



Viewing compliance details about the data stored in your organization

Cloud Manager

Tom Onacki
July 19, 2021

Table of Contents

- Viewing compliance details about the data stored in your organization. 1
 - Viewing files that contain personal data 1
 - Viewing files that contain sensitive personal data 3
 - Viewing files by categories 4
 - Viewing files by file types 5
 - Viewing file metadata 5
 - Viewing permissions for files 6
 - Checking for duplicate files in your storage systems 7
 - Viewing Dashboard data for specific working environments 8
 - Filtering data in the Data Investigation page 9
 - What's included in each file list report (CSV file) 11

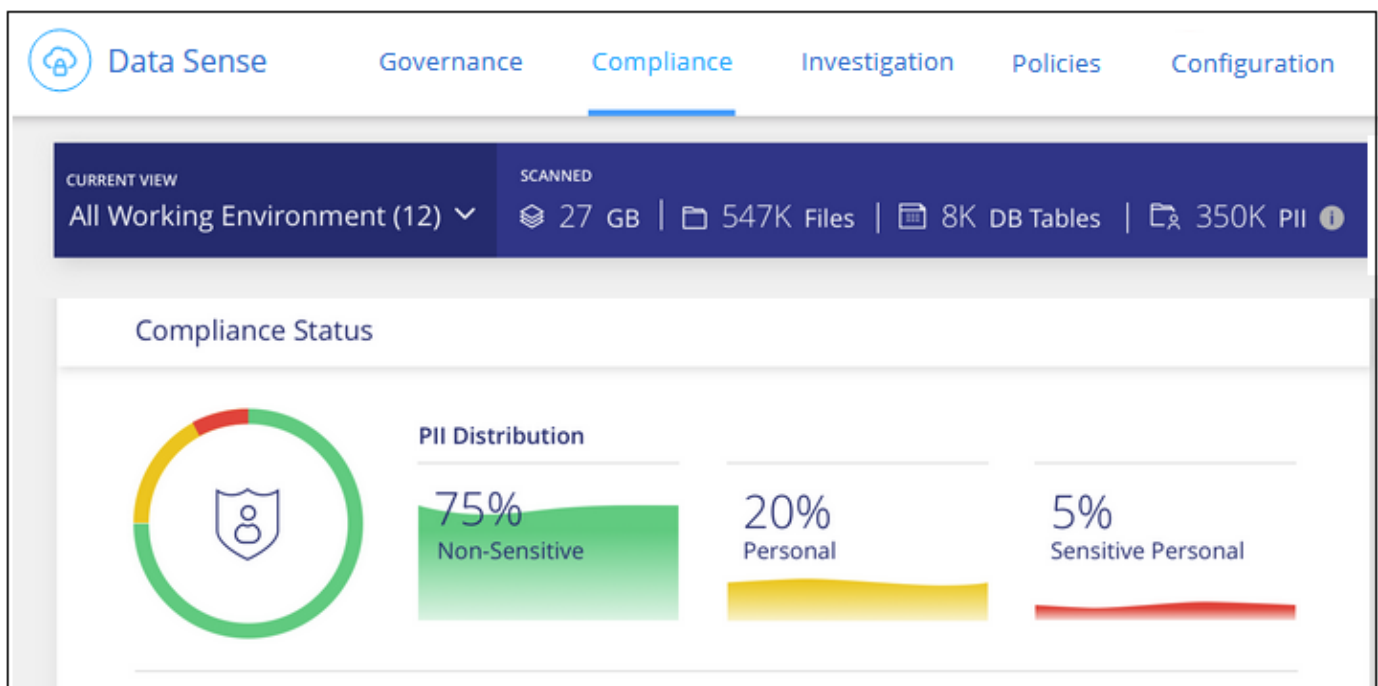
Viewing compliance details about the data stored in your organization

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

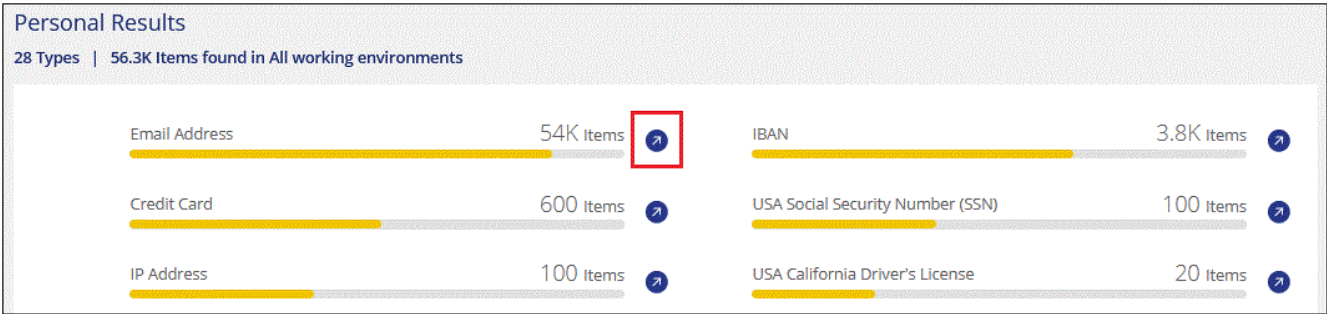
occurs by looking for one or more predefined keywords in proximity to the personal data that was found. For example, Data Sense identifies a U.S. social security number (SSN) as a SSN if it sees a proximity word next to it—for example, *SSN* or *social security*. [The table of personal data](#) shows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

Steps

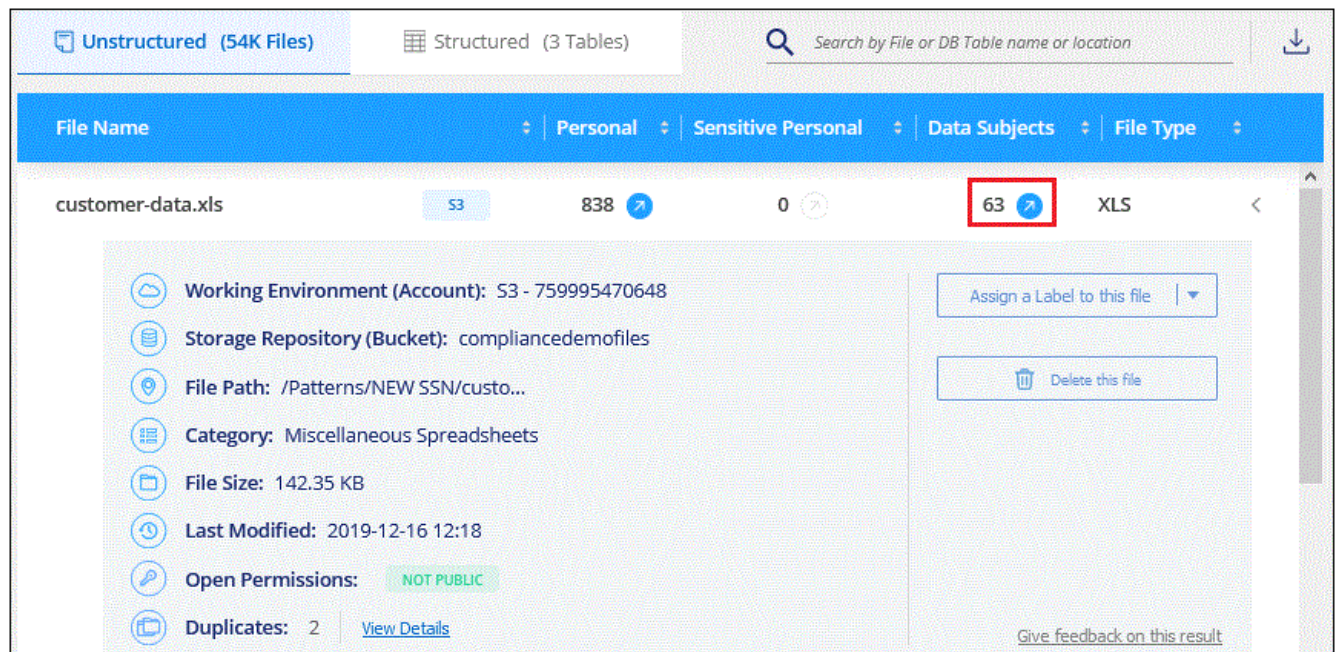
- 1. At the top of Cloud Manager, click **Data Sense** and click the **Compliance** tab.
- 2. To investigate the details for all personal data, click the icon next to the personal data percentage.



- 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.

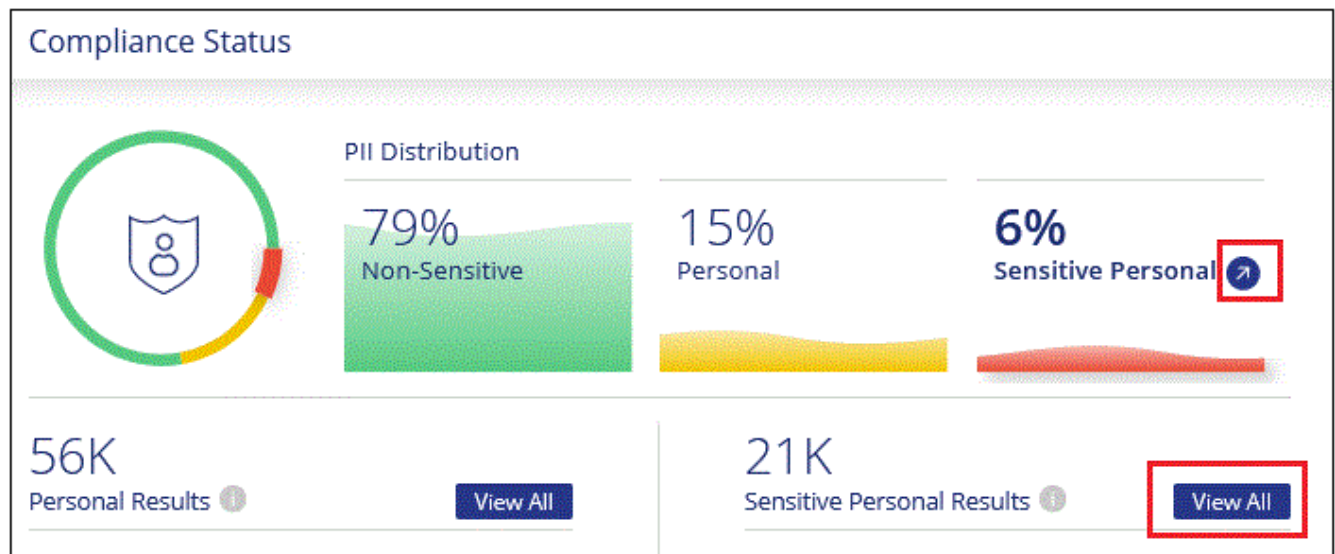
For example, one sensitive GDPR data category is ethnic origin. Because of its NLP abilities, Data Sense can distinguish the difference between a sentence that reads "George is Mexican" (indicating sensitive data as specified in article 9 of the GDPR), versus "George is eating Mexican food."



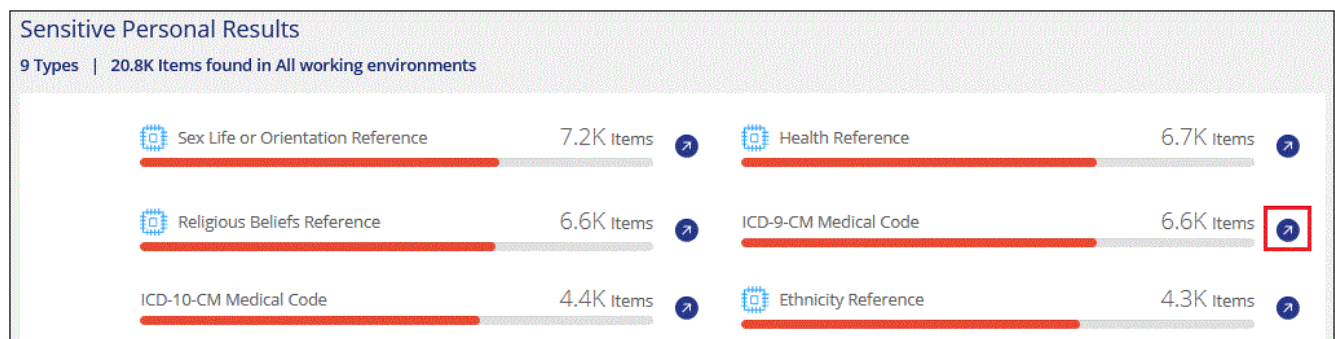
Only English is supported when scanning for sensitive personal data. Support for more languages will be added later.

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.](#)

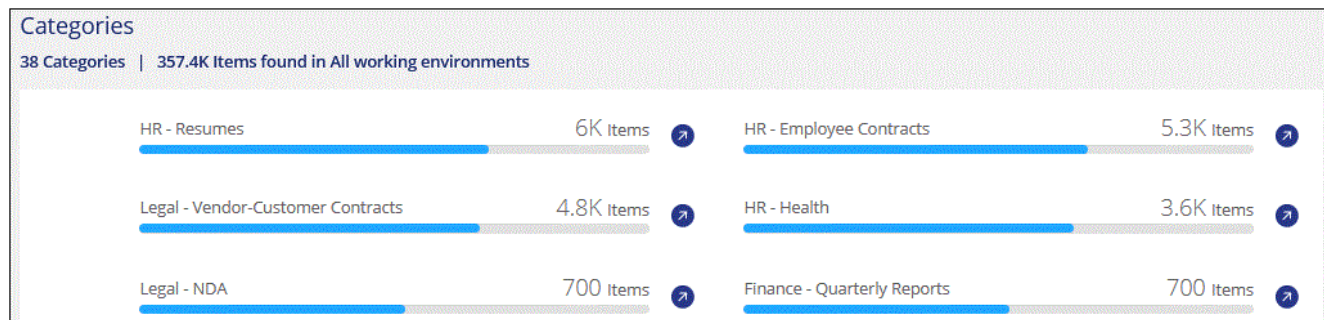
Categories can help you understand what's happening with your data by showing you the types of information that you have. For example, a category like resumes or employee contracts can include sensitive data. When you investigate the results, you might find that employee contracts are stored in an insecure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



Only English is 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.

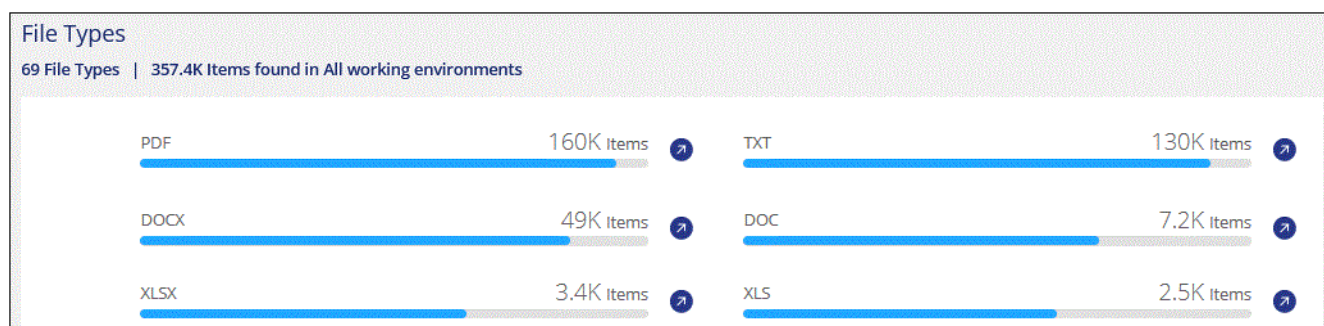
Viewing files by file types

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](#)

Status: To Check

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 the file to any NFS share. See [Moving 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**.

The screenshot shows a file management interface. At the top, there's a blue header with filters: File Name, Personal, Sensitive Personal, Data Subjects, and File Type. Below the header, the file 'Expense Report TPO-1060.pdf' is selected, showing its category (cvo), size (6), and other metadata. The file details pane on the left lists: Working Environment: WorkingEnvironment1, Repository: Volume Name, File Path: /Prod/labs-base/Expense Report TPO-1060.pdf, Category: Legal, File Size: 22 MB, Last Modified: 2019-08-06 07:51, Open Permissions: NO OPEN PERMISSIONS, and File Owner: Avy. A red box highlights the 'View all Permissions' button. To the right, a modal titled 'Permissions list for file Expense Report TPO-1060.pdf' is open, displaying a table of permissions.

Group or User	Read	Write
user1@company.com	✓	✗
user2@company.com	✓	✓
dist_list_IT@company.com	✓	✗
user4@company.com	✓	✓

This button is available only for files in CIFS shares.

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: xxxxxxxx

[View Duplicates](#)

[Close](#)

3 items

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


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.

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.

Steps




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.

- 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
- Storage repository
- Protocol type
- File path
- File type
- Created time
- Last modified
- Last accessed
- File size
- 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
- Storage repository
- Column count
- Row count
- Personal information

- Sensitive personal information

Copyright Information

Copyright © 2021 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.