File Access Manager Data Classification

Version: 8.4

Revised: March 27, 2023

Copyright and Trademark Notices

Copyright © 2023 SailPoint Technologies, Inc. All Rights Reserved.

All logos, text, content, including underlying HTML code, designs, and graphics used and/or depicted on these written materials or in this Internet website are protected under United States and international copyright and trademark laws and treaties, and may not be used or reproduced without the prior express written permission of SailPoint Technologies, Inc.

"SailPoint," "SailPoint & Design," "SailPoint Technologies & Design," "Identity Cube," "Identity IQ," "IdentityAI," "IdentityNow," "SailPoint Predictive Identity" and "SecurityIQ" are registered trademarks of SailPoint Technologies, Inc. None of the foregoing marks may be used without the prior express written permission of SailPoint Technologies, Inc. All other trademarks shown herein are owned by the respective companies or persons indicated.

SailPoint Technologies, Inc. makes no warranty of any kind with regard to this manual or the information included therein, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. SailPoint Technologies shall not be liable for errors contained herein or direct, indirect, special, incidental or consequential damages in connection with the furnishing, performance, or use of this material.

Patents Notice. https://www.sailpoint.com/patents

Restricted Rights Legend. All rights are reserved. No part of this document may be published, distributed, reproduced, publicly displayed, used to create derivative works, or translated to another language, without the prior written consent of SailPoint Technologies. The information contained in this document is subject to change without notice.

Use, duplication or disclosure by the U.S. 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 252.227-7013 for DOD agencies, and subparagraphs (c)(1) and (c)(2) of the Commercial Computer Software Restricted Rights clause at FAR 52.227-19 for other agencies.

Regulatory/Export Compliance. The export and re-export of this software is controlled for export purposes by the U.S. Government. By accepting this software and/or documentation, licensee agrees to comply with all U.S. and foreign export laws and regulations as they relate to software and related documentation. Licensee will not export or re-export outside the United States software or documentation, whether directly or indirectly, to any Prohibited Party and will not cause, approve or otherwise intentionally facilitate others in so doing. A Prohibited Party includes: a party in a U.S. embargoed country or country the United States has named as a supporter of international terrorism; a party involved in proliferation; a party identified by the U.S. Government as a Denied Party; a party named on the U.S. Department of Commerce’s Entity List in Supplement No. 4 to 15 C.F.R. § 744; a party prohibited from participation in export or re-export transactions by a U.S. Government General Order; a party listed by the U.S. Government’s Office of Foreign Assets Control as ineligible to participate in transactions subject to U.S. jurisdiction; or any party that licensee knows or has reason to know has violated or plans to violate U.S. or foreign export laws or regulations. Licensee shall ensure that each of its software users complies with U.S. and foreign export laws and regulations as they relate to software and related documentation.

Contents

Data Classification 1

General 1

Classification Architecture and Flow Architecture 2

Supported Applications 5

Supported File Types 6

Optical Character Recognition (OCR) 7

Enabling Optical Character Recognition 7

Classification Types 8

Content-Based Classification 8

Behavior-Based Classification 8

Composite Classification 9

Data Classification Components 10

Data Categories 10

Data Classification Policy 11

Rules 11

File Properties 11

Encrypted files 12

Local Classification 12

Policy Objects 12

Classification Types 14

Creating a Data Classification Policy 18

Content-Based Classification Rules 21

Creating a Content-based Classification Rule 21

Creating a Behavioral-based Classification Rule 23

Scheduling Classify Behavioral Rules Task 25

Composite Rule 26

Triggering the Composite Rules 26

Creating a Composite Classification Rule 27

Global Rules 28

Creating a Global Rule 28

Global Rule Options 30

Global Rule – Rules Screen 31

Filter 32

Data Classification Verification Algorithms 33

Out of the Box Verification Algorithms 33

Creating a Verification Algorithm 33

Verification Algorithms screen 35

Table fields 36

Uploading a New Verification Algorithm 37

Deleting a Custom Verification Algorithm 38

Application Scope 39

Editing the Application Scope 39

Policy Scope 42

Run Resource Classification 44

Import Data Classification Results 45

Data Classification Results 47

Data Classification Results – Report 47

Data Remediation Policy 48

Create a Data Remediation Rule 49

Edit a Data Remediation Rule 50

Delete a Data Remediation Rule 51

Log Reports 51

Writing a PowerShell Script for Data Remediation 51

Transferring Data Classification Policies Between Systems 52

Exporting Data Classification Policies 53

Import Data Classification Policies 54

Data Classification

Data Classification categorizes and tags business resources (BRs) based on the following:

- Content
- Behavior
- Imported designation

Classification is done by identifying resources with specific data or resources accessed by specific user types, according to standard and user defined policies.

This section describes the data classification feature in File Access Manager and the operations available on the web application, which can be found by navigating to **Compliance > Data Classification**.

General

File Access Manager's Data Classification engine is a mechanism to classify organizational data and to apply categories to it based on both a content and behavioral analysis of files and Business Resources (Data Assets) residing on the various applications.

You can use Data Classification to create policies and rules to address well-known or widely used regulation compliance requirements such as GDPR, CCPA, HIPAA, ICD, LPGD and more.

The Data Classification mechanism provides both a content-based and a behavioral-based analysis of files and BRs residing on the various applications, which facilitates their classification into categories, based on those analysis. Content-Based Classification parses and indexes the files' textual content and searches for specific patterns, according to predefined sets of rules. These patterns can consist of sensitive keywords or keyword lists, complex regular expressions representing patterns such as Social Security Numbers (SSNs) and credit card numbers, and other user-defined parameters.

Behavioral-Based Classification analyzes the activity information gathered by File Access Manager, and can be used to classify business resources (BR) based on the type of users who access the files frequently.

Content-Based Classification

Searches files for specific content of interest, such as SSNs, credit card numbers, health records, etc.

Behavioral-Based Classification

Analyzes BRs according to properties of users who access this data. For example, if members of the board of directors or members of the finance department regularly use these BRs.

The classification of both content and behavioral data depends upon user-configurable criteria. Classification results can serve as a data source on their own, and can form the basis of queries on the forensics screens (See the chapter on forensics. However, classification results also serve as an additional information layer, associated with activities and permission data.

The classification results layer connects all other layers with data.

The Data Classification module supports using external classification of files in one of the following methods:

- DC Import—importing a spreadsheet into File Access Manager listing files and directories assigned to categories
- Writing to file properties, and creating rules in File Access Manager, assigning categories to files that contain those properties.

These methods can even be used for encrypted files without File Access Manager reading the file content.

Classification Architecture and Flow Architecture

The Data Classification content indexing is performed by the Central Data Classification services and their associated Collectors. Architecture has additional information on the possible deployment models and how to scale the Data Classification Collectors to achieve greater speed and performance.

The Central Classification service reads the BRs eligible for indexing and sends them to the Collectors. The Collectors index the files in the received BRs according to the defined data classification policy, and send the results back to the Central Service to be saved in the database.

The Collectors no longer keep a persisted full text index on disk, since all the processing is done in-memory.

Content Classification Process

The classification processes (run concurrently and independently) include:

- Classification Policy Management and Update
- Running a Content Indexing task
- Querying and Retrieving Results

Classification Policy Management and Updates

Once a Content Indexing task is issued, the Data Classification Engine reads the most updated policy definition. That policy definition will persist through the duration of the Content Indexing task. Any changes made to the policy definition after the Content Indexing task has been started will not be reflected in the current classification process.

Indexing Flow

The classification engine Content Indexing Task:

1. The central service retrieves the BRs to be indexed from the File Access Manager database but only when:
   a. This is the first indexing run of a business resource
   b. The last modified business resource date is more recent than the last business resource indexing date
   c. The business resource is included in the Scope of the Application
   d. The business resource is not contained in a de-duplicated share
   e. If the data classification policy was changed from the last indexing tasks, all the BRs will be re-indexed
2. The central service sends the BRs to the Collectors.
3. The Collector retrieves the list of files in each business resource.
4. Reads the content of each file.
5. Indexes and classifies the file content and sends the results to the Central Data Classification to be saved into the database.

Data Classification Deduplication Scan

In CIFS systems it is possible for multiple shares to point to the same physical address (where they are considered “duplicate shares”).

To minimize the running time of the Data Classification task, these duplicate shares are identified, and shared data is scanned only once.

When a user queries the Forensics tab of Data Classification, the classification results are reflected through all duplicate shares.

The following scenario involves four shares in a Windows server:

- Share1 points to D:\
- Share2 points to D:\folder1
- Share3 points to D:\
- Share4 points to E:\

The results of the deduplication scan will be:

- Share1 will be scanned completely.
- Share2 will be skipped, since Share1 contains Share2
- Share3 will be skipped, since Share1 is equal to Share3.
- Share4 will be scanned completely.

When a user queries the Forensics tab of Data Classification, the user will receive the results of all shares.

Limitations and Known Issues:

If the Crawler excludes BRs in contained shares, Data Classification will not classify those BRs.

Re-Indexing Scenarios

Every data classification policy change will cause all the BRs to be re-indexed on the next indexing task. The assumption is that the policy remains static and unchanged after the implementation and testing phase are completed. File Access Manager provides different features to limit the scope of the indexed BRs to be able to test the policy changes faster, such as Scoping and Run a Specific Resource Classification task.

Enabling Optical Character Recognition

In the case of OCR scanning, *enabling* will cause the next task to re-index the Data Classification. Disabling the OCR capability **will not** initiate re-indexing. This means that once files are marked as sensitive, we can turn off the resource intensive optical character recognition process without removing this indication, until any other filtering setting is changed.

File Access Manager Text Search

The Data Classification engine uses Lucene as its primary, text-based optimized database. The Lucene database provides term-based search capabilities, based on the textual content extracted and analyzed from indexed files.

To extract textual content from various file types and formats, the classification engine uses a proprietary text extraction library, which is able to extract the file content based on its type. Based on the extracted content, the Lucene indexing service parses and analyzes file content into an index of searchable terms. The full content of the files itself is not saved as part of the index, which allows the index to remain relatively small, highly efficient, and optimized for term-based textual searches.

When the system compares a Content Classification request with the textual index, it parses and translates the various policy rules into term-based search queries. Query results, representing files that correspond to the rules' requirements, consist of file names, extensions, and full path locations, along with other attributes. In certain cases, results may include an actual term or phrase that matches the rule-based query, rather than the full content of the file.

Regular-Expressions

Regular-expression-based rules involves matching regular-expression patterns with a file during the process of reading the file content, and not comparing the pattern with a term-based index.

Lucene's Indexing Process

While it parses and analyzes the content data, the Lucene index analyzer eliminates white spaces, certain punctuation characters, and "stop-words" from the content. Stop-words are a predetermined set of frequently used words with diminished semantic significance, such as pronouns and prepositions. Lucene filters stop-words to keep the index manageable, to eliminate "white noise," and to improve search heuristics. Lucene analyzes and tokenizes file content into searchable terms based on the white spaces and stop-words omitted from the original text. The tokenizing algorithm affects Data Classification policy rules.

Multi-term Phrase-Based Rules

The Data Classification engine allows both single-term keyword searches and multi-term phrase searches. Lucene omits any "stop-word" contained in a multi-term search phrase. For example, a rule containing the phrase "It was the best of times, it was the worst of times," will classify the file containing the entire sentence, as well as any file containing a contiguous phrase, such as "best times worst times." To avoid possible false-positive classification, it is best to restrict multi-term phrase searches to meaningful, contiguous terms.

Chinese and Logogrammatic Languages

Some scripts, such as Chinese, represent words by symbols (logograms), and a single word may consist of one or more logograms. Furthermore, while most languages use white spaces to separate words, Chinese, as well as other logogrammatic scripts often do not separate words by spaces. The combination of these two phenomena, along with Lucene's omission of white spaces, will cause phrase searches in Chinese (with multiple logograms, separated by spaces) to return positive matches for files containing the same sequence of logograms, regardless of the spaces between them. Thus, a rule containing the phrase "*胥 莚 慮贊 蹝 轈*", will classify files containing phrases that consist of these logograms, regardless of spaces. Therefore, the phrases "***胥莚慮贊蹝**轈,*" "*胥莚慮贊蹝轈*" or "*胥莚慮贊 蹝轈*" , and "***胥 莚 慮贊 蹝** 轈,*" will all be classified by the rule defined above. However, single term keyword searches of words consisting of multiple logograms, and phrases not separated by spaces, will return correct, exact match results: a rule containing the term "*胥莚慮*" will only classify files containing that exact term. If more complex phrases are required, a rule containing multiple phrases with the "Contains All" operator will give the desired results.

Supported Applications

Data Classification supports the following applications:

Target System	Products and Supported Versions
On-premises File Storage	Microsoft Windows
	Microsoft SharePoint
	NFS v3/v4
NAS File Storage	NetApp for CIFS
	NetApp for NFS
	EMC Celerra/VNX/Unity for CIFS
	EMC Celerra/VNX for NFS
	EMC Isilon for CIFS
	Hitachi HNAS
	DFS for CIFS
	Generic CIFS
O365 File Storage	Microsoft OneDrive for Business
	Microsoft SharePoint Online (Office 365)
Cloud File Storage	Box
	Dropbox
	Google Drive
	Ctera
	Azure Files

Supported File Types

The privacy engine indexes data based on a file’s content and attributes. The system also supports file properties and custom properties for all supported file types. The privacy engine reads file content based on the file extension.

Image files can be analyzed and searched for keywords using an optical character recognition (OCR) capability. This is a resource heavy process and is configured separately. See section Optical Character Recognition (OCR).

The Data Classification engine supports the following file types /extensions:

File Extension	Expected file type
docx doc xls xlsx ppt pptx	Microsoft Office files
txt csv	Plain Text (including Comma Separated Values files)

File Extension	Expected file type
htm html xml	Web files
cs js sql	Code script files
pdf	
zip gzip tar rar 7zip	Archive files
Jpeg jpg tif tiff gif png wmf emf bmp pdf	Image files analyzed by the OCR module*

The system downloads files from cloud-based content stores and non-CIFS application (for example, Box, DropBox, Google Drive, OneDrive, SharePoint and NFS) to a local directory on the server. Once the indexing process finishes, the system deletes the downloaded files from the indexing server.

Optical Character Recognition (OCR)

File Access Manager can identify text from within image files either directly or embedded in other files – such as a scanned documents or a collection of scans stored in a zip file. Files less than 1000 pixels across will not be scanned to avoid less reliable results from low resolution images.

The data privacy engine can analyze files containing sensitive data in image form.

> Note: The optical character recognition process is resource intensive and should be configured carefully taking the run-time into consideration. It is disabled by default.

OCR Capability can be added to the scope selected in the DSAR Scope screen.

Enabling Optical Character Recognition

By default, optical character recognition is disabled on the entire scope of the DSAR. To enable optical character recognition on a resource, edit the application scope line.

1. Find the desired application from the DSAR Scope screen.
2. Click **Edit**.
3. Click **Optical Character Recognition (OCR)** to enable OCR analysis for this application.

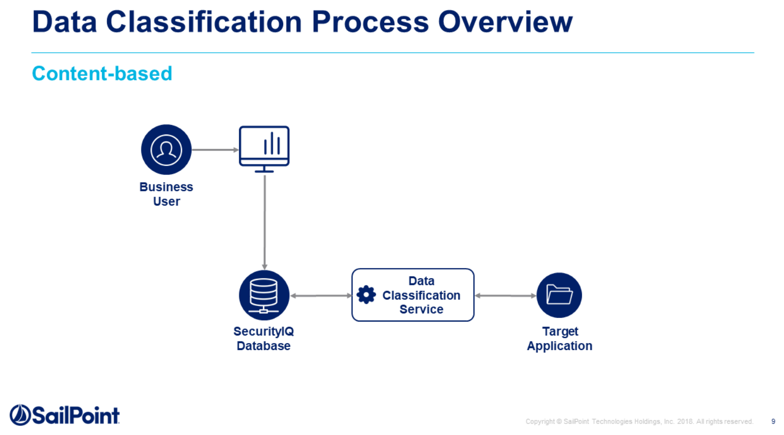
Classification Types

Data Classification types include:

- Content-Based Classification
- Behavior-Based Classification
- Composite Classification

Content-Based Classification

The classification engine indexes data based on file attributes and file contents (for text, office, and PDF files). The classification engine determines the file type by the file extension.



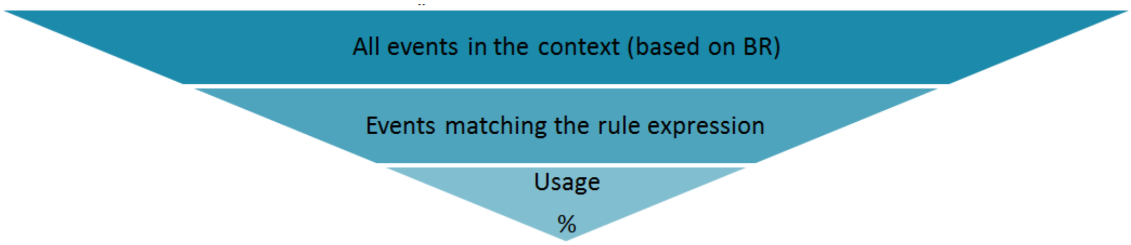
Data Classification Process Overview

Content-based

Behavior-Based Classification

Behavior-based classification classifies BRs based on actual user activity.

An example of a behavior-based classification rule would be to classify BRs as “Finance” if more than 80% of the activities in the last month were issued by users whose department is defined as Finance in Active Directory.



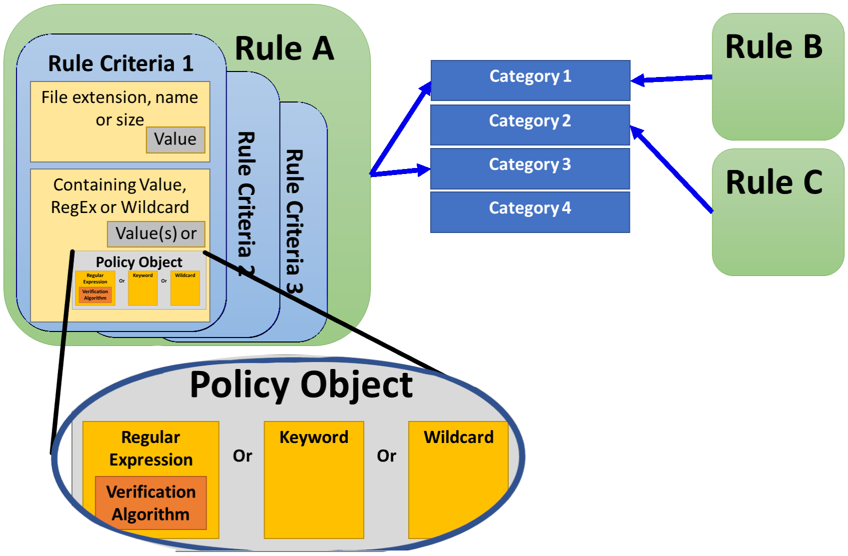
All events in the context (based on BR)
Events matching the rule expression
Usage
%

Composite Classification

In order to comply with complex regulations, it is sometimes required to use additional logic with data classification rules by combining the results of several classifications. The composite data classification rule is based on the File Access Manager categories already classified by the Content and Behavioral rules.

> Note: A Data Classification policy cannot have both composite rules and other types of rules.

Data Classification Components



- The Data Classification process assigns **categories** to **business resources** according to **rules**.
  - Rules are composed of one or more rule criteria

Rule criteria consist of finding a match within files to one or more string or pattern.

The strings can be defined as free text, regular expressions, or one stored as a **policy object**.

A regular expression in a policy object may be accompanied by a verification algorithm to further narrow down the search.

> Note: There are policy objects and verification algorithms out of the box for standard searches, or you can create your own to fit your needs.

The classification rule is the main data classification component. Rules also contain subcomponents that complete the rule structure, simplify the rule management task, and provide extended functions.

File properties can be used for classification of files that is performed by the customer manually or using a third party application. File Access Manager will read the metadata on the files, and can use them for data classification rules. This will include reading metadata from encrypted files.

Data Categories

The data category (the basic component of data classification) is the tag used when a classification rule is satisfied.

To define a data category, open the **Manage Categories** panel from any of the Data Classification screens

For example:

1. Navigate to **Compliance > Data Classification > Policies > Actions > Manage Categories**
   or
   Compliance > Data Classification > Rules > Actions > Manage Categories.
2. In the Manage Categories window, type the category name in the *Add New Category* section.
3. Click **Add**.

The system adds a new data category to the Current Categories list. Users can edit and delete existing user-defined categories from the Current Categories list. Users can also search categories either by name or by checking the **Show user defined categories only** checkbox.

Data Classification Policy

The Data Classification Policy is a logical container for data classification rules. For example, all the rules that belong to HIPAA should be located under the HIPAA policy. The system already contains several predefined policies, and users can create additional user-defined policies.

Rules

Policies set the rules for detecting sensitive data to be protected by compliance regulation or by organizational procedure.

File Properties

File Access Manager indexes standard attributes, including extension, size, and file name and also index attributes for office files. All the file properties are discovered and created during the indexing process.

1. In the web client, navigate to **Compliance > Data Classification > Rules > Actions > Manage File Properties**
   or
   Compliance > Data Classification > Policies > Actions > Manage File Properties to open the Manage File Properties window.
2. Type in the file property details.
3. Check the **Custom Properties** checkbox, if relevant.
4. Click **Add**.

Encrypted files

In order to classify encrypted files without File Access Manager reading the file contents, you can tag the files locally according to your classification rules, and use these tags for classification rules (See Local Classification).

> Note: If you choose to tag the file using the Tag's property, it will be called Keywords after being uploaded to File Access Manager.

Local Classification

You can use a local classification for files, tagging files with relevant tags. The metadata of the files are uploaded to the File Access Manager database as file properties in the scanning process. These properties can be used to create classification rules manually.

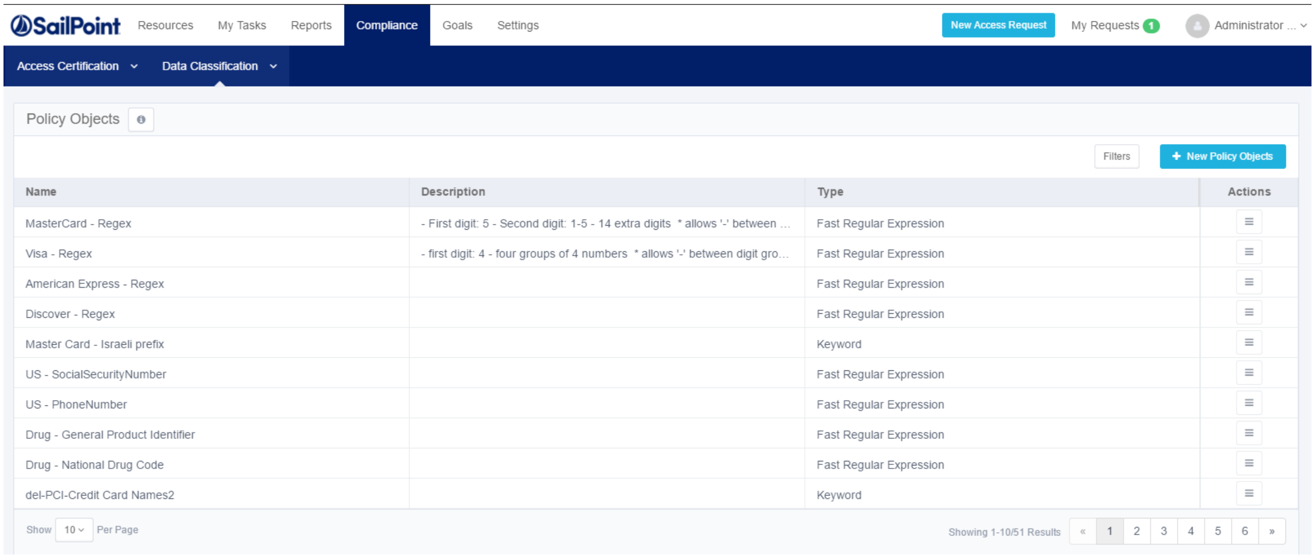
The file properties found will be added automatically to the list of available properties for filtering after the first iteration. In order to have these properties available in the initial run of the Data Classification, add the properties to the property list, as described in File Properties above.

Policy Objects

Policy objects are searches, saved for use in rules.

For example, predefined policy objects can search for credit cards.

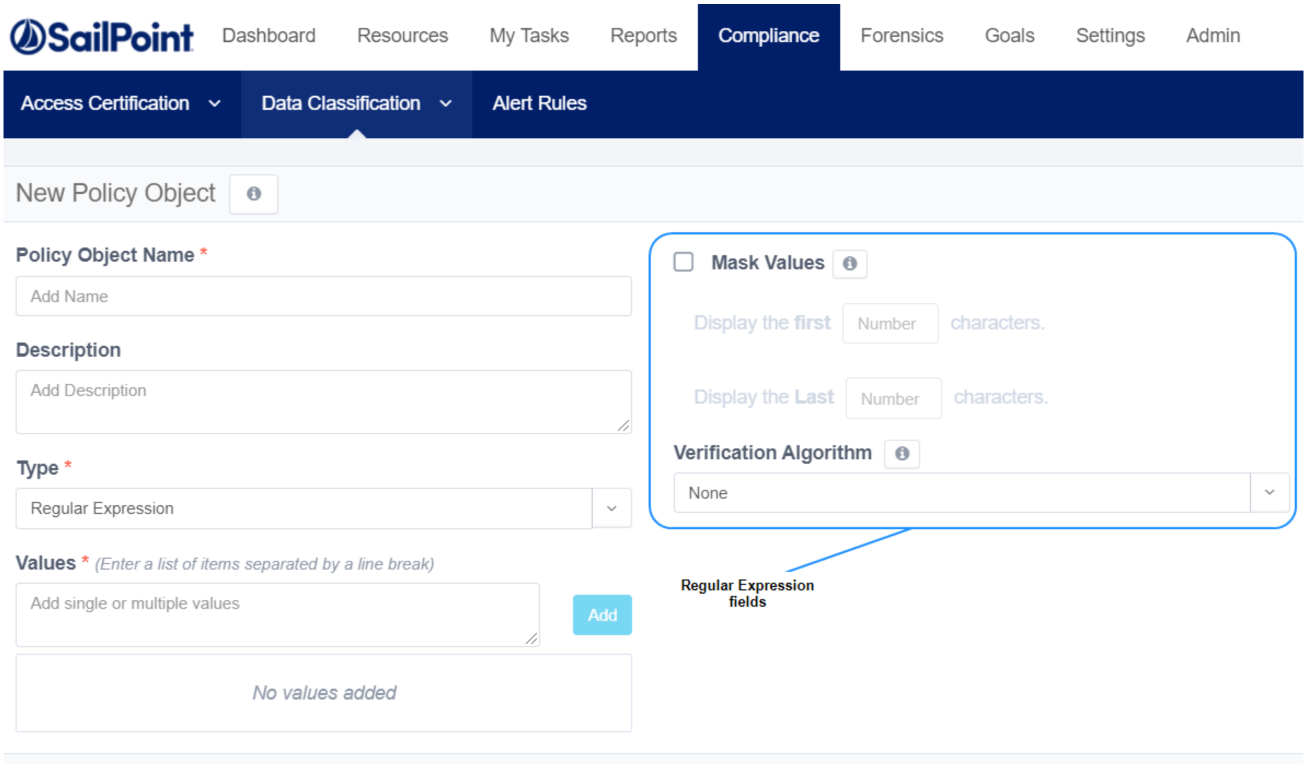
1. Navigate to **Compliance > Data Classification > Policy Objects.** to open the Data Classification – Policy Objects page.



Policy Objects

Name	Description	Type	Actions
MasterCard - Regex	- First digit: 5 - Second digit: 1-5 - 14 extra digits * allows '-' between ...	Fast Regular Expression	≡
Visa - Regex	- first digit: 4 - four groups of 4 numbers * allows '-' between digit gro...	Fast Regular Expression	≡
American Express - Regex		Fast Regular Expression	≡
Discover - Regex		Fast Regular Expression	≡
Master Card - Israeli prefix		Keyword	≡
US - SocialSecurityNumber		Fast Regular Expression	≡
US - PhoneNumber		Fast Regular Expression	≡
Drug - General Product Identifier		Fast Regular Expression	≡
Drug - National Drug Code		Fast Regular Expression	≡
del-PCI-Credit Card Names2		Keyword	≡

2. Click **New Policy Object** to open the New Policy Object page.



Data classification policy object fields include:

Policy Object Name

Name of the policy object

Description

Free text

Type

The type of search the policy object performs:

Keyword

A keyword may be one or more words. If multiple words are involved, the entire phrase will be searched.

Note that stop words such as "a" or "and" are stripped from the search keywords. If you want to include stop keywords in the phrase, you can use a regex phrase instead. (For a nerd-level description of ignoring stop words, see https://www.elastic.co/guide/en/elasticsearch/guide/current/stopwords.html)

Wildcard

Supports the following special characters:

***** any number of characters

? only one character

Regular Expression

Using standard regex for defining policies

Values

Values to search for:

- Single Value
- List - A list of matching values

Mask Values (Regular Expression policy objects only)

Masking portions of matched values.

- **Display the first characters**—number of characters from the left displayed in the matched value
- **Display the last characters**—number of characters from the right displayed in the matched value

Verification Algorithm

A code based algorithm to enable more complex filtering. See Data Classification Verification Algorithms for further details.

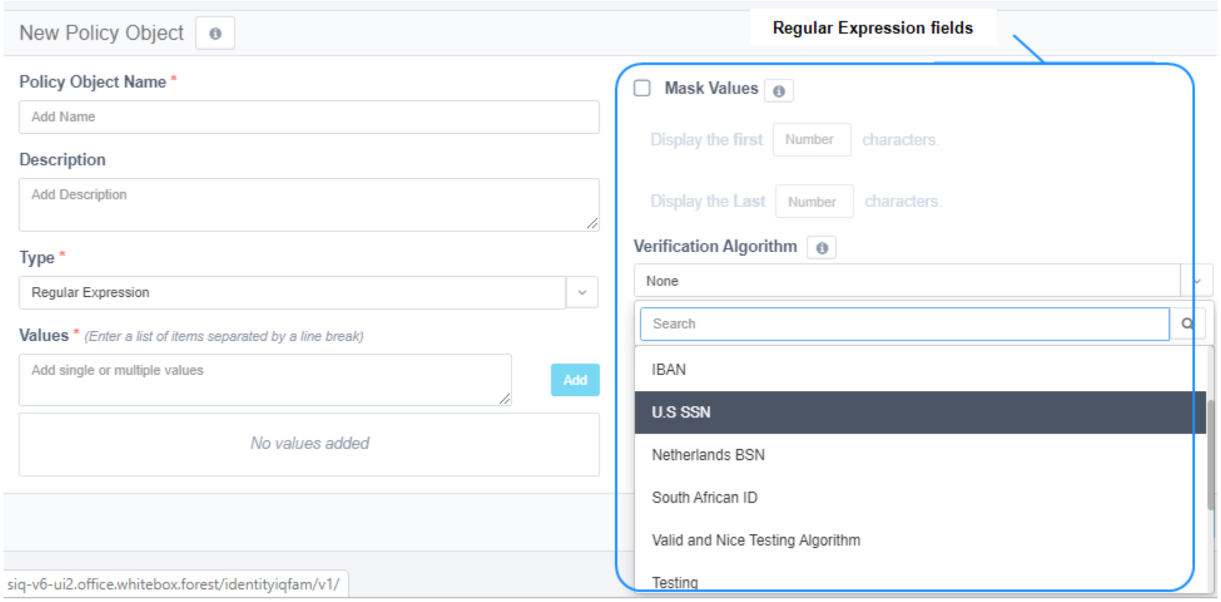
Policy objects are a good way to reuse searches containing complex definitions.

Click **Save** to end the New Policy Object process.

Classification Types

Regular Expressions Within Policy Objects

Regular expressions form the basis for many content pattern searches. File Access Manager uses the *.net regular expression engine* as its underlying engine for regular expressions searches. All regular-expression definitions and searches must conform to the engine’s restrictions, limitations, and standards.



New Policy Object

Policy Object Name *

Add Name

Description

Add Description

Type *

Regular Expression

Values * *(Enter a list of items separated by a line break)*

Add single or multiple values

No values added

siq-v6-ui2.office.whitebox.forest/identityiqfam/v1/

Regular Expression fields

□ Mask Values

Display the first Number characters.

Display the Last Number characters.

Verification Algorithm

None

When selecting a policy of type Regular Expression, the New Policy Object panel adds the following fields to the New Policy Object panel (see image above).

Verification Algorithm

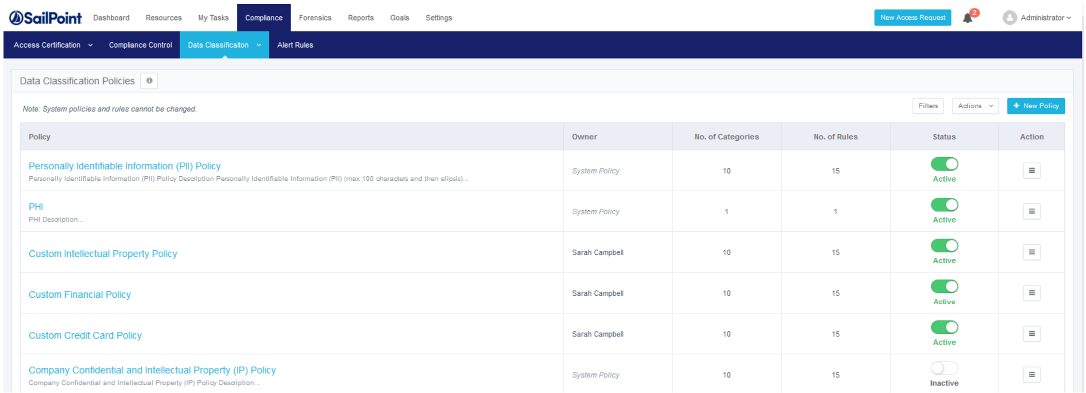
A standard, out of the box example, is the Luhn verification algorithm. This algorithm ensures that all phrases classified as credit cards are, indeed, valid credit card numbers (as far as an algorithm can validate without contacting the bank, of course). When selected, this verification will only be run on strings that conform with the credit card regular expression entered, for example:

“^3[47][0-9]{13}$”

See Data Classification Verification Algorithms for a full description on creating verification algorithms.

Mask Values

By default, the regular-expression matches are saved as part of the results. It is recommended to mask the values of the matches to avoid exposing sensitive data in the File Access Manager database.



Data Classification Policies

Note: System policies and rules cannot be changed.

Policy	Owner	No. of Categories	No. of Rules	Status	Action
Personally Identifiable Information (PII) Policy Personally Identifiable Information (PII) Policy Description Personally Identifiable Information (PII) (max 100 characters and then elipsis)...	*System Policy*	10	15	Active	≡
PHI PHI Description...	*System Policy*	1	1	Active	≡
Custom Intellectual Property Policy	Sarah Campbell	10	15	Active	≡
Custom Financial Policy	Sarah Campbell	10	15	Active	≡
Custom Credit Card Policy	Sarah Campbell	10	15	Active	≡
Company Confidential and Intellectual Property (IP) Policy Company Confidential and Intellectual Property (IP) Policy Description...	*System Policy*	10	15	Inactive	≡

Regex Matching and Case

Please note that regex matching is case sensitive by default. To make a regex ignore case, use the prefix “(?!)”
For example: “home” will find “home”, but ignore “Home”
The regex “(?!)home” will find “Home”, “HOME” and “HoMe”

Identifying Line Breaks using Regex in File Access Manager

For parsed files, line breaks are represented by a single CR (\r), instead of (\r\n) or (\n), and therefore not identified by the regex line boundaries ^ and $.

if we take the following regex:

```
(?m)(^|\s)up($|\s)
```

And try to match it with the following text (assuming the line breaks are \r):

```
going
```

```
up
```

```
up
```

```
and away!
```

It will not match anything since the line breaks are not \n as expected by the regex.

In order to identify the start and end of a line, we have to check for the CR explicitly. The issue is that once we identify an end of line character, the cursor has moved past this character, and we can't use this to identify the start of the next line.

If we change the regex to look like this:

```
(\r|\s)up(\r|\s)
```

It’s going to match only the first up, since the \r character will be part of the match and thus not part of the evaluation for the next “up.”

We need to check the previous and next characters, without moving the cursor.

If we try the following regex:

```
(?<=(\r|\s))up(?=\r|\s)
```

Both “up” strings will be matched. This is because of two modifications:

(?<=...) positive lookbehind,

When there’s a match, it moves back to assert whether the regex that replaces “...“ is matched, but then discards the match and moves forward to where it was to continue matching.

(?=...) positive lookahead

When there’s a match, it moves forward to assert whether the regex that replaces “...“ is matched, but then discards the match and moves back to where it was to continue matching.

Combining those two means the match contains only “up” without the preceding or following \r, so they can be used for more matches.

These non-capturing matches are known as zero-length assertions. For more information on lookahead and lookbehind assertions (collectively called lookaround) see https://www.regular-expressions.info/lookaround.html.

Examples

To look for rows starting with "John," you could use: `(?<=\r|^)John.*(?=\r|$)`

To look for rows ending in "Doe," you could use: `(?<=\r|^).*Doe(?=\r|$)`

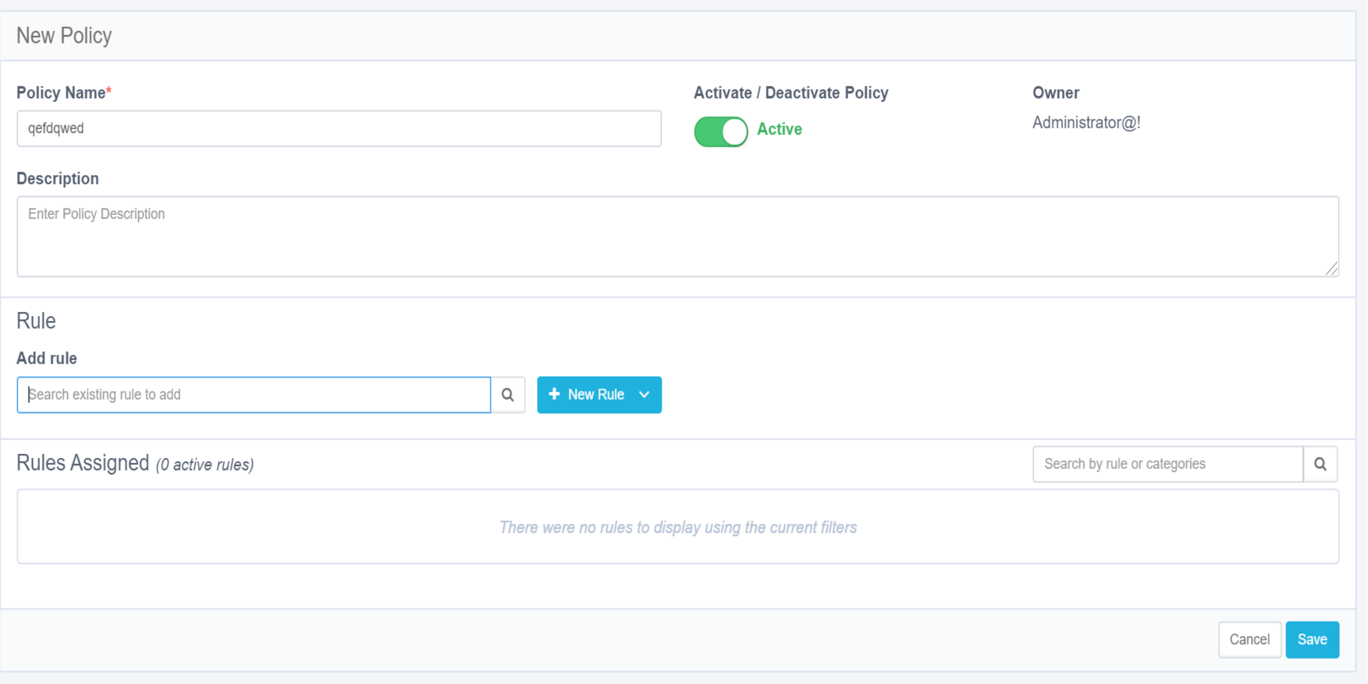
Creating a Data Classification Policy

Creating a data classification policy involves defining several policy details to make the policy unique. Any new policy can be used as a template and the basis for additional policies.

To create a new policy:

1. In the web client, navigate to **Compliance > Data Classification > Policies > New Policy**.

   A New Policy window displays.



New Policy

Policy Name*: qefdqwed

Activate / Deactivate Policy: Active

Owner: Administrator@!

Description: Enter Policy Description

Rule

Add rule: Search existing rule to add

Rules Assigned (0 active rules)

There were no rules to display using the current filters

The available Classification Policy fields / buttons that display in this window include:

Policy Name

Policy names are unique. It is best to create a naming convention that avoids using the same name twice.

Activate/Deactivate Policy

Users can activate or deactivate a policy using this button

Owner

The login user is the creator of the policy. (This field is read-only.)

Description

Free text

2. Users can add existing rules or create a new rule for a policy.

   a. Add an existing rule, using the Add Rule search field.

   b. Click **+New Rule** to add a new rule.

      The rule you added displays in the Rules Assigned list

   Users can perform the following actions on rules:

   - Activate/deactivate
   - Edit (only user-defined rules)
   - Remove

3. Click **Save** to save the new policy.

4. The system adds the policy to the Policies list.

To search for an existing policy:

1. Navigate to: **Compliance > Data Classification > Policies**

   The Policy window displays.

2. Search for existing policies by typing a name or part of a name in the following search fields:

   - Policy Name
   - OwnerSearch by status by selecting an option from the Status dropdown menu.

3. Fine tune the search even further by selecting an option from the Scope Type dropdown menu or by typing a name or part of a name in the Application Type search field.

4. You can perform the following actions on a selected policy:

   - Activate/deactivate
   - Edit (only user-defined policies)

- Duplicate
- Delete (only user-defined policies)

Content-Based Classification Rules

A content-based classification rule specifies file attributes, as well as data patterns within the files, that fit a particular type of data. For example, credit card numbers, driver's license numbers, text files created last month by user X@domain.com. Each such rule is associated with a category.

Creating a Content-based Classification Rule

In the process of creating a content-based classification rule, File Access Manager performs an AND operation between each expression. However, some operators act as an internal OR (for example, the IN operator).

To create a Content-Based rule, perform the following steps:

1. Navigate to ***Compliance > Data Classification > Rules***.

2. Click **+ New Rule** >Content-Based Rule

   A New Content-Based Rule window displays.

   The available Content-Based Rule fields include:

   Rule Name (mandatory field)

   Rule names are unique. It is best to create a naming convention that avoids using the same name twice.

   Categories

   One or more categories to tag files that meet rule requirements.

   To add a new category to the Categories list, click **Manage Categories** and add a new item.

3. In the web client, navigate to **Compliance > Data Classification > Rules > New Rule >** .

4. In the Rule Criteria section, add the general details to the Content-Based Classification rule.

   Users can search for existing rules, using filters.

   Users can perform the following actions on rules:

   - Edit (only user-defined rules)
   - Duplicate
   - Delete (only user-defined rules)

5. Create an expression and click **Save**.

> Note: Users can edit or delete existing rule criteria.

6. Add additional rule requirements as needed.

7. Click **Save** to save the new content-based rule.

   The system adds the rules to the Rules list.

Creating a Behavioral-based Classification Rule

> Note: You must enable the “Classify behavioral rules” task in order to run behavioral based rules.

In the process of creating a content-based classification rule, File Access Manager performs an AND operation between the expressions. However, some operators act as an internal OR (for example, the IN operator).

To create a Behavioral-based rule:

1. Open the rules page.

   Compliance > Data Classification > Rules

2. Click **+ New Rule** > Behavioral Based Rule.

   #### *Rule Name*

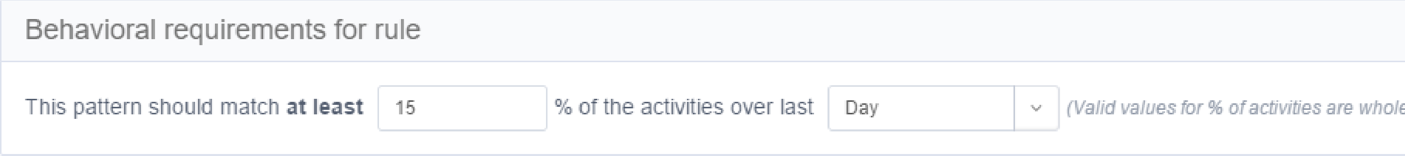
   Rule names are unique. It is best to create a naming convention that avoids using the same name twice.

   #### *Categories*

   Enter one or more categories for the rule.

   To add a new category to the Categories list, click **Manage Categories** and add a new item.

3. Behavioral Requirements for Rule specifies the threshold and timeframe for categorizing BRs according to the users accessing these files. An example of a threshold configuration is “at least 25% of the users with activities on files in this folder are members of the Finance department.”



Behavioral requirements for rule

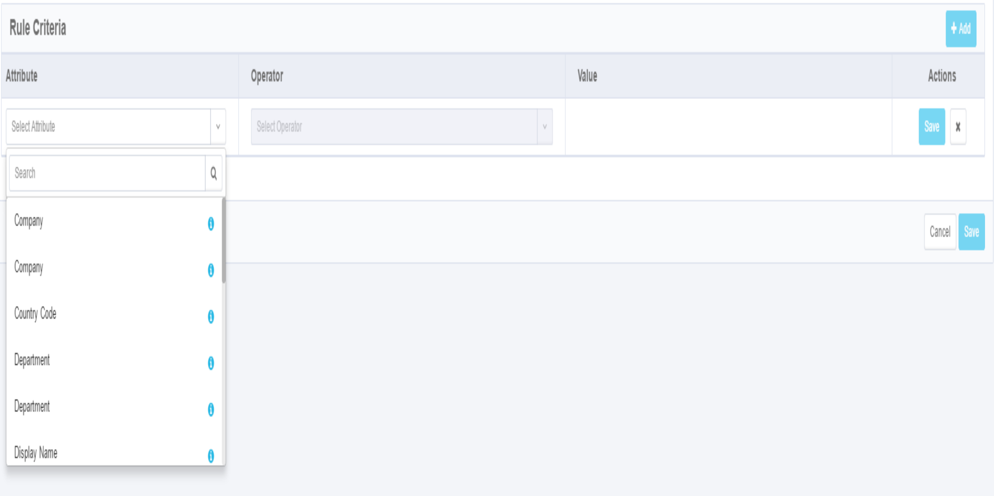
This pattern should match **at least** 15 % of the activities over last Day *(Valid values for % of activities are whole*

4. Define the timeframe and required usage:

   - *Value* - Percentage required to meet the rule.

   - *Timeframe* - The timeframe during which to check the rule

5. In the Rule Criteria section, add the general details to the Behavioral Based Classification rule.



6. Select an Attribute, an Operator, and a Value (optional) from the dropdown menus.

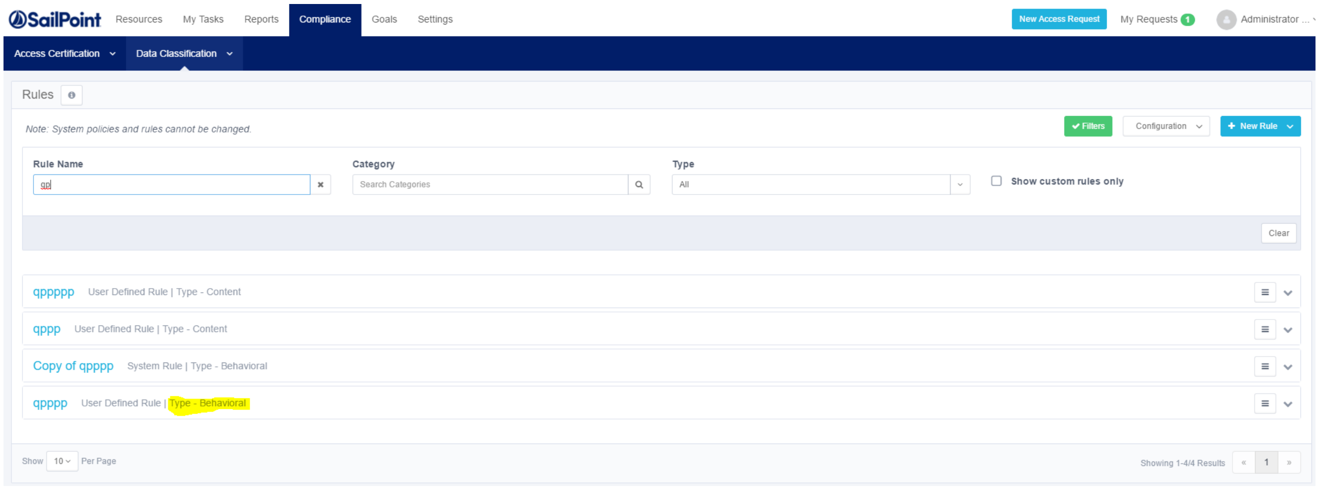
7. Create an expression and click **Save**.

> Note: Users can edit or delete existing rule criteria.

8. Add additional rule requirements as needed.

9. Click **Save** to save the new content-based rule.

10. The system adds the rules to the Rules list.



Note: System policies and rules cannot be changed.

Rule Name: qp

qppppp User Defined Rule | Type - Content

qppp User Defined Rule | Type - Content

Copy of qpppp System Rule | Type - Behavioral

qpppp User Defined Rule | Type - Behavioral

Scheduling Classify Behavioral Rules Task

The Classify Behavioral Rules task is a global scheduled task. It is created out of the box, and is disabled by default. This task runs on all scope on supported applications.

All applications, besides the ones listed below, support the classify behavioral rules task.

File Extension	Expected file type
NFS (Generic)	No activities
CIFS (Generic)	No activities
Generic database table	No file, but if one is added it will work
SQL Server DB	No file
AD	No file
DFS	No activities
Home-Grown	No activities
NFS (Generic)	No activities

To schedule the Classify Behavioral Rules task:

1. Navigate to **Settings > Task Management > Scheduled Tasks**.
2. Select the task Classify Behavioral Rules on the Scheduled Tasks table on the tickbox on the task row. This will open the options buttons.
3. Click **Edit** to open the scheduling edit panel.
4. Set the task to active / inactive.
5. Change the scheduling parameters as required.

Composite Rule

A composite classification rule lets you combine several rules together to form a more complex criterion. This can include content and behavioral type rules, and is defined by category.

- The data classification matches content or behavioral patterns to rules, and assign categories to resources according to these rules.
- After running data classification, composite rules use combinations of categories to define complex combinations of simple rules

Examples:

> You can combine Personal Identification Information (PII) in conjunction with health-related information (ICD), to define a rule to identify Personal Health Information (PHI).
>
> or
>
> You can create a rule to list files that have at least two out of one list of categories, and must contain another specific category.
>
> or
>
> Identify all resources that would be defined by rules that belong to category **X**.

To define a composite classification rule, select one or more categories, and the created rule will be triggered for any existing rules within the selected categories.

In the first example above, if we define a rule as follows:

```
Contain at least 2 of  PII, ICD
```

This will add all business resources that fill any of the rules in *PII category*, and any of the rules in *ICD category*.

- The value column allows selecting one or more categories from the category repository.

Triggering the Composite Rules

- Composite rule tasks are trigger after each data classification task, and evaluate results from that application only.
- The Composite rule runs after of all content and behavioral rules, as it is based on their results.
- If you change a composite rule, this change will take effect only when a new classification task is executed, and

triggers the composite rule.

- This task can not be scheduled.

Creating a Composite Classification Rule

To create a composite classification rule:

1. Open the rules page

   Compliance > Data Classification > Rules

2. Click **+ New Rule** > Composite Classification Rule

   Rule Name

   Rule names are unique. It is best to create a naming convention that avoids using the same name twice.

   Categories

   Enter one or more categories for the rule

   To add a new category to the Categories list, click **Manage Categories** and add a new item

3. In the Rule Criteria section, add the desired combination of categories to trigger the rule.

   Operator

   Enter the number of concurrence of categories in the business resource tested for this criterion.

   For example, if you went the rule to collect BRs that fit both criteria C1 and C2, set

   Operator: Contain at least 2 of

   Value: C1, C2

   Value

   Enter one or more categories from the search box

4. Click **Save** to save the criterion.

5. Click **+ Add** to add another criterion. All criteria will be combined with an AND operator.

6. Click **Save** to save the new content-based rule.

7. The system adds the rules to the Rules list.

Global Rules

Global Rules are policy-level classifications, designed to enable complex searches for advanced classification scenarios.

They allow Compliance Managers and users to define content classifications criteria that span across multiple data points and categorizes or labels files only when they span multiple data dimensions and satisfy multiple rules.

Thus, Global Rules allow Compliance Managers to apply more sophisticated policies to classify their data more accurately, get more targeted results, and reduce compliance clutter. By supporting adjustable thresholds for classifications, Global Rules give compliance professionals the flexibility and agility they need to tailor their compliance suites to the needs of their organization, their internal criteria, and regulatory requirements.

> Note: Policies can only contain one Global Rule.

> Note: Policies provided by SailPoint like the GDPR and PII Policies, have built-in Global Rules. These can be adjusted based on the organizations needs.

> Note: Users are able to add and adjust Global Rules to all user-created policies.

> Note: Adding or removing Global Rules from SailPoint delivered, out of the box, policies is possible.

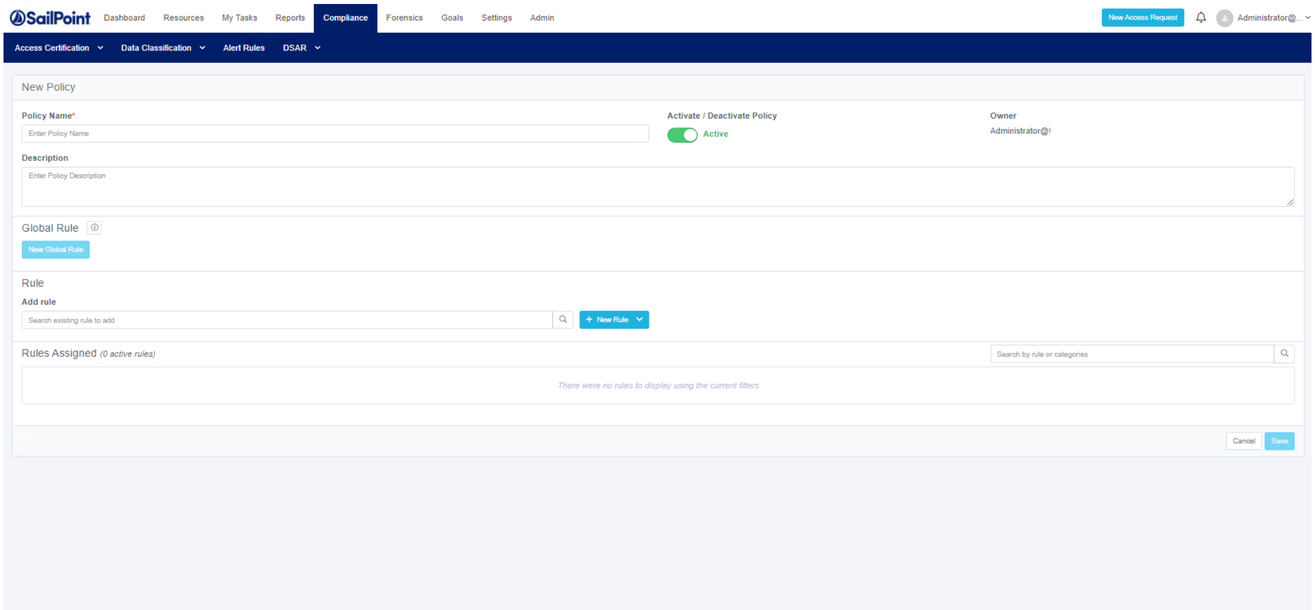
Creating a Global Rule

To create a Global rule within a user-defined policy, perform the following steps:

1. Navigate to **Compliance > Data Classification > Policies** to open the rules page.

2. Click **+ New Policy**.

   A new Policy window displays.



New Policy

Policy Name*
Enter Policy Name

Activate / Deactivate Policy
Active

Owner
Administrator@!

Description
Enter Policy Description

Global Rule

Rule
Add rule
Search existing rule to add

Rules Assigned *(0 active rules)*
Search by rule or categories

There were no rules to display using the current filters

The available Policy fields include:

Policy Name

Policy names are unique. It is best to create a naming convention that avoids using the same name twice.

Activate/Deactivate Policy

Users can activate or deactivate a policy using this button

Owner

The logged in user is the creator of the policy. (This field is read-only.)

Description

Users can provide descriptions to the policies they create to better explain what the policy is meant for and designed to do. You can use this description to describe the logic of your Global rules, as well for additional readability.

> Note: To add a Global rule to a policy, a user must first add at least one Content rule. This can be done by either adding a new rule or by adding a pre-existing rule using the Search option.
> We recommend adding Global rules at the end after the Content rules have been added.

> Note: Global rule settings are enabled once Content rules have been added.

3. After adding at least one Content rule, select **New Global Rule**.

   A New Global Rule window displays.

   Provide the following information:

   Rule Name

   Unique name for the Global rule

   Categories

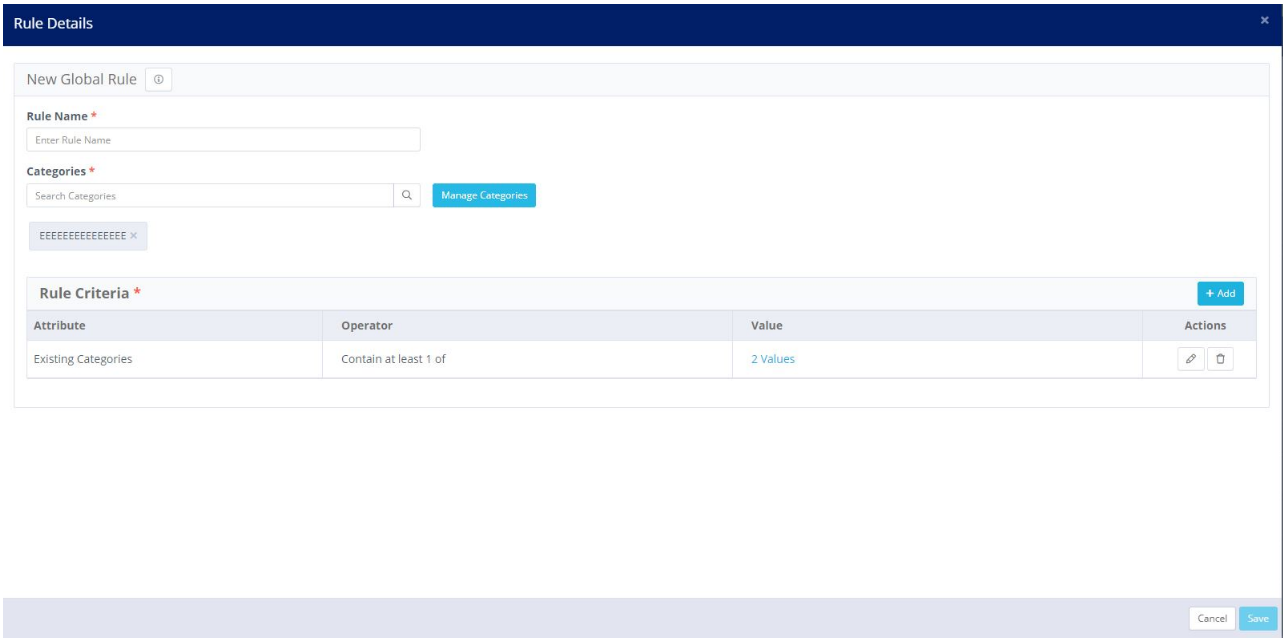
   The resource will be classified by the Categories when the rule is satisfied

4. For the Rule Criteria, add the type of categories in the Value field. The dropdown will provide a list of only the categories that are set by the content rules defined within the policy.

   Creating rule criteria allows you to set a condition that will be satisfied when a set of categories are applied.

5. Select **Save** within the Rule Criteria block.

6. Either add another set of rule criteria or click **Save** to save the new Global rule.



Rule Details

New Global Rule

Rule Name *

Enter Rule Name

Categories *

Search Categories

Rule Criteria *

Attribute	Operator	Value	Actions
Existing Categories	Contain at least 1 of	2 Values	

Global Rule Options

Within the Policy screen, a user can either Edit or Remove a Global Rule from the policy.

Select the hamburger menu within the Global rule to see these options.

- Edit–all options can be edited
- Remove–the Global rule will be removed from the policy

> Note: Global rules cannot be created or removed from out of the box policies. They can only be edited.

> Caution: If you remove a Global rule from a policy, you cannot add it back. I new Global rule will have to be created and added to the policy.

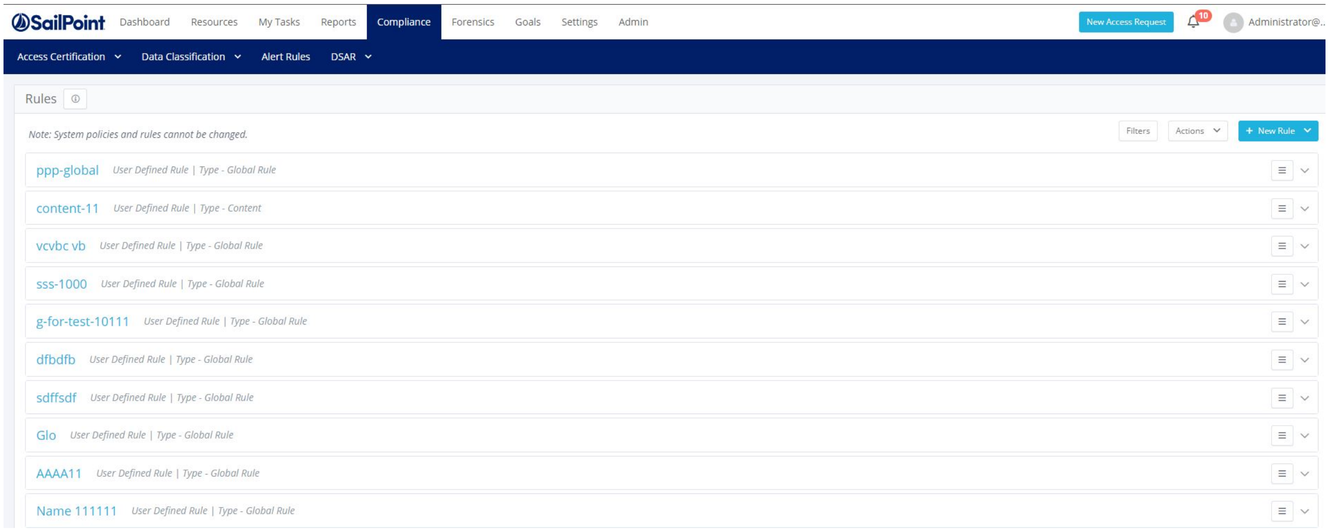
Global Rule – Rules Screen

By navigating to the Rules screen (**Compliance > Rules**), a user can see all of the various rules including global rules that have been created. These rules will also state what type of rule they are.

From the Rules screen, Global Rules can only be deleted (with the exception of OOTB global rules). You cannot edit a Global Rule from this screen.

> Note: If a Global Rule is deleted from the Rules screen, it will automatically be deleted from any corresponding policy.

Next to the hamburger menu is a downward arrow. A user can select this arrow to view details, such as categories and criteria, about the Global Rule.



Note: System policies and rules cannot be changed.

ppp-global User Defined Rule | Type - Global Rule

content-11 User Defined Rule | Type - Content

vcvbc vb User Defined Rule | Type - Global Rule

sss-1000 User Defined Rule | Type - Global Rule

g-for-test-10111 User Defined Rule | Type - Global Rule

dfbdfb User Defined Rule | Type - Global Rule

sdffsdf User Defined Rule | Type - Global Rule

Glo User Defined Rule | Type - Global Rule

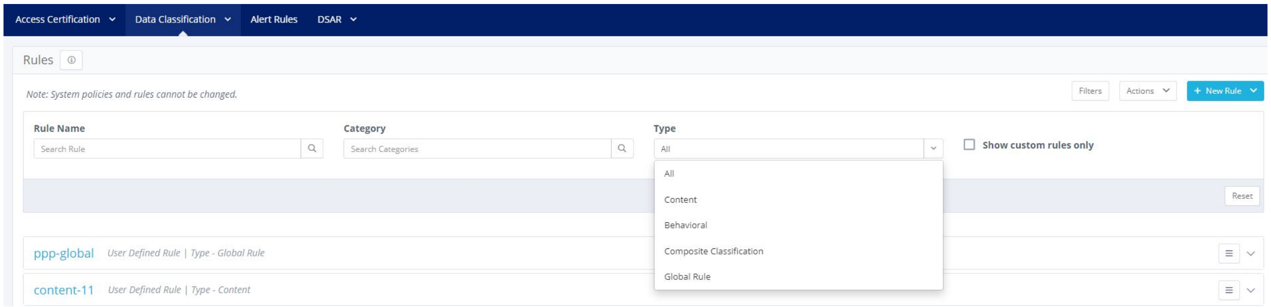
AAAA11 User Defined Rule | Type - Global Rule

Name 111111 User Defined Rule | Type - Global Rule

Filter

1. To view only Global Rules, select the Filter icon.

2. In the Type dropdown, select Global Rules.



Note: System policies and rules cannot be changed.

ppp-global *User Defined Rule | Type - Global Rule*

content-11 *User Defined Rule | Type - Content*

Data Classification Verification Algorithms

You can use verification algorithms in a Data Classification policy object of type Regular Expression to filter the regular expression results. This will enforce additional restrictions and validations on matched phrases. The verification algorithm will take as an input each one of the data classification policy objects' regular expression match result strings, and will remove results that do not meet the criteria defined within the algorithm.

File Access Manager comes with a set of verification algorithms out of the box for standard verifications, such as Luhn, for credit card numbers, or SSN algorithms. In addition, you can write a verification algorithm, upload it to the File Access Manager website, and use it in data classification policy objects.

Out of the Box Verification Algorithms

Verification algorithms for common rules are pre-loaded in File Access Manager:

- Luhn (Credit Card Number)
- US SSN
- Netherlands BSN
- Israeli ID
- IBAN
- South African ID

The dropdown list of verification algorithms in the Rule Criteria screen includes out of the box algorithms, as well as algorithms uploaded by the user.

Creating a Verification Algorithm

Guidelines

- The assembly must target .NET Standard 2.1 or .NET 6.0. These will be referred to as the supported .NET platforms.
- You may write only one implementation class of the IDataClassificationVerifier interface per assembly.
- It is only possible to upload one assembly per verification algorithm. In case your code requires usage of additional referenced assemblies, you must pack them all into one assembly.

> Note: Verification algorithm assemblies written in previous versions of File Access Manager (in .NET Framework 4.5) must be removed, and re-written to target one of the supported .NET platforms as mentioned above, and uploaded again.

Walkthrough

1. Create a new .NET Framework Class Library targeting a supported .NET platform.
2. In your project, add a reference to the assembly FAM.DataClassification.Verifiers.dll. This assembly is provided by SailPoint, and contains the IDataClassificationVerifier interface. This assembly can be downloaded from Compass.
3. Create a new class that implements the IDataClassificationVerifier interface.
4. This class must provide an implementation of the only public method defined in the interface named “Verify.” This method takes as an argument a match result string and returns a boolean that denotes if the verification passed or failed.
5. Build your project, and upload the output assembly as described in Verification Algorithms screen.
6. This uploaded verification algorithm will now be available in the verification algorithm dropdown list of the Policy Object screen, alongside the other built in or uploaded algorithms.

Examples

Below is an example of code to create a verification dll that verifies that the number passed is even.

```
using FAM.DataClassification.Verifiers;

namespace VerificationAlgorithmExample

{

  public class EvenNumberVerificationAlgorithm : IDataClassificationVerifier

  {

    /// <summary>

     /// Example for a custom verifier that verifies that the input is an even number

    /// </summary>
    /// <param name="value">A regular expression match result</param>

    /// <returns>True if passed verification, False if failed</returns>

    public bool Verify(string value)

    {

        if (long.TryParse(value, out long parsedLong))

        {

            return parsedLong % 2 == 0;

        }

        return false;

    }

  }

}
```

Verification Algorithms screen

Description

The Verification Algorithms table shows the custom verification algorithms uploaded by the users, or as part of a policy upload from another File Access Manager system.
This table does not contain the standard out of the box verification algorithms.

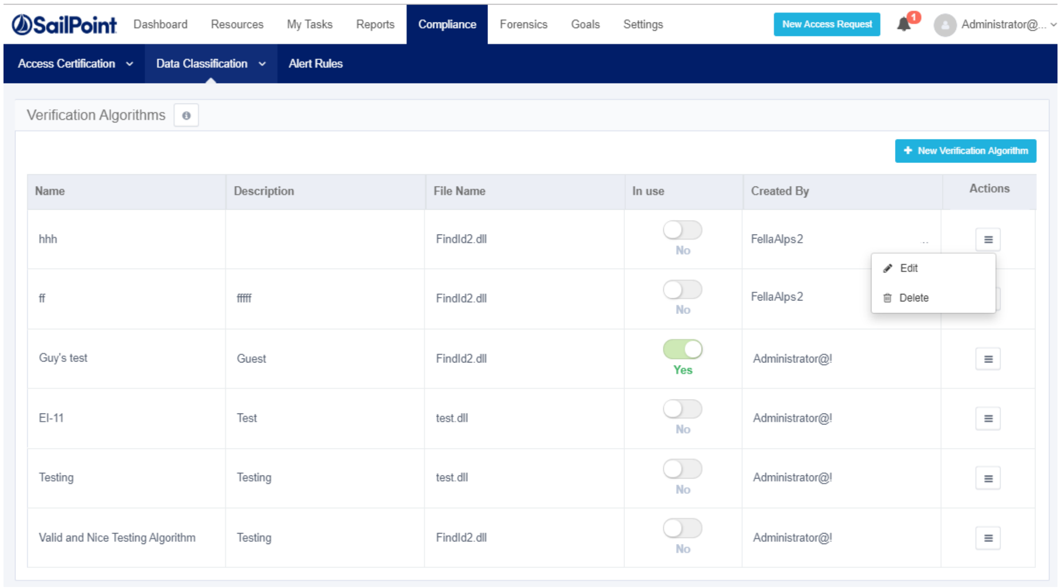
Access

File Access Manager website:

Compliance > Data Classification > Verification Algorithms

Permission

By default, this page is accessible only to Administrators.



Name	Description	File Name	In use	Created By	Actions
hhh		FindId2.dll	No	FellaAlps2	≡
ff	fffff	FindId2.dll	No	FellaAlps2	
Guy's test	Guest	FindId2.dll	Yes	Administrator@!	≡
EI-11	Test	test.dll	No	Administrator@!	≡
Testing	Testing	test.dll	No	Administrator@!	≡
Valid and Nice Testing Algorithm	Testing	FindId2.dll	No	Administrator@!	≡

Table fields

Name

Verification algorithm name. This name will also appear in the dropdown list of verifications, along with the existing, out of the box verification algorithms.

Description

Added when the verification algorithm is uploaded.

File name

The verification algorithm dll file created by the user and uploaded to File Access Manager.

In use

This flag indicates whether this algorithm is part of a policy object, that is used in an active policy.

Created by

The user uploading the algorithm. Verification algorithms that are uploaded to the system using the policy upload tool, will be listed in the verification algorithms list as Created By “Conversion.”

See Transferring Data Classification Policies Between Systems for further details on imported policies.

This screen can be used to:

- View custom built verification algorithms.
- See whether an algorithm is in use
- Edit an algorithm details: Update the name, upload a new file or update the description.
- Upload new verification algorithms (See below how to create an algorithm dll),
- Delete verification algorithms.

Uploading a New Verification Algorithm

A new verification algorithm must follow the guidelines below:

- Extension: .dll
- File size: Up to 5 MB
- The verifier name must be unique in the list of verification algorithms.

1. Open the Verification Algorithms panel.
2. Click **+ New Verification Algorithm**.
3. Select **File**.
4. Select a .dll file from your computer.
5. Enter the name and description of the verification algorithm (see description above).

   Name

   Verification algorithm name. This name will appear in the dropdown list of verifications, along with the existing, out of the box verification algorithms

   Description

   Free text description of the verification algorithm.
6. Click **Save** or **Cancel** to continue.

Deleting a Custom Verification Algorithm

1. Open the Verification Algorithms panel.

2. Click the menu icon on the row of the verification algorithm you want to delete, to open the action menu.

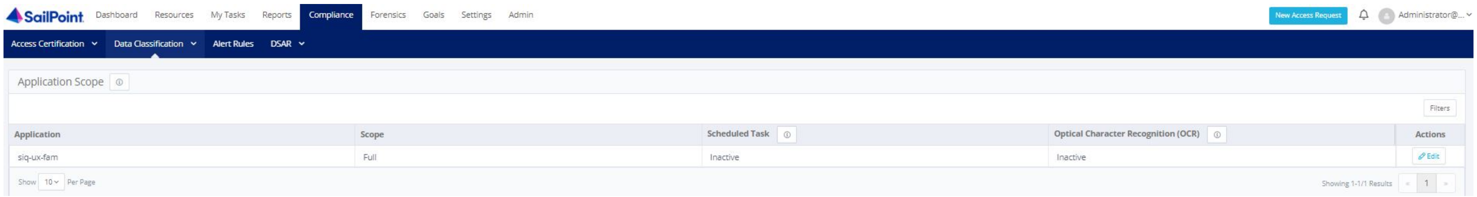
3. Click **Delete**.

> Note: If the algorithm is currently part of a policy object that is used in an active policy, a popup message will warn the user before deleting.

Application Scope

Use this screen to view and set the scope of applications and resources on which to apply Data Classification policies.

In the web client, navigate to **Compliance > Data Classification > Application Scope**.



Application Scope

Application	Scope	Scheduled Task	Optical Character Recognition (OCR)	Actions
siq-ux-fam	Full	Inactive	Inactive	Edit

Show 10 Per Page

Showing 1-1/1 Results « 1 »

The scope list includes only applications with installed Data Classification. It is only possible to install Data Classification on an application in the administrative client.

The column Optical Character Recognition (OCR) indicates whether the application has OCR activated on part or all part of its resources.

The scope definition directly affects the time required for Data Classification indexing.

Activating Optical Character Recognition on resources is a resource intensive process, and should be configured carefully.

For example, to reduce Data Classification indexing, an Administrator can:

- Exclude an application from Data Classification indexing (if “non-sensitive” data is saved by default on that application).
- Include a specific resource (one with very important data) for Data Classification indexing.

You can specify which resources to use (and which to exclude) from a selected application by clicking the Edit button to the right of the application.

While the Data Classification status (Active/Inactive) can only be changed from the administrative client, non-administrator users can view the status in the Web application.

The Scope definition only takes effect after the next run of the Data Classification task.

Only the business resources of an application selected for editing display on the list.

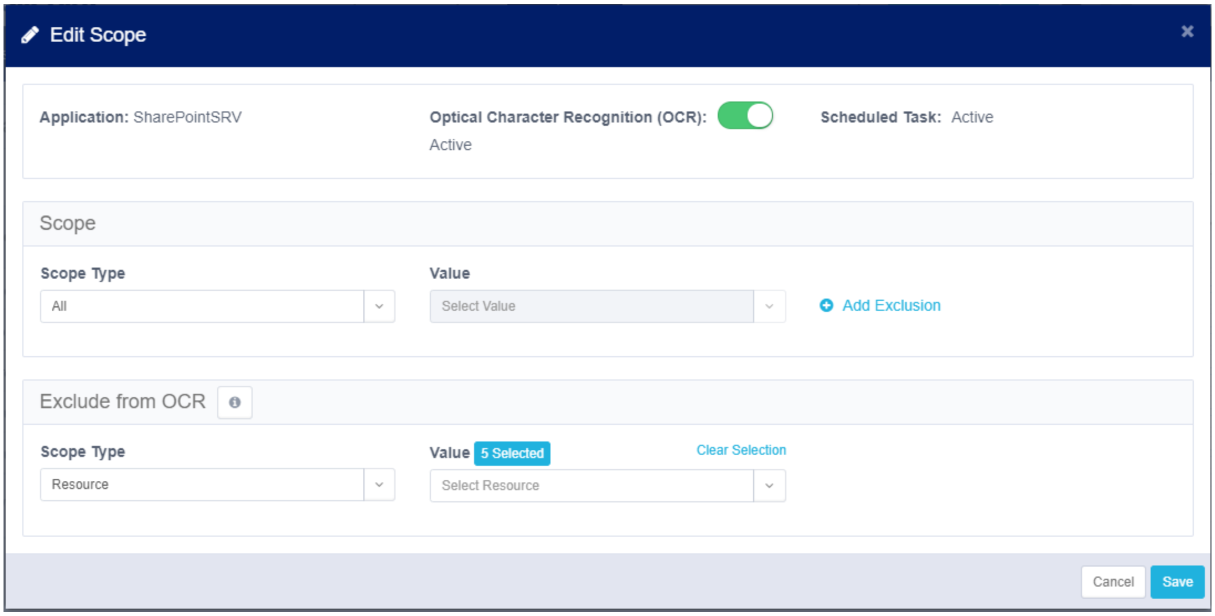
Editing the Application Scope

1. Click **Edit** to modify the scope (such as folders), and/or the OCR setting of the Data Classification per application.

2. Find the desired application from the Application Scope screen.

3. Click **Edit**.

This will open the Application Scope Edit screen.



Edit Scope

Application: SharePointSRV

Optical Character Recognition (OCR): Active

Scheduled Task: Active

Scope

Scope Type: All

Value: Select Value

Exclude from OCR

Scope Type: Resource

Value: 5 Selected

Select Resource

To change the scope to include in the Data Classification process:

1. Select the scope type.

   - All – Run Data Classification on all the resources in the application

   - Resource – select from a list of resource to include

To exclude resources from the Data Classification process:

a. Click **Add Exclusion** to open the exclusion entry field.

b. Select resources to exclude from the dropdown list.

To remove the exclusion of resources from the Data Classification process:

a. Click **Remove Exclusion**.

To enable OCR:

a. Click **Optical Character Recognition (OCR)** to enable / disable OCR analysis for this application.

b. Select the resources to exclude from the OCR analysis from the drop down resource tree.

> Note: All resources selected in the Data Classification scope will include all subfolders (or parallel resources, per application) as well. The checkbox “Including subfolders” cannot be unselected.

> Note: Changes to the scope or activating the OCR on an application will trigger a re-indexing in the next run of the Data Classification task.

> Note: Deactivating OCR on an application will not trigger re-indexing.

Section Scope has additional information on scope inclusion and exclusion.

Policy Scope

Use this page to view and define the Data Classification scope of your existing policies and adjust each one by specific applications or application types.

By default, every policy will include all applications and application types in the scope.

1. In the web client, navigate to **Compliance > Data Classification > Policy Scope**.

2. Select the **Edit** icon on the desired policy.

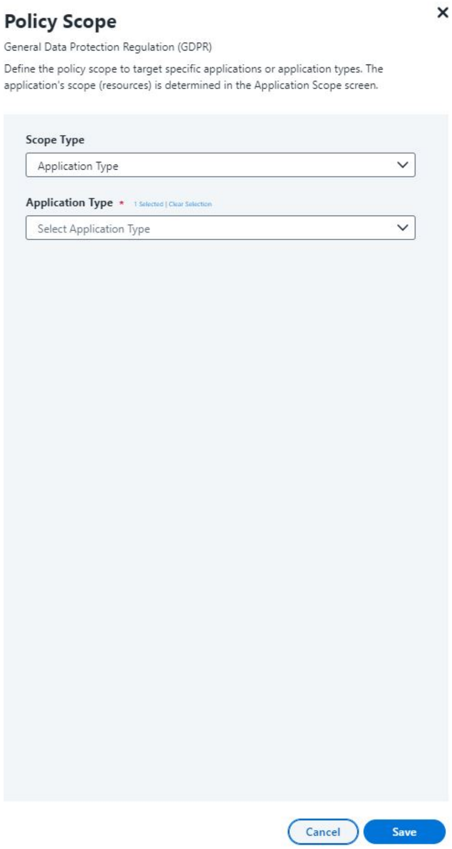
   The Policy Scope overlay displays. From here, the Scope Type and the Application Type or specific Application can be edited.

   If Application Type is selected, supported application within File Access Manager will display.

   If Application is selected, a list of the added applications will display.

   > Note: If a Scope Type is selected, the Application Type of Application field is mandatory.

3. Select **Save**.



Policy Scope

General Data Protection Regulation (GDPR)

Define the policy scope to target specific applications or application types. The application's scope (resources) is determined in the Application Scope screen.

Scope Type: Application Type

Application Type * 1 Selected | Clear Selection: Select Application Type

After saving the policy scope, a task is created.

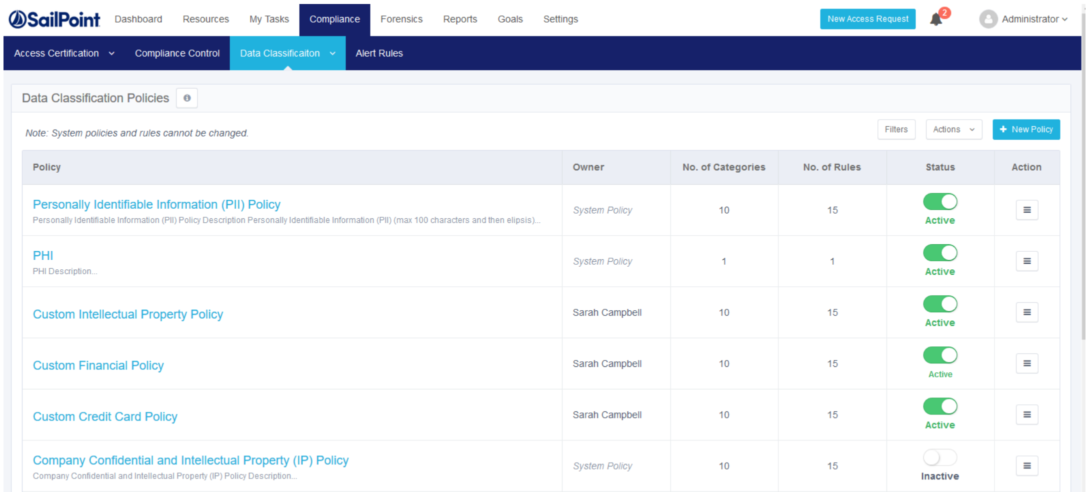
4. Rescan all applications associated with the updated policy.

Run Resource Classification

Use this feature to run the Data Classification process on a specific business resource, rather than on an entire application. You can test the Data Classification process faster, since you will only be testing a single resource. In addition, you can run Data Classification faster on a single sensitive resource (for example, one on which many changes were made), than on multiple resources.

To run Resource Classification, perform the following step:

Navigate to **Compliance > Data Classification > Policies or Rules > Actions > Run Resource Classification**.



Data Classification Policies

Note: System policies and rules cannot be changed.

Policy	Owner	No. of Categories	No. of Rules	Status	Action
Personally Identifiable Information (PII) Policy Personally Identifiable Information (PII) Policy Description Personally Identifiable Information (PII) (max 100 characters and then elipsis)...	*System Policy*	10	15	Active	≡
PHI PHI Description...	*System Policy*	1	1	Active	≡
Custom Intellectual Property Policy	Sarah Campbell	10	15	Active	≡
Custom Financial Policy	Sarah Campbell	10	15	Active	≡
Custom Credit Card Policy	Sarah Campbell	10	15	Active	≡
Company Confidential and Intellectual Property (IP) Policy Company Confidential and Intellectual Property (IP) Policy Description...	*System Policy*	10	15	Inactive	≡

Import Data Classification Results

To import external data classification results, select the data source that contains the results. This data source must have the following fields:

- Category
- Application name
- Full path
- File name

An additional field, **match count**, is optional.

> Note: The final content of the data classification table might contain duplicate categories, if the import process, and data classification process contain identical categories. These are added as additional lines in the table.

To import data classification results from other data sources:

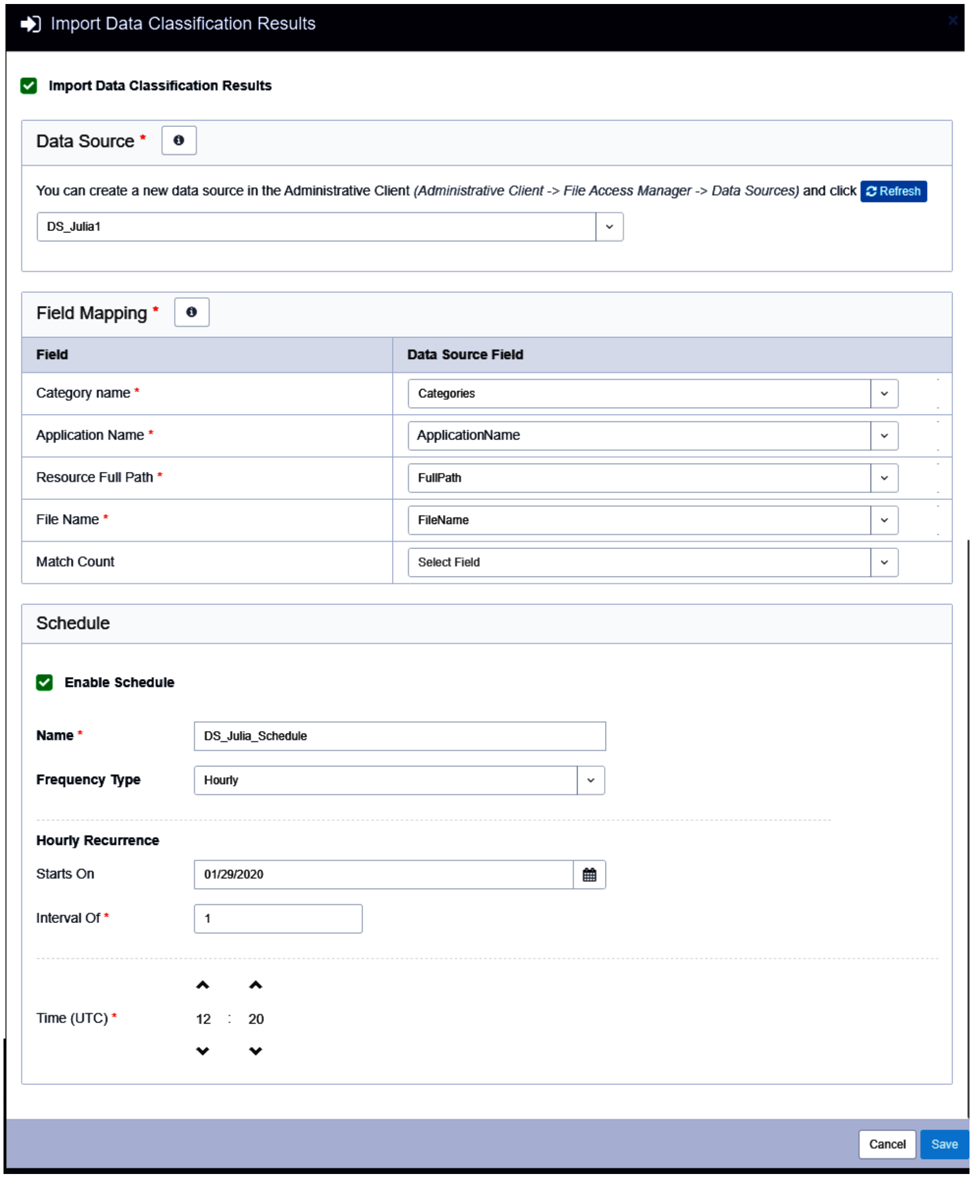
1. Open the File Access Manager website
2. Navigate to **Compliance > Data Classification > Policies**
   or
   Compliance > Data Classification > Rules
3. Click the **Actions** menu > **Import Data Classification results**
4. Configure the import fields by mapping the data source filed to the File Access Manager fields.

> Note: The import task will import the match count from the external source, in addition to the fields of the categories. The match count field is imported as a number. If the field mapped to Match Count is empty, or is not a number, the process will load a null into this field.Set a schedule for refreshing the database from the external source.
> The schedule can be any of the following frequency types:

- Once
- Daily

- Weekly - default value
- Monthly

5. Click **Save** to store the field mapping and scheduling.



Import Data Classification Results

☑ **Import Data Classification Results**

Data Source *

You can create a new data source in the Administrative Client *(Administrative Client -> File Access Manager -> Data Sources)* and click Refresh

DS_Julia1

Field Mapping *

Field	Data Source Field
Category name *	Categories
Application Name *	ApplicationName
Resource Full Path *	FullPath
File Name *	FileName
Match Count	Select Field

Schedule

☑ **Enable Schedule**

Name * DS_Julia_Schedule

Frequency Type Hourly

Hourly Recurrence

Starts On 01/29/2020

Interval Of * 1

Time (UTC) * 12 : 20

To follow the task progress, go to **Settings > Tasks Management > Tasks >** “Value from the Scheduled task name field” task.

To cancel the setup of the import, close the window.

Data Classification Results

Data Classification Results – Report

File Access Manager provides reports of data classification results. You can filter the results be various parameters, including a “match count” – having a certain sensitive category at either more than, or less than a given threshold.

To generate data classification report, perform the following steps:

1. In the web client, navigate to **Reports > Report Templates**.
2. Use the **Classified Data** tag to locate a specific report.
3. To apply a different filter than one of the existing templates:
   a. Create a duplicate template by selecting **Duplicate** from the template drop down menu
   b. Set the filter parameters, and **run now** or **Save** the template for future runs.
4. The report will be available in the **My Reports** screen.

> Note: File Access Manager can import data classification results using the Import Classification Result capability. This allows for results to be imported from a created data source.

Data Remediation Policy

A Data Remediation policy is a set of policy rules, which govern actions that are run on the basis of the Data Classification process results.

Each File Access Manager deployment has a data remediation policy that spans all the deployment's applications.

Each Data Remediation rule consists of:

- Categories - the data classifications of a file that triggers the specific rule
- Scope - whether the rule should be triggered by application, by application type, or should not be limited by either [unlimited]
- Script path - The path to a script to be executed on the files that match this category.

> Note: The script must be written in PowerShell and can accept both the filename and the category as parameters, and return an error message in case it fails.

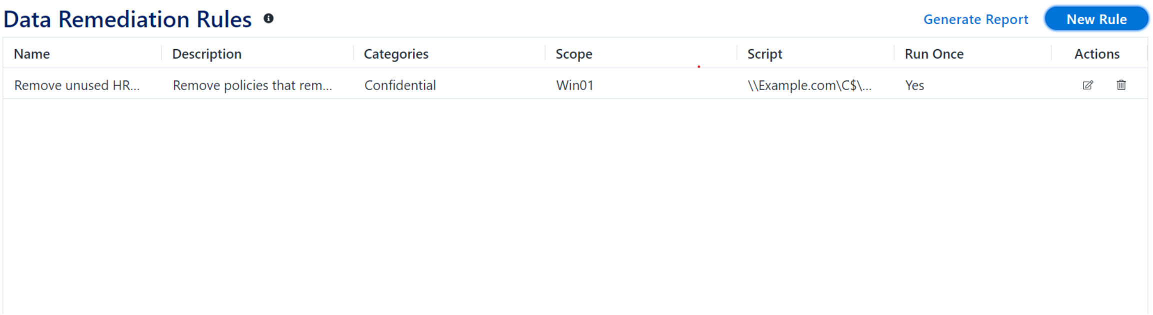
A Data Remediation script is executed on a file that matches one of the Data Remediation rules. Each rule can run a single script.

The Data Remediation scripts are executed by the installed Application's Data Classification service. The service periodically queries the database for new scripts which are pending for execution, and in turn executes them and writes the execution results to the logs.

You can track the execution of the Data Remediation rules by generating log reports.

To set a Data Remediation Policy:

1. Navigate to **Compliance > Data Classification > Data Remediation**



Data Remediation Rules

Name	Description	Categories	Scope	Script	Run Once	Actions
Remove unused HR...	Remove policies that rem...	Confidential	Win01	\\Example.com\C$\...	Yes	✎ 🗑

2. The data remediation has the following options:

   a. **Generate Report**: Run or schedule a report based on the remediation rules, according to the requested time period.

   b. **New Rule**: Create a data remediation rule

   c. Each Data Remediation line has the options **Edit** and **Delete**.

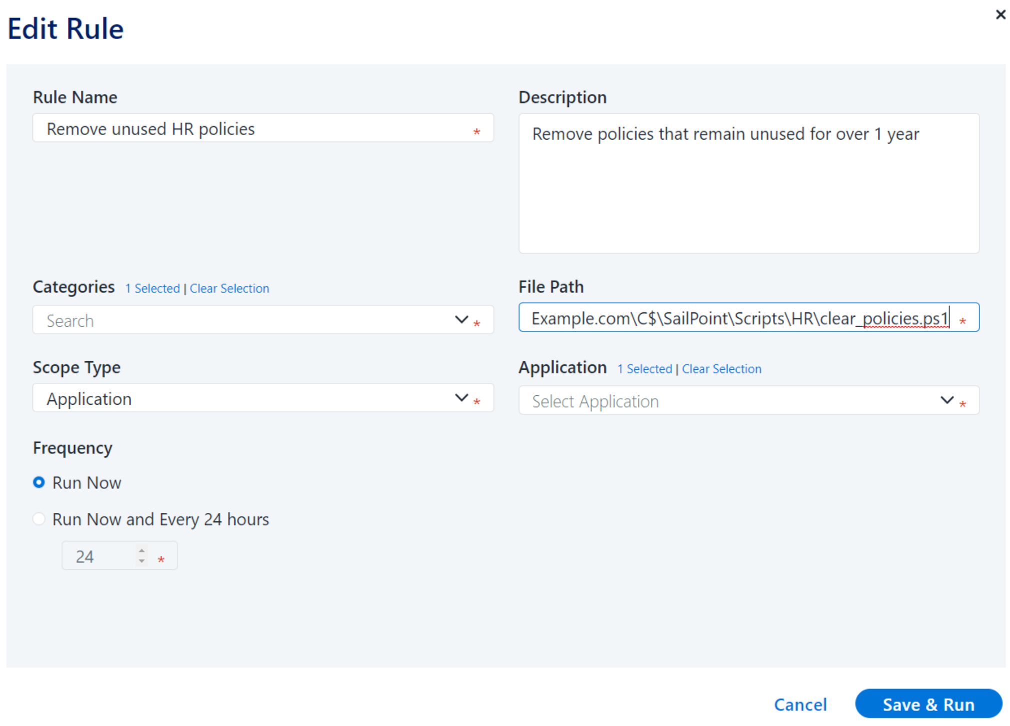
> Note: Data Remediation allows you to run any operation on classified files. This also includes encrypting.

Create a Data Remediation Rule

To set a new Data Remediation rule:

1. Navigate to **Compliance > Data Classification > Data Remediation**

The New Data Remediation Rule screen displays.



Edit Rule

Rule Name: Remove unused HR policies *

Description: Remove policies that remain unused for over 1 year

Categories 1 Selected | Clear Selection: Search *

File Path: Example.com\C$\SailPoint\Scripts\HR\clear_policies.ps1 *

Scope Type: Application *

Application 1 Selected | Clear Selection: Select Application *

Frequency

● Run Now

○ Run Now and Every 24 hours

24 *

2. Fill in the following fields:

- *Rule Name* (mandatory)
- *Description*
- *Categories*—select at least one category from the dropdown list.
- *Script Path*—the path to the PowerShell script to run. Since the script is executed by the data classification service, the path must be relative to the server in which the data classification service is installed. If this action will be run by multiple data classification services serving different applications, all services must be able to access the path.
- *Scope Type*: Select the scope to apply to the rule by selecting one of the following:
  - All (default)
  - Application Type
  - By Application
- *Application*: Select one or more applications by marking the tickboxes in the dropdown list.
- *Application type*: Select one or more application types by marking the tickboxes in the dropdown list.
- *Frequency*: Select an execution interval.
  - Run Now—one time run.
  - Run now and Every X Hours—the default is 24 hours. Set an interval between 1-99 hours.

3. Click **Save & Run**, or **Cancel**.

Edit a Data Remediation Rule

To edit a Data Remediation rule:

1. Navigate to **Compliance > Data Classification > Data Remediation** [Select policy] **Edit Rule icon**.
2. The Edit Data Remediation Rule screen displays. Follow the steps described above.
3. Click **Save & Run** or **Cancel**. If you click **Save & Run** at any stage of editing a data remediation rule, it will cause the assigned actions to execute immediately. This is true even if no changes were made.

Delete a Data Remediation Rule

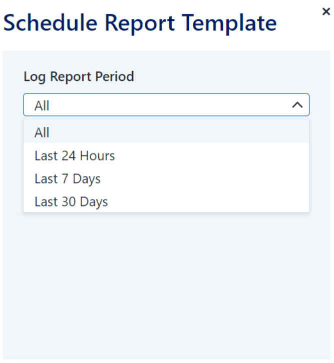
To delete a Data Remediation rule, navigate to **Compliance > Data Classification > Data Remediation** [Select policy] **Delete Rule icon**.

Log Reports

You can track the execution of the Data Remediation rules and actions by generating log reports.

To view Data Remediation reports navigate to **Compliance > Data Classification > Data Remediation**.

1. Click the **Generate Report** menu option. This will open the report dialog box.



Schedule Report Template

Log Report Period: All

2. Click **Produce Now** to produce the report now, or click **Schedule a New Report** to schedule the report.

3. If you selected Schedule a New Report in the previous step, select one of the following scheduling options:

    a. Last Day

    b. Last 7 Days

    c. Last 30 Days

    d. All

Writing a PowerShell Script for Data Remediation

The File Access Manager administrator must provide a path to a valid script to perform the desired action.

That script must be written in PowerShell and return either nothing (or an empty string) to indicate success, or a string message to specify an error in case of failure.

The script receives 2 parameters when its executed:

- A string which represents the full path of the file upon which the action should act
- A string which represents the category which caused the action to be executed

Any credentials needed for the script to operate must be provided within the script.

Transferring Data Classification Policies Between Systems

File Access Manager provides an easy way to transfer data classification policies from one system to another, through a command line interface. Administrators can use the import/export tool to import/export custom policies from one server to another.

> Note: Importing Data Classification Policies can only be done between versions listed in the Data Classification Importer section within Import Data Classification Policies

> Note: You must be defined as an Administrator in the File Access Manager administrative client.

> Note: You can only execute the import/export tool in its file working directory.

To run the Import/Export tool, perform the following steps:

1. Use the Windows command line to navigate to the following directory:

   CD % SAILPOINT_HOME%\FileAccessManager\Server Installer\Tools\PolicyExporter

   CD % SAILPOINT_HOME%\FileAccessManager\Server Installer\Tools\PolicyImporter

2. In the Windows command line, type:

   ```
   cd {path to the tool directory}
   ```

   ```
   PolicyExporter.exe {options}
   ```

   OR

   ```
   PolicyImporter.exe {options}
   ```

   > Note: The tool argument can be a minus sign (-) followed by a letter in upper case, or two minus signs (--) followed by a word in lower case letters.

   For example:

   ```
   -U DOMAIN\USER
   ```

OR

```
--user DOMAIN\USER
```

The tool validates arguments before performing any action, and the system alerts the user if one or more arguments are missing or are invalid. If you do not provide arguments, a Help screen displays.

Each Data Classification Policy is assigned with a unique global ID (GUID). When new policies are imported, File Access Manager compares the GUID's on both policies to identify them uniquely.

> Note: While the name of the tool is Import/Export, the procedural order is to export data classification policies first.

Exporting Data Classification Policies

Data classification policies are exported with their rules, policy objects, categories, file properties, and rule criteria. The tool transfers an output file to the target server for import. The tool also creates a log file, which File Access Manager technical support team can use as a reference for troubleshooting.

If a policy object includes a verification algorithm created by the user, this dll file will be exported as well.

As noted in Transferring Data Classification Policies Between Systems , you must have administrative rights in File Access Manager and use the file working directory.

To export data classification policies, perform the following steps:

1. Run the tool with the following selected options:

   a. -O, --output (Default: output_policies.bin) (Output file location)

      > Note: The output file is in binary format and cannot be edited.

      The file location can be both either absolute (c:\program files\Sailpoint\outputs) or relative (..\..\outputs).

   b. -A, --all

   c. The tool exports all policies available from current system.

   d. -L, --policies

      The tool exports specific policies (each policy specified by its policy name (not case sensitive) and with a comma separating the name of one policy from the other.

      Policy names that contain spaces ( ), should be in quotation marks (“) Example: PolicyExporter.exe -U domain\user -L “policy1 – my policy”,”POLICY2 – HIS POLICY”

> Note: Select either -A or -L, since they are mutually exclusive.

e. -U, --user (Required.)

2. This is the name of the user to whom data classification policies are exported, and should include both the user name and the domain name (if there is one).

   a. -P, --password

   b. The user password validates the export. The system will only prompt you three times to provide a password.

   c. --help

   d. The Help screen displays.

   e. –version
   The version information displays.

Import Data Classification Policies

Data classification policies are exported with their rules, policy objects, categories, file properties, and rule criteria. The tool creates a file with a summary of what was imported and what was not imported. The tool also creates a log file, which File Access Manager technical support team can use as a reference for troubleshooting.

As noted inTransferring Data Classification Policies Between Systems, you must have administrative rights and use the file working directory.

To import data classification policies, perform the following steps:

> Note: The only way to run an import or export on the tools is by the command line.

1. Run the tool with the following selected options:

   a. CD %SAILPOINT_HOME%\FileAccessManager\Server Installer\Tools\PolicyImporter

   b. -I, --input (Input file location)

   c. The exported output file path

      The file location can be either absolute (c:\program files\Sailpoint\outputs) or relative (..\..\outputs).

d. -R, --override (Default: false)

The system recognizes a policy by its unique ID, not by its policy name. Override refers to overriding existing data classification policies and policy rules.

e. -C, --activate (Default: false)

Activate refers to activation of all policies immediately after migration.

> Note: The option to activate supersedes the policy and policy rule association on the exported server - if the option to activate is specified will all be activated, otherwise will all be deactivated.

f. -O, --output (Default: output_stats.txt)

The output summary file is in the selected location.

The file location can be absolute location (c:\program files\Sailpoint\outputs) or relative (..\..\outputs).

Examples:

--output ..\..\imported.log

-O c:\temp\stats.txt

-T, --test (Default: false)

Any changes made during this simulation of the importation of policies and policy rules are rolled back afterward, so you can see what has been changed without altering any policies or policy rules.

f. -M, --multi-output (Default: false)

g. The output summary is written in one or more files, with a time stamp appended to the file name.

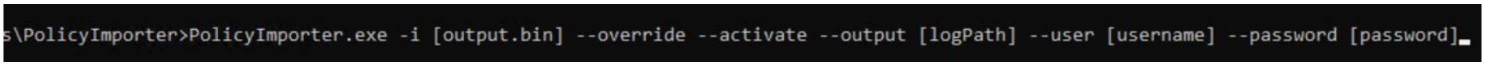
Example: output_stats.180507091022.txt

> Note: When this option is not used, append the content of the result to the same file, along with the time stamp.

h. U, --user (Required).

i. This is the name of the user to whom data classification policies are exported, and should include both the user name and the domain name (if there is one).

j. -P, --password



```
s\PolicyImporter>PolicyImporter.exe -i [output.bin] --override --activate --output [logPath] --user [username] --password [password]
```

After inserting all parameters and executing the command, the tool will indicate either a success or fail message (displayed in the command line). It will also create a log file which the File Access Manager Technical Support Team can use as a reference for troubleshooting.

2. If the user needs more information about the File Access Manager version, complete the following in the command line.

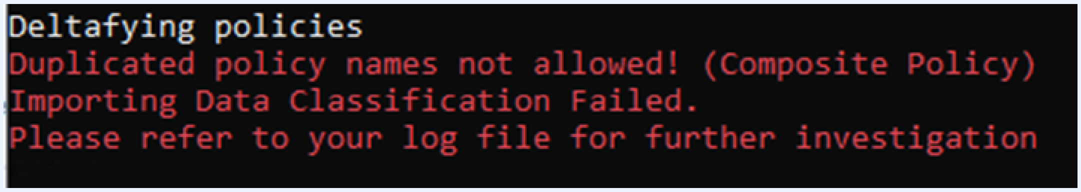
   a. --help

   b. The Help screen displays.

   c. –version

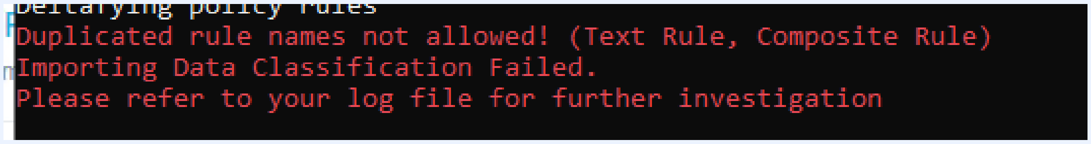
      The version information displays.

> Note: File Access Manager cannot import a Data Classification policy if the policy name exists. In this case, the following error message will display. Rename the existing policy and rerun the import procedure.



```
Deltafying policies
Duplicated policy names not allowed! (Composite Policy)
Importing Data Classification Failed.
Please refer to your log file for further investigation
```

> Note: File Access Manager cannot import a Data Classification rule if the rule name exists. In this case, the following error message will display. Rename the existing rule and rerun the import procedure.



```
Deltafying policy rules
Duplicated rule names not allowed! (Text Rule, Composite Rule)
Importing Data Classification Failed.
Please refer to your log file for further investigation
```