TECHNICAL WHITE PAPER

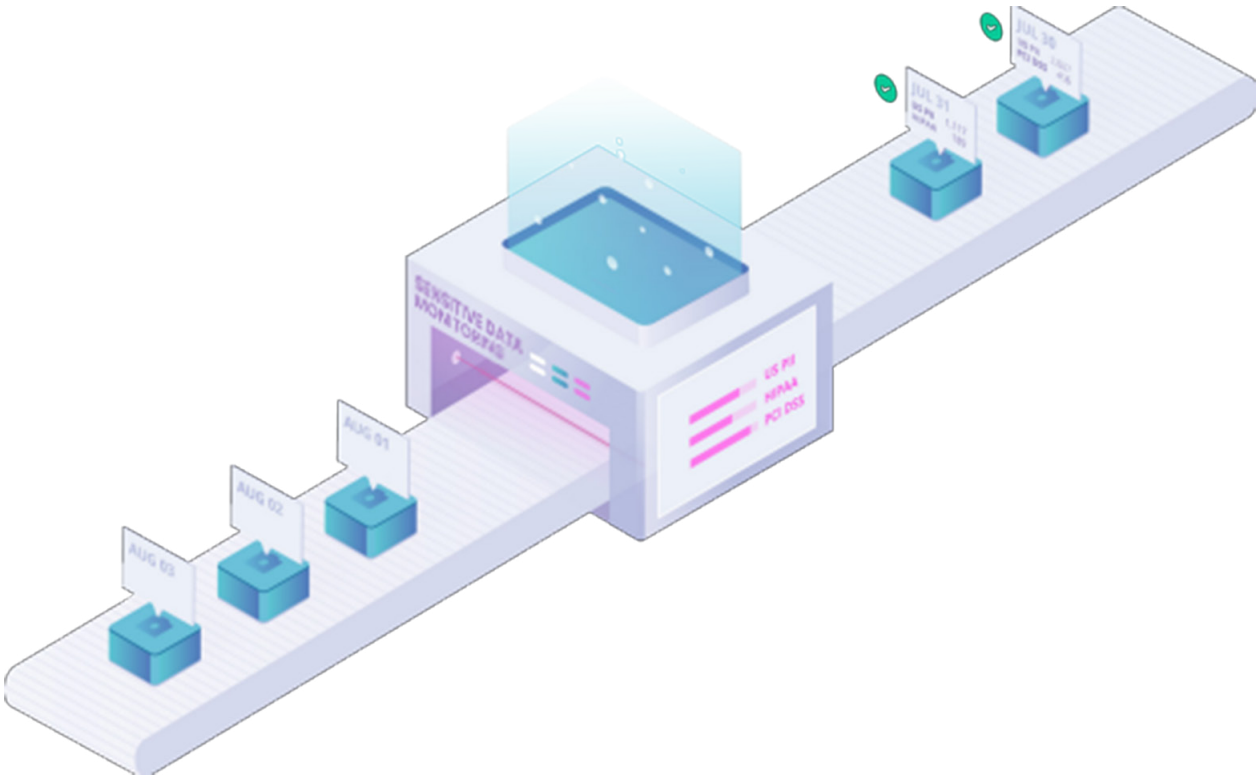
# An Introduction to Rubrik Sensitive Data Monitoring

Rubrik Technical Marketing
September 2023
RWP-0568

# Table of Contents

## WHAT IS SENSITIVE DATA MONITORING?

As businesses adopt the cloud, they grapple with massive data fragmentation, making it impossible to know where sensitive data resides. At the same time, the increasing risk of data privacy breaches and non-compliance with regulations impose serious financial penalties. Sensitive Data Monitoring is a SaaS application, hosted by Rubrik Security Cloud, that discovers, classifies, and then reports on sensitive data without any impact to production. By leveraging the data on your existing Rubrik deployments, users get up and running in just a few minutes with zero additional infrastructure required.




Sensitive Data Monitoring has two main concepts that are important to understand—Policies and Analyzers. Analyzers are where a user defines the type of sensitive data (ex. credit card numbers) that Sensitive Data Monitoring should be discovering while Policies are a logical grouping of one or more analyzers that also associates those analyzers with the specific objects (ex. VMware VM). Sensitive Data Monitoring scans. In addition to VMware vSphere, Microsoft Hyper-V, Azure Stack HCI and Nutanix AHV VMs, policies can be associated with NAS filesets, Windows filesets, Linux filesets, Volume Groups, Microsoft 365 OneDrive and Sharepoint, Rubrik Cloud Vault hosted NAS Cloud Direct, and Rubrik Cloud Vault hosted Azure VMs.

## SENSITIVE DATA MONITORING LIFECYCLE

The following steps in the Sensitive Data Monitoring Lifecycle are divided between both the SaaS infrastructure and the customer owned Rubrik cluster which ensures customer data is secure by only syncing customer metadata to Sensitive Data Monitoring.

1. Configure Sensitive Data Monitoring through the Rubrik Security Cloud UI

   All configuration items, such as creating Policies and Analyzers, are controlled through the Rubrik Security Cloud interface. Once changes are made to a Policy or Analyzers, they are automatically synced to the relevant Rubrik cluster where the classification jobs will be run. Progress of these sync jobs can be monitored on the Sensitive Data Monitoring Events page.

2. As part of the standard Rubrik cluster workflow, a snapshot (either SLA based or On Demand) is taken and then indexed.

3. After indexing has been completed, a Sensitive Data Monitoring specific job will read the raw version of every file and process that information into text for further processing.

   This step represents the "bottleneck" of the Sensitive Data Monitoring lifecycle so the jobs are automatically paralyzed across all nodes in the Rubrik cluster. The throughput for this step can be calculated as 10 MB/s * # of Rubrik cluster nodes. The first time a snapshot is processed, every file will be processed. Subsequent snapshots of the same object will be processed incrementally (i.e. only changed files will be processed) afterwards.

4. Once the text of each indexed file is available, the Analyzers that were previously synced from Sensitive Data Monitoring will check for classification hits.

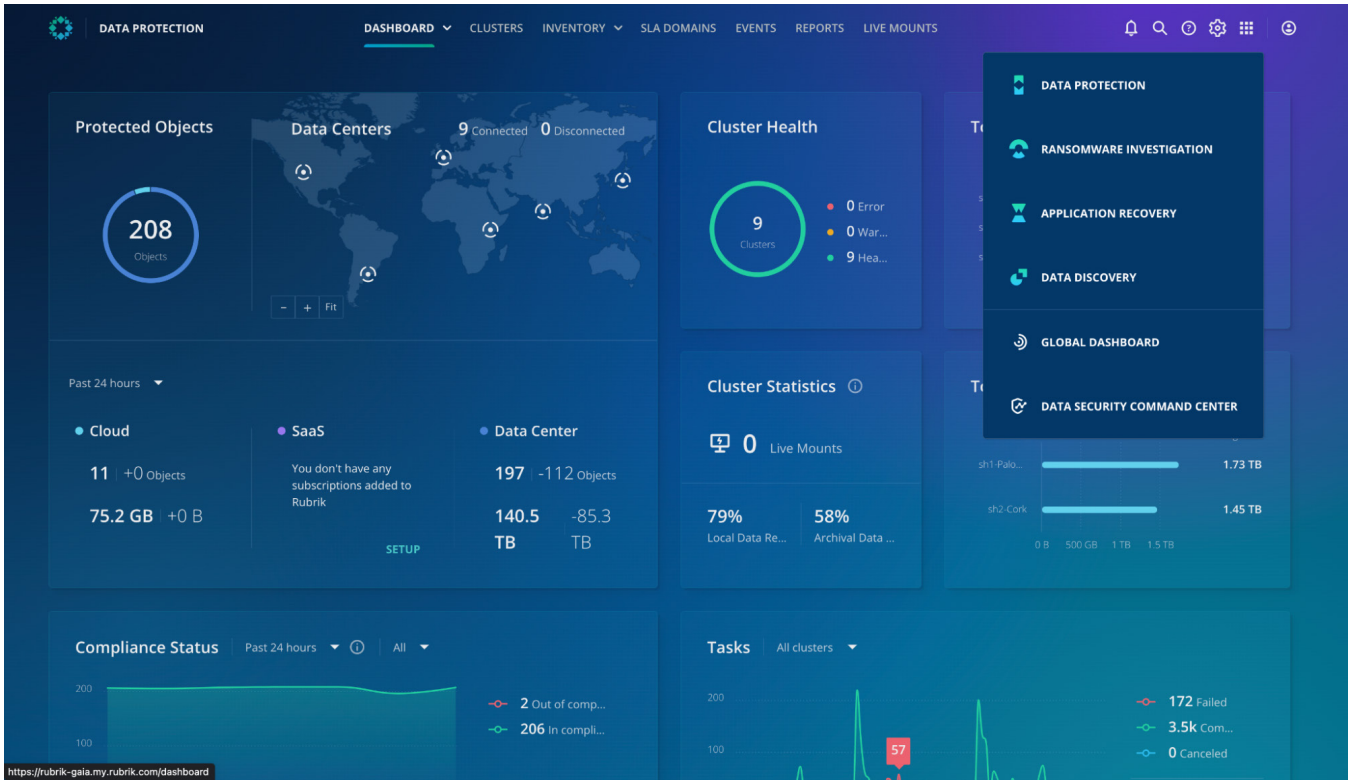
   The output of this process is metadata similar to "Sensitive Data Monitoring found X number of classification hits in this file"

5. The metadata created in Step 4 is synced to the Rubrik Security Cloud platform where it is post-processed into usable information.

   Since the results of the classification jobs result in file level information, Sensitive Data Monitoring will aggregate all the metadata to created directory and object level results. Additionally, the results for changed files are merged into previous fulls to maintain a complete snapshot view of sensitive data. This information is then presented through the UI.

## GETTING STARTED

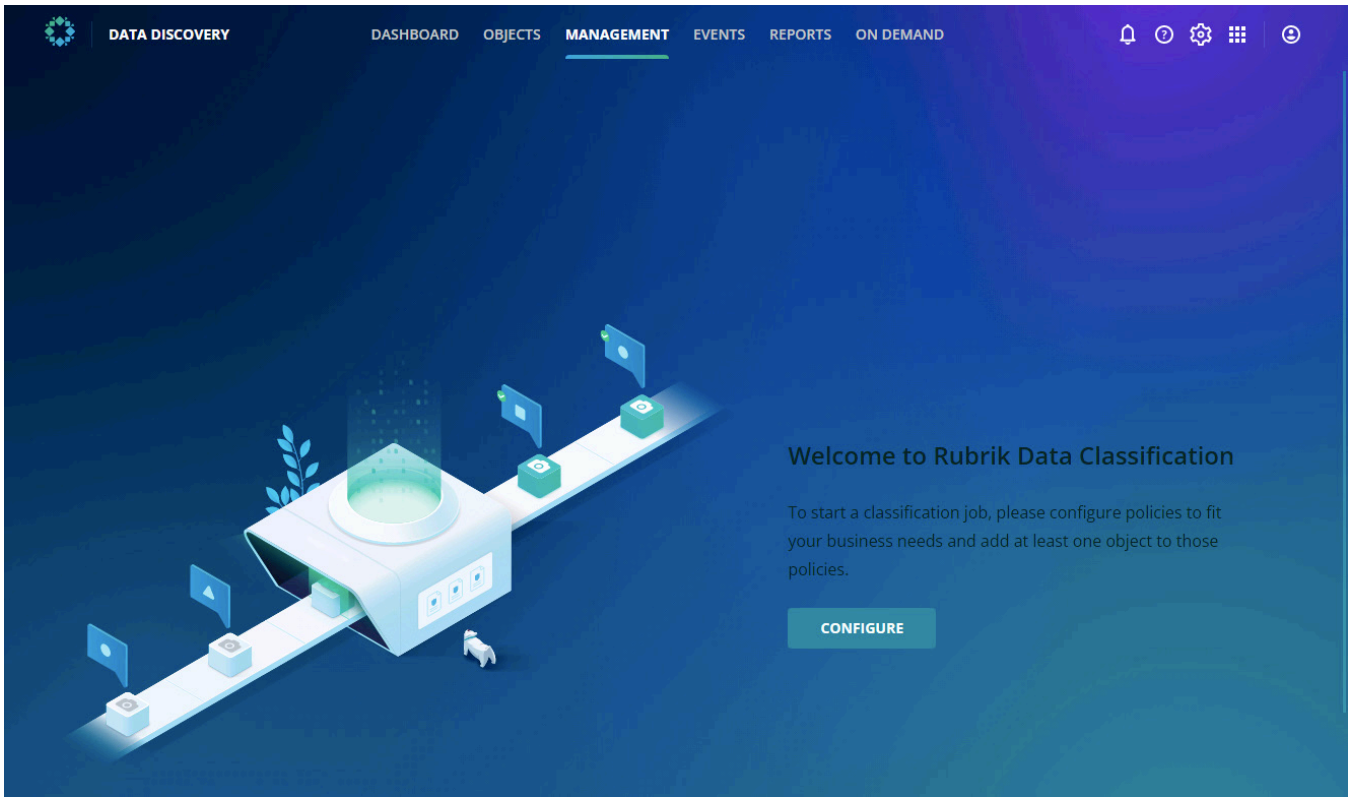
After being enabled, the Sensitive Data Monitoring application can be accessed from the application switcher icon, by selecting “Data Discovery” or by browsing directly to https://yourDomain.my.rubrik.com/sonar/



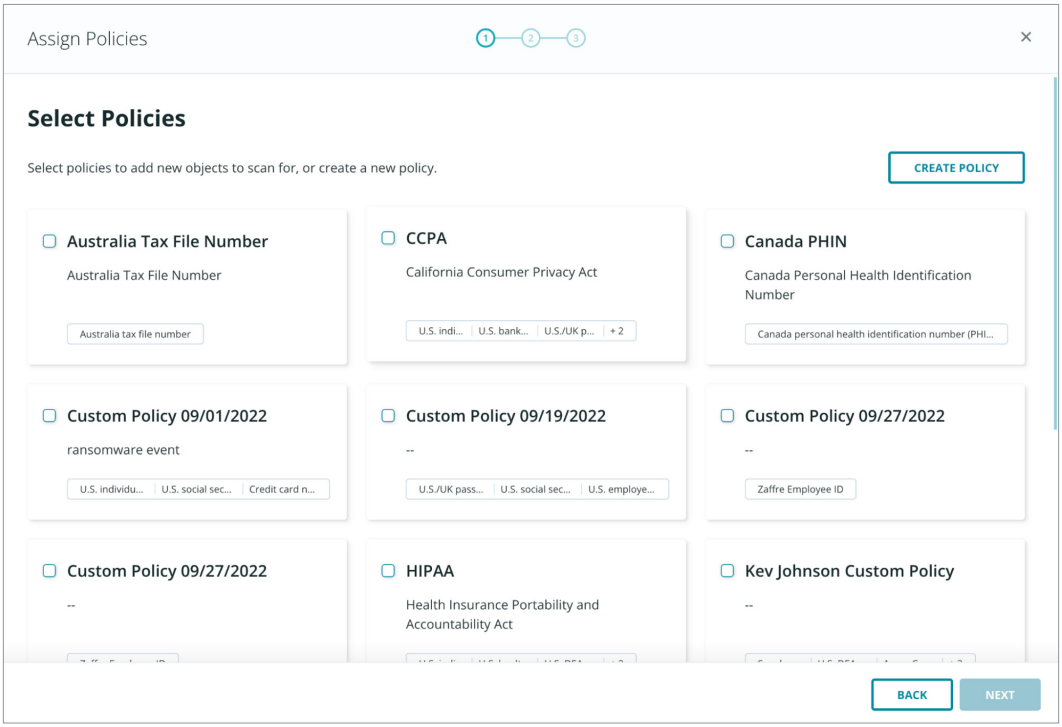

You will then be able to use the predefined Policies and Analyzers, which are covered in more detail below, or define your own to begin looking for sensitive data in your environment.

### Procedure

1. Select the **CONFIGURE** button.



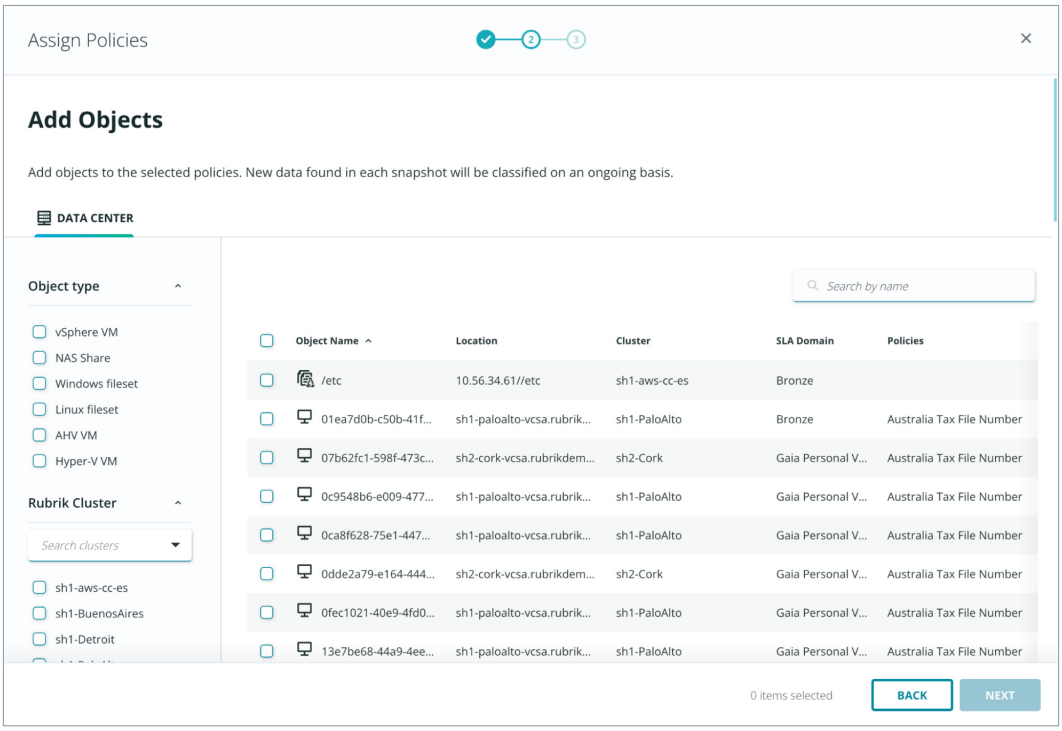
2. Choose a predefined policy or select the **Create new policy** option and then select the **NEXT** button.




3. Select the objects Sensitive Data Monitoring should scan as part of the Policy.

   You have the option of filtering by Object type, the Rubrik cluster where the object lives, or by Searching for the object name.

   Once all objects have been selected, click the **NEXT** button.



4. Review your configuration and then select the **CONFIGURE** button to save the Policy.

   Once a Policy has been defined, or later updated, the policy will automatically be synced to your Rubrik cluster where it will be used to analyze the selected objects on their next snapshot. This process is further detailed in the Lifecycle of an Analyzer section.

## POLICIES

Sensitive Data Monitoring uses predefined policies mapped to industry regulations for quick discovery of sensitive data or custom policies configured to address unique sensitive data discovery needs.

## ANALYZERS

Analyzers define which specific data patterns are searched for in indexed snapshots. They can be predefined by Rubrik or custom created by you.

### PREDEFINED

Each predefined analyzer utilizes a regular expression to detect a specific pattern relevant to the analyzer. Once that pattern has been detected, Sensitive Data Monitoring will utilize several optional layers to validate the matched pattern and prevent false flags from occurring.

The most common of these layers is a "keyword" validation that will check the 300 characters before and after the pattern match for a list of keywords that changes based on the analyzer being run. For example, the U.S./UK passport number analyzer will look for the word "Passport" before or after the main match. If "Passport" is found, the match will be marked as valid.

In addition to the keyword validation, an analyzer may use a checksum formula to validate the match. The most common of which is the Luhn algorithm. More information on the Luhn algorithm and checksum validation can be found in the glossary section.

A list of predefined policies can be found in the support portal.

### CUSTOM

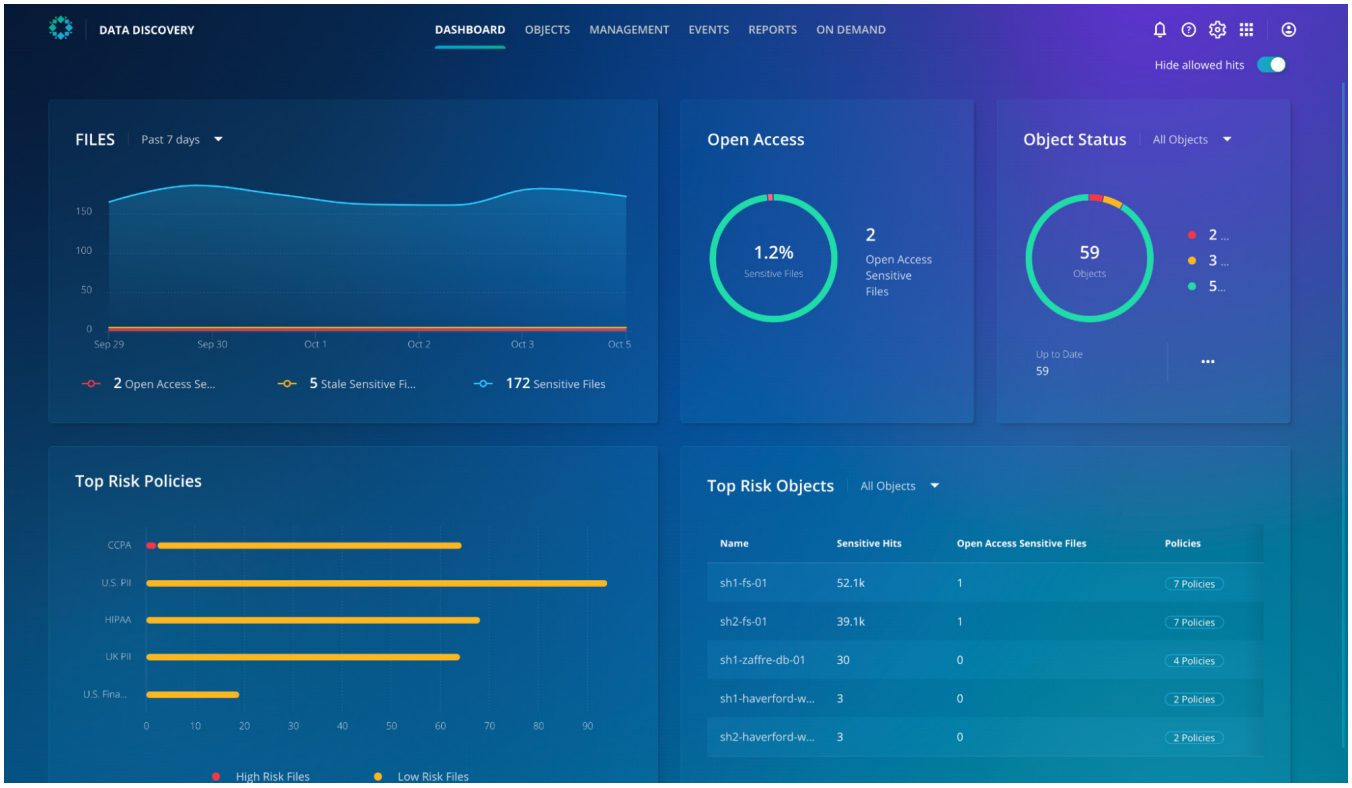
Custom analyzers support regular expressions (Perl Compatible Regular Expressions) and dictionary terms. When using dictionary terms, you can use double quotes to enclose any search term that should be in quotes or contains a separator character (comma or line break). For example, if you wanted to search for "Rubrik" (i.e. Rubrik in quotes) you use ""Rubrik"" as the dictionary term.

## OBJECT CLASSIFICATION HITS

When Sensitive Data Monitoring detects a specific piece of sensitive data in an object, as defined by a Policy, a classification hit will be shown in the UI. These hits can be viewed through the Dashboard or at the Individual Object level.

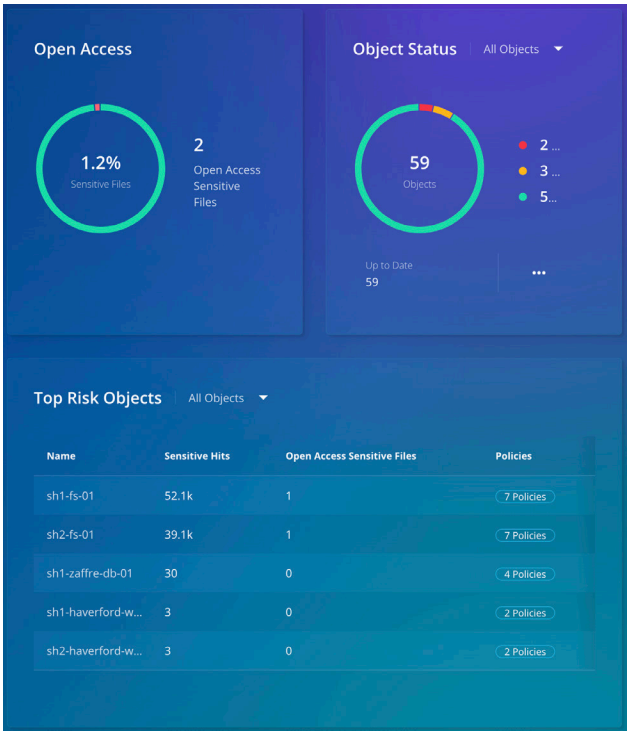
### DASHBOARD

The Sensitive Data Monitoring Dashboard provides an overview of a user's Sensitive Data Monitoring environment.



The left side of the Dashboard shows a trend graph for both the total number of **Sensitive Files**, **Stale Sensitive Files**, and **Open Access Sensitive Files** as well as the **Top Risk Policies**.

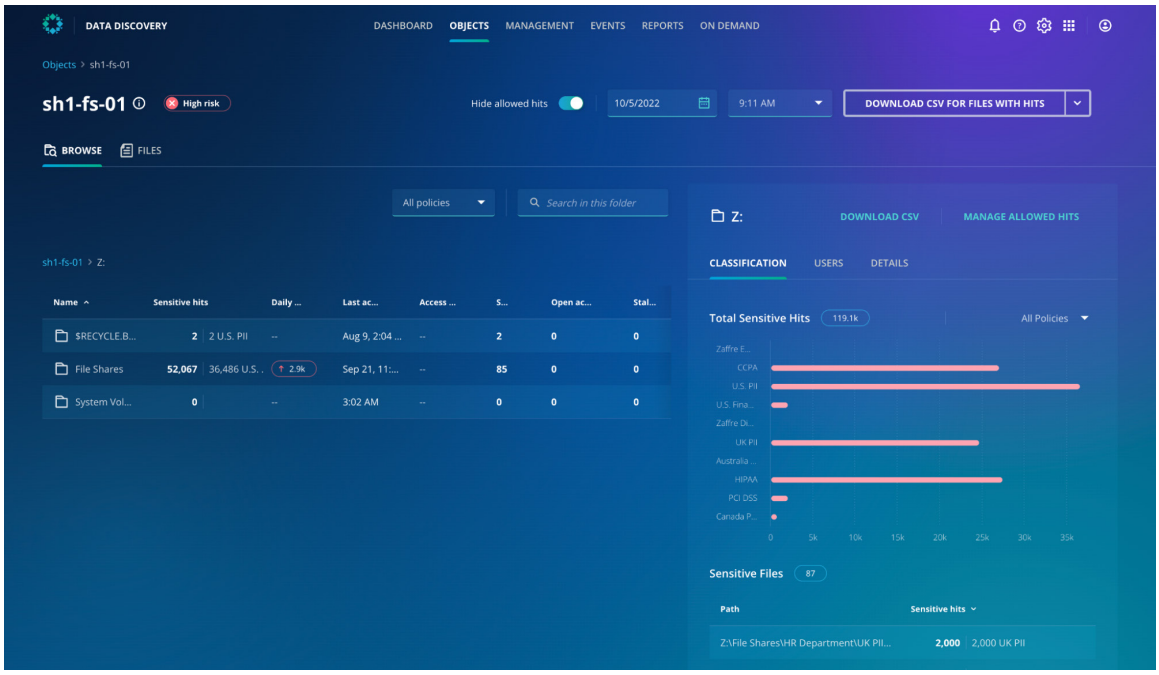

The right side of the Dashboard provides number and percentage of **Open Access Sensitive Files**, number of **Objects by status**, and **Top Risk Objects**.



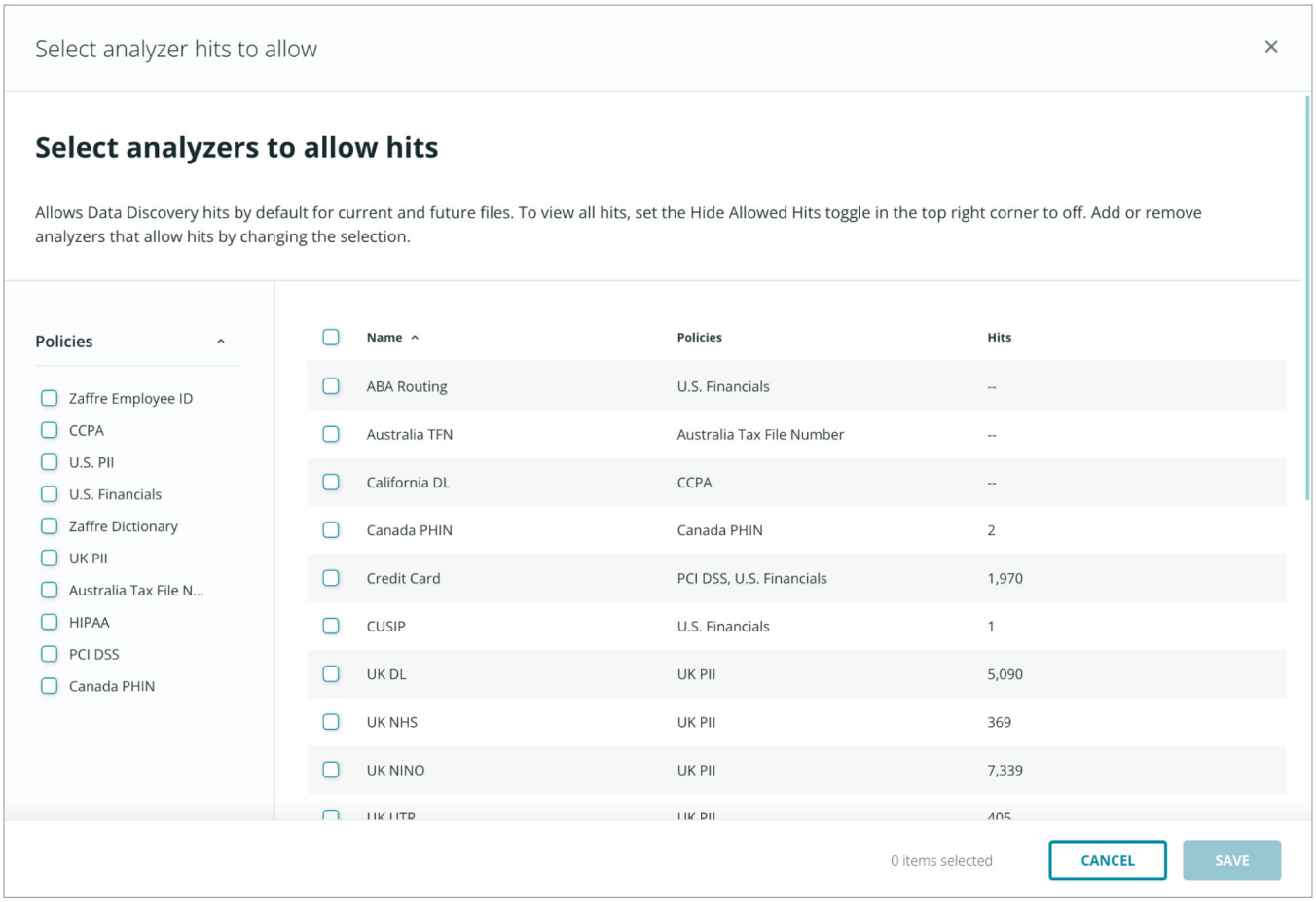
## INDIVIDUAL OBJECTS

The hits for an individual object can be viewed by selecting the links in the Dashboard, Object, selecting a Top Risk Policies, selecting a Top Risk Object, or by browsing directly to https://yourDomain.my.rubrik.com/sonar/dashboard/objects.


When viewing a specific object, you have the ability to Browse the object's filesystem and view the classification hits at each level of the object's hierarchy. For example, you can view the hits for a Windows VM entire `C:` drive or view results all the way down to an individual file.

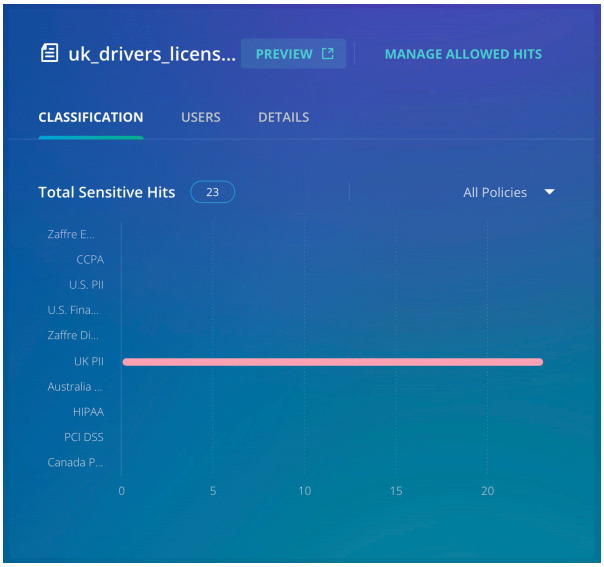


At each level of the object's hierarchy, you have the ability to **MANAGE ALLOWED HITS** for the object which allows you to "hide" a specific Analyzers results, for that object, from the UI. This is useful when you have a hierarchy object (drive, folder, file, etc.) that contains sensitive information that Sensitive Data Monitoring will hit on but that can be "ignored". For example, if you have an Excel file with credit card information that you do not need to be shown, you can update the Allowed Hits list for that Excel file to allow hits from the Credit Card Analyzer.



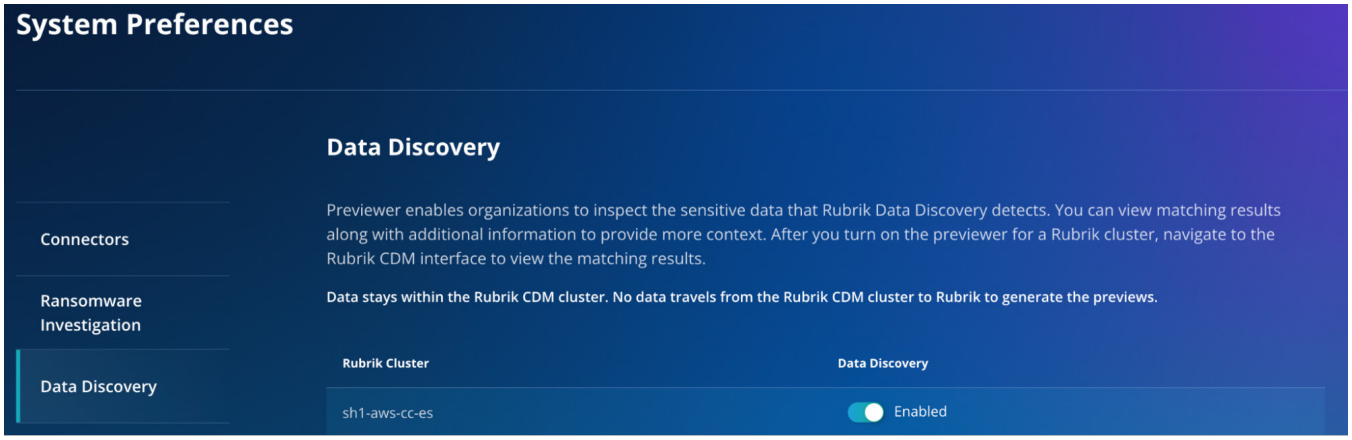

If needed, the **Hide Allowed Hits** toggle, which is found on both the Dashboard and Individual Objects pages, can be set to the off position to temporarily show any allowed hits.

When you select an individual file, a **PREVIEW** button will appear in the UI.




When selected, the **PREVIEW** button will open a link to the Rubrik cluster where you can view specific data that caused a classification hit.


**NOTE:** This information is only available on the Rubrik cluster and is not shared with or accessible by the Rubrik Security Cloud platform. This functionality can be disabled through the Rubrik Security Cloud **System preferences** page (Settings Icon > System preferences).

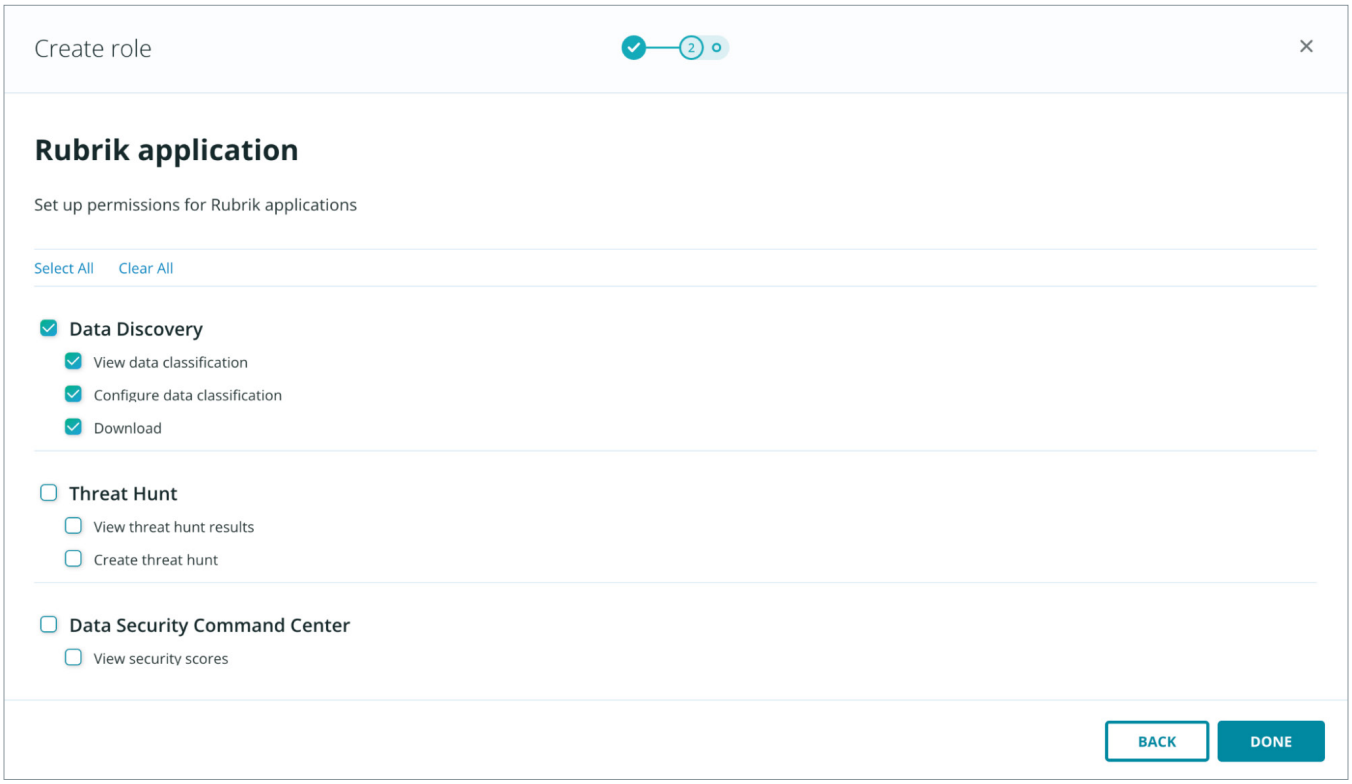


## USER ROLES AND PERMISSIONS

User management includes three Role permissions that can be used to create new Sensitive Data Monitoring specific roles which then can be applied to a Users account.

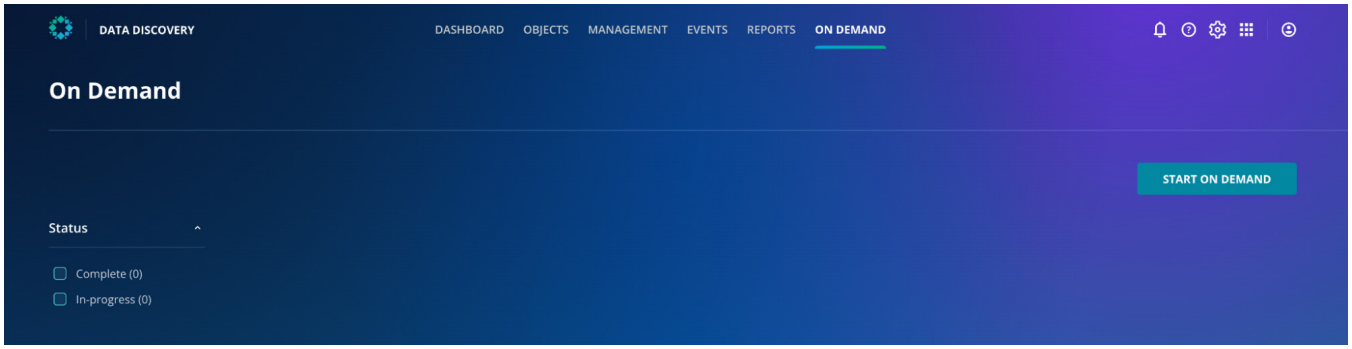
### PERMISSIONS

- **View –** Allows the user to view all Sensitive Data Monitoring information
- **Download –** Allows the user to Download any Sensitive Data Monitoring classification hit information
- **Configuration –** Allows the user to make configuration changes to Sensitive Data Monitoring

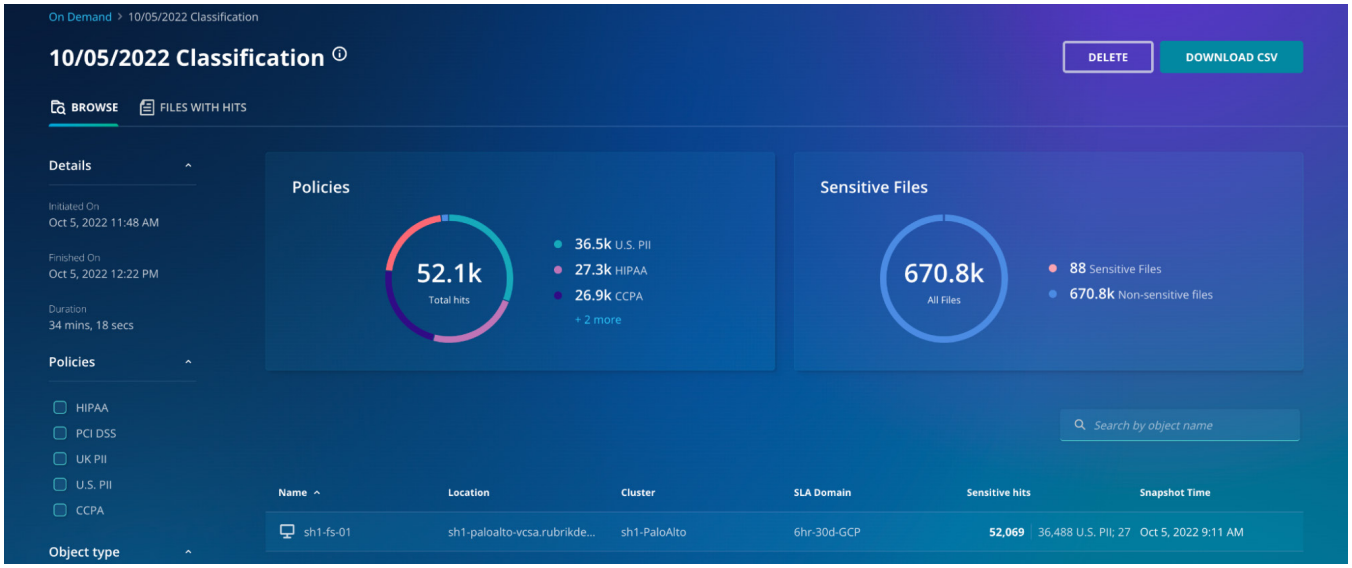


## ON DEMAND CLASSIFICATION

The On Demand page enables users to create a single use Sensitive Data Monitoring Policy. To create a new on demand classification job, select "**START ON DEMAND**" icon and then select the relevant Analyzers and Objects.

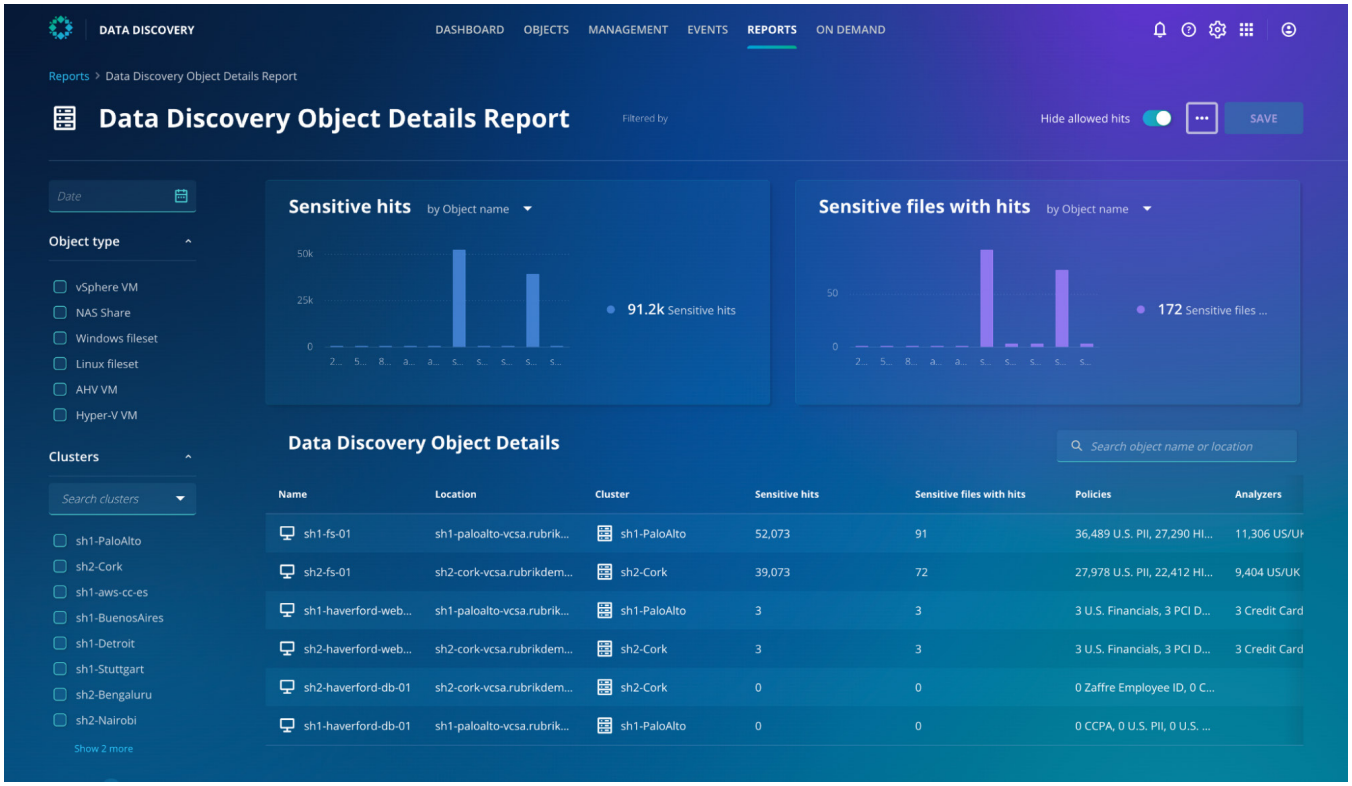



Once completed you will be able to view various results from classification jobs, such as the classification job time, the number of hits in files, and the location of the data being searched.



## REPORTING

The Sensitive Data Monitoring Object Details Report can be created on the Sensitive Data Monitoring Reports page. The report includes the ability to filter by **Object type**, **Clusters**, and **Policies** and will include the total number of **Sensitive hits** and **Sensitive files with hits** sorted by **Object name**. You can also view various detailed information on individual objects.



## GLOSSARY

**GRAMM-LEACH-BLILEY ACT (GLBA)**

Requires financial institutions—companies that offer consumers financial products or services like loans, financial or investment advice, or insurance—to explain their information-sharing practices to their customers and to safeguard sensitive data.[1]

**HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT OF (HIPAA)**

The Health Insurance Portability and Accountability Act of 1996 (HIPAA) is a federal law that required the creation of national standards to protect sensitive patient health information from being disclosed without the patient's consent or knowledge. The US Department of Health and Human Services (HHS) issued the HIPAA Privacy Rule to implement the requirements of HIPAA. The HIPAA Security Rule protects a subset of information covered by the Privacy Rule.[2]

**PCI DSS**

Outlines requirements for the way that you store, process, and submit card-based transactions. These parameters are meant to help prevent fraud and keep information secure enough to deter data breaches. While there is no absolute prevention for data breaches—even some of the biggest brands have been hit with a security issue—meeting the PCI standard helps defend against hackers and others who may access payment card information with malicious intent.[3]

**LUHN ALGORITHM**

Luhn formula, also known as the "modulus 10" or "mod 10" algorithm, named after its creator, IBM scientist Hans Peter Luhn, is a simple checksum formula used to validate a variety of identification numbers, such as credit card numbers."[4]

## SOURCES


1 Gramm-Leach-Bliley Act https://www.ftc.gov/tips-advice/business-center/privacy-and-security/gramm-leach-bliley-act

2 Health Insurance Portability and Accountability Act of 1996 (HIPAA) https://www.cdc.gov/phlp/publications/topic/hipaa.html

3 PCI DSS Compliance Guide https://www.pcisecuritystandards.org/

4 Luhn algorithm https://en.wikipedia.org/wiki/Luhn_algorithm


## VERSION HISTORY

| Version | Date | Summary of Changes | Owner |
|---|---|---|---|
| 1.0 | September 2020 | Initial Release | |
| 2.0 | November 2022 | Naming, Policy, and Analyzer updates | Joshua Robinson |
| 2.1 | May 2023 | Added reference to Rubrik DSaaS support | Kev Johnson |
| 2.2 | September 2023 | Product naming and boilerplate updates | Alpika Singh |


**Global HQ**
3495 Deer Creek Road
Palo Alto, CA 94304
United States

1-844-4RUBRIK
inquiries@rubrik.com
**www.rubrik.com**





rwp-an-introduction-to-rubrik-sensitive-data-monitoring / 20230913