Manage Cloud Data Sense

Cloud Data Sense

NetApp
February 07, 2023

Table of Contents

Manage Cloud Data Sense 1
Adding personal data identifiers to your Data Sense scans 1
Viewing the status of your compliance actions 11
Auditing the history of Data Sense actions 11
Reducing the Data Sense scan speed 13
Removing data sources from Cloud Data Sense 14
Uninstalling Cloud Data Sense 16

Manage Cloud Data Sense

Adding personal data identifiers to your Data Sense scans

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

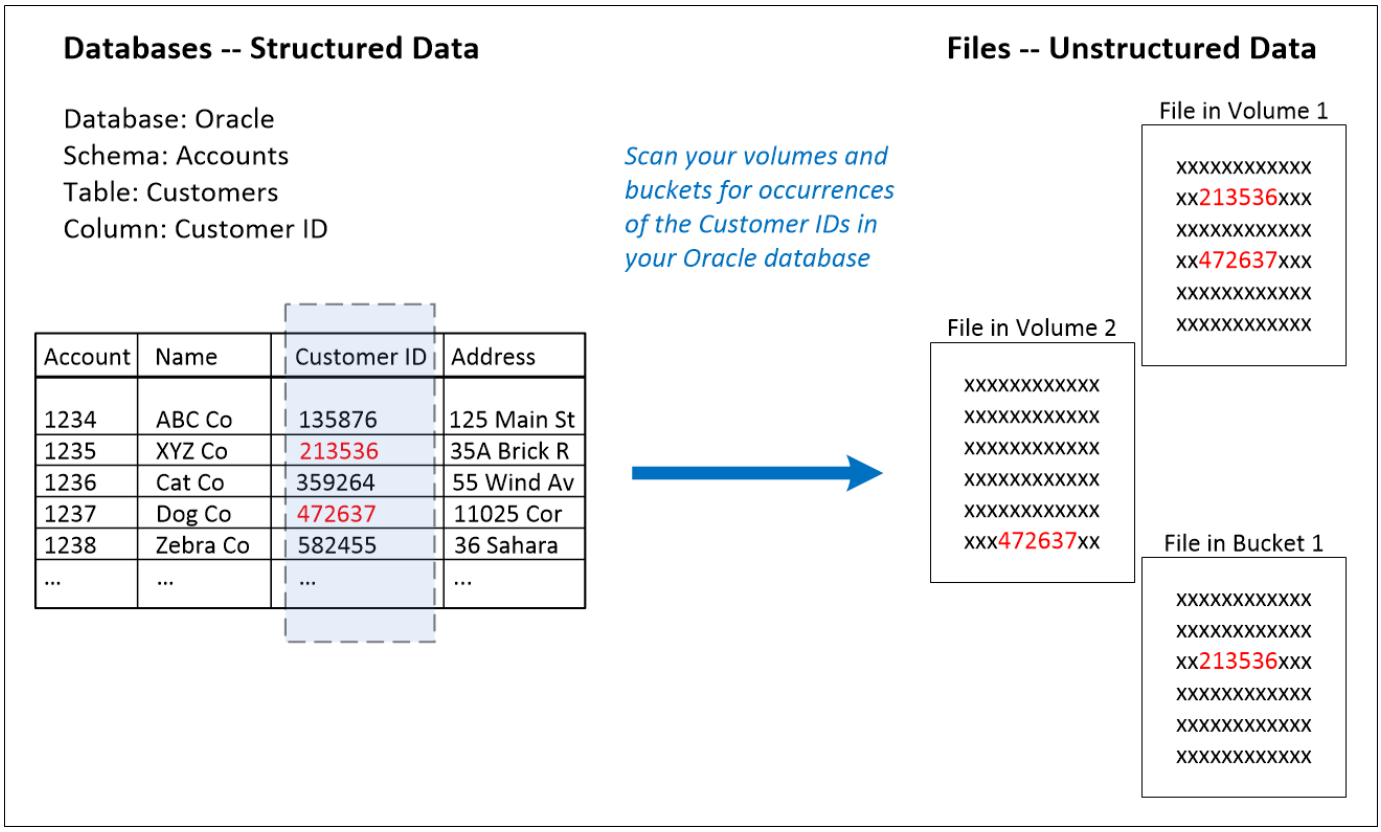
- You can add unique identifiers based on specific columns in databases you are scanning.
- You can add custom keywords from a text file — these words are identified within your data.
- You can add a personal pattern using a regular expression (regex) — the regex is added to the existing predefined patterns.

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

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

Add custom personal data identifiers from your databases

A feature we call *Data Fusion* allows you to scan your organizations' data to identify whether unique identifiers from your databases are found in any of your other data sources. You can choose the additional identifiers that 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.



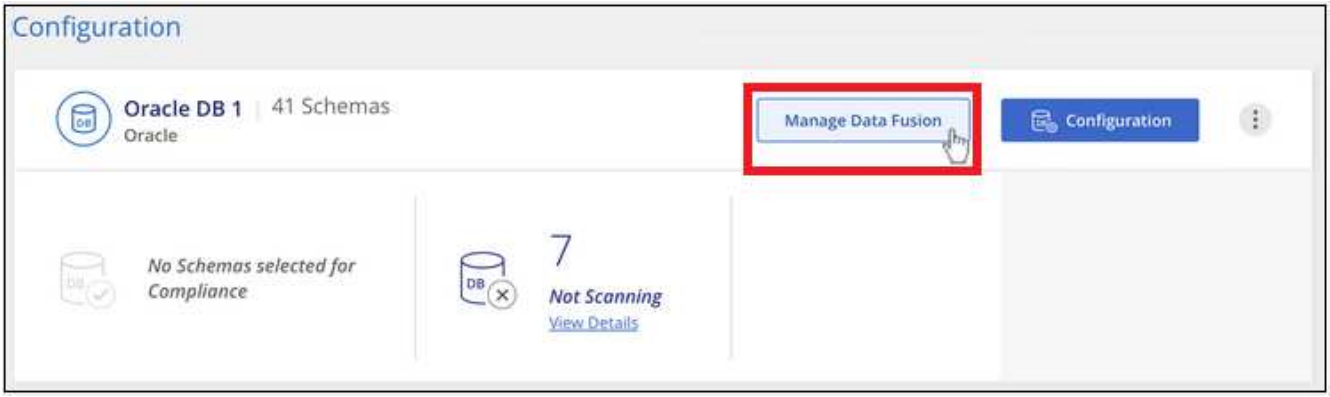
As you can see, two unique Customer IDs have been found in two volumes and in one S3 bucket. Any matches in database tables will also be identified.

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

Steps

You must have added at least one database server to 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.

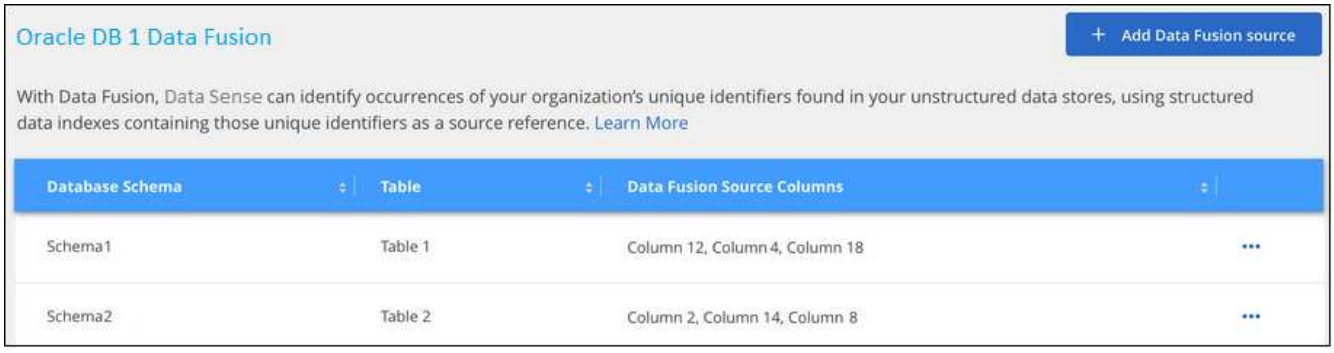


2. Click **Add Data Fusion source** on the next page.
3. In the *Add Data Fusion Source* page:
   a. Select the Database Schema from the drop-down menu.
   b. Enter the Table name in that schema.
   c. Enter the Column, or Columns, that contain the unique identifiers you want to use.

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

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

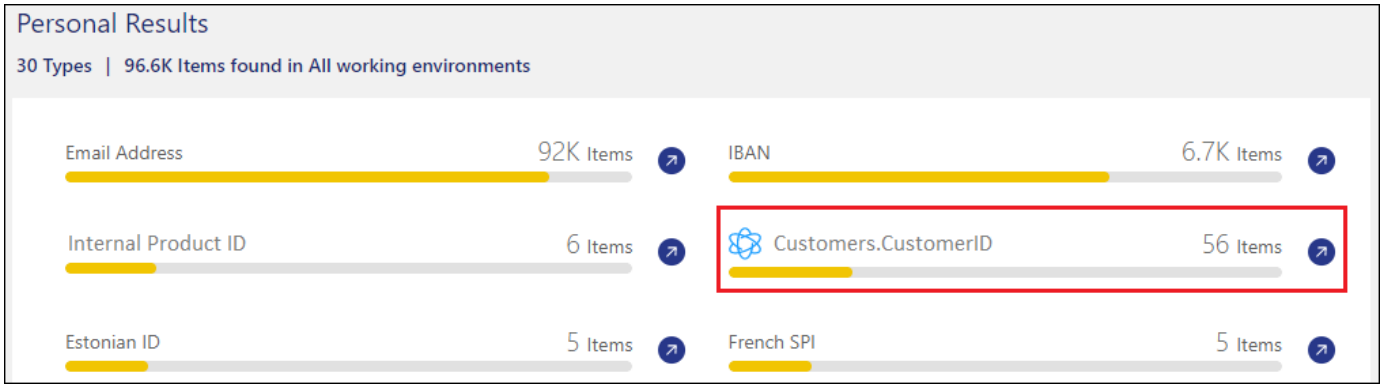
   The Data Fusion inventory page displays the database source columns that you have configured Data Sense to scan.



Results

After the next scan, the results will include this new information in the Compliance Dashboard under the

"Personal Results" section, and in the Investigation page in the "Personal Data" filter. Each source column you added appears in the filter list in the format "Table.Column", for example `Customers.CustomerID`.



Delete a Data Fusion source

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

Add custom personal data identifiers using a regex

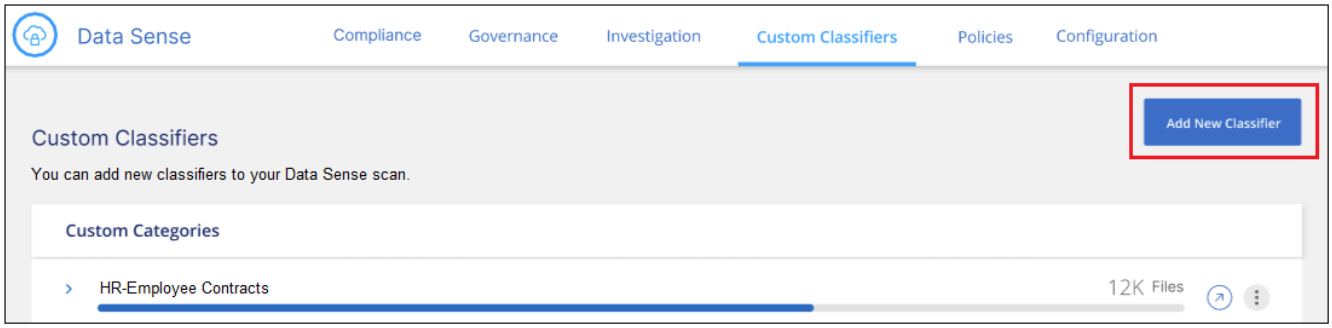
You can add a personal pattern to identify specific information in your data using a custom regular expression (regex). This allows you to create a new custom regex to identify new personal information elements that don't yet exist in the system. The regex is added to the existing predefined patterns that Data Sense already uses, and the results will be visible under the personal patterns section.

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

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

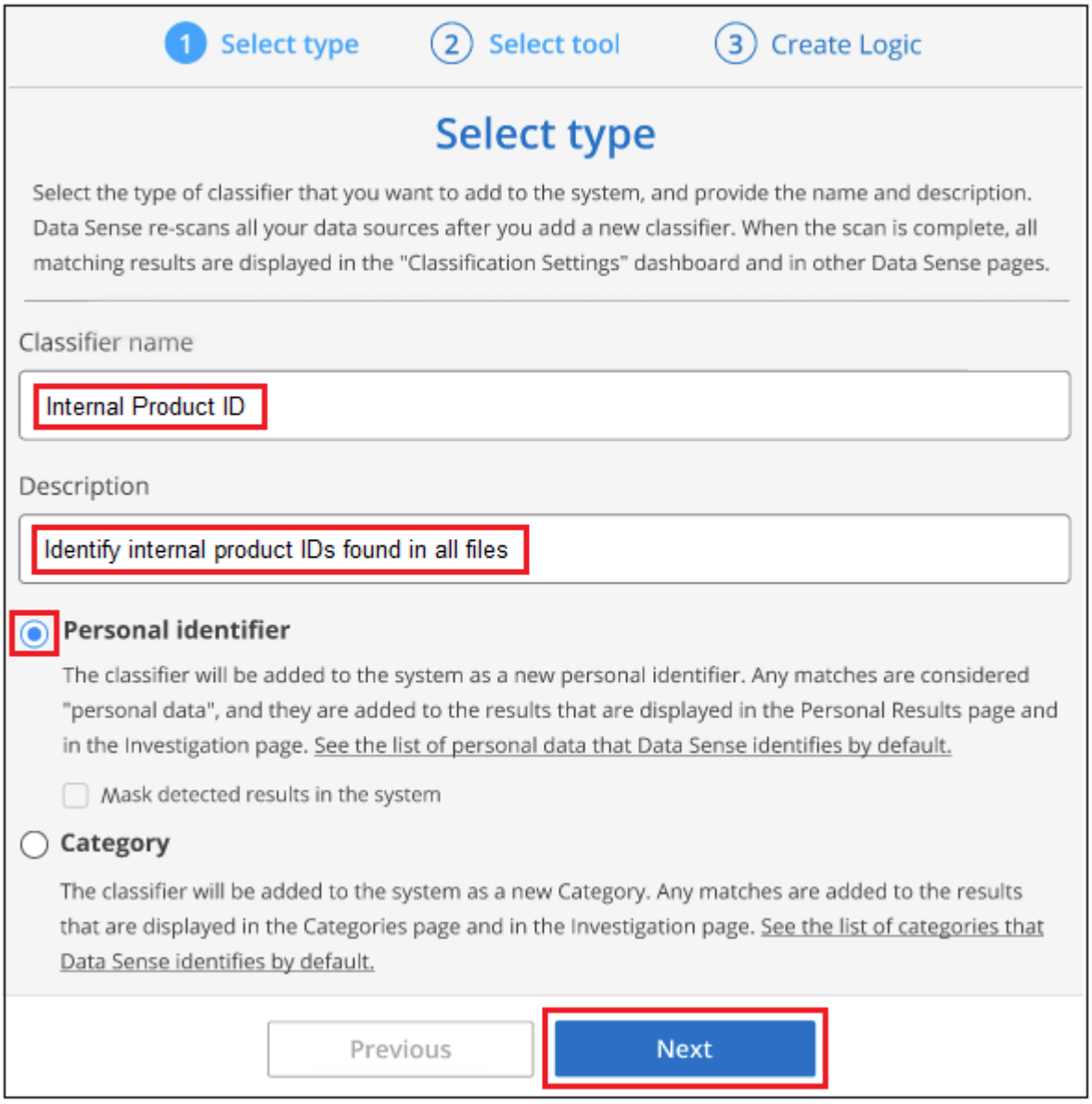
Steps

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

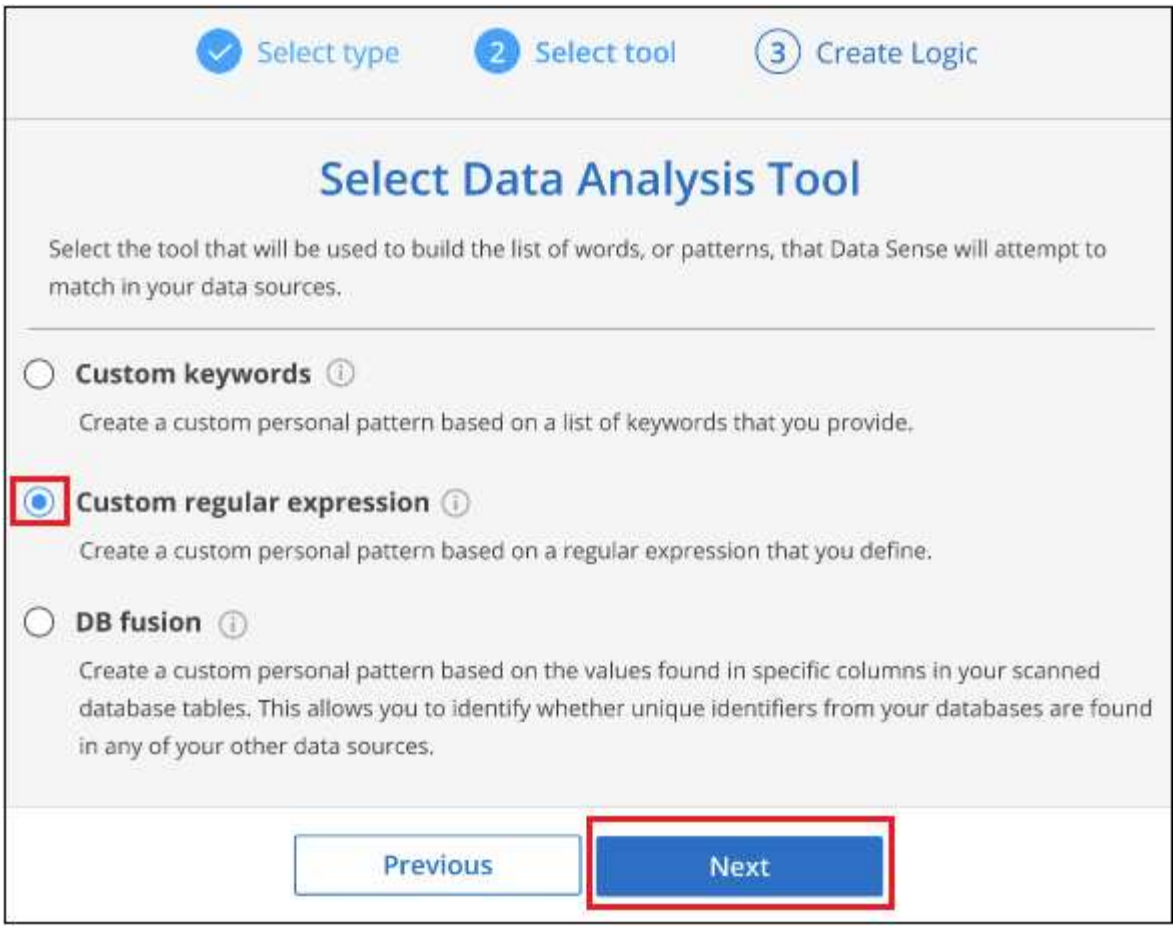


2. In the *Select type* page, enter the name of the classifier, provide a brief description, select **Personal identifier**, and then click **Next**.

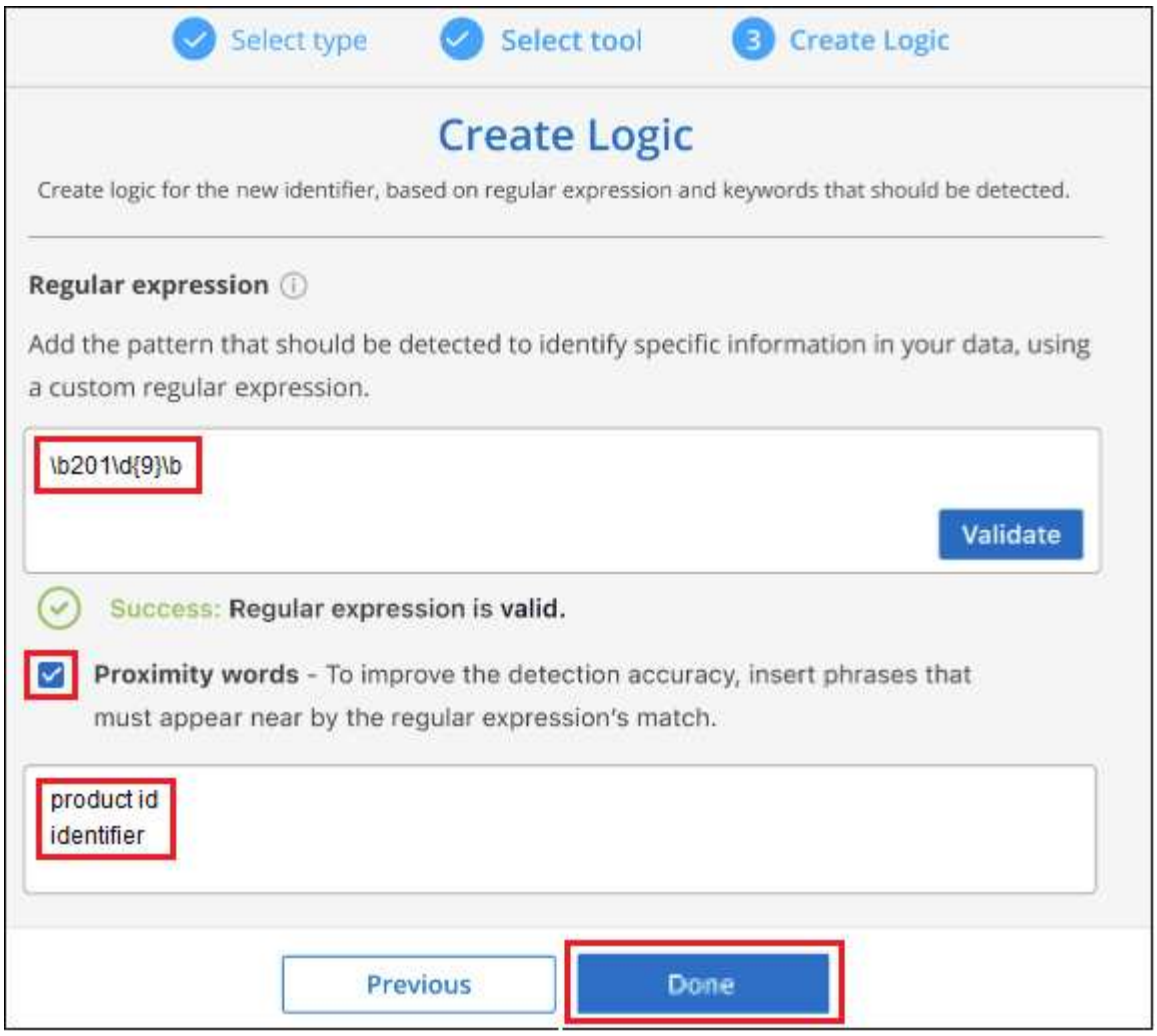
   The name you enter will appear in the Data Sense UI as the heading for scanned files that match the classifier requirements. You can also check the box to "Mask detected results in the system" so the full result won't appear in the UI. For example, you may want to do this to hide full credit card numbers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



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

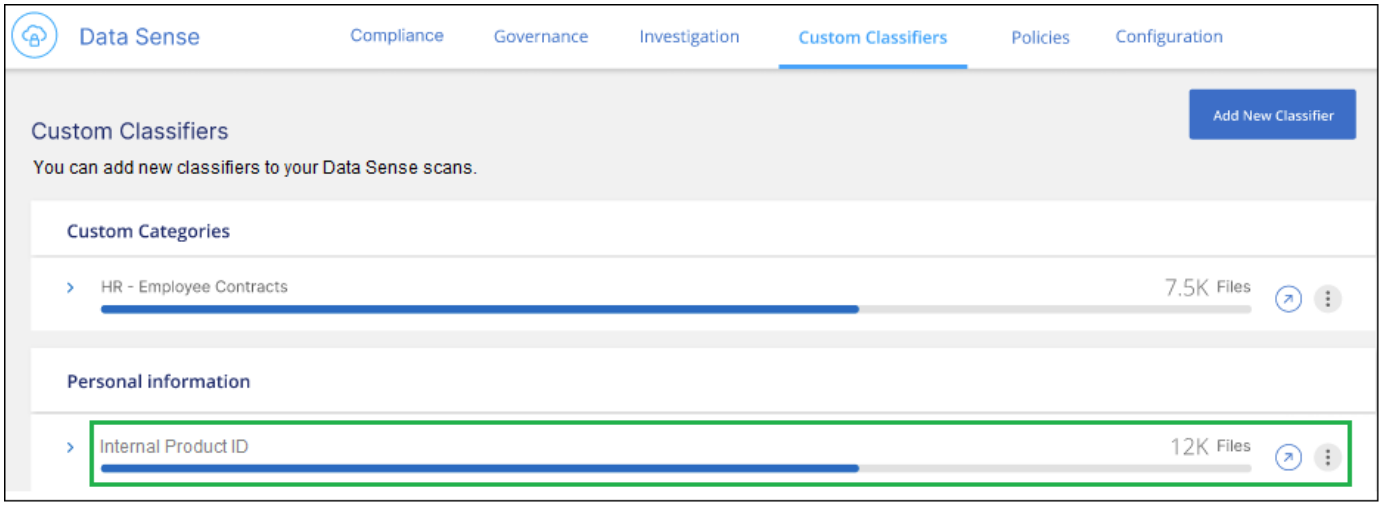


4. In the *Create Logic* page, enter the regular expression and any proximity words, and click **Done**.
   a. You can enter any legal regular expression. Click the **Validate** button to have Data Sense verify that the regular expression is valid, and that it is not too broad — meaning it will return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



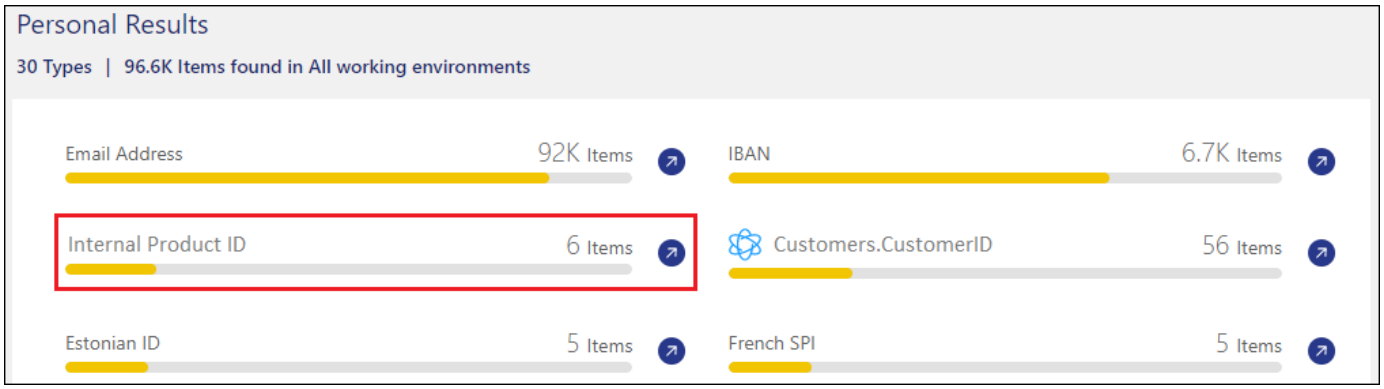
Result

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



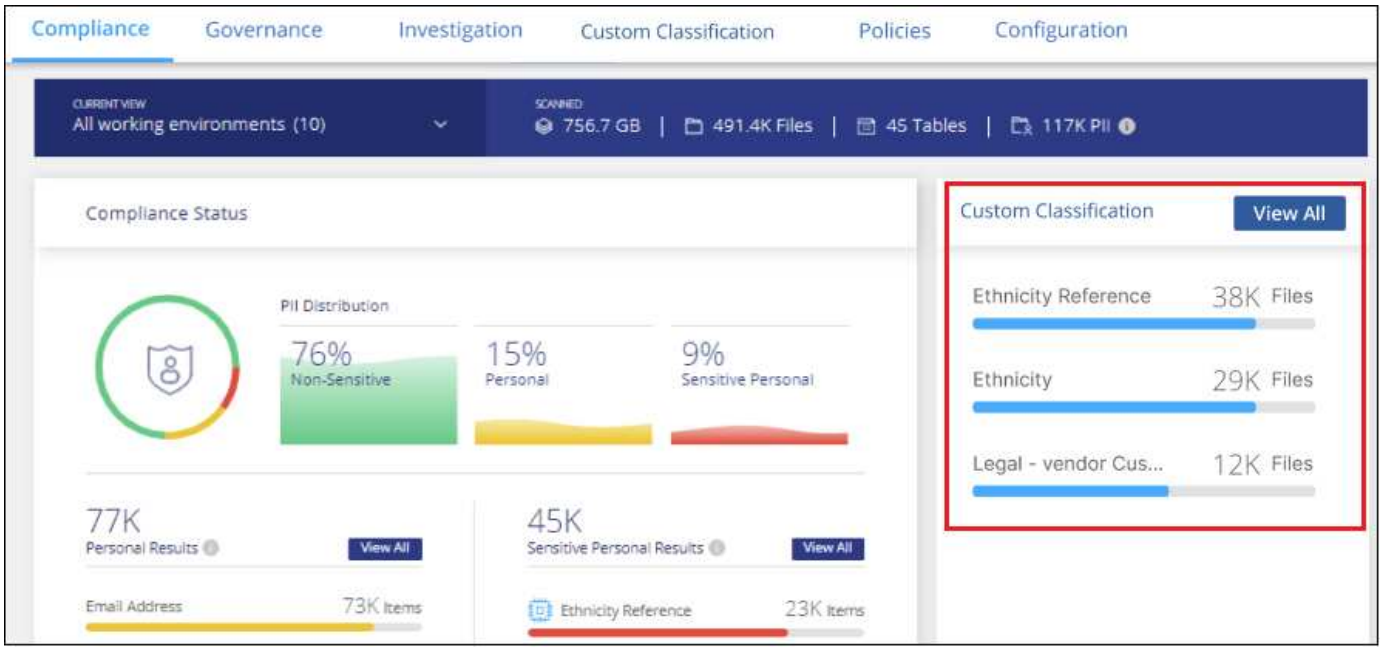
View results from your custom classifiers

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



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

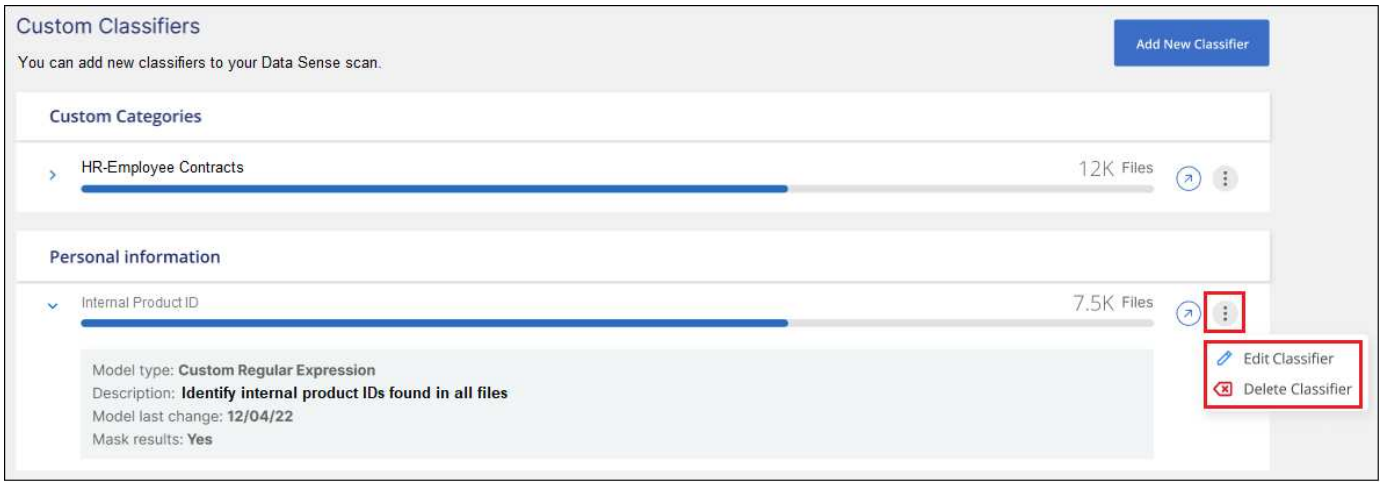
Additionally, all of your custom classifier results appear in the Custom Classifiers tab, and the top 6 custom classifier results are displayed in the Compliance Dashboard, as shown below.



Manage custom classifiers

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

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



Add custom keywords from a text file

You can add custom keywords to Data Sense so that it will identify specific information in your data. You add the keywords from a text file that you define. The keywords are added to the existing predefined keywords that Data Sense already uses, and the results will be visible under the personal patterns section.

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

After updating the custom keywords, Data Sense will restart scanning all data sources - the new results will appear in Data Sense after the scan has completed.

You must add, or create, the text files that include the custom keywords in the following location on the Data Sense system:

```
/opt/netapp/Datasense/tools/datascience/custom_keywords/keywords_sets
```

You can create a single file with multiple keywords, or you can add many files that each contain certain keywords. The format for the file is one word on each line, for example, internal Product Names which are types of owls are listed below:

internal_product_names.txt

```
barred
barn
horned
snowy
screech
```

The Data Sense search for these items is not case sensitive.

Note the following requirements:

- The file name should not contain digits.
- Each file can contain a maximum of 100,000 words. If there are more words, only the first 100,000 are

added.

- Each word must be at least 3 characters long. Shorter words are ignored.
- Duplicate words are only added once.

Access the command line

You'll need to access the Data Sense system in order to initiate the command to add custom keywords.

When Data Sense is installed on your premises, you can access the command line directly.

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

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

- <path_to_the_ssh_key> = location of ssh authentication keys
- <machine_user>:
  - For AWS: use the <ec2-user>
  - For Azure: use the user created for the BlueXP instance
  - For GCP: use the user created for the BlueXP instance
- <datasense_ip> = IP address of the virtual machine instance

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

- Security group rules in AWS
- Security group rules in Azure
- Firewall rules in Google Cloud

Command syntax to add custom keywords

The command syntax to add custom keywords from a file is:

```
sudo bash tools/datascience/custom_keywords/upload_custom_keywords.sh -s
activate -f <file_name>.txt
```

- <file_name> = This is the name of the file that contains the keywords.

You run the command from the path **/opt/netapp/Datasense/**.

If you have created many files that contain custom keywords, you can add the keywords from all the files at once using this command:

```
sudo bash tools/datascience/custom_keywords/upload_custom_keywords.sh -s
activate
```

Example

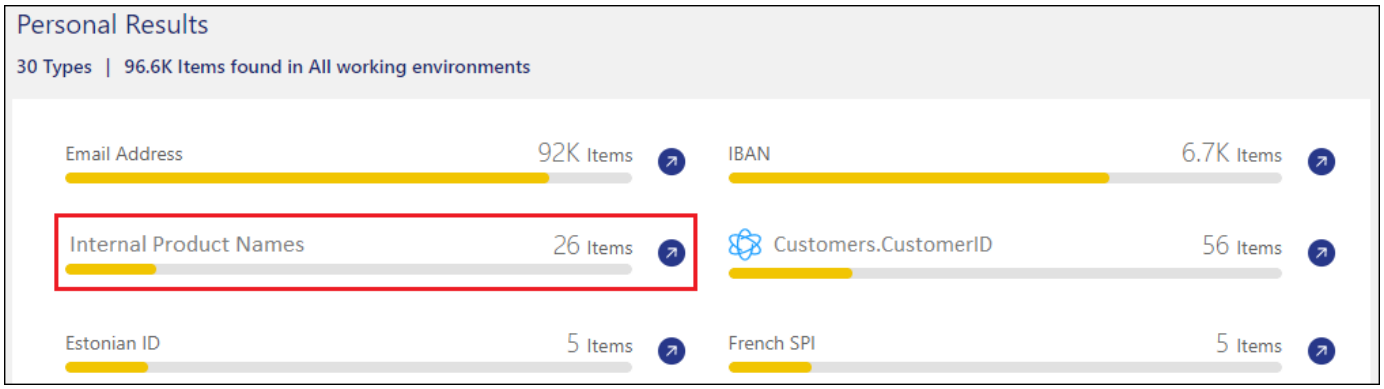
To see where your internal Product Names are mentioned in all of your files, enter the following command.

```
[user ~]$ cd /opt/netapp/Datasense/
[user Datasense]$ sudo bash
tools/datascience/custom_keywords/upload_custom_keywords.sh -s activate -f
internal_product_names.txt
```

```
log v1.0 | 2022-08-24 08:16:25,332 | INFO | ds_logger |
upload_custom_keywords | 126 | 1 | None | upload_custom_keywords_126 | All
legal keywords were successfully inserted
```

Results

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



As you can see, the name of the text file is used as the name in the Personal Results panel. In this manner you can activate keywords from different text files and see the results for each type of keyword.

Deactivate custom keywords

If you decide at some later point that you don't need Data Sense to identify certain custom keywords that you previously added, use the **deactivate** option in the command to remove the keywords that are defined in the text file.

```
sudo bash tools/datascience/custom_keywords/upload_custom_keywords.sh -s
deactivate -f <file_name>.txt
```

For example, to remove the keywords defined in the file **internal_product_names.txt**:

```
[user ~]$ cd /opt/netapp/Datasense/
[user Datasense]$ sudo bash
tools/datascience/custom_keywords/upload_custom_keywords.sh -s deactivate
-f internal_product_names.txt
```

```
log v1.0 | 2022-08-24 08:16:25,332 | INFO | ds_logger |
upload_custom_keywords | 87 | 1 | None | upload_custom_keywords_87 |
Deactivated keyword pattern from internal_product_names.txt successfully
```

Viewing the status of your compliance actions

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

This allows you to see the actions that have completed successfully, those currently in progress, and those that have failed so you can diagnose and fix any problems.

The status can be:

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

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

Steps

1. In the bottom-right of the 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> | ds_audit_logger | <module> | 0 | 0 | File <full file path> deleted from device <device path> - <result>
```

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

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

```
2022-06-06 15:23:08,910 | INFO | ds_audit_logger | es_scanned_file | 237 | 49 | Copy file /CIFS_share/data/dop1/random_positives.tsv from device 10.31.133.183 (type: SMB_SHARE) to device 10.31.130.133:/export_reports (NFS_SHARE) - SUCCESS
2022-06-06 15:23:08,968 | WARNING | ds_audit_logger | es_scanned_file | 239 | 153 | Copy file /CIFS_share/data/compliance-netapp.tar.gz from device 10.31.133.183 (type: SMB_SHARE) to device 10.31.130.133:/export_reports (NFS_SHARE) - FAILURE
```

Accessing the log file

The audit log files are located on the Data Sense machine in: `/opt/netapp/audit_logs/`

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

When Data Sense is installed on your premises, you can navigate directly to the log file.

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

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

- <path_to_the_ssh_key> = location of ssh authentication keys
- <machine_user>:
  - For AWS: use the <ec2-user>
  - For Azure: use the user created for the BlueXP instance
  - For GCP: use the user created for the BlueXP instance
- <datasense_ip> = IP address of the virtual machine instance

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

- Security group rules in AWS
- Security group rules in Azure
- Firewall rules in Google Cloud

Reducing the Data Sense scan speed

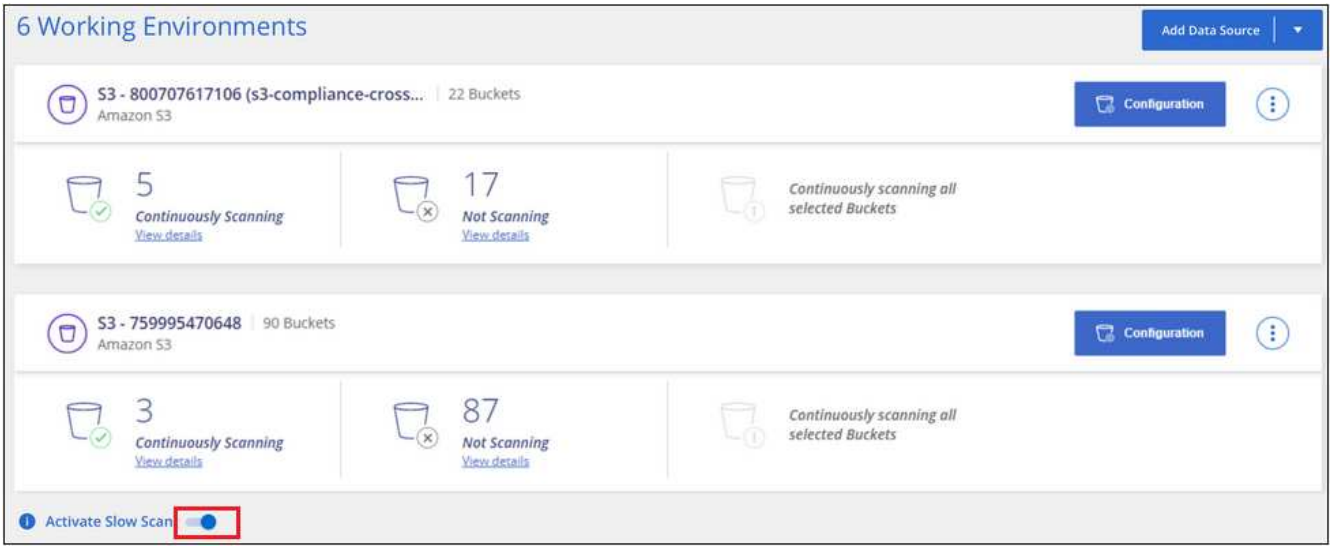
Data scans have a negligible impact on your storage systems and on your data. However, if you are concerned with even a very small impact, you can configure Data Sense to perform "slow" scans.

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

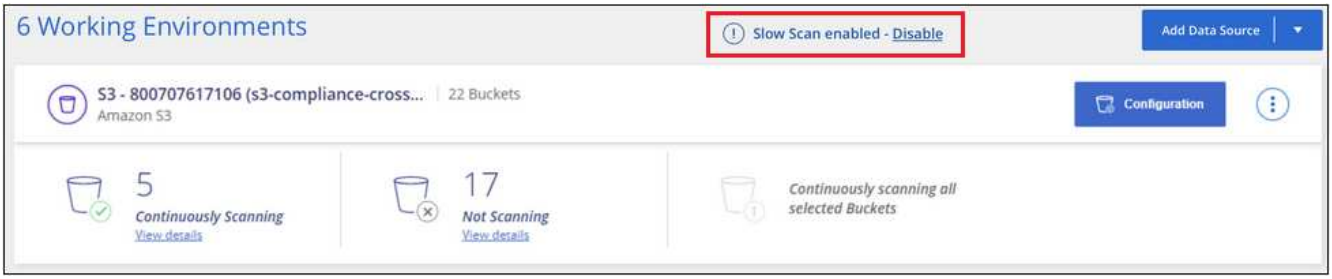
The scan speed can't be reduced when scanning databases.

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.

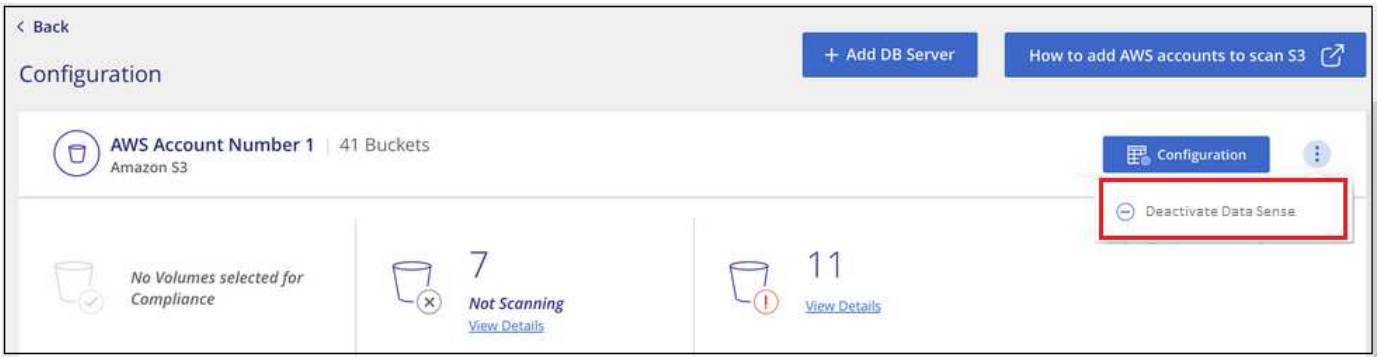
Removing data sources from Cloud Data Sense

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 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 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
```

Copyright information

Trademark information