



Gain insight into data privacy

Cloud Manager

NetApp
March 04, 2022

Table of Contents

- Gain insight into data privacy 1
 - Learn about Cloud Data Sense 1
 - Get started 7
 - Integrate your Active Directory with Cloud Data Sense 67
 - Set up licensing for Cloud Data Sense 69
 - Viewing governance details about the data stored in your organization 76
 - Viewing compliance details about the data stored in your organization. 80
 - Organizing your private data 91
 - Managing your private data 107
 - Adding personal data identifiers using Data Fusion 118
 - Viewing compliance reports 120
 - Responding to a Data Subject Access Request. 126
 - Categories of private data 128
 - Removing data sources from Cloud Data Sense 134
 - Frequently asked questions about Cloud Data Sense 137

Gain insight into data privacy

Learn about Cloud Data Sense

Cloud Data Sense is a data governance service for Cloud Manager that scans your corporate on-premises and cloud data sources and working environments to map and classify data, and to identify private information. This can help reduce your security and compliance risk, decrease storage costs, and assist with your data migration projects.

[Learn about the use cases for Cloud Data Sense.](#)

Features

Cloud Data Sense provides several tools that can help you with your compliance efforts. You can use Data Sense to:

- Identify Personal Identifiable Information (PII)
- Identify a wide scope of sensitive information as required by GDPR, CCPA, PCI, and HIPAA privacy regulations
- Respond to Data Subject Access Requests (DSAR)
- Notify Cloud Manager users through email when files contain certain PII (you define this criteria using [Policies](#))
- View and modify [Azure Information Protection \(AIP\) labels](#) in your files
- Add custom tags to files (for example, "needs to be moved") and assign a Cloud Manager user so that person can own updates to the files
- Copy, move, and delete files

Cloud Data Sense also provides tools that can help with your governance efforts. You can use Cloud Data Sense to:

- Identify the stale data, non-business data, duplicate files, and very large files in your systems.

You can use this information to decide whether you want to move, delete, or tier some files to less expensive object storage.

- View the size of data and whether any of the data contains sensitive information prior to moving it.

This is useful if you are planning to migrate data from on-premises locations to the cloud.

Supported working environments and data sources

Cloud Data Sense can scan data from the following types of working environments and data sources:

Working environments:

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

- Amazon FSx for ONTAP
- Amazon S3

Data sources:

- Non-NetApp file shares
- Object storage (that uses S3 protocol)
- Databases
- OneDrive accounts
- SharePoint accounts

Data Sense supports NFS versions 3.x, 4.0, and 4.1, and CIFS versions 1.x, 2.0, 2.1, and 3.0.



A Beta feature released in 2021 allows you to run compliance scans on the backup files (created using [Cloud Backup](#)) from your on-prem ONTAP volumes. This gives you the option to have Cloud Data Sense scan your on-prem ONTAP volumes directly, or scan the backup files made from those volumes.

Cost

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

[Learn how to license Cloud Data Sense.](#)

Note: When scanning backup files created from your on-prem ONTAP systems, the scanned data does not count against the capacity purchased in your PAYGO subscription or BYOL license.

- Installing Cloud Data Sense in the cloud requires deploying a cloud instance, which results in charges from the cloud provider where it is deployed. See [the type of instance that is deployed for each cloud provider](#). There is no cost if you install Data Sense on an on-premises system.
- Cloud Data Sense requires that you have deployed a Connector. In many cases you already have a Connector because of other storage and services you are using in Cloud Manager. The Connector instance results in charges from the cloud provider where it is deployed. See the [type of instance that is deployed for each cloud provider](#). There is no cost if you install the Connector on an on-premises system.

Data transfer costs

Data transfer costs depend on your setup. If the Cloud Data Sense instance and data source are in the same Availability Zone and region, then there are no data transfer costs. But if the data source, such as a Cloud Volumes ONTAP cluster or S3 Bucket, is in a *different* Availability Zone or region, then you'll be charged by your cloud provider for data transfer costs. See these links for more details:

- [AWS: Amazon EC2 Pricing](#)
- [Microsoft Azure: Bandwidth Pricing Details](#)
- [Google Cloud: Storage Transfer Service pricing](#)

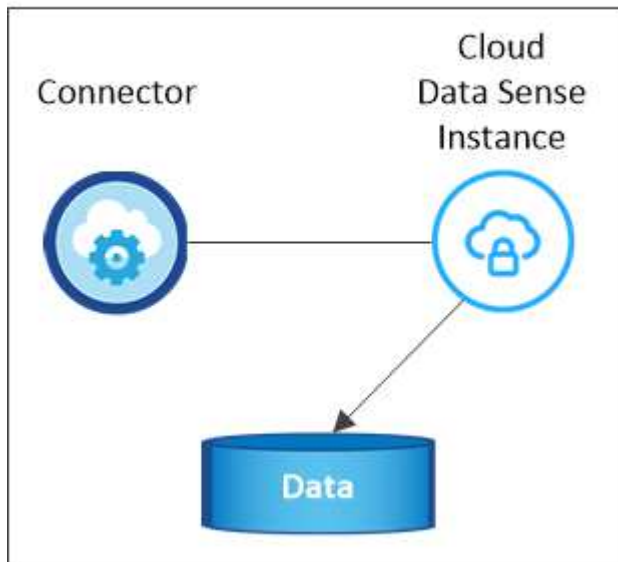
The Cloud Data Sense instance

When you deploy Data Sense in the cloud, Cloud Manager deploys the instance in the same subnet as the Connector. [Learn more about Connectors.](#)



If the Connector is installed on-prem, it deploys the Cloud Data Sense instance in same VPC or VNet as the first Cloud Volumes ONTAP system in the request. You can install Data Sense on-prem as well.

VPC or VNet



Note the following about the default instance:

- In AWS, Cloud Data Sense runs on an [m5.4xlarge instance](#) with a 500 GB GP2 disk. The operating system image is Amazon Linux 2 (Red Hat 7.3.1).

In regions where m5.4xlarge isn't available, Data Sense runs on an m4.4xlarge instance instead.

- In Azure, Cloud Data Sense runs on a [Standard_D16s_v3 VM](#) with a 512 GB disk. The operating system image is CentOS 7.8.
- In GCP, Cloud Data Sense runs on an [n2-standard-16 VM](#) with a 512 GB Standard persistent disk. The operating system image is CentOS 7.9.

In regions where n2-standard-16 isn't available, Data Sense runs on an n2d-standard-16 or n1-standard-16 VM instead.

- The instance is named *CloudCompliance* with a generated hash (UUID) concatenated to it. For example: *CloudCompliance-16bb6564-38ad-4080-9a92-36f5fd2f71c7*
- Only one Data Sense instance is deployed per Connector.
- Upgrades of Data Sense software is automated as long as the instance has internet access.



The instance should remain running at all times because Cloud Data Sense continuously scans the data.

Using a smaller instance type

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

System size	Specs	Limitations
Extra Large (default)	16 CPUs, 64 GB RAM, 500 GB SSD	None
Medium	8 CPUs, 32 GB RAM, 200 GB SSD	Slower scanning, and can only scan up to 1 million files.
Small	8 CPUs, 16 GB RAM, 100 GB SSD	Same limitations as "Medium", plus the ability to identify data subject names inside files is disabled.

When deploying Data Sense in the cloud, email ng-contact-data-sense@netapp.com for assistance if you want to use one of these smaller systems. We'll need to work with you to deploy these smaller cloud configurations.

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

How Cloud Data Sense works

At a high-level, Cloud Data Sense works like this:

1. You deploy an instance of Data Sense in Cloud Manager.
2. You enable high-level mapping or deep-level scanning on one or more working environments or data sources.
3. Data Sense scans the data using an AI learning process.
4. You use the provided dashboards and reporting tools to help in your compliance and governance efforts.

How scans work

After you enable Cloud Data Sense and select the volumes, buckets, database schemas, or OneDrive or SharePoint user data you want to scan, it immediately starts scanning the data to identify personal and sensitive data. It maps your organizational data, categorizes each file, and identifies and extracts entities and predefined patterns in the data. The result of the scan is an index of personal information, sensitive personal information, data categories, and file types.

Data Sense connects to the data like any other client by mounting NFS and CIFS volumes. NFS volumes are automatically accessed as read-only, while you need to provide Active Directory credentials to scan CIFS volumes.



After the initial scan, Data Sense continuously scans your data to detect incremental changes (this is why it's important to keep the instance running).

You can enable and disable scans at the volume level, at the bucket level, at the database schema level, at the OneDrive user level, and at the SharePoint site level.

What's the difference between Mapping and Classification scans

Cloud Data Sense enables you to run a general "mapping" scan on selected working environments and data sources. Mapping provides only a high-level overview of your data, whereas Classification provides deep-level scanning of your data. Mapping can be done on your data sources very quickly because it does not access files to see the data inside.

Many users like this functionality because they want to quickly scan their data to identify the data sources that require more research - and then they can enable classification scans only on those data sources or volumes.

The table below shows some of the differences:

Feature	Classification	Mapping
Scan speed	Slow	Fast
List of file types and used capacity	Yes	Yes
Number of files and used capacity	Yes	Yes
Age and size of files	Yes	Yes
Ability to run a Data Mapping Report	Yes	Yes
Data Investigation page to view file details	Yes	No
Search for names within files	Yes	No
Create policies that provide custom search results	Yes	No
Categorize data using AIP labels and Status tags	Yes	No
Copy, delete, and move source files	Yes	No

Feature	Classification	Mapping
Ability to run other reports	Yes	No

Information that Cloud Data Sense indexes

Data Sense collects, indexes, and assigns categories to your data (files). The data that Data Sense indexes includes the following:

Standard metadata

Cloud Data Sense collects standard metadata about files: the file type, its size, creation and modification dates, and so on.

Personal data

Personally identifiable information such as email addresses, identification numbers, or credit card numbers. [Learn more about personal data.](#)

Sensitive personal data

Special types of sensitive information, such as health data, ethnic origin, or political opinions, as defined by GDPR and other privacy regulations. [Learn more about sensitive personal data.](#)

Categories

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. [Learn more about categories.](#)

Types

Cloud Data Sense takes the data that it scanned and breaks it down by file type. [Learn more about types.](#)

Name entity recognition

Cloud Data Sense uses AI to extract natural persons' names from documents. [Learn about responding to Data Subject Access Requests.](#)

Networking overview

Cloud Manager deploys the Cloud Data Sense instance with a security group that enables inbound HTTP connections from the Connector instance.

When using Cloud Manager in SaaS mode, the connection to Cloud Manager is served over HTTPS, and the private data sent between your browser and the Data Sense instance are secured with end-to-end encryption, which means NetApp and third parties can't read it.

If you need to use the local user interface instead of the SaaS user interface for any reason, you can still [access the local UI](#).

Outbound rules are completely open. Internet access is needed to install and upgrade the Data Sense software and to send usage metrics.

If you have strict networking requirements, [learn about the endpoints that Cloud Data Sense contacts.](#)

User access to compliance information

The role each user has been assigned provides different capabilities within Cloud Manager and within Cloud Data Sense:

- An **Account Admin** can manage compliance settings and view compliance information for all working environments.
- A **Workspace Admin** can manage compliance settings and view compliance information only for systems that they have permissions to access. If a Workspace Admin can't access a working environment in Cloud Manager, then they can't see any compliance information for the working environment in the Data Sense tab.
- Users with the **Compliance Viewer** role can only view compliance information and generate reports for systems that they have permission to access. These users cannot enable/disable scanning of volumes, buckets, or database schemas. These users can't copy, move, or delete files either.

[Learn more about Cloud Manager roles](#) and how to [add users with specific roles](#).

Get started

Deploy Cloud Data Sense

Deploy Cloud Data Sense in the cloud

Complete a few steps to deploy Cloud Data Sense in the cloud.

Note that you can also [deploy Data Sense on a Linux host that has internet access](#). The type of installation may be a good option if you prefer to scan on-premises ONTAP systems using a Data Sense instance that's also located on premises — but this is not a requirement. The software functions exactly the same way regardless of which installation method you choose.

Quick start

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

1

Create a Connector

If you don't already have a Connector, create a Connector now. See [creating a Connector in AWS](#), [creating a Connector in Azure](#), or [creating a Connector in GCP](#).

You can also [deploy the Connector on-premises](#) on a Linux host in your network or in the cloud.

2

Review prerequisites

Ensure that your environment can meet the prerequisites. This includes outbound internet access for the instance, connectivity between the Connector and Cloud Data Sense over port 443, and more. [See the complete list](#).

The default configuration requires 16 vCPUs for the Cloud Data Sense instance. See [more details about the instance type](#).

3

Deploy Cloud Data Sense

Launch the installation wizard to deploy the Cloud Data Sense instance in the cloud.

4

Subscribe to the Cloud Data Sense service

The first 1 TB of data that Cloud Data Sense scans in Cloud Manager is free. A Cloud Manager subscription through your cloud provider Marketplace, or a BYOL license from NetApp, is required to continue scanning data after that point.

Create a Connector

If you don't already have a Connector, create a Connector in your cloud provider. See [creating a Connector in AWS](#) or [creating a Connector in Azure](#), or [creating a Connector in GCP](#). In most cases you will probably have a Connector set up before you attempt to activate Cloud Data Sense because most [Cloud Manager features require a Connector](#), but there are cases where you'll need to set one up now.

There are some scenarios where you have to use a Connector that's deployed in a specific cloud provider:

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

On-prem ONTAP systems, non-NetApp file shares, generic S3 Object storage, databases, OneDrive folders, and SharePoint accounts can be scanned when using any of these cloud Connectors.

Note that you can also [deploy the Connector on-premises](#) on a Linux host in your network or in the cloud. Some users planning to install Data Sense on-prem may also choose to install the Connector on-prem.

As you can see, there may be some situations where you need to use [multiple Connectors](#).



If you're planning on scanning Azure NetApp Files volumes, you need to make sure you're deploying in the same region as the volumes you wish to scan.

Review prerequisites

Review the following prerequisites to make sure that you have a supported configuration before you deploy Cloud Data Sense in the cloud.

Enable outbound internet access from Cloud Data Sense

Cloud Data Sense requires outbound internet access. If your virtual or physical network uses a proxy server for internet access, ensure that the Data Sense instance has outbound internet access to contact the following endpoints. When you deploy Data Sense in the cloud, it's located in the same subnet as the Connector.

Review the appropriate table below depending on whether you are deploying Cloud Data Sense in AWS, Azure, or GCP.

Required endpoints for AWS deployments:

Endpoints	Purpose
https://cloudmanager.cloud.netapp.com	Communication with the Cloud Manager service, which includes NetApp accounts.

Endpoints	Purpose
https://netapp-cloud-account.auth0.com https://auth0.com	Communication with NetApp Cloud Central for centralized user authentication.
https://cloud-compliance-support-netapp.s3.us-west-2.amazonaws.com https://hub.docker.com https://auth.docker.io https://registry-1.docker.io https://index.docker.io/ https://dseasb33srrn.cloudfront.net/ https://production.cloudflare.docker.com/	Provides access to software images, manifests, and templates.
https://kinesis.us-east-1.amazonaws.com	Enables NetApp to stream data from audit records.
https://cognito-idp.us-east-1.amazonaws.com https://cognito-identity.us-east-1.amazonaws.com https://user-feedback-store-prod.s3.us-west-2.amazonaws.com https://customer-data-production.s3.us-west-2.amazonaws.com	Enables Cloud Data Sense to access and download manifests and templates, and to send logs and metrics.

Required endpoints for Azure and GCP deployments:

Endpoints	Purpose
https://cloudmanager.cloud.netapp.com	Communication with the Cloud Manager service, which includes NetApp accounts.
https://netapp-cloud-account.auth0.com https://auth0.com	Communication with NetApp Cloud Central for centralized user authentication.
https://support.compliance.cloudmanager.cloud.netapp.com/ https://hub.docker.com https://auth.docker.io https://registry-1.docker.io https://index.docker.io/ https://dseasb33srrn.cloudfront.net/ https://production.cloudflare.docker.com/	Provides access to software images, manifests, templates, and to send logs and metrics.
https://support.compliance.cloudmanager.cloud.netapp.com/	Enables NetApp to stream data from audit records.

Ensure that Cloud Manager has the required permissions

Ensure that Cloud Manager has permissions to deploy resources and create security groups for the Cloud Data Sense instance. You can find the latest Cloud Manager permissions in [the policies provided by NetApp](#).

Note: If you created the Connector in GCP using Cloud Manager 3.9.10 or greater, then you're all set. If you created the Connector using an earlier version, then you'll need to add the following permissions to the GCP service account associated with the Connector to deploy Cloud Data Sense to GCP.

```
compute.instances.addAccessConfig
compute.subnetworks.use
compute.subnetworks.useExternalIp
```

Check your vCPU limits

Ensure that your cloud provider's vCPU limit allows for the deployment of an instance with 16 cores. You'll need to verify the vCPU limit for the relevant instance family in the region where Cloud Manager is running. [See the required instance types](#).

See the following links for more details on vCPU limits:

- [AWS documentation: Amazon EC2 service quotas](#)
- [Azure documentation: Virtual machine vCPU quotas](#)
- [Google Cloud documentation: Resource quotas](#)

Note that you can deploy Data Sense on a system with fewer CPUs and less RAM, but there are limitations when using these systems. See [Using a smaller instance type](#) for details.

Ensure that the Cloud Manager Connector can access Cloud Data Sense

Ensure connectivity between the Connector and the Cloud Data Sense instance. The security group for the Connector must allow inbound and outbound traffic over port 443 to and from the Data Sense instance. This connection enables deployment of the Data Sense instance and enables you to view information in the Compliance and Governance tabs.

Additional inbound and outbound rules are required for AWS and AWS GovCloud deployments. See [Rules for the Connector in AWS](#) for details.

Ensure that you can keep Cloud Data Sense running

The Cloud Data Sense instance needs to stay on to continuously scan your data.

Ensure web browser connectivity to Cloud Data Sense

After Cloud Data Sense is enabled, ensure that users access the Cloud Manager interface from a host that has a connection to the Data Sense instance.

The Data Sense instance uses a private IP address to ensure that the indexed data isn't accessible to the internet. As a result, the web browser that you use to access Cloud Manager must have a connection to that private IP address. That connection can come from a direct connection to your cloud provider (for example, a VPN), or from a host that's inside the same network as the Data Sense instance.

Deploy Data Sense in the cloud

Follow these steps to deploy an instance of Cloud Data Sense in the cloud.

Steps

1. In Cloud Manager, click **Data Sense**.
2. Click **Activate Data Sense**.



3. Click **Activate Data Sense** to start the cloud deployment wizard.



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



5. When the instance is deployed, click **Continue to configuration** to go to the *Configuration* page.

Result

Cloud Manager deploys the Cloud Data Sense instance in your cloud provider.

What's Next

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

You can also [set up licensing for Cloud Data Sense](#) at this time. You will not be charged until the amount of data exceeds 1 TB.

Deploy Cloud Data Sense on a Linux host that has internet access

Complete a few steps to deploy Cloud Data Sense on a Linux host in your network, or in the cloud, that has internet access.

The on-prem installation may be a good option if you prefer to scan on-premises ONTAP systems using a Data Sense instance that's also located on premises — but this is not a requirement. The software functions exactly the same way regardless of which installation method you choose.

Note that you can also [deploy Data Sense in an on-premises site that doesn't have internet access](#) for completely secure sites.

Quick start

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

1

Create a Connector

If you don't already have a Connector, create a Connector now. See [creating a Connector in AWS](#), [creating a Connector in Azure](#), or [creating a Connector in GCP](#).

You can also [deploy the Connector on-premises](#) on a Linux host in your network or in the cloud.

2

Review prerequisites

Ensure that your environment can meet the prerequisites. This includes outbound internet access for the instance, connectivity between the Connector and Cloud Data Sense over port 443, and more. [See the complete list](#).

You also need a Linux system that meets the [following requirements](#).

3

Deploy Cloud Data Sense

Download the Cloud Data Sense software from the NetApp Support Site and copy the installer file to the Linux host you plan to use. Then launch the installation wizard and follow the prompts to deploy the Data Sense instance.

4

Subscribe to the Cloud Data Sense service

The first 1 TB of data that Cloud Data Sense scans in Cloud Manager is free. A subscription to your cloud provider Marketplace, or a BYOL license from NetApp, is required to continue scanning data after that point.

Create a Connector

If you don't already have a Connector, create a Connector in your cloud provider. See [creating a Connector in AWS](#) or [creating a Connector in Azure](#), or [creating a Connector in GCP](#). In most cases you will probably have a Connector set up before you attempt to activate Cloud Data Sense because most [Cloud Manager features require a Connector](#), but there are cases where you'll need to set one up now.

There are some scenarios where you have to use a Connector that's deployed in a specific cloud provider:

- When scanning data in Cloud Volumes ONTAP in AWS, Amazon FSx for ONTAP, or in AWS S3 buckets, you use a connector in AWS.
- When scanning data in Cloud Volumes ONTAP in Azure or in Azure NetApp Files, you use a connector in Azure.
- When scanning data in Cloud Volumes ONTAP in GCP, you use a connector in GCP.

On-prem ONTAP systems, non-NetApp file shares, generic S3 Object storage, databases, OneDrive folders, and SharePoint accounts can be scanned using any of these cloud Connectors.

Note that you can also [deploy the Connector on-premises](#) on a Linux host in your network or in the cloud. Some users planning to install Data Sense on-prem may also choose to install the Connector on-prem.

As you can see, there may be some situations where you need to use [multiple Connectors](#).



If you're planning on scanning Azure NetApp Files volumes, you need to make sure you're deploying in the same region as the volumes you wish to scan.

Prepare the Linux host system

Data Sense software must run on a host that meets specific operating system requirements, RAM requirements, software requirements, and so on.

- Operating system: Red Hat Enterprise Linux or CentOS version 8.0 or 8.1
 - Version 7.8 can be used, but the Linux kernel version must be 4.14 or greater
 - The OS must be capable of installing the docker engine (for example, disable the *firewalld* service if needed)
- Disk: SSD with 500 GiB available on /, or
 - 100 GiB available on /opt
 - 400 GiB available on /var
- RAM: 64 GB (swap memory must be disabled on the host)
- CPU: 16 cores

Note that you can deploy Data Sense on a system with fewer CPUs and less RAM, but there are limitations when using these systems. See [Using a smaller instance type](#) for details.

- A Red Hat Enterprise Linux system must be registered with Red Hat Subscription Management. If it's not registered, the system can't access repositories to update required 3rd party software during installation.
- The following software must be installed on the host. If it doesn't already exist on the host, then the installer will install the software for you:
 - Docker Engine version 19 or later. [View installation instructions](#).

- Python 3 version 3.6 or later. [View installation instructions.](#)

Verify Cloud Manager and Data Sense prerequisites

Review the following prerequisites to make sure that you have a supported configuration before you deploy Cloud Data Sense on a Linux system.

Enable outbound internet access from Cloud Data Sense

Cloud Data Sense requires outbound internet access. If your virtual or physical network uses a proxy server for internet access, ensure that the Data Sense instance has outbound internet access to contact the following endpoints.

Endpoints	Purpose
https://cloudmanager.cloud.netapp.com	Communication with the Cloud Manager service, which includes NetApp accounts.
https://netapp-cloud-account.auth0.com https://auth0.com	Communication with NetApp Cloud Central for centralized user authentication.
https://support.compliance.cloudmanager.cloud.netapp.com/ https://hub.docker.com https://auth.docker.io https://registry-1.docker.io https://index.docker.io/ https://dseasb33srrn.cloudfront.net/ https://production.cloudflare.docker.com/	Provides access to software images, manifests, templates, and to send logs and metrics.
https://support.compliance.cloudmanager.cloud.netapp.com/	Enables NetApp to stream data from audit records.
https://github.com/docker https://download.docker.com http://mirror.centos.org http://mirrorlist.centos.org http://mirror.centos.org/centos/7/extras/x86_64/Packages/container-selinux-2.107-3.el7.noarch.rpm	Provides prerequisite packages for installation.

Ensure that Cloud Manager has the required permissions

Ensure that Cloud Manager has permissions to deploy resources and create security groups for the Cloud Data Sense instance. You can find the latest Cloud Manager permissions in [the policies provided by NetApp](#).

Note: If you created the Connector in GCP using Cloud Manager 3.9.10 or greater, then you're all set. If you created the Connector using an earlier version, then you'll need to add the following permissions to the GCP service account associated with the Connector to deploy Cloud Data Sense to GCP.

```
compute.instances.addAccessConfig
compute.subnetworks.use
compute.subnetworks.useExternalIp
```


Ensure that the Cloud Manager Connector can access Cloud Data Sense

Ensure connectivity between the Connector and the Cloud Data Sense instance. The security group for the Connector must allow inbound and outbound traffic over port 443 to and from the Data Sense instance.

This connection enables deployment of the Data Sense instance and enables you to view information in the Compliance and Governance tabs.

Make sure port 8080 is open so you can see the installation progress in Cloud Manager.

Ensure that you can keep Cloud Data Sense running

The Cloud Data Sense instance needs to stay on to continuously scan your data.

Ensure web browser connectivity to Cloud Data Sense

After Cloud Data Sense is enabled, ensure that users access the Cloud Manager interface from a host that has a connection to the Data Sense instance.

The Data Sense instance uses a private IP address to ensure that the indexed data isn't accessible to the internet. As a result, the web browser that you use to access Cloud Manager must have a connection to that private IP address. That connection can come from a direct connection to your cloud provider (for example, a VPN), or from a host that's inside the same network as the Data Sense instance.

Deploy Data Sense on premises

For typical configurations you'll install the software on a single host system. [See those steps here.](#)

For very large configurations where you'll be scanning petabytes of data, you can include multiple hosts to provide additional processing power. [See those steps here.](#)

See [Preparing the Linux host system](#) and [Reviewing prerequisites](#) for the full list of requirements before you deploy Cloud Data Sense.

Upgrades to Data Sense software is automated as long as the instance has internet connectivity.



Cloud Data Sense is currently unable to scan S3 buckets, Azure NetApp Files, or FSx for ONTAP when the software is installed on premises. In these cases you'll need to deploy a separate Connector and instance of Data Sense in the cloud and [switch between Connectors](#) for your different data sources.

Single-host installation for typical configurations

Follow these steps when installing Data Sense software on a single on-premises host.

What you'll need

- Verify that your Linux system meets the [host requirements](#).
- (Optional) Verify that the system has the two prerequisite software packages installed (Docker Engine and Python 3). The installer will install this software if it is not already on the system.
- Make sure you have root privileges on the Linux system.
- Verify that your offline environment meets the required [permissions and connectivity](#).

Steps

1. Download the Cloud Data Sense software from the [NetApp Support Site](#). The file you should select is named `cc_onprem_installer_<version>.tar.gz`.

2. Copy the installer file to the Linux host you plan to use (using `scp` or some other method).
3. In Cloud Manager, click **Data Sense**.
4. Click **Activate Data Sense**.



5. Click **Activate Data Sense** to start the on-prem deployment wizard.



6. In the *Deploy Data Sense On Premises* dialog, copy the provided command and paste it in a text file so you can use it later, and click **Close**. For example:

```
sudo ./install.sh -a 12345 -c 27AG75 -t 2198qq
```

7. Unzip the installer file on the host machine, for example:

```
tar -xzf cc_onprem_installer_1.7.2.tar.gz
```

8. When prompted by the installer, you can enter the required values in a series of prompts, or you can enter the complete command in the first prompt:

Enter parameters as prompted:	Enter the full command:
<ol style="list-style-type: none"> 1. Paste the information you copied from step 6: <code>sudo ./install.sh -a <account_id> -c <agent_id> -t <token></code> 2. Enter the IP address or host name of the Data Sense host machine so it can be accessed by the Connector instance. 3. Enter the IP address or host name of the Cloud Manager Connector host machine so it can be accessed by the Data Sense instance. 4. Enter proxy details as prompted. If your Cloud Manager already uses a proxy, there is no need to enter this information again here since Data Sense will automatically use the proxy used by Cloud Manager. 	<p>Alternatively, you can create the whole command in advance and enter it in the first prompt:</p> <pre>sudo ./install.sh -a <account_id> -c <agent_id> -t <token> --host <ds_host> --cm-host <cm_host> --proxy-host <proxy_host> --proxy-port <proxy_port> --proxy-scheme <proxy_scheme> --proxy -user <proxy_user> --proxy-password <proxy_password></pre>

Variable values:

- *account_id* = NetApp Account ID
- *agent_id* = Connector ID
- *token* = jwt user token
- *ds_host* = IP address or host name of the Data Sense Linux system.
- *cm_host* = IP address or host name of the Cloud Manager Connector system.
- *proxy_host* = IP or host name of the proxy server if the host is behind a proxy server.
- *proxy_port* = Port to connect to the proxy server (default 80).
- *proxy_scheme* = Connection scheme: https or http (default http).
- *proxy_user* = Authenticated user to connect to the proxy server, if basic authentication is required.
- *proxy_password* = Password for the user name that you specified.

Result

The Cloud Data Sense installer installs packages, installs docker, registers the installation, and installs Data Sense. Installation can take 10 to 20 minutes.

If there is connectivity over port 8080 between the host machine and the Connector instance, you will see the installation progress in the Data Sense tab in Cloud Manager.

What's Next

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

You can also [set up licensing for Cloud Data Sense](#) at this time. You will not be charged until the amount of data exceeds 1 TB.

Multi-host installation for large configurations

For very large configurations where you'll be scanning petabytes of data, you can include multiple hosts to provide additional processing power. When using multiple host systems, the primary system is called the

Manager node and the additional systems that provide extra processing power are called *Scanner nodes*.

Follow these steps when installing Data Sense software on multiple on-premises hosts.

What you'll need

- Verify that all your Linux systems for the Manager and Scanner nodes meet the [host requirements](#).
- (Optional) Verify that the systems have the two prerequisite software packages installed (Docker Engine and Python 3). The installer will install this software if it is not already on the systems.
- Make sure you have root privileges on the Linux systems.
- Verify that your environment meets the required [permissions and connectivity](#).
- You must have the IP addresses of the scanner node hosts that you plan to use.
- The following ports and protocols must be enabled on all hosts:

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

Steps

1. Follow steps 1 through 7 from the [Single-host installation](#) on the manager node.
2. As shown in step 8, when prompted by the installer, you can enter the required values in a series of prompts, or you can enter the complete command in the first prompt.

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

For example, this command adds 3 scanner nodes:

```
sudo ./install.sh -a <account_id> -c <agent_id> -t <token> --host <ds_host>
--cm-host <cm_host> -n <node_ip1>,<node_ip2>,<node_ip3> --proxy-host
<proxy_host> --proxy-port <proxy_port> --proxy-scheme <proxy_scheme> --proxy
-user <proxy_user> --proxy-password <proxy_password>
```

3. Before the manager node installation completes, a dialog displays the installation command needed for the scanner nodes. Copy the command and save it in a text file. For example:

```
sudo ./node_install.sh -m 10.11.12.13 -t ABCDEF-1-3u69m1-1s35212
```

4. On **each** scanner node host:
 - a. Copy the Data Sense installer file (**cc_onprem_installer_<version>.tar.gz**) to the host machine (using `scp` or some other method).
 - b. Unzip the installer file.

- c. Paste and execute the command that you copied in step 3.

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

Result

The Cloud Data Sense installer finishes installing packages, docker, and registers the installation. Installation can take 10 to 20 minutes.

What's Next

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

You can also [set up licensing for Cloud Data Sense](#) at this time. You will not be charged until the amount of data exceeds 1 TB.

Deploy Cloud Data Sense on prem without internet access

Complete a few steps to deploy Cloud Data Sense on a host in an on-premises site that doesn't have internet access. This type of installation is perfect for your secure sites.

Note that you can also [deploy Data Sense in an on-premises site that has internet access](#).

Supported data sources

When installed in this manner (sometimes called an "offline" or "dark" site), Data Sense can only scan data from data sources that are also local to the on-premises site. At this time, Data Sense can scan the following local data sources:

- On-premises ONTAP systems
- Database schemas
- Non-NetApp NFS or CIFS file shares
- Object Storage that uses the Simple Storage Service (S3) protocol

There is no support currently for scanning Cloud Volumes ONTAP, Azure NetApp Files, FSx for ONTAP, OneDrive accounts, SharePoint accounts, or AWS S3 accounts.

Limitations

Most Data Sense features work when it is deployed in a site with no internet access. However, certain features that require internet access are not supported, for example:

- Managing Microsoft Azure Information Protection (AIP) labels
- Automated software upgrades from Cloud Manager

Both the Cloud Manager Connector and Data Sense will require periodic manual upgrades to enable new features. You can see the Data Sense version at the bottom of the Data Sense UI pages. Check the [Cloud Manager What's New](#) to see the new features in each release and for when a software update package is available. Then you can follow the steps to [upgrade your Data Sense software](#).

Quick start

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

1

Install the Cloud Manager Connector

If you don't already have a Connector installed at your offline on-premises site, [deploy the Connector](#) on a Linux host now.

2

Review Data Sense prerequisites

Ensure that your Linux system meets the [host requirements](#), that it has all required software installed, and that your offline environment meets the required [permissions and connectivity](#).

3

Deploy Data Sense

Download the Cloud Data Sense software from the NetApp Support Site and copy the installer file to the Linux host you plan to use. Then launch the installation wizard and follow the prompts to deploy the Cloud Data Sense instance.

4

Subscribe to the Cloud Data Sense service

The first 1 TB of data that Cloud Data Sense scans in Cloud Manager is free. A BYOL license from NetApp is required to continue scanning data after that point.

Install the Cloud Manager Connector

If you don't already have a Cloud Manager Connector installed at your offline on-premises site, [deploy the Connector](#) on a Linux host in your offline site.

Prepare the Linux host system

Data Sense software must run on a host that meets specific operating system requirements, RAM requirements, software requirements, and so on.

- Operating system: Red Hat Enterprise Linux or CentOS version 8.0 or 8.1
 - Version 7.8 can be used, but the Linux kernel version must be 4.14 or greater
 - The OS must be capable of installing the Docker Engine (for example, disable the *firewalld* service if needed)
- Disk: SSD with 500 GiB available on /, or
 - 100 GiB available on /opt
 - 400 GiB available on /var
- RAM: 64 GB (swap memory must be disabled on the host)
- CPU: 16 cores

Note that you can deploy Data Sense on a system with fewer CPUs and less RAM, but there are limitations when using these systems. See [Using a smaller instance type](#) for details.

The following software must be installed on the host before you install Data Sense:

- Docker Engine version 19 or later. [View installation instructions.](#)
- Python 3 version 3.6 or later. [View installation instructions.](#)

Verify Cloud Manager and Data Sense prerequisites

Review the following prerequisites to make sure that you have a supported configuration before you deploy Cloud Data Sense.

- Ensure that Cloud Manager has permissions to deploy resources and create security groups for the Cloud Data Sense instance.
- Ensure that the Cloud Manager Connector can access the Data Sense instance. The security group for the Connector must allow inbound and outbound traffic over port 443 to and from the Data Sense instance.

This connection enables deployment of the Data Sense instance and enables you to view compliance and governance information.

Make sure port 8080 is open so you can see the installation progress in Cloud Manager.

- Ensure that you can keep Cloud Data Sense running. The Cloud Data Sense instance needs to stay on to continuously scan your data.
- Ensure web browser connectivity to Cloud Data Sense. After Cloud Data Sense is enabled, ensure that users access the Cloud Manager interface from a host that has a connection to the Data Sense instance.

The Data Sense instance uses a private IP address to ensure that the indexed data isn't accessible to others. As a result, the web browser that you use to access Cloud Manager must have a connection to that private IP address. That connection can come from a host that's inside the same network as the Data Sense instance.

Deploy Data Sense

For typical configurations you'll install the software on a single host system. [See those steps here.](#)

For very large configurations where you'll be scanning petabytes of data, you can include multiple hosts to provide additional processing power. [See those steps here.](#)

Single-host installation for typical configurations

Follow these steps when installing Data Sense software on a single on-premises host in an offline environment.

What you'll need

- Verify that your Linux system meets the [host requirements](#).
- Verify that you have installed the two prerequisite software packages (Docker Engine and Python 3).
- Make sure you have root privileges on the Linux system.
- Verify that your offline environment meets the required [permissions and connectivity](#).

Steps

1. On an internet-configured system, download the Cloud Data Sense software from the [NetApp Support Site](#). The file you should select is named **DataSense-offline-bundle-`<version>`.tar.gz**.

2. Copy the installer bundle to the Linux host you plan to use in the dark site.
3. Unzip the installer bundle on the host machine, for example:

```
tar -xzf DataSense-offline-bundle-v1.7.2.tar.gz
```

This extracts required software and the actual installation file **cc_onprem_installer_<version>.tar.gz**.

4. Launch Cloud Manager and click the **Data Sense** tab.
5. Click **Activate Data Sense**.



6. Click **Deploy** to start the on-prem deployment wizard.



7. In the *Deploy Data Sense On Premises* dialog, copy the provided command and paste it in a text file so you can use it later, and click **Close**. For example:

```
sudo ./install.sh -a 12345 -c 27AG75 -t 2198qq --darksite
```

8. Unzip the installation file on the host machine, for example:

```
tar -xzf cc_onprem_installer_1.7.2.tar.gz
```


- When prompted by the installer, you can enter the required values in a series of prompts, or you can enter the complete command in the first prompt:

Enter parameters as prompted:	Enter the full command:
<ol style="list-style-type: none"> Paste the information you copied from step 7: <code>sudo ./install.sh -a <account_id> -c <agent_id> -t <token> --darksite</code> Enter the IP address or host name of the Data Sense host machine so it can be accessed by the Connector instance. Enter the IP address or host name of the Cloud Manager Connector host machine so it can be accessed by the Data Sense instance. Enter proxy details as prompted. If your Cloud Manager already uses a proxy, there is no need to enter this information again here since Data Sense will automatically use the proxy used by Cloud Manager. 	<p>Alternatively, you can create the whole command in advance and enter it in the first prompt:</p> <pre>sudo ./install.sh -a <account_id> -c <agent_id> -t <token> --host <ds_host> --cm-host <cm_host> --proxy-host <proxy_host> --proxy-port <proxy_port> --proxy-scheme <proxy_scheme> --proxy -user <proxy_user> --proxy-password <proxy_password> --darksite</pre>

Variable values:

- *account_id* = NetApp Account ID
- *agent_id* = Connector ID
- *token* = jwt user token
- *ds_host* = IP address or host name of the Data Sense Linux system.
- *cm_host* = IP address or host name of the Cloud Manager Connector system.
- *proxy_host* = IP or host name of the proxy server if the host is behind a proxy server.
- *proxy_port* = Port to connect to the proxy server (default 80).
- *proxy_scheme* = Connection scheme: https or http (default http).
- *proxy_user* = Authenticated user to connect to the proxy server, if basic authentication is required.
- *proxy_password* = Password for the user name that you specified.

Result

The Data Sense installer installs packages, registers the installation, and installs Data Sense. Installation can take 10 to 20 minutes.

If there is connectivity over port 8080 between the host machine and the Connector instance, you will see the installation progress in the Data Sense tab in Cloud Manager.

What's Next

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

Multi-host installation for large configurations

For very large configurations where you'll be scanning petabytes of data, you can include multiple hosts to

provide additional processing power. When using multiple host systems, the primary system is called the *Manager node* and the additional systems that provide extra processing power are called *Scanner nodes*.

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

What you'll need

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

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

Steps

1. Follow steps 1 through 8 from the [Single-host installation](#) on the manager node.
2. As shown in step 9, when prompted by the installer, you can enter the required values in a series of prompts, or you can enter the complete command in the first prompt.

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

For example, this command adds 3 scanner nodes:

```
sudo ./install.sh -a <account_id> -c <agent_id> -t <token> --host <ds_host>
--cm-host <cm_host> -n <node_ip1>,<node_ip2>,<node_ip3> --proxy-host
<proxy_host> --proxy-port <proxy_port> --proxy-scheme <proxy_scheme> --proxy
-user <proxy_user> --proxy-password <proxy_password> --darksite
```

3. Before the manager node installation completes, a dialog displays the installation command needed for the scanner nodes. Copy the command and save it in a text file. For example:

```
sudo ./node_install.sh -m 10.11.12.13 -t ABCDEF-1-3u69m1-1s35212
```

4. On **each** scanner node host:
 - a. Copy the Data Sense installer file (**cc_onprem_installer_<version>.tar.gz**) to the host machine.
 - b. Unzip the installer file.

- c. Paste and run the command that you copied in step 3.

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

Result

The Cloud Data Sense installer finishes installing packages, and registers the installation. Installation can take 15 to 25 minutes.

What's Next

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

Upgrade Data Sense software

Since Data Sense software is updated with new features on a regular basis, you should get into a routine to check for new versions periodically to make sure you're using the newest software and features. You'll need to upgrade Data Sense software manually because there's no internet connectivity to perform the upgrade automatically.

Before you begin

- Data Sense software can be upgraded one major version at a time. For example, if you have version 1.7.x installed, you can upgrade only to 1.8.x. If you are a few major versions behind, you'll need to upgrade the software multiple times.
- Verify that your on-prem Connector software has been upgraded to the newest available version. [See the Connector upgrade steps](#).

Steps

1. On an internet-configured system, download the Cloud Data Sense software from the [NetApp Support Site](#). The file you should select is named **DataSense-offline-bundle-<version>.tar.gz**.
2. Copy the software bundle to the Linux host where Data Sense is installed in the dark site.
3. Unzip the software bundle on the host machine, for example:

```
tar -xvf DataSense-offline-bundle-v1.8.2.tar.gz
```

This extracts the upgrade script **start_darksite_upgrade.sh** and any required third-party software.

4. Run the upgrade script on the host machine, for example:

```
start_darksite_upgrade.sh
```

Result

The Data Sense software is upgraded on your host. The update can take 5 to 10 minutes.

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

You can verify that the software has been updated by checking the version at the bottom of the Data Sense UI

pages.

Activate scanning on your data sources

Getting started with Cloud Data Sense for Cloud Volumes ONTAP and on-premises ONTAP

Complete a few steps to start scanning your Cloud Volumes ONTAP and on-premises ONTAP volumes using Cloud Data Sense.

Quick start

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

1

Discover the data sources that you want to scan

Before you can scan volumes, you must add the systems as working environments in Cloud Manager:

- For Cloud Volumes ONTAP systems, these working environments should already be available in Cloud Manager
- For on-premises ONTAP systems, [Cloud Manager must discover the ONTAP clusters](#)

2

Deploy the Cloud Data Sense instance

[Deploy Cloud Data Sense](#) if there isn't already an instance deployed.

3

Enable Cloud Data Sense and select the volumes to scan

Click **Data Sense**, select the **Configuration** tab, and activate compliance scans for volumes in specific working environments.

4

Ensure access to volumes

Now that Cloud Data Sense is enabled, ensure that it can access all volumes.

- The Cloud Data Sense instance needs a network connection to each Cloud Volumes ONTAP subnet or on-prem ONTAP system.
- Security groups for Cloud Volumes ONTAP must allow inbound connections from the Data Sense instance.
- Make sure these ports are open to the Data Sense instance:
 - For NFS – ports 111 and 2049.
 - For CIFS – ports 139 and 445.
- NFS volume export policies must allow access from the Data Sense instance.
- Data Sense needs Active Directory credentials to scan CIFS volumes.

Click **Compliance > Configuration > Edit CIFS Credentials** and provide the credentials.

5

Manage the volumes you want to scan

Select or deselect the volumes that you want to scan and Cloud Data Sense will start or stop scanning them.

Discovering the data sources that you want to scan

If the data sources you want to scan are not already in your Cloud Manager environment, you can add them to the canvas at this time.

Your Cloud Volumes ONTAP systems should already be available in the Canvas in Cloud Manager. For on-premises ONTAP systems, you'll need to have [Cloud Manager discover these clusters](#).

Deploying the Cloud Data Sense instance

Deploy Cloud Data Sense if there isn't already an instance deployed.

If you are scanning Cloud Volumes ONTAP and on-premises ONTAP systems that are accessible over the internet, you can [deploy Cloud Data Sense in the cloud](#) or [in an on-premises location that has internet access](#).

If you are scanning on-premises ONTAP systems that have been installed in a dark site that has no internet access, you need to [deploy Cloud Data Sense in the same on-premises location that has no internet access](#). This also requires that the Cloud Manager Connector is deployed in that same on-premises location.

Upgrades to Data Sense software is automated as long as the instance has internet connectivity.

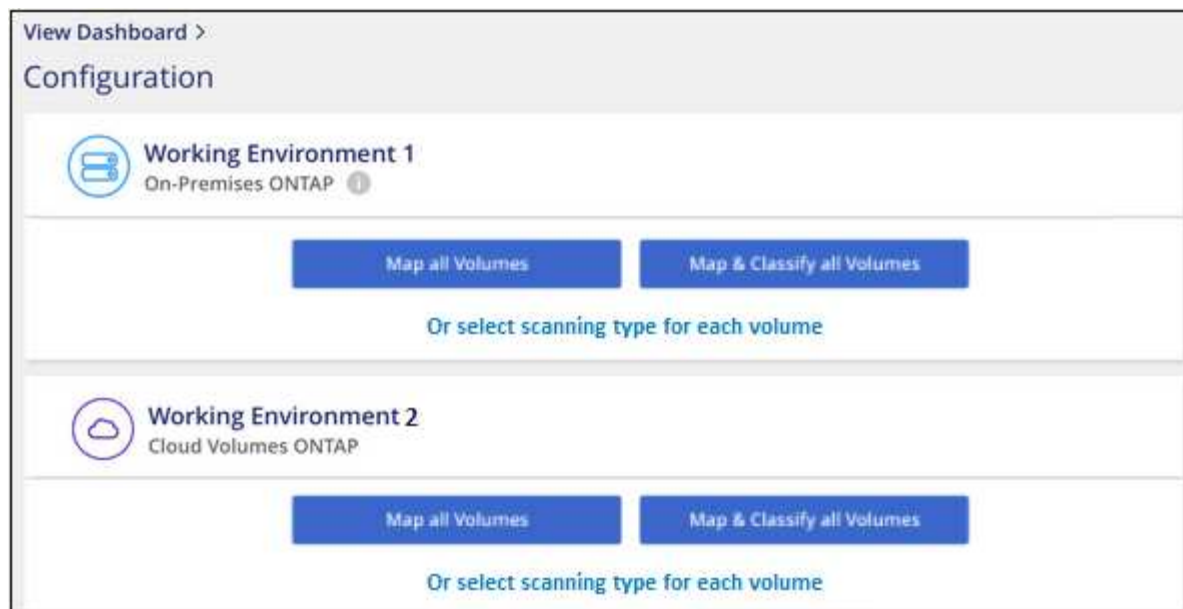
Enabling Cloud Data Sense in your working environments

You can enable Cloud Data Sense on Cloud Volumes ONTAP systems (in AWS, Azure, and GCP) and on on-premises ONTAP clusters.



Following these steps for on-prem ONTAP clusters scans the volumes directly on the on-prem ONTAP system. If you are already creating backup files from those on-prem systems using [Cloud Backup](#), you can scan the backup files in the cloud instead. Go to [Scanning backup files from on-premises ONTAP systems](#) for details.

1. At the top of Cloud Manager, click **Data Sense** and then select the **Configuration** tab.



2. Select how you want to scan the volumes in each working environment. [Learn about mapping and classification scans](#):
 - To map all volumes, click **Map all Volumes**.
 - To map and classify all volumes, click **Map & Classify all Volumes**.
 - To customize scanning for each volume, click **Or select scanning type for each volume**, and then choose the volumes you want to map and/or classify.

See [Enabling and disabling compliance scans on volumes](#) for details.

3. In the confirmation dialog box, click **Approve** to have Data Sense start scanning your volumes.

Result

Cloud Data Sense starts scanning the volumes you selected in the working environment. Results will be available in the Compliance dashboard as soon as Cloud Data Sense finishes the initial scans. The time that it takes depends on the amount of data—it could be a few minutes or hours.

Verifying that Cloud Data Sense has access to volumes

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

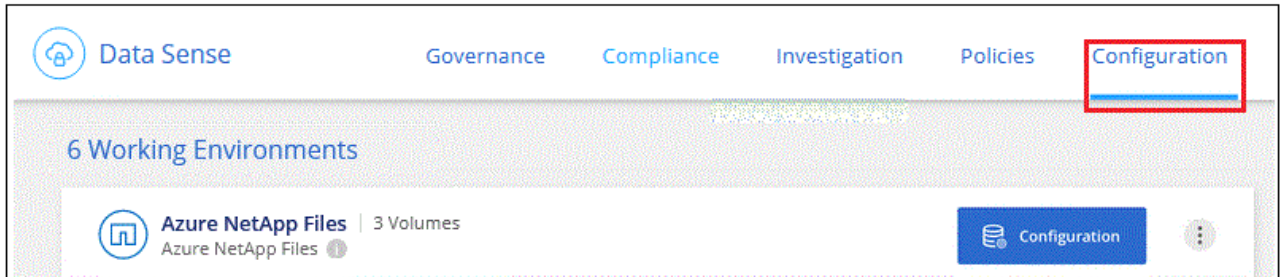
Steps

1. Make sure that there's a network connection between the Cloud Data Sense instance and each network that includes volumes for Cloud Volumes ONTAP or on-prem ONTAP clusters.
2. Ensure that the security group for Cloud Volumes ONTAP allows inbound traffic from the Data Sense instance.

You can either open the security group for traffic from the IP address of the Data Sense instance, or you can open the security group for all traffic from inside the virtual network.

3. Ensure the following ports are open to the Data Sense instance:
 - For NFS – ports 111 and 2049.

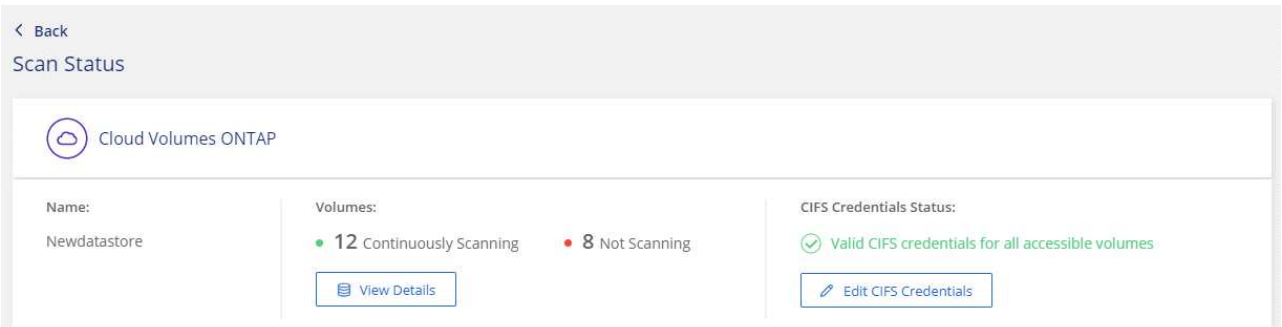
- For CIFS – ports 139 and 445.
- 4. Ensure that NFS volume export policies include the IP address of the Data Sense instance so it can access the data on each volume.
- 5. If you use CIFS, provide Data Sense with Active Directory credentials so it can scan CIFS volumes.
 - a. At the top of Cloud Manager, click **Data Sense**.
 - b. Click the **Configuration** tab.



- c. For each working environment, click **Edit CIFS Credentials** and enter the user name and password that Data Sense needs to access CIFS volumes on the system.

The credentials can be read-only, but providing admin credentials ensures that Data Sense can read any data that requires elevated permissions. The credentials are stored on the Cloud Data Sense instance.

After you enter the credentials, you should see a message that all CIFS volumes were authenticated successfully.



6. On the *Configuration* page, click **View Details** to review the status for each CIFS and NFS volume and correct any errors.

For example, the following image shows four volumes; one of which Cloud Data Sense can't scan due to network connectivity issues between the Data Sense instance and the volume.

cognitoWE Scan Configuration
44/79 Volumes selected for Data Sense scan

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

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

Enabling and disabling compliance scans on volumes

You can start or stop mapping-only scans, or mapping and classification scans, in a working environment at any time from the Configuration page. You can also change from mapping-only scans to mapping and classification scans, and vice-versa. We recommend that you scan all volumes.

cognitoWE Scan Configuration
44/79 Volumes selected for Data Sense scan

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

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

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



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

Scanning data protection volumes

By default, data protection (DP) volumes are not scanned because they are not exposed externally and Cloud Data Sense cannot access them. These are the destination volumes for SnapMirror operations from an on-premises ONTAP system or from a Cloud Volumes ONTAP system.

Initially, the volume list identifies these volumes as *Type DP* with the *Status Not Scanning* and the *Required Action Enable Access to DP volumes*.

The screenshot shows the 'Working Environment Name' Configuration page. At the top, it says '22/28 Volumes selected for compliance scan'. There are buttons for 'Off', 'Map', 'Map & Classify', and 'Custom'. A red box highlights the 'Enable Access to DP Volumes' button. Below the buttons is a table with columns: Scan, Storage Repository (Volume), Type, Status, and Required Action.

Scan	Storage Repository (Volume)	Type	Status	Required Action
<input type="button" value="Off"/> <input type="button" value="Map"/> <input type="button" value="Map & Classify"/>	VolumeName1	DP	Not Scanning	Enable access to DP Volumes ⓘ
<input type="button" value="Off"/> <input type="button" value="Map"/> <input type="button" value="Map & Classify"/>	VolumeName2	NFS	Continuously Scanning	
<input type="button" value="Off"/> <input type="button" value="Map"/> <input type="button" value="Map & Classify"/>	VolumeName3	CIFS	Not Scanning	

Steps

If you want to scan these data protection volumes:

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

The screenshot shows the 'Provide Active Directory Credentials' dialog box. The 'Use existing CIFS Scanning Credentials (user1@domain2)' radio button is selected and highlighted with a red box. Below it are fields for 'Active Directory Domain' and 'DNS IP Address'. A text block explains that DP Volumes created from a SnapMirror relationship do not allow external access by default and that continuing will create NFS shares. At the bottom are 'Enable Access to DP Volumes' and 'Cancel' buttons.

The screenshot shows the 'Provide Active Directory Credentials' dialog box. The 'Use Custom Credentials' radio button is selected and highlighted with a red box. Below it are fields for 'Username' and 'Password'. Below those are fields for 'Active Directory Domain' and 'DNS IP Address'. A text block explains that DP Volumes created from a SnapMirror relationship do not allow external access by default and that continuing will create NFS shares. At the bottom are 'Enable Access to DP Volumes' and 'Cancel' buttons.

3. Activate each DP volume that you want to scan [the same way you enabled other volumes](#).

Result

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

Note: If you had no CIFS data protection volumes when you initially enabled access to DP volumes, and later

add some, the button **Enable Access to CIFS DP** appears at the top of the Configuration page. Click this button and add CIFS credentials to enable access to these CIFS DP volumes.



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

Scanning backup files from on-premises ONTAP systems

If you don't want Cloud Data Sense to scan volumes directly on your on-prem ONTAP systems, a Beta feature released in January 2021 allows you to run compliance scans on backup files created from your on-prem ONTAP volumes. So if you're already creating backup files to the cloud using [Cloud Backup](#), you can use this feature to run compliance scans on those backup files.

When scanning backup files created from your on-prem ONTAP systems, the scanned data does not count against the capacity purchased in your PAYGO subscription or BYOL license.

Note: When Data Sense scans backup files it uses permissions granted through the Cloud Restore instance to access the backup files. Typically the Restore instance powers down when not actively restoring files, but it remains **On** when scanning backup files. See [more information about the Restore instance](#).

Steps

If you want to scan the backup files from on-prem ONTAP systems:

1. At the top of Cloud Manager, click **Data Sense** and then select the **Configuration** tab.
2. From the list of working environments, click the **BACKUP** button from the list of filters.

All the on-premises ONTAP working environments that have backup files are listed. If you don't have any backup files from an on-prem system, then the working environment is not shown.

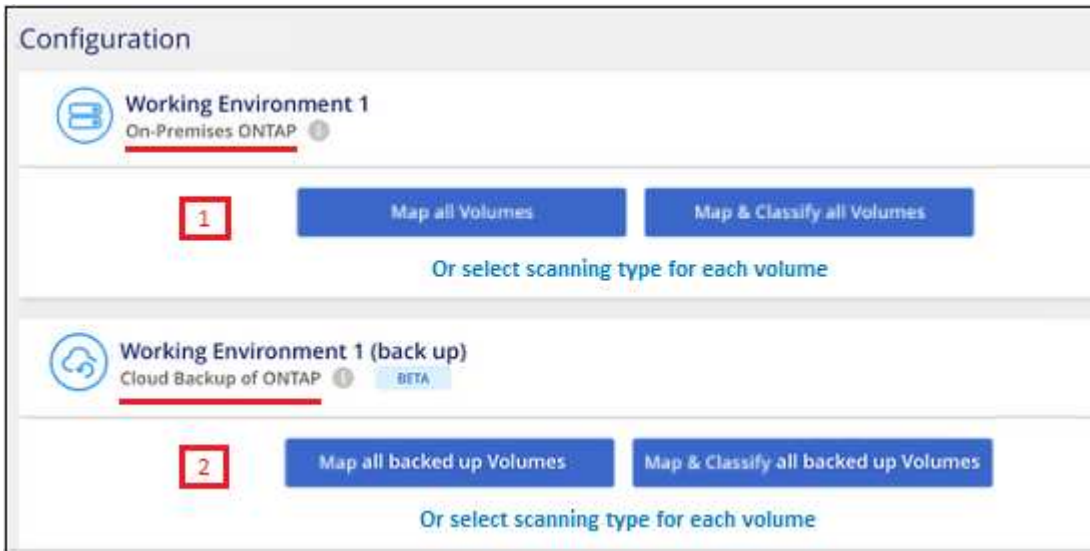


3. Select how you want to scan the backup volumes:
 - To map all volumes, click **Map all backed up Volumes**.
 - To map and classify all volumes, click **Map & Classify all backed up Volumes**.
 - To customize scanning for each backed up volume, click **Or select scanning type for each volume**, and then choose the volumes you want to map and/or classify.

See [Enabling and disabling compliance scans on volumes](#) for details.

Scanning on-prem volumes versus backups of those volumes

When you view the entire list of working environments you will see two listings for each on-prem cluster if they have backed up files.



The first item is the on-prem cluster and the actual volumes.

The second item is the backup files of those volumes from that same on-prem cluster.

Choose the first option to scan the volumes on the on-prem system. Choose the second option to scan the backup files from those volumes. Do not scan both on-prem volumes and backup files of the same cluster.

Getting started with Cloud Data Sense for Azure NetApp Files

Complete a few steps to get started with Cloud Data Sense for Azure NetApp Files.

Quick start

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

1

Discover the Azure NetApp Files systems you want to scan

Before you can scan Azure NetApp Files volumes, [Cloud Manager must be set up to discover the configuration](#).

2

Deploy the Cloud Data Sense instance

[Deploy Cloud Data Sense in Cloud Manager](#) if there isn't already an instance deployed.

3

Enable Cloud Data Sense and select the volumes to scan

Click **Compliance**, select the **Configuration** tab, and activate compliance scans for volumes in specific working environments.

4

Ensure access to volumes

Now that Cloud Data Sense is enabled, ensure that it can access all volumes.

- The Cloud Data Sense instance needs a network connection to each Azure NetApp Files subnet.
- Make sure these ports are open to the Data Sense instance:
 - For NFS – ports 111 and 2049.
 - For CIFS – ports 139 and 445.
- NFS volume export policies must allow access from the Data Sense instance.
- Data Sense needs Active Directory credentials to scan CIFS volumes.

Click **Compliance > Configuration > Edit CIFS Credentials** and provide the credentials.

5

Manage the volumes you want to scan

Select or deselect the volumes that you want to scan and Cloud Data Sense will start or stop scanning them.

Discovering the Azure NetApp Files system that you want to scan

If the Azure NetApp Files system you want to scan is not already in Cloud Manager as a working environment, you can add it to the canvas at this time.

[See how to discover the Azure NetApp Files system in Cloud Manager.](#)

Deploying the Cloud Data Sense instance

[Deploy Cloud Data Sense](#) if there isn't already an instance deployed.

Data Sense must be deployed in the cloud when scanning Azure NetApp Files volumes, and it must be deployed in the same region as the volumes you wish to scan.

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

Upgrades to Data Sense software is automated as long as the instance has internet connectivity.

Enabling Cloud Data Sense in your working environments

You can enable Cloud Data Sense on your Azure NetApp Files volumes.

1. At the top of Cloud Manager, click **Data Sense** and then select the **Configuration** tab.



2. Select how you want to scan the volumes in each working environment. [Learn about mapping and classification scans](#):

- To map all volumes, click **Map all Volumes**.
- To map and classify all volumes, click **Map & Classify all Volumes**.
- To customize scanning for each volume, click **Or select scanning type for each volume**, and then choose the volumes you want to map and/or classify.

See [Enabling and disabling compliance scans on volumes](#) for details.

3. In the confirmation dialog box, click **Approve** to have Data Sense start scanning your volumes.

Result

Cloud Data Sense starts scanning the volumes you selected in the working environment. Results will be available in the Compliance dashboard as soon as Cloud Data Sense finishes the initial scans. The time that it takes depends on the amount of data—it could be a few minutes or hours.

Verifying that Cloud Data Sense has access to volumes

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

Steps

1. Make sure that there's a network connection between the Cloud Data Sense instance and each network that includes volumes for Azure NetApp Files.



For Azure NetApp Files, Cloud Data Sense can only scan volumes that are in the same region as Cloud Manager.

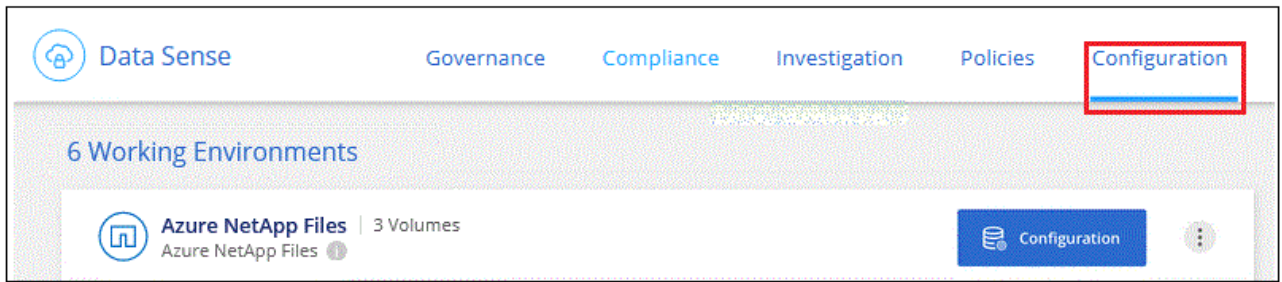
2. Ensure the following ports are open to the Data Sense instance:

- For NFS – ports 111 and 2049.
- For CIFS – ports 139 and 445.

3. Ensure that NFS volume export policies include the IP address of the Data Sense instance so it can access the data on each volume.

4. If you use CIFS, provide Data Sense with Active Directory credentials so it can scan CIFS volumes.

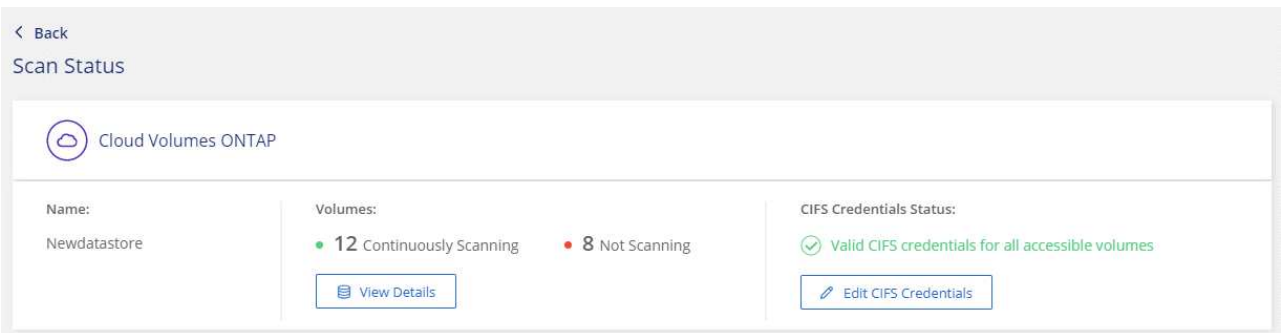
- a. At the top of Cloud Manager, click **Data Sense**.
- b. Click the **Configuration** tab.



- c. For each working environment, click **Edit CIFS Credentials** and enter the user name and password that Data Sense needs to access CIFS volumes on the system.

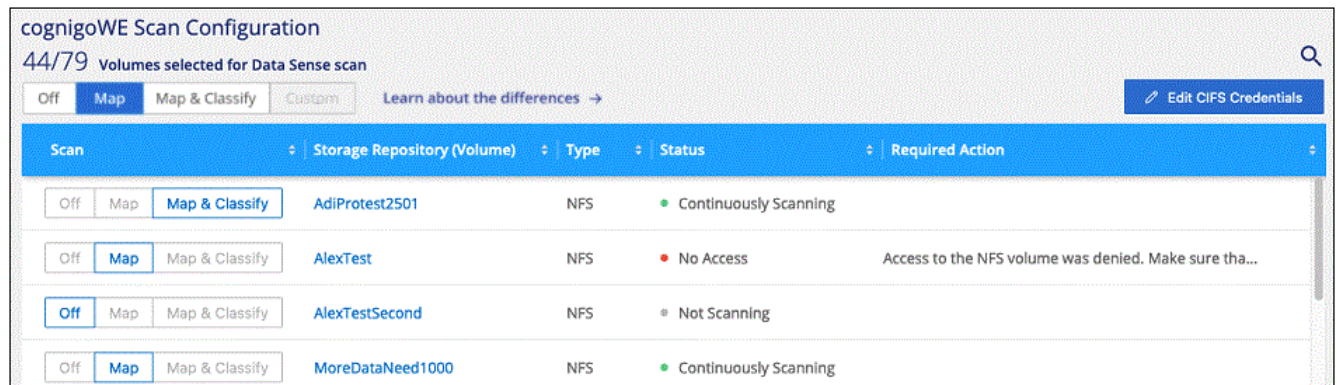
The credentials can be read-only, but providing admin credentials ensures that Data Sense can read any data that requires elevated permissions. The credentials are stored on the Cloud Data Sense instance.

After you enter the credentials, you should see a message that all CIFS volumes were authenticated successfully.



5. On the *Configuration* page, click **View Details** to review the status for each CIFS and NFS volume and correct any errors.

For example, the following image shows four volumes; one of which Cloud Data Sense can't scan due to network connectivity issues between the Data Sense instance and the volume.



Enabling and disabling compliance scans on volumes

You can start or stop mapping-only scans, or mapping and classification scans, in a working environment at any time from the Configuration page. You can also change from mapping-only scans to mapping and classification scans, and vice-versa. We recommend that you scan all volumes.

cognigoWE Scan Configuration
44/79 Volumes selected for Data Sense scan

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

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

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



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

Get started with Cloud Data Sense for Amazon FSx for ONTAP

Complete a few steps to get started scanning Amazon FSx for ONTAP volume with Cloud Data Sense.

Before you begin

- You need an active Connector in AWS to deploy and manage Data Sense.
- The security group you selected when creating the working environment must allow traffic from the Cloud Data Sense instance. You can find the associated security group using the ENI connected to the FSx for ONTAP file system and edit it using the AWS Management Console.

[AWS security groups for Linux instances](#)

[AWS security groups for Windows instances](#)

[AWS elastic network interfaces \(ENI\)](#)

Quick start

Get started quickly by following these steps or scroll down for full details.

1

Discover the FSx for ONTAP file systems you want to scan

Before you can scan FSx for ONTAP volumes, [you must have an FSx working environment with volumes configured](#).

2

Deploy the Cloud Data Sense instance

[Deploy Cloud Data Sense in Cloud Manager](#) if there isn't already an instance deployed.

3

Enable Cloud Data Sense and select the volumes to scan

Click **Data Sense**, select the **Configuration** tab, and activate compliance scans for volumes in specific working environments.

4

Ensure access to volumes

Now that Cloud Data Sense is enabled, ensure that it can access all volumes.

- The Cloud Data Sense instance needs a network connection to each FSx for ONTAP subnet.
- Make sure the following ports are open to the Data Sense instance:
 - For NFS – ports 111 and 2049.
 - For CIFS – ports 139 and 445.
- NFS volume export policies must allow access from the Data Sense instance.
- Data Sense needs Active Directory credentials to scan CIFS volumes.

Click **Compliance > Configuration > Edit CIFS Credentials** and provide the credentials.

5

Manage the volumes you want to scan

Select or deselect the volumes you want to scan and Cloud Data Sense will start or stop scanning them.

Discovering the FSx for ONTAP file system that you want to scan

If the FSx for ONTAP file system you want to scan is not already in Cloud Manager as a working environment, you can add it to the canvas at this time.

[See how to discover or create the FSx for ONTAP file system in Cloud Manager](#).

Deploying the Cloud Data Sense instance

[Deploy Cloud Data Sense](#) if there isn't already an instance deployed.

You should deploy Data Sense in the same AWS network as the Connector for AWS and the FSx volumes you wish to scan.

Note: Deploying Cloud Data Sense in an on-premises location is not currently supported when scanning FSx volumes.

Upgrades to Data Sense software is automated as long as the instance has internet connectivity.

Enabling Cloud Data Sense in your working environments

You can enable Cloud Data Sense for FSx for ONTAP volumes.

1. At the top of Cloud Manager, click **Data Sense** and then select the **Configuration** tab.



2. Select how you want to scan the volumes in each working environment. [Learn about mapping and classification scans](#):
 - To map all volumes, click **Map all Volumes**.
 - To map and classify all volumes, click **Map & Classify all Volumes**.
 - To customize scanning for each volume, click **Or select scanning type for each volume**, and then choose the volumes you want to map and/or classify.

See [Enabling and disabling compliance scans on volumes](#) for details.

3. In the confirmation dialog box, click **Approve** to have Data Sense start scanning your volumes.

Result

Cloud Data Sense starts scanning the volumes you selected in the working environment. Results will be available in the Compliance dashboard as soon as Cloud Data Sense finishes the initial scans. The time that it takes depends on the amount of data—it could be a few minutes or hours.

Verifying that Cloud Data Sense has access to volumes

Make sure Cloud Data Sense can access volumes by checking your networking, security groups, and export policies.

You'll need to provide Data Sense with CIFS credentials so it can access CIFS volumes.

Steps

1. On the *Configuration* page, click **View Details** to review the status and correct any errors.

For example, the following image shows a volume Cloud Data Sense can't scan due to network connectivity issues between the Data Sense instance and the volume.

Scan	Storage Repository (Volume)	Type	Status	Required Action
Off Map Map & Classify	jrmclone	NFS	No Access	Check network connectivity between the Data Sense ...

- Make sure there's a network connection between the Cloud Data Sense instance and each network that includes volumes for FSx for ONTAP.



For FSx for ONTAP, Cloud Data Sense can scan volumes only in the same region as Cloud Manager.

- Ensure the following ports are open to the Data Sense instance.
 - For NFS – ports 111 and 2049.
 - For CIFS – ports 139 and 445.
- Ensure NFS volume export policies include the IP address of the Data Sense instance so it can access the data on each volume.
- If you use CIFS, provide Data Sense with Active Directory credentials so it can scan CIFS volumes.
 - At the top of Cloud Manager, click **Data Sense**.
 - Click the **Configuration** tab.
 - For each working environment, click **Edit CIFS Credentials** and enter the user name and password that Data Sense needs to access CIFS volumes on the system.

The credentials can be read-only, but providing admin credentials ensures that Data Sense can read any data that requires elevated permissions. The credentials are stored on the Cloud Data Sense instance.

After you enter the credentials, you should see a message that all CIFS volumes were authenticated successfully.

Enabling and disabling compliance scans on volumes

You can start or stop mapping-only scans, or mapping and classification scans, in a working environment at any time from the Configuration page. You can also change from mapping-only scans to mapping and classification scans, and vice-versa. We recommend that you scan all volumes.

cognitoWE Scan Configuration				
44/79 Volumes selected for Data Sense scan				
<div> Off Map Map & Classify Custom </div> <div> Learn about the differences → </div> <div> Edit CIFS Credentials </div>				
Scan	Storage Repository (Volume)	Type	Status	Required Action
Off Map Map & Classify	AdiNFSVol_copy	NFS	No Access	Access to the NFS volume was denied. Make sure tha...
Off Map Map & Classify	AdiProtest2501	NFS	Continuously Scanning	
Off Map Map & Classify	AlexTest	NFS	No Access	Access to the NFS volume was denied. Make sure tha...
Off Map Map & Classify	AlexTestSecond	NFS	Not Scanning	
Off Map Map & Classify	MoreDataNeed1000	NFS	Continuously Scanning	

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



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

Scanning data protection volumes

By default, data protection (DP) volumes are not scanned because they are not exposed externally and Cloud Data Sense cannot access them. These are the destination volumes for SnapMirror operations from an FSx for ONTAP file system.

Initially, the volume list identifies these volumes as *Type DP* with the *Status Not Scanning* and the *Required Action Enable Access to DP volumes*.

The screenshot shows the 'Working Environment Name' Configuration page. At the top, there's a search icon and a button 'Enable Access to DP Volumes' which is highlighted with a red box. Below it is a button 'Edit CIFS Credentials'. The page has tabs for 'Off', 'Map', 'Map & Classify', and 'Custom', with 'Map' currently selected. A link 'Learn about the differences' is also present. Below the tabs is a table with columns: 'Scan', 'Storage Repository (Volume)', 'Type', 'Status', and 'Required Action'. The table lists three volumes: 'VolumeName1' (Type: DP, Status: Not Scanning, Required Action: Enable access to DP Volumes), 'VolumeName2' (Type: NFS, Status: Continuously Scanning), and 'VolumeName3' (Type: CIFS, Status: Not Scanning). Each volume row has its own set of 'Off', 'Map', and 'Map & Classify' buttons.

Steps

If you want to scan these data protection volumes:

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

3. Activate each DP volume that you want to scan [the same way you enabled other volumes](#).

Result

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

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



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

Getting started with Cloud Data Sense for Amazon S3

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

Quick start

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

1

Set up the S3 requirements in your cloud environment

Ensure that your cloud environment can meet the requirements for Cloud Data Sense, including preparing an IAM role and setting up connectivity from Data Sense to S3. [See the complete list](#).

2

Deploy the Cloud Data Sense instance

[Deploy Cloud Data Sense](#) if there isn't already an instance deployed.

3

Activate Data Sense on your S3 working environment

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

permissions.

4

Select the buckets to scan

Select the buckets that you'd like to scan and Cloud Data Sense will start scanning them.

Reviewing S3 prerequisites

The following requirements are specific to scanning S3 buckets.

Set up an IAM role for the Cloud Data Sense instance

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

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

Provide connectivity from Cloud Data Sense to Amazon S3

Cloud Data Sense needs a connection to Amazon S3. The best way to provide that connection is through a VPC Endpoint to the S3 service. For instructions, see [AWS Documentation: Creating a Gateway Endpoint](#).

When you create the VPC Endpoint, be sure to select the region, VPC, and route table that corresponds to

the Cloud Data Sense instance. You must also modify the security group to add an outbound HTTPS rule that enables traffic to the S3 endpoint. Otherwise, Data Sense can't connect to the S3 service.

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

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



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

Deploying the Cloud Data Sense instance

[Deploy Cloud Data Sense in Cloud Manager](#) if there isn't already an instance deployed.

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

Note: Deploying Cloud Data Sense in an on-premises location is not currently supported when scanning S3 buckets.

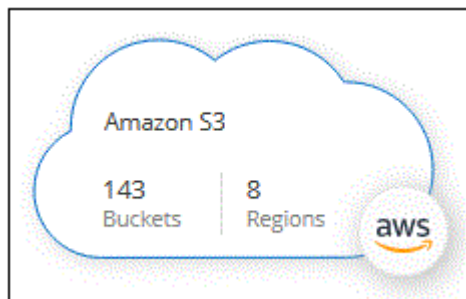
Upgrades to Data Sense software is automated as long as the instance has internet connectivity.

Activating Data Sense on your S3 working environment

Enable Cloud Data Sense on Amazon S3 after you verify the prerequisites.

Steps

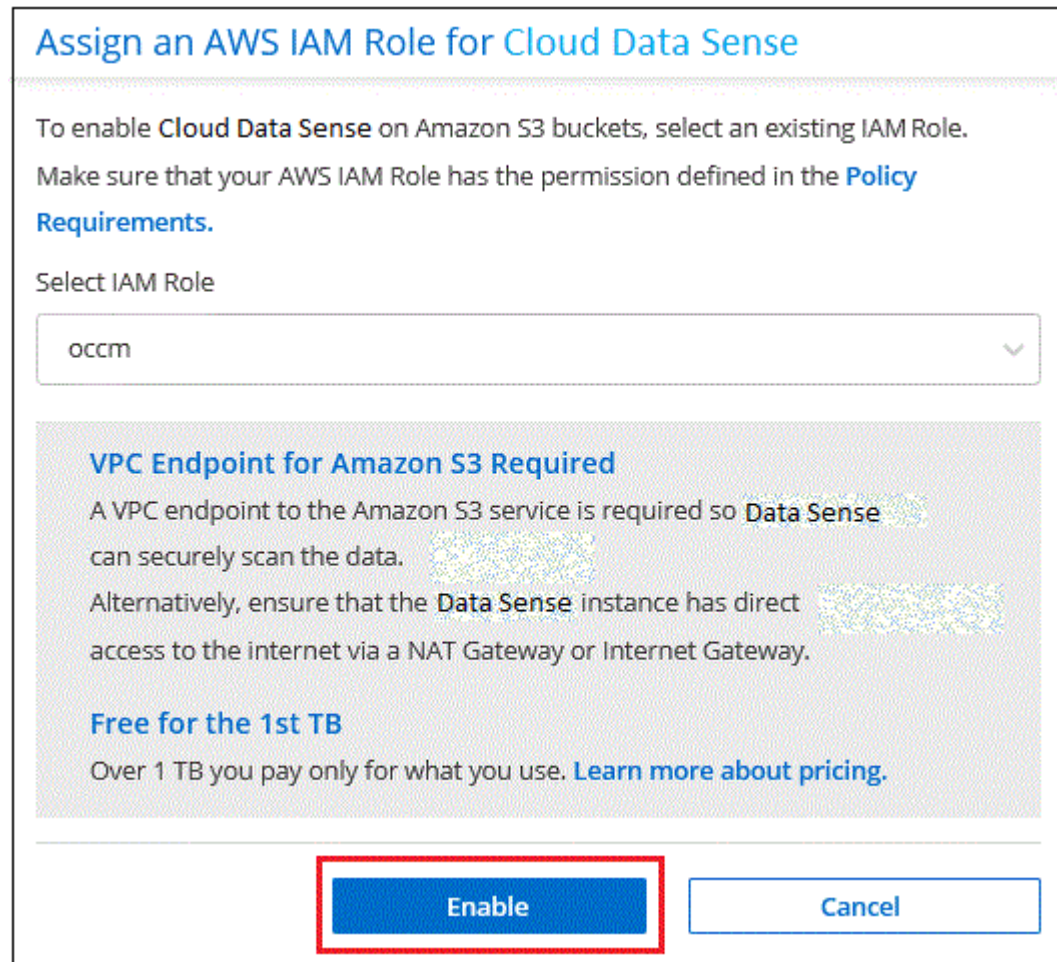
1. At the top of Cloud Manager, click **Canvas**.
2. Select the Amazon S3 working environment.



3. In the Data Sense pane on the right, click **Enable**.



4. When prompted, assign an IAM role to the Cloud Data Sense instance that has [the required permissions](#).



5. Click **Enable**.



You can also enable compliance scans for a working environment from the Configuration page by clicking the  button and selecting **Activate Data Sense**.

Result

Cloud Manager assigns the IAM role to the instance.

Enabling and disabling compliance scans on S3 buckets

After Cloud Manager enables Cloud Data Sense on Amazon S3, the next step is to configure the buckets that you want to scan.

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

Cloud Data Sense can also [scan S3 buckets that are in different AWS accounts](#).

Steps

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



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

Amazon S3 Configuration			
15/28 Buckets in Scan Scope.			
Scan	Bucket Name	Status	Required Action
<div>OffMapMap & Classify</div>	BucketName1	Not Scanning	Add Credentials
<div>OffMapMap & Classify</div>	BucketName2	Continuously Scanning	
<div>OffMapMap & Classify</div>	BucketName3	Not Scanning	

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

To:	Do this:
Disable scanning on a bucket	Click Off

Result

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

Scanning buckets from additional AWS accounts

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

Steps

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

Create role



Select type of trusted entity

AWS service
EC2, Lambda and others

Another AWS account
Belonging to you or 3rd party

Web identity
Cognito or any OpenID provider

SAML 2.0 federation
Your corporate directory

Allows entities in other accounts to perform actions in this account. [Learn more](#)

Specify accounts that can use this role

Account ID*

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

Be sure to do the following:

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

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

2. Go to the source AWS account where the Data Sense instance resides and select the IAM role that is attached to the instance.
 - a. Change the **Maximum CLI/API session duration** from 1 hour to 12 hours and save that change.
 - b. Click **Attach policies** and then click **Create policy**.
 - c. Create a policy that includes the "sts:AssumeRole" action and specify the ARN of the role that you created in the target account.

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

The Cloud Data Sense instance profile account now has access to the additional AWS account.

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



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

Result

Cloud Data Sense starts scanning the new S3 buckets that you enabled.

Scanning database schemas

Complete a few steps to start scanning your database schemas with Cloud Data Sense.

Quick start

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

1

Review database prerequisites

Ensure that your database is supported and that you have the information necessary to connect to the database.

2

Deploy the Cloud Data Sense instance

[Deploy Cloud Data Sense](#) if there isn't already an instance deployed.

3

Add the database server

Add the database server that you want to access.

4

Select the schemas

Select the schemas that you want to scan.

Reviewing prerequisites

Review the following prerequisites to make sure that you have a supported configuration before you enable Cloud Data Sense.

Supported databases

Cloud Data Sense can scan schemas from the following databases:

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



The statistics gathering feature **must be enabled** in the database.

Database requirements

Any database with connectivity to the Cloud Data Sense instance can be scanned, regardless of where it is hosted. You just need the following information to connect to the database:

- IP Address or host name
- Port

- Service name (only for accessing Oracle databases)
- Credentials that allow read access to the schemas

When choosing a user name and password, it's important to choose one that has full read permissions to all the schemas and tables you want to scan. We recommend that you create a dedicated user for the Cloud Data Sense system with all the required permissions.

Note: For MongoDB, a read-only Admin role is required.

Deploying the Cloud Data Sense instance

Deploy Cloud Data Sense if there isn't already an instance deployed.

If you are scanning database schemas that are accessible over the internet, you can [deploy Cloud Data Sense in the cloud](#) or [deploy Data Sense in an on-premises location that has internet access](#).

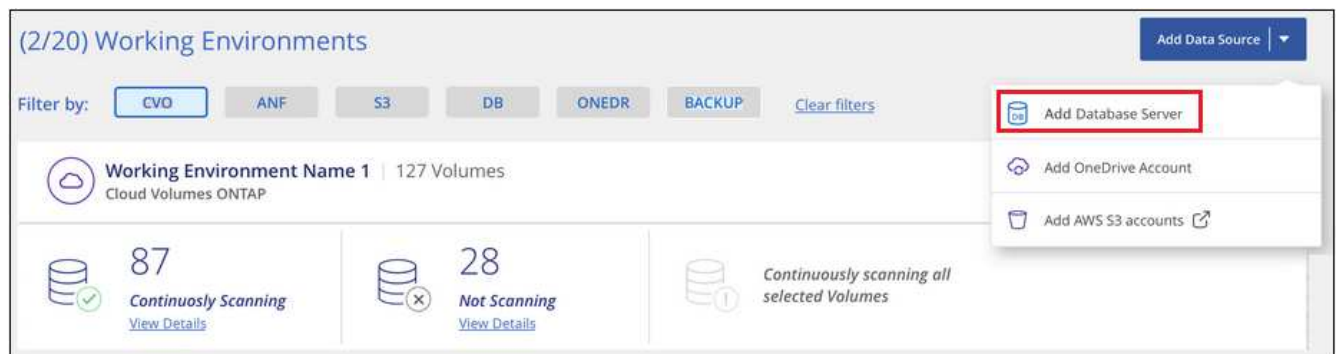
If you are scanning database schemas that have been installed in a dark site that has no internet access, you need to [deploy Cloud Data Sense in the same on-premises location that has no internet access](#). This also requires that the Cloud Manager Connector is deployed in that same on-premises location.

Upgrades to Data Sense software is automated as long as the instance has internet connectivity.

Adding the database server

Add the database server where the schemas reside.

1. From the Working Environments Configuration page, click **Add Data Source > Add Database Server**.



2. Enter the required information to identify the database server.
 - a. Select the database type.
 - b. Enter the port and the host name or IP address to connect to the database.
 - c. For Oracle databases, enter the Service name.
 - d. Enter the credentials so that Cloud Data Sense can access the server.
 - e. Click **Add DB Server**.

Add DB Server

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

Database

Database Type

Host Name or IP Address

Port

Service Name

Credentials

Username

Password

Add DB Server **Cancel**

The database is added to the list of working environments.

Enabling and disabling compliance scans on database schemas

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



There is no option to select mapping-only scans for database schemas.

1. From the *Configuration* page, click the **Configuration** button for the database you want to configure.

Configuration

Oracle DB 1 | 41 Schemas
Oracle

Configuration

No Schemas selected for Compliance

7
Not Scanning
[View Details](#)

2. Select the schemas that you want to scan by moving the slider to the right.

'Working Environment Name' Configuration			
28/28 Schemas selected for compliance scan		<input type="text"/> Edit Credentials	
Scan	Schema Name	Status	Required Action
<input type="checkbox"/>	DB1 - SchemaName1	Not Scanning	Add Credentials ⓘ
<input checked="" type="checkbox"/>	DB1 - SchemaName2	Continuously Scanning	
<input checked="" type="checkbox"/>	DB1 - SchemaName3	Continuously Scanning	
<input checked="" type="checkbox"/>	DB1 - SchemaName4	Continuously Scanning	

Result

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

Scanning OneDrive accounts

Complete a few steps to start scanning files in your user's OneDrive folders with Cloud Data Sense.

Quick start

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

1

Review OneDrive prerequisites

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

2

Deploy the Cloud Data Sense instance

[Deploy Cloud Data Sense](#) if there isn't already an instance deployed.

3

Add the OneDrive account

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

4

Add the users and select the users to scan

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

Reviewing OneDrive requirements

Review the following prerequisites to make sure that you have a supported configuration before you enable Cloud Data Sense.

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

Deploying the Cloud Data Sense instance

Deploy Cloud Data Sense if there isn't already an instance deployed.

Data Sense can be [deployed in the cloud](#) or [in an on-premises location that has internet access](#).

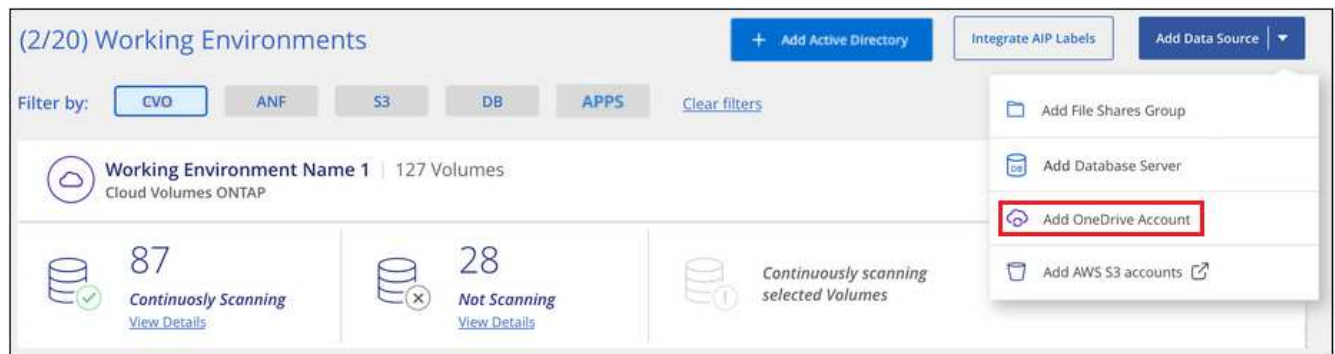
Upgrades to Data Sense software is automated as long as the instance has internet connectivity.

Adding the OneDrive account

Add the OneDrive account where the user files reside.

Steps

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



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

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

Adding OneDrive users to compliance scans

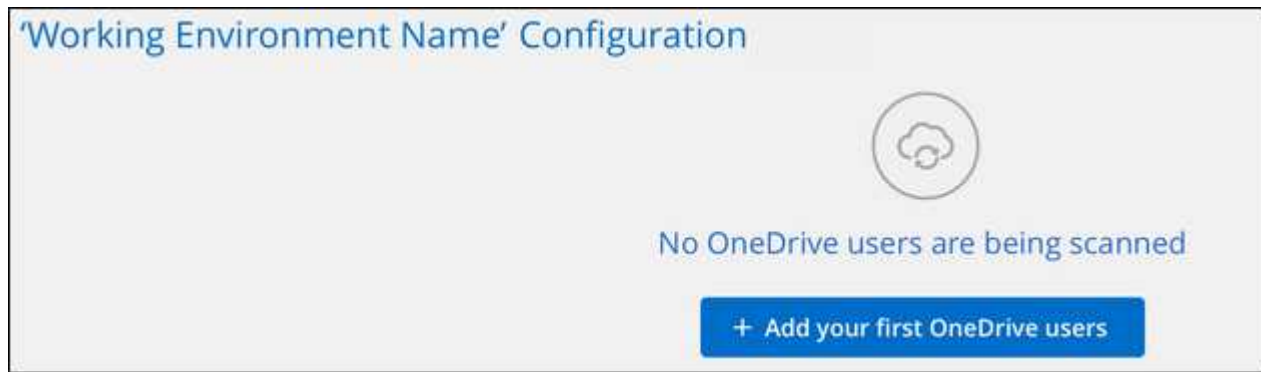
You can add individual OneDrive users, or all of your OneDrive users, so that their files will be scanned by Cloud Data Sense.

Steps

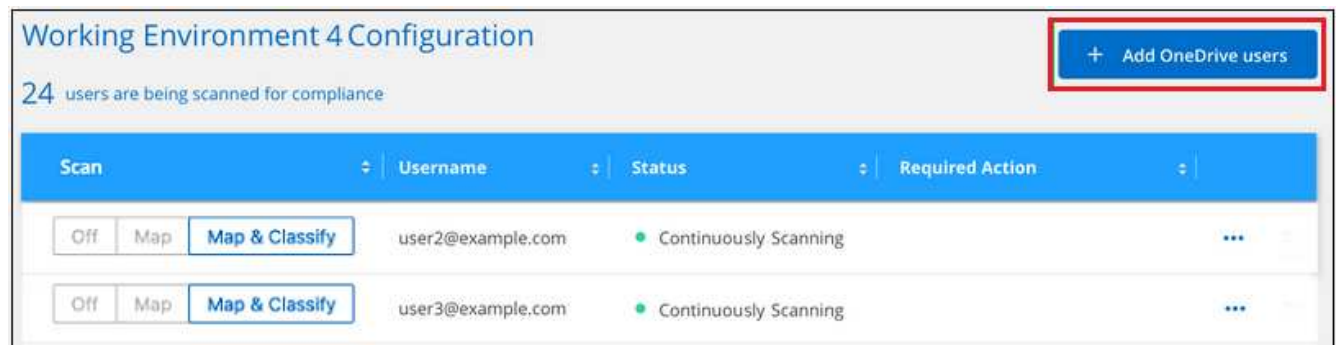
1. From the *Configuration* page, click the **Configuration** button for the OneDrive account.



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



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



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

The screenshot shows a dialog box titled "Add OneDrive users". It contains the text: "Provide a list of OneDrive users for Cloud Data Sense to scan their data, line-separated. You can add up to 100 users at a time." Below this, there is a section titled "Type or paste below the OneDrive user accounts to add". Underneath, there is a text area labeled "User Accounts" containing several lines of "user@example.com". At the bottom, there are two buttons: "Add Users" and "Cancel".

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

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

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

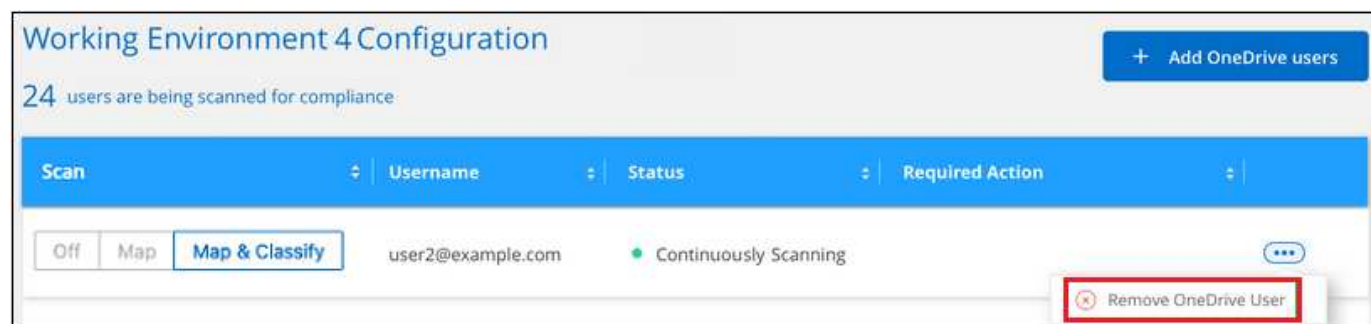
To:	Do this:
Enable mapping-only scans on user files	Click Map
Enable full scans on user files	Click Map & Classify
Disable scanning on user files	Click Off

Result

Cloud Data Sense starts scanning the files for the users you added, and the results are displayed in the Dashboard and in other locations.

Removing a OneDrive user from compliance scans

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



Scanning SharePoint accounts

Complete a few steps to start scanning files in your SharePoint accounts with Cloud Data Sense.

Quick start

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

1

Review SharePoint prerequisites

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

2

Deploy the Cloud Data Sense instance

[Deploy Cloud Data Sense](#) if there isn't already an instance deployed.

3

Log into the SharePoint account

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

4

Add the SharePoint site URLs to scan

Add the list of SharePoint site URLs that you want to scan in the SharePoint account, and select the type of scanning. You can add up to 100 URLs at time.

Reviewing SharePoint requirements

Review the following prerequisites to make sure you are ready to enable Cloud Data Sense on a SharePoint account.

- You must have the Admin login credentials for the SharePoint account that provides read access to all SharePoint sites.
- You will need a line-separated list of the SharePoint site URLs for all the data you want to scan.

Deploying the Cloud Data Sense instance

Deploy Cloud Data Sense if there isn't already an instance deployed.

Data Sense can be [deployed in the cloud](#) or [in an on-premises location that has internet access](#).

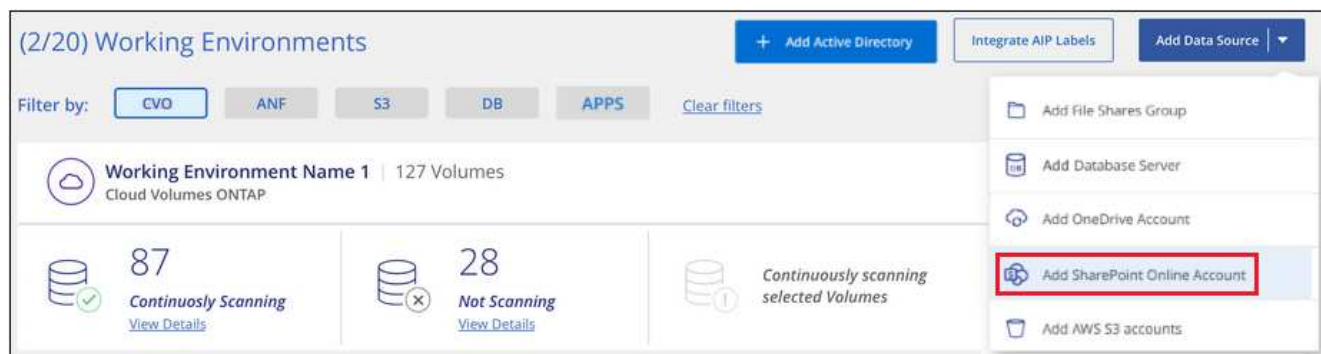
Upgrades to Data Sense software is automated as long as the instance has internet connectivity.

Adding the SharePoint account

Add the SharePoint account where the user files reside.

Steps

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



2. In the Add a SharePoint Online Account dialog, click **Sign in to SharePoint**.
3. In the Microsoft page that appears, select the SharePoint account and enter the required Admin user and password, then click **Accept** to allow Cloud Data Sense to read data from this account.

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

Adding SharePoint sites to compliance scans

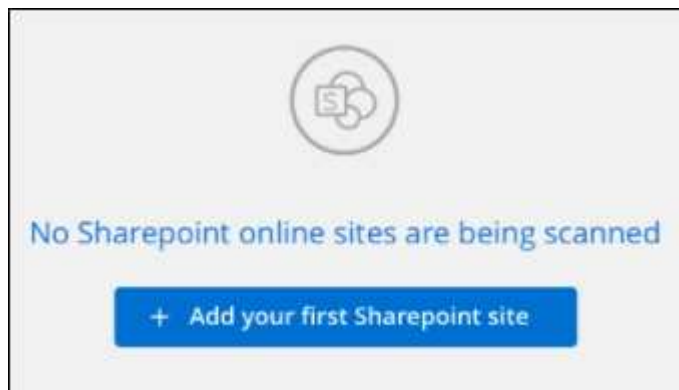
You can add individual SharePoint sites, or all of the SharePoint sites in the account, so that the associated files will be scanned by Cloud Data Sense.

Steps

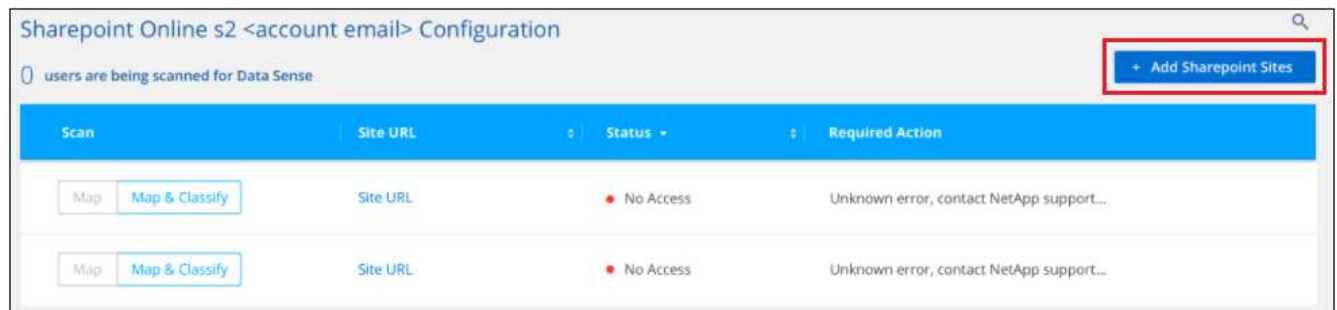
1. From the *Configuration* page, click the **Configuration** button for the SharePoint account.



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



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



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

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

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

4. Enable mapping-only scans, or mapping and classification scans, on the files in the SharePoint sites.

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

Result

Cloud Data Sense starts scanning the files in the SharePoint sites you added, and the results are displayed in the Dashboard and in other locations.

Removing a SharePoint site from compliance scans

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

Scan	Site URL	Status	Required Action
Off Map Map & Classify	Site URL	Continuously Scanning	...
Off Map Map & Classify	Site URL	Continuously Scanning	Remove SharePoint Site

Scanning file shares

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

Quick start

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

1

Review file share prerequisites

For CIFS (SMB) shares, ensure that you have credentials to access the shares.

2

Deploy the Cloud Data Sense instance

[Deploy Cloud Data Sense](#) if there isn't already an instance deployed.

3

Create a group to hold the file shares

The group is a container for the file shares that you want to scan, and it is used as the working environment name for those file shares.

4

Add the file shares and select the shares to scan

Add the list of file shares that you want to scan and select the type of scanning. You can add up to 100 file shares at a time.

Reviewing file share requirements

Review the following prerequisites to make sure that you have a supported configuration before you enable Cloud Data Sense.

- The shares can be hosted anywhere, including in the cloud or on-premises. These are file shares that reside on non-NetApp storage systems.
- There needs to be network connectivity between the Data Sense instance and the shares.
- Make sure these ports are open to the Data Sense instance:
 - For NFS – ports 111 and 2049.
 - For CIFS – ports 139 and 445.
- You will need the list of shares you want to add in the format `<host_name>:/<share_path>`. You can

enter the shares individually, or you can supply a line-separated list of the file shares you want to scan.

- For CIFS (SMB) shares, ensure that you have Active Directory credentials that provide read access to the shares. Admin credentials are preferred in case Cloud Data Sense needs to scan any data that requires elevated permissions.

Deploying the Cloud Data Sense instance

Deploy Cloud Data Sense if there isn't already an instance deployed.

Data Sense can be [deployed in the cloud](#) or [in an on-premises location that has internet access](#).

Upgrades to Data Sense software is automated as long as the instance has internet connectivity.

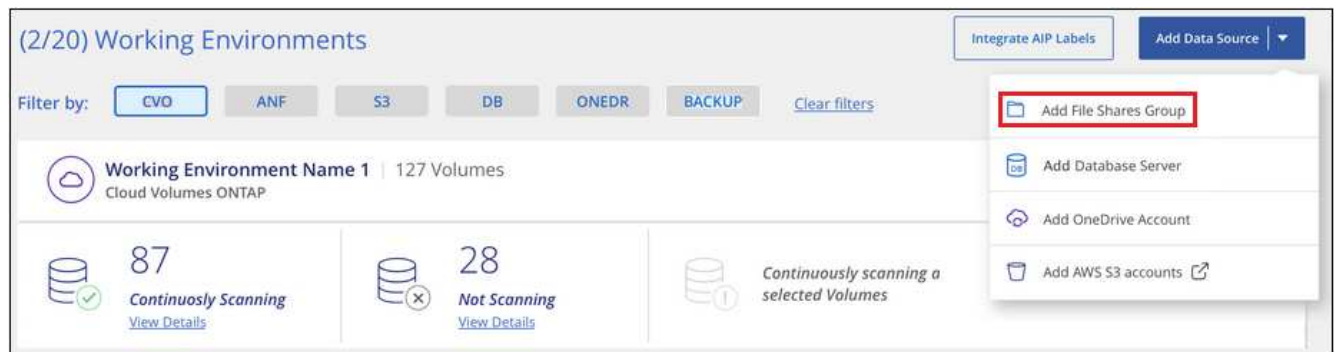
Creating the group for the file shares

You must add a files shares "group" before you can add your file shares. The group is a container for the file shares that you want to scan, and the group name is used as the working environment name for those file shares.

You can mix NFS and CIFS shares in the same group, however, all CIFS file shares in a group need to be using the same Active Directory credentials. If you plan to add CIFS shares that use different credentials, you must make a separate group for each unique set of credentials.

Steps

1. From the Working Environments Configuration page, click **Add Data Source > Add File Shares Group**.



2. In the Add Files Shares Group dialog, enter the name for the group of shares and click **Continue**.

The new File Shares Group is added to the list of working environments.

Adding file shares to a group

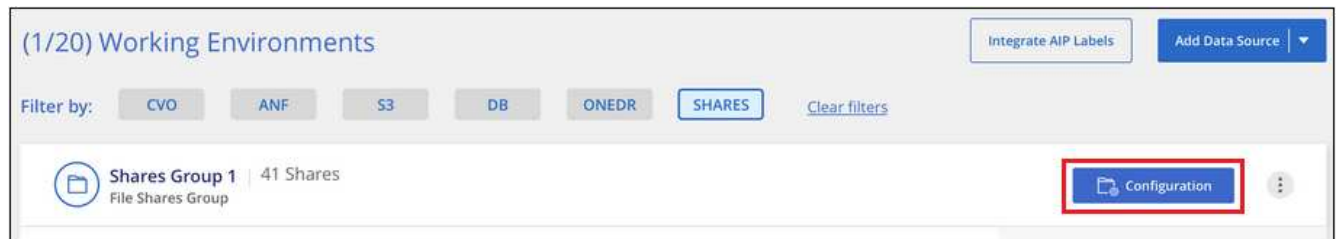
You add file shares to the File Shares Group so that the files in those shares will be scanned by Cloud Data Sense. You add the shares in the format `<host_name>:/<share_path>`.

You can add individual file shares, or you can supply a line-separated list of the file shares you want to scan. You can add up to 100 shares at a time.

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

Steps

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



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



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



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

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

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

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

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

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

Result

Cloud Data Sense starts scanning the files in the file shares you added, and the results are displayed in the Dashboard and in other locations.

Removing a file share from compliance scans

If you no longer need to scan certain file shares, you can remove individual file shares from having their files scanned at any time. Just click **Remove Share** from the Configuration page.



Scanning object storage that uses S3 protocol

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

Quick start

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

1

Review object storage prerequisites

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

You need to have the Access Key and Secret Key from the object storage provider so that Cloud Data Sense can access the buckets.

2

Deploy the Cloud Data Sense instance

[Deploy Cloud Data Sense](#) if there isn't already an instance deployed.

3

Add the Object Storage Service

Add the object storage service to Cloud Data Sense.

4

Select the buckets to scan

Select the buckets that you'd like to scan and Cloud Data Sense will start scanning them.

Reviewing object storage requirements

Review the following prerequisites to make sure that you have a supported configuration before you enable Cloud Data Sense.

- You need to have the endpoint URL to connect with the object storage service.
- You need to have the Access Key and Secret Key from the object storage provider so that Data Sense can

access the buckets.

- Support for Azure Blob requires that you use the [MinIO service](#).

Deploying the Cloud Data Sense instance

Deploy Cloud Data Sense if there isn't already an instance deployed.

Data Sense can be [deployed in the cloud](#) or [in an on-premises location that has internet access](#).

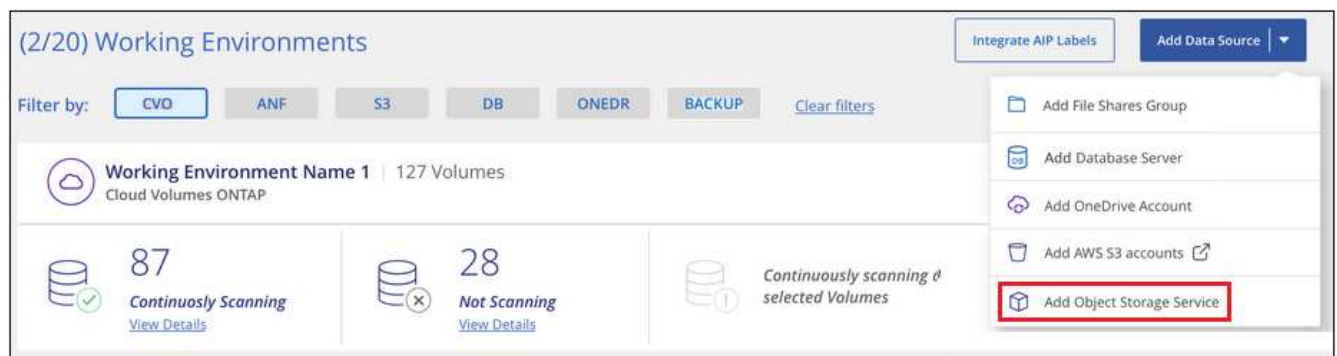
Upgrades to Data Sense software is automated as long as the instance has internet connectivity.

Adding the object storage service to Cloud Data Sense

Add the object storage service.

Steps

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



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

Add Object Storage Service

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

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

Name the Working Environment	Endpoint URL
<input type="text" value="object_myIBM"/>	<input type="text" value="http://my.endpoint.com"/>
Access Key	Secret Key
<input type="text" value="AJUKDO574NDJG86795"/>	<input type="password" value="....."/>

Result

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

Enabling and disabling compliance scans on object storage buckets

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

Steps

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

(1/20) Working Environments

Integrate AIP Labels

Add Data Source

Filter by:

CVO

ANF

S3

DB

ONEDR

BCKUP

OB.STG

Clear filters

Rstor Integrated | 41 Buckets
Object Storage Service

Configuration

23
Continuously Scanning
[View Details](#)

All Buckets selected for Compliance

Continuously scanning all selected Buckets

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

Rstor Integrated Configuration

3/55 Buckets selected for Compliance scan

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

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

Result

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

Integrate your Active Directory with Cloud Data Sense

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

When you set up certain data sources (listed below), you need to enter Active Directory credentials in order for Data Sense to scan CIFS volumes. This integration provides Data Sense with file owner and permissions details for those data sources. The Active Directory entered for those data sources may be different than the global Active Directory credentials you enter here. Data Sense will look in all integrated Active Directories for user and permission details.

This helps in the following locations in Data Sense:

- You can use the "File Owner" [filter](#) and see results in the file's metadata in the Investigation pane. Instead of the file owner containing the SID (Security IDentifier), it is populated with the actual user name.
- You can see [full file permissions](#) for each file when you click the "View all Permissions" button.
- In the [Governance dashboard](#), the Open Permissions panel will show a greater level of detail about your data.



Local user SIDs, and SIDs from unknown domains, are not translated to the actual user name.

Supported data sources

An Active Directory integration with Cloud Data Sense can identify data from within the following data sources:

- On-premises ONTAP systems
- Cloud Volumes ONTAP
- Azure NetApp Files
- FSx for ONTAP
- Non-NetApp CIFS file shares (not for NFS file shares)

There is no support for identifying user and permission information from Database schemas, OneDrive accounts, SharePoint accounts, Amazon S3 accounts, or Object Storage that uses the Simple Storage Service (S3) protocol.

Connecting to your Active Directory server

After you've deployed Data Sense and have activated scanning on your data sources, you can integrate Data Sense with your Active Directory.

The Active Directory credentials can be read-only, but providing admin credentials ensures that Data Sense can read any data that requires elevated permissions. The credentials are stored on the Cloud Data Sense instance.

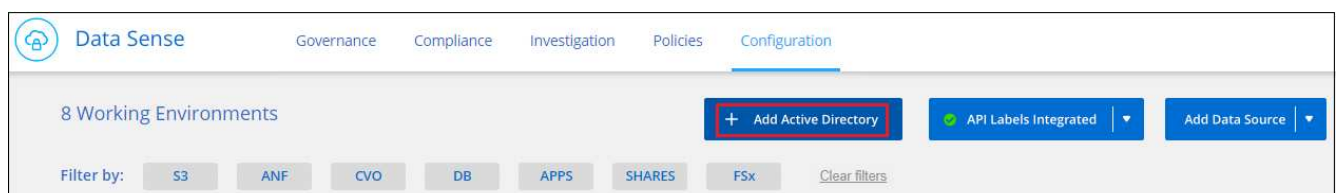
Requirements

- You must have an Active Directory already set up for the users in your company.
- You must have the information for the Active Directory:
 - DNS Server IP Address, or multiple IP Addresses
 - User Name and Password for the server
 - Domain Name (Active Directory Name)
 - Whether you are using secure LDAP (LDAPS) or not
 - LDAP Server Port (typically 389 for LDAP, and 636 for secure LDAP)
- The following ports must be open for outbound communication by the Data Sense instance:

Protocol	Port	Destination	Purpose
TCP & UDP	389	Active Directory	LDAP
TCP	636	Active Directory	LDAP over SSL
TCP	3268	Active Directory	Global Catalog
TCP	3269	Active Directory	Global Catalog over SSL

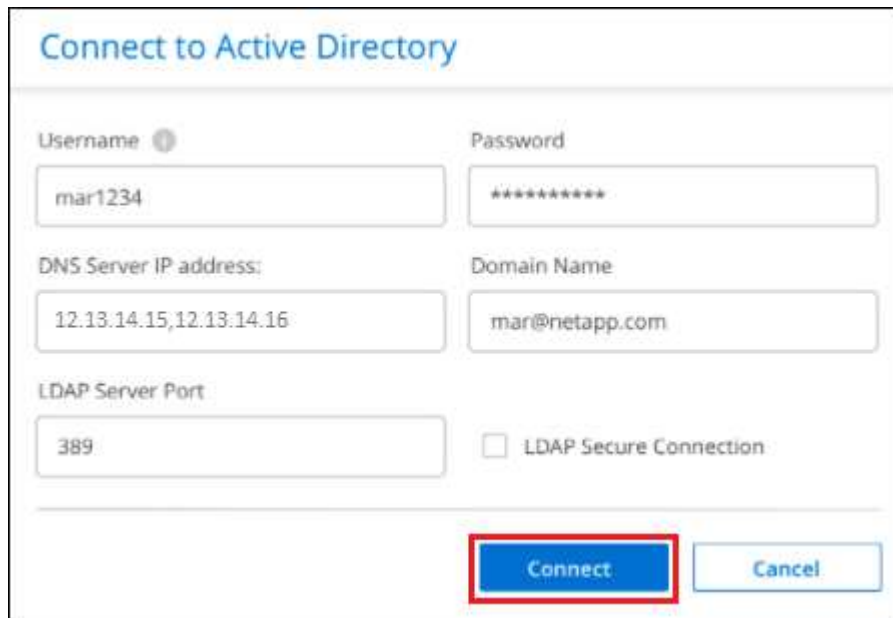
Steps

1. From the Cloud Data Sense Configuration page, click **Add Active Directory**.

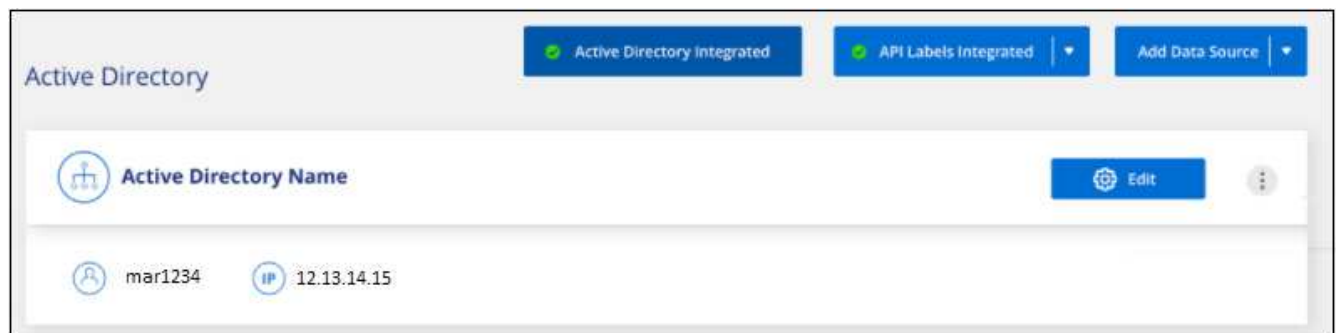


2. In the Connect to Active Directory dialog, enter the Active Directory details and click **Connect**. Multiple IP

Addresses can be added by adding a comma between entries.


A screenshot of a web form titled "Connect to Active Directory". The form contains several input fields: "Username" with the value "mar1234", "Password" with masked characters "*****", "DNS Server IP address:" with the value "12.13.14.15,12.13.14.16", "Domain Name" with the value "mar@netapp.com", and "LDAP Server Port" with the value "389". There is also an unchecked checkbox labeled "LDAP Secure Connection". At the bottom right, there are two buttons: "Connect" (highlighted with a red rectangle) and "Cancel".

Data Sense integrates to the Active Directory, and a new section is added to the Configuration page.

A screenshot of the "Active Directory" configuration page. At the top, there are three status indicators: "Active Directory Integrated" (green checkmark), "API Labels Integrated" (green checkmark), and "Add Data Source" (blue button). Below this, there is a section titled "Active Directory Name" with a tree icon. To the right of this section is an "Edit" button (blue) and a three-dot menu button (grey). Below the "Active Directory Name" section, there are two entries: "mar1234" (with a person icon) and "12.13.14.15" (with an IP icon).

Managing your Active Directory integration

If you need to modify any values in your Active Directory integration, click the **Edit** button and make the changes.

You can also delete the integration if you no longer need it by clicking the  button and then **Remove Active Directory**.

Set up licensing for Cloud Data Sense

The first 1 TB of data that Cloud Data Sense scans in a Cloud Manager workspace is free. A BYOL license from NetApp, or a Cloud Manager subscription from your cloud provider's marketplace, is required to continue scanning data after that point.

A few notes before you read any further:

- If you've already subscribed to the Cloud Manager pay-as-you-go (PAYGO) subscription in your cloud provider's marketplace, then you're automatically subscribed to Cloud Data Sense as well. You won't need

to subscribe again.

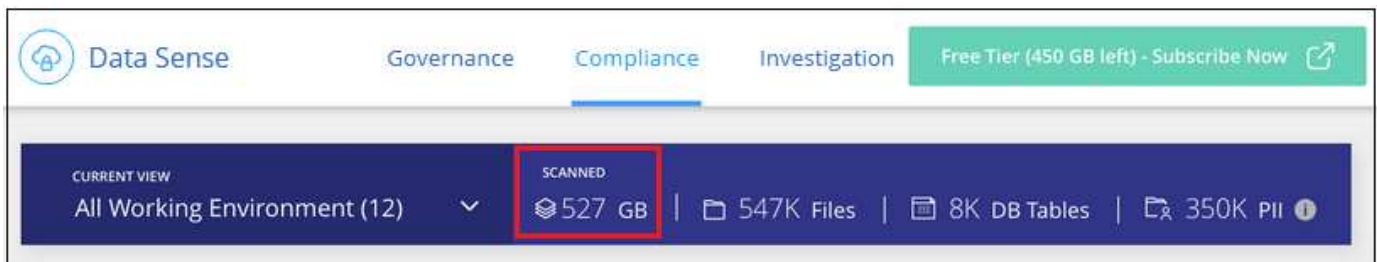
- The Cloud Data Sense bring-your-own-license (BYOL) is a *floating* license that you can use across all the working environments and data sources in the workspace that you plan to scan. You'll see an active subscription in the Digital Wallet.
- There is no charge for scanning backup files created from your on-prem ONTAP volumes (created using [Cloud Backup](#)). This data doesn't count against the capacity purchased in your license.

[Learn more about the licensing and costs related to Cloud Data Sense.](#)

Use a Cloud Data Sense PAYGO subscription

Pay-as-you-go subscriptions from your cloud provider's marketplace enable you to license the use of Cloud Volumes ONTAP systems and many Cloud Data Services, such as Cloud Data Sense.

You can subscribe at any time and you will not be charged until the amount of data exceeds 1 TB. You can always see the total amount of data that is being scanned from the Data Sense Dashboard. And the *Subscribe Now* button makes it easy to subscribe when you are ready.



Note: If you are prompted by Cloud Data Sense to subscribe, but you already have an Azure subscription, you're probably using the old **Cloud Manager** subscription and you need to change to the new **NetApp Cloud Manager** subscription. See [how to change to the new NetApp Cloud Manager plan in Azure](#).

Steps

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

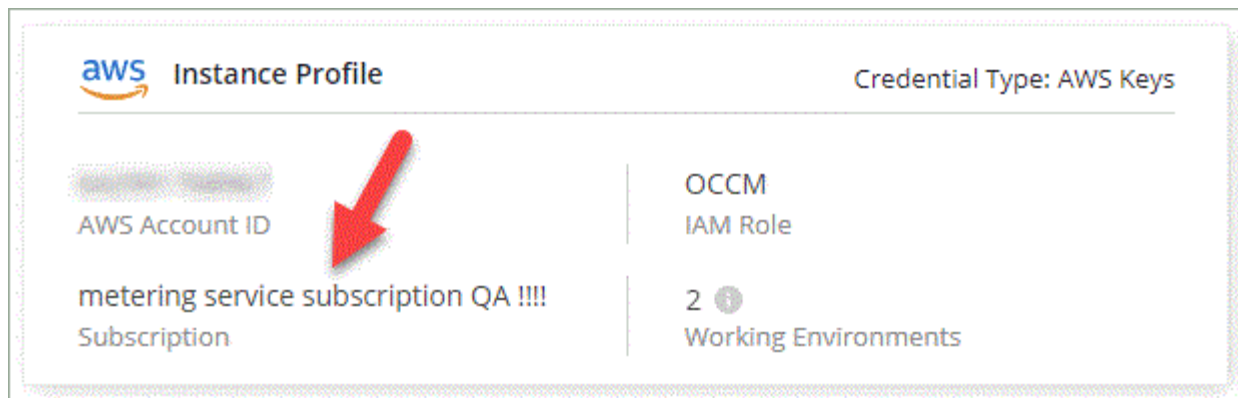
1. In the upper right of the Cloud Manager console, click the Settings icon, and select **Credentials**.



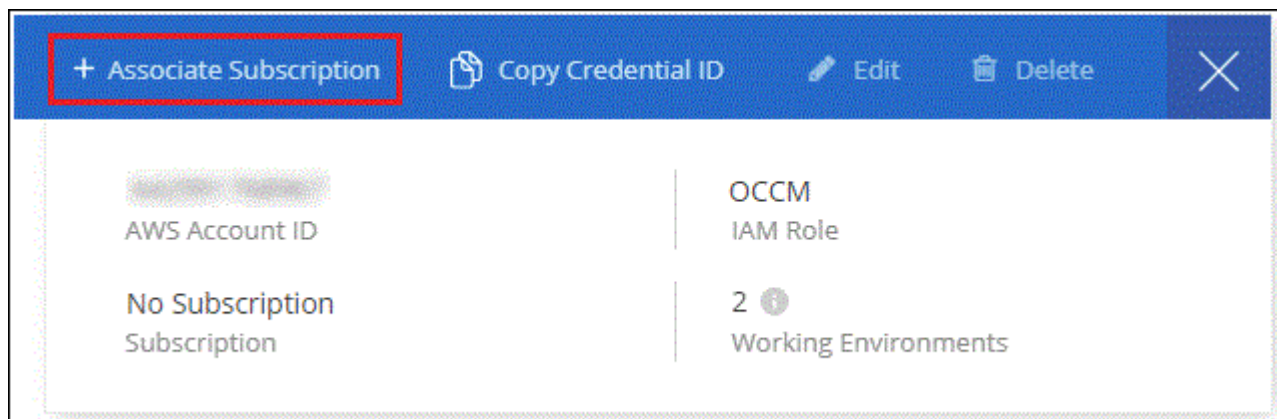
2. Find the credentials for the AWS Instance Profile, Azure Managed Service Identity, or Google Project.

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

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



3. If you don't have a subscription yet, hover over the credentials, click the action menu, and click **Associate Subscription**.



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

The following video shows how to associate an [AWS Marketplace](#) subscription to an AWS subscription:

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

The following video shows how to associate an [Azure Marketplace](#) subscription to an Azure subscription:

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

The following video shows how to associate a [GCP Marketplace](#) subscription to a GCP subscription:

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

Change to the new Cloud Manager plan in Azure

Cloud Data Sense (Cloud Compliance) was added to the Azure Marketplace subscription named **NetApp Cloud Manager** as of October 2020. If you already have the original Azure **Cloud Manager** subscription it will not allow you to use Cloud Data Sense.

You need to follow these steps to change to the new **NetApp Cloud Manager** subscription before you can start using Cloud Data Sense.



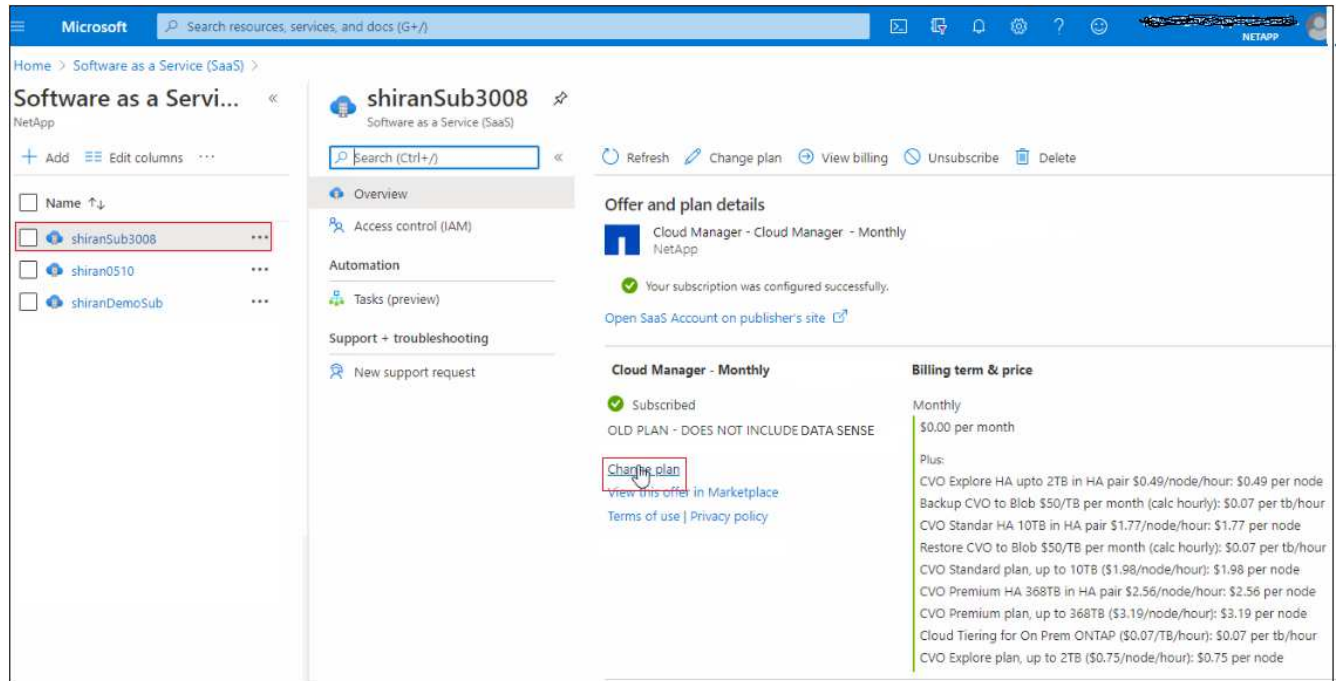
If your existing Subscription was issued with a special private offer, you need to contact NetApp so that we can issue a new special private offer with Data Sense included.

Steps

1. In the upper right of the Cloud Manager console, click the Settings icon, and select **Credentials**.
2. Find the credentials for the Azure Managed Service Identity that you want to change the subscription for and hover over the credentials and click **Associate Subscription**.

The details for your current Marketplace Subscription are displayed.

3. Log in to the [Azure portal](#) and select **Software as a Service (SaaS)**.
4. Select the subscription for which you want to change the plan and click **Change Plan**.



5. In the Change Plan page, select the **NetApp Cloud Manager** plan and click the **Change Plan** button.

Change plan

Subscription plans

i You can only change plans within the billing term of the current plan. If you would like to make further changes please view this offer in the Marketplace, you might need to unsubscribe and resubscribe to a new subscription plan to accomodate more changes.

Billing term ☒ Monthly ☐ Yearly

Software plan	Description	Price
<input checked="" type="radio"/> NetApp Cloud Manager	PLAN - INCLUDES DATA SENSE	\$0.00 per month Plus: CVO Explore HA upto 2TB in HA pair \$0.49/node/hour: \$0.49 per node CVO Premium plan, up to 368TB (\$3.19/node/hour): \$3.19 per node CVO Standard plan, up to 10TB (\$1.98/node/hour): \$1.98 per node Cloud Data Sense \$50/TB/Month: \$0.068 per tb/hour CVO Premium HA 368TB in HA pair \$2.56/node/hour: \$2.56 per node CVO Standar HA 10TB in HA pair \$1.77/node/hour: \$1.77 per node Cloud Tiering for On Prem ONTAP (\$0.07/TB/hour): \$0.07 per tb/hour Backup CVO to Blob \$50/TB per month (calc hourly): \$0.07 per tb/hour CVO Explore plan, up to 2TB (\$0.75/node/hour): \$0.75 per node
<input type="radio"/> Cloud Manager	OLD PLAN - DOES NOT INCLUDE DATA SENSE	\$0.00 per month Plus: CVO Explore HA upto 2TB in HA pair \$0.49/node/hour: \$0.49 per node Backup CVO to Blob \$50/TB per month (calc hourly): \$0.07 per tb/hour CVO Standar HA 10TB in HA pair \$1.77/node/hour: \$1.77 per node Restore CVO to Blob \$50/TB per month (calc hourly): \$0.07 per tb/hour CVO Standard plan, up to 10TB (\$1.98/node/hour): \$1.98 per node CVO Premium HA 368TB in HA pair \$2.56/node/hour: \$2.56 per node CVO Premium plan, up to 368TB (\$3.19/node/hour): \$3.19 per node Cloud Tiering for On Prem ONTAP (\$0.07/TB/hour): \$0.07 per tb/hour CVO Explore plan, up to 2TB (\$0.75/node/hour): \$0.75 per node

☒ Current plan

- Return to Cloud Manager, select the subscription, and hover over the “i” above subscription in the Credentials card to verify your subscription has changed.

Use a Cloud Data Sense BYOL license

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

If you don't have a Cloud Data Sense license, contact us to purchase one:

- [Send email to purchase a license.](#)
- Click the chat icon in the lower-right of Cloud Manager to request a license.

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

You use the Digital Wallet page in Cloud Manager to manage Cloud Data Sense BYOL licenses. You can add new licenses and update existing licenses.

Obtain your Cloud Data Sense license file

After you have purchased your Cloud Data Sense license, you activate the license in Cloud Manager by entering the Cloud Data Sense serial number and NSS account, or by uploading the NLF license file. The steps below show how to get the NLF license file if you plan to use that method.

Steps

1. Sign in to the [NetApp Support Site](#) and click **Systems > Software Licenses**.
2. Enter your Cloud Data Sense license serial number.

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

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

Get License

SERIAL NUMBER: 4810

LICENSE: SUBS-CLD-DAT-SENSE-TB-2Y

SALES ORDER: 3005

TENANT ID:

Example: account-xxxxxxx

[Cancel](#) [Submit](#)

You can find your Cloud Manager Account ID by selecting the **Account** drop-down from the top of Cloud Manager, and then clicking **Manage Account** next to your account. Your Account ID is in the Overview tab.

Add Cloud Data Sense BYOL licenses to your account

After you purchase a Cloud Data Sense license for your Cloud Manager account, you need to add the license to Cloud Manager to use the Cloud Data Sense service.

Steps

1. Click **All Services > Digital Wallet > Data Services Licenses**.
2. Click **Add License**.
3. In the *Add License* dialog, enter the license information and click **Add License**:
 - If you have the Data Sense license serial number and know your NSS account, select the **Enter Serial Number** option and enter that information.

If your NetApp Support Site account isn't available from the drop-down list, [add the NSS account to Cloud Manager](#).

- If you have the Data Sense license file, select the **Upload License File** option and follow the prompts to attach the file.

Add License

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

☒ Enter Serial Number
 ☐ Upload License File

Serial Number

NetApp Support Site Account

Add License **Cancel**

☐ Enter Serial Number
 ☒ Upload License File

To install a license, follow these instructions:

- 1 Obtain the license file from the "System > Software Licenses" tab at [NetApp Support Site](#). You will need to provide your cloud service serial number and Cloud Manager Account ID.
- 2 Click Upload File and then select the file.

Upload License File

 Upload

Add License **Cancel**

Result

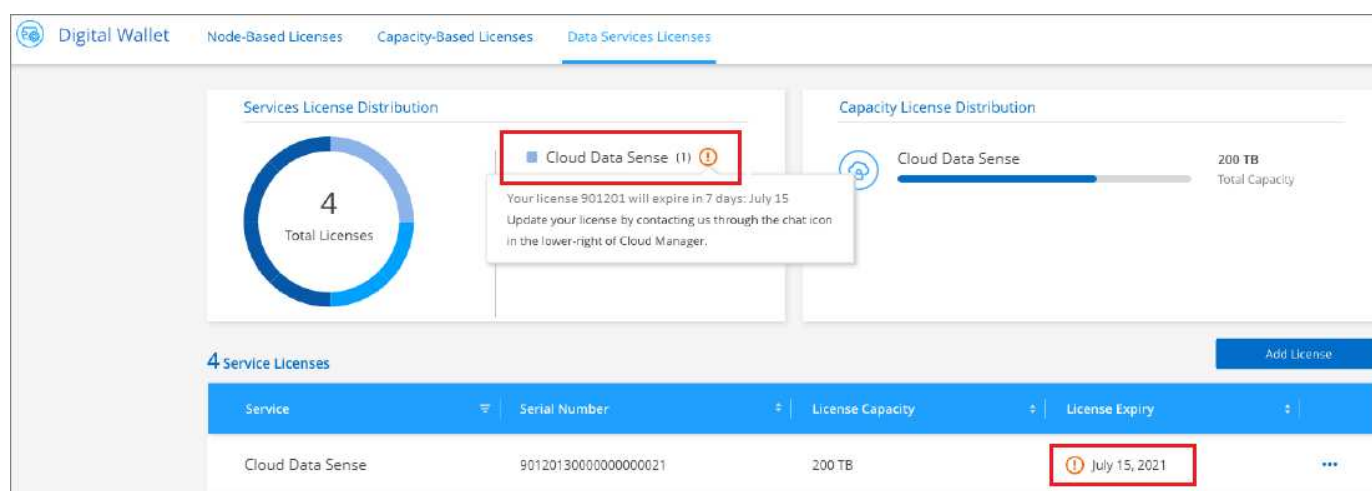
Cloud Manager adds the license so that your Cloud Data Sense service is active.

Update a Cloud Data Sense BYOL license

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



This status also appears in the Digital Wallet page.



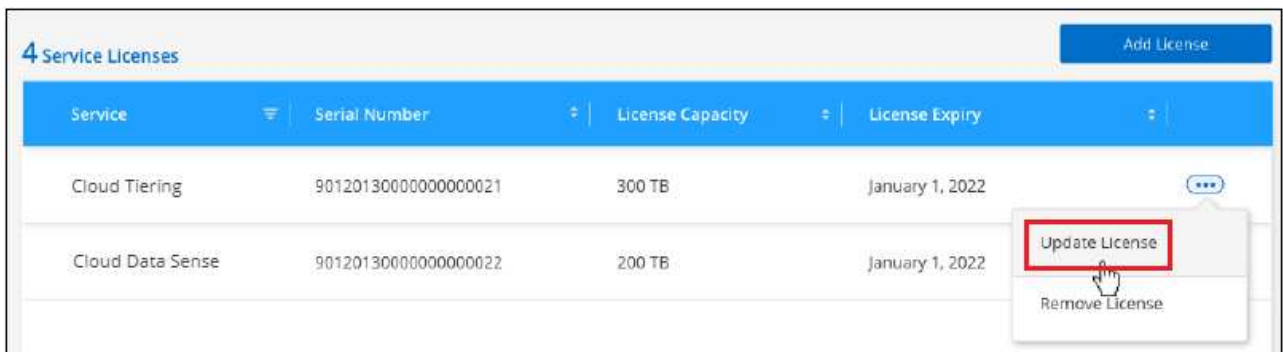
You can update your Cloud Data Sense license before it expires so that there is no interruption in your ability to scan your data.

Steps

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

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

2. If Cloud Manager can't automatically update the license (for example, when installed in a dark site), then you'll need to manually upload the license file.
 - a. You can [obtain the license file from the NetApp Support Site](#).
 - b. On the Digital Wallet page in the *Data Services Licenses* tab, click ... for the service serial number you are updating, and click **Update License**.



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

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

Result

Cloud Manager updates the license so that your Cloud Data Sense service continues to be active.

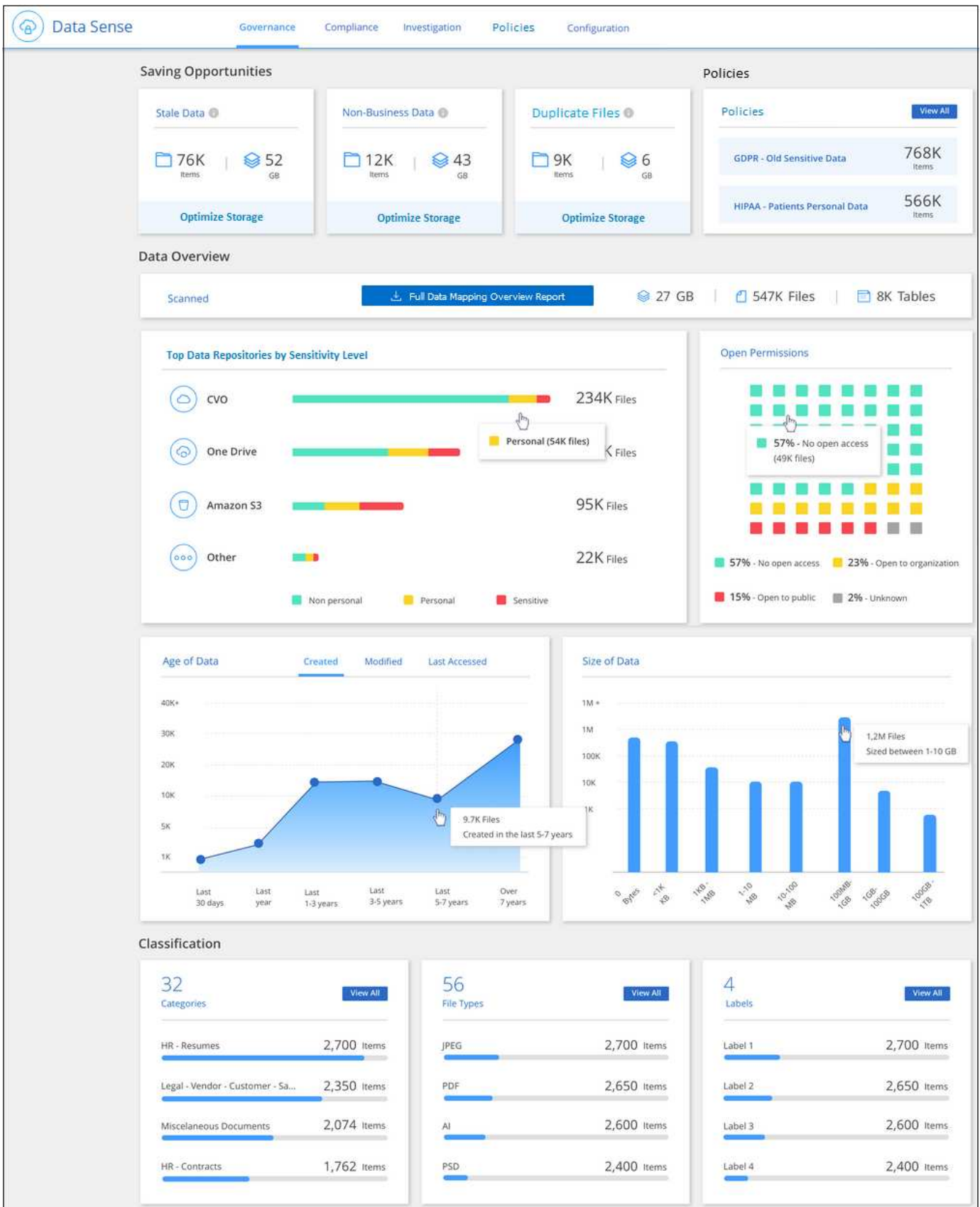
Viewing governance details about the data stored in your organization

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.



Files in OneDrive accounts, SharePoint accounts, and in databases are not represented in this chart.

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.

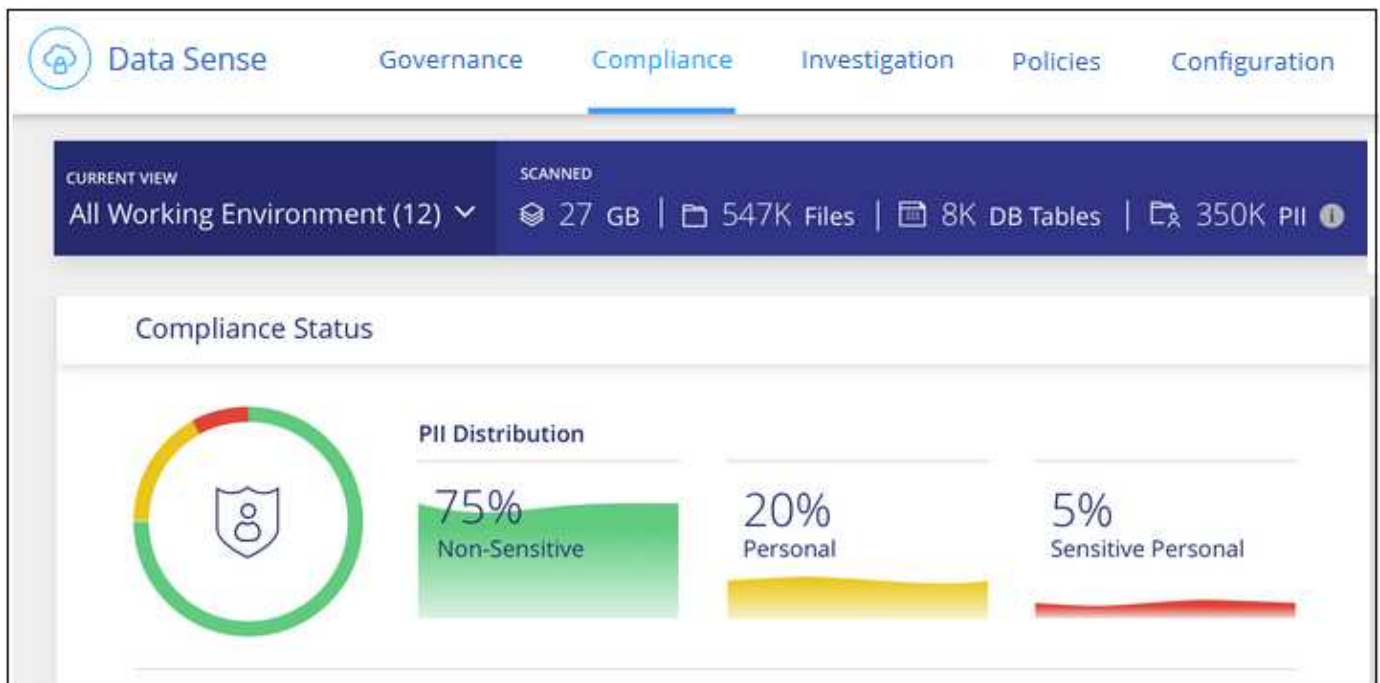
Viewing compliance details about the data stored in your organization

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, 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

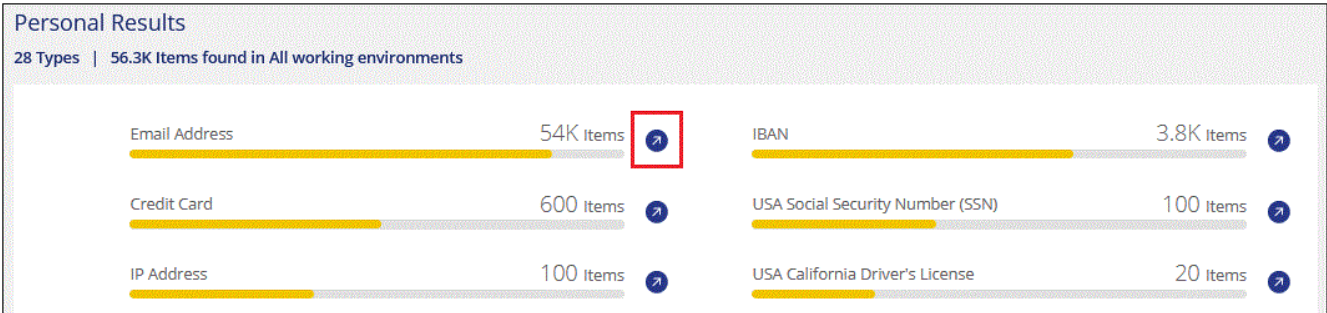
example, Data Sense identifies a U.S. social security number (SSN) as a SSN if it sees a proximity word next to it—for example, *SSN* or *social security*. [The table of personal data](#) shows when Data Sense uses proximity validation.

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."



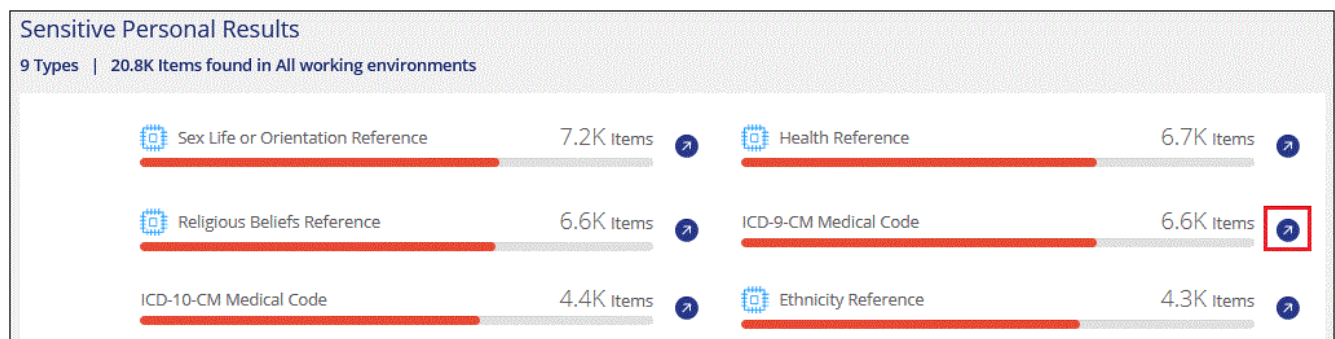
Only English is supported when scanning for sensitive personal data. Support for more languages will be added later.

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.](#)

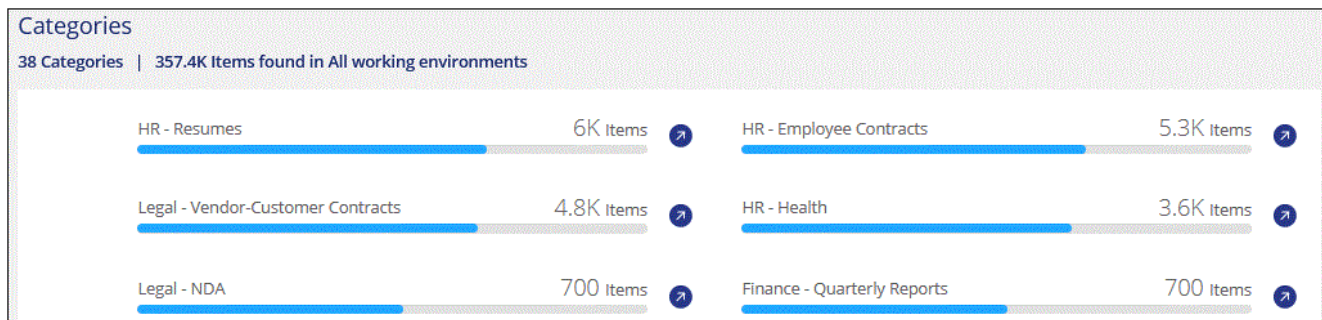
Categories can help you understand what's happening with your data by showing you the types of information that you have. For example, a category like resumes or employee contracts can include sensitive data. When you investigate the results, you might find that employee contracts are stored in an insecure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

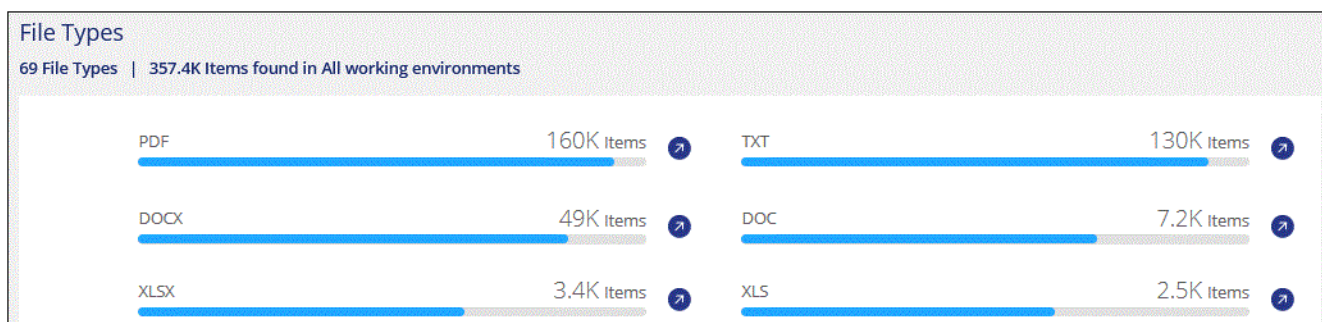
Viewing files by file types

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.

Unstructured (32K Files)

Structured (323 DB Tables)

File Name

Personal

Sensitive Personal

Data Subjects

File Type

Expense Report EXP-TPO-106038887654

cvo

6

3

16

PDF

Expense Report EXP-TPO-106038887654

cvo

6

3

16

PDF

Working Environment: WorkingEnvironment1

Repository: Volume Name

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

Category: Legal

File Size: 22 MB

Created: 2013-01-05 08:22

Last Modified: 2019-08-06 07:51

Last Accessed: 2019-08-06 07:51

Open Permissions: NO OPEN PERMISSIONS

View all Permissions

File Owner: Asaf Ley

Duplicates: 3

View Details

Tags: 0 tags

Assign a Label to this file

Delete this file

Give feedback on this result

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.

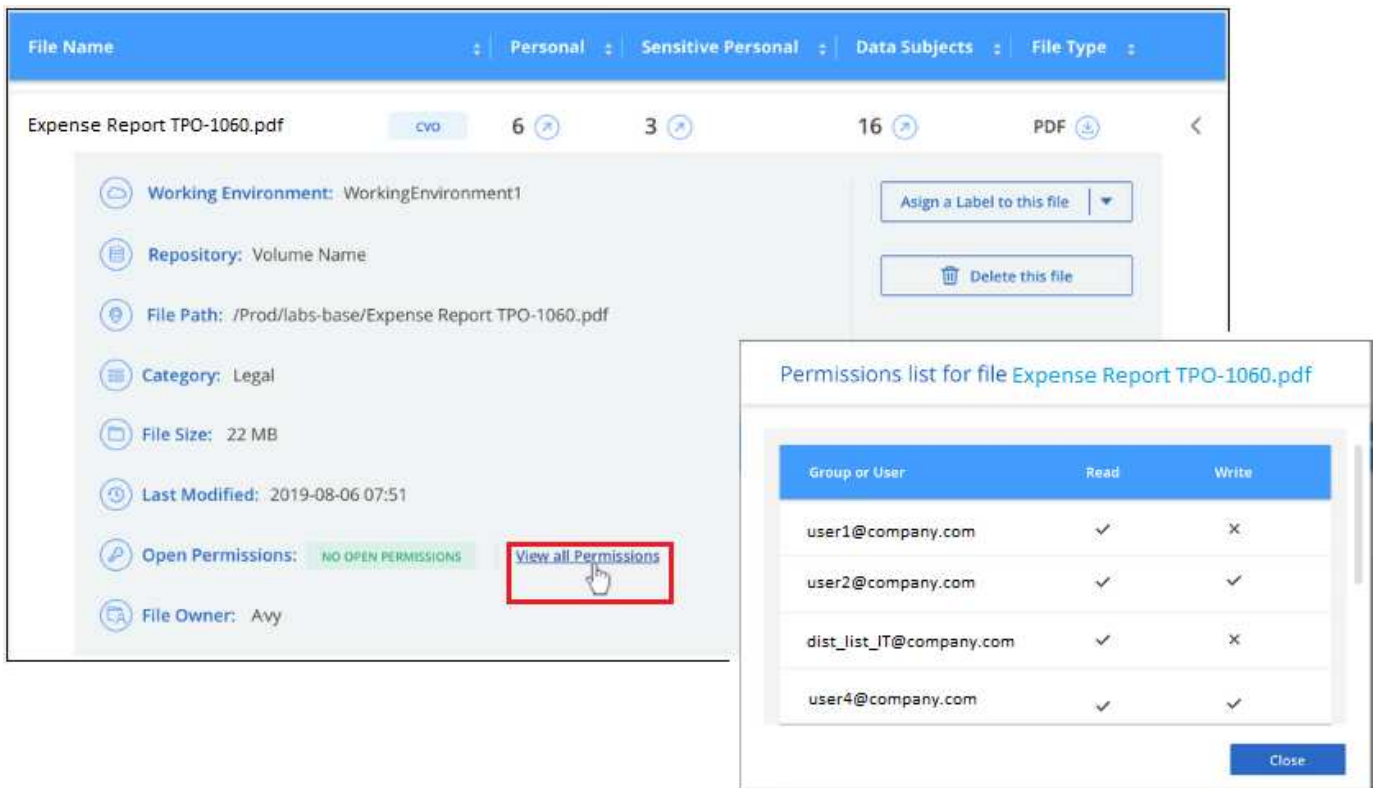
Note that not all information is available for all data sources - just what is appropriate for that data source. For example, volume name, permissions, and AIP labels are not relevant for database files.

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 Status 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.

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.](#)

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.

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.

Last Modified: 2019-08-06 07:51

Open Permissions: NO OPEN PERMISSIONS [View all Permissions](#)

File Owner: Asaf Ley

Duplicates: 3 [View Details](#)

Duplicates of File 'Name 1'

Duplicates: 3

Total Size of all Duplicates: 1GB

File Hash: xxxxxx

[View Duplicates](#)

[Close](#)

3 items

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



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

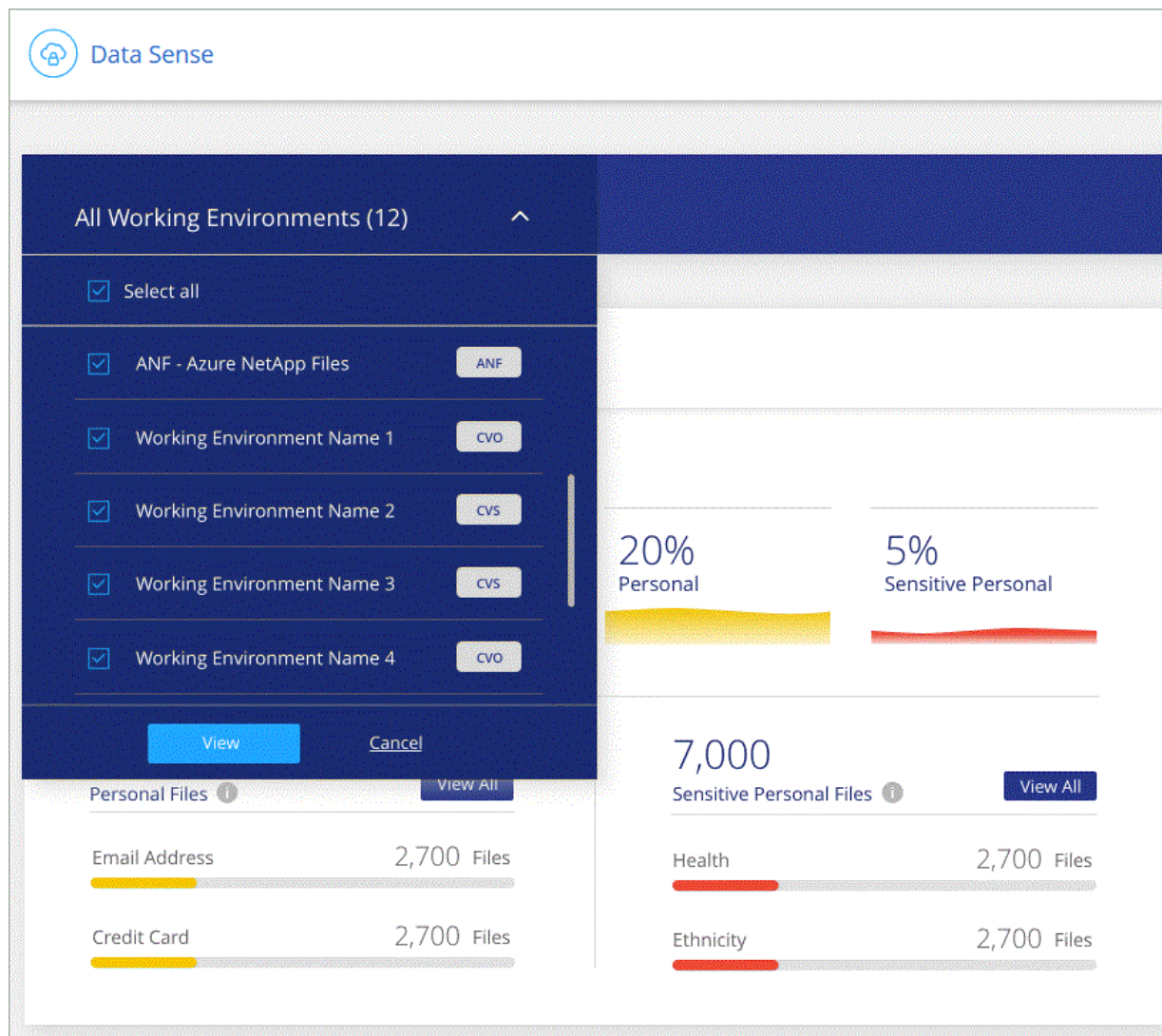
Viewing Dashboard data for specific working environments

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. If you want to save a CSV version of the content as a report after you have refined it, click the  button.

Data Investigation		Unstructured (252K Files)	Structured (0 Tables)	Search by File or DB Table name or location			Download
FILTERS: Clear All		252K items					
		Tags Assign to Label Move Copy Delete					
Policies + Open Permissions + File Owner + Label + Working Environment Type 2 + Working Environment + Storage Repository 2 +		<input type="checkbox"/> File Name	<input type="checkbox"/> Personal	<input type="checkbox"/> Sensitive Personal	<input type="checkbox"/> Data Subjects	<input type="checkbox"/> 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) 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:
 - Policies
 - Open Permissions
 - File Owner
 - Label
 - Working Environment Type (note that OneDrive and SharePoint are categorized under "Cloud Apps")
 - Working Environment name
 - Storage Repository (for example, a volume or a schema)
 - File Path
 - Category
 - Sensitivity Level
 - Personal Data
 - Sensitive Personal Data
 - Data Subject
 - File Type
 - File Size
 - Created Time
 - Last Modified
 - Last Accessed (For the [types of files that we scan for content](#) this is the last time Data Sense scanned the file.)
 - Duplicates
 - File Hash

- Tags
- Assigned To
- The *Policies* filter at the top of the Filters pane lists the custom filters that provide commonly requested combinations of filters; like a saved database query or Favorites list. Go [here](#) to view the list of predefined Policies and to see how you can create your own custom Policies.

What's included in each file list report (CSV file)

From each Investigation page you can click the  button to download file lists (in CSV format) that include details about the identified files. If Data Sense is scanning both Structured (database tables) and Unstructured (files) data, there are two reports contained in the downloaded ZIP file.

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

The **Unstructured Data Report** includes the following information:

- File name
- Location type
- Working environment type
- Working environment name
- Storage repository (for example, a volume)
- Protocol type
- File path
- File type
- File size
- Created time
- Discovered time (when Data Sense first discovered the file)
- 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:

- DB Table name
- Location type
- Working environment type

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

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).
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 and SharePoint 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

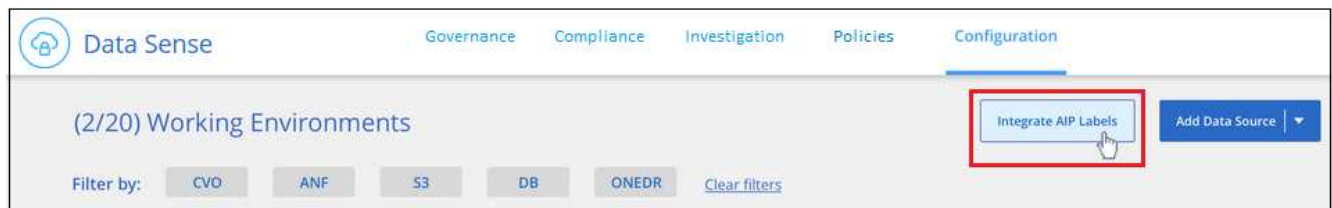
Before you can manage AIP labels, you need to integrate the AIP label functionality into Cloud Data Sense by signing into your existing Azure account. Once enabled, you can manage AIP labels within files for all [working environments and data sources](#) in your Cloud Manager workspace.

Requirements

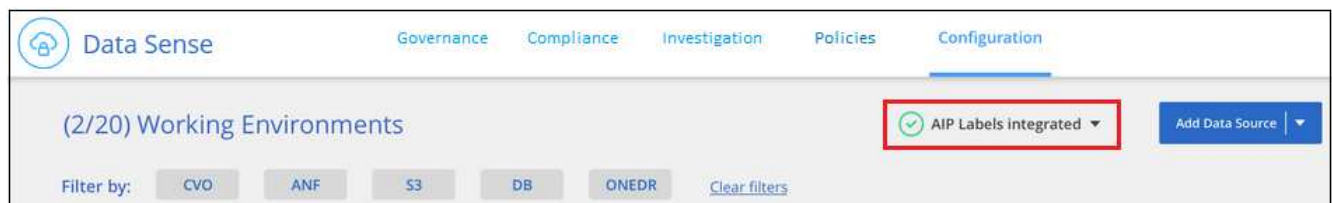
- You must have an account and an Azure Information Protection license.
- You must have the login credentials for the Azure account.
- If you plan to change labels in files that reside in Amazon S3 buckets, ensure that the permission `s3:PutObject` is included in the IAM role. See [setting up the IAM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

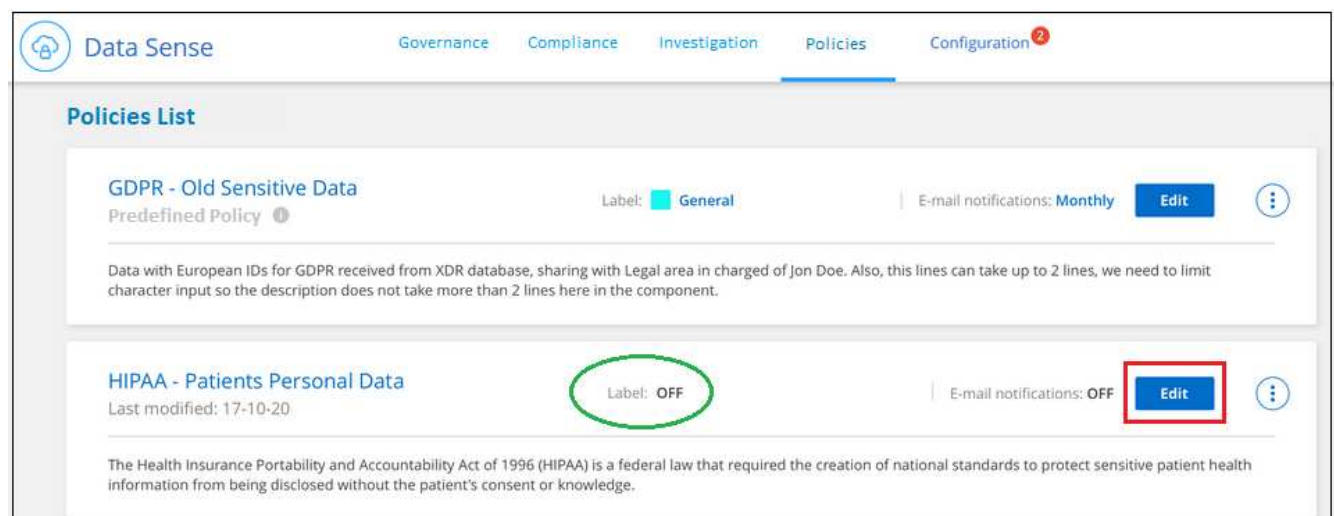
Depending on whether a label is already applied to a file, and the classification level of the label, the following actions are taken when changing a label:

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.

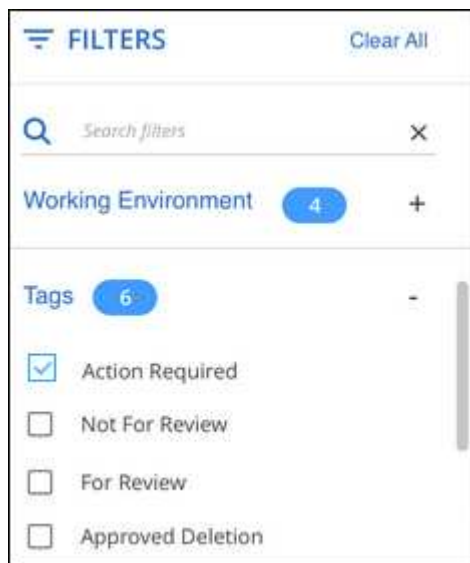


Tags assigned to files in Cloud Data Sense are not related to the tags you can add to resources, such as volumes or virtual machine instances. Data Sense tags are applied at the file level.

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.

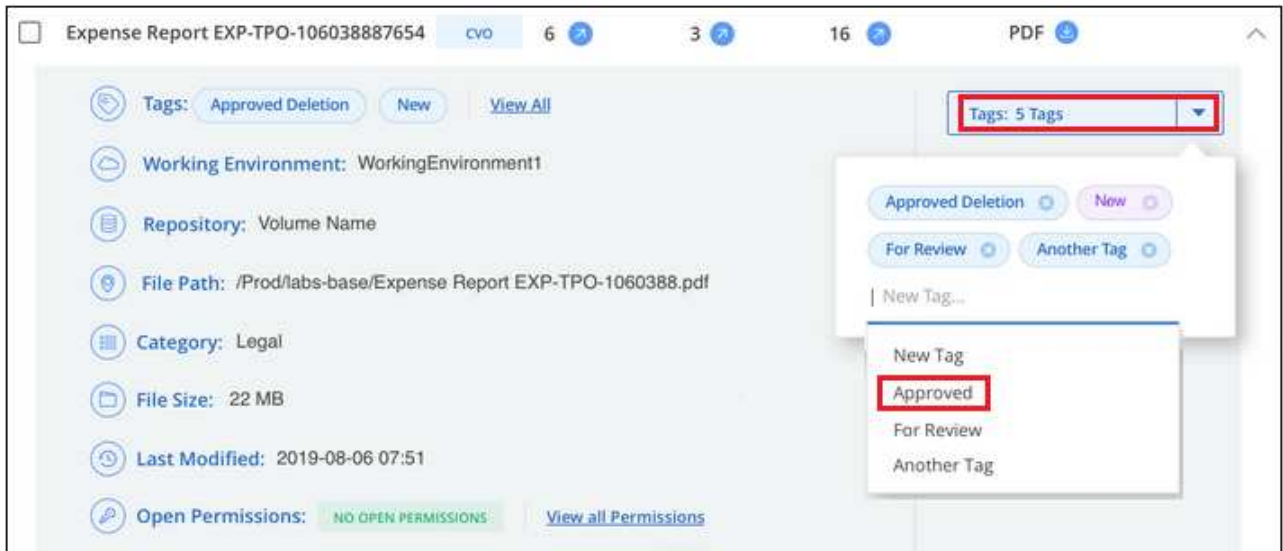
To add a tag to a single file:

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-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	▼

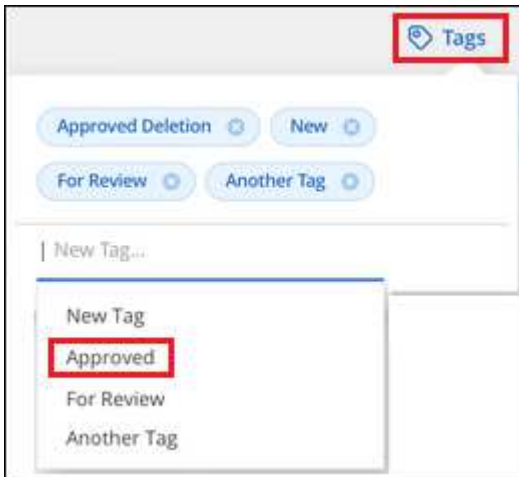
- To select individual files, check the box for each file (☒ Volume_1).
- To select all files on the current page, check the box in the title row (☒ File Name).

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

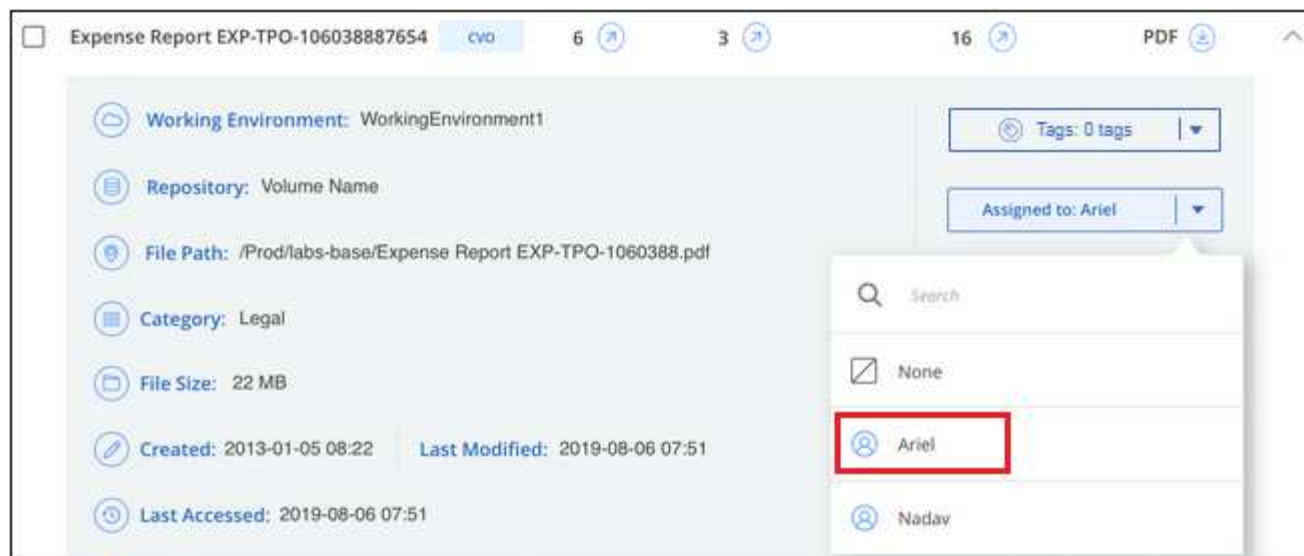
Note that the user name is not added to the file as part of the file metadata - it is just seen by Cloud Manager users when using Cloud Data Sense.

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:

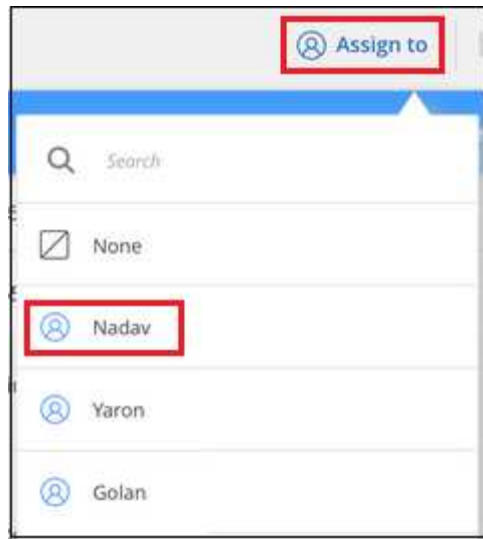
Steps

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.

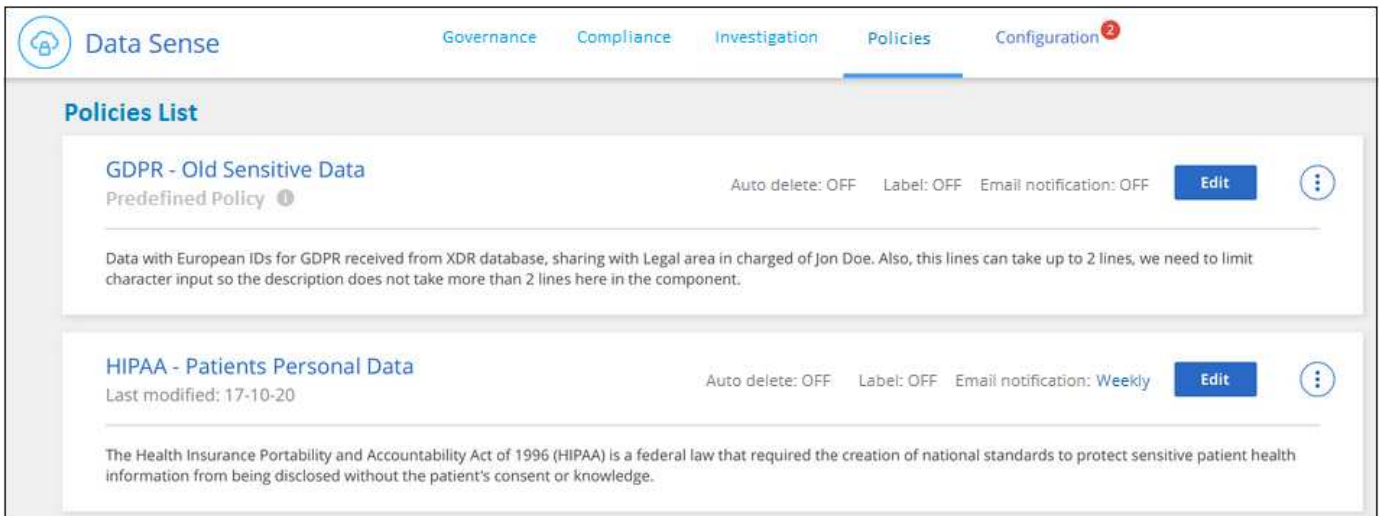
Controlling your data using Policies

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

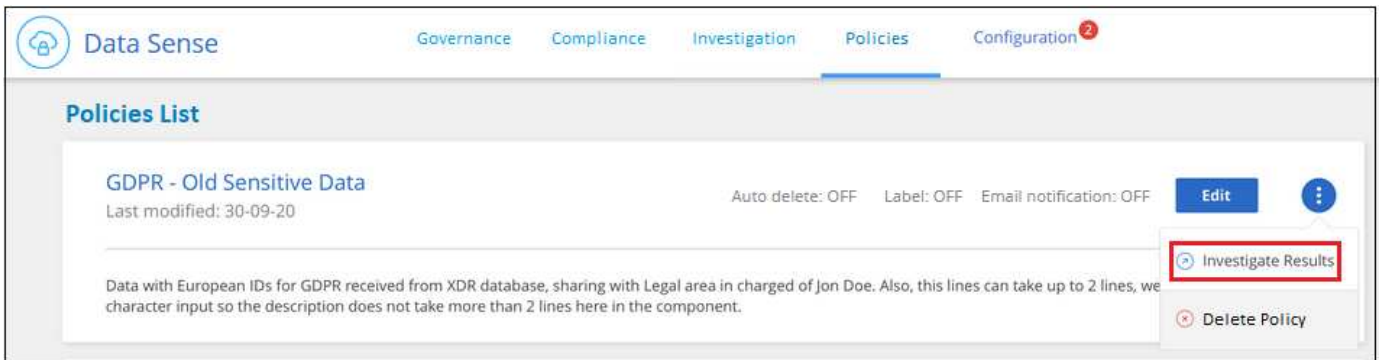
The **Policies** tab in the Compliance Dashboard lists all the predefined and custom Policies available on this instance of Cloud Data Sense.



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**.

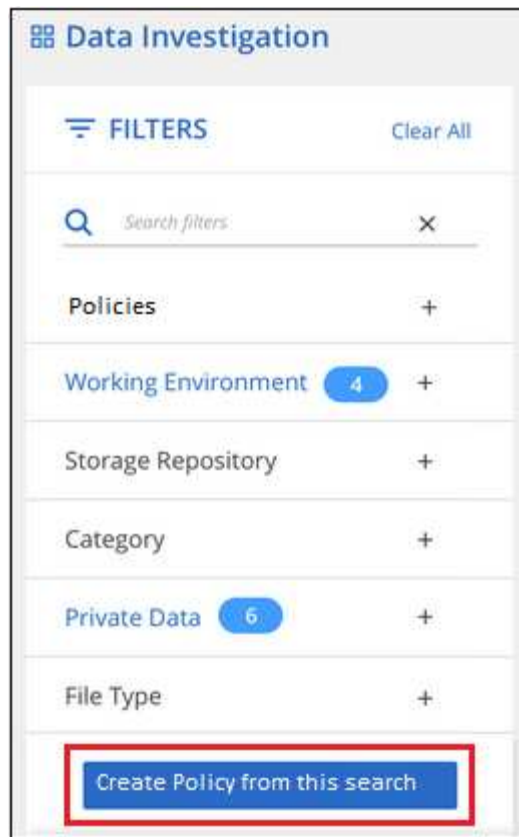


Creating custom Policies

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

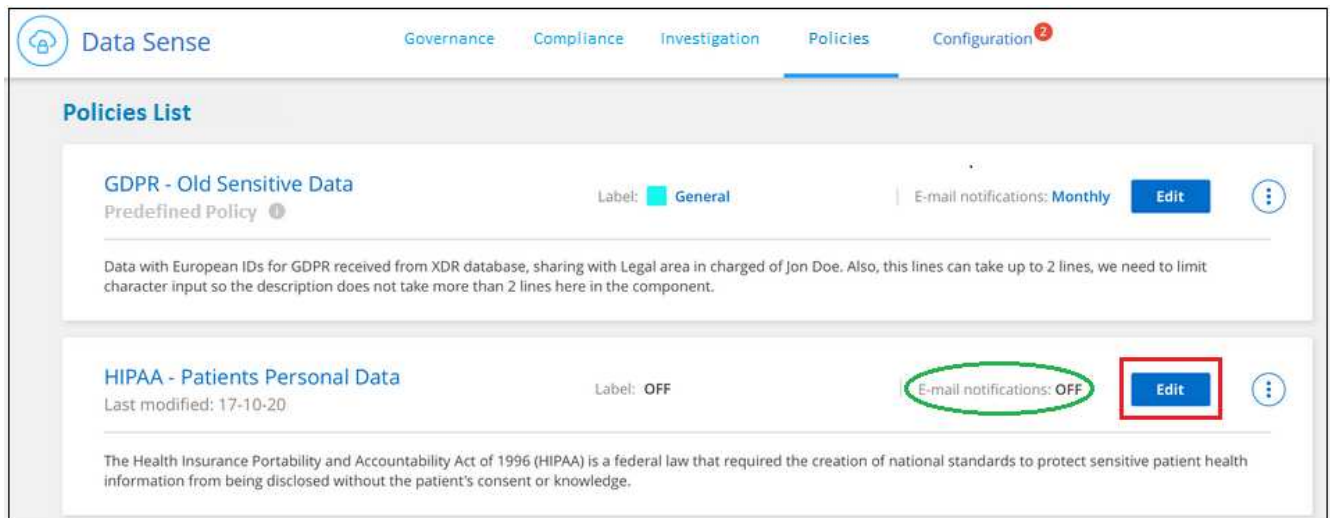
Cloud Data Sense can send email alerts to Cloud Manager users when certain critical Policies return results so you can get notifications to protect your data. You can choose to send the email notifications on a daily, weekly, or monthly basis.

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.

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 or in volume backups.

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.

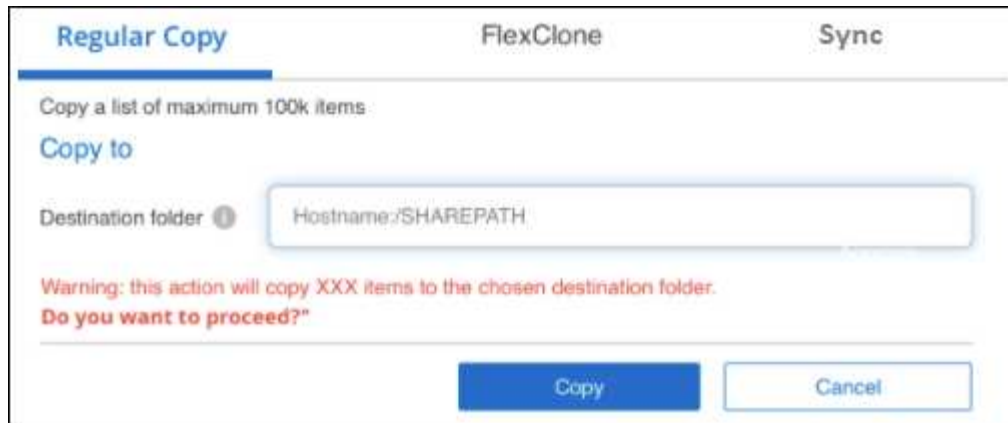
Steps

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.

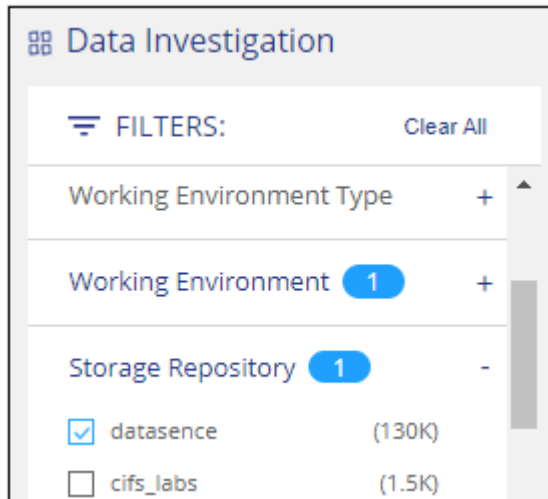
Requirements

- 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

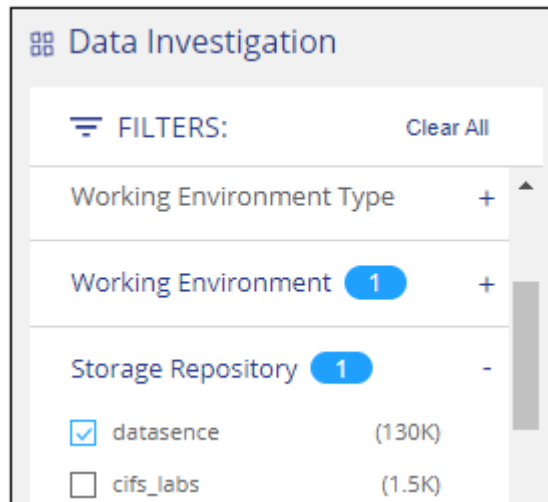
- You must have the Account Admin or Workspace Admin role to copy and sync files.
- All selected files must be from the same source repository (ONTAP volume, S3 bucket, NFS or CIFS share, etc.).
- You can copy a maximum of 200,000 files at a time.

- 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.

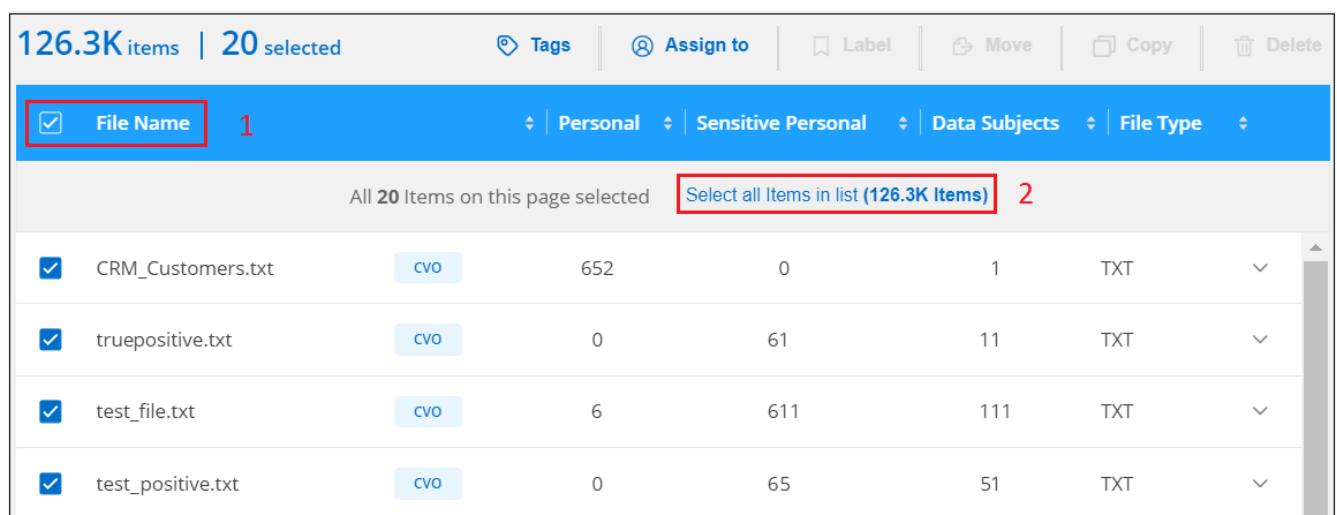
Steps

- 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.

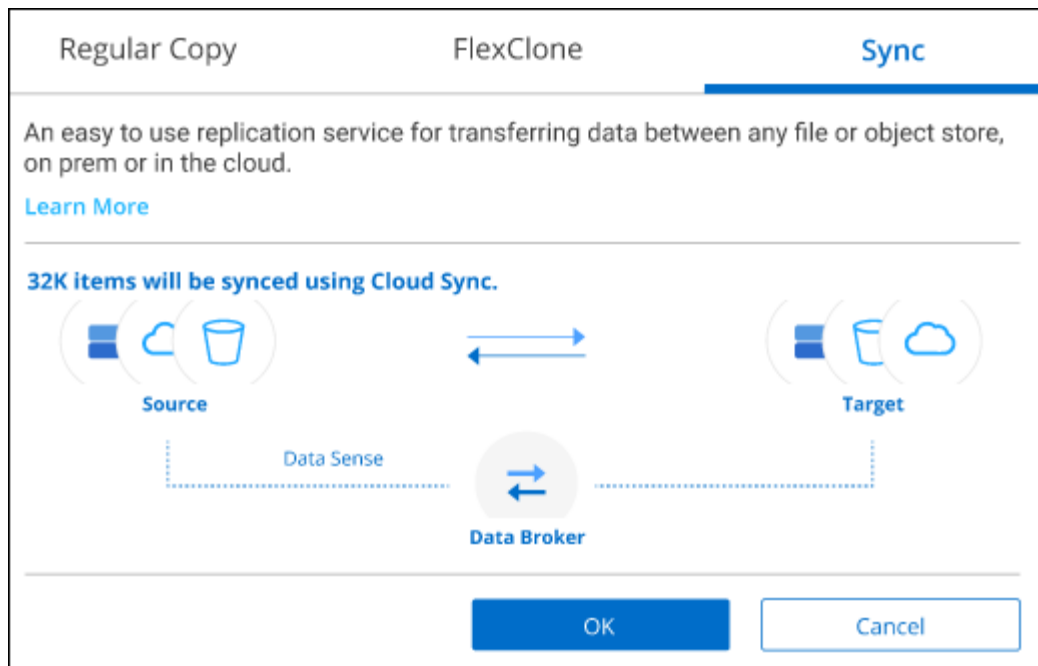


Apply any other filters so that you're seeing only the files that you want to copy and sync to the destination system.

- 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**.



- 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 or files that reside in volume backups.

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

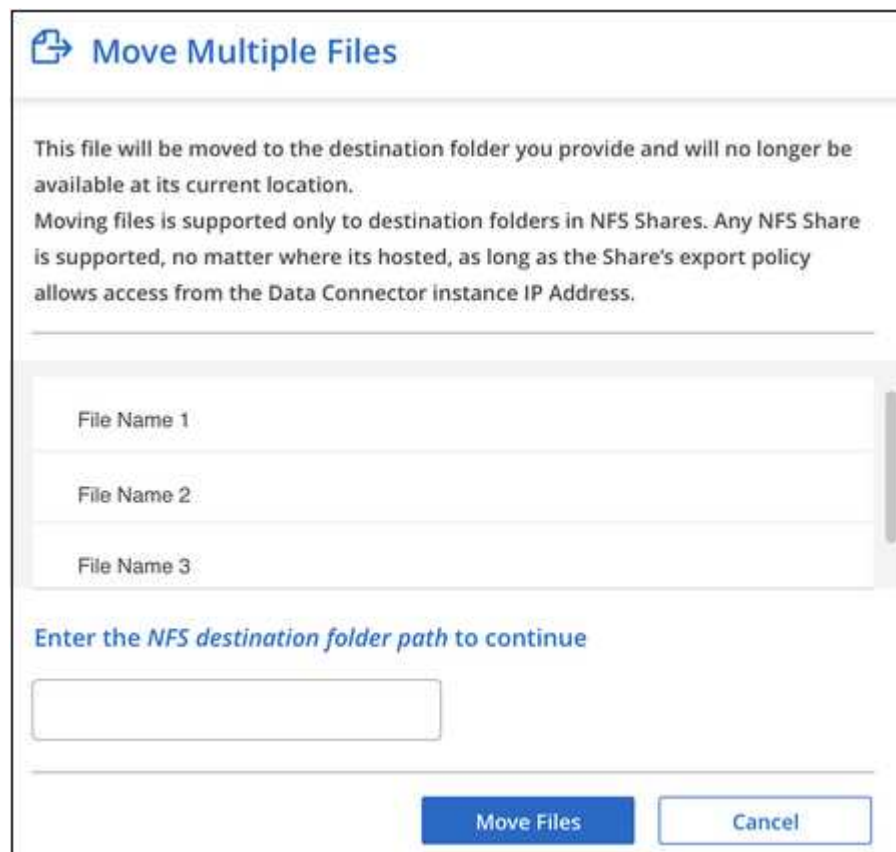
Copy

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 or files that reside in volume backups.

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

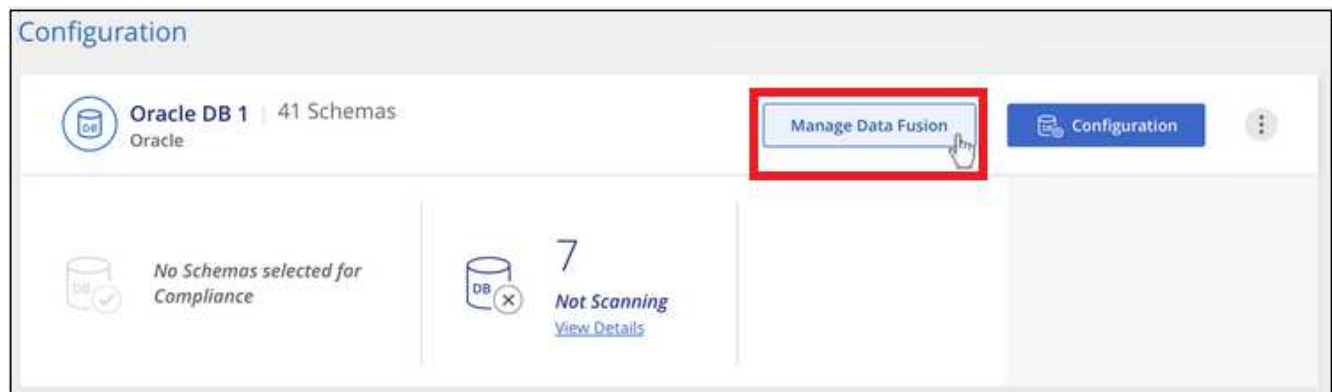
```
XXXXXXXXXXXXX
XXXXXXXXXXXXX
xx213536xxx
XXXXXXXXXXXXX
XXXXXXXXXXXXX
XXXXXXXXXXXXX
```

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.

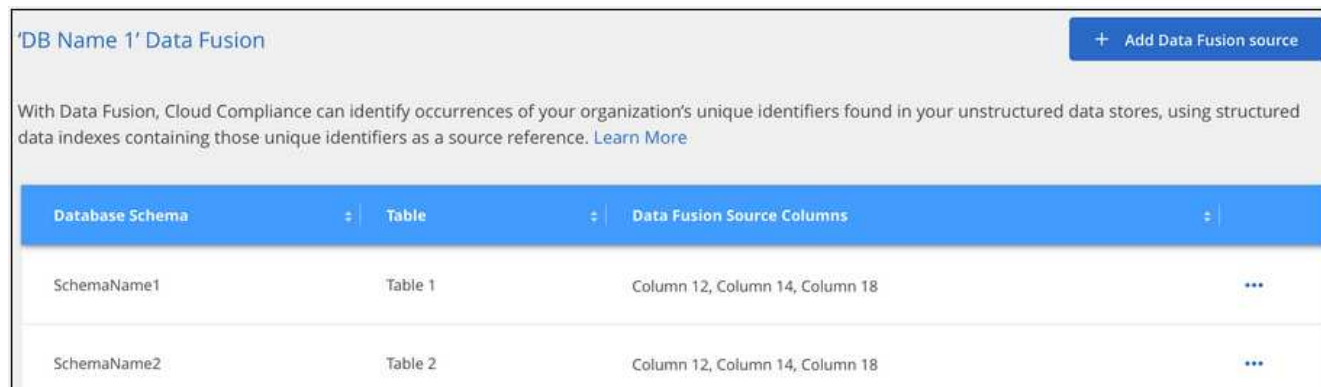


2. Click **Add Data Fusion source** on the next page.
3. In the *Add Data Fusion Source* page:
 - a. Select the Database Schema from the drop-down menu.
 - b. Enter the Table name in that schema.
 - c. Enter the Column, or Columns, that contain the unique identifiers you want to use.

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 and databases. 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

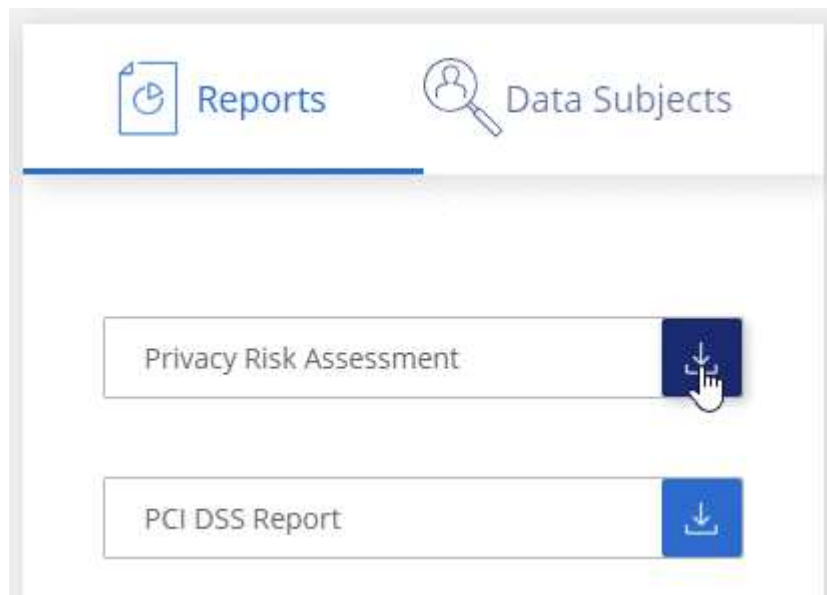
The number of people, by location, for which national identifiers were found.

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

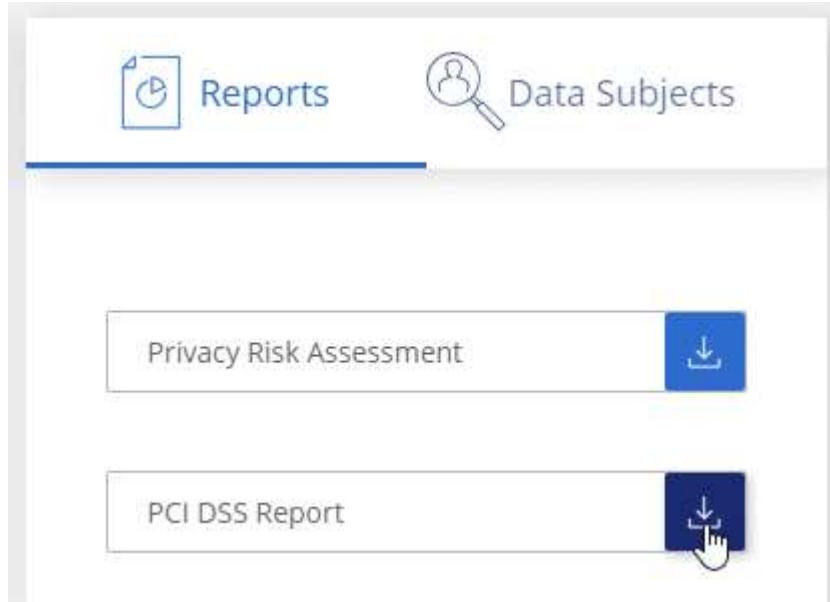
The working environments where the credit card information was found and whether encryption and ransomware protection are enabled.

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

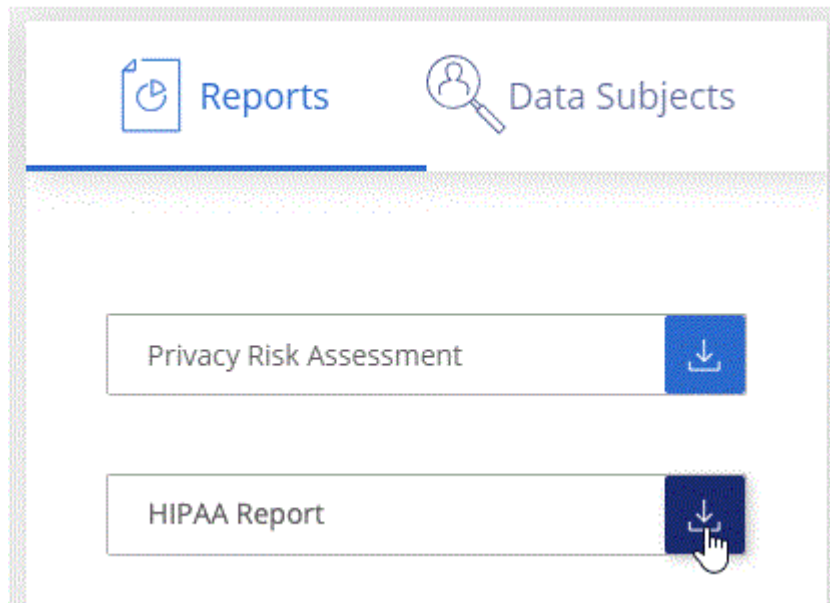
The working environments where the health information was found and whether encryption and ransomware protection are enabled.

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.

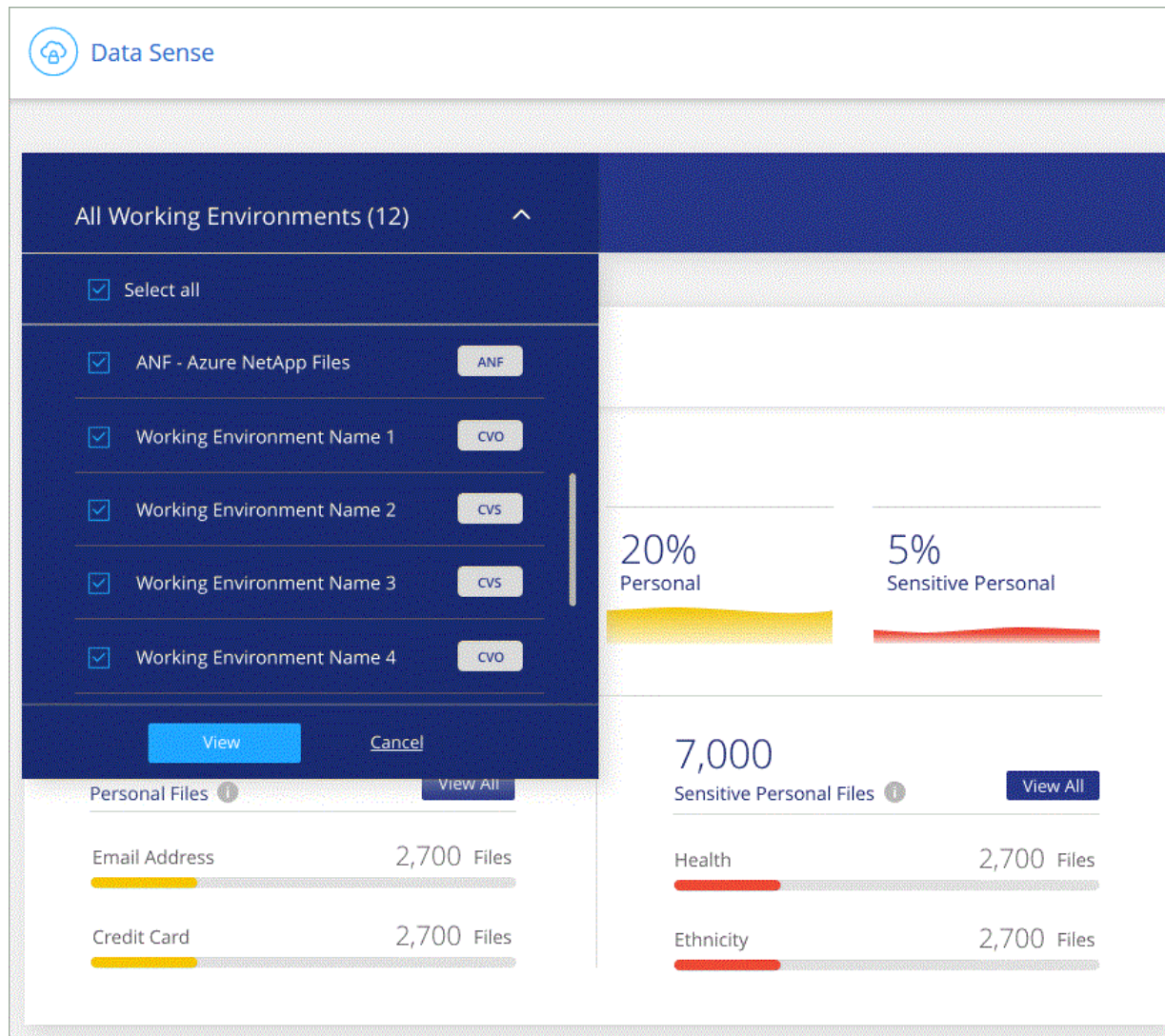
Selecting the working environments for reports

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](#).

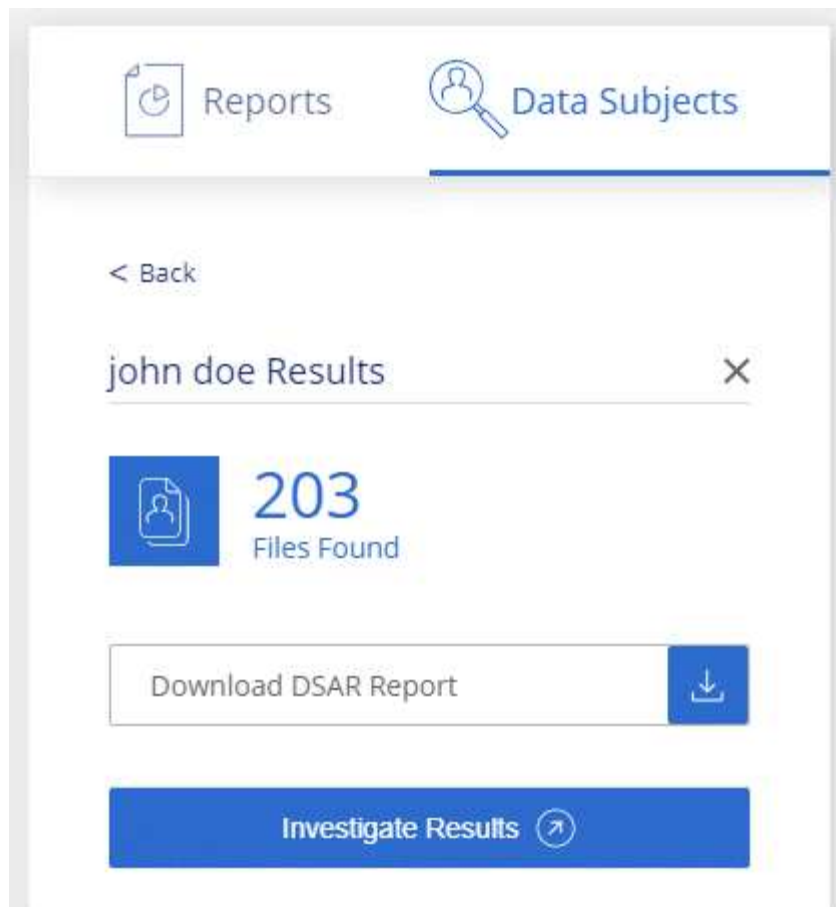


English, German, and Spanish are supported when searching for the names of data subjects. Support for more languages will be added later.

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, and SharePoint 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

Type	Identifier	Proximity validation?
National Identifiers		
130		

Type	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

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, and .XLSX.

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%


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

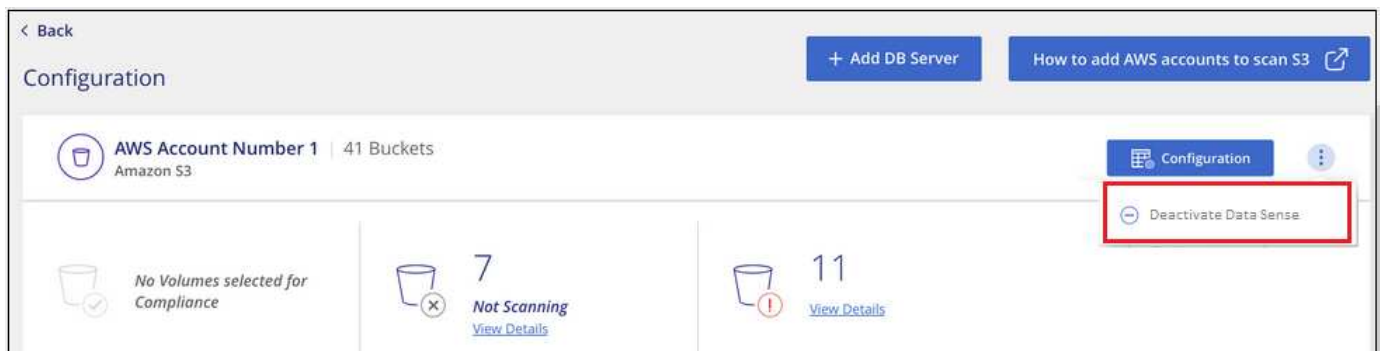
Removing data sources from Cloud Data Sense

If you need to, you can stop Cloud Data Sense from scanning one or more working environments, databases, file share groups, OneDrive accounts, 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**.

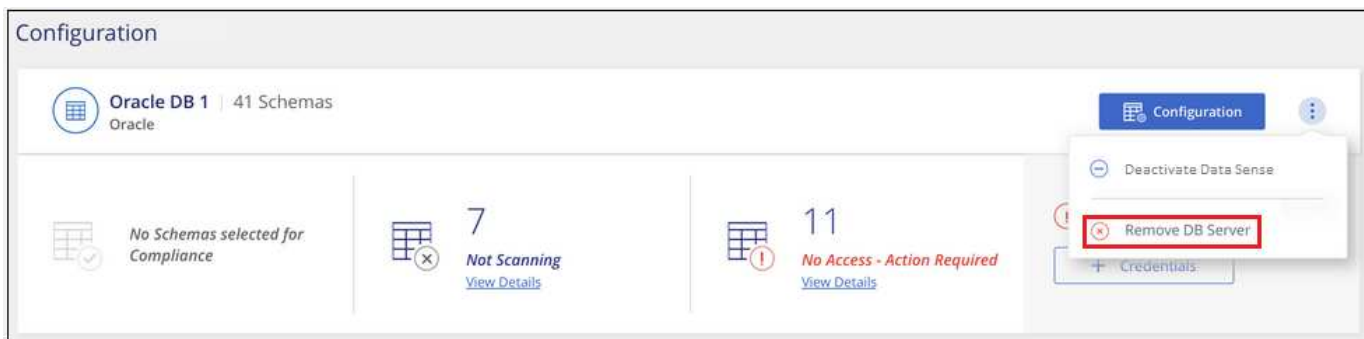


You can also disable compliance scans for a working environment from the Services panel when you select the working environment.

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 or SharePoint account from Cloud Data Sense

If you no longer want to scan user files from a certain OneDrive account, or from a specific SharePoint 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 or SharePoint account, and then click **Remove OneDrive Account** or **Remove SharePoint 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.



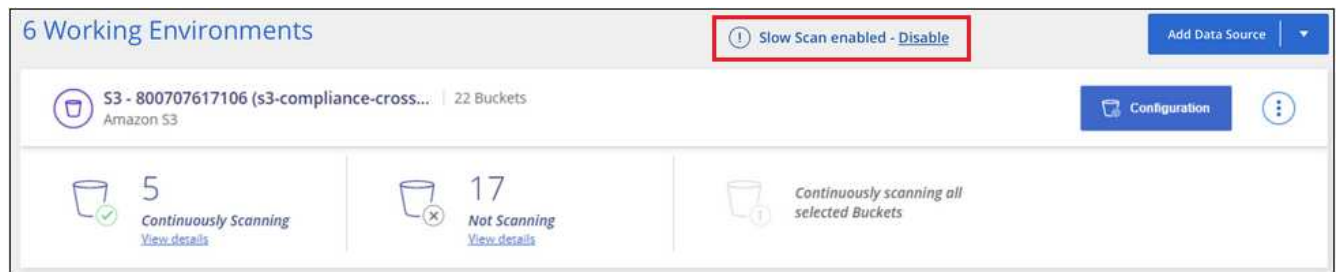
The scan speed can't be reduced when scanning databases.

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*

Frequently asked questions about Cloud Data Sense

This FAQ can help if you're just looking for a quick answer to a question.

What is Cloud Data Sense?

Cloud Data Sense is a cloud offering that uses Artificial Intelligence (AI) driven technology to help organizations understand data context and identify sensitive data across your storage systems. The systems can be Azure NetApp Files configurations, Amazon FSx for ONTAP, Cloud Volumes ONTAP systems (hosted in AWS, Azure, or GCP), Amazon S3 buckets, on-prem ONTAP systems, non-NetApp file shares, generic S3 object storage, databases, OneDrive accounts, and SharePoint accounts.

Cloud Data Sense provides pre-defined parameters (such as sensitive information types and categories) to address new data compliance regulations for data privacy and sensitivity, such as GDPR, CCPA, HIPAA, and more.

Why should I use Cloud Data Sense?

Cloud Data Sense can empower you with data to help you:

- Comply with data compliance and privacy regulations.
- Comply with data retention policies.
- Easily locate and report on specific data in response to data subjects, as required by GDPR, CCPA, HIPAA, and other data privacy regulations.

What are the common use cases for Cloud Data Sense?

- Identify Personal Identifiable Information (PII).
- Identify a wide scope of sensitive information as required by GDPR and CCPA privacy regulations.
- Comply with new and upcoming data privacy regulations.

[Learn more about the use cases for Cloud Data Sense.](#)

What types of data can be scanned with Cloud Data Sense?

Cloud Data Sense supports scanning of unstructured data over NFS and CIFS protocols that are managed by Cloud Volumes ONTAP, Azure NetApp Files, Amazon FSx for ONTAP, on-prem ONTAP systems, and in non-NetApp file shares. Data Sense supports NFS versions 3.x, 4.0, and 4.1, and CIFS versions 1.x, 2.0, 2.1, and 3.0.

Data Sense can also scan data stored on Amazon S3 buckets and in generic S3 object storage.

Additionally, Data Sense can scan databases that are located anywhere, and user files from OneDrive and SharePoint accounts.

[Learn how scans work.](#)

Which cloud providers are supported?

Cloud Data Sense operates as part of Cloud Manager and supports AWS, Azure, and GCP. This provides your organization with unified privacy visibility across different cloud providers.

How do I access Cloud Data Sense?

Cloud Data Sense is operated and managed through Cloud Manager. You can access Data Sense features from the **Data Sense** tab in Cloud Manager.

How does Cloud Data Sense work?

Cloud Data Sense deploys another layer of Artificial Intelligence alongside your Cloud Manager system and storage systems. It then scans the data on volumes, buckets, databases, and OneDrive accounts and indexes the data insights that are found.

[Learn more about how Cloud Data Sense works.](#)

How much does Cloud Data Sense cost?

The cost to use Cloud Data Sense depends on the amount of data that you're scanning. The first 1 TB of data that Data Sense scans in a Cloud Manager workspace is free. A subscription to the AWS, Azure, or GCP Marketplace, or a BYOL license from NetApp, is required to continue scanning data after that point. See [pricing](#) for details.

What type of instance or VM is required for Cloud Data Sense?

When [deployed in the cloud](#):

- In AWS, Cloud Data Sense runs on an m5.4xlarge instance with a 500 GB GP2 disk.
- In Azure, Cloud Data Sense runs on a Standard_D16s_v3 VM with a 512 GB disk.
- In GCP, Cloud Data Sense runs on an n2-standard-16 VM with a 512 GB Standard persistent disk.

You can install Data Sense software on a Linux host that has internet access in your network or in the cloud. Everything works the same and you continue to manage your scan configuration and results through Cloud Manager. See [Deploying Cloud Data Sense on premises](#) for system requirements and installation details.

Additionally, you can [deploy Data Sense in an on-premises site that doesn't have internet access](#) for completely secure sites.

Note that you can deploy Data Sense on a system with fewer CPUs and less RAM, but there are limitations when using these systems. See [Using a smaller instance type](#) for details.

[Learn more about how Cloud Data Sense works.](#)

How often does Cloud Data Sense scan my data?

Data changes frequently, so Cloud Data Sense scans your data continuously with no impact to your data. While the initial scan of your data might take longer, subsequent scans only scan the incremental changes, which reduces system scan times.

[Learn how scans work.](#)

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. [See how to reduce the scan speed.](#)

Does Cloud Data Sense offer reports?

Yes. The information offered by Cloud Data Sense can be relevant to other stakeholders in your organizations, so we enable you to generate reports to share the insights.

The following reports are available for Data Sense:

Privacy Risk Assessment report

Provides privacy insights from your data and a privacy risk score. [Learn more.](#)

Data Subject Access Request report

Enables you to extract a report of all files that contain information regarding a data subject's specific name or personal identifier. [Learn more.](#)

PCI DSS report

Helps you identify the distribution of credit card information across your files. [Learn more.](#)

HIPAA report

Helps you identify the distribution of health information across your files. [Learn more.](#)

Data Mapping report

Provides information about the size and number of files in your working environments. This includes usage capacity, age of data, size of data, and file types. [Learn more.](#)

Reports on a specific information type

Reports are available that include details about the identified files that contain personal data and sensitive personal data. You can also see files broken down by category and file type. [Learn more.](#)

Does scan performance vary?

Scan performance can vary based on the network bandwidth and the average file size in your environment. It can also depend on the size characteristics of the host system (either in the cloud or on-premises). See [The Cloud Data Sense instance](#) and [Deploying Cloud Data Sense](#) for more information.

When initially adding new data sources you can also choose to only perform a "mapping" scan instead of a full "classification" scan. Mapping can be done on your data sources very quickly because it does not access files to see the data inside. [See the difference between a mapping and classification scan.](#)

Which file types are supported?

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

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, and .XLSX.

How do I enable Cloud Data Sense?

First you need to deploy an instance of Cloud Data Sense in Cloud Manager. Once the instance is running, you can enable the service on existing working environments and databases from the **Data Sense** tab or by selecting a specific working environment.

[Learn how to get started.](#)



Activating Cloud Data Sense results in an immediate initial scan. Scan results display shortly after.

How do I disable Cloud Data Sense?

You can disable Cloud Data Sense from scanning an individual working environment, database, file share group, OneDrive account, or SharePoint account from the Data Sense Configuration page.

[Learn more.](#)



To completely remove the Cloud Data Sense instance, you can manually remove the Data Sense instance from your cloud provider's portal or on-prem location.

What happens if data tiering is enabled on your ONTAP volumes?

You might want to enable Cloud Data Sense on ONTAP systems that tier cold data to object storage. If data tiering is enabled, Data Sense scans all of the data—data that's on disks and cold data tiered to object storage.

The compliance scan doesn't heat up the cold data—it stays cold and tiered to object storage.

Can I use Cloud Data Sense to scan on-premises ONTAP storage?

Yes. As long as you have discovered the on-prem ONTAP cluster as a working environment in Cloud Manager, you can scan any of the volume data.

Alternatively, you can run compliance scans on backup files created from your on-prem ONTAP volumes. So if you're already creating backup files from your on-prem systems using [Cloud Backup](#), you can run compliance scans on those backup files.

[Learn more.](#)

Can Cloud Data Sense send notifications to my organization?

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

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

Can I customize the service to my organization's needs?

Cloud Data Sense provides out-of-the-box insights to your data. These insights can be extracted and used for your organization's needs.

Additionally, you can use the **Data Fusion** capability to have Data Sense scan all your data based on criteria found in specific columns in databases you are scanning — essentially allowing you to make your own custom personal data types.

[Learn more.](#)

Can Cloud Data Sense work with the AIP labels I have embedded in my files?

Yes. You can manage AIP labels in the files that Cloud Data Sense is scanning if you have subscribed to [Azure Information Protection \(AIP\)](#). You can view the labels that are already assigned to files, add labels to files, and change existing labels.

[Learn more.](#)

Can I limit Cloud Data Sense information to specific users?

Yes, Cloud Data Sense is fully integrated with Cloud Manager. Cloud Manager users can only see information for the working environments they are eligible to view according to their workspace privileges.

Additionally, if you want to allow certain users to just view Data Sense scan results without having the ability to manage Data Sense settings, you can assign those users the *Cloud Compliance Viewer* role.

[Learn more.](#)

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.