



Use Cloud Data Sense

Cloud Data Sense

NetApp
May 12, 2022

Table of Contents

- Use Cloud Data Sense. 1
 - Viewing governance details about the data stored in your organization 1
 - Viewing compliance details about the data stored in your organization. 4
 - Organizing your private data 15
 - Managing your private data 31
 - Adding personal data identifiers using Data Fusion. 42
 - Viewing compliance reports 44
 - Responding to a Data Subject Access Request. 51
 - Categories of private data 53
 - Removing data sources from Cloud Data Sense 59

Use Cloud Data Sense

Viewing governance details about the data stored in your organization

Gain control of the costs related to the data on your organizations' storage resources. Cloud Data Sense 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 are 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.



Saving 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.](#)

Policies with the largest number of results

Click the name of a Policy in the *Policy* area 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

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 working environments and data sources. See [Data Mapping Report](#) for complete details.

Top data repositories listed by data sensitivity

The *Top Data Repositories by Sensitivity Level* area lists up to 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-Personal 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 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 Access
- 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.

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.

Viewing 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 Cloud Data Sense 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 Cloud Data Sense 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.

Viewing files that contain personal data

Cloud Data Sense 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](#).

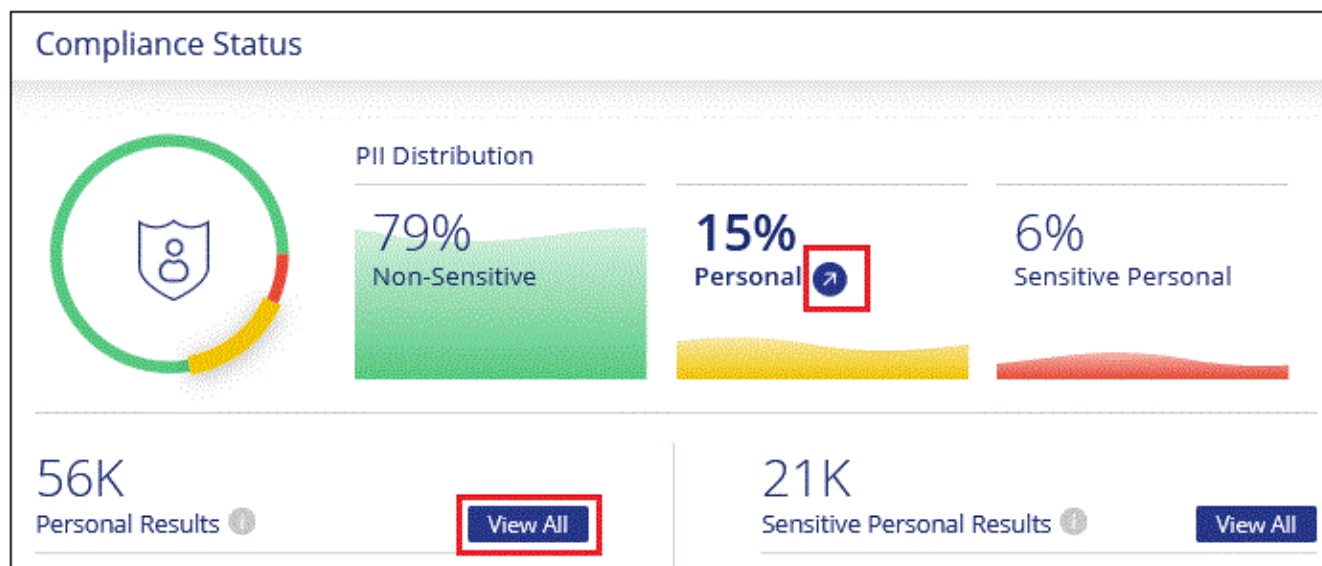
Additionally, if you have added a database server to be scanned, the *Data Fusion* feature allows you to scan your files to identify whether unique identifiers from your databases are found in those files or other databases. See [Adding personal data identifiers using Data Fusion](#) for details.

For some types of personal data, Data Sense 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, Data Sense 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 Data Sense uses proximity validation.

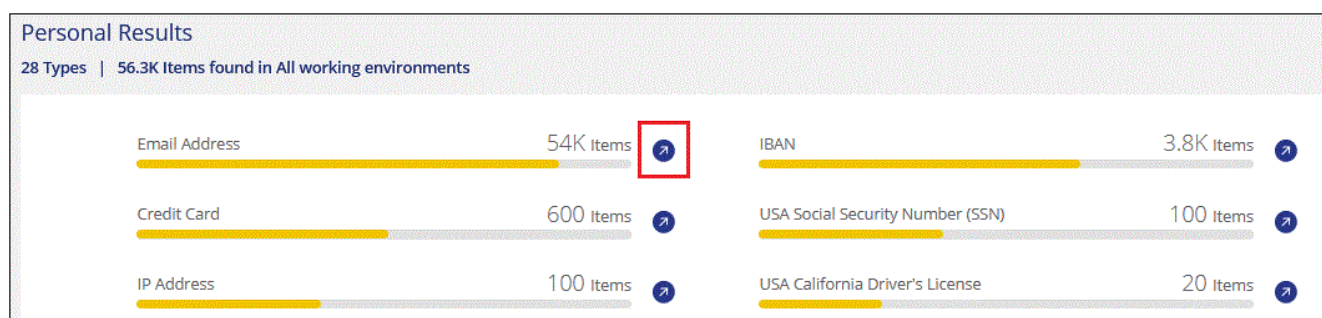
Steps

1. At the top of Cloud Manager, click **Data Sense** and 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.



Viewing files that contain sensitive personal data

Cloud Data Sense 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](#).

Cloud Data Sense 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, Data Sense 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."



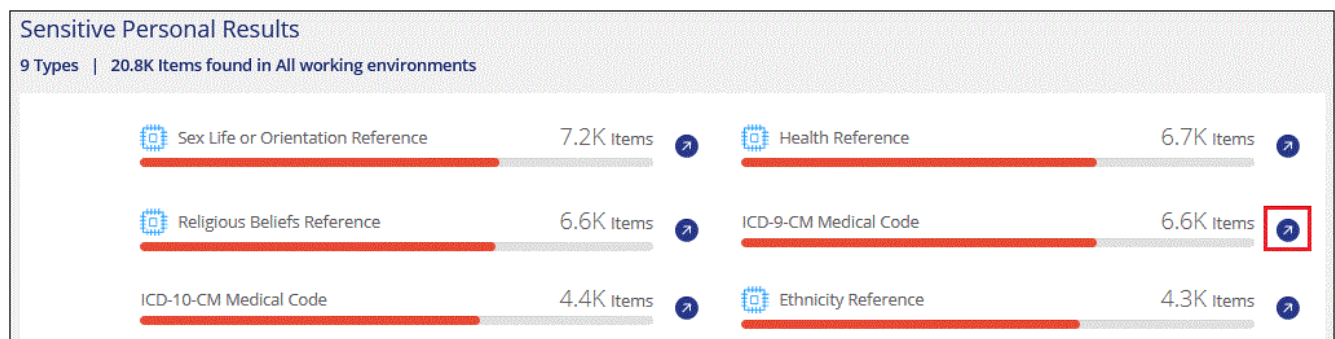
Only English is supported when scanning for sensitive personal data. Support for more languages will be added later.

Steps

1. At the top of Cloud Manager, click **Data Sense** and 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.

Viewing files by categories

Cloud Data Sense 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.](#)

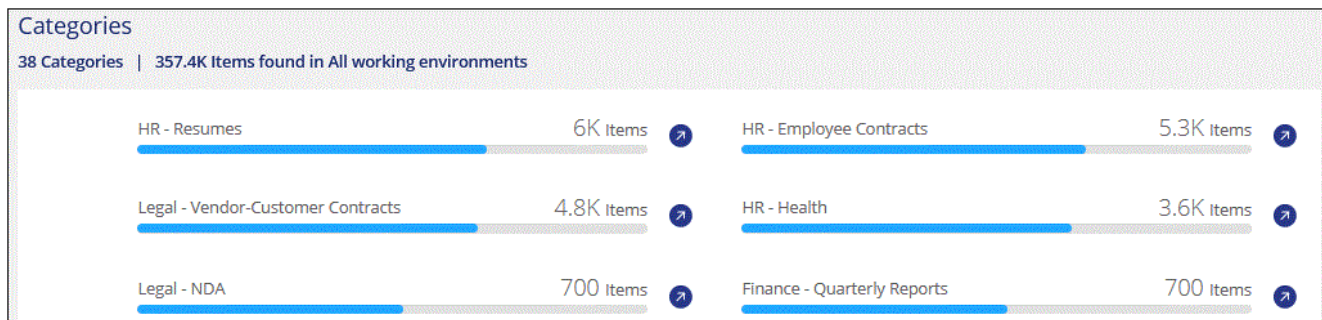
Categories can help you understand what's happening with your data by showing you the types of information that you have. For example, a category like resumes or employee contracts can include sensitive data. When you investigate the results, you might find that employee contracts are stored in an insecure 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. At the top of Cloud Manager, click **Data Sense** and 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.



- 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.

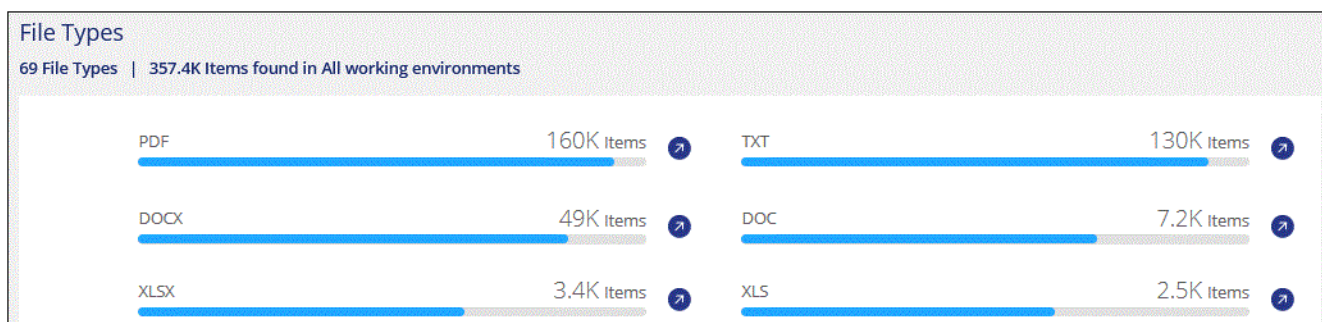
Viewing files by file types

Cloud Data Sense 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

- At the top of Cloud Manager, click **Data Sense** and click the **Compliance** tab.
- 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.

Viewing file metadata

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

The screenshot shows the Data Sense interface. At the top, there's a header with '364.9K items' and navigation 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. The table lists two files: '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 view. The modal view includes metadata such as Tags (Decathlon, gidi, IS NOT OK, and 6 more), 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 view.

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 Cloud Data Sense](#)). 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 Cloud Manager 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 Cloud Data Sense, 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.

Viewing permissions for files

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

Note that if you see SIDs (Security IDentifiers) instead of user and group names, you should integrate your Active Directory into Data Sense. [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 [View all 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 the name of a user or a group and the Investigation page is displayed with the name of that user or group in the “User / Group Permissions” filter so you can see all the files that the user or group has access to.

Additionally, you can click for any group to see the list of users who are part of the group.

Checking 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.

Data Sense 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.

Viewing 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.

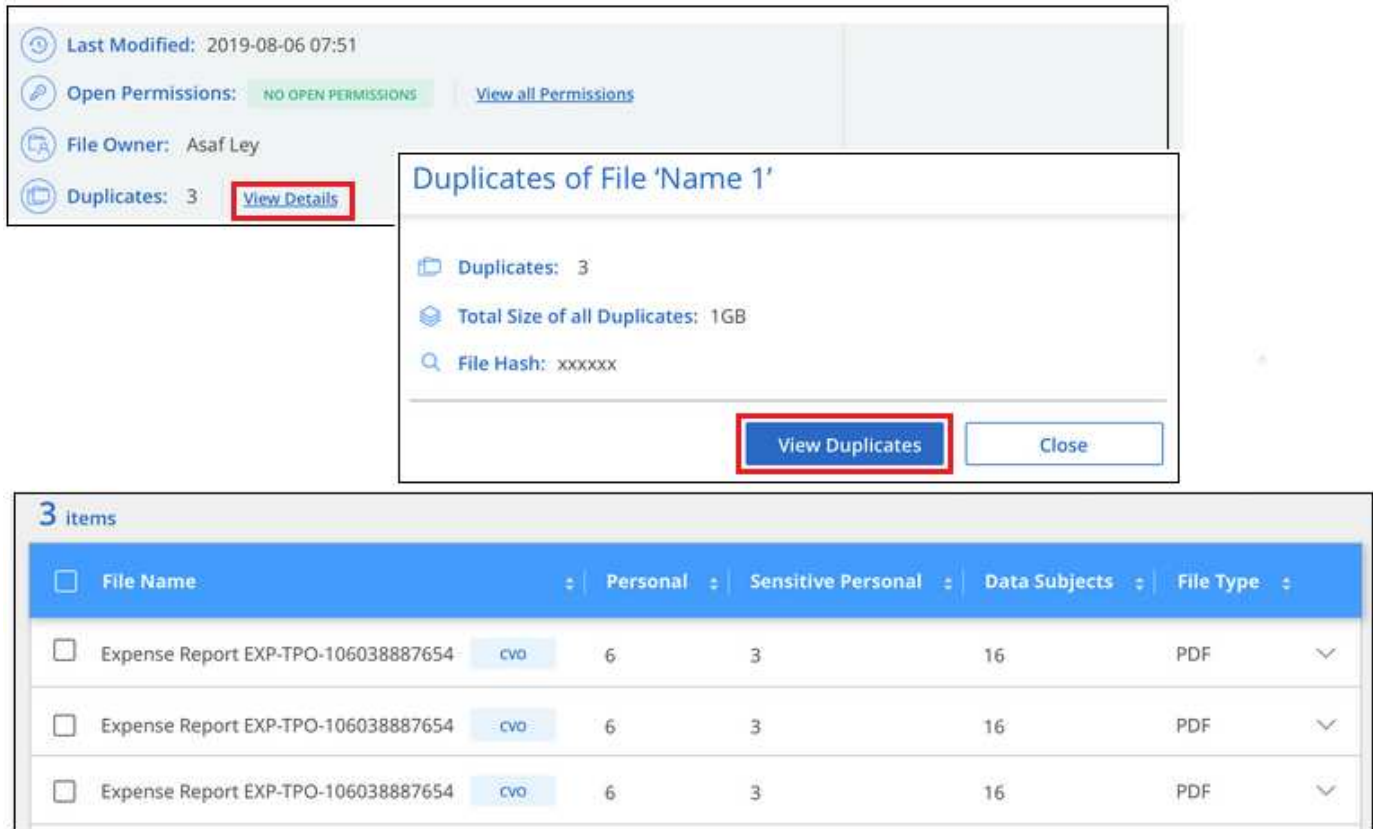
All files with duplicates from all file types (not including databases), with a minimum size of 50 MB, and/or containing personal or sensitive personal information, will show in the Results page.

Viewing 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 a 'View Details' modal window. At the top, it says 'Duplicates of File 'Name 1''. Below this, it lists 'Duplicates: 3', 'Total Size of all Duplicates: 1GB', and 'File Hash: xxxxxx'. A 'View Duplicates' button is highlighted with a red box. In the background, a table shows 3 items with columns for File Name, Personal, Sensitive Personal, Data Subjects, and File Type.

File Name	Personal	Sensitive Personal	Data Subjects	File Type
Expense Report EXP-TPO-106038887654	6	3	16	PDF
Expense Report EXP-TPO-106038887654	6	3	16	PDF
Expense Report EXP-TPO-106038887654	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.

Viewing Dashboard data for specific working environments

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

When you filter the dashboard, Data Sense 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**.



Filtering 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 to save a .CSV file.

Data Investigation

Unstructured (364K Files)

Directories (64 Folders)

Structured (45 Tables)

Search by file or DB table

FILTERS:

Clear All

Policies

+

Open Permissions

+

File Owner

+

Label

+

Working Environment Type

2

+

Working Environment

+

Storage Repository

2

+

364K items

Tags

Assign to

Label

Move

Copy

Delete

<input type="checkbox"/>	File Name		Personal	Sensitive Personal	Data Subjects	File Type	
<input type="checkbox"/>	cgdpr_yes_adam.txt	ANF	0	797	111	TXT	
<input type="checkbox"/>	cgdpr_yes_adam.txt	ANF	0	797	111	TXT	
<input type="checkbox"/>	true positive.txt	ANF	0	611	111	TXT	
<input type="checkbox"/>	cgdpr_yes_adam.txt	ANF	0	611	111	TXT	
<input type="checkbox"/>	true positive.txt	ANF	0	611	111	TXT	
<input type="checkbox"/>	true positive.txt	ANF	0	611	111	TXT	
<input type="checkbox"/>	cgdpr_yes_adam.txt	ANF	0	611	111	TXT	
<input type="checkbox"/>	cgdpr_yes_adam.txt	ANF	0	611	111	TXT	

- The top-level tabs allow you to view data from files (unstructured data), directories (unstructured data), or from databases (structured data).
- The controls at the top of each column allow you to sort the results in numerical or alphabetical order.
- The left-pane filters enable you to refine the results by selecting from the following attributes:

Filter	Details
Policies	Select a policy or policies. Go here to view the list of existing policies and to create your own custom policies.
Open Permissions	Select the type of permissions
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
Label	Select AIP labels that are assigned to your files
Working Environment Type	Select the type of working environment. OneDrive, SharePoint, and Google Drive are categorized under "Cloud Apps".
Working Environment name	Select specific working environments
Storage Repository	Select the storage repository, for example, a volume or a schema
File Path	Enter a partial or full path
Category	Select the types of categories
Sensitivity Level	Select the sensitivity level
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
Directory Type	Select the directory type; either "Share" or "Folder"

Filter	Details
File Type	Select the types of files
File Size	Select the file size range
Created Time	Select a range when the file was created
Discovered Time	Select a range when Data Sense discovered the file
Last Modified	Select a range when the file was last modified
Last Accessed	Select a range when the file was last accessed. For the types of files that Data Sense scans, this is the last time Data Sense scanned the file.
Duplicates	Select whether the file is duplicated in the repositories
File Hash	Enter the file's hash to find a specific file, even if the name is different
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

Organizing your private data

Cloud Data Sense 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 Cloud Data Sense to manage those AIP labels.
- You can add Tags to files that you want to mark for organization or for some type of follow-up.
- You can assign a Cloud Manager 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 Cloud Manager users 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 Data Sense tagging and Azure Information Protection labeling.

Tags	Labels
File tags are an integrated part of Data Sense.	Requires that you have subscribed to Azure Information Protection (AIP).

Tags	Labels
The tag is only kept in the Data Sense 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 Data Sense 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.	

Categorizing your data using AIP labels

You can manage AIP labels in the files that Cloud Data Sense 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. Data Sense enables you to view the labels that are already assigned to files, add labels to files, and change labels when a label already exists.

Cloud Data Sense 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, Cloud Data Sense considers it as a file without a label.
- If you have deployed the Data Sense instance in an on-prem location that has no internet access (also known as a dark site), then the AIP label functionality is unavailable.

Integrating AIP labels in your workspace

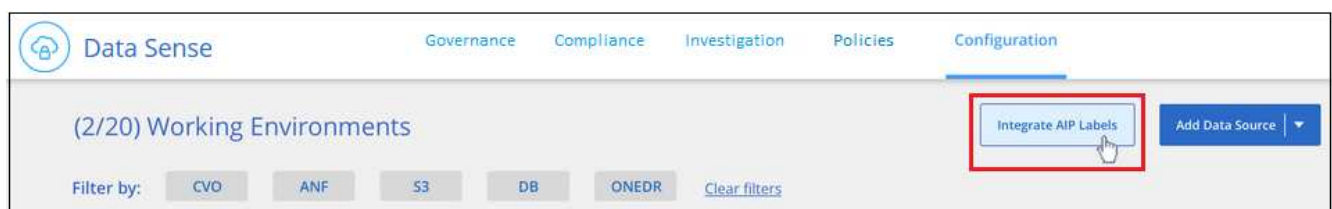
Before you can manage AIP labels, you need to integrate the AIP label functionality into Cloud Data Sense by signing into your existing Azure account. Once enabled, you can manage AIP labels within files for all [working environments and data sources](#) in your Cloud Manager workspace.

Requirements

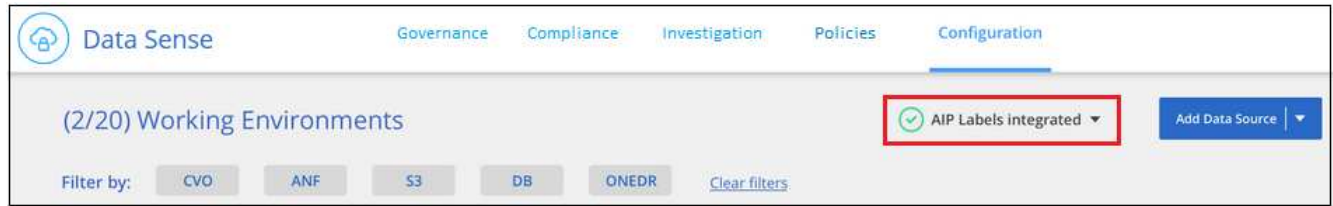
- You must have an account and an Azure Information Protection license.
- You must have the login credentials for the Azure account.
- If you plan to change labels in files that reside in Amazon S3 buckets, ensure that the permission `s3:PutObject` is included in the IAM role. See [setting up the IAM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

Steps

1. From the Cloud Data Sense 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 Cloud Data Sense 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.

Viewing 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.



Assigning AIP labels manually

You can add, change, and remove AIP labels from your files using Cloud Data Sense.

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.

To assign an AIP label to multiple files:

Steps

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



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

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



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

Assigning 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 Cloud Data Sense scans your files.

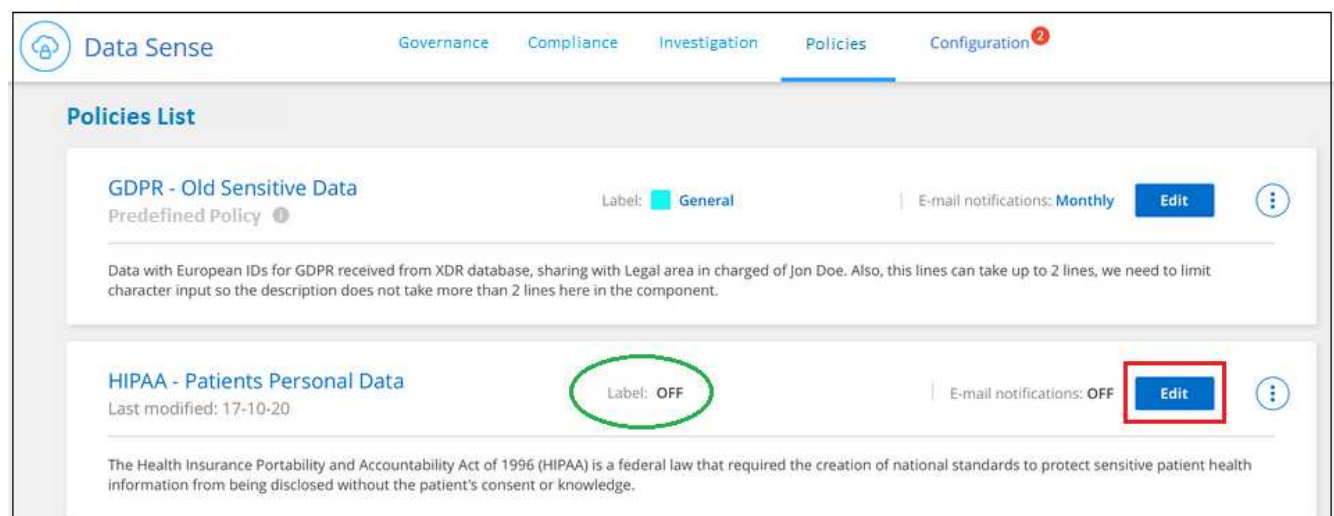
Depending on whether a label is already applied to a file, and the classification level of the label, the following actions are taken when changing a label:

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
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**).

Edit Policy

Saving this filtered view will create a new Policy, you can view/edit it in the "Policy" tab

Name this Policy

HIPAA - Patient Personal Data

Give it a description to quickly identify it

Files containing patient health information that is more than 30 days old

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

☒ Automatically label matches of this Policy with: select label

General

Finance

Confidential

Cancel

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.

Removing the AIP integration

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

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

Steps

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

Configuration

AIP Labels integrated

Add Data Source

Remove Integration

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

Applying 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.

Data Sense 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 Cloud Manager users using Cloud Data Sense so you can see if a file needs to be deleted or checked for some type of follow-up.

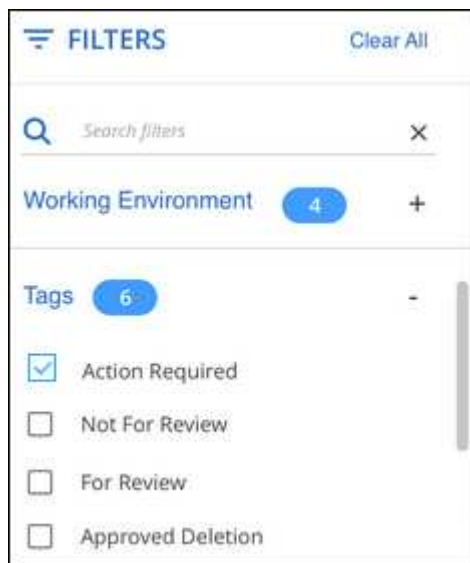


Tags assigned to files in Cloud Data Sense are not related to the tags you can add to resources, such as volumes or virtual machine instances. Data Sense tags are applied at the file level.

Viewing files that have certain tags applied

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

1. Click the **Investigation** tab from Cloud Data Sense.
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.

Assigning tags to files

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

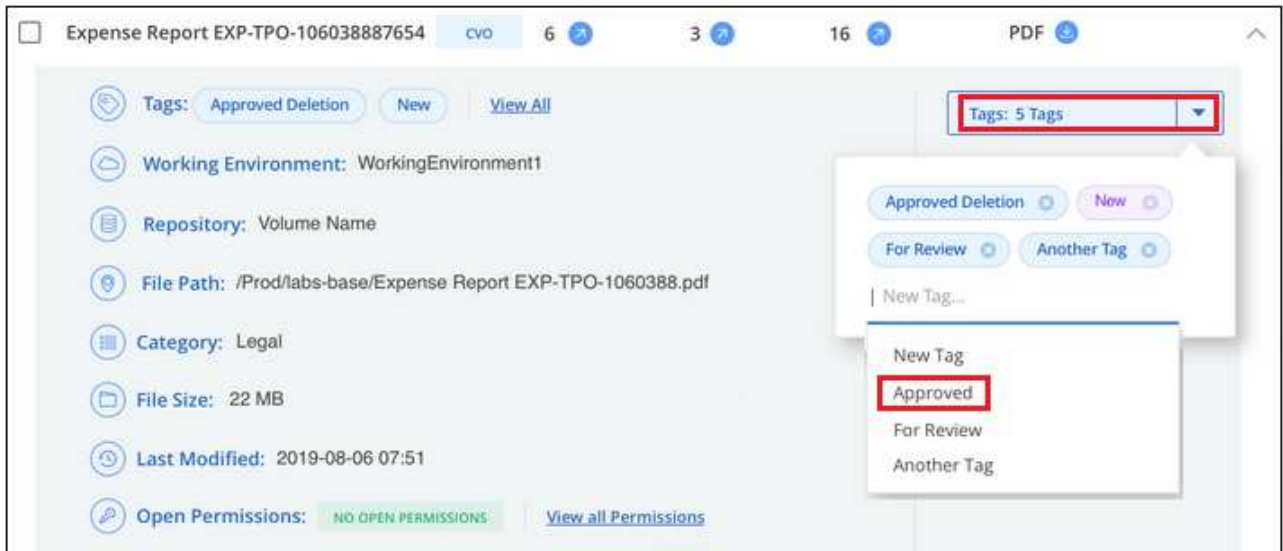
To add a tag to a single file:

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.

2345 items							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-10603887654	cvo	6	3	16	PDF						
<input checked="" type="checkbox"/>	Expense Report EXP-TPO-10603887654	cvo	6	3	6	PDF						
<input type="checkbox"/>	Expense Report EXP-TPO-10603887654	cvo	6	3	6	PDF						
<input type="checkbox"/>	Expense Report EXP-TPO-10603887654	cvo	6	3	6	PDF						

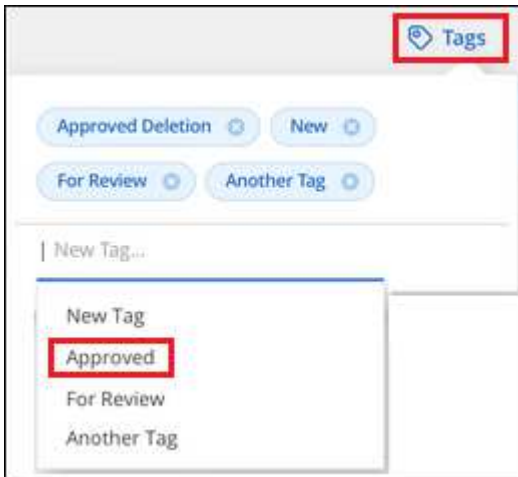
- To select individual files, check the box for each file (☒ Volume_1).
- To select all files on the current page, check the box in the title row (☒ File Name).

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

3. Add the tag or tags:

- To assign an existing tag, click in the **New Tag...** field and start typing the name of the tag. When the tag you are looking for appears, select it and press **Enter**.
- To create a new tag and assign it to the file, click in the **New Tag...** field, enter the name of the new

tag, and press **Enter**.



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

Deleting 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.

Assigning users to manage certain files

You can assign a Cloud Manager 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 Cloud Manager users when using Cloud Data Sense.

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

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.

To assign a user to multiple files:

Steps

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

2345 items

Tags

Assign to

Label

Copy

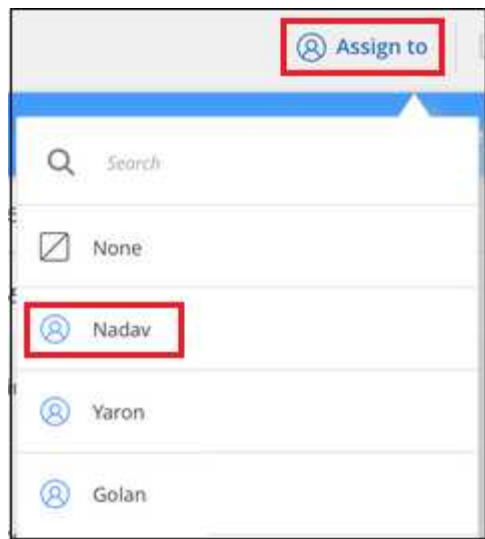
Move

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.

Controlling your data using Policies

Policies are like a favorites list of custom filters that provide search results in the Investigation page for commonly requested compliance queries. Cloud Data Sense 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 Cloud Manager users 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

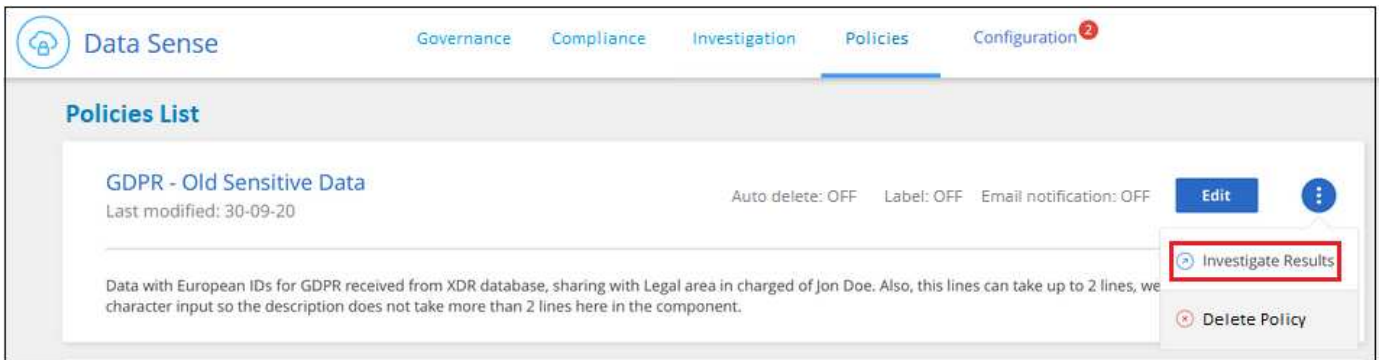
The **Policies** tab in the Compliance Dashboard lists all the predefined and custom Policies available on this instance of Cloud Data Sense.



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

Viewing 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**.

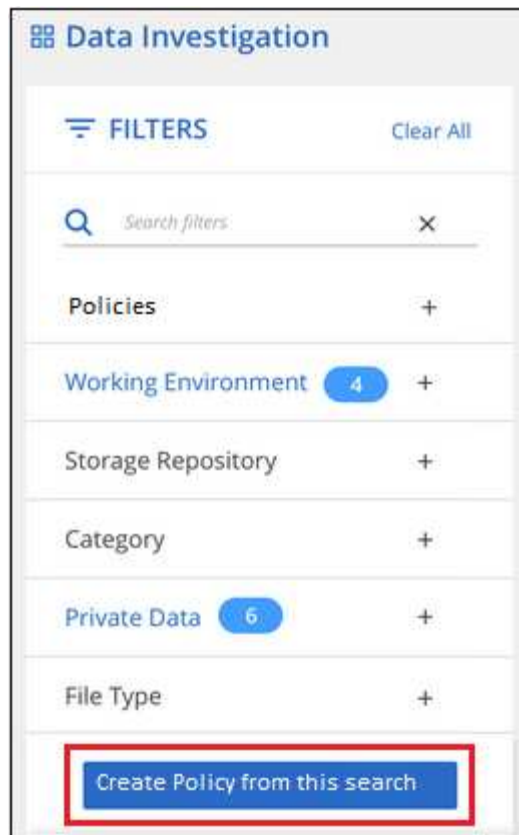


Creating custom Policies

You can create your own custom Policies that provide results for searches specific to your organization.

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. 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 Cloud Manager users, 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 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](#).)
 - e. 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

New Policy to view all files that were created over 60 days ago

Give it a detailed description that explains what it searches for

See if any files greater than 60 days old should be deleted from the system

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

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

☐ Automatically label this Policy's matches with: Select a label

Create Policy Cancel

Result

The new Policy appears in the Policies tab.

Sending email alerts when non-compliant data is found

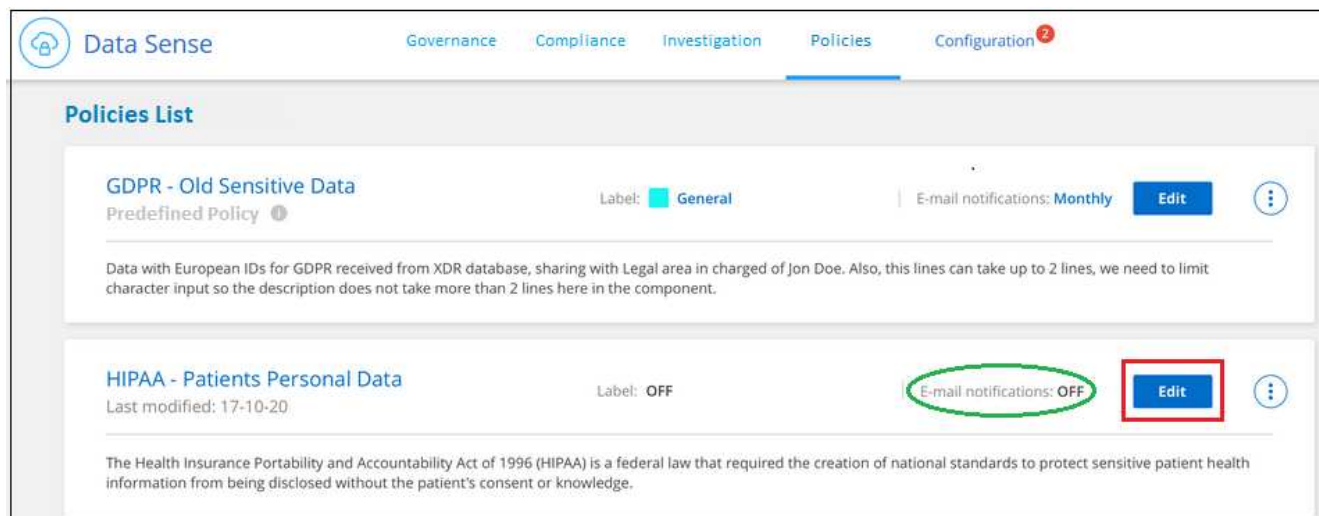
Cloud Data Sense can send email alerts to Cloud Manager users when certain critical Policies return results so you can get notifications to protect your data. You can choose to send the email notifications on a daily, weekly, or monthly basis.

You can 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 box if you want notification emails sent to Cloud Manager users, and choose the interval at which the email is sent (for example, every **Week**).

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.

Editing Policies

You can modify certain parts of a Policy depending on the type of Policy:

- Custom Policies - You can modify the *Name*, the *Description*, whether email notifications are sent, and whether AIP labels are added.
- Predefined Policies - You can modify only whether email notifications are sent and whether AIP labels are

added.




If you need to change the filter parameters for a custom Policy, you'll need to create a new Policy with the parameters you want, and then delete the old Policy.

To modify a Policy, click the **Edit** button, enter your changes on the *Edit Policy* page, and click **Save Policy**.

Deleting 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

Cloud Data Sense 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

Name	Description	Logic
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
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

Managing your private data

Cloud Data Sense 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 Data Sense 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.

Copying source files

You can copy any source files that Data Sense 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 Cloud Sync](#) functionality to copy and sync data from a source to a target.

Copying source files to an NFS share

You can copy source files that Data Sense is scanning to any NFS share. The NFS share doesn't need to be integrated with Data Sense, 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>`.



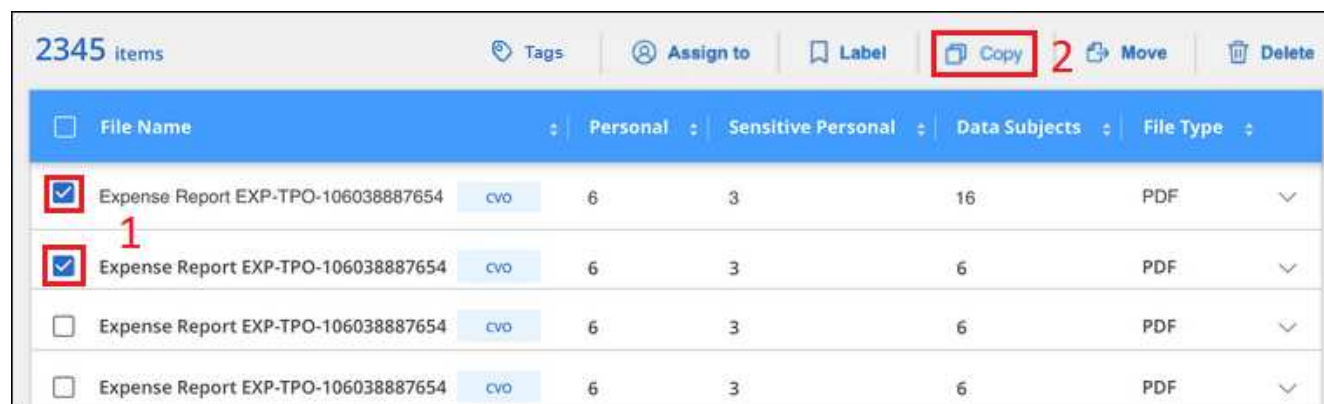
You can't copy files that reside in databases.

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 Data Sense instance.
- You can copy a maximum of 100,000 files at a time.

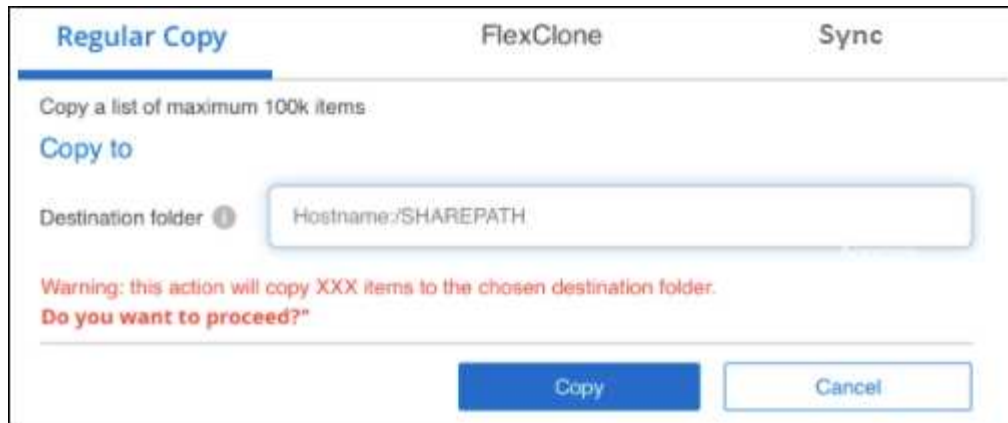
Steps

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



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

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



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

A dialog appears with the status of the copy operation.

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

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



Cloning volume data to a new volume

You can clone an existing ONTAP volume that Data Sense 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.

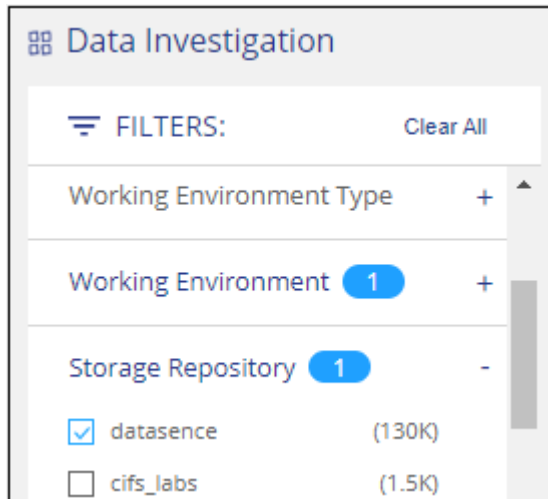
Requirements

- You must have the Account Admin or Workspace Admin role to copy files.
- 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 Data Sense for the working environment where the source volume resides, then Data Sense 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](#).

Copying and synchronizing source files to a target system

You can copy source files that Data Sense is scanning from any supported unstructured data source to a directory in a specific target destination location ([target locations that are supported by Cloud 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 Cloud 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

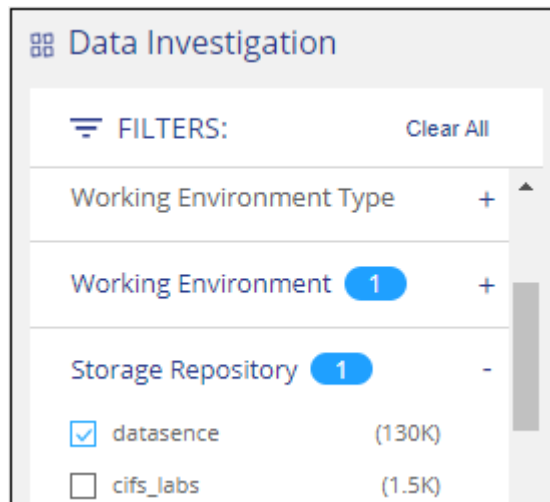
- You must have the Account Admin or Workspace Admin role to copy and sync files.
- All selected files must be from the same source repository (ONTAP volume, S3 bucket, NFS or CIFS share, etc.).
- You'll need to activate the Cloud Sync service and configure a minimum of one data broker that can be

used to transfer files between the source and target systems. Review the Cloud Sync requirements beginning with the [Quick Start description](#).

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

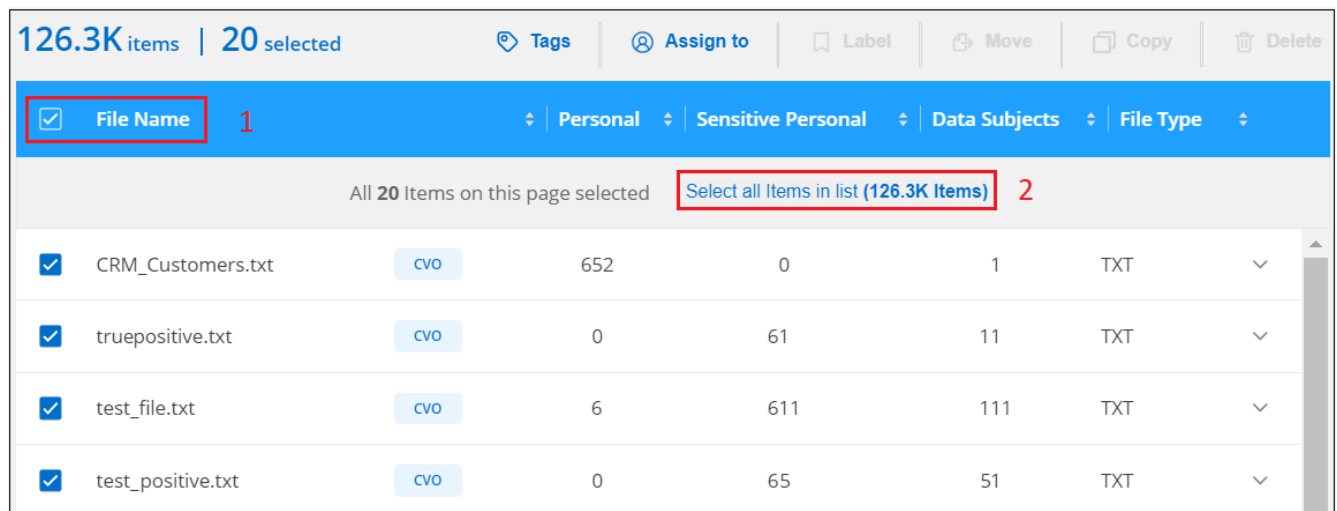
Steps

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.

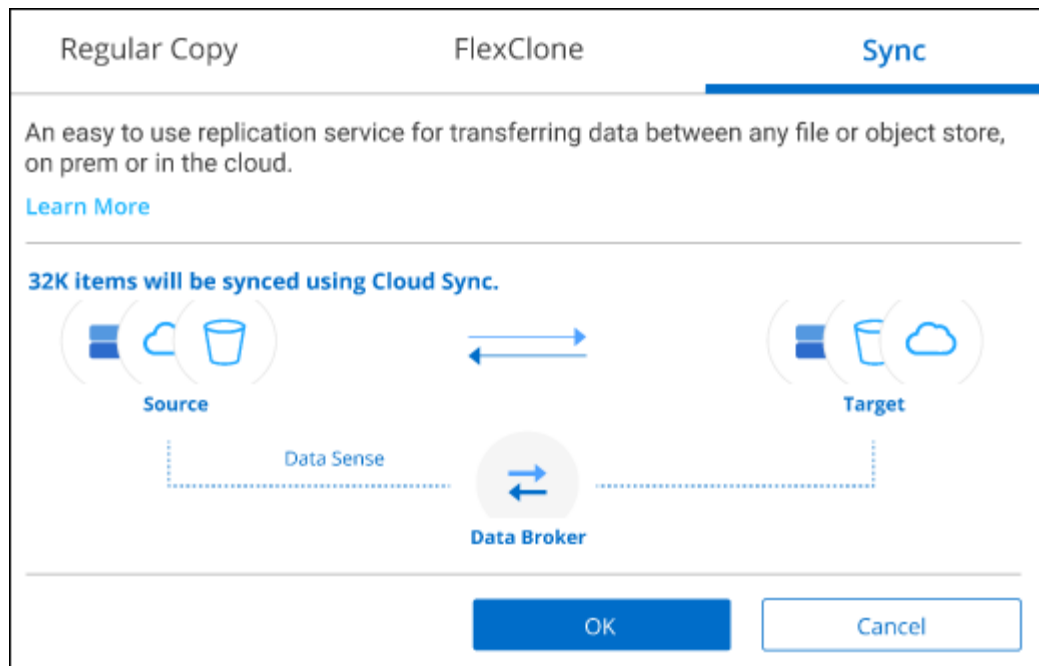


Apply any other filters so that you're seeing only the files that you want to copy and sync to the destination system.

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 Cloud Sync UI is opened in Cloud Manager.

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

5. You'll need to select the Target system and then select (or create) the Data Broker you plan to use. Review the Cloud 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 Cloud Sync operations are disabled when it is invoked from Data Sense:

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

Moving source files to an NFS share

You can move source files that Data Sense is scanning to any NFS share. The NFS share doesn't need to be integrated with Data Sense (see [Scanning file shares](#)).



You can't move files that reside in databases.

Requirements

You must have the Account Admin or Workspace Admin role to move files.

Moving files requires that the NFS share allows access from the Data Sense instance.

Steps

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

2345 items

Tags

Assign to

Label

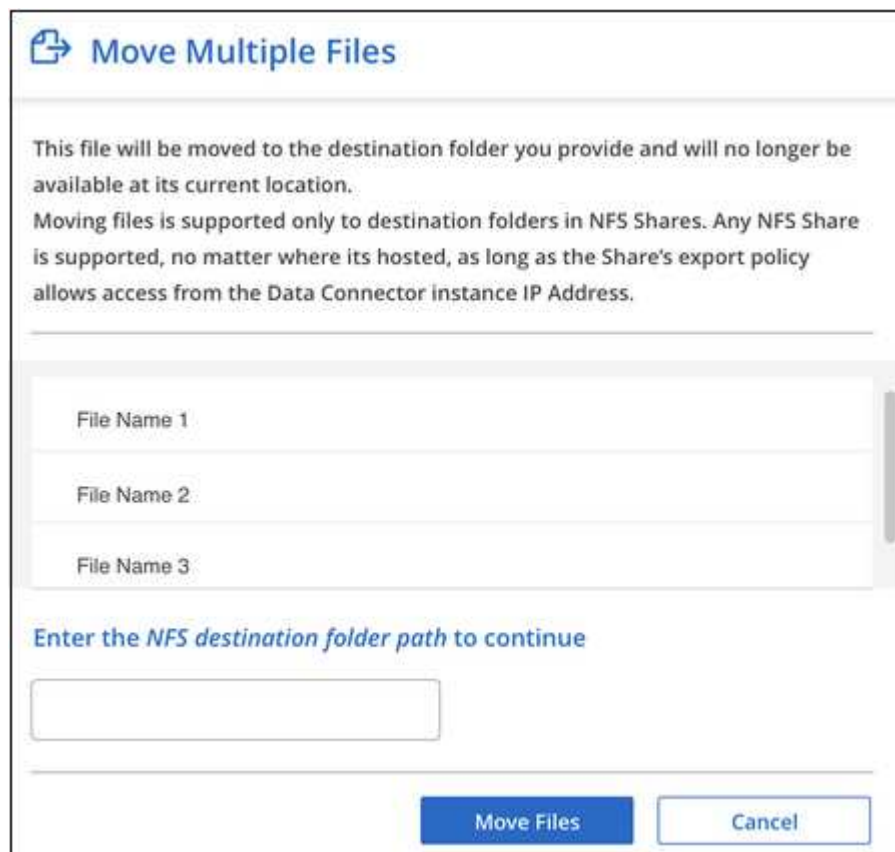
Copy

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 **Move**.



Move Multiple Files

This file will be moved to the destination folder you provide and will no longer be available at its current location.

Moving files is supported only to destination folders in NFS Shares. Any NFS Share is supported, no matter where its hosted, as long as the Share's export policy allows access from the Data Connector instance IP Address.

File Name 1

File Name 2

File Name 3

Enter the NFS destination folder path to continue

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>, and click **Move Files**.

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



Deleting 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.

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`.

Deleting 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**.



Deleting 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 Data Sense 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

Delete files with sensitive data

Give it a detailed description that explains what it searches for

Delete files that contain sensitive information and that were discovered in the past 30 days

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

Type "permanently delete" to continue with the deletion.

permanently delete

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

☐ Automatically label this Policy's matches with: Select a label

Create Policy Cancel

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](#).

Viewing the status of your compliance actions

When you run an action from the Investigation Results pane across many files, for example, deleting 100 files, the process can take some time. You can monitor the status of these asynchronous actions in the *Action Status* pane so you'll know when it has been applied to all files. This allows you to see the actions that have completed successfully, those currently in progress, and those that have failed so you can diagnose and fix any problems.

The status can be:

- Finished
- In Progress
- Queued
- Canceled

- Failed

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

Steps

1.

In the bottom-right of the Data Sense UI you can see the **Actions Status** button



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

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

Adding personal data identifiers using Data Fusion

A feature we call *Data Fusion* allows you to scan your organizations' data to identify whether unique identifiers from your databases are found in files or other databases - basically making your own list of "personal data" that is identified in Cloud Data Sense scans. This gives you the full picture about where potentially sensitive data resides in *all* your files.

Since you are scanning your own databases, whatever language your data is stored in will be used to identify data in future Cloud Data Sense scans.



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.

Creating custom personal data identifiers from your databases

You can choose the additional identifiers that Cloud Data Sense will look for in its' scans by selecting a specific column, or columns, in a database table. For example, the diagram below shows how data fusion is used to scan your volumes, buckets, and databases for occurrences of all your Customer IDs from your Oracle database.

Databases -- Structured Data

Database: Oracle
Schema: Accounts
Table: Customers
Column: Customer ID

Account	Name	Customer ID	Address
1234	ABC Co	135876	125 Main St
1235	XYZ Co	213536	35A Brick R
1236	Cat Co	359264	55 Wind Av
1237	Dog Co	472637	11025 Cor
1238	Zebra Co	582455	36 Sahara
...

Scan your volumes and buckets for occurrences of the Customer IDs in your Oracle database

Files -- Unstructured Data

File in Volume 1

```
XXXXXXXXXXXXX
xx213536xxx
XXXXXXXXXXXXX
xx472637xxx
XXXXXXXXXXXXX
XXXXXXXXXXXXX
```

File in Volume 2

```
XXXXXXXXXXXXX
XXXXXXXXXXXXX
XXXXXXXXXXXXX
XXXXXXXXXXXXX
XXXXXXXXXXXXX
xxx472637xx
```

File in Bucket 1

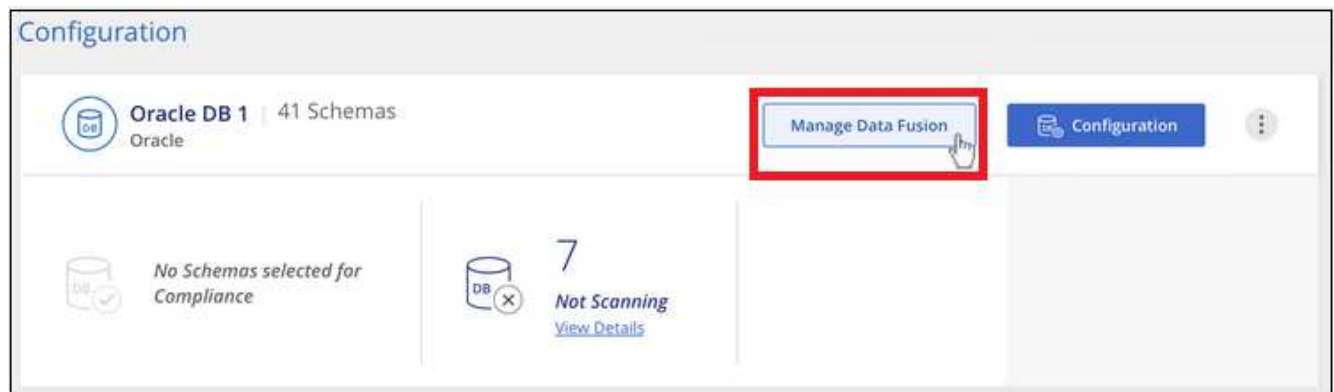
```
XXXXXXXXXXXXX
XXXXXXXXXXXXX
xx213536xxx
XXXXXXXXXXXXX
XXXXXXXXXXXXX
XXXXXXXXXXXXX
```

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

Steps

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

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

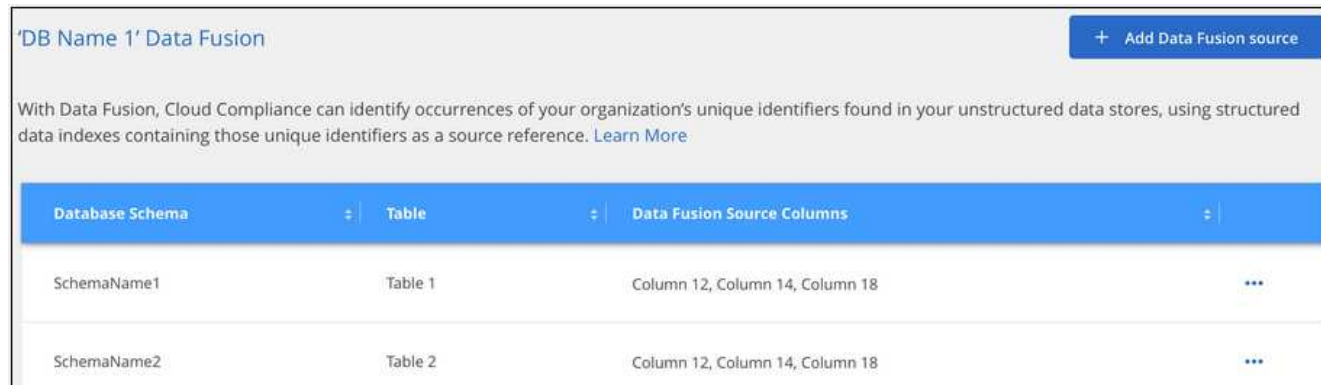


2. Click **Add Data Fusion source** on the next page.
3. In the *Add Data Fusion Source* page:
 - a. Select the Database Schema from the drop-down menu.
 - b. Enter the Table name in that schema.
 - c. Enter the Column, or Columns, that contain the unique identifiers you want to use.

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

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

The Data Fusion inventory page displays the database source columns that you have configured for Cloud Data Sense to scan.



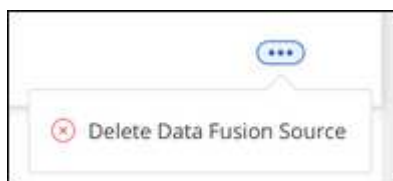
Database Schema	Table	Data Fusion Source Columns
SchemaName1	Table 1	Column 12, Column 14, Column 18
SchemaName2	Table 2	Column 12, Column 14, Column 18

Results

After the next scan, the results will include this new information in the Dashboard under the "Personal" results section, and in the Investigation page in the "Personal Data" filter. Each source column you added appears in the filter list as "Table.Column", for example `Customers.Customer ID`.

Deleting a Data Fusion source

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



Viewing compliance reports

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

By default, the Cloud Data Sense 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 Cloud Data Sense 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

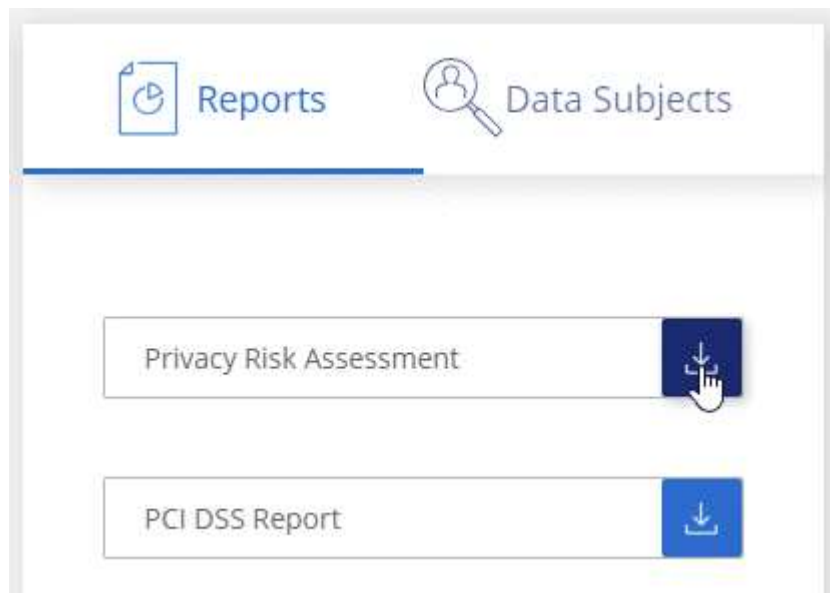
The number of people, by location, for which national identifiers were found.

Generating the Privacy Risk Assessment Report

Go to the Data Sense tab to generate the report.

Steps

1. At the top of Cloud Manager, click **Data Sense**.
2. Click **Compliance**, and then click the download icon next to **Privacy Risk Assessment** under **Reports**.



Result

Cloud Data Sense generates a PDF report that you can review and send to other groups as needed.

Severity score

Cloud Data Sense 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

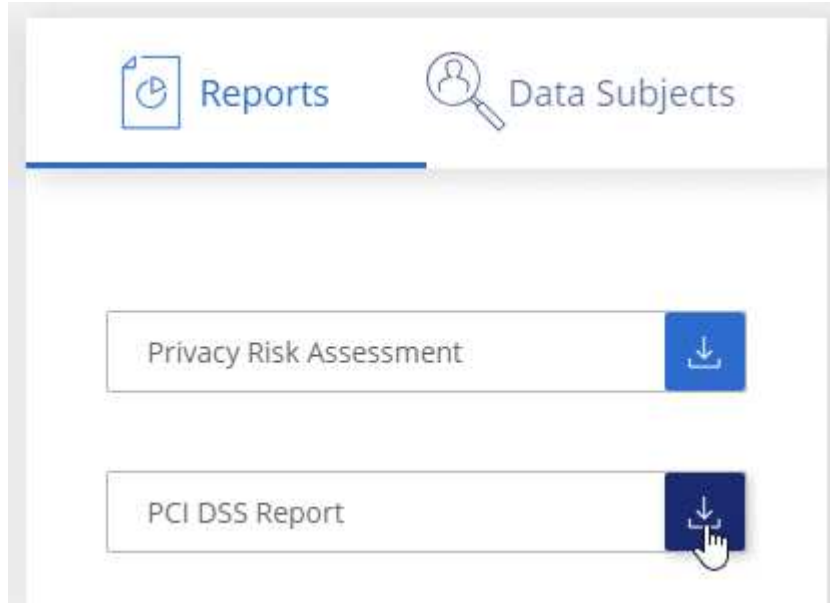
The working environments where the credit card information was found and whether encryption and ransomware protection are enabled.

Generating the PCI DSS Report

Go to the Data Sense tab to generate the report.

Steps

1. At the top of Cloud Manager, click **Data Sense**.
2. Click **Compliance**, and then click the download icon next to **PCI DSS Report** under **Reports**.



Result

Cloud Data Sense 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 Cloud Data Sense 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

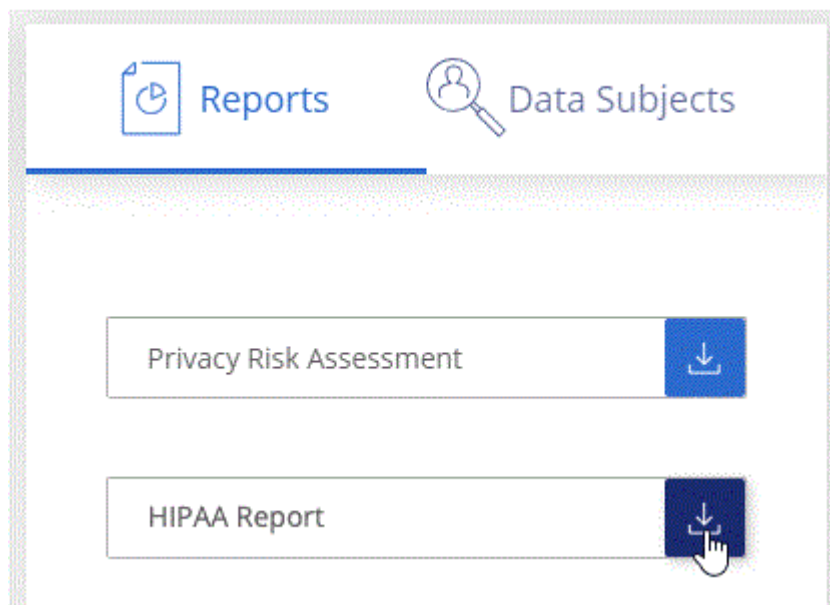
The working environments where the health information was found and whether encryption and ransomware protection are enabled.

Generating the HIPAA Report

Go to the Data Sense tab to generate the report.

Steps

1. At the top of Cloud Manager, click **Data Sense**.
2. Click **Compliance**, and then click the download icon next to **HIPAA Report** under **Reports**.



Result

Cloud Data Sense generates a PDF report that you can review and send to other groups as needed.

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 report summarizing all your working environments and data sources, and then provides a breakdown for each working environment.

The report includes the following information:

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.

Generating the Data Mapping Report

Go to the Data Sense tab to generate the report.

Steps

1. At the top of Cloud Manager, click **Data Sense**.
2. Click **Governance**, and then click the **Full Data Mapping Overview Report** button from the Governance Dashboard.



Result

Cloud Data Sense generates a PDF report that you can review and send to other groups as needed.

Data Investigation Report

The Data Investigation Report is a download of the contents of the Data Investigation page. [Learn more about the Data Investigation page.](#)

You can save the report on the local machine as a .CSV file; which can include up to 5,000 rows of data. If Data Sense is scanning database tables (structured data) and files (unstructured data), there can be up to two .CSV reports downloaded.

Generating the Data Investigation Report

Steps

1. From the Data Investigation page, click the  button on the top, right of the page.

A dialog displays a message that the report is being downloaded.

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)
- Working environment type
- File path
- File type
- File size
- Created time
- Last modified
- Last accessed
- File owner
- Category
- Personal information
- Sensitive personal information
- 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 **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

Note: Information about Folders is not currently available in the reports.

Selecting the working environments for reports

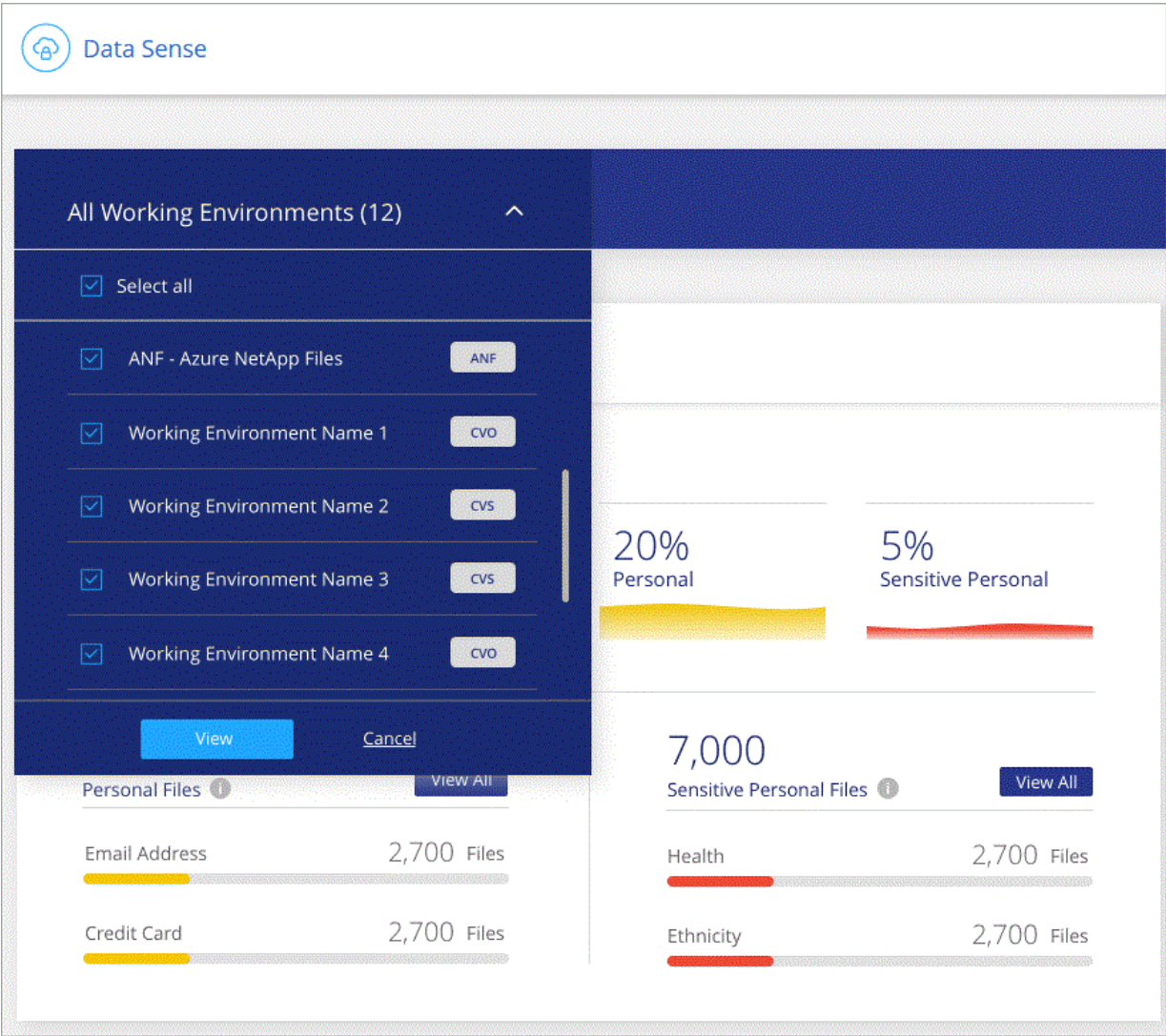
You can filter the contents of the Cloud Data Sense Compliance dashboard to see compliance data for all working environments and databases, or for just specific working environments.

When you filter the dashboard, Data Sense 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**.



Responding to a Data Subject Access Request

Respond to a Data Subject Access Request (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.



The DSAR capabilities 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 provide file-level details.



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

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.

How can Cloud Data Sense help you respond to a DSAR?

When you perform a data subject search, Cloud Data Sense finds all of the files, buckets, OneDrive, and SharePoint accounts that have that person's name or identifier in it. Data Sense 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.

Searching for data subjects and downloading 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](#).

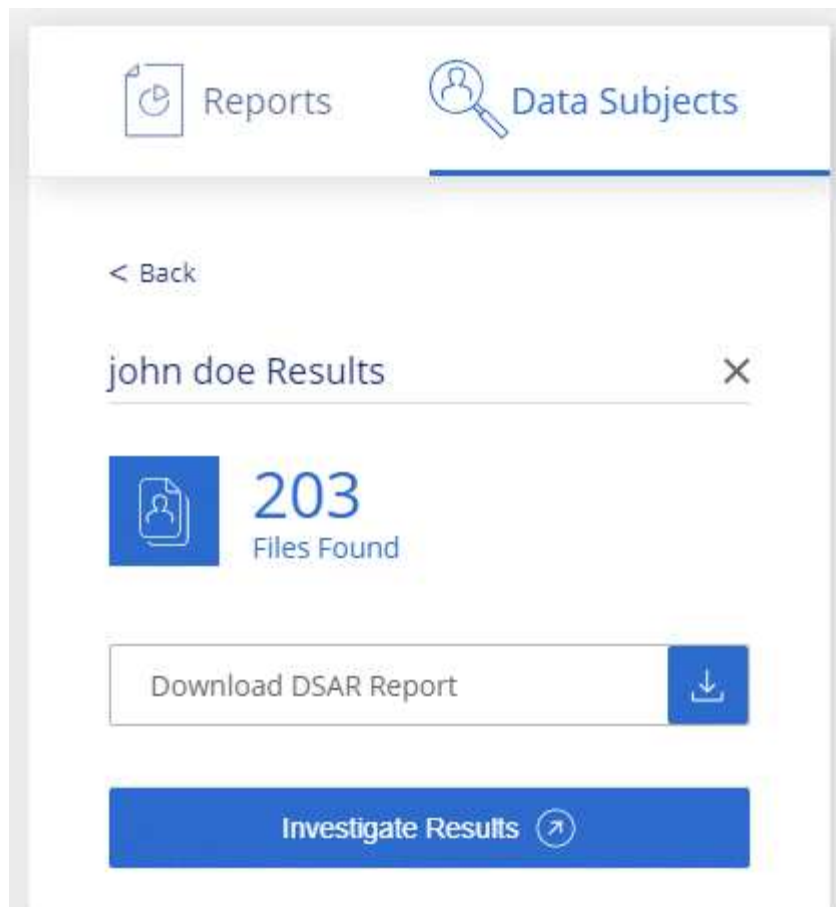


English, German, and Spanish are supported when searching for the names of data subjects. Support for more languages will be added later.

Steps

1. At the top of Cloud Manager, click **Data Sense**.
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 Cloud Data Sense 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.

Categories of private data

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



If you need Cloud Data Sense 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 identifies whether Cloud Data Sense uses [proximity validation](#) to validate its findings for the identifier.

The items in this category can be recognized in any language.

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 allows you to choose the additional identifiers that Cloud Data Sense will look for in its' scans by selecting columns in a database table. See [Adding personal data identifiers using Data Fusion](#) for details.

Type	Identifier	Proximity validation?
General	Email address	No
	Credit card number	No
	IBAN number (International Bank Account Number)	No
	IP address	No
	Password	Yes

Type	Identifier	Proximity validation?
National Identifiers		

Type	Lithuanian ID	Yes
	Luxembourg ID	Yes
	Maltese ID	Yes
Identifier	National Health Service (NHS) Number	Yes
	Polish ID (PESEL)	Yes
	Portuguese Tax Identification Number (NIF)	Yes
	Romanian ID (CNP)	Yes
	Slovenian ID (EMSO)	Yes
	South African ID	Yes
	Spanish Tax Identification Number	Yes
	Swedish ID	Yes
	U.K. ID (NINO)	Yes
	USA Social Security Number (SSN)	Yes
Proximity		
Validation?		

Types of sensitive personal data

The sensitive personal data that Cloud Data Sense 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

Cloud Data Sense 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	NDA's	✓	✓	✓
	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
- Design Files
- Email Application Data
- Encrypted
- Executables
- Financial Application Data
- Health Application Data
- Images
- Logs
- Miscellaneous Documents
- Miscellaneous Presentations
- Miscellaneous Spreadsheets
- Miscellaneous "Unknown"
- Structured Data
- Videos
- Zero-Byte Files

Types of files

Cloud Data Sense scans all files for category and metadata insights and displays all file types in the file types section of the dashboard.

But when Data Sense 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 Cloud Data Sense 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 Data Sense finds. We break it down by *precision* and *recall*:

Precision

The probability that what Data Sense 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 Data Sense to find what it should. For example, a recall rate of 70% for personal data means that Data Sense can identify 7 out of 10 files that actually contain personal information in your organization. Data Sense 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 Data Sense 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%

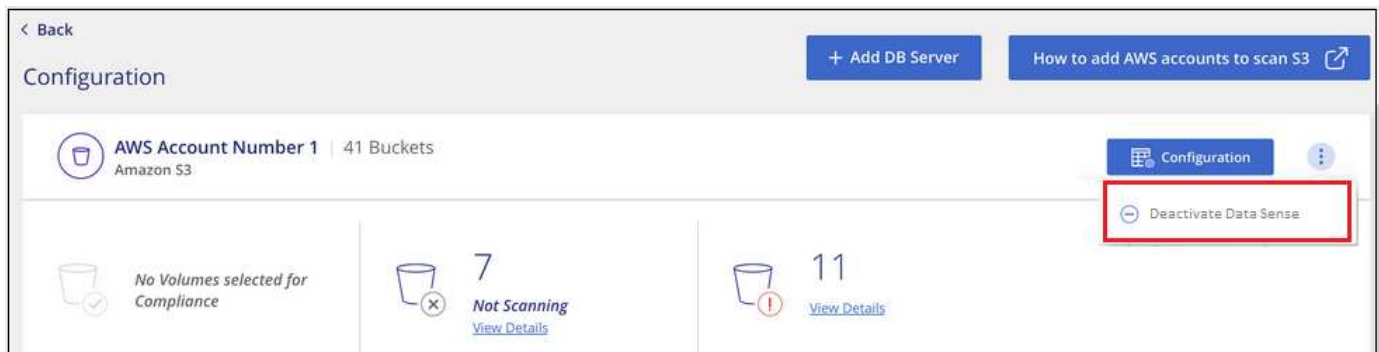
Removing data sources from Cloud Data Sense

If you need to, you can stop Cloud Data Sense from scanning one or more working environments, databases, file share groups, OneDrive accounts, Google Drive accounts, or SharePoint accounts. You can also delete the Cloud Data Sense instance if you no longer want to use Data Sense with your working environments.

Deactivating compliance scans for a working environment

When you deactivate scans, Cloud Data Sense no longer scans the data on the working environment and it removes the indexed compliance insights from the Data Sense instance (the data from the working environment itself isn't deleted).


1. From the *Configuration* page, click the  button in the row for the working environment, and then click **Deactivate Data Sense**.



You can also disable compliance scans for a working environment from the Services panel when you select the working environment.

Removing a database from Cloud Data Sense

If you no longer want to scan a certain database, you can delete it from the Cloud Data Sense interface and stop all scans.


1. From the *Configuration* page, click the  button in the row for the database, and then click **Remove DB Server**.



Removing a OneDrive, SharePoint, or Google Drive account from Cloud Data Sense

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

Steps

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




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

Removing a group of file shares from Cloud Data Sense

If you no longer want to scan user files from a file shares group, you can delete the File Shares Group from the Cloud Data Sense interface and stop all scans.

Steps

1. From the *Configuration* page, click the  button in the row for the File Shares Group, and then click **Remove File Shares Group**.



2. Click **Delete Group of Shares** from the confirmation dialog.

Reducing the Data Sense scan speed

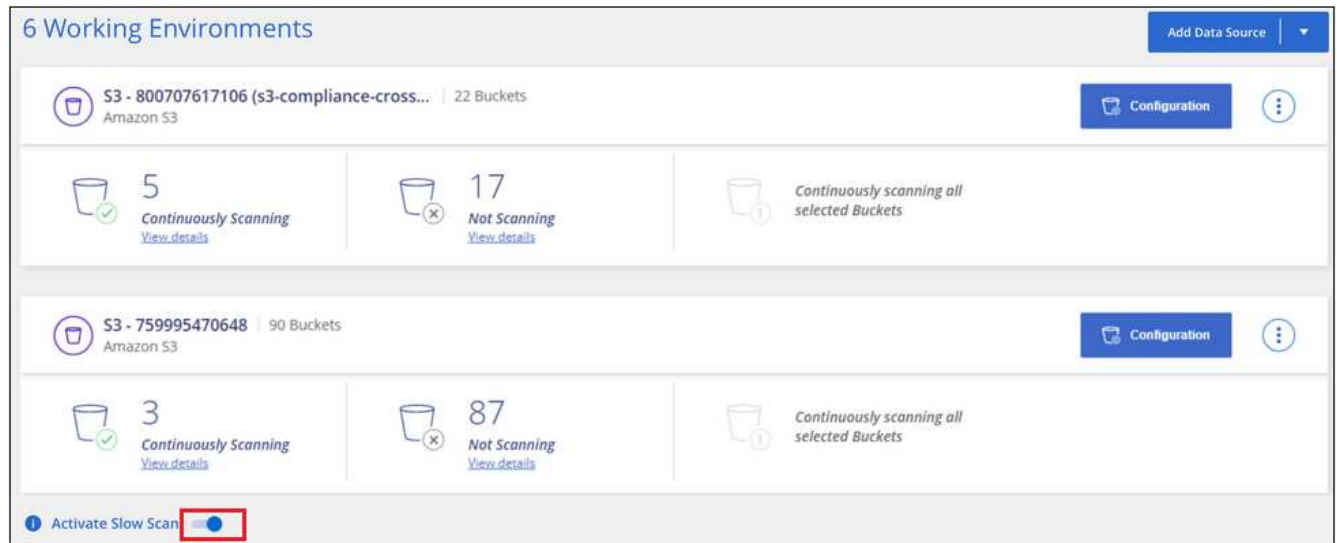
Data scans have a negligible impact on your storage systems and on your data. However, if you are concerned with even a very small impact, you can configure Data Sense to perform "slow" scans. When enabled, slow scanning is used on all data sources - you can't configure slow scanning for a single working environment or data source.



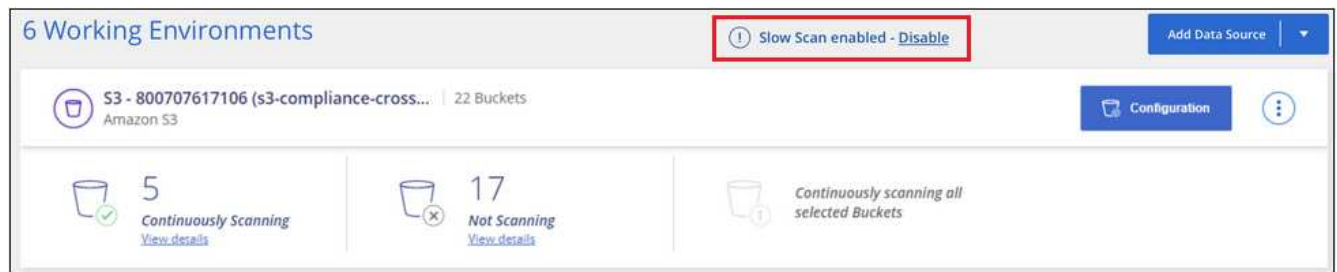
The scan speed can't be reduced when scanning databases.

Steps

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



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



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

Deleting the Cloud Data Sense instance

You can delete the Cloud Data Sense instance if you no longer want to use Data Sense. Deleting the instance also deletes the associated disks where the indexed data resides.

1. Go to your cloud provider's console and delete the Cloud Data Sense instance.

The instance is named *CloudCompliance* with a generated hash (UUID) concatenated to it. For example: *CloudCompliance-16bb6564-38ad-4080-9a92-36f5fd2f71c7*

Copyright Information

Copyright © 2022 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.

RESTRICTED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (c)(1)(ii) of the Rights in Technical Data and Computer Software clause at DFARS 252.277-7103 (October 1988) and FAR 52-227-19 (June 1987).

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.