



Use BlueXP classification

BlueXP classification

NetApp
March 14, 2024

Table of Contents

- Use BlueXP classification 1
 - View governance details about the data stored in your organization 1
 - View compliance details about the data stored in your organization 7
 - Categories of private data 14
 - Investigate the data stored in your organization 21
 - Organize your private data 29
 - Assign policies to your data 38
 - Manage your private data 49
 - Monitor and manage file access events 59
 - View compliance reports 63

Use BlueXP classification

View governance details about the data stored in your organization

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

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

The Governance dashboard

The Governance dashboard provides information so that you can increase the efficiency and control the costs related to the data stored on your storage resources.

Save Opportunities

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

- **Stale Data** - Data that was last modified over 3 years ago.
- **Non-Business Data** - Data considered not to be business related, based on their Category or File Type. This includes:
 - Application Data
 - Audio
 - Executables
 - Images
 - Logs
 - Videos
 - Miscellaneous (general "other" category)
- **Duplicate Files** - Files that are duplicated in other locations in the data sources you are scanning. [See what types of duplicate files are displayed.](#)

NOTE

If any of your data sources implement data tiering, old data that already resides in object storage may be identified in the *Stale Data* category.

Policies with the largest number of results

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

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

Data Overview

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

Top data repositories listed by data sensitivity

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

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

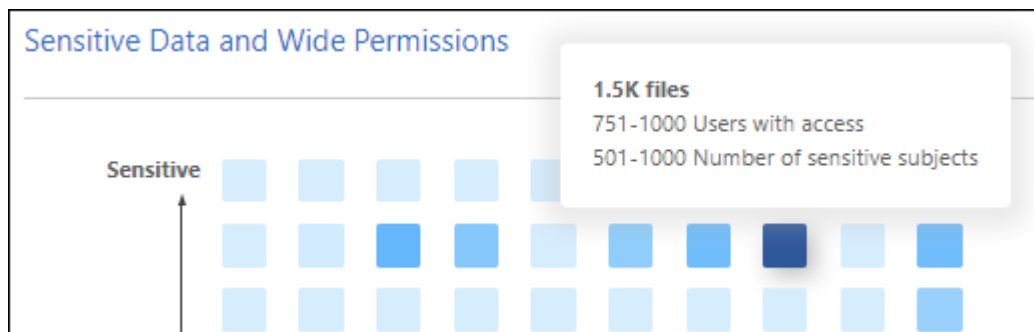
You can hover over each section to see the total number of items in each category.

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

Data listed by sensitivity and wide permissions

The *Sensitive Data and Wide Permissions* area 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.

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 hover 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.

Data listed by types of Open Permissions

The *Open Permissions* area shows the percentage for each type of permissions that exist for all files that are being scanned. The chart shows the following types of permissions:

- No Open Permissions
- Open to Organization
- Open to Public
- Unknown Access

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

Age of Data and Size of Data graphs

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

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

- **Age of Data graph** - Categorizes data based on the time it was created, the last time it was accessed, or the last time it was modified.
- **Size of Data graph** - Categorizes data based on size.

NOTE

If any of your data sources implement data tiering, old data that already resides in object storage may be identified in the *Age of Data* graph.

Most identified data Classifications

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

Categories

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

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

File types

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

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

AIP labels

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.

Data Mapping Report

The Data Mapping Report provides an overview of the data being stored in your corporate data sources to assist you with decisions of migration, back up, security, and compliance processes. The report first lists an overview that summarizes all your working environments and data sources, and then it provides a breakdown for each working environment.

The report includes the following information:

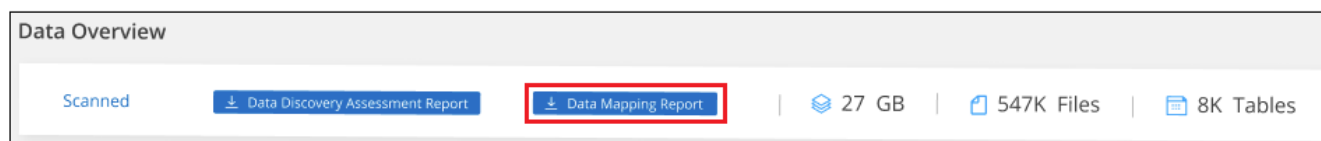
Category	Description
Usage Capacity	For all working environments: Lists the number of files and the used capacity for each working environment. For single working environments: Lists the files that are using the most capacity.
Age of Data	Provides three charts and graphs for when files were created, last modified, or last accessed. Lists the number of files, and their used capacity, based on certain date ranges.
Size of Data	Lists the number of files that exist within certain size ranges in your working environments.
File Types	Lists the total number of files and the used capacity for each type of file being stored in your working environments.

Generate the Data Mapping Report

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

Steps


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



Result

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

If the report is larger than 1 MB, the PDF file is retained on the BlueXP classification instance and you'll see a pop-up message about the exact location. When BlueXP classification is installed on a Linux machine on your premises, or on a Linux machine you deployed in the cloud, you can navigate directly to the PDF file. When BlueXP classification is deployed in the cloud, you'll need to SSH to the BlueXP classification instance to download PDF file. [See how to access data on the Classification instance.](#)

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

Data Discovery Assessment Report

The Data Discovery Assessment Report provides a high-level analysis of the scanned environment to highlight the system's findings and to show areas of concern and potential remediation steps. The results are based on both mapping and classifying your data. The goal of this report is to raise awareness of three significant aspects of your data set:

Feature	Description
Data governance concerns	A detailed picture of all the data you own and areas where you may be able to reduce the amount of data to save costs.
Data security exposures	Areas where your data is accessible to internal or external attacks because of broad access permissions.
Data compliance gaps	Where your personal or sensitive personal information is located for both security and for DSARs (data subject access requests).

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

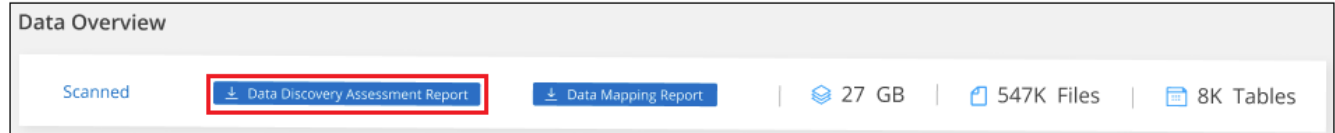
- Reduce storage costs by changing your retention policy, or by moving or deleting certain data (stale, duplicate, or non-business data)
- Protect your data that has broad permissions by revising global group management policies
- Protect your data that has personal or sensitive personal information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

Generate the Data Discovery Assessment Report

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


Steps

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



Result

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

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

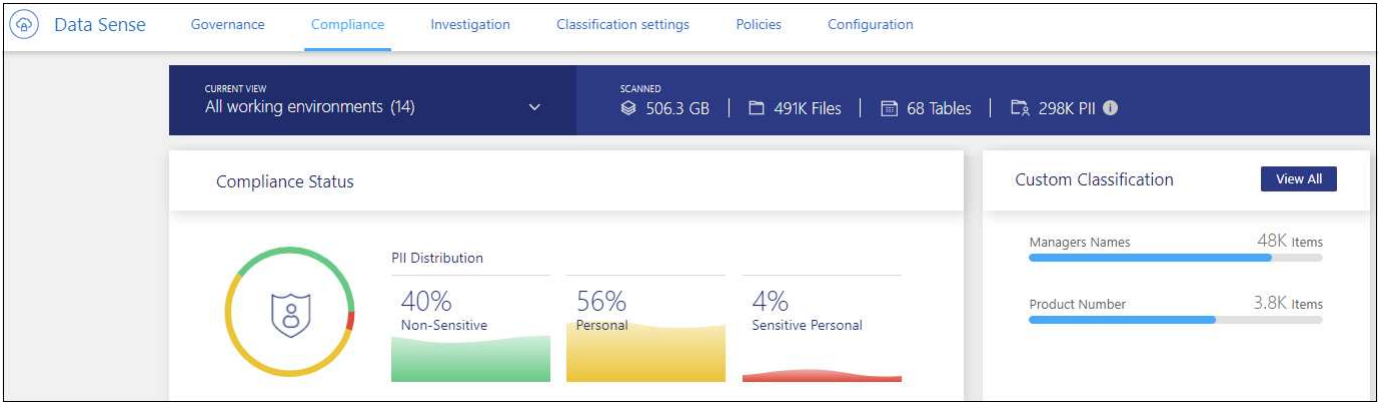
View compliance details about the data stored in your organization

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



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

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



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

You can also filter the results from the Data Investigation page and download a report of the results as a CSV file. See [Filtering data in the Data Investigation page](#) for details.

View files that contain personal data

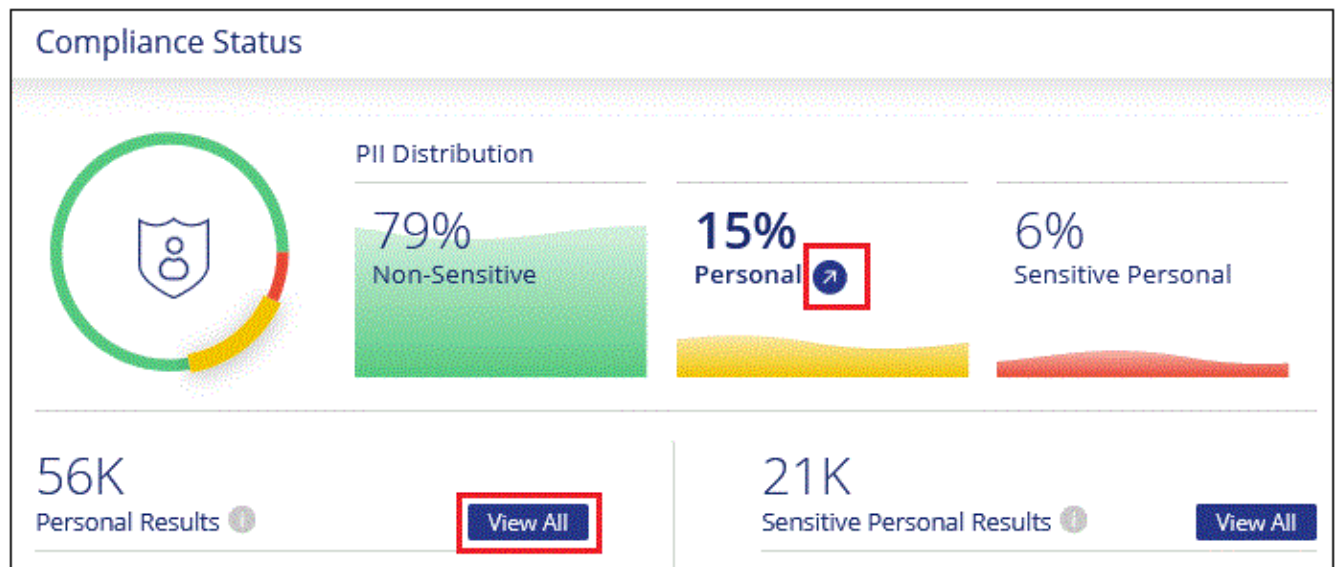
BlueXP classification automatically identifies specific words, strings, and patterns (Regex) inside the data. For example, Personal Identification Information (PII), credit card numbers, social security numbers, bank account numbers, passwords, and more. [See the full list](#). BlueXP classification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

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

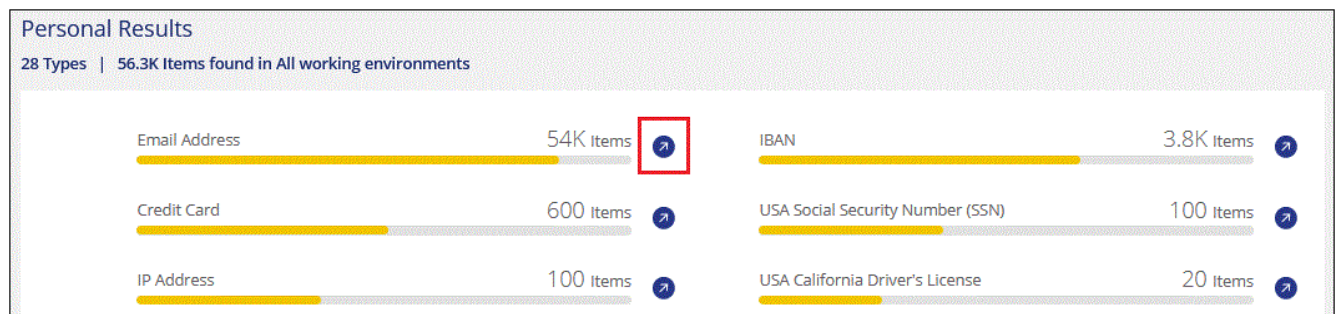
For some types of personal data, BlueXP classification uses *proximity validation* to validate its findings. The validation occurs by looking for one or more predefined keywords in proximity to the personal data that was found. For example, BlueXP classification identifies a U.S. social security number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

Steps

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



3. To investigate the details for a specific type of personal data, click **View All** and then click the **Investigate Results** icon for a specific type of personal data; for example, email addresses.



4. Investigate the data by searching, sorting, expanding details for a specific file, clicking **Investigate Results** to see masked information, or by downloading the file list.

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

The screenshot shows the BlueXP interface for a file named 'customer-data.xls'. The top navigation bar includes tabs for 'Unstructured (54.6K Files)', 'Directories (6 Folders)', and 'Structured (3 Tables)'. Below this, a summary bar shows '54.6K items | 1.95 GB' and action buttons like 'Tags', 'Assign to', 'Label', 'Move', 'Copy', and 'Delete'. A blue header bar contains filters: 'File Name', 'Personal', 'Sensitive Personal', 'Data Subjects', and 'File Type'. The file entry 'customer-data.xls' is highlighted, with a red box around the number '63' in the 'Sensitive Personal' column. To the right of the file name are icons for 'S3', '688', '0', and 'XLS'. Below the file name, a detailed sidebar shows metadata: Tags (Credit Cards, gidi, tartanpion), Working Environment (Account: S3 - 759995470648), Storage Repository (Bucket: compliancedemofiles), File Path (/Patterns/NEW SSN/customer-data.xls), Category (Miscellaneous Spreadsheets), File Size (142.35 KB), Discovered Time (2020-11-16 12:40), Created Time (2019-12-16 12:18), Last Modified (2019-12-16 12:18), Open Permissions (NOT PUBLIC), and Duplicates (2). On the right side of the sidebar, there are buttons for 'Tags: 3 tags', 'Assigned to: Alona Tyupa', 'Assign a Label to this file', 'Copy File', 'Move File', and 'Delete File'. At the bottom right, a link says 'Give feedback on this result'.

The screenshot shows the BlueXP interface for a directory named '/datasensecopy/C\$/...'. The top navigation bar includes tabs for 'Unstructured (491.4K Files)', 'Directories (60.7K Folders)', and 'Structured (45 Tables)'. Below this, a summary bar shows '60.7K items | 2.3 GB' and action buttons like 'Tags', 'Assign to', 'Label', 'Move', 'Copy', and 'Delete'. A blue header bar contains filters: 'Directory Name', 'Storage Repository', 'Personal', 'Sensitive Personal', and 'Type'. The directory entry is highlighted, with a red box around the number '10' in the 'Sensitive Personal' column. To the right of the directory name are icons for 'CVO', 'ANF', and 'Folder'. Below the directory name, a detailed sidebar shows metadata: Working Environment (Azure NetApp Files), Storage Repository (Volume: datasensecopy), Directory Path (/datasensecopy/copy_63/contextual_data/C\$/Users/shraga.WESTEROS/Desktop/...), Discovered Time (2022-07-10 22:58), and Last Modified (2020-02-06 09:57).

View files that contain sensitive personal data

BlueXP classification automatically identifies special types of sensitive personal information, as defined by privacy regulations such as [articles 9 and 10 of the GDPR](#). For example, information regarding a person's health, ethnic origin, or sexual orientation. [See the full list](#). BlueXP classification identifies this type of information in individual files, in files within directories (shares and folders), and in database tables.

BlueXP classification uses artificial intelligence (AI), natural language processing (NLP), machine learning

(ML), and cognitive computing (CC) to understand the meaning of the content that it scans in order to extract entities and categorize it accordingly.

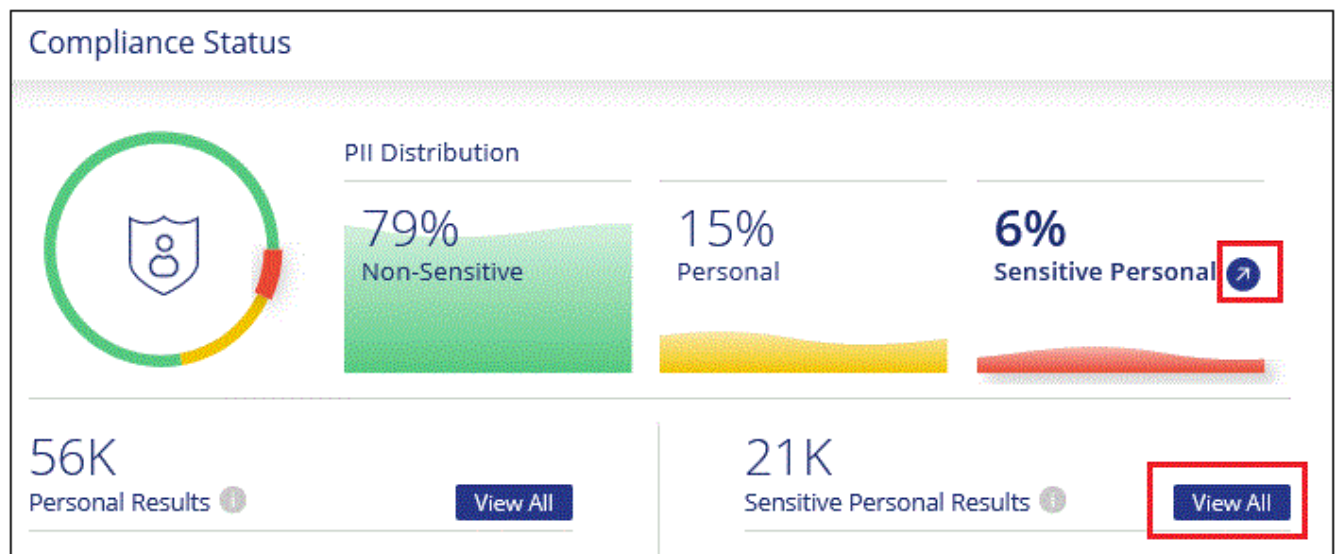
For example, one sensitive GDPR data category is ethnic origin. Because of its NLP abilities, BlueXP classification can distinguish the difference between a sentence that reads "George is Mexican" (indicating sensitive data as specified in article 9 of the GDPR), versus "George is eating Mexican food."



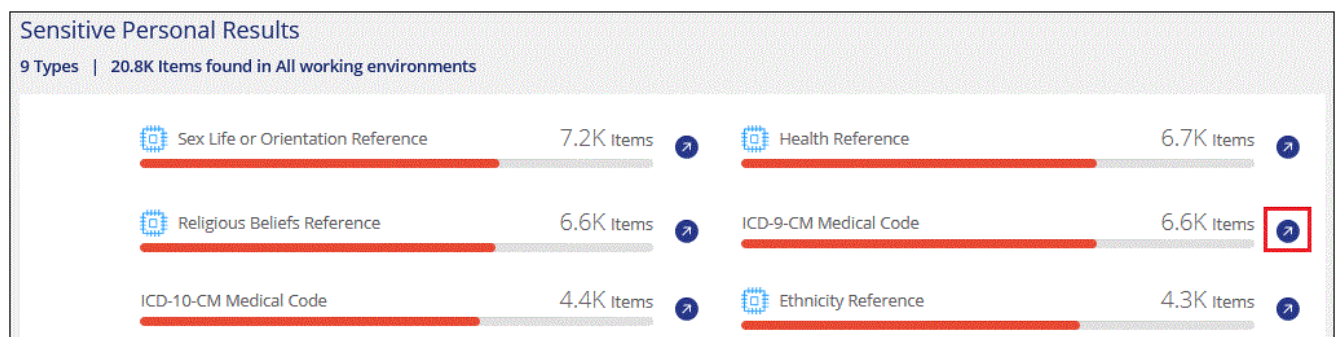
Only English is supported when scanning for sensitive personal data. Support for more languages will be added later.

Steps

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



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



4. Investigate the data by searching, sorting, expanding details for a specific file, clicking **Investigate Results** to see masked information, or by downloading the file list.

View files by categories

BlueXP classification takes the data that it scanned and divides it into different types of categories. Categories

are topics based on AI analysis of the content and metadata of each file. [See the list of categories.](#)

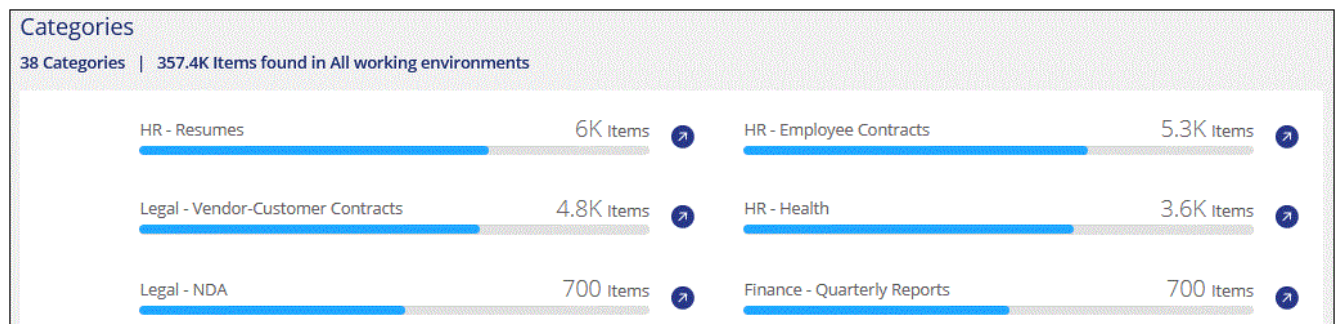
Categories can help you understand what's happening with your data by showing you the types of information that you have. For example, a category like resumes or employee contracts can include sensitive data. When you investigate the results, you might find that employee contracts are stored in an insecure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



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

Steps

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



3. Investigate the data by searching, sorting, expanding details for a specific file, clicking **Investigate Results** to see masked information, or by downloading the file list.

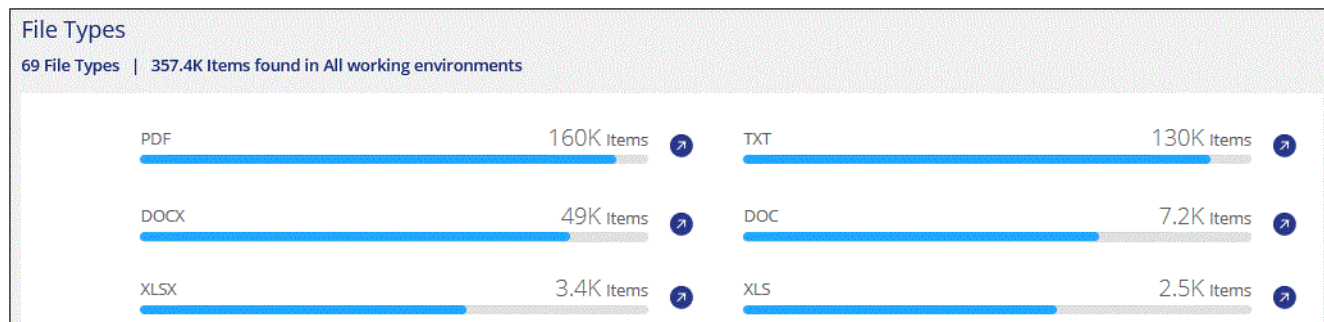
View files by file types

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

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

Steps

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



- Investigate the data by searching, sorting, expanding details for a specific file, clicking **Investigate Results** to see masked information, or by downloading the file list.

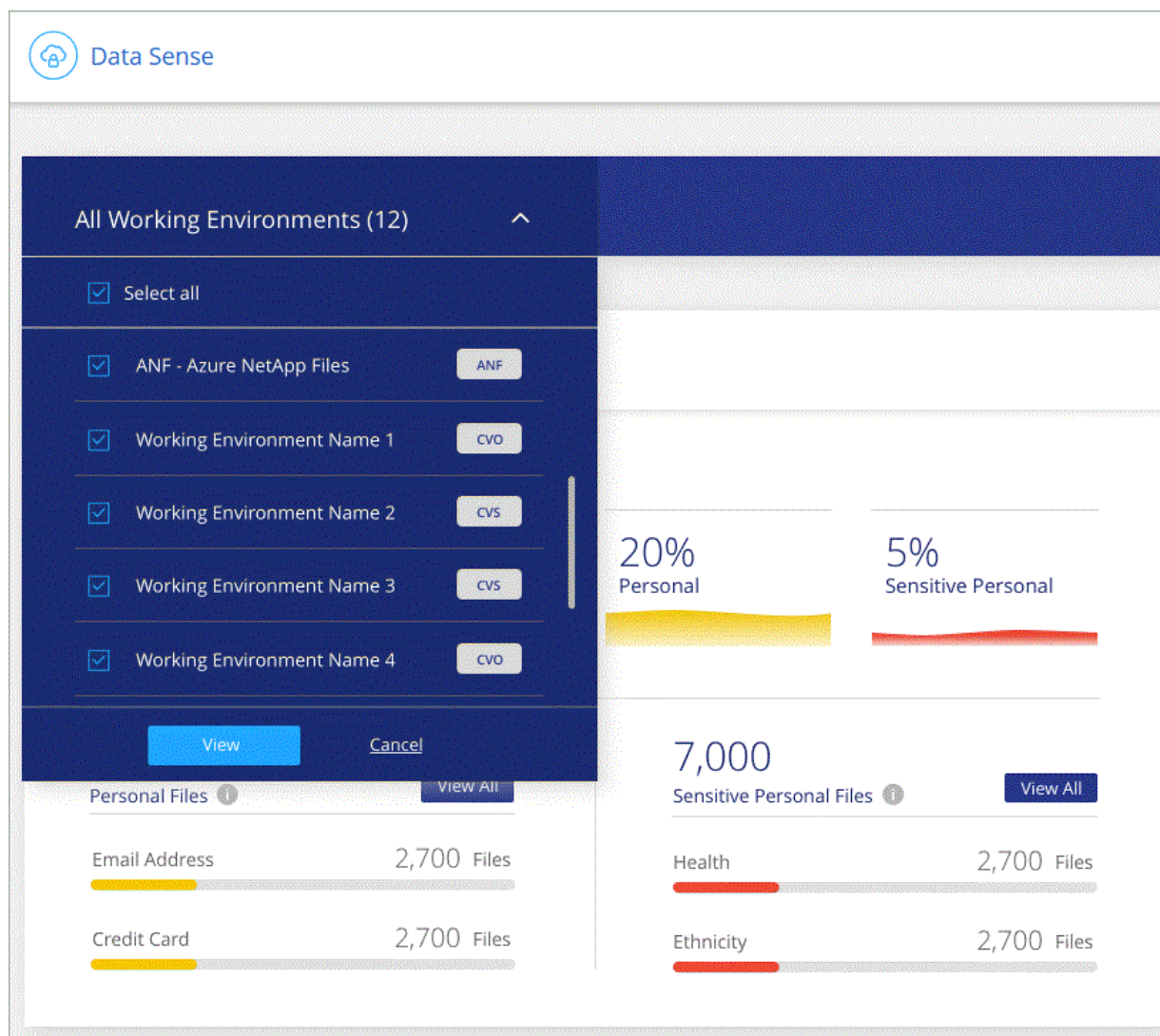
View Dashboard data for specific working environments

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

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

Steps

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



Categories of private data

There are many types of private data that BlueXP classification can identify in your volumes, Amazon S3 buckets, databases, OneDrive folders, SharePoint accounts, and Google Drive accounts. See the categories below.



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

Types of personal data

The personal data found in files can be general personal data or national identifiers. The third column in the table below identifies whether BlueXP classification uses [proximity validation](#) to validate its findings for the identifier.

The languages in which these items can be recognized are identified in the table.

Note that you can add to the list of personal data that is found in your files. If you are scanning a database

server, the *Data Fusion* feature enables you to choose additional identifiers that BlueXP classification will look for in its' scans by selecting columns in a database table. You can also add custom keywords from a text file, or custom patterns using a regular expression. See [Adding personal data identifiers to your BlueXP classification scans](#) for details.

Type	Identifier	Proximity validation?	English	German	Spanish	French	Japanese
General	Credit card number	No	✓	✓	✓		✓
	Data Subjects	No	✓	✓	✓		
	Email Address	No	✓	✓	✓		✓
	IBAN Number (International Bank Account Number)	No	✓	✓	✓		✓
	IP Address	No	✓	✓	✓		✓
	Password	Yes	✓	✓	✓		✓

Type	Identifier	Proximity validation?	English	German	Spanish	French	Japanese
------	------------	--------------------------	---------	--------	---------	--------	----------

National Identifiers							
-------------------------	--	--	--	--	--	--	--

Type	Identifier	Proximity validation?	English	German	Spanish	French	Japanese
------	------------	--------------------------	---------	--------	---------	--------	----------

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

Types of sensitive personal data

The sensitive personal data that BlueXP classification can find in files includes the following list.

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

Criminal Procedures Reference

Data concerning a natural person's criminal convictions and offenses.

Ethnicity Reference

Data concerning a natural person's racial or ethnic origin.

Health Reference

Data concerning a natural person's health.

ICD-9-CM Medical Codes

Codes used in the medical and health industry.

ICD-10-CM Medical Codes

Codes used in the medical and health industry.

Philosophical Beliefs Reference

Data concerning a natural person's philosophical beliefs.

Political Opinions Reference

Data concerning a natural person's political opinions.

Religious Beliefs Reference

Data concerning a natural person's religious beliefs.

Sex Life or Orientation Reference

Data concerning a natural person's sex life or sexual orientation.

Types of categories

BlueXP classification categorizes your data as follows.

Most of these categories can be recognized in English, German, and Spanish.

Category	Type	English	German	Spanish
Finance	Balance Sheets	✓	✓	✓
	Purchase Orders	✓	✓	✓
	Invoices	✓	✓	✓
	Quarterly Reports	✓	✓	✓
HR	Background Checks	✓		✓
	Compensation Plans	✓	✓	✓
	Employee Contracts	✓		✓
	Employee Reviews	✓		✓
	Health	✓		✓
	Resumes	✓	✓	✓
Legal	NDAs	✓	✓	✓
	Vendor-Customer contracts	✓	✓	✓
Marketing	Campaigns	✓	✓	✓
	Conferences	✓	✓	✓
Operations	Audit Reports	✓	✓	✓
Sales	Sales Orders	✓	✓	
Services	RFI	✓		✓
	RFP	✓		✓
	SOW	✓	✓	✓
	Training	✓	✓	✓
Support	Complaints and Tickets	✓	✓	✓

The following Metadata is also categorized, and are identified in the same supported languages:

- Application Data
- Archive Files
- Audio
- Business Application Data
- CAD Files
- Code
- Corrupted
- Database and index files
- BlueXP classification Breadcrumbs
- Design Files
- Email Application Data

- Encrypted (files with a high entropy score)
- Executables
- Financial Application Data
- Health Application Data
- Images
- Logs
- Miscellaneous Documents
- Miscellaneous Presentations
- Miscellaneous Spreadsheets
- Miscellaneous "Unknown"
- Password Protected files
- Structured Data
- Videos
- Zero-Byte Files

Types of files

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

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

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

Accuracy of information found

NetApp can't guarantee 100% accuracy of the personal data and sensitive personal data that BlueXP classification identifies. You should always validate the information by reviewing the data.

Based on our testing, the table below shows the accuracy of the information that BlueXP classification finds. We break it down by *precision* and *recall*:

Precision

The probability that what BlueXP classification finds has been identified correctly. For example, a precision rate of 90% for personal data means that 9 out of 10 files identified as containing personal information, actually contain personal information. 1 out of 10 files would be a false positive.

Recall

The probability for BlueXP classification to find what it should. For example, a recall rate of 70% for personal data means that BlueXP classification can identify 7 out of 10 files that actually contain personal information in your organization. BlueXP classification would miss 30% of the data and it won't appear in the dashboard.

We are constantly improving the accuracy of our results. Those improvements will be automatically available in future BlueXP classification releases.

Type	Precision	Recall
Personal data - General	90%-95%	60%-80%
Personal data - Country identifiers	30%-60%	40%-60%
Sensitive personal data	80%-95%	20%-30%
Categories	90%-97%	60%-80%

Investigate the data stored in your organization

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



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

Filter data in the Data Investigation page

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

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

- The top-level tabs enable you to view data from files (unstructured data), directories (folders and file shares), or from databases (structured data).

- The controls at the top of each column enable you to sort the results in numerical or alphabetical order.
- The left-pane filters enable you to refine the results by selecting the attributes described in the next sections.

Filter data by sensitivity and content

Use the following filters to view how much sensitive information is contained in your data.

Filter	Details
Category	Select the types of categories .
Sensitivity Level	Select the sensitivity level: Personal, Sensitive personal, or Non sensitive.
Number of identifiers	Select the range of detected sensitive identifiers per file. Includes personal data and sensitive personal data. When filtering in Directories, BlueXP classification totals the matches from all files in each folder (and sub-folders). NOTE: The December 2023 (version 1.26.6) release temporarily removed the option to calculate the number of personal identifiable information (PII) data by Directories.
Personal Data	Select the types of personal data .
Sensitive Personal Data	Select the types of sensitive personal data .
Data Subject	Enter a data subject's full name or known identifier. Learn more about data subjects here .

Filter data by user owner and user permissions

Use the following filters to view file owners and permissions to access your data.

Filter	Details
Open Permissions	Select the type of permissions within the data and within folders/shares.
User / Group Permissions	Select one or multiple user names and/or group names, or enter a partial name.
File Owner	Enter the file owner name.
Number of users with access	Select one or multiple category ranges to show which files and folders are open to a certain number of users.

Filter data by time

Use the following filters to view data based on time criteria.

Filter	Details
Created Time	Select a time range when the file was created. You can also specify a custom time range to further refine the search results.

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

Filter data by metadata

Use the following filters to view data based on location, size, and directory or file type.

Filter	Details
File Path	Enter up to 20 partial or full paths that you want to include or exclude from the query. If you enter both include paths and exclude paths, BlueXP classification finds all files in the included paths first, then it removes files from excluded paths, and then it displays the results. Note that using "*" in this filter has no effect, and that you can't exclude specific folders from the scan - all the directories and files under a configured share will be scanned.
Directory Type	Select the directory type; either "Share" or "Folder".
File Type	Select the types of files .
File Size	Select the file size range.
File Hash	Enter the file's hash to find a specific file, even if the name is different.

Filter data by storage type

Use the following filters to view data by storage type.

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

Filter data by tags, labels, assigned users, and policies

Use the following filters to view data by AIP labels or tags.

Filter	Details
Policies	Select a policy or policies. Go here to view the list of existing policies and to create your own custom policies.
Label	Select AIP labels that are assigned to your files.
Tags	Select the tag or tags that are assigned to your files.
Assigned To	Select the name of the person to which the file is assigned.

Filter data by analysis status

Use the following filter to view data by the BlueXP classification scan status.

Filter	Details
Analysis Status	Select an option to show the list of files that are Pending First Scan, Completed being scanned, Pending Rescan, or that have Failed to be scanned.
Scan Analysis Event	Select whether you want to view files that were not classified because BlueXP classification couldn't revert last accessed time, or files that were classified even though BlueXP classification couldn't revert last accessed time.

See [details about the "last accessed time" timestamp](#) for more information about the items that appear in the Investigation page when filtering using the Scan Analysis Event.

Filter data by duplicates

Use the following filter to view files that are duplicated in your storage.

Filter	Details
Duplicates	Select whether the file is duplicated in the repositories.

View file metadata

In the Data Investigation results pane you can click  for any single file to view the file metadata.

The screenshot shows a file management interface. At the top, there's a header with '365K items | 14 GB' and several action buttons: Tags, Assign to, Label, Move, Copy, and Delete. Below this is a table with columns: File Name, Personal, Sensitive Personal, Data Subjects, and File Type. Two files are listed: 'ground truth.xlsx' and 'GM_PD 12-1-09 SP.xls.pdf'. The second file is selected, and its details are shown in a modal window. The details include: Tags (Decathlon, gidi, IS NOT OK, And 6 more, View All), Working Environment (OneDrive daylabs.onmicrosoft.com), Storage Repository (User: ruh@daylabs.onmicrosoft.com), File Path (/scattered/26/GM_PD 12-1-09 SP.xls.pdf), Category (Miscellaneous Documents), File Size (427.46 KB), Discovered Time (2021-01-12 10:37), Created Time (2018-05-22 12:38), Last Modified (2018-10-22 13:28), and Duplicates (None). On the right side of the modal, there are buttons for Tags (9 tags), Assigned to (Amit Ashbel), Assign a Label to this file, Copy File, Move File, and Delete File. A red box highlights a back arrow icon in the top right corner of the modal.

In addition to showing you the working environment and volume where the file resides, the metadata shows much more information, including the file permissions, file owner, whether there are duplicates of this file, and assigned AIP label (if you have [integrated AIP in BlueXP classification](#)). This information is useful if you're planning to [create Policies](#) because you can see all the information that you can use to filter your data.

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

When viewing the details for a single file there are a few actions you can take on the file:

- You can move or copy the file to any NFS share. See [Moving source files to an NFS share](#) and [Copying source files to an NFS share](#) for details.
- You can delete the file. See [Deleting source files](#) for details.
- You can assign a certain Status to the file. See [Applying tags](#) for details.
- You can assign the file to a BlueXP user to be responsible for any follow-up actions that need to be done on the file. See [Assigning users to a file](#) for details.
- If you have integrated AIP labels with BlueXP classification, you can assign a label to this file, or change to a different label if one already exists. See [Assigning AIP labels manually](#) for details.

View permissions for files and directories

To view a list of all users or groups who have access to a file or to a directory, and the types of permissions they have, click **View all Permissions**. This button is available only for data in CIFS shares, SharePoint Online, SharePoint On-Premise, and OneDrive.

Note that if you see SIDs (Security IDentifiers) instead of user and group names, you should integrate your

Active Directory into BlueXP classification. [See how to do this.](#)

File Name

Personal Sensitive Personal Data Subjects File Type

Expense Report TPO-1060.pdf

cvo 6 3 16 PDF

Working Environment: WorkingEnvironment1

Repository: Volume Name

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

Category: Legal

File Size: 22 MB

Last Modified: 2019-08-06 07:51

Open Permissions: NO OPEN PERMISSIONS

File Owner: Avy

Assign a Label to this file

Delete this file

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

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

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

Additionally, you can click the name of a user or a group and the Investigation page is displayed with the name of that user or group populated in the “User / Group Permissions” filter so you can see all the files and directories that the user or group has access to.

Check for duplicate files in your storage systems

You can view if duplicate files are being stored in your storage systems. This is useful if you want to identify areas where you can save storage space. It can also be helpful to make sure certain files that have specific permissions or sensitive information are not unnecessarily duplicated in your storage systems.

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

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


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

View all duplicated files

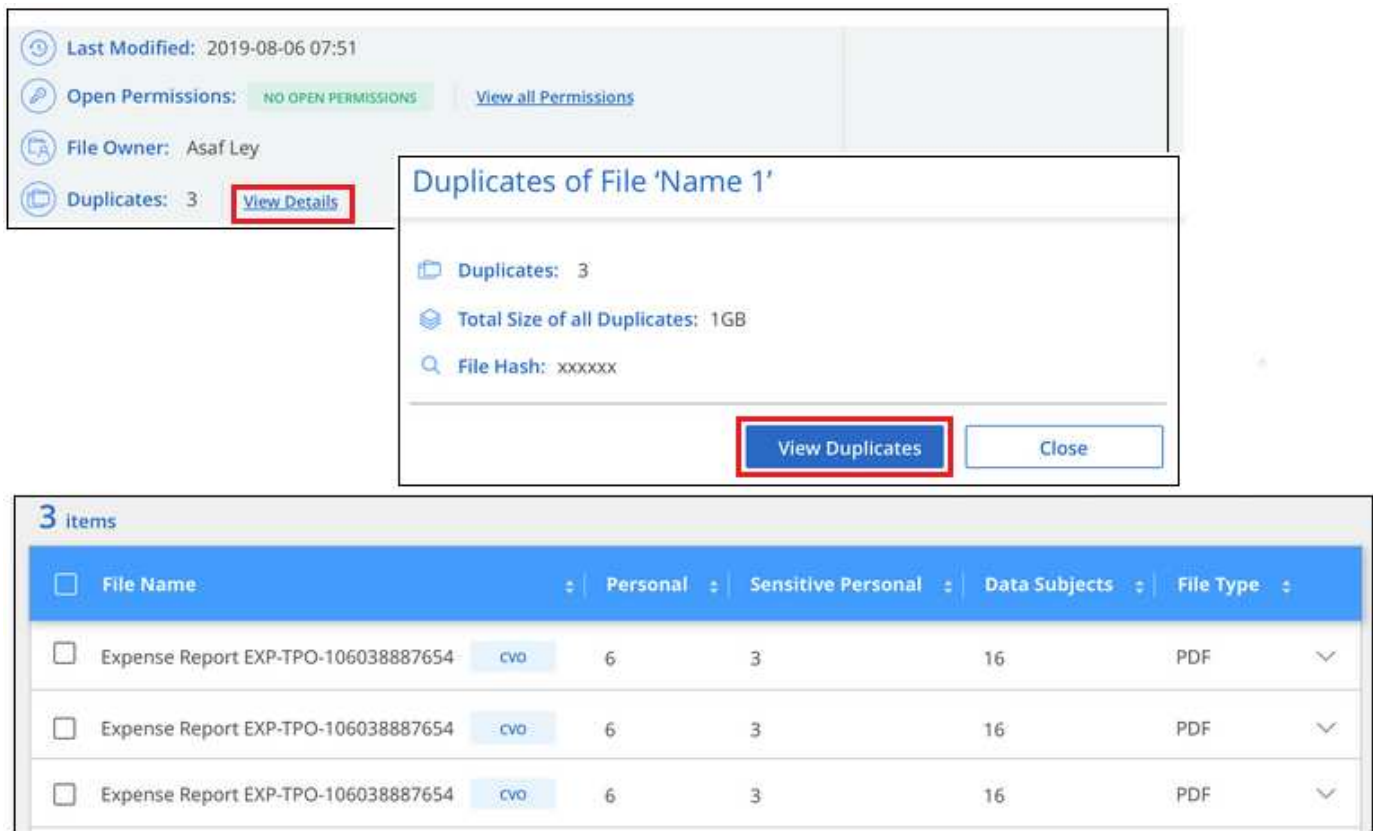
If you want a list of all files that are duplicated in the working environments and data sources you are scanning, you can use the filter called **Duplicates > Has duplicates** in the Data Investigation page.

All duplicated files are displayed in the Results page.

View if a specific file is duplicated

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

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



The screenshot shows the 'File Owner' section with fields for 'Last Modified', 'Open Permissions', 'File Owner', and 'Duplicates'. The 'Duplicates' field shows '3' and a 'View Details' button. A modal window titled 'Duplicates of File 'Name 1'' is open, displaying 'Duplicates: 3', 'Total Size of all Duplicates: 1GB', and 'File Hash: xxxxxx'. It has 'View Duplicates' and 'Close' buttons. Below the modal is a table with 3 items.

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



You can use the "file hash" value provided in this page and enter it directly in the Investigation page to search for a specific duplicate file at any time - or you can use it in a Policy.

Data Investigation Report

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

The report is available in two different formats:

- As a .CSV file that you can save to the local machine.

This report can include a maximum of 10,000 rows of data.

- As a .JSON file that you export to an NFS Share.


If there are more than 250,000 rows of data, additional .JSON files are created.

When exporting to a file share, make sure BlueXP classification has the correct permissions for export access.

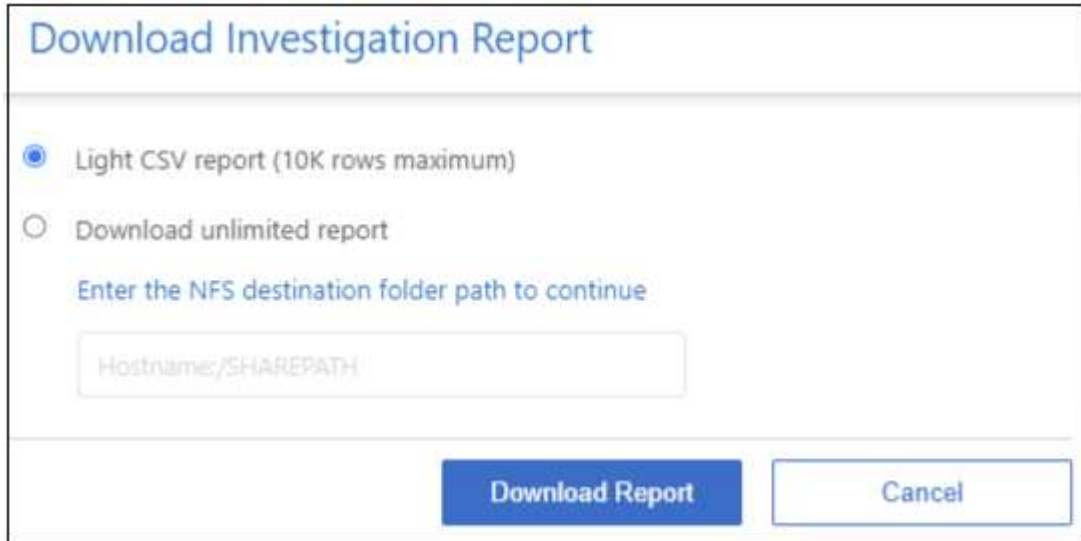
There can be up to three report files downloaded if BlueXP classification is scanning files (unstructured data), directories (folders and file shares), and databases (structured data).

Generate the Data Investigation Report

Steps

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

When selecting a .JSON report, enter the name of the NFS share where the report will be downloaded in the format <host_name>:/<share_path>.



The dialog box titled "Download Investigation Report" contains two radio button options. The first option, "Light CSV report (10K rows maximum)", is selected. The second option is "Download unlimited report". Below these options is a text input field with the placeholder text "Enter the NFS destination folder path to continue" and "Hostname:/SHAREPATH". At the bottom right of the dialog are two buttons: "Download Report" and "Cancel".

Result

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

You can view the progress of JSON report generation in the [Actions Status pane](#).

What's included in each Data Investigation Report

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

- File name
- Location type
- Working environment name
- Storage repository (for example, a volume, bucket, shares)
- Repository type
- File path
- File type
- File size (in MB)
- Created time
- Last modified

- Last accessed
- File owner
- Category
- Personal information
- Sensitive personal information
- Open permissions
- Scan Analysis Error
- Deletion detection date

A deletion detection date identifies the date that the file was deleted or moved. This enables you to identify when sensitive files have been moved. Deleted files aren't part of the file number count that appears in the dashboard or on the Investigation page. The files only appear in the CSV reports.

The **Unstructured Directories Data Report** includes the following information about your folders and file shares:

- Working environment type
- Working environment name
- Directory name
- Storage repository (for example, a folder or file shares)
- Directory owner
- Created time
- Discovered time
- Last modified
- Last accessed
- Open permissions
- Directory type

The **Structured Data Report** includes the following information about your database tables:

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

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.

- 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.



The December 2023 (v1.26.6) release temporarily removed the option to integrate data using Azure Information Protection (AIP) labels.

- You can add Tags to files that you want to mark for organization or for some type of follow-up.
- You can assign a BlueXP user to a specific file, or to multiple files, so that person can be responsible for managing the file.
- Using the "Policy" functionality you can create your own custom search queries so that you can easily see the results by clicking one button.
- You can send email alerts to BlueXP users, or any other email address, when certain critical Policies return results.



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

Should I use tags or labels?

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

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

Categorize your data using AIP labels

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

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



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

Integrate AIP labels in your workspace

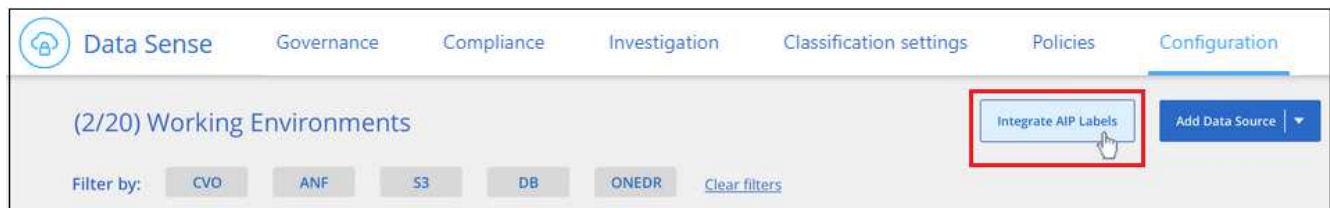
Before you can manage AIP labels, you need to integrate the AIP label functionality into BlueXP classification by signing into your existing Azure account. Once enabled, you can manage AIP labels within files for all [data sources](#) in your BlueXP workspace.

Requirements

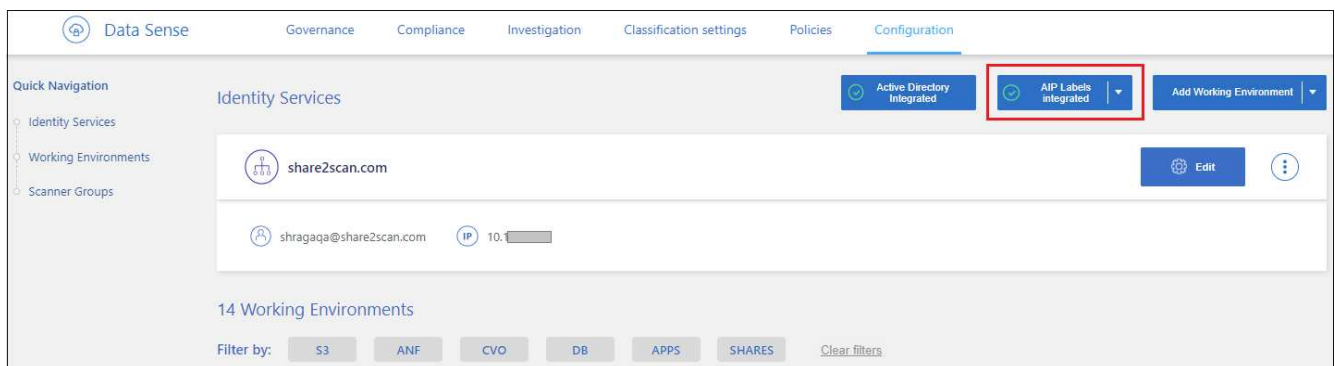
- You must have an account and an Azure Information Protection license.
- You must have the login credentials for the Azure account.
- If you plan to change labels in files that reside in Amazon S3 buckets, ensure that the permission `s3:PutObject` is included in the IAM role. See [setting up the IAM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

Steps

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



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



Result

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

labels to files using Policies.

View AIP labels in your files

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

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



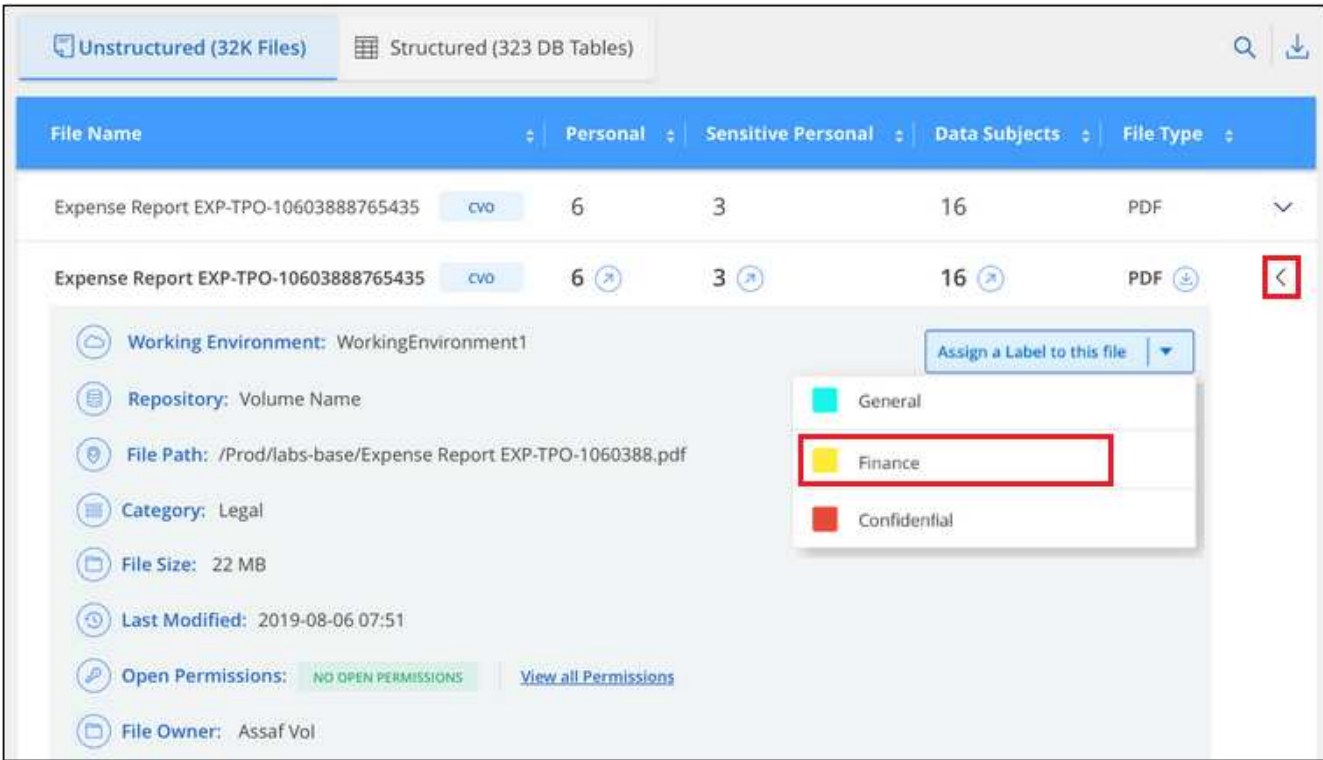
Assign AIP labels manually

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

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

Steps

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



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

The label appears in the file metadata.

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

Steps

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

255 items 1.2 GB | 2 Selected 3 MB

Tags

Assign to

Label

Copy

Move

Delete

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

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

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



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

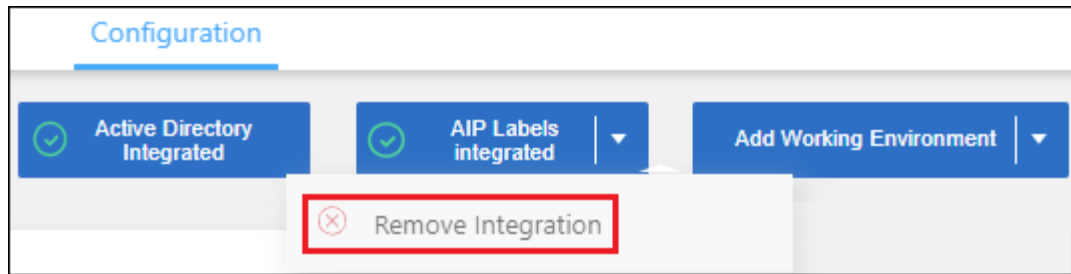
Remove the AIP integration

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.

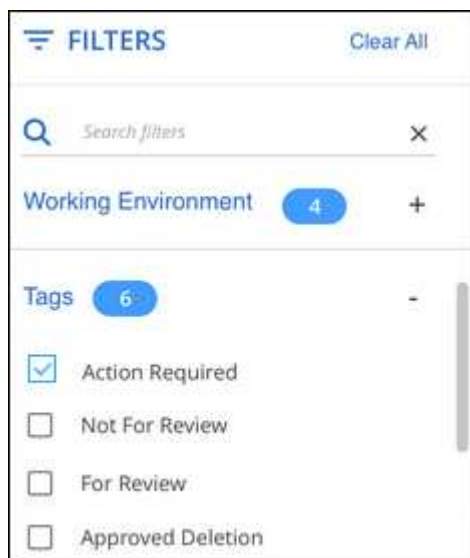


Tags assigned to files in BlueXP classification are not related to the tags you can add to resources, such as volumes or virtual machine instances. BlueXP classification tags are applied at the file level.

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.

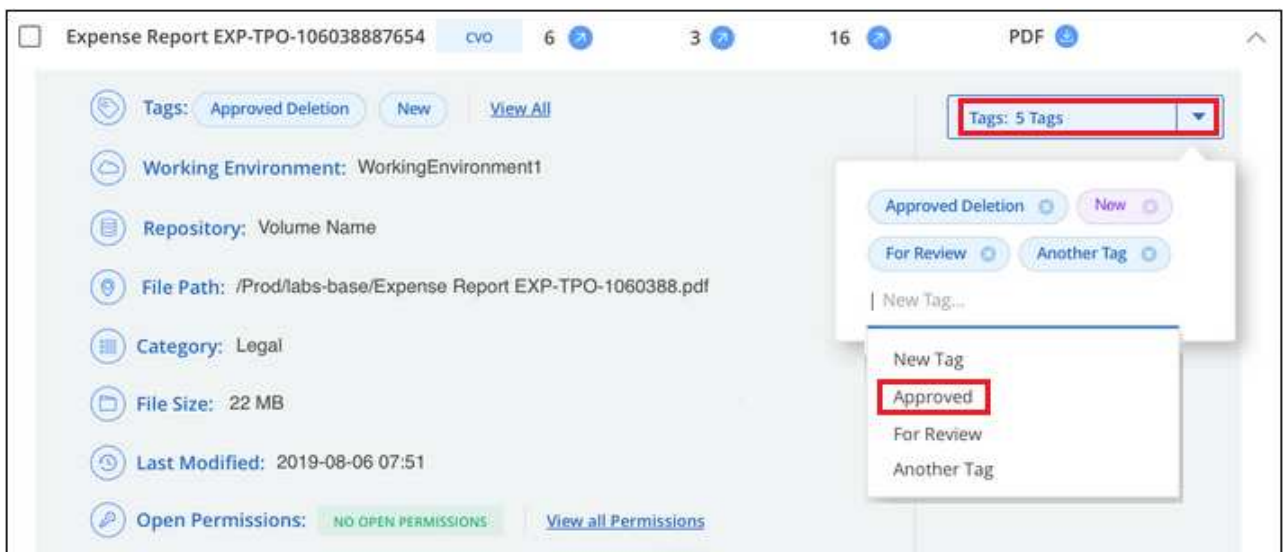
Assign tags to files

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

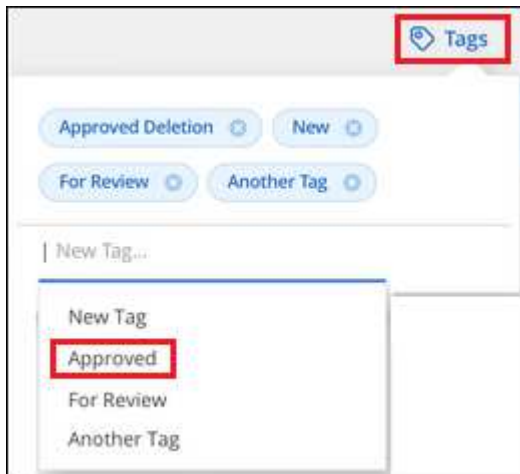
1. In the Data Investigation results pane, select the file, or files, that you want to tag.

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.

- From the button bar, click **Tags** and the currently assigned tags are displayed.
- 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**.



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

Delete tags from files

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

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

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

Assign users to manage certain files

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


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

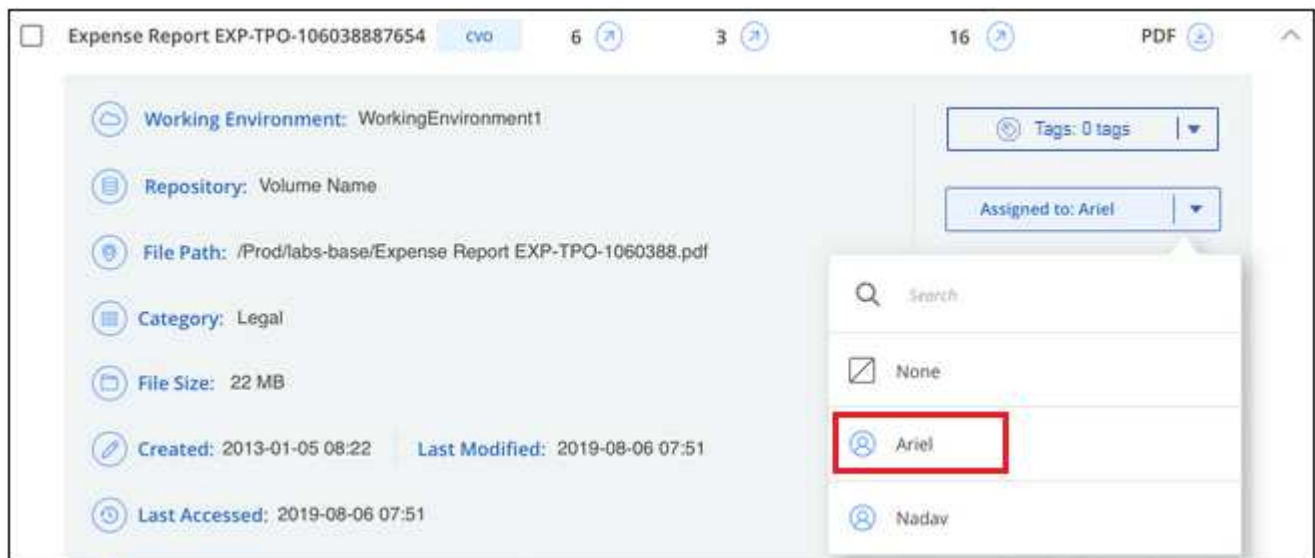
Note that the user name is not added to the file as part of the file metadata - it is just seen by BlueXP users when using BlueXP classification.

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

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

Steps

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



The User name appears in the file metadata.

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

Steps

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

255 items 1.2 GB | 2 Selected 3 MB

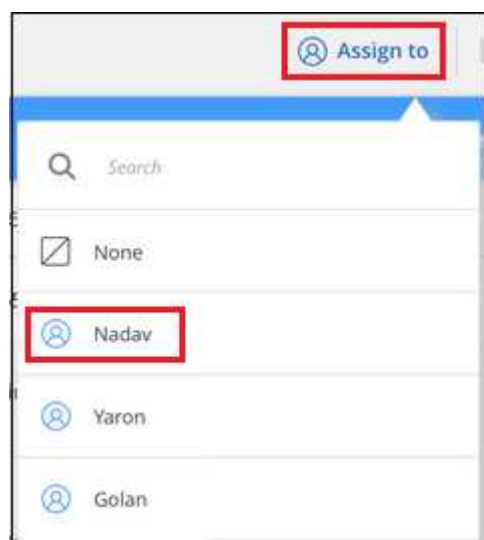
Tags Assign to Label Copy Move Delete

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

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

◦ To select all files on the current page, check the box in the title row (☒ File Name).

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



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

Assign policies to your data

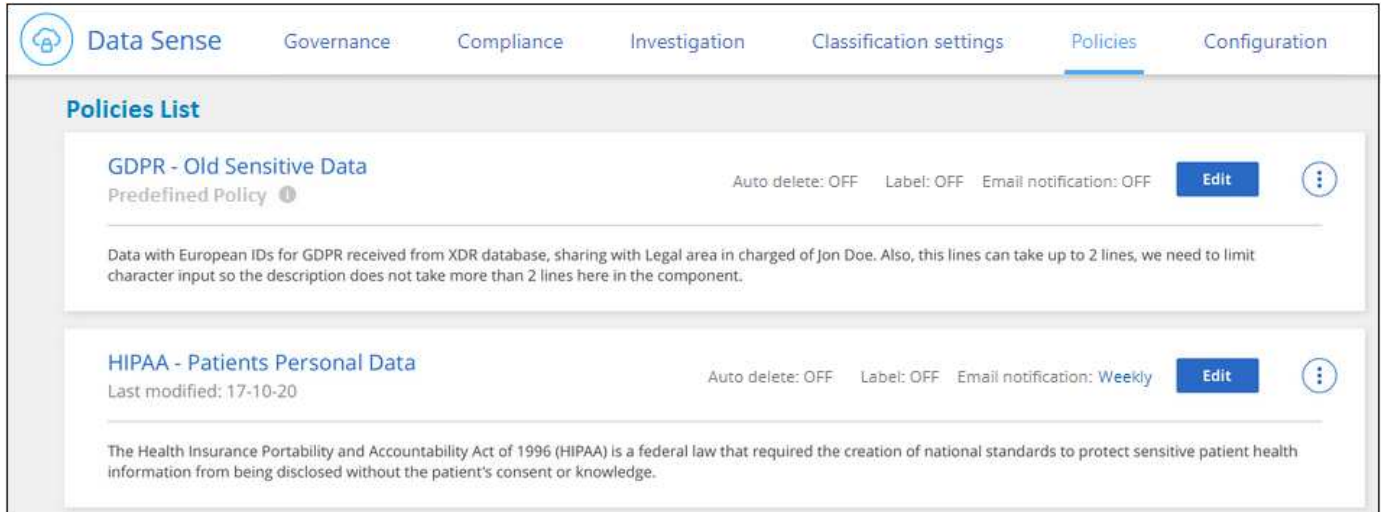
Policies are like a favorites list of custom filters that provide search results in the Investigation page for commonly requested compliance queries. BlueXP classification provides a set of predefined Policies based on common customer requests. You can create custom Policies that provide results for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

Policies provide the following functionality:

- [Predefined Policies](#) from NetApp based on user requests
- Ability to create your own custom Policies
- Launch the Investigation page with the results from your Policies in one click
- Send email alerts to BlueXP users, or any other email addresses, when certain critical Policies return results so you can get notifications to protect your data

- Assign AIP (Azure Information Protection) labels automatically to all files that match the criteria defined in a Policy
- Delete files automatically (once per day) when certain Policies return results so you can protect your data automatically

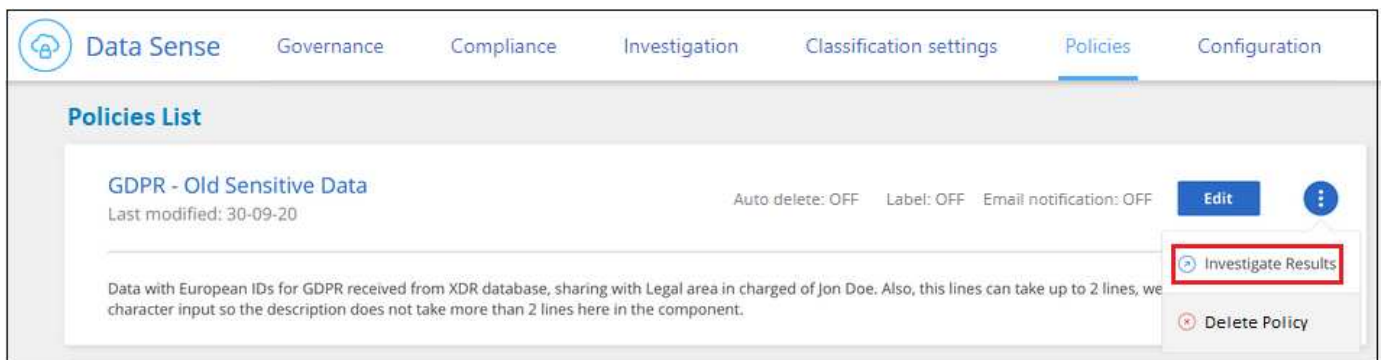
The **Policies** tab in the Compliance Dashboard lists all the predefined and custom Policies available on this instance of BlueXP classification.



In addition, Policies appear in the list of Filters in the Investigation page.

View Policy results in the Investigation page

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



Create custom Policies

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

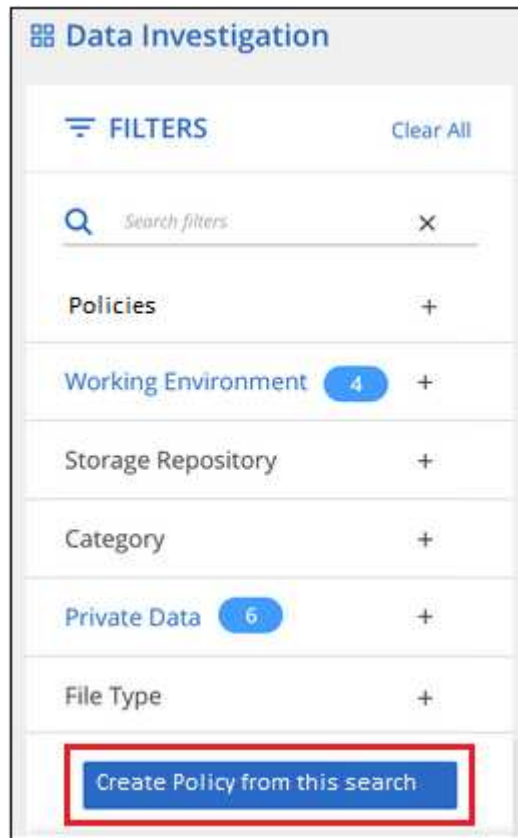
Note that the actions for deleting data and assigning AIP labels based on the policy results are valid only for files. Directories that match the search criteria can't be deleted automatically or assigned AIP labels.

Steps

1. From the Data Investigation page, define your search by selecting all the filters you want to use. See

Filtering data in the [Data Investigation](#) page for details.

2. Once you have all the filter characteristics just the way you want them, click **Create Policy from this search**.



The screenshot shows the 'Data Investigation' interface. At the top, there's a header with a grid icon and the text 'Data Investigation'. Below this is a 'FILTERS' section with a 'Clear All' link. A search bar labeled 'Search filters:' with a magnifying glass icon and a close 'X' button is present. Below the search bar, there are several filter categories, each with a plus sign to expand it: 'Policies', 'Working Environment' (which shows a blue pill with the number '4'), 'Storage Repository', 'Category', 'Private Data' (which shows a blue pill with the number '6'), and 'File Type'. At the bottom of the filters section, a blue button with the text 'Create Policy from this search' is highlighted with a red rectangular border.

3. Name the Policy and select other actions that can be performed by the Policy:
 - a. Enter a unique name and description.
 - b. Optionally, check the box to automatically delete files that match the Policy parameters. Learn more about [deleting source files using a policy](#).
 - c. Optionally, check the box if you want notification emails sent to BlueXP users in your account, and choose the interval at which the email is sent. Learn more about [sending email alerts based on policy results](#).
 - d. Optionally, check the box if you want notification emails sent to other users, enter up to 20 email addresses, and choose the interval at which the email is sent.
 - e. Optionally, check the box to automatically assign AIP labels to files that match the Policy parameters, and select the label. (Only if you have already integrated AIP labels. Learn more about [AIP labels](#).)
 - f. Click **Create Policy**.

Create Policy

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

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

Name this Policy

Files with personal data created over 60 days ago

Give it a detailed description that explains what it searches for

See if any old files with personal data should be moved or deleted

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

Email updates about this Policy:

☒ Email all the users in this account Every Day

☐ Send Email Every Day to:

Label:

☐ Automatically label this Policy's matches with: New Personal

Cancel
Create Policy

Result

The new Policy appears in the Policies tab.

Send email alerts when non-compliant data is found

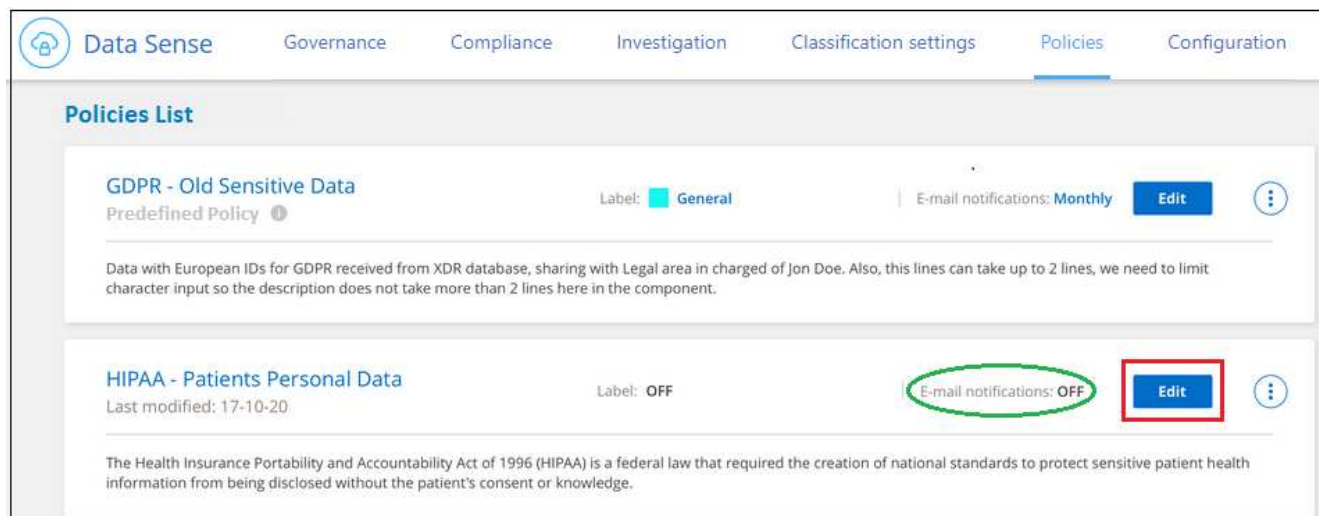
BlueXP classification can send email alerts to BlueXP users in your account when certain critical Policies return results so you can get notifications to protect your data. You can choose to send the email notifications on a daily, weekly, or monthly basis. You can also choose to send email alerts to any other email address - up to 20 email addresses - not in your BlueXP account.

You can configure this setting when creating the Policy or when editing any Policy.

Follow these steps to add email updates to an existing Policy.

Steps

1. From the Policies List page, click **Edit** for the Policy where you want to add (or change) the email setting.



2. In the Edit Policy page:

- Check the "Email all the users in this account" box if you want notification emails sent to users in your BlueXP account, and choose the interval at which the email is sent (for example, **Every Day**).
- Check the "Send Email" box if you want notification emails sent to additional users, choose the interval at which the email is sent, and enter up to 20 email addresses.

The screenshot shows the 'Edit Policy' page. It includes a header 'Edit Policy' and a sub-header 'Saving this filtered view will create a new Policy, you can view/edit it in the "Policy" tab'. The form contains several fields: 'Name this Policy' (with the value 'HIPAA - Patient Personal Data'), 'Give it a description to quickly identify it' (with the value 'Files containing patient health information that is more than 30 days old'), and a section for 'Email updates about this Policy:'. In this section, the 'Email all the users in this account' checkbox is checked, and the 'Send Email' checkbox is also checked. The 'Send Email' section shows a dropdown for 'Every Day' and a text input for 'to:' with the value 'email@gmail.com'. At the bottom, there is a 'Label:' section with a checkbox for 'Automatically label this Policy's matches with: New Personal'. The 'Save Policy' button is highlighted with a red box.

3. Click **Save Policy** and the interval at which the email is sent appears in the Policy description.

Result

The first email is sent now if there are any results from the Policy - but only if any files meet the Policy criteria. No personal information is sent in the notification emails. The email indicates that there are files that match the Policy criteria, and it provides a link to the Policy results.

Delete source files automatically using Policies

You can create a custom Policy to delete files that match the policy. For example, you may want to delete files that contain sensitive information and were discovered by BlueXP classification in the past 30 days.

Only Account Admins can create a policy to automatically delete files.



All files that match the policy will be permanently deleted once a day.

Steps

1. From the Data Investigation page, define your search by selecting all the filters you want to use. See [Filtering data in the Data Investigation page](#) for details.
2. Once you have all the filter characteristics just the way you want them, click **Create Policy from this search**.
3. Name the Policy and select other actions that can be performed by the Policy:
 - a. Enter a unique name and description.
 - b. Check the box to "Automatically delete files that match this policy" and type **permanently delete** to confirm that you want files permanently deleted by this policy.
 - c. Click **Create Policy**.

Create Policy

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

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

Name this Policy

Give it a detailed description that explains what it searches for

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

Type "permanently delete" to continue with the deletion.

☐ Send email updates about this Policy to Cloud Manager users on this account every

☐ Automatically label this Policy's matches with:

Result

The new Policy appears in the Policies tab. Files that match the policy are deleted once per day when the policy runs.

You can view the list of files that have been deleted in the [Actions Status pane](#).

Assign AIP labels automatically with Policies

You can assign an AIP label to all the files that meet the criteria of the Policy. You can specify the AIP label when creating the Policy, or you can add the label when editing any Policy.

Labels are added or updated in files continuously as BlueXP classification scans your files.

Depending on whether a label is already applied to a file, and the classification level of the label, the following actions are taken when changing a label:

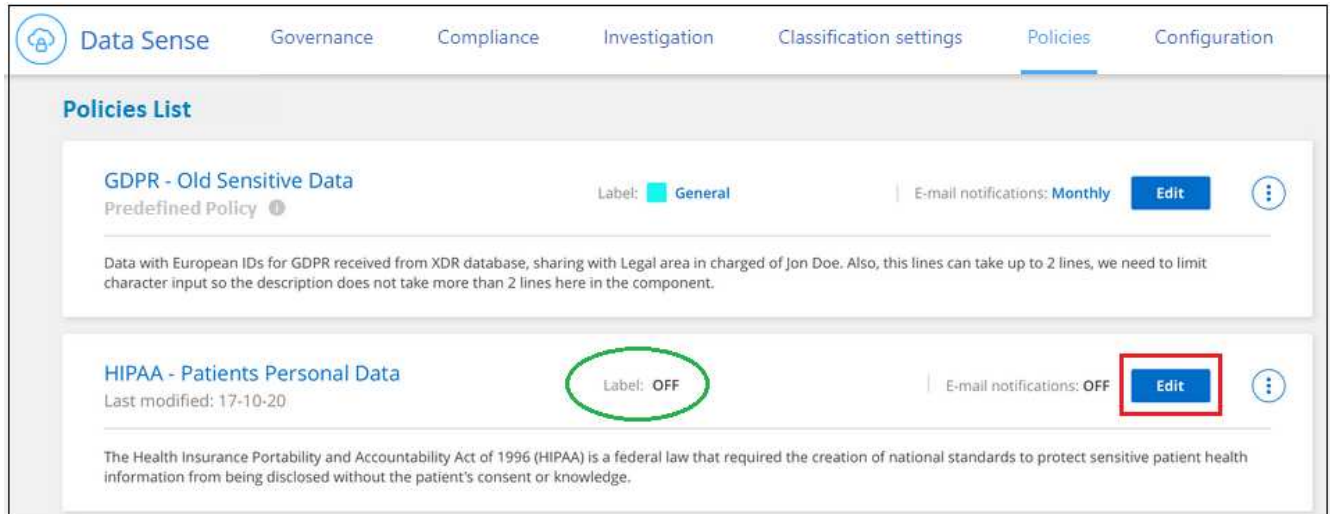
If the file...	Then...
Has no label	The label is added
Has an existing label of a lower level of classification	The higher level label is added
Has an existing label of a higher level of classification	The higher level label is retained

If the file...	Then...
Is assigned a label both manually and by a Policy	The higher level label is added
Is assigned two different labels by two Policies	The higher level label is added

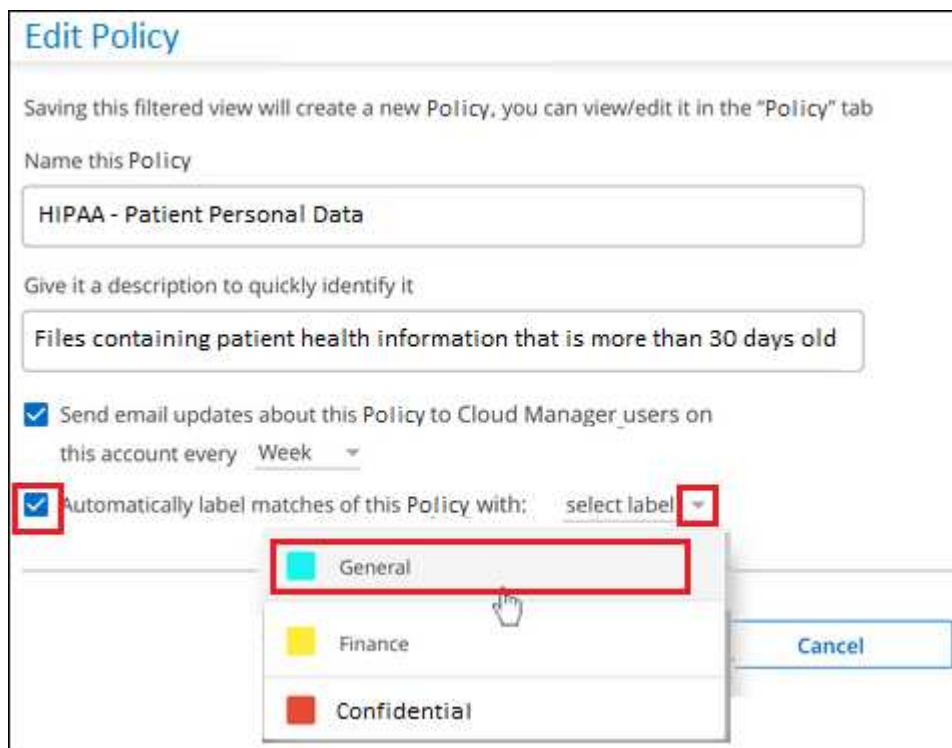
Follow these steps to add an AIP label to an existing Policy.

Steps

1. From the Policies List page, click **Edit** for the Policy where you want to add (or change) the AIP label.



2. In the Edit Policy page, check the box to enable automatic labels for files that match the Policy parameters, and select the label (for example, **General**).



3. Click **Save Policy** and the label appears in the Policy description.



If a Policy was configured with a label, but the label has since been removed from AIP, the label name is turned to OFF and the label is not assigned anymore.

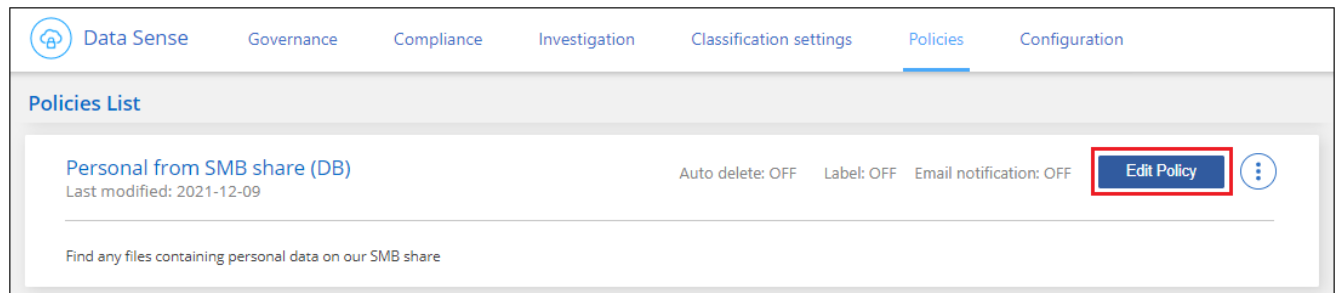
Edit Policies

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

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

Steps

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



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

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

Edit Policy

Edit Query

Name this Policy

Personal from SMB share (DB)

Give it a detailed description that explains what it searches for:

Find any files containing personal data on our SMB share

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

Email updates about this Policy:

☐ Email all the users in this account

Every Day

☐ Send Email

Every Day

 to:

Label:

☐ Automatically label this Policy's matches with:

Cancel

Save Policy

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

Data Investigation

Unstructured (16 Files)

Directories (0 Folders)

Structured (0 Tables)

Search by File, Table or loca

FILTERS:

Clear All

Policies 1

Open Permissions

User / Group Permissions

File Owner

Label

Working Environment Type

Working Environment

Save Changes

Cancel Edit Query

16 items | 250.2 MB

Tags

Assign to

Label

Move

Copy

Delete

<input type="checkbox"/>	File Name		Personal	Sensitive Personal	Data Subjects	File Type
<input type="checkbox"/>	cifs2.json	SHARES	1	0	0	JSON
<input type="checkbox"/>	cifs12.json	SHARES	1	0	0	JSON
<input type="checkbox"/>	TableTextServiceYi.txt	SHARES	1	0	0	TXT
<input type="checkbox"/>	testpass.json	SHARES	1	0	0	JSON
<input type="checkbox"/>	urlp.txt	SHARES	1	0	0	TXT
<input type="checkbox"/>	License.sharpen.txt	SHARES	1	0	1	TXT
<input type="checkbox"/>	TableTextServiceYi.txt	SHARES	1	0	0	TXT
<input type="checkbox"/>	Notice.txt	SHARES	1	0	0	TXT
<input type="checkbox"/>	urlp.txt	SHARES	1	0	0	TXT
<input type="checkbox"/>	Notice.txt	SHARES	1	0	0	TXT

1-16 of 16

Result

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

Delete Policies

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

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

List of predefined Policies

BlueXP classification provides the following system-defined Policies:

Name	Description	Logic
S3 publicly - Exposed private data	S3 Objects containing personal or sensitive personal information, with open Public read access.	S3 Public AND contains personal OR sensitive personal info
PCI DSS - Stale data over 30 days	Files containing Credit Card information, last modified over 30 days ago.	Contains credit card AND last modified over 30 days
HIPAA - Stale data over 30 days	Files containing Health information, last modified over 30 days ago.	Contains health data (defined same way as in HIPAA report) AND last modified over 30 days
Private data - Stale over 7 years	Files containing personal or sensitive personal information, last modified over 7 years ago.	Files containing personal or sensitive personal information, last modified over 7 years ago
GDPR - European citizens	Files containing more than 5 identifiers of an EU country's citizens or DB Tables containing identifiers of an EU country's citizens.	Files containing over 5 identifiers of an (one) EU citizens or DB Tables containing rows with over 15% of columns with one country's EU identifiers. (any one of the national identifiers of the European countries. Does not include Brazil, California, USA SSN, Israel, South Africa)
CCPA - California residents	Files containing over 10 California Driver's License identifiers or DB Tables with this identifier.	Files containing over 10 California Driver's License identifiers OR DB Tables containing California Driver's license
Data Subject names - High risk	Files with over 50 Data Subject names.	Files with over 50 Data Subject names
Email Addresses - High risk	Files with over 50 Email Addresses, or DB Columns with over 50% of their rows containing Email Addresses	Files with over 50 Email Addresses, or DB Columns with over 50% of their rows containing Email Addresses

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

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.

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



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

Copy source files

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

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

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

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

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

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

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

Copy source files to an NFS share

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



You can't copy files that reside in databases.

Requirements

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

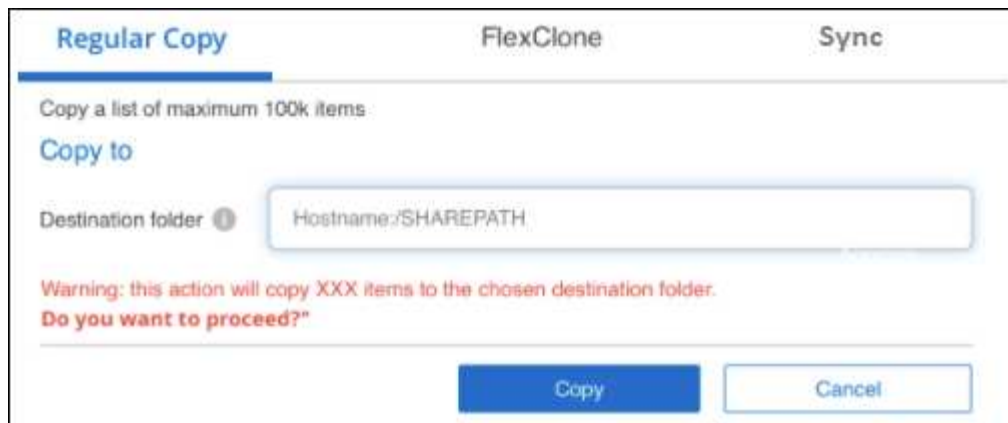
Steps

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

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

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

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



Regular Copy FlexClone Sync

Copy a list of maximum 100k items

Copy to

Destination folder ⓘ Hostname:/SHAREPATH

Warning: this action will copy XXX items to the chosen destination folder.
Do you want to proceed?"

Copy Cancel

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**.



Unstructured (32K Files) Structured (323 DB Tables)

File Name	Personal	Sensitive Personal	Data Subjects	File Type
<input type="checkbox"/> Expense Report EXP-TPO-1060388765435	cvo 6	3	16	PDF
<input type="checkbox"/> Expense Report EXP-TPO-1060388765435	cvo 6	3	16	PDF

Working Environment: WorkingEnvironment1

Repository: Volume Name

File Path: /Prod/labs-base/Expense Report EXP-TPO-1060388.pdf

Assign a Label to this file

Move File

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.

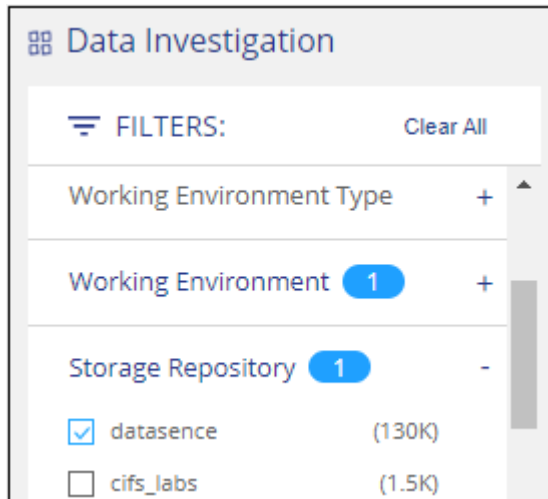
Requirements

- 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)

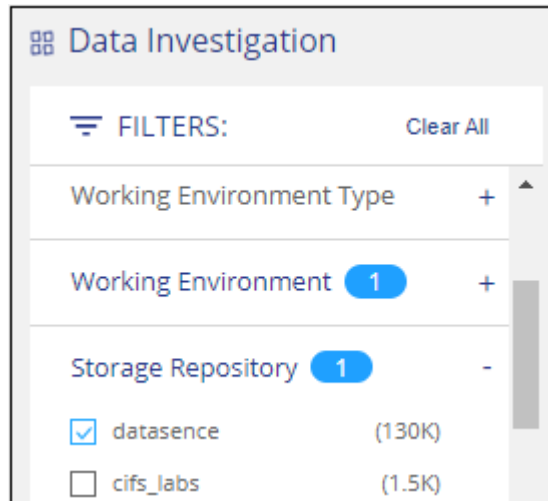
share, etc.).

- 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.

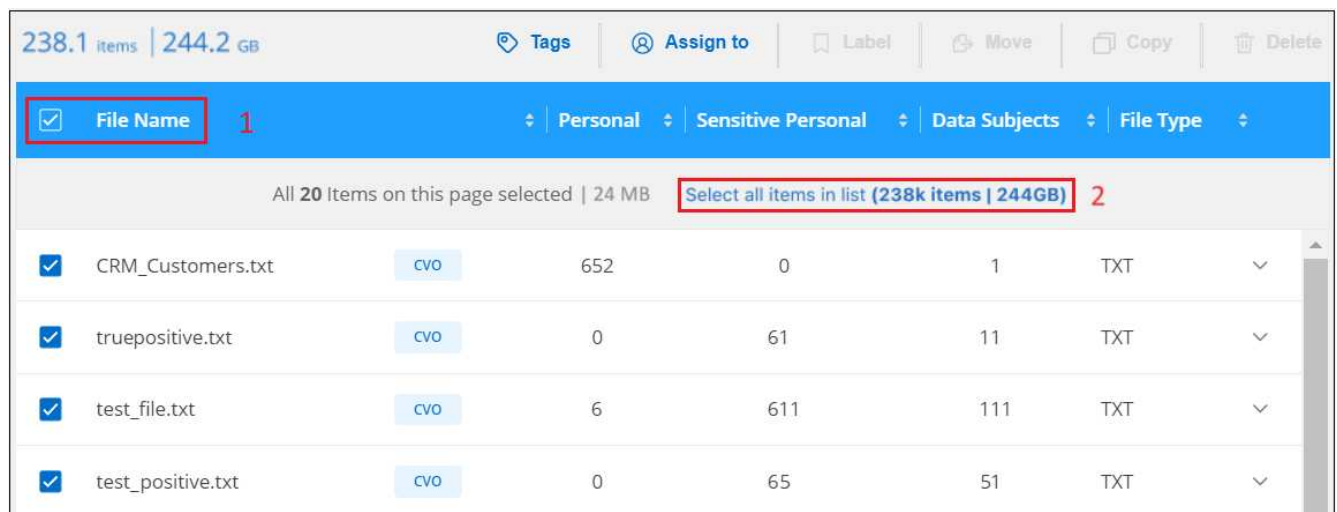
Steps

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.

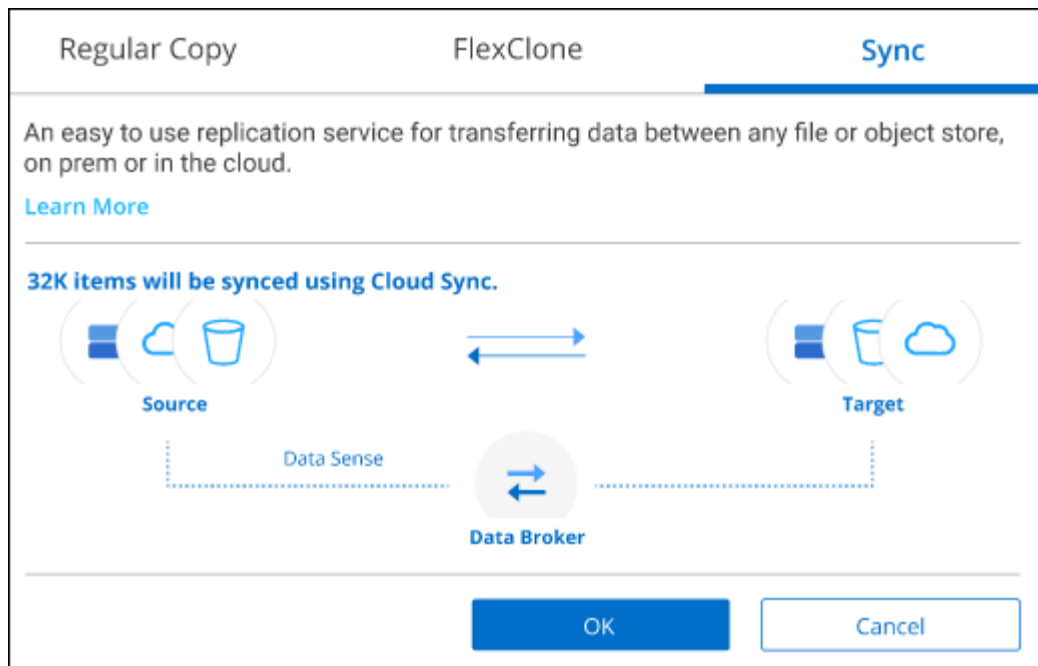


Apply any other filters so that you're seeing only the files that you want to copy and sync to the destination system.

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.

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)**.

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

☒ **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**.

 **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**.



Monitor and manage file access events

You can configure BlueXP classification to log events to a file whenever one of the files you are scanning has been accessed or changed. Then you can view the contents of this log file (the File Access Audit Log), or download it, to see what file changes have occurred, and by whom.

For example, if you want to track when any changes are made to your sensitive personnel files or payroll files, you can enable this feature on the volumes where these files reside. Then you can view the File Access Audit Log to see if these files have been changed. If they have been changed, you'll be able to see the security identifier (SID) of the person who made the change and when the change was made.

You can enable this feature on any critical volumes in your working environments that are running ONTAP software. When enabled on a volume, file access events are tracked for all files and directories in the volume. This capability is based on the [FPolicy feature](#) from your ONTAP systems, and it uses your BlueXP classification system as an FPolicy server to receive notification messages (events) from ONTAP. BlueXP classification can process approximately 990 events per second from ONTAP.



The December 2023 (version 1.26.6) release temporarily removed the option to activate audit log collection.

BlueXP classification currently captures events for the following actions on your files and directories:

- Create
- Read
- Write
- Rename
- Delete
- Change owner/permissions
- Change SACL/DACL (access control list (ACL) changes)

BlueXP users with the "Account Admin" or "Workspace Admin" roles can configure volumes for file access audit logging and can view and download the audit logs. Users with the "Compliance Viewer" role can only view and download the audit logs.

Supported data sources

BlueXP classification can log file access events for files that reside on ONTAP volumes in the following data sources:

- On-premises ONTAP
- Cloud Volumes ONTAP
- Amazon FSx for NetApp ONTAP

The data sources must be running ONTAP version 9.11 or greater software.

Configure volumes for file access audit logging

You can enable file access audit logging on individual volumes to track changes to files and directories.



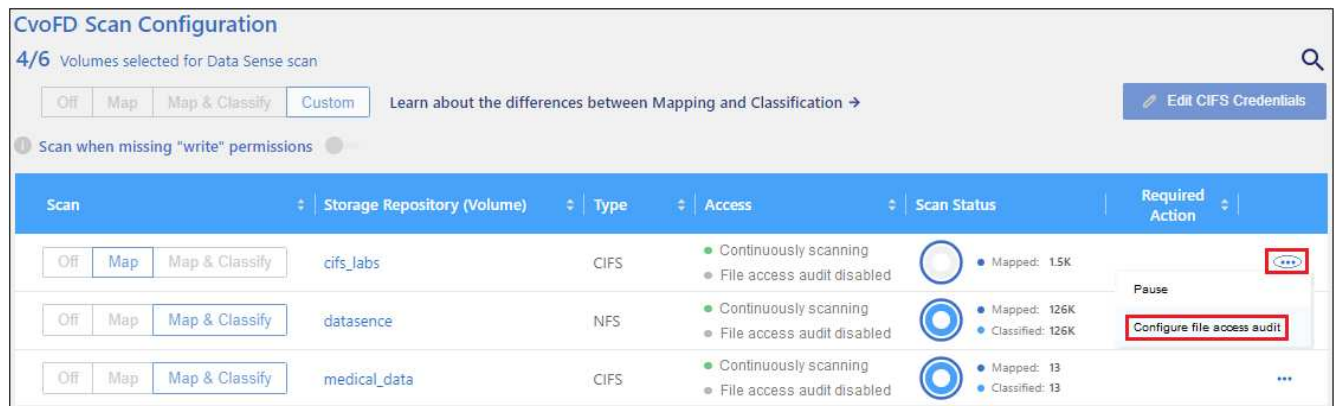
You can configure file access auditing on a maximum of 8 volumes on each data source at this time.

Before you begin

- You should have Active Directory configured for the data sources so that BlueXP classification can list the ID of the person who accessed the files.
- Port 5018 must be open for outbound access from the BlueXP classification system.

Steps

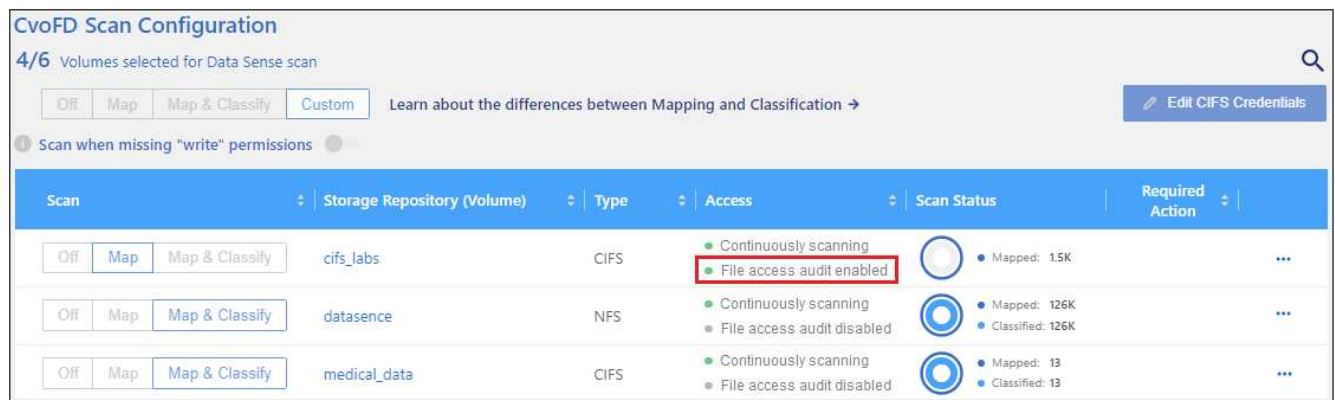
1. From the *Scan Configuration* page for the data source, click **...** for the volume and select **Configure File Access Audit**.



2. In the *Configure File Access Audit* dialog, click the checkbox for **Enable logging of file access events** and click **Save**.

Note at this time that all the potential logging actions are selected by default — they are not editable.

3. You'll see that file access auditing is now enabled for that volume.



Result

Any file access events that are generated for files on the enabled volumes will be added to the log file. Check the log file occasionally to see they types of events you are receiving.

Log file contents

A log file is created for every volume that you have configured to track file access events. The log files are named "<volume_name>_<volume_uuid>_log"; for example "fpolicy_cifs_838c1727-dd2d-11ed-a3ec-590ce4991.log". Each line in the audit log contains information in this format:

- Timestamp - date and time of the event
- Client IP - IP address of the instance/PC/Proxy in which the file operation was performed
- Volume Name - name of the volume
- Volume UUID - UUID of the volume
- File Type - type of file: FILE or DIR
- File Size - size of file in bytes

- Path - full path and name of the affected file or directory
- Volume Type - type of volume: SMB or NFS
- User ID - security identifier (SID) of the person who performed the action
- File Owner ID - security identifier (SID) of the file owner
- Event Type - Create, Read, Write, Rename, Delete, Change owner/permissions, or Change SACL/DACL
- Action Details - what was done: depends on the action

For example, the following line from the log file shows that a "Create" action has occurred in the volume "fpolicy_cifs" - a new file "f14" has been created in the volume.

```
{ "Timestamp": "2023-04-24 13:57", "Client_IP": "172.31.14.35",
  "Volume_Name": "fpolicy_cifs", "Volume_UUID": "838c1727-dd2d-11ed-a3ec-590ce4991", "File_Type": "FILE", "File_Size": 100, "Path":
  "\\FPOLICY_CVO\\fpolicy_cifs_share\\dbs\\f14, "Volume_Type": "SMB", "User_ID":
  "S-1-5-21-459977447-2546672318-3630509715-500", "File_Owner_ID": "S-1-5-32-544", "Event_Type": "CREATE", "Action_Details": {details}}
```

You can use the BlueXP classification Investigation page to search for the volume (using the "Storage Repository" filter) or file (using the "File / Directory Path" filter) to see more details about the affected volume and file.

Access the File Access Audit Log files

The File Access Audit Log files are located on the BlueXP classification machine in:
/opt/netapp/fpolicy/logs

Each file is configured by default to contain a maximum of 50,000 events. [You can customize this value in the File Access Audit Log Configuration page](#). After this maximum has been reached, older entries in the log file are overwritten.

The total size of all the log files in the directory is set by default to a maximum of 50 GB. [You can customize this value in the File Access Audit Log Configuration page](#). When that limit is reached, the oldest log files are deleted as new log files are added. Additionally, any log files that are older than 14 days will be overwritten as that is the maximum retention time.

When BlueXP classification is installed on a Linux machine on your premises, or on a Linux machine you deployed in the cloud, you can navigate directly to the log files.

When BlueXP classification is deployed in the cloud, you'll need to SSH to the BlueXP classification instance. You SSH to the system by entering the user and password, or by using the SSH key you provided during the BlueXP Connector installation. The SSH command is:

```
ssh -i <path_to_the_ssh_key> <machine_user>@<datasense_ip>
```

- <path_to_the_ssh_key> = location of ssh authentication keys
- <machine_user>:

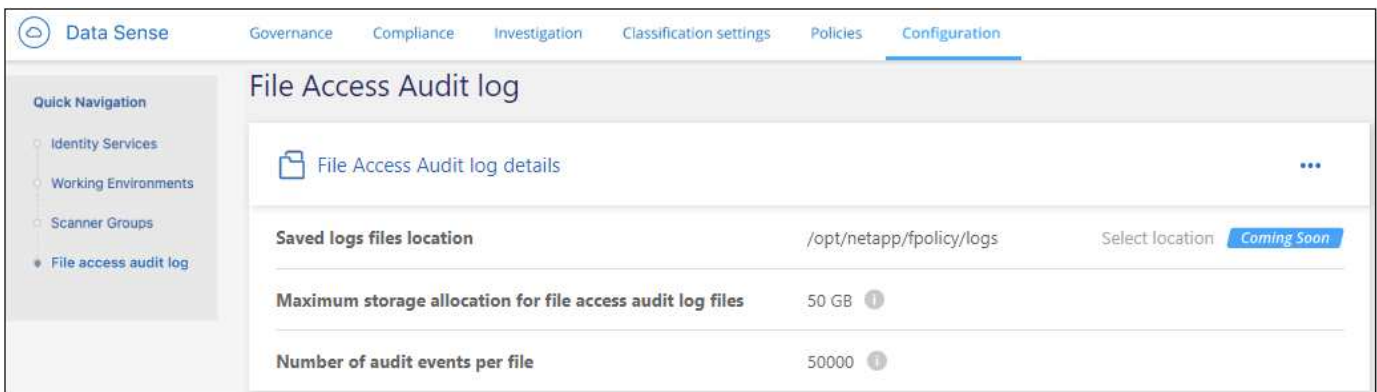
- For AWS: use the <ec2-user>
- For Azure: use the user created for the BlueXP instance
- For GCP: use the user created for the BlueXP instance
- <datasense_ip> = IP address of the BlueXP classification virtual machine instance

Note that you'll need to modify the security group inbound rules to access the system in the cloud. For details, see:

- [Security group rules in AWS](#)
- [Security group rules in Azure](#)
- [Firewall rules in Google Cloud](#)

Configure File Access Audit Log settings

There are three options that you can configure for the file access audit file logs. These settings apply to all data sources that have configured file access audit logging on this BlueXP classification instance. You configure these settings from the *File Access Audit Log* section of the BlueXP classification *Configuration* page.



Audit Log Option	Description
Log file location	The location is currently hardcoded to write the log files to <code>/opt/netapp/fpolicy/logs</code>
Maximum storage allocation for audit logs	The total size of all the log files in the directory is currently hardcoded to a default value of 50 GB. When that limit is reached, the oldest log files are deleted automatically.
Maximum number of audit events per audit file	Each file is currently hardcoded to contain a maximum of 50,000 events. After this maximum has been reached, old events are deleted as new events are added.

Note that these settings are currently hardcoded to default settings. They can't be changed.

View compliance reports

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

By default, the BlueXP classification dashboards display compliance and governance data for all working

environments, databases, and data sources. If you want to view reports that contain data for only some of the working environments, [select those working environments](#).



- The reports described in this section are available only if you have chosen to perform a full classification scan on your data sources. Data sources that have had a mapping-only scan can only generate the Data Mapping Report.
- NetApp can't guarantee 100% accuracy of the personal data and sensitive personal data that BlueXP classification identifies. You should always validate the information by reviewing the data.

Privacy Risk Assessment Report

The Privacy Risk Assessment Report provides an overview of your organization's privacy risk status, as required by privacy regulations such as GDPR and CCPA. The report includes the following information:

Compliance status

A [severity score](#) and the distribution of data, whether it's non-sensitive, personal, or sensitive personal.

Assessment overview

A breakdown of the types of personal data found, as well as the categories of data.

Data subjects in this assessment

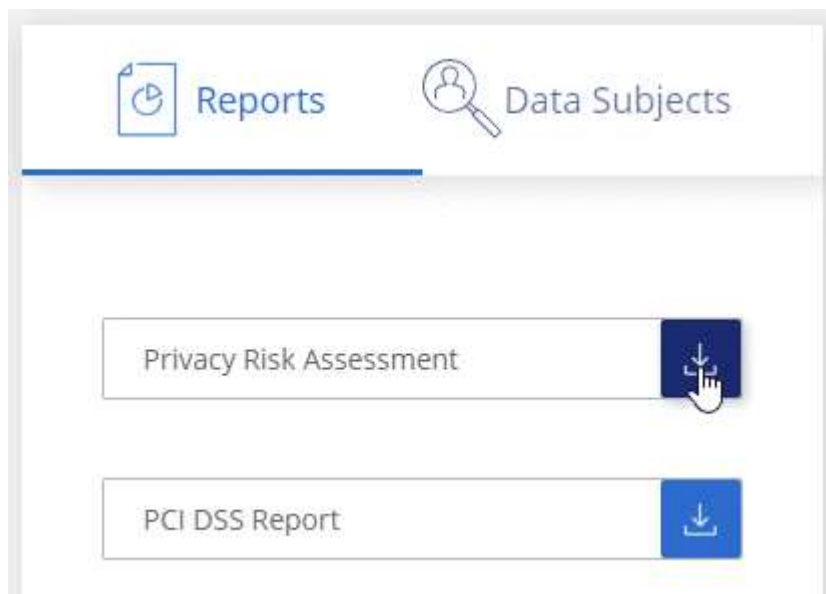
The number of people, by location, for which national identifiers were found.

Generate the Privacy Risk Assessment Report

Go to the Compliance tab to generate the report.

Steps

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



Result

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

Severity score

BlueXP classification calculates the severity score for the Privacy Risk Assessment Report on the basis of three variables:

- The percentage of personal data out of all data.
- The percentage of sensitive personal data out of all data.
- The percentage of files that include data subjects, determined by national identifiers such as national IDs, Social Security numbers, and tax ID numbers.

The logic used to determine the score is as follows:

Severity score	Logic
0	All three variables are exactly 0%
1	One of the variables are larger than 0%
2	One of the variables are larger than 3%
3	Two of the variables are larger than 3%
4	Three of the variables are larger than 3%
5	One of the variables are larger than 6%
6	Two of the variables are larger than 6%
7	Three of the variables are larger than 6%
8	One of the variables are larger than 15%
9	Two of the variables are larger than 15%
10	Three of the variables are larger than 15%

PCI DSS Report

The Payment Card Industry Data Security Standard (PCI DSS) Report can help you identify the distribution of credit card information across your files. The report includes the following information:

Overview

How many files contain credit card information and in which working environments.

Encryption

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

Ransomware Protection

The percentage of files containing credit card information that are on working environments that do or don't have ransomware protection enabled. This information is specific to Cloud Volumes ONTAP.

Retention

The timeframe in which the files were last modified. This is helpful because you shouldn't keep credit card information for longer than you need to process it.

Distribution of Credit Card Information

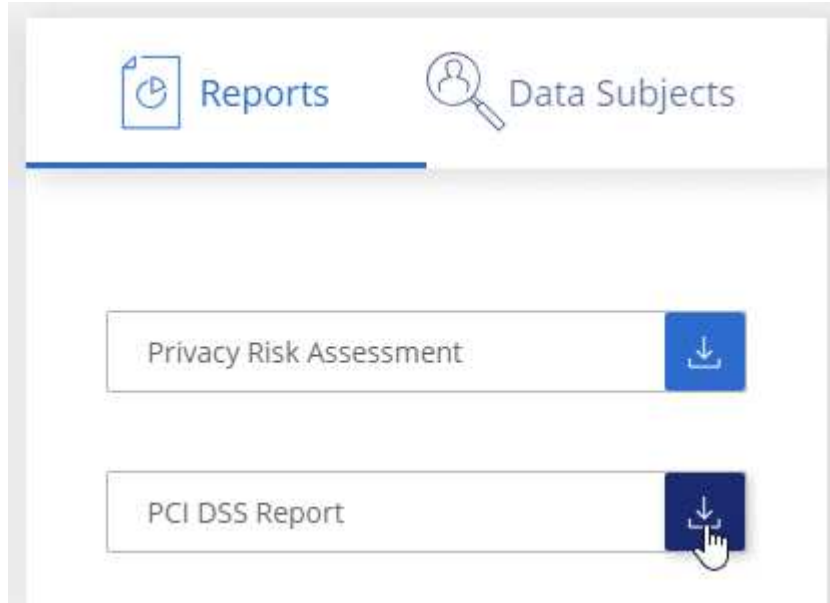
The working environments where the credit card information was found and whether encryption and ransomware protection are enabled.

Generate the PCI DSS Report

Go to the Compliance tab to generate the report.

Steps

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



Result

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

HIPAA Report

The Health Insurance Portability and Accountability Act (HIPAA) Report can help you identify files containing health information. It is designed to aid in your organization's requirement to comply with HIPAA data privacy laws. The information BlueXP classification looks for includes:

- Health reference pattern
- ICD-10-CM Medical code
- ICD-9-CM Medical code
- HR - Health category
- Health Application Data category

The report includes the following information:

Overview

How many files contain health information and in which working environments.

Encryption

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

Ransomware Protection

The percentage of files containing health information that are on working environments that do or don't have ransomware protection enabled. This information is specific to Cloud Volumes ONTAP.

Retention

The timeframe in which the files were last modified. This is helpful because you shouldn't keep health information for longer than you need to process it.

Distribution of Health Information

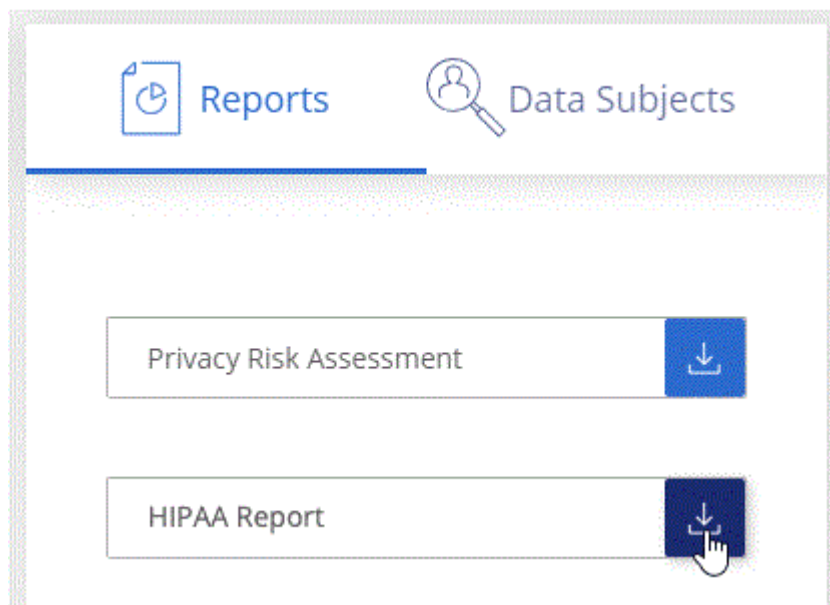
The working environments where the health information was found and whether encryption and ransomware protection are enabled.

Generate the HIPAA Report

Go to the Compliance tab to generate the report.

Steps

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



Result

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

What is a Data Subject Access Request?

Privacy regulations such as the European GDPR grant data subjects (such as customers or employees) the right to access their personal data. When a data subject requests this information, this is known as a DSAR (data subject access request). Organizations are required to respond to these requests "without undue delay", and at the latest within one month of receipt.

You can respond to a DSAR by searching for a subject's full name or known identifier (such as an email address) and then downloading a report. The report is designed to aid in your organization's requirement to comply with GDPR or similar data privacy laws.

How can BlueXP classification help you respond to a DSAR?

When you perform a data subject search, BlueXP classification finds all of the files, buckets, OneDrive, and SharePoint accounts that have that person's name or identifier in it. BlueXP classification checks the latest pre-indexed data for the name or identifier. It doesn't initiate a new scan.

After the search is complete, you can then download the list of files for a Data Subject Access Request report. The report aggregates insights from the data and puts it into legal terms that you can send back to the person.



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

Search for data subjects and download reports

Search for the data subject's full name or known identifier and then download a file list report or DSAR report. You can search by [any personal information type](#).

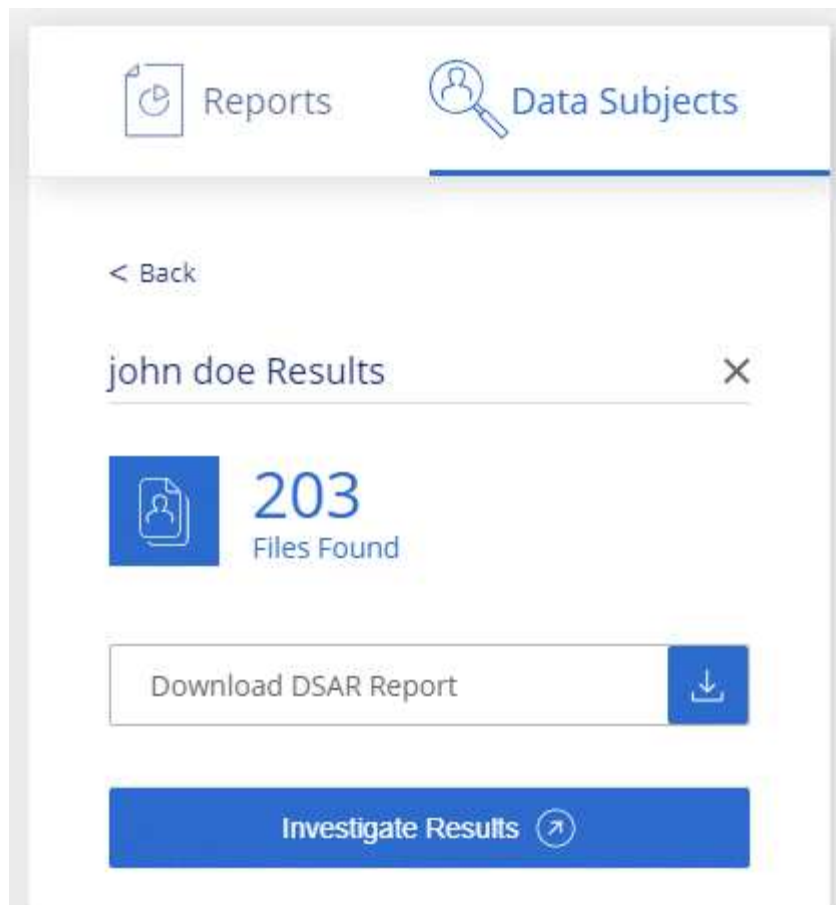


English, German, Japanese, and Spanish are supported when searching for the names of data subjects. Support for more languages will be added later.

Steps

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

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



4. Choose one of the available options:

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



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

Select the working environments for reports

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

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

Steps

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

All Working Environments (12) ^

☒ Select all

☒ ANF - Azure NetApp Files

ANF

☒ Working Environment Name 1

CVO

☒ Working Environment Name 2

CVS

☒ Working Environment Name 3

CVS

☒ Working Environment Name 4

CVO

View

Cancel

Personal Files ⓘ

View All

Email Address 2,700 Files



Credit Card 2,700 Files



20%
Personal



5%
Sensitive Personal



7,000

Sensitive Personal Files ⓘ

View All

Health 2,700 Files



Ethnicity 2,700 Files



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.