Manage Cloud Data Sense

Cloud Data Sense

NetApp
June 20, 2022

Table of Contents

Manage Cloud Data Sense

Adding personal data identifiers using Data Fusion

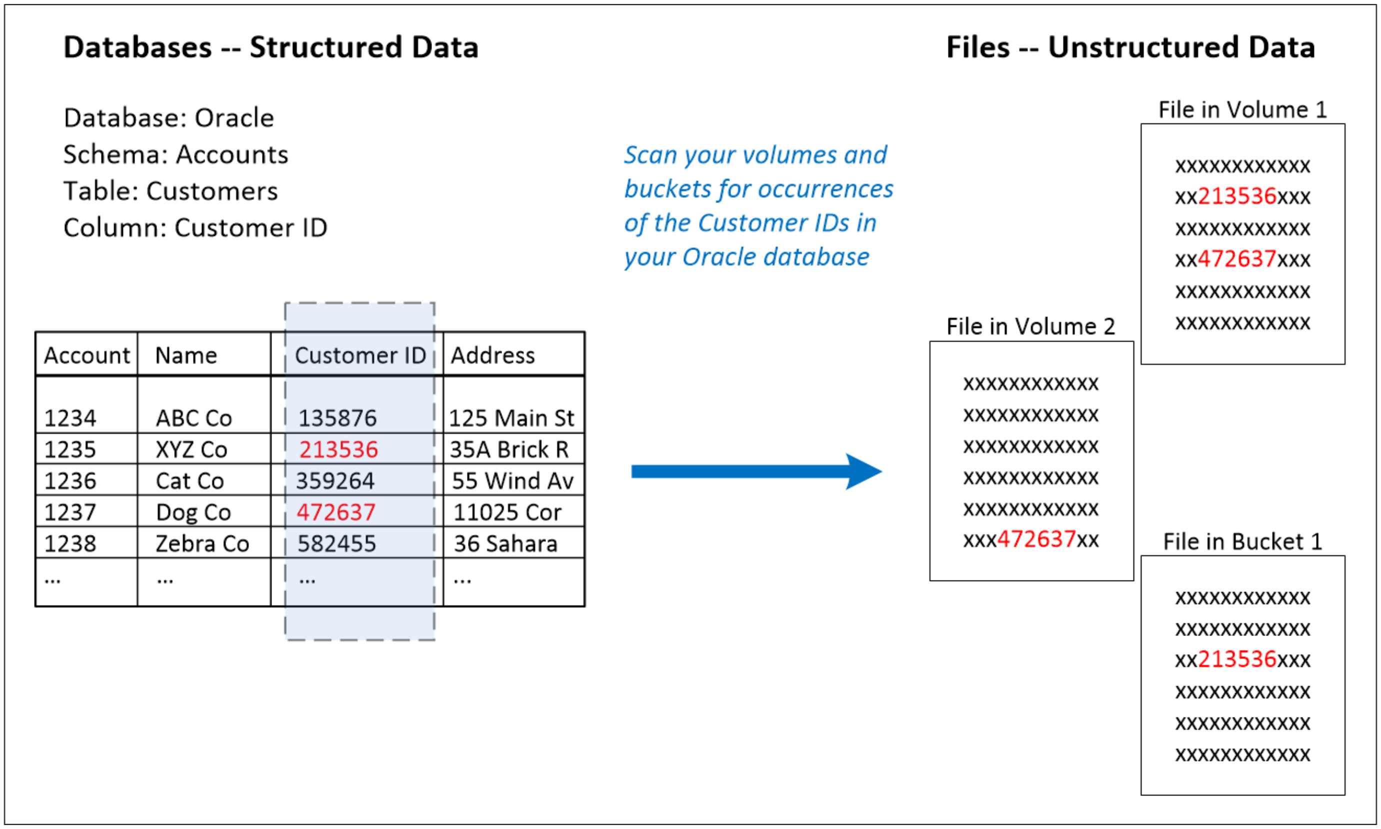
A feature we call *Data Fusion* allows you to scan your organizations' data to identify whether unique identifiers from your databases are found in files or other databases - basically making your own list of "personal data" that is identified in Cloud Data Sense scans. This gives you the full picture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



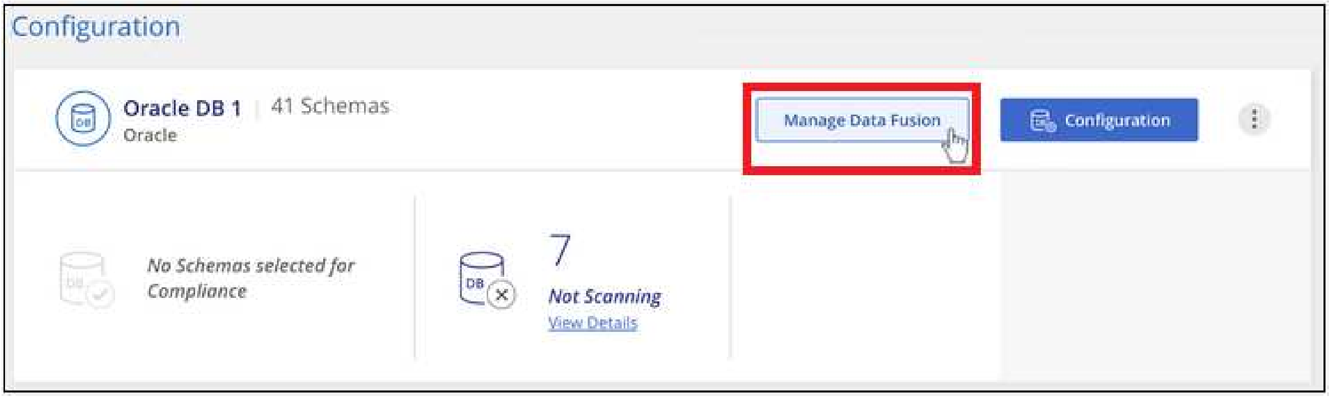
Account	Name	Customer ID	Address
1234	ABC Co	135876	125 Main St
1235	XYZ Co	213536	35A Brick R
1236	Cat Co	359264	55 Wind Av
1237	Dog Co	472637	11025 Cor
1238	Zebra Co	582455	36 Sahara
...	...	...	...

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.

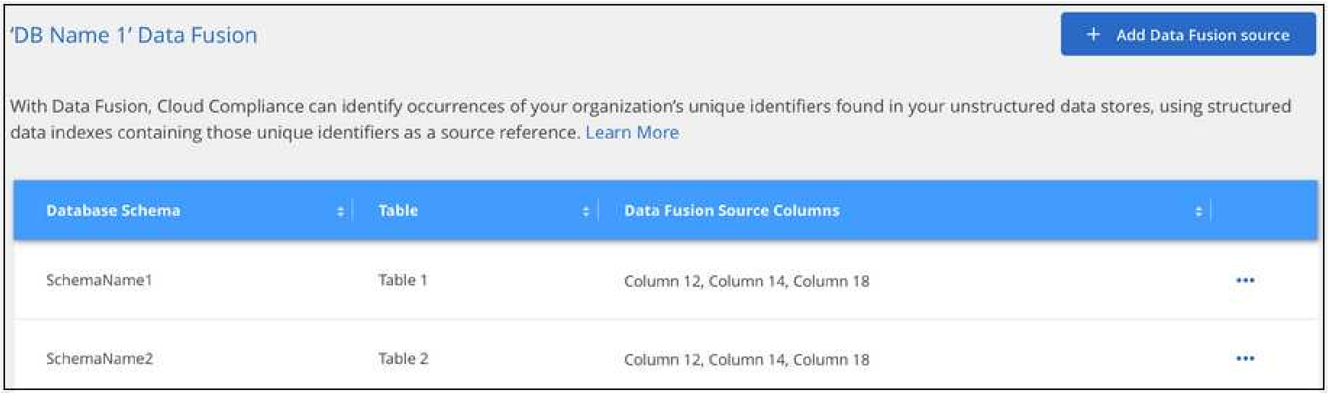


2. Click **Add Data Fusion source** on the next page.
3. In the *Add Data Fusion Source* page:
    a. Select the Database Schema from the drop-down menu.
    b. Enter the Table name in that schema.
    c. Enter the Column, or Columns, that contain the unique identifiers you want to use.

       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.



With Data Fusion, Cloud Compliance can identify occurrences of your organization's unique identifiers found in your unstructured data stores, using structured data indexes containing those unique identifiers as a source reference. Learn More

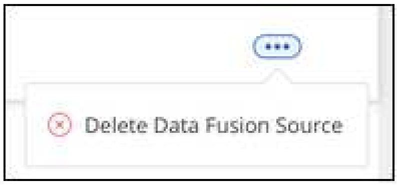
Database Schema	Table	Data Fusion Source Columns	
SchemaName1	Table 1	Column 12, Column 14, Column 18	•••
SchemaName2	Table 2	Column 12, Column 14, Column 18	•••

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 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.

Auditing the history of Data Sense actions

Data Sense logs management activities that have been performed on files from all the working environments and data sources that Data Sense is scanning. You can view the contents of the Data Sense audit log files, or download them, to see what file changes have occurred, and when.

For example, you can see what request was issued, the time of the request, and details such as source location in case a file was deleted, or source and destination location in case a file was moved.

Log file contents

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

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

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

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

```
2022-06-06 15:23:08,910 | INFO | datasense_audit_logger | es_scanned_file
| 237 | 49 | Copy file /idanCIFS_share/data/dop1/random_positives.tsv from
device 172.31.133.183 (type: SMB_SHARE) to device
172.31.130.133:/export_reports (NFS_SHARE) - SUCCESS
2022-06-06 15:23:08,968 | WARNING | datasense_audit_logger |
es_scanned_file | 239 | 153 | Copy file /idanCIFS_share/data/compliance-
netapp.tar.gz from device 172.31.133.183 (type: SMB_SHARE) to device
172.31.130.133:/export_reports (NFS_SHARE) - FAILURE
```

Accessing the log file

The audit log is located on the Data Sense machine in:
`/opt/netapp/audit_logs/<date>/DataSense_audit_log_<date>_<process_name>.log`

For on-premises deployments you can navigate directly to the log files.

When Data Sense is deployed in the cloud, you can SSH to the Data Sense instance. You can SSH to the system by entering the user and password, or by using the SSH key you provided during the Cloud Manager Connector installation. The SSH command is:

```
ssh -i <path to the ssh key> <machine user>@<datasense ip>
```

- <path to the ssh key> = location of ssh authentication keys
- <machine user>:
    - For AWS = use the <ec2-user>
    - For Azure: use the user created for the Cloud Manager instance
    - For GCP: use the user created for the Cloud Manager instance
- <data sense ip> = IP address of the virtual machine instance

Note that you’ll need to modify the security group inbound rules to access the system. See Ports and security groups for more information.

Reducing the Data Sense scan speed

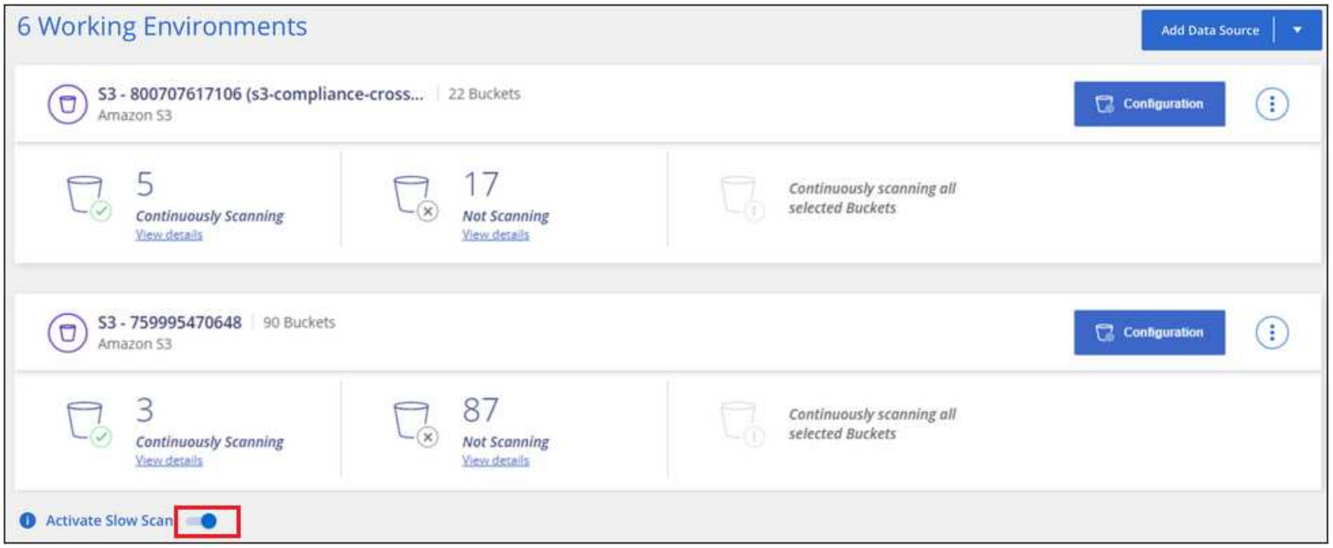
Data scans have a negligible impact on your storage systems and on your data. However, if you are concerned with even a very small impact, you can configure Data Sense to perform "slow" scans.

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

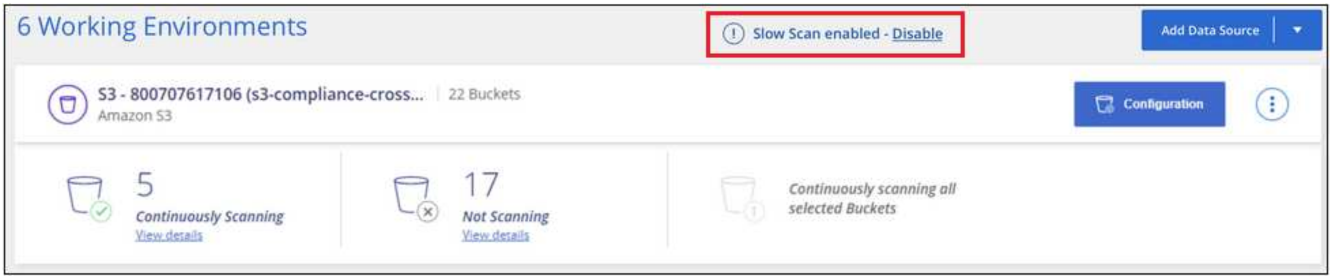
> The scan speed can’t be reduced when scanning databases.

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.

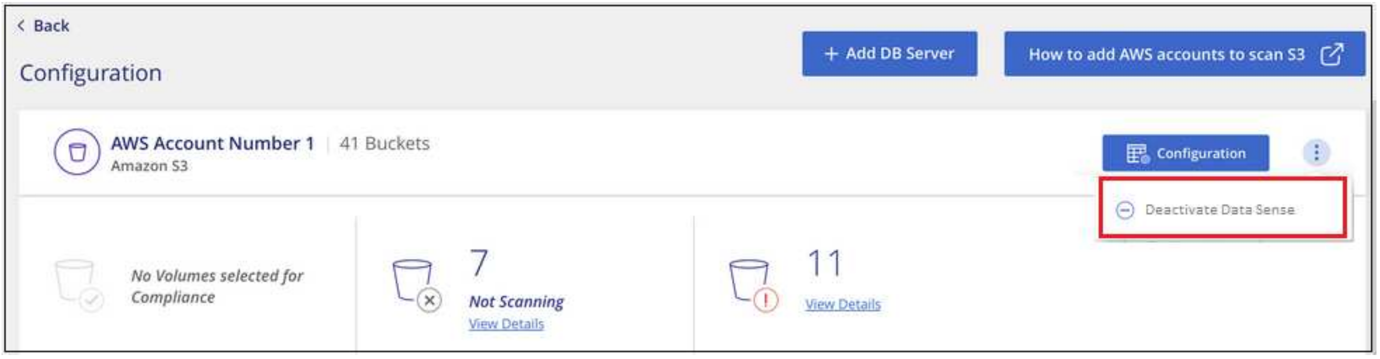
Removing data sources from Cloud Data Sense

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

Deactivating compliance scans for a working environment

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

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

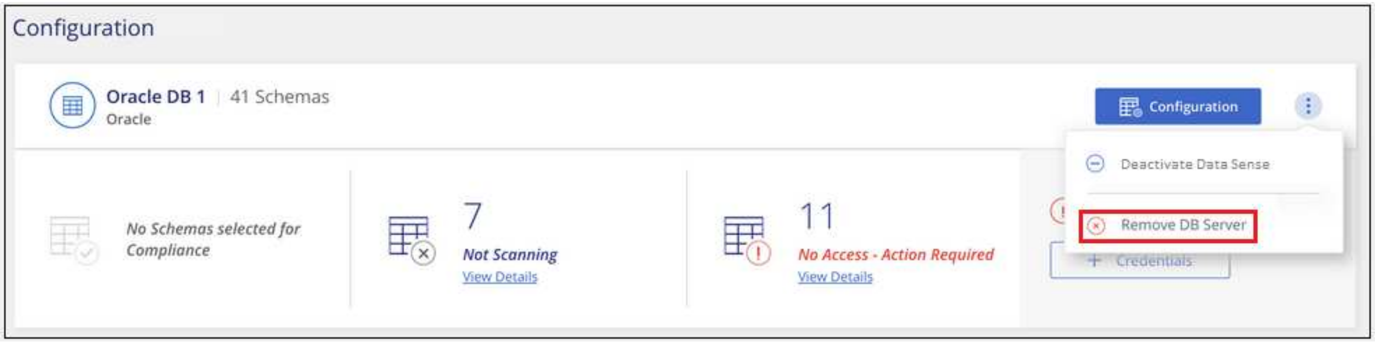


You can also disable compliance scans for a working environment from the Services panel when you select the working environment.

Removing a database from Cloud Data Sense

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

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

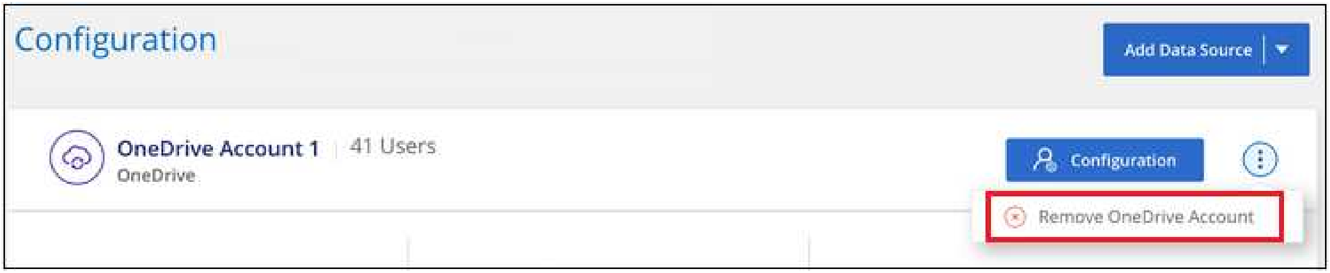


Removing a OneDrive, SharePoint, or Google Drive account from Cloud Data Sense

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

Steps

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



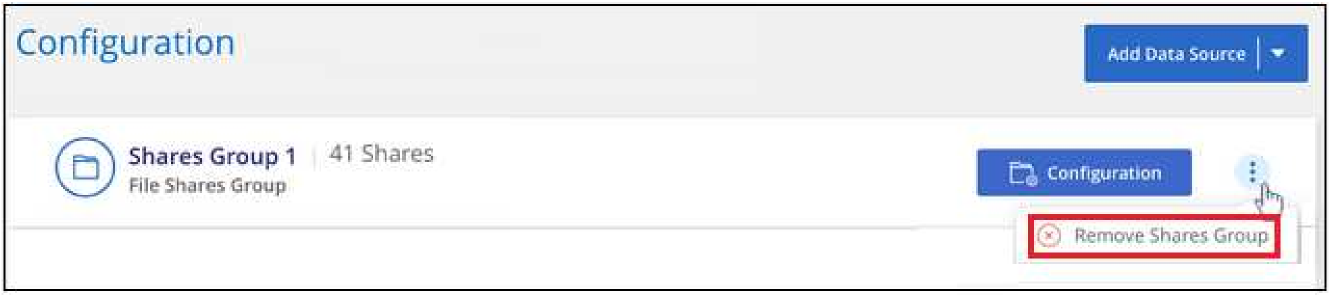
2. Click **Delete Account** from the confirmation dialog.

Removing a group of file shares from Cloud Data Sense

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

Steps

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



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

Uninstalling Cloud Data Sense

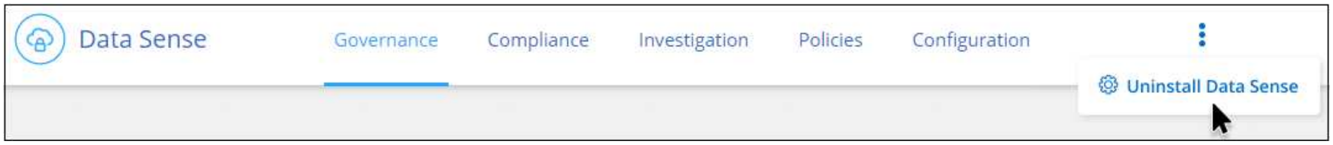
You can uninstall the Data Sense software to troubleshoot issues or to permanently remove the software from the host. Deleting the instance also deletes the associated disks where the indexed data resides - all the information Data Sense has scanned will be permanently deleted.

The steps that you need to use depend on whether you deployed Data Sense in the cloud or on an on-premises host.

Uninstall Data Sense from a cloud deployment

You can uninstall and delete the Cloud Data Sense instance from the cloud provider if you no longer want to use Data Sense.

1. At the top of the Data Sense page, click the ⋮ button and then click **Uninstall Data Sense**.



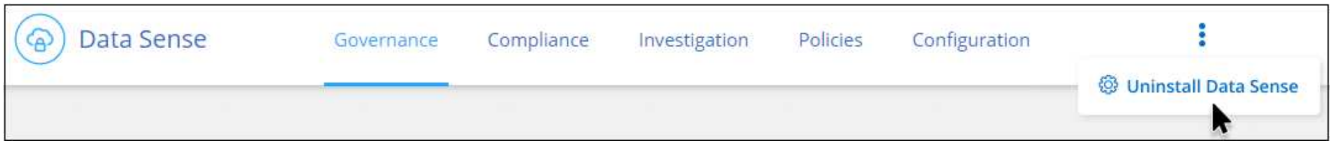
2. In the *Uninstall Data Sense* dialog, type **uninstall** to confirm that you want to delete the instance and all associated data, and then click **Uninstall**.

Note that you can go to your cloud provider's console and delete the Cloud Data Sense instance from there as well. The instance is named *CloudCompliance* with a generated hash (UUID) concatenated to it. For example: *CloudCompliance-16bb6564-38ad-4080-9a92-36f5fd2f71c7*

Uninstall Data Sense from an on-premises deployment

You can uninstall Data Sense from a host if you no longer want to use Data Sense, or if you had an issue that requires reinstallation.

1. At the top of the Data Sense page, click the ⋮ button and then click **Uninstall Data Sense**.



2. In the *Uninstall Data Sense* dialog, type **uninstall** to confirm that you want to clear all the configuration information, and then click **Uninstall**.
3. To complete the uninstallation from the host, run the uninstall script on the host machine, for example:

```
uninstall.sh
```