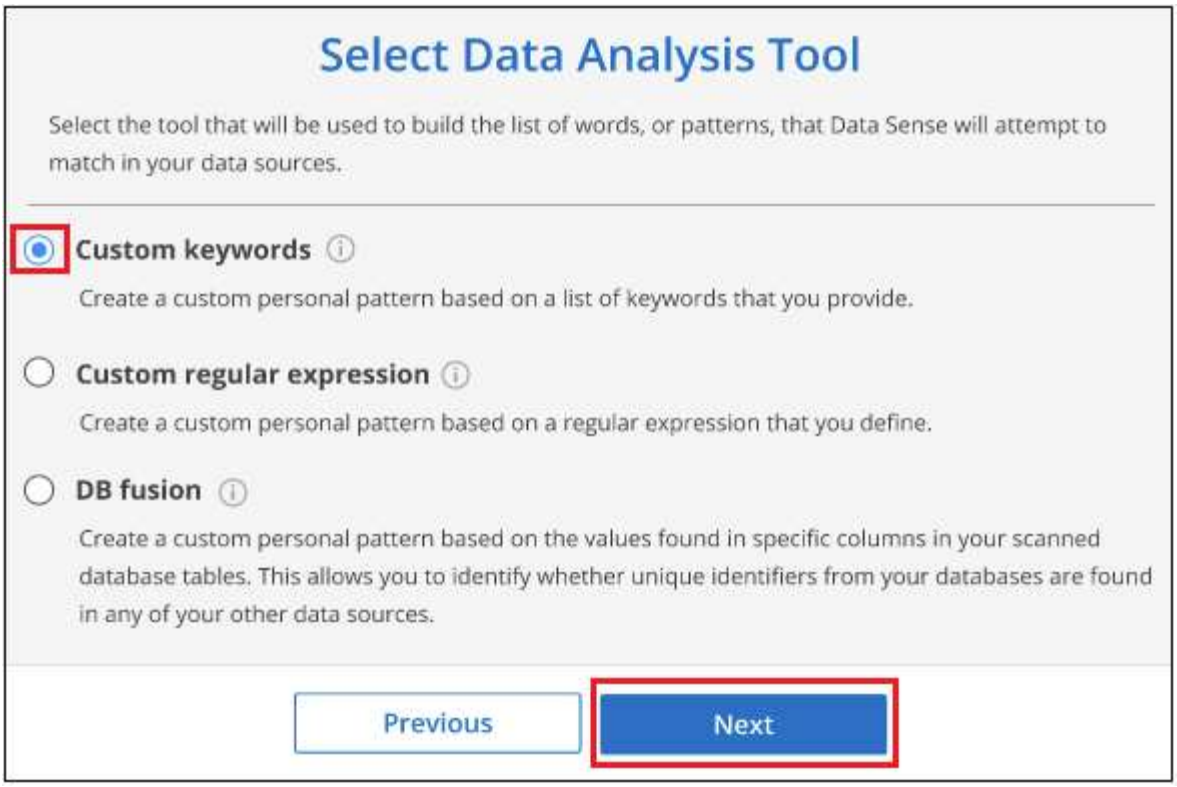
2. In the *Select type* page, enter the name of the classifier, provide a brief description, select **Personal identifier**, and then click **Next**.

   The name you enter will appear in the BlueXP classification UI as the heading for scanned files that match the classifier requirements, and as the name of the filter in the Investigation page.

   You can also check the box to "Mask detected results in the system" so the full result won't appear in the UI. For example, you may want to do this to hide full credit card numbers or similar personal data (the mask would appear in the UI like this: "**** **** ****" 3434).

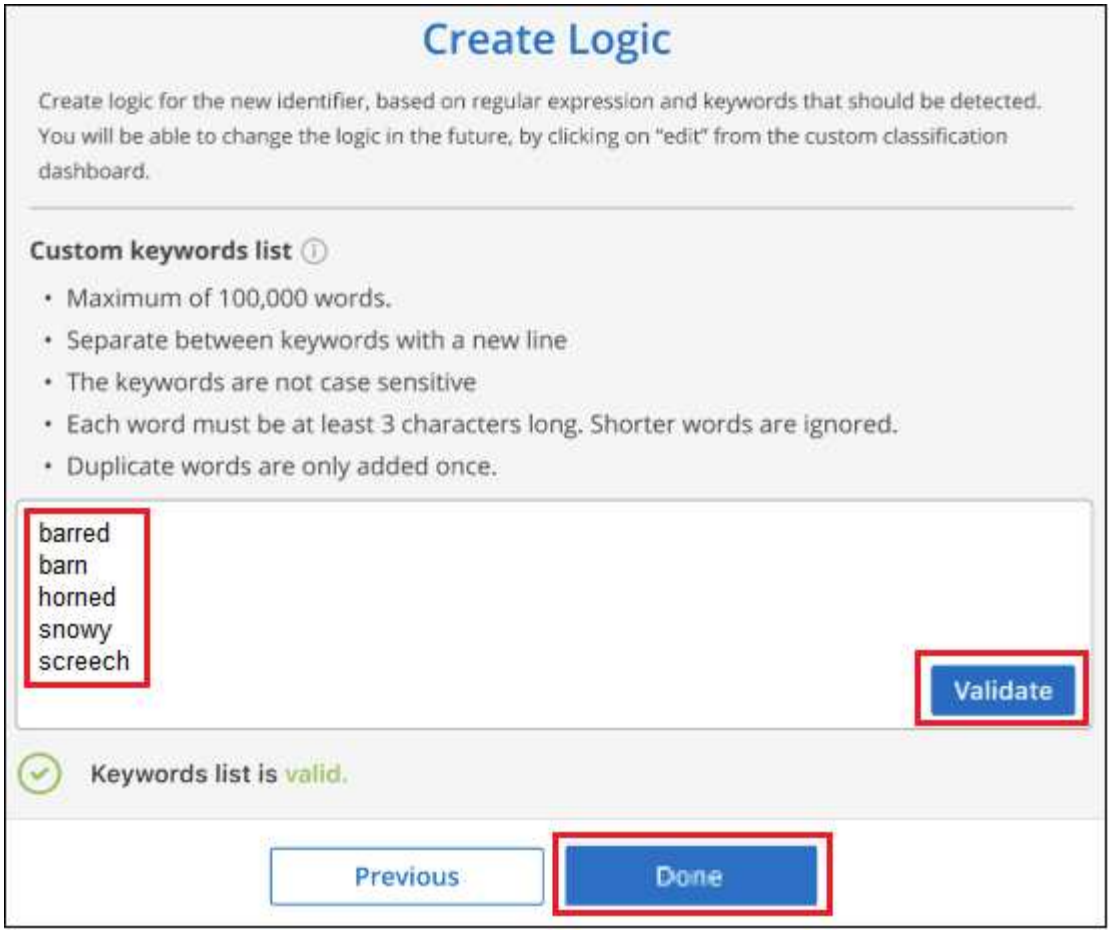


3. In the *Select Data Analysis Tool* page, select **Custom keywords** as the method you want to use to define the classifier, and then click **Next**.



4. In the *Create Logic* page, enter the keywords you want to recognize - each word on a separate line - and click **Validate**.

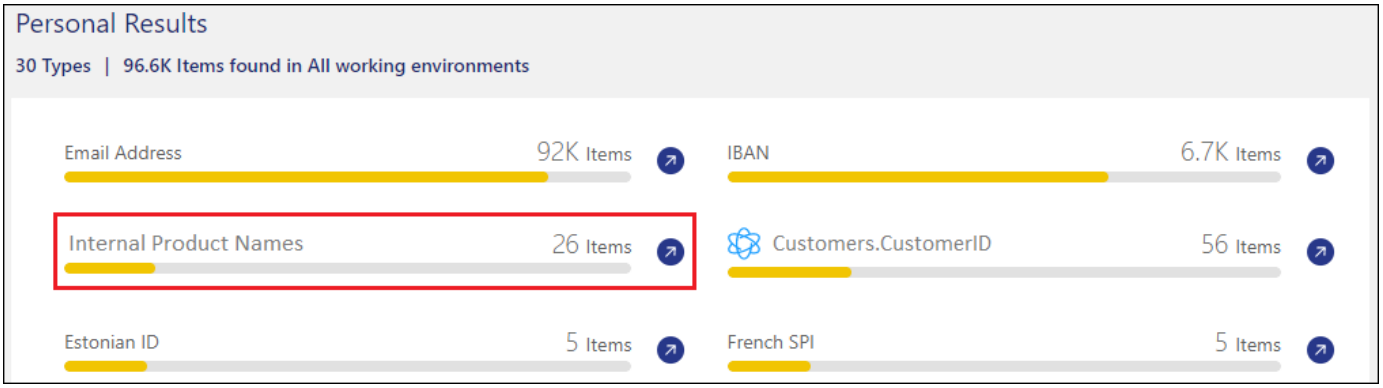
   The screenshot below shows internal Product Names (different types of owls). The BlueXP classification search for these items is not case sensitive.



5. Click **Done** and BlueXP classification starts to rescan your data.

Results

After the scan is complete, the results will include this new information in the Compliance Dashboard under the "Personal Results" section, and in the Investigation page in the "Personal Data" filter.



As you can see, the name of the classifier is used as the name in the Personal Results panel. In this manner you can activate many different groups of keywords and see the results for each group.

Add custom personal data identifiers using a regex

You can add a personal pattern to identify specific information in your data using a custom regular expression (regex). This allows you to create a new custom regex to identify new personal information elements that don't yet exist in the system. The regex is added to the existing predefined patterns that BlueXP classification

already uses, and the results will be visible under the personal patterns section.

For example, you may want to see where your internal Product IDs are mentioned in all of your files. If the Product ID has a clear structure, for example, it is a 12-digit number that starts with 201, you can use the custom regex feature to search for it in your files. The regular expression for this example is **\b201\d{9}\b**.

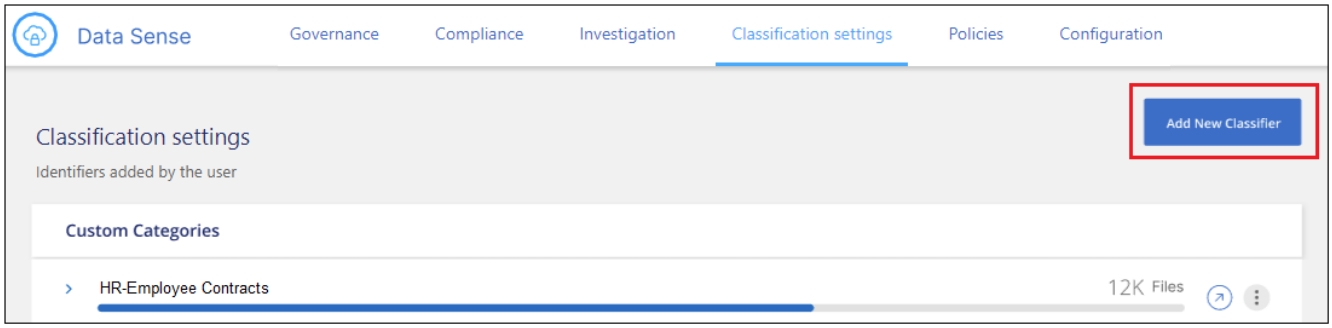
After adding the regex, BlueXP classification will restart scanning all data sources. After the scan has completed, the new results will appear in the BlueXP classification Compliance Dashboard under the "Personal Results" section, and in the Investigation page in the "Personal Data" filter.

If you need assistance in building the regular expression, refer to Regular expressions 101. Choose **Python** for the Flavor to see the types of results BlueXP classification will match from the regular expression. The Python Regex Tester page is also useful by displaying a graphical representation of your patterns.

BlueXP classification doesn't support pattern flags when creating a regex. This means you should not use "/".

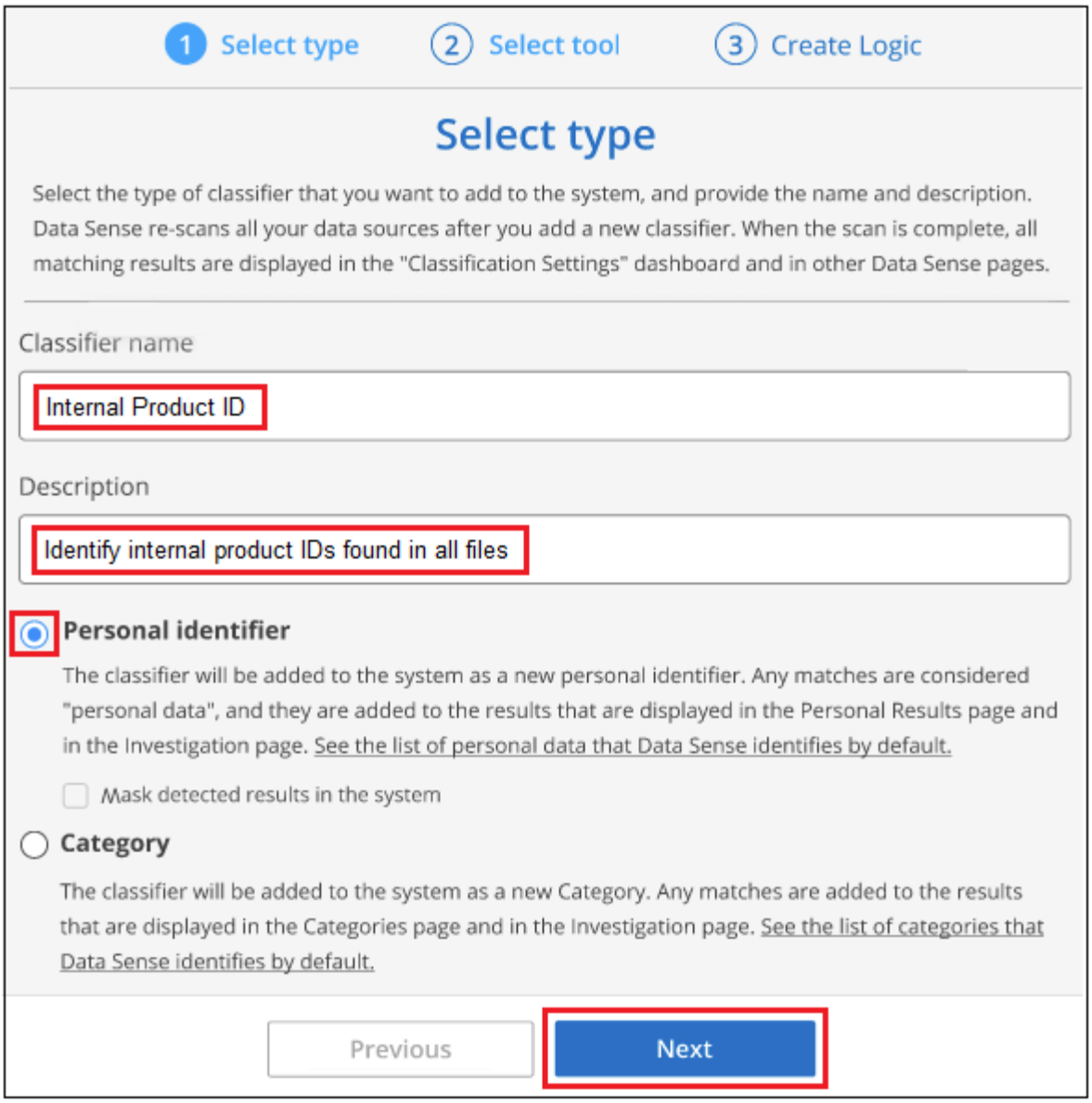
Steps

1. From the *Classification settings* tab, click **Add New Classifier** to launch the *Add Custom Classifier* wizard.

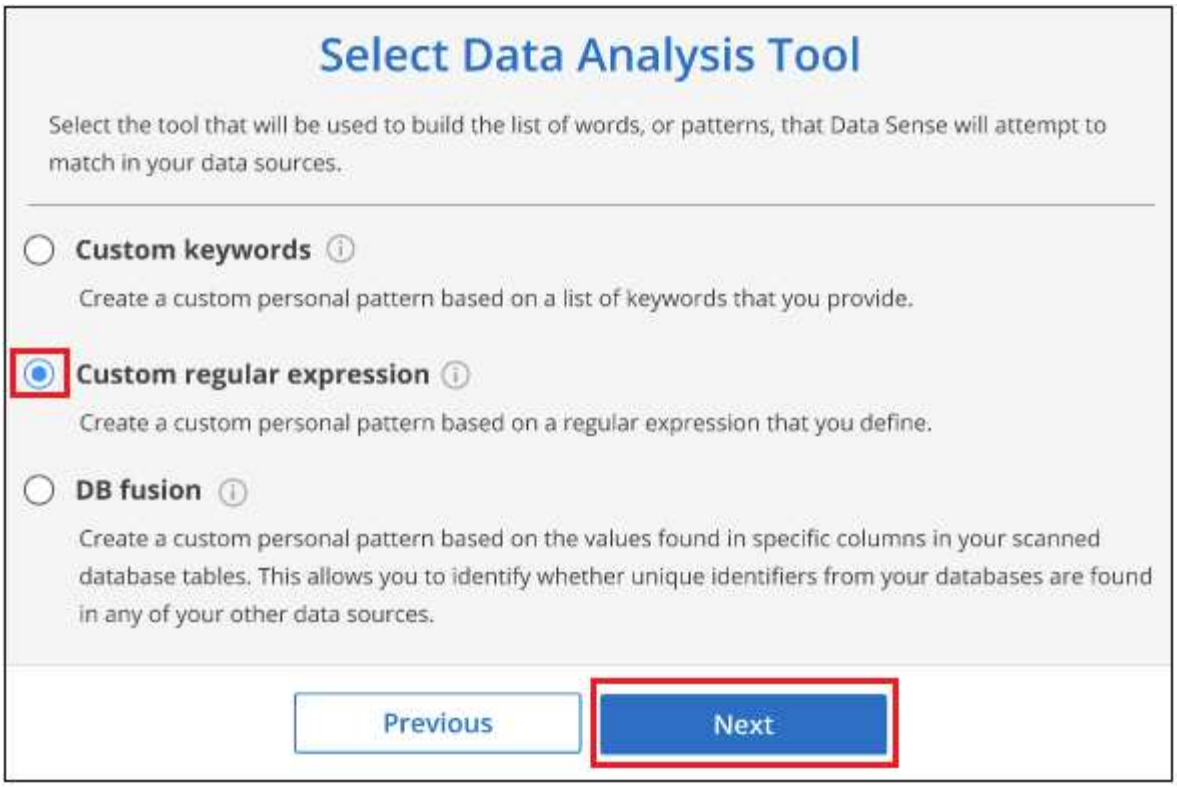


2. In the *Select type* page, enter the name of the classifier, provide a brief description, select **Personal identifier**, and then click **Next**.

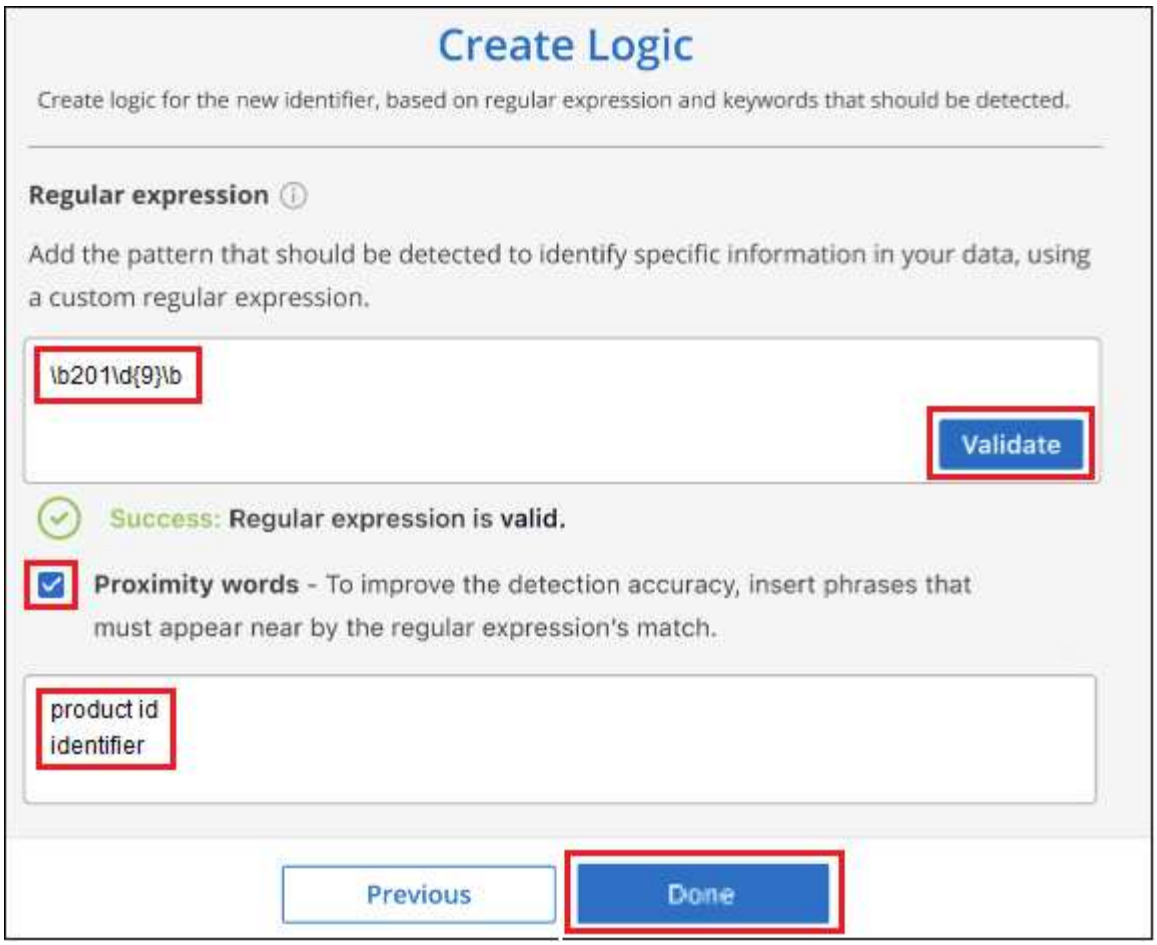
   The name you enter will appear in the BlueXP classification UI as the heading for scanned files that match the classifier requirements, and as the name of the filter in the Investigation page. You can also check the box to "Mask detected results in the system" so the full result won't appear in the UI. For example, you may want to do this to hide full credit card numbers or similar personal data.



3. In the *Select Data Analysis Tool* page, select **Custom regular expression** as the method you want to use to define the classifier, and then click **Next**.

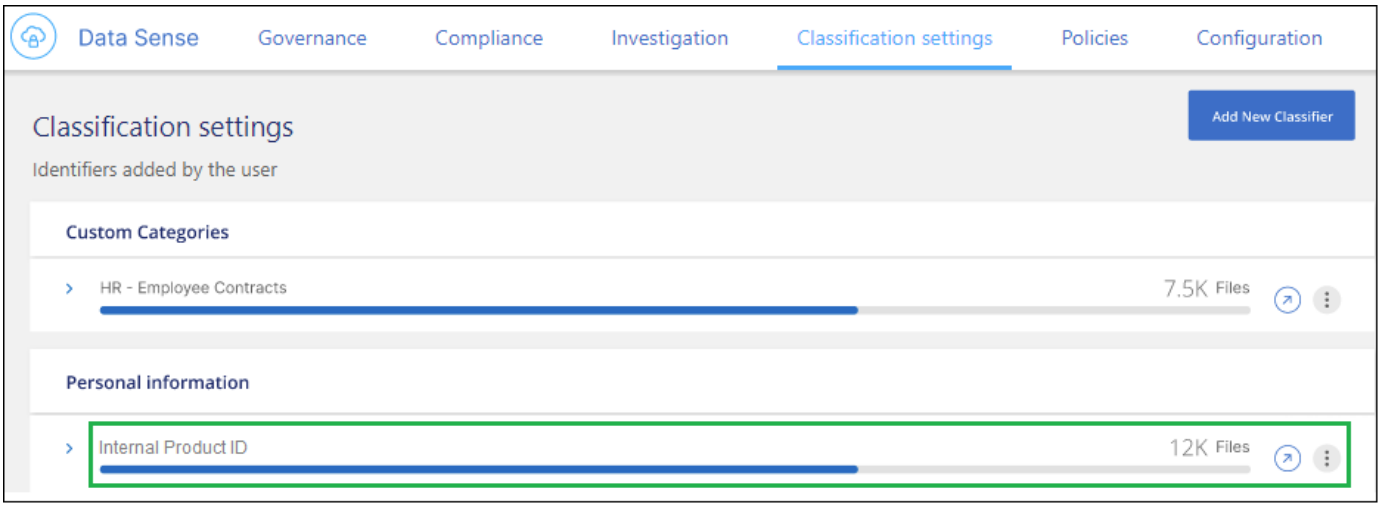


4. In the *Create Logic* page, enter the regular expression and any proximity words, and click **Done**.
   a. You can enter any legal regular expression. Click the **Validate** button to have BlueXP classification verify that the regular expression is valid, and that it is not too broad — meaning it will return too many results.
   b. Optionally, you can enter some proximity words to help refine the accuracy of the results. These are words that will typically be found within 300 characters of the pattern you are searching for (either before or after the found pattern). Enter each word, or phrase, on a separate line.



Results

The classifier is added and BlueXP classification starts to rescan all your data sources. You are returned to the Custom Classifiers page where you can view the number of files that have matched your new classifier. Results from scanning all of your data sources will take some time depending on the number of files that need to be scanned.



Add custom categories

BlueXP classification takes the data that it scans and divides it into different types of categories. Categories are topics based on artificial intelligence analysis of the content and metadata of each file. See the list of

predefined categories.

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

You can add custom categories to BlueXP classification so you can identify where categories of information that are unique for your data estate are found in your data. You add each category by creating "training" files that contain the categories of data that you want to identify, and then have BlueXP classification scan those files to "learn" through AI so that it can identify that data in your data sources. The categories are added to the existing predefined categories that BlueXP classification already identifies, and the results are visible under the Categories section.

For example, you may want to see where compressed installation files in .gz format are located in your files so that you can remove them, if necessary.

After updating the custom categories, BlueXP classification will restart scanning all data sources. After the scan has completed, the new results will appear in the BlueXP classification Compliance Dashboard under the "Categories" section, and in the Investigation page in the "Category" filter. See how to view files by categories.

Before you begin

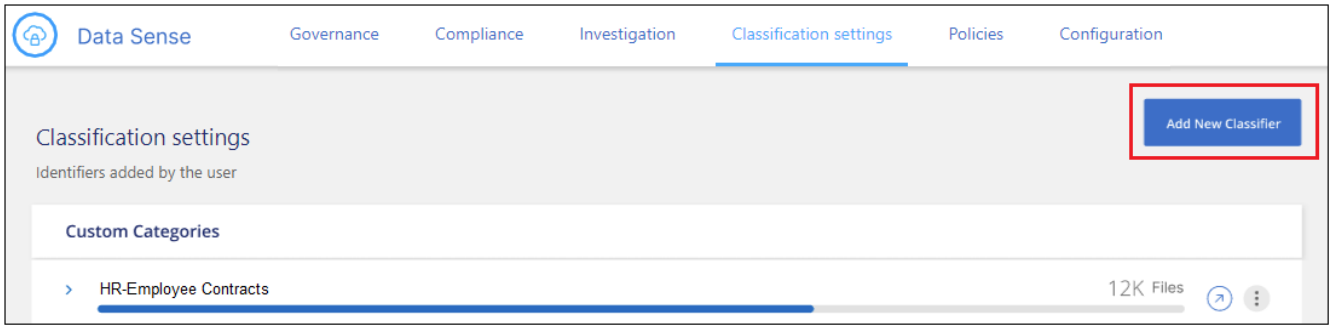
You'll need to create a minimum of 25 training files that contain samples of the categories of data that you want BlueXP classification to recognize. The following file types are supported:

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

The files must be a minimum of 100 bytes, and they must be located in a folder that is accessible by BlueXP classification.

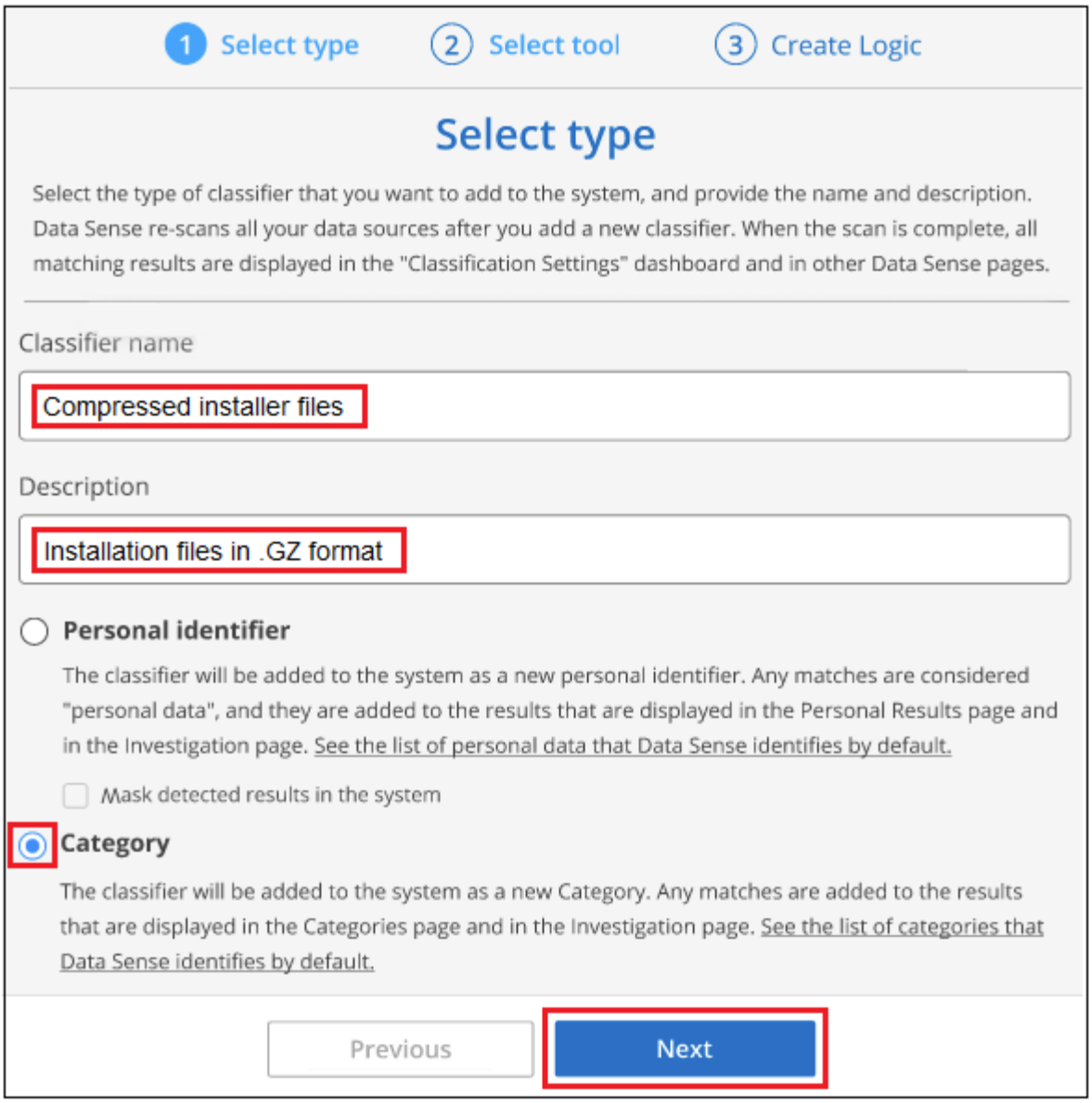
Steps

1. From the *Classification settings* tab, click **Add New Classifier** to launch the *Add Custom Classifier* wizard.

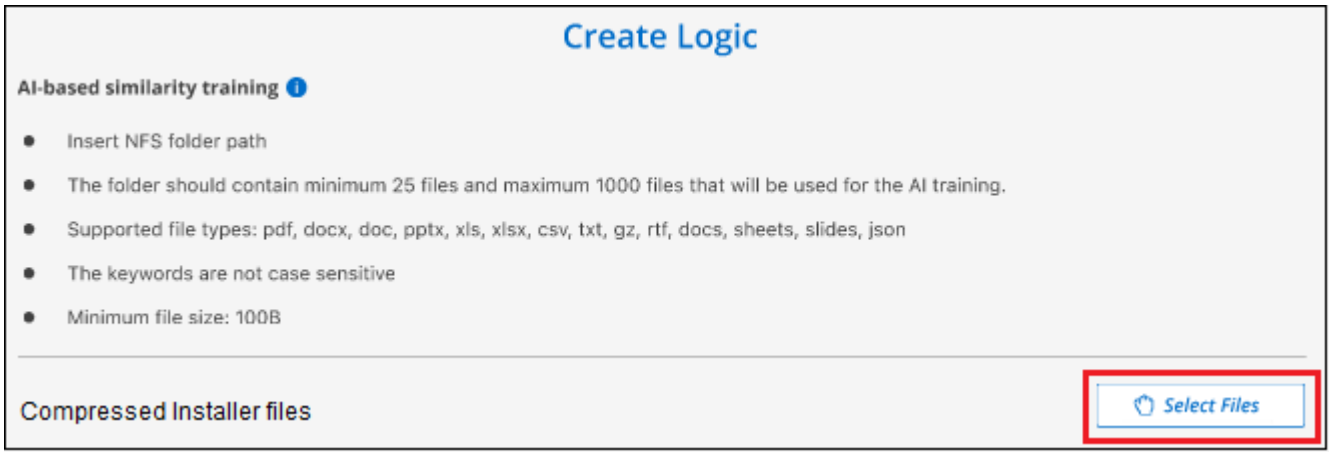


2. In the *Select type* page, enter the name of the classifier, provide a brief description, select **Category**, and then click **Next**.

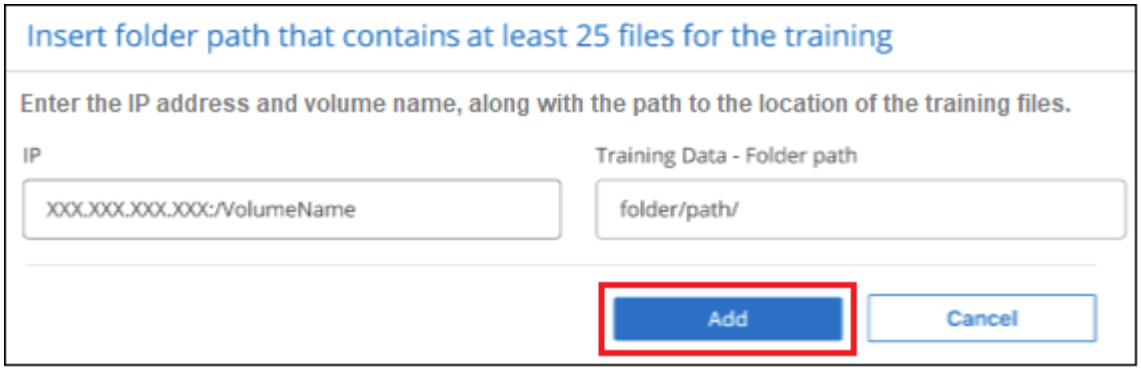
   The name you enter will appear in the BlueXP classification UI as the heading for scanned files that match the category of data you are defining, and as the name of the filter in the Investigation page.



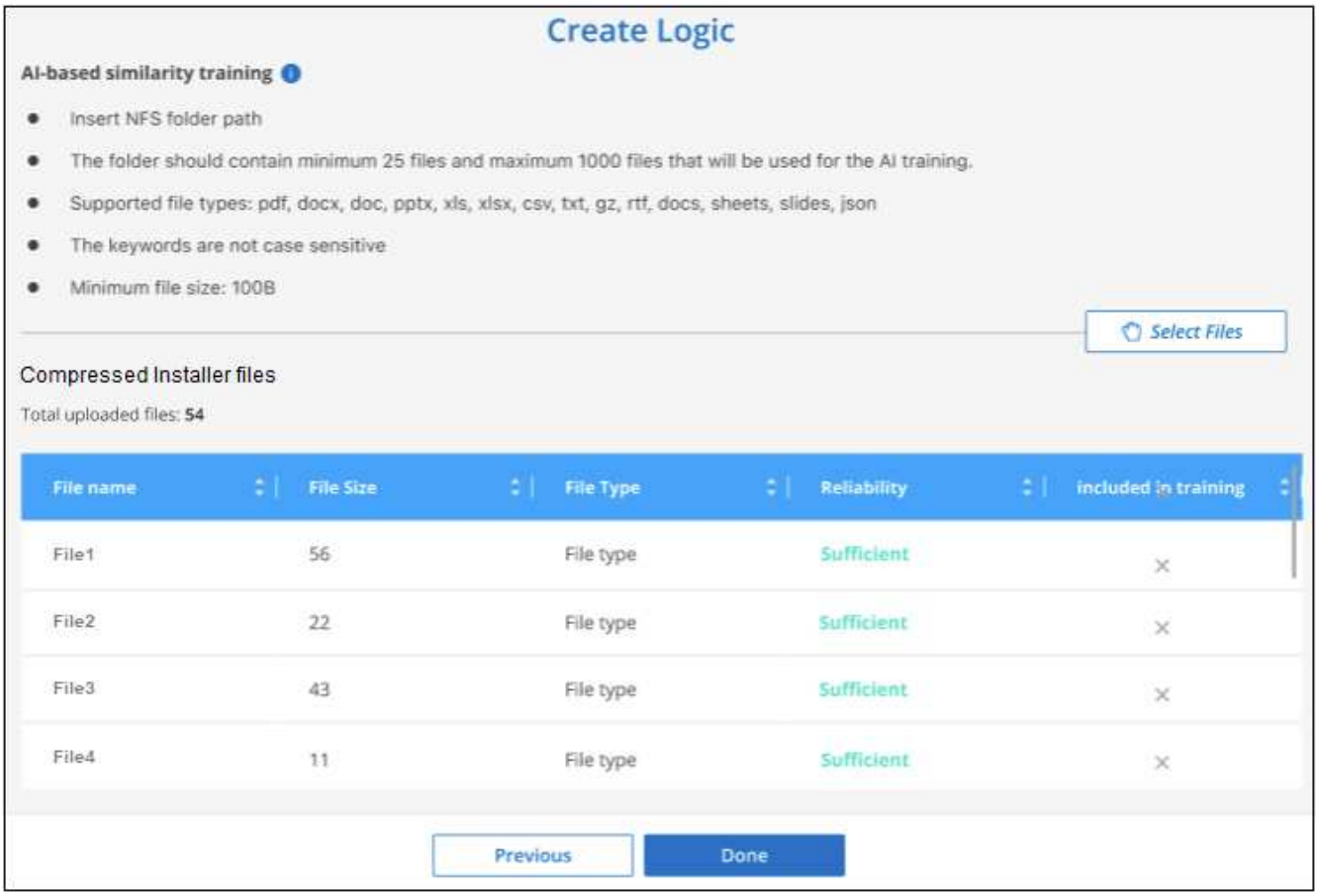
3. In the *Create Logic* page, make sure you have the learning files prepared, and then click **Select files**.



4. Enter the IP address of the volume, and the path where the training files are located, and click **Add**.



5. Verify that the training files were recognized by BlueXP classification. Click the **x** to remove any training files that do not meet the requirements. Then click **Done**.

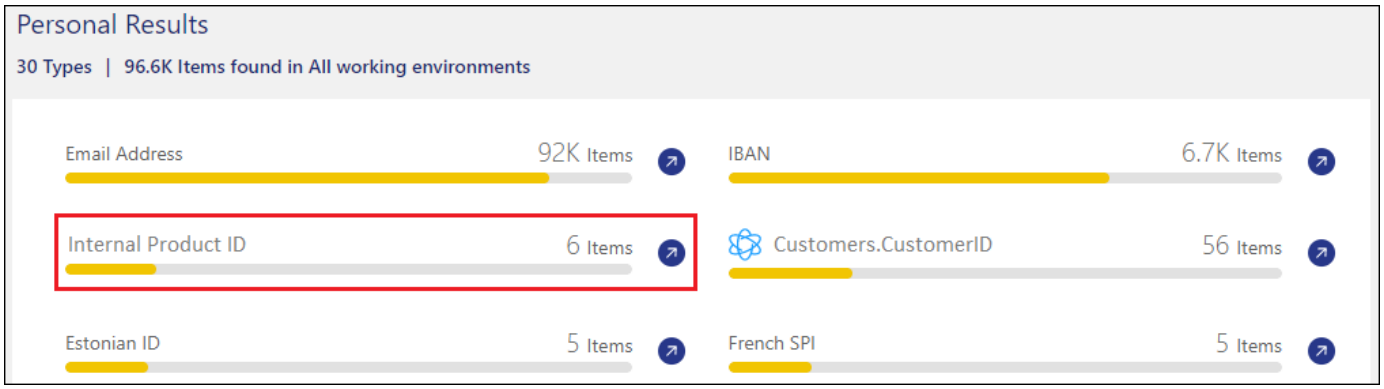


Results

The new category is created as defined by the training files and added to BlueXP classification. Then BlueXP classification starts to rescan all your data sources to identify files that fit into this new category. You are returned to the Custom Classifiers page where you can view the number of files that have matched your new category. Results from scanning all of your data sources will take some time depending on the number of files that need to be scanned.

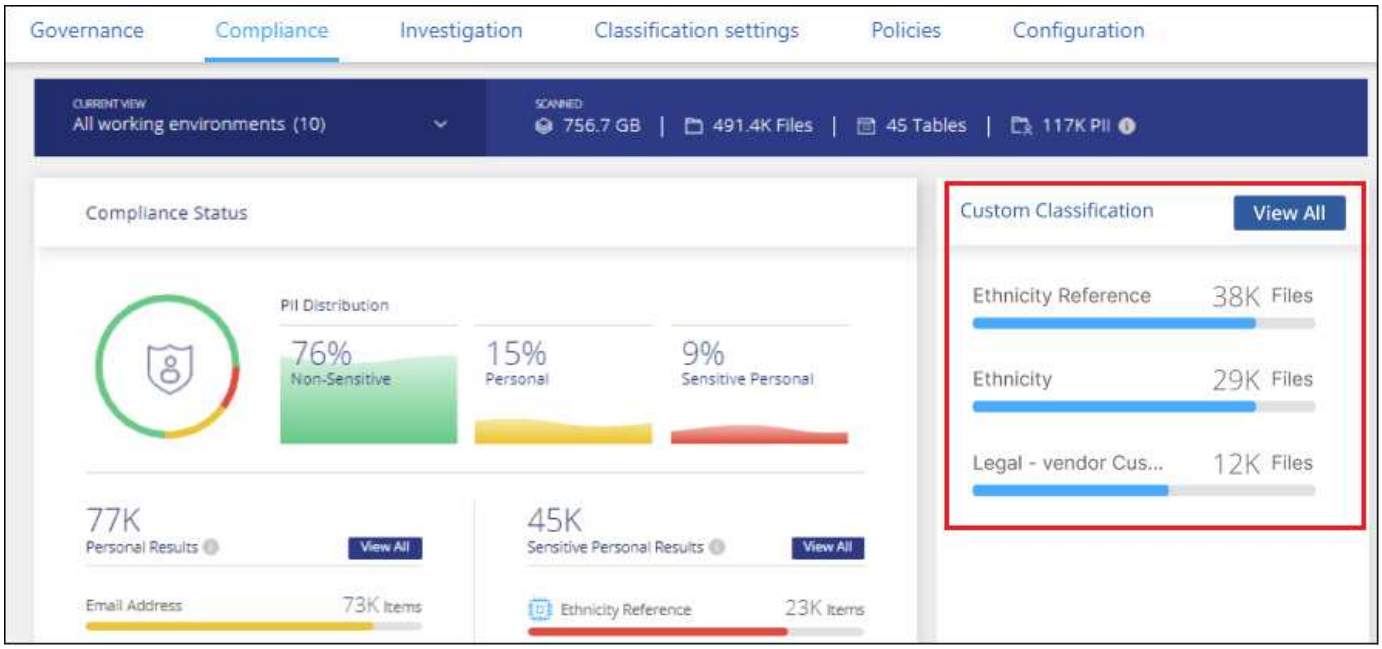
View results from your custom classifiers

You can view the results from any of your custom classifiers in the Compliance Dashboard and in the Investigation page. For example, this screenshot shows the matched information in the Compliance Dashboard under the "Personal Results" section.



Click the button to see the detailed results in the Investigation page.

Additionally, all of your custom classifier results appear in the Custom Classifiers tab, and the top 6 custom classifier results are displayed in the Compliance Dashboard, as shown below.

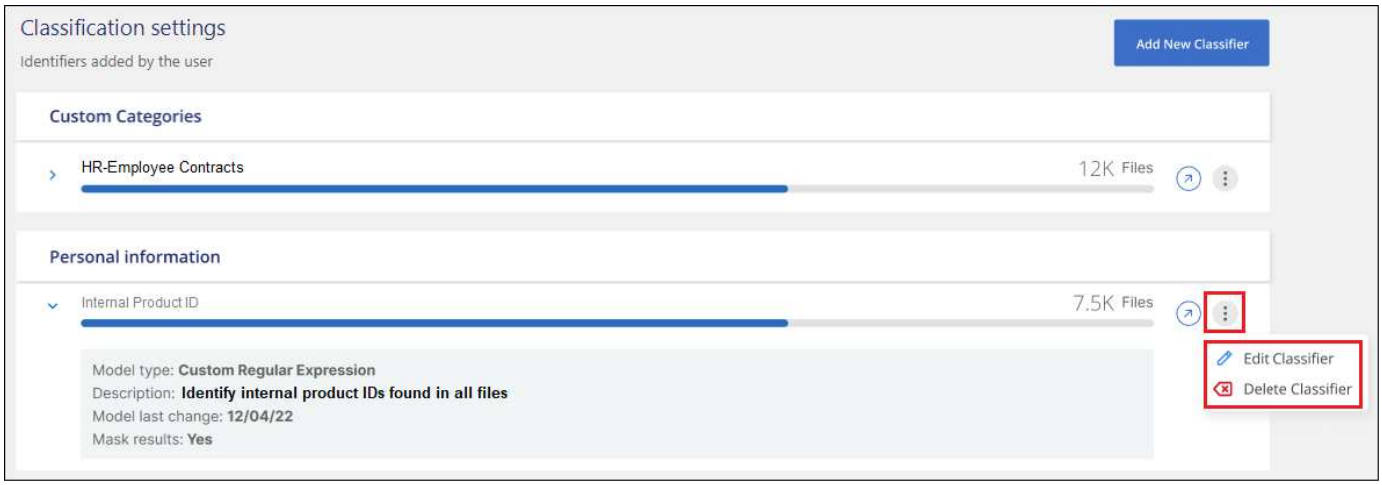


Manage custom classifiers

You can change any of the custom classifiers that you have created by using the **Edit Classifier** button.

You can’t edit Data Fusion classifiers at this time.

And if you decide at some later point that you don’t need BlueXP classification to identify the custom patterns that you added, you can use the **Delete Classifier** button to remove each item.



View the status of your compliance actions in BlueXP classification

When you run an asynchronous action from the Investigation Results pane across many files, for example, moving or deleting 100 files, the process can take some time. You can monitor the status of these actions in the *Action Status* pane so you'll know when it has been applied to all files.

This allows you to see the actions that have completed successfully, those currently in progress, and those that have failed so you can diagnose and fix any problems. Note that short operations that complete quickly, such as moving a single file, do not appear in the Actions Status pane.

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

The status can be:

- Success - A BlueXP classification action is finished and all items succeeded.
- Partial Success - A BlueXP classification action is finished and some items failed and some succeeded.
- In Progress - The action is still in progress.
- Queued - The action has not started.
- Canceled - The action has been canceled.
- Failed - The action failed.

Note that you can Cancel any actions that have the "Queued" or "In Progress" status.

Steps

1. In the bottom-right of the BlueXP classification UI you can see the **Actions Status** button .
2. Click this button and the most recent 20 actions are listed.

You can click the name of an action to view details corresponding to that operation.

Audit the history of BlueXP classification actions

BlueXP classification logs management activities that have been performed on files from all the working environments and data sources that BlueXP classification is scanning. BlueXP classification also logs the activities when deploying the BlueXP classification instance.

You can view the contents of the BlueXP classification audit log files, or download them, to see what file changes have occurred, and when. For example, you can see what request was issued, the time of the request, and details such as source location in case a file was deleted, or source and destination location in case a file was moved.

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

Log file contents

Each line in the audit log contains information in this format:

```
<full date> | <status> | ds_audit_logger | <module> | 0 | 0 | File <full file path> deleted from device <device path> - <result>
```

- Date and time - full timestamp for the event
- Status - INFO, WARNING
- Action type (delete, copy, move, create saved search, update saved search, rescan files, download JSON report, etc.)
- File name (if the action is relevant to a file)
- Details for the action - what was done: depends on the action
  - Saved search name
  - For move - Source and destination
  - For copy - Source and destination
  - For tag - tag name
  - For assign to - user name
  - For email alert - email address / account

For example, the following lines from the log file show a successful copy operation and a failed copy operation.

```
2022-06-06 15:23:08,910 | INFO | ds_audit_logger | es_scanned_file | 237 | 49 | Copy file /CIFS_share/data/dop1/random_positives.tsv from device 10.31.133.183 (type: SMB_SHARE) to device 10.31.130.133:/export_reports (NFS_SHARE) - SUCCESS
2022-06-06 15:23:08,968 | WARNING | ds_audit_logger | es_scanned_file | 239 | 153 | Copy file /CIFS_share/data/compliance-netapp.tar.gz from device 10.31.133.183 (type: SMB_SHARE) to device 10.31.130.133:/export_reports (NFS_SHARE) - FAILURE
```

Log file locations

The management audit log files are located on the BlueXP classification machine in: `/opt/netapp/audit_logs/`

The installation audit log files are written to `/opt/netapp/install_logs/`

Each log file can be a maximum of 10 MB in size. When that limit is reached, a new log file is started. The log files are named "DataSense_audit.log", "DataSense_audit.log.1", "DataSense_audit.log.2", and so on. A maximum of 100 log files are retained on the system - older log files are deleted automatically after the maximum has been reached.

Access the log files

You'll need to log into the BlueXP classification system to access the log files. See how to log in to the BlueXP classification system depending on whether you manually installed the software on a Linux machine or if you deployed the instance in the cloud.

Reducing the BlueXP classification scan speed

Data scans have a negligible impact on your storage systems and on your data. However, if you are concerned with even a very small impact, you can configure BlueXP classification to perform "slow" scans.

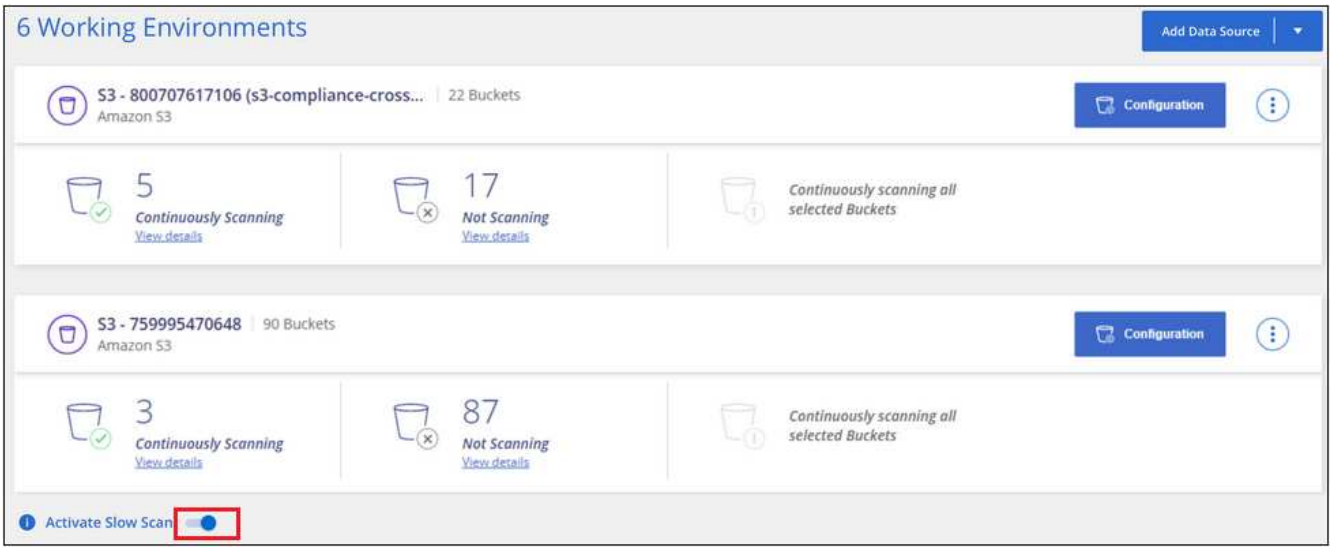
When enabled, slow scanning is used on all data sources - you can't configure slow scanning for a single working environment or data source.

The scan speed can't be reduced when scanning databases.

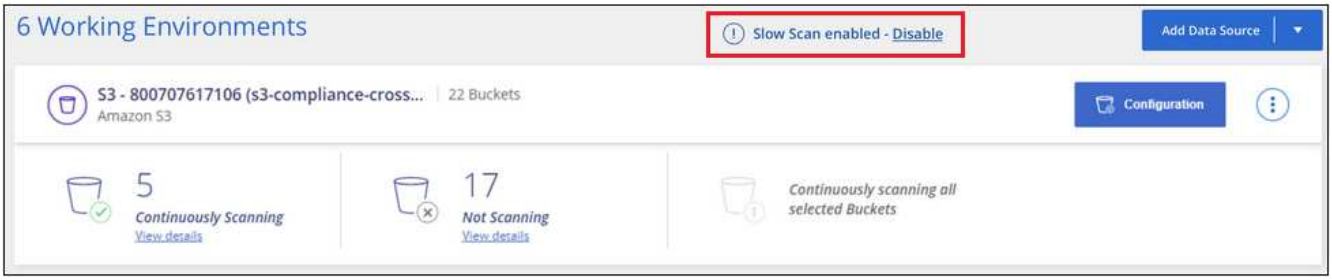
This information is relevant only for BlueXP classification legacy versions 1.30 and earlier.

Steps

1. From the bottom of the *Configuration* page, move the slider to the right to activate slow scanning.



The top of the Configuration page indicates that slow scanning is enabled.



2. You can disable slow scanning by clicking **Disable** from this message.

Remove a OneDrive, SharePoint, or Google Drive account from BlueXP classification

If you no longer want to scan user files from a certain OneDrive account, from a specific SharePoint account, or from a Google Drive account, you can delete the account from the BlueXP classification interface and stop all scans.

Steps

1. From the *Configuration* page, select the button in the row for the OneDrive, SharePoint, or Google Drive account, then select **Remove OneDrive Account**, **Remove SharePoint Account**, or **Remove Google Drive account**.

2. Click **Delete Account** from the confirmation dialog.

Copyright information

Trademark information