



Manage data deprecations

BlueXP classification

NetApp
June 20, 2024

Table of Contents

- Manage data deprecations 1
 - View governance details about your data using the Governance dashboard 1
 - Organize your private data 2
 - Manage your private data 11
 - Add personal data identifiers to your BlueXP classification scans 21
 - Viewing the status of your compliance actions 36
 - Audit the history of BlueXP classification actions 37
 - Reducing the BlueXP classification scan speed 38
 - Remove a OneDrive, SharePoint, or Google Drive account from BlueXP classification 39

Manage data deprecations

View governance details about your data using the Governance dashboard

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

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

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

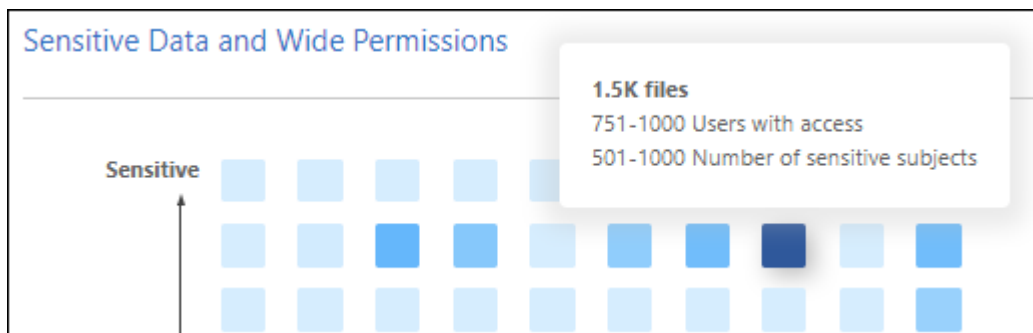
Data listed by sensitivity and wide permissions on the Governance dashboard

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



This applies to BlueXP classification versions 1.30 and earlier.

Files are rated based on the number of users with permission to access the files on the X axis (lowest to highest), and the number of sensitive identifiers within the files on the Y axis (lowest to highest). The blocks represent the number of files that match the items from the X and Y axes. The lighter colored blocks are good; with fewer users able to access the files, and with fewer sensitive identifiers per file. The darker blocks are the items you may want to investigate. For example, the screen below shows the tooltip text for the dark blue block. It shows that you have 1,500 files where 751-1000 users have access, and where there are 501-1000 sensitive identifiers per file.



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

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



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

Classification area on the dashboard showing AIP labels

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

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

See [AIP Labels](#) for more information.

Organize your private data

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

NOTE This information is relevant only for BlueXP classification legacy versions 1.30 and earlier. The December 2023 (v1.26.6) release removed the option to integrate data using Azure Information Protection (AIP) labels.

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



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

Should I use tags or labels?

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

| Tags | Labels |
|--|--|
| File tags are an integrated part of BlueXP classification. | Requires that you have subscribed to Azure Information Protection (AIP). |

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

Categorize your data using AIP labels

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

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



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

Integrate AIP labels in your workspace

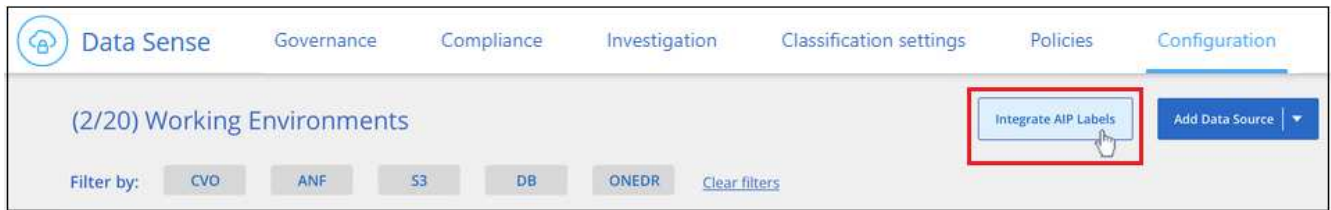
Before you can manage AIP labels, you need to integrate the AIP label functionality into BlueXP classification by signing into your existing Azure account. Once enabled, you can manage AIP labels within files for all [data sources](#) in your BlueXP workspace.

Requirements

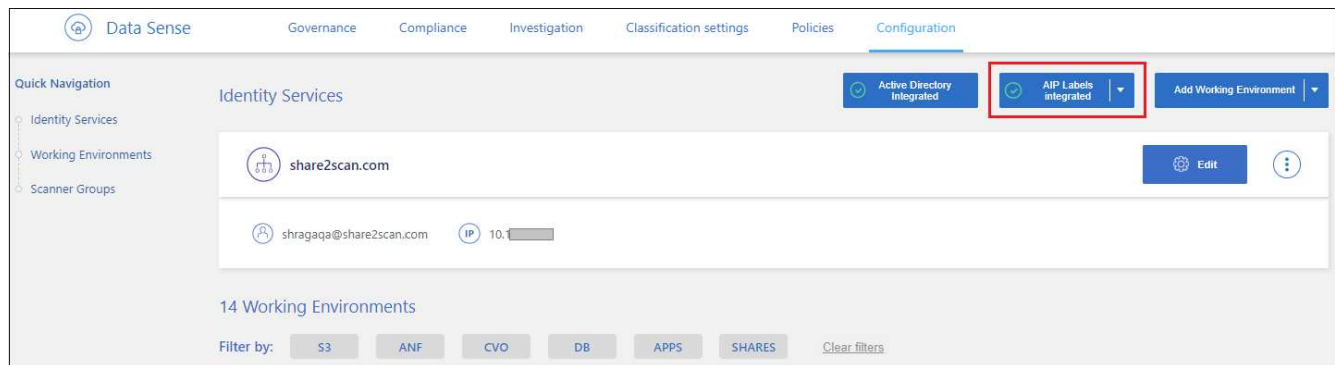
- You must have an account and an Azure Information Protection license.
- You must have the login credentials for the Azure account.
- If you plan to change labels in files that reside in Amazon S3 buckets, ensure that the permission `s3:PutObject` is included in the IAM role. See [setting up the IAM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

Steps

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



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



Result

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

View AIP labels in your files

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

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



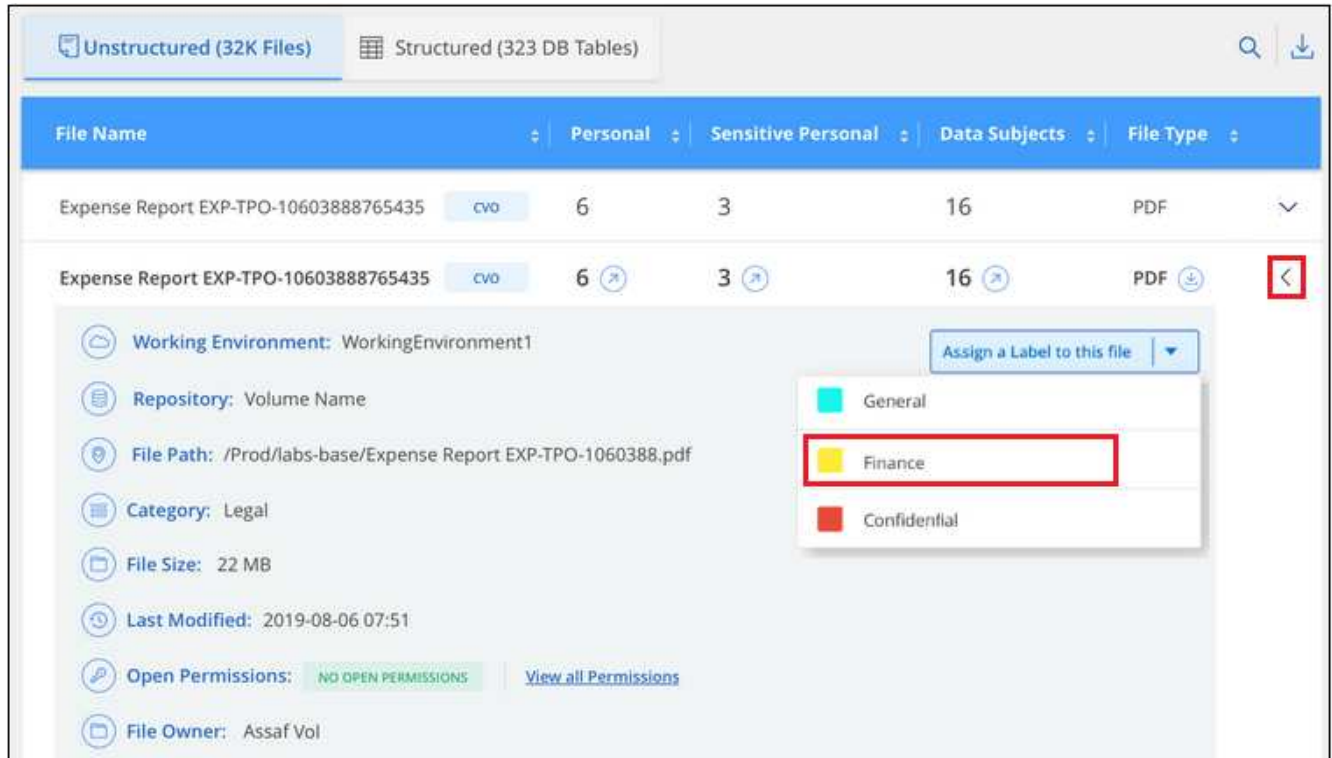
Assign AIP labels manually

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

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

Steps

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



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

The label appears in the file metadata.

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

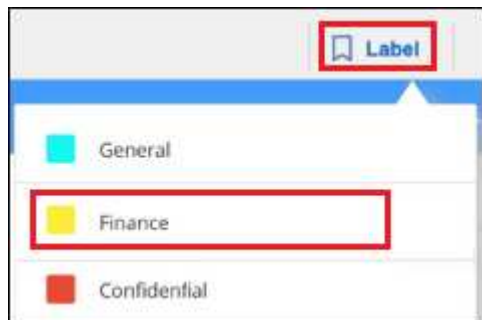
Steps

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



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

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



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

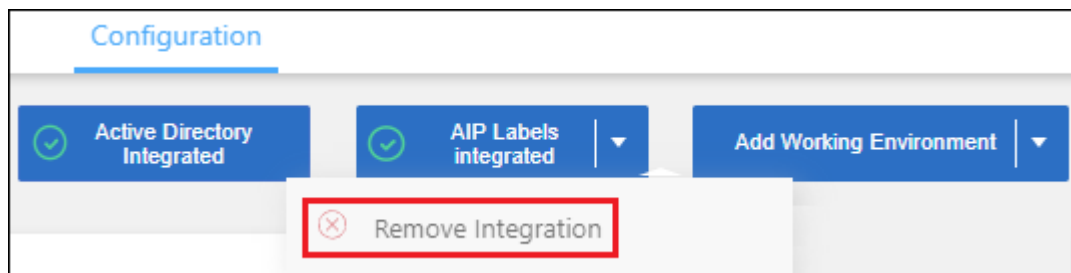
Remove the AIP integration

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

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

Steps

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



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

Apply tags to manage your scanned files

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

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

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

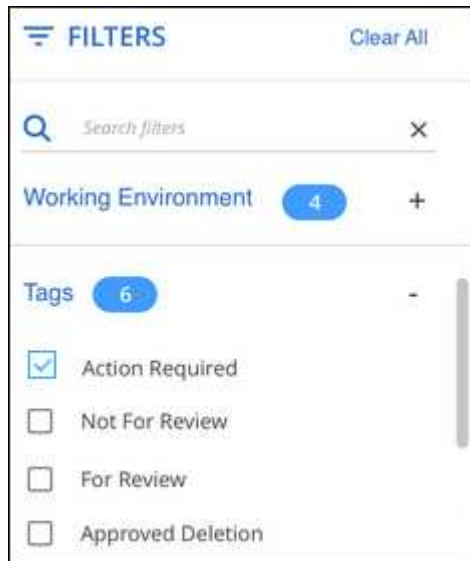


Tags assigned to files in BlueXP classification are not related to the tags you can add to resources, such as volumes or virtual machine instances. BlueXP classification tags are applied at the file level.

View files that have certain tags applied

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

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




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

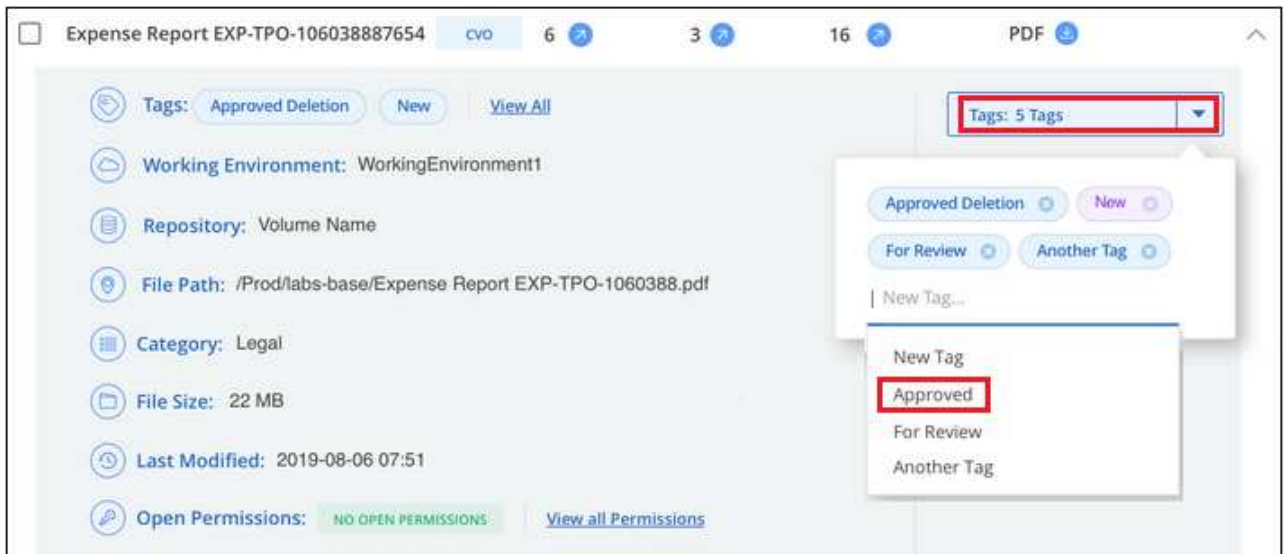
Assign tags to files

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

To add a tag to a single file:

Steps

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



The tag appears in the file metadata.

To add a tag to multiple files:

Steps

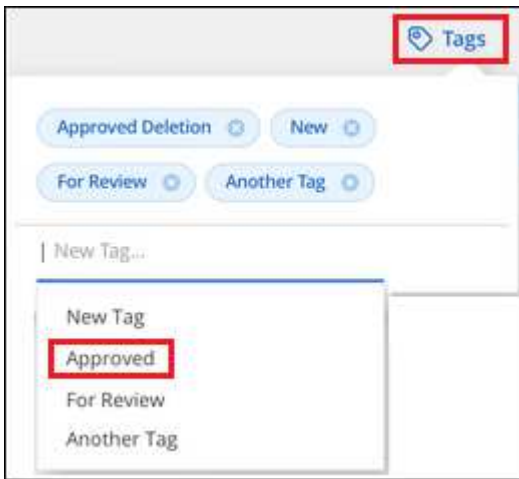
1. In the Data Investigation results pane, select the file, or files, that you want to tag.

| 255 items 1.2 GB 2 Selected 3 MB | | | | | | | Tags | Assign to | Label | Copy | Move | Delete |
|-------------------------------------|-------------------------------------|----------|--------------------|---------------|-----------|-----|------|-----------|-------|------|------|--------|
| <input type="checkbox"/> | File Name | Personal | Sensitive Personal | Data Subjects | File Type | | | | | | | |
| <input checked="" type="checkbox"/> | Expense Report EXP-TPO-106038887654 | cvo | 6 | 3 | 16 | PDF | | | | | | |
| <input checked="" type="checkbox"/> | Expense Report EXP-TPO-106038887654 | cvo | 6 | 3 | 6 | PDF | | | | | | |
| <input type="checkbox"/> | Expense Report EXP-TPO-106038887654 | cvo | 6 | 3 | 6 | PDF | | | | | | |
| <input type="checkbox"/> | Expense Report EXP-TPO-106038887654 | cvo | 6 | 3 | 6 | PDF | | | | | | |

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

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

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



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

Delete tags from files

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

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



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

Assign users to manage certain files

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


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

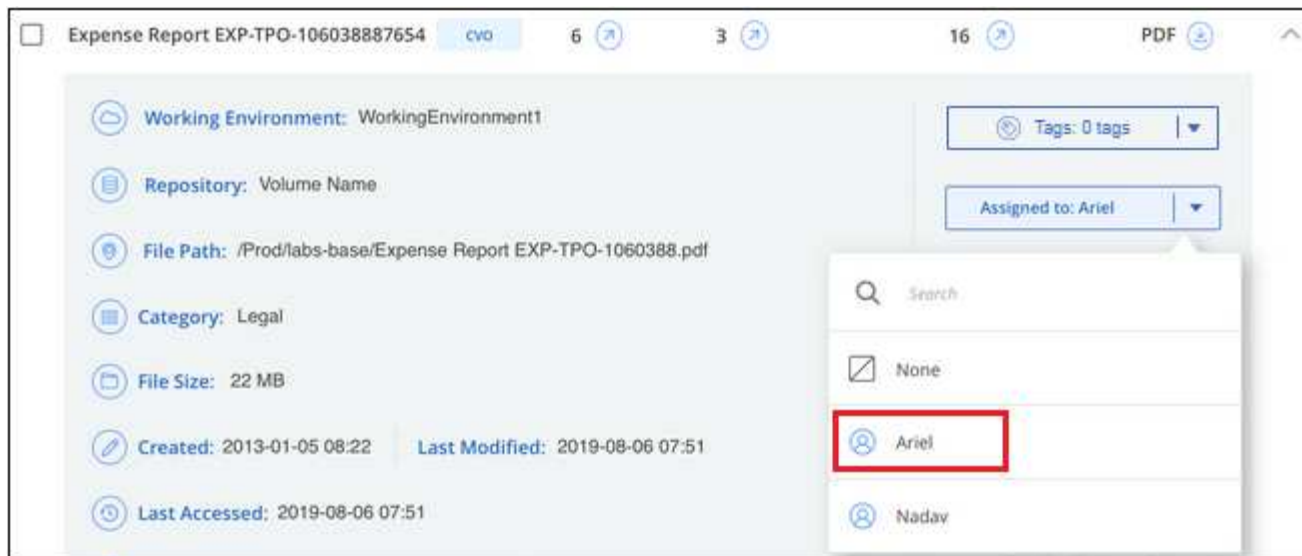
Note that the user name is not added to the file as part of the file metadata - it is just seen by BlueXP users when using BlueXP classification.

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

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

Steps

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

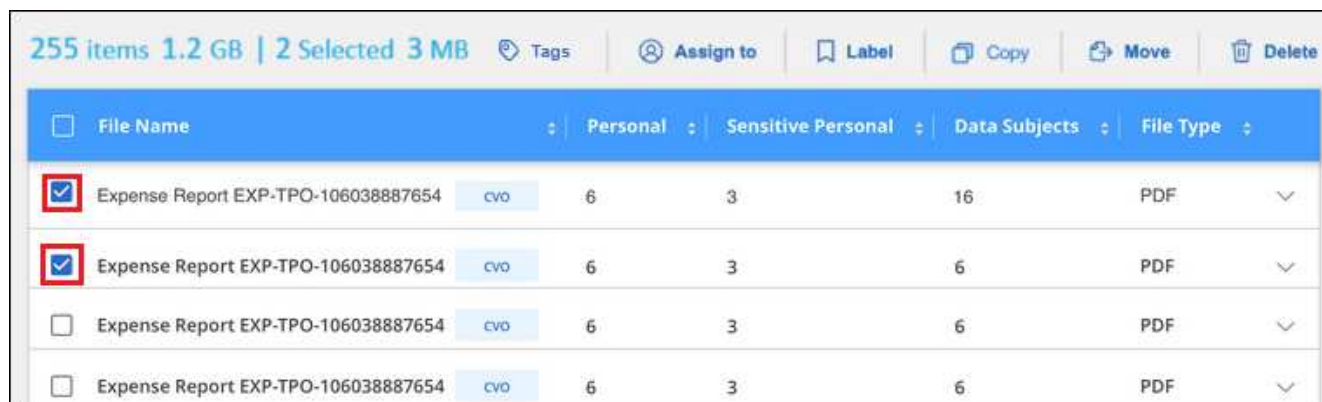


The User name appears in the file metadata.

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

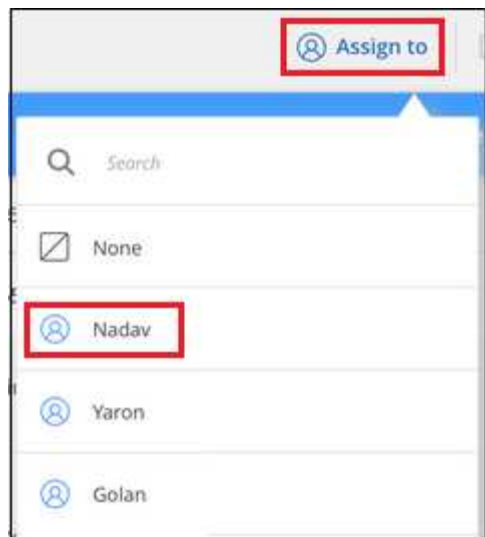
Steps

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



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

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



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

Manage your private data

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

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

- You can copy files to a destination NFS share if you want to make a copy of certain data and move it to a different NFS location.
- You can clone an ONTAP volume to a new volume, while including only selected files from the source volume in the new cloned volume. This is useful for situations where you're migrating data and you want to exclude certain files from the original volume.
- You can copy and synchronize files from a source repository to a directory in a specific destination location. This is useful for situations where you're migrating data from one source system to another while there is still some final activity on the source files.
- You can move source files that BlueXP classification is scanning to any NFS share.
- You can delete files that seem insecure or too risky to leave in your storage system, or that you have identified as duplicate.



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

Copy source files

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

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

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

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

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

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

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

Copy source files to an NFS share

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



You can't copy files that reside in databases.

Requirements

- You must have the Account Admin or Workspace Admin role to copy files.
- Copying files requires that the destination NFS share allows access from the BlueXP classification instance.
- You can copy between 1 and 100,000 files at a time.

Steps

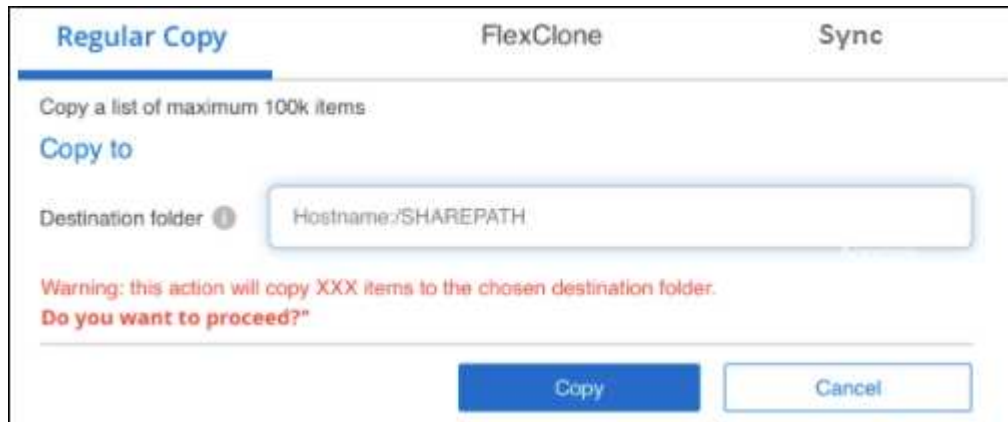
1. In the Data Investigation results pane, select the file, or files, that you want to copy, and click **Copy**.



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

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

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



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

A dialog appears with the status of the copy operation.

You can view the progress of the copy operation in the [Actions Status](#) pane.

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



Clone volume data to a new volume

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

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

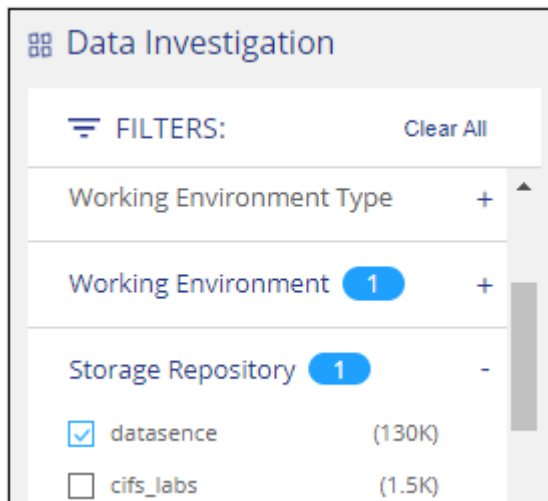
Note: FlexGroup volumes can't be cloned because they're not supported by FlexClone.

Requirements

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

Steps

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



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

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



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

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

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

A dialog appears with the status of the clone operation.

Result

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

You can view the progress of the clone operation in the [Actions Status pane](#).

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

Copy and synchronize source files to a target system

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

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



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

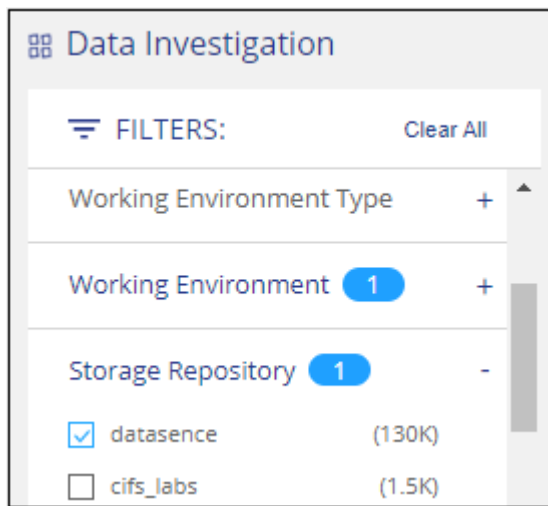
Requirements

- You must have the Account Admin or Workspace Admin role to copy and sync files.
- You must select a minimum of 20 files.
- All selected files must be from the same source repository (ONTAP volume, S3 bucket, NFS or CIFS share, etc.).
- You'll need to activate the BlueXP copy and sync service and configure a minimum of one data broker that can be used to transfer files between the source and target systems. Review the BlueXP copy and sync requirements beginning with the [Quick Start description](#).

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

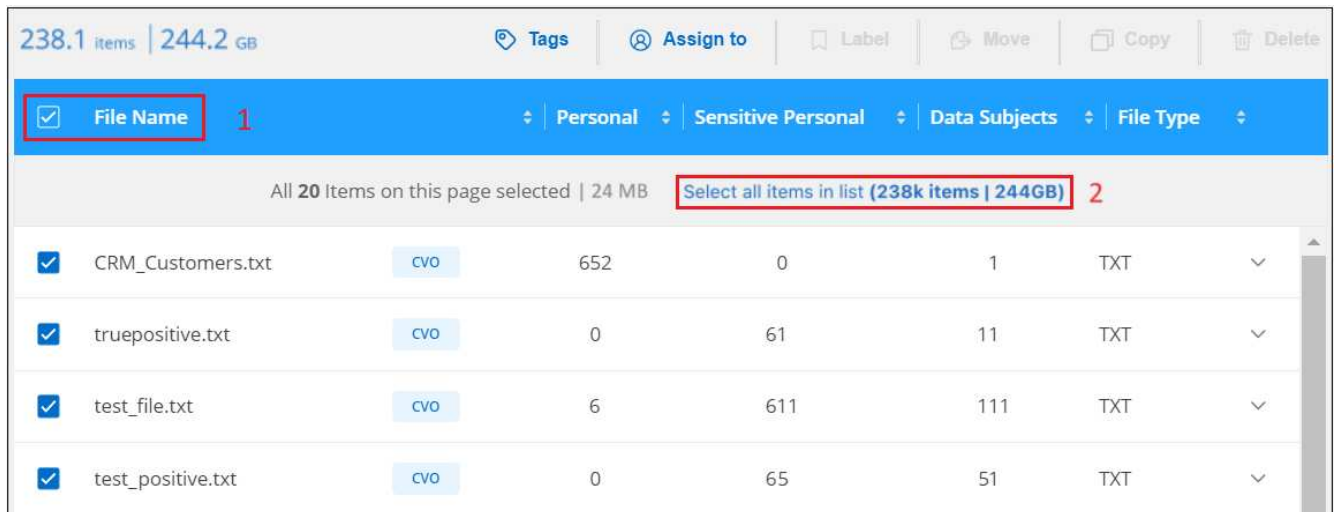
Steps

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

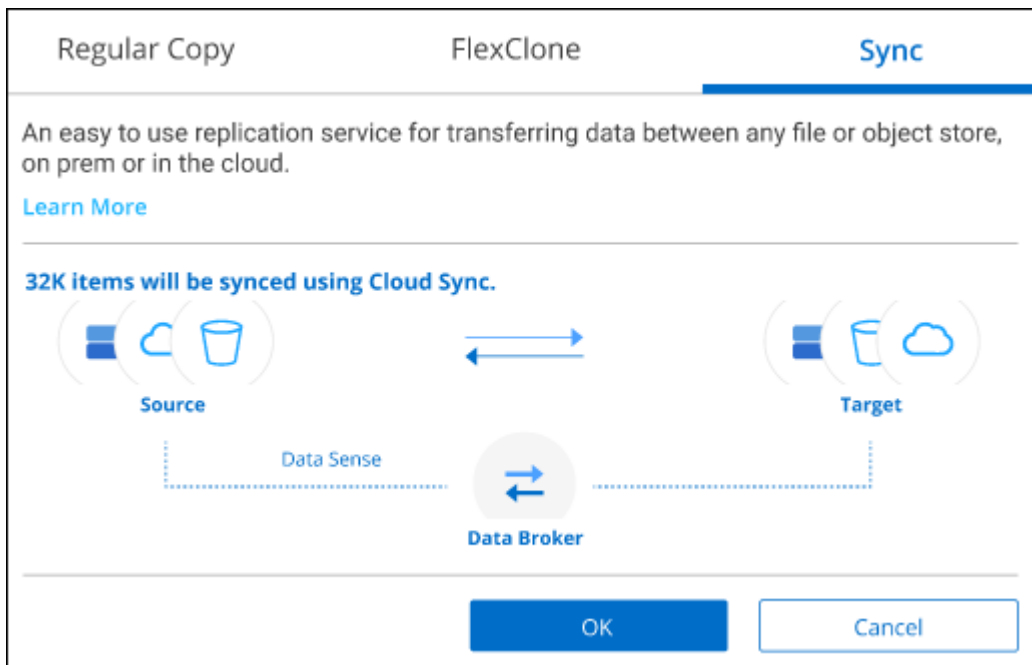


Apply any other filters so that you're seeing only the files that you want to copy and sync to the destination system.

2. In the Investigation results pane, select all files on all pages by checking the box in the title row (File Name), then in the pop-up message **All 20 Items on this page selected** [Select all Items in list \(63K Items\)](#) click **Select all items in list (xxx items)**, and then click **Copy**.



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

Steps

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

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

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

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

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

Move Files (63)

The files will be moved to the destination folder you provide and will no longer be available at their current location.

Moving files is supported only to destination folders in NFS Shares. Any NFS Share is supported, no matter where it is hosted, as long as the share's export policy allows access from the data connector instance IP address.

The status of this action will appear in the Action Status.

Enter the NFS destination folder path to continue

Hostname:/SHAREPATH

Leave breadcrumb

A breadcrumb file helps your users understand why a file was moved from its original location. For each moved file, the system creates a breadcrumb file in the source location named **<filename>-breadcrumb-<date>.txt**.

Enter the content of your breadcrumb

Max length should be maximum 400 characters

Move Files **Cancel**

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 delete files manually from the Investigation pane, or [automatically using Policies](#).



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

Deleting files requires the following permissions:

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

Delete source files manually

Requirements

- You must have the Account Admin or Workspace Admin role to delete files.
- You can delete a maximum of 100,000 files at a time.

Steps

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



- To select individual files, check the box for each file (Volume_1).

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

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

You can view the progress of the delete operation in the [Actions Status pane](#).

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



Add personal data identifiers to your BlueXP classification scans

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

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

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

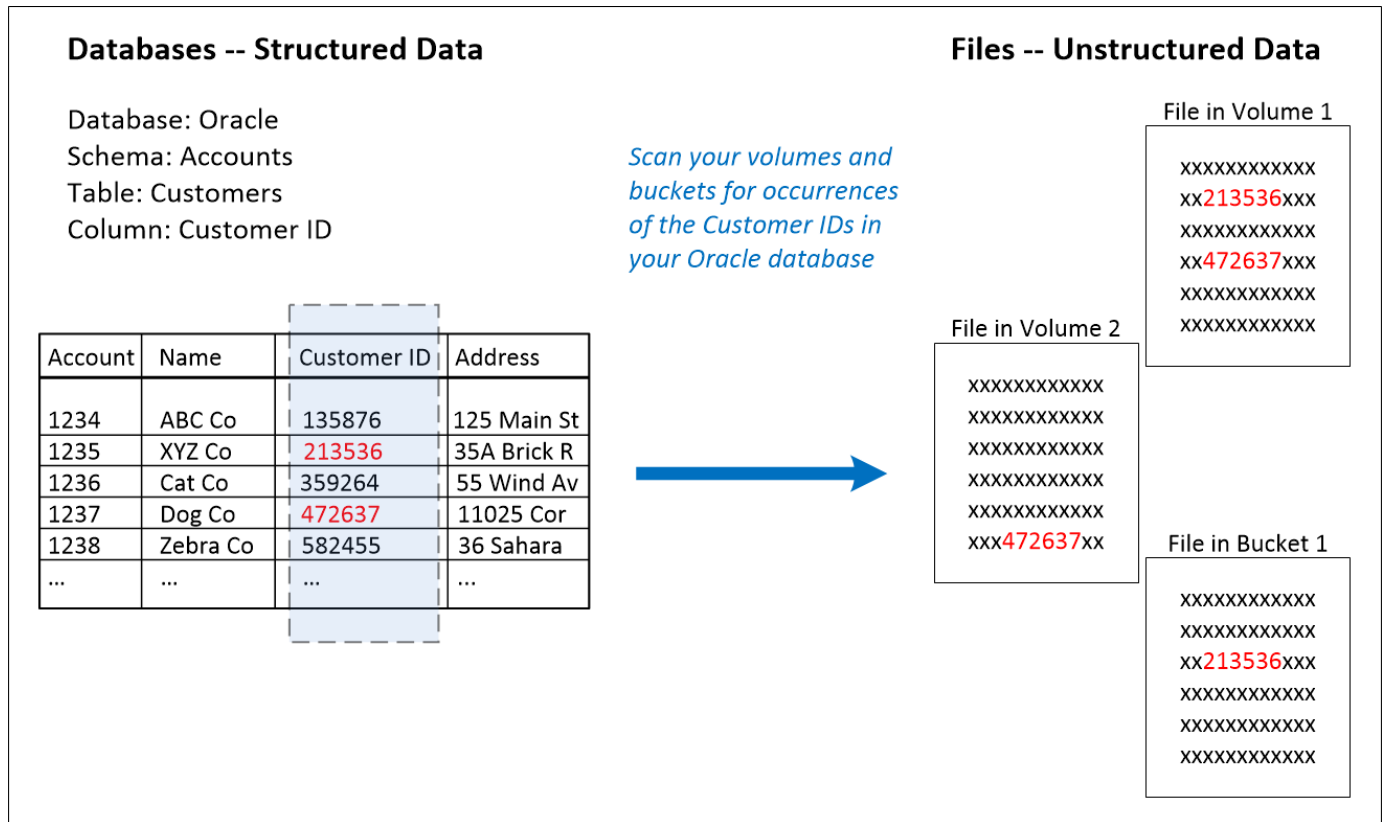
All of these mechanisms to add custom scanning criteria are supported in all languages.



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

Add custom personal data identifiers from your databases

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



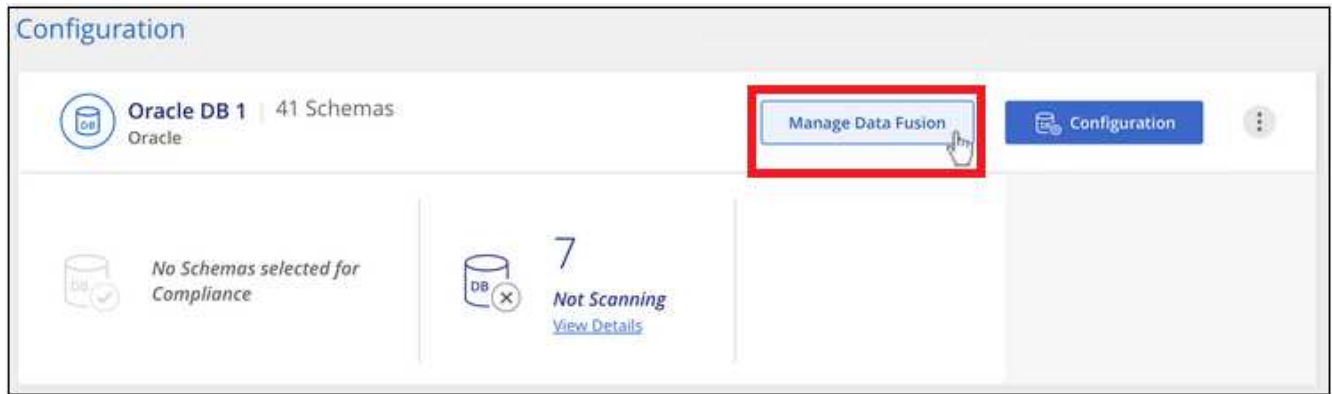
As you can see, two unique Customer IDs have been found in two volumes and in one S3 bucket. Any matches in database tables will also be identified.

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

Steps

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

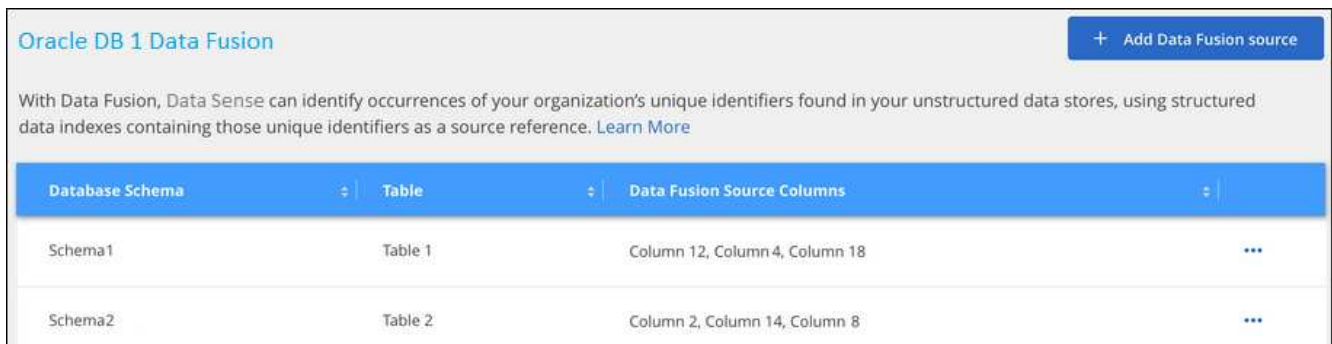
1. In the Configuration page, click **Manage Data Fusion** in the database where the source data resides.



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

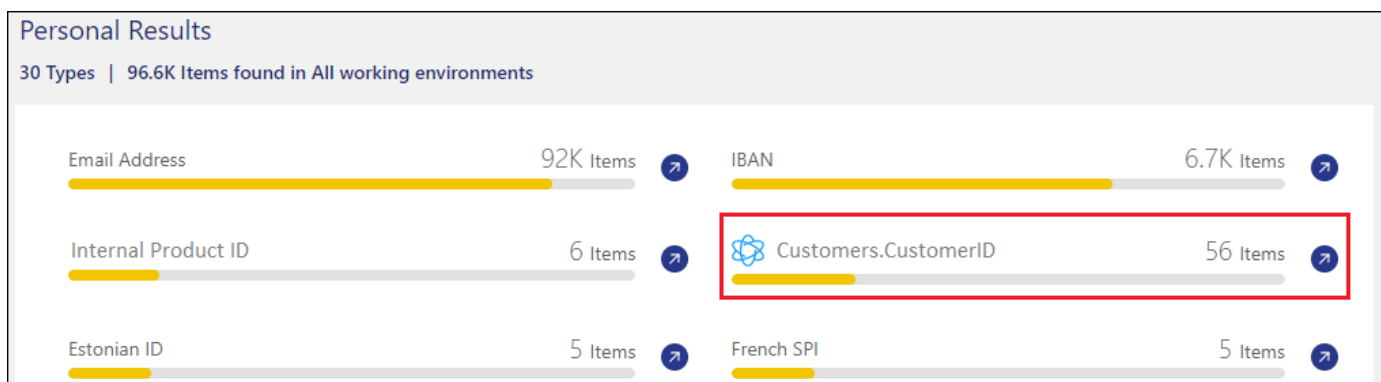
When adding multiple columns, enter each column name, or table view name, on a separate line.

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



Results

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



Delete a Data Fusion source

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



Add custom keywords from a list of words

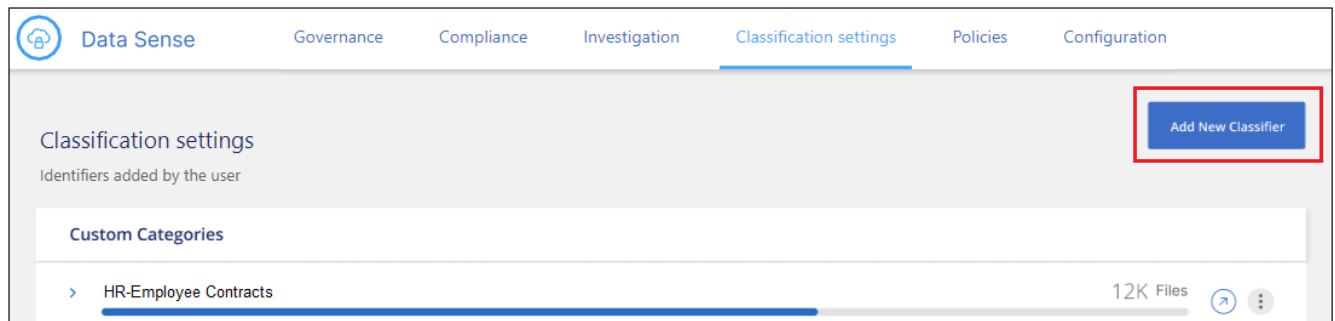
You can add custom keywords to BlueXP classification so that it will identify where that information is found in your data. You add the keywords just by entering each word you want BlueXP classification to recognize. The keywords are added to the existing predefined keywords that BlueXP classification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

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

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

Steps

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



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

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

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

1 Select type 2 Select tool 3 Create Logic

Select type

Select the type of classifier that you want to add to the system, and provide the name and description. Data Sense re-scans all your data sources after you add a new classifier. When the scan is complete, all matching results are displayed in the "Classification Settings" dashboard and in other Data Sense pages.

Classifier name

Description

Personal identifier

The classifier will be added to the system as a new personal identifier. Any matches are considered "personal data", and they are added to the results that are displayed in the Personal Results page and in the Investigation page. [See the list of personal data that Data Sense identifies by default.](#)

Mask detected results in the system

Category

The classifier will be added to the system as a new Category. Any matches are added to the results that are displayed in the Categories page and in the Investigation page. [See the list of categories that Data Sense identifies by default.](#)

Previous **Next**

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**.

Select Data Analysis Tool

Select the tool that will be used to build the list of words, or patterns, that Data Sense will attempt to match in your data sources.

Custom keywords ⓘ
Create a custom personal pattern based on a list of keywords that you provide.

Custom regular expression ⓘ
Create a custom personal pattern based on a regular expression that you define.

DB fusion ⓘ
Create a custom personal pattern based on the values found in specific columns in your scanned database tables. This allows you to identify whether unique identifiers from your databases are found in any of your other data sources.

[Previous](#) [Next](#)

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

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

Create Logic

Create logic for the new identifier, based on regular expression and keywords that should be detected. You will be able to change the logic in the future, by clicking on "edit" from the custom classification dashboard.

Custom keywords list ⓘ

- Maximum of 100,000 words.
- Separate between keywords with a new line
- The keywords are not case sensitive
- Each word must be at least 3 characters long. Shorter words are ignored.
- Duplicate words are only added once.

barred
 barn
 horned
 snowy
 screech

Validate

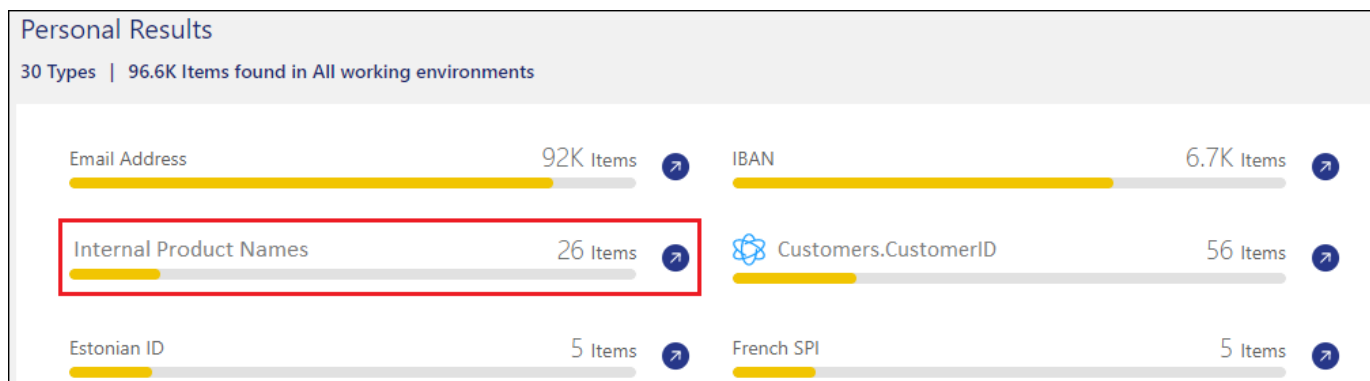
✔ Keywords list is valid.

Previous
Done

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

Results

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



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

Add custom personal data identifiers using a regex

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

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

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

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

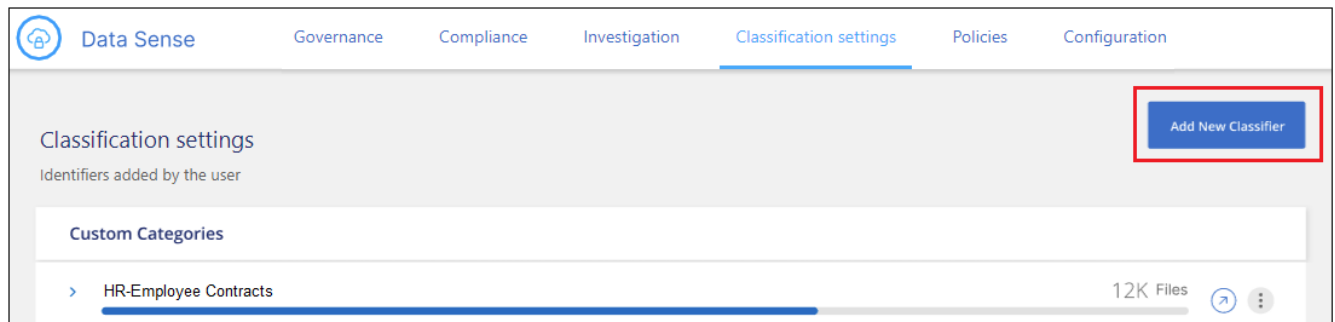
If you need assistance in building the regular expression, refer to [Regular expressions 101](#). Choose **Python** for the Flavor to see the types of results BlueXP classification will match from the regular expression. The [Python Regex Tester page](#) is also useful by displaying a graphical representation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



Currently we do not allow the use of pattern flags when creating a regex - this means you should not use "/".

Steps

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



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

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

1 Select type 2 Select tool 3 Create Logic

Select type

Select the type of classifier that you want to add to the system, and provide the name and description. Data Sense re-scans all your data sources after you add a new classifier. When the scan is complete, all matching results are displayed in the "Classification Settings" dashboard and in other Data Sense pages.

Classifier name

Description

Personal identifier

The classifier will be added to the system as a new personal identifier. Any matches are considered "personal data", and they are added to the results that are displayed in the Personal Results page and in the Investigation page. [See the list of personal data that Data Sense identifies by default.](#)

Mask detected results in the system

Category

The classifier will be added to the system as a new Category. Any matches are added to the results that are displayed in the Categories page and in the Investigation page. [See the list of categories that Data Sense identifies by default.](#)

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

Select Data Analysis Tool

Select the tool that will be used to build the list of words, or patterns, that Data Sense will attempt to match in your data sources.

Custom keywords ⓘ
Create a custom personal pattern based on a list of keywords that you provide.

Custom regular expression ⓘ
Create a custom personal pattern based on a regular expression that you define.

DB fusion ⓘ
Create a custom personal pattern based on the values found in specific columns in your scanned database tables. This allows you to identify whether unique identifiers from your databases are found in any of your other data sources.

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

Create Logic

Create logic for the new identifier, based on regular expression and keywords that should be detected.

Regular expression ⓘ

Add the pattern that should be detected to identify specific information in your data, using a custom regular expression.

Validate

✔ **Success:** Regular expression is valid.

Proximity words - To improve the detection accuracy, insert phrases that must appear near by the regular expression's match.

Previous
Done

Results

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

Data Sense
Governance
Compliance
Investigation
Classification settings
Policies
Configuration

Add New Classifier

Classification settings

Identifiers added by the user

Custom Categories

> HR - Employee Contracts 7.5K Files

Personal information

> Internal Product ID 12K Files

Add custom categories

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

[predefined categories](#).

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

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

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

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

What you'll need

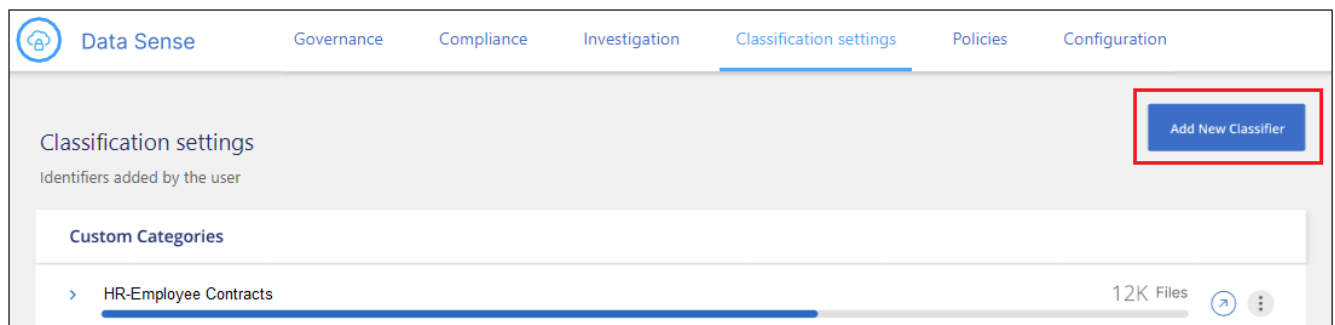
You'll need to create a minimum of 25 training files that contain samples of the categories of data that you want BlueXP classification to recognize. The following file types are supported:

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

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

Steps

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



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

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

1 Select type 2 Select tool 3 Create Logic

Select type

Select the type of classifier that you want to add to the system, and provide the name and description. Data Sense re-scans all your data sources after you add a new classifier. When the scan is complete, all matching results are displayed in the "Classification Settings" dashboard and in other Data Sense pages.

Classifier name

Compressed installer files

Description

Installation files in .GZ format

Personal identifier

The classifier will be added to the system as a new personal identifier. Any matches are considered "personal data", and they are added to the results that are displayed in the Personal Results page and in the Investigation page. [See the list of personal data that Data Sense identifies by default.](#)

Mask detected results in the system

Category

The classifier will be added to the system as a new Category. Any matches are added to the results that are displayed in the Categories page and in the Investigation page. [See the list of categories that Data Sense identifies by default.](#)

Previous Next

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

Create Logic

AI-based similarity training ⓘ

- Insert NFS folder path
- The folder should contain minimum 25 files and maximum 1000 files that will be used for the AI training.
- Supported file types: pdf, docx, doc, pptx, xls, xlsx, csv, txt, gz, rtf, docs, sheets, slides, json
- The keywords are not case sensitive
- Minimum file size: 100B

Compressed Installer files

Select Files

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

Insert folder path that contains at least 25 files for the training

Enter the IP address and volume name, along with the path to the location of the training files.

IP:

Training Data - Folder path:

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**.

Create Logic

AI-based similarity training

- Insert NFS folder path
- The folder should contain minimum 25 files and maximum 1000 files that will be used for the AI training.
- Supported file types: pdf, docx, doc, pptx, xls, xlsx, csv, txt, gz, rtf, docs, sheets, slides, json
- The keywords are not case sensitive
- Minimum file size: 100B

Compressed Installer files

Total uploaded files: 54

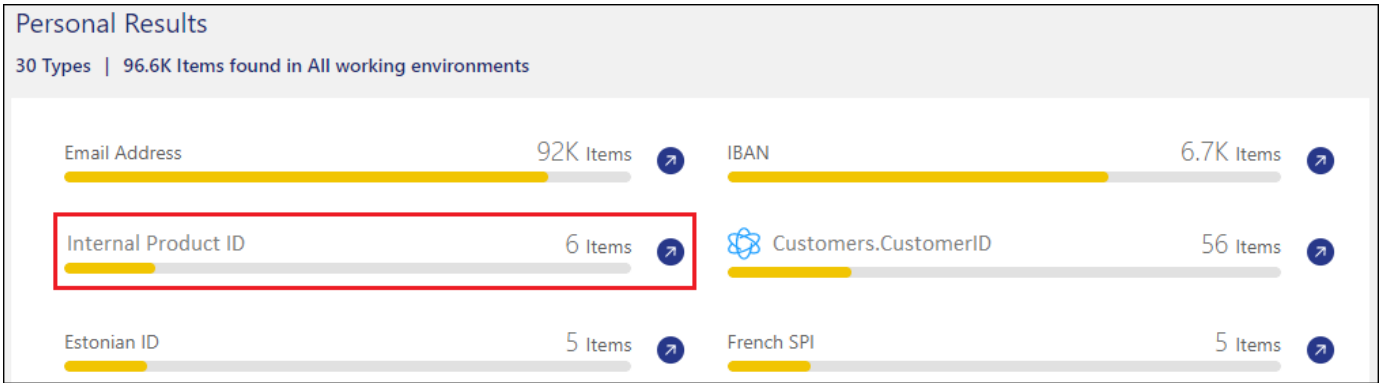
| File name | File Size | File Type | Reliability | included in training |
|-----------|-----------|-----------|-------------|----------------------|
| File1 | 56 | File type | Sufficient | x |
| File2 | 22 | File type | Sufficient | x |
| File3 | 43 | File type | Sufficient | x |
| File4 | 11 | File type | Sufficient | x |

Results

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

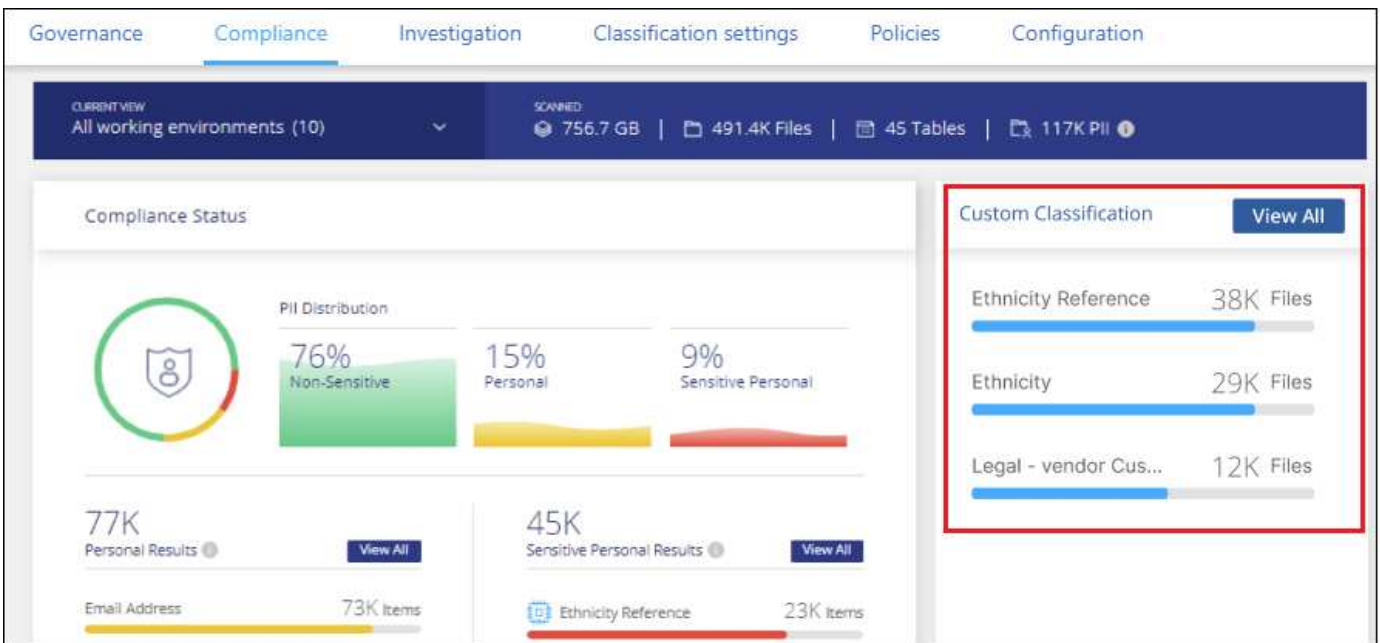
View results from your custom classifiers

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



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

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



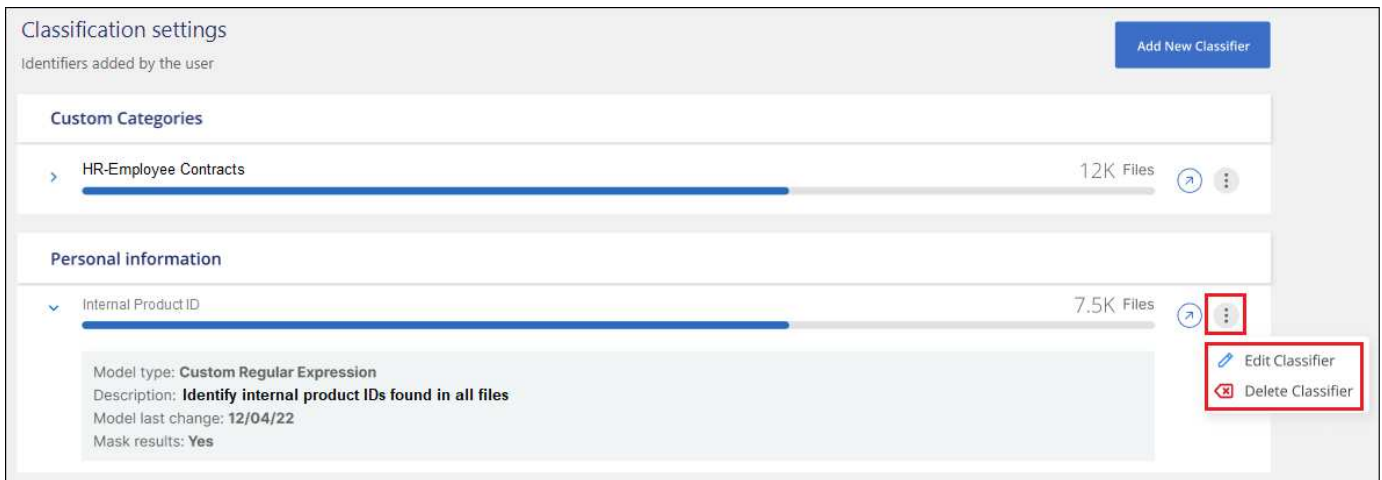
Manage custom classifiers

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



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

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



Viewing the status of your compliance actions

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

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

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

The status can be:

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

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

Steps

1.

In the bottom-right of the BlueXP classification UI you can see the **Actions Status** button



2. Click this button and the most recent 20 actions are listed.

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

Audit the history of BlueXP classification actions

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

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

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

Log file contents

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

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

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

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

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

Log file locations

The management audit log files are located on the BlueXP classification machine in:

`/opt/netapp/audit_logs/`

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

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

Access the log files

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

Reducing the BlueXP classification scan speed

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

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



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

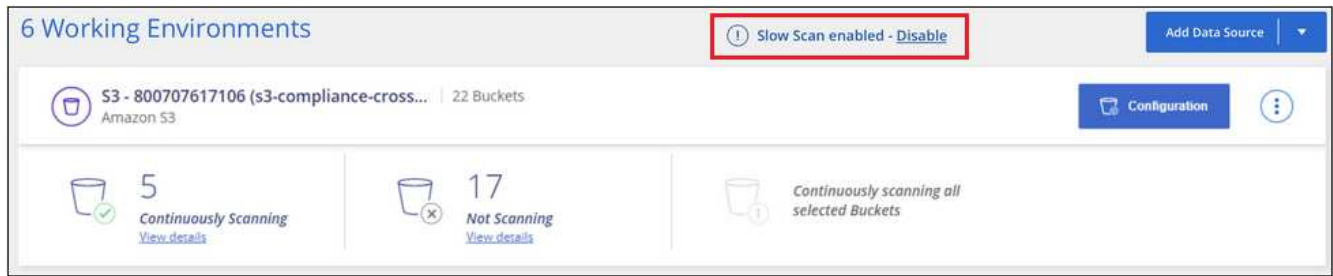
NOTE This information is relevant only for BlueXP classification legacy versions 1.30 and earlier.

Steps

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

The screenshot shows the '6 Working Environments' configuration page. At the top, there is a header '6 Working Environments' and a button 'Add Data Source'. Below this, there are two main sections for Amazon S3 buckets. The first section is for 'S3 - 800707617106 (s3-compliance-cross... | 22 Buckets)'. It shows three status indicators: '5 Continuously Scanning' (with a green checkmark), '17 Not Scanning' (with a red X), and 'Continuously scanning all selected Buckets' (with a blue 'i' icon). The second section is for 'S3 - 759995470648 | 90 Buckets', showing '3 Continuously Scanning', '87 Not Scanning', and 'Continuously scanning all selected Buckets'. At the bottom of the page, there is a toggle switch labeled 'Activate Slow Scan' which is currently turned on (blue). A red box highlights this toggle switch.

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




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

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

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

Steps

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



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

Copyright information

Copyright © 2024 NetApp, Inc. All Rights Reserved. Printed in the U.S. No part of this document covered by copyright may be reproduced in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or storage in an electronic retrieval system—without prior written permission of the copyright owner.

Software derived from copyrighted NetApp material is subject to the following license and disclaimer:

THIS SOFTWARE IS PROVIDED BY NETAPP "AS IS" AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, WHICH ARE HEREBY DISCLAIMED. IN NO EVENT SHALL NETAPP BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

NetApp reserves the right to change any products described herein at any time, and without notice. NetApp assumes no responsibility or liability arising from the use of products described herein, except as expressly agreed to in writing by NetApp. The use or purchase of this product does not convey a license under any patent rights, trademark rights, or any other intellectual property rights of NetApp.

The product described in this manual may be protected by one or more U.S. patents, foreign patents, or pending applications.

LIMITED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (b)(3) of the Rights in Technical Data -Noncommercial Items at DFARS 252.227-7013 (FEB 2014) and FAR 52.227-19 (DEC 2007).

Data contained herein pertains to a commercial product and/or commercial service (as defined in FAR 2.101) and is proprietary to NetApp, Inc. All NetApp technical data and computer software provided under this Agreement is commercial in nature and developed solely at private expense. The U.S. Government has a non-exclusive, non-transferrable, nonsublicensable, worldwide, limited irrevocable license to use the Data only in connection with and in support of the U.S. Government contract under which the Data was delivered. Except as provided herein, the Data may not be used, disclosed, reproduced, modified, performed, or displayed without the prior written approval of NetApp, Inc. United States Government license rights for the Department of Defense are limited to those rights identified in DFARS clause 252.227-7015(b) (FEB 2014).

Trademark information

NETAPP, the NETAPP logo, and the marks listed at <http://www.netapp.com/TM> are trademarks of NetApp, Inc. Other company and product names may be trademarks of their respective owners.