Forcepoint Data Security Posture Management

Powered by Getvisibility

GQL Reference Guide

Forcepoint

Report

Forcepoint
April 23, 2024

Table of Contents

Introduction

GQL (Getvisibility Query Language) is a query language designed to enhance the flexibility and efficiency of querying data through the Forcepoint DSPM platform. It enables users to craft custom queries without the need for hard coding, significantly simplifying the process of filtering through and analysing data. Based on Apache Lucene query language, GQL supports boolean, term, and range queries. This flexibility allows the language to seamlessly integrate with the platform's Analytics software to produce elegant and insightful visualisations. Once mastered, GQL offers maximum flexibility, enabling both broad and precise data analysis. This adaptability ensures that users can leverage the full potential of the Forcepoint DSPM solution, whether they're conducting comprehensive overviews or detailed investigations.

Usage

Terms

There are separate sets of terms used for the different datasets within the Forcepoint DSPM platform. Each of the datasets allow for unique GQL terms relating to this data:

- **Files** - Unstructured data discovered and classified on-prem and in the cloud file storage locations. GQL term examples: `path, ingestedAt, flow`
- **Trustees** - Users and groups that are discovered in on-prem and in cloud IAM systems. GQL term examples: `type, isAdmin, outdatedPassword`
- **Activity** - User activities tracked by the endpoint classification platform. GQL term examples: `recipients, operation, agentId`
- **Management** - Administrative data from individual classification endpoints. GQL term examples: `lastSeen, status, os`

For the full sets of terms, see tables below.

Operations

Operations are performed on or between terms to help filter data.

The operations that are available are:

- **AND** - Combines queries to match items meeting all conditions
- **OR** - Matches items meeting any listed conditions
- **( )** - Groups queries to clarify operation order
- **=** - Equal to
- **!=** - Not equal to
- **>** - Greater than
- **<** - Less than
- **>=** - Greater than or equal to
- **<=** - Less than or equal to

Formation

Queries are formed using terms, their values, and operations. They can be as simple as a query looking for High Risk HR Data:

```
dataAttributeName=HR AND risk=2
```

To complex queries specifying Health, Safety, and Compliance Documents as a data asset in Forcepoint DSPM:

```
complianceTag=PII AND dataAttributeName=HR AND (dataAttributeName=Record OR
dataAttributeName=Legal OR dataAttributeName=Safety) AND (detectorHits="Health
Insurance" OR detectorHits="Risk assessment" OR detectorHits="Policy and Procedure" OR
detectorHits="Compliance report" OR detectorHits="Safety Policies" OR
detectorHits="Security Policies")
```

The UI will give suggestions as you type to help.

Figure 1.

Figure 2.

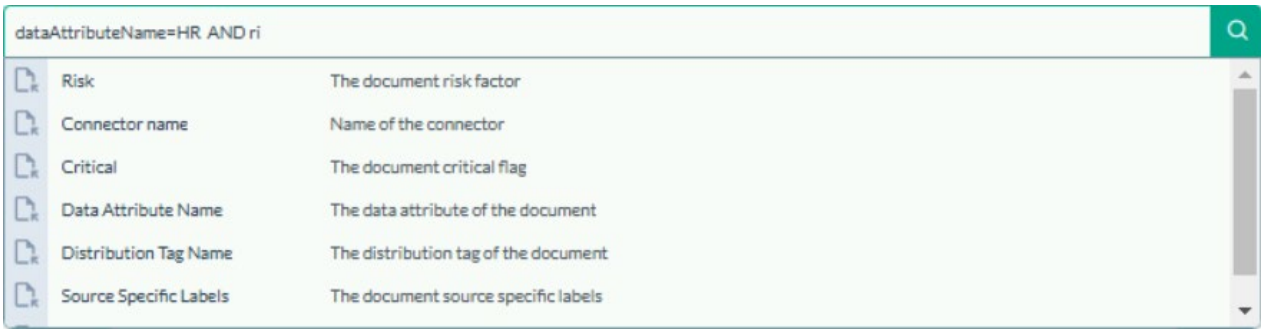

Figure 3.

Figure 4.

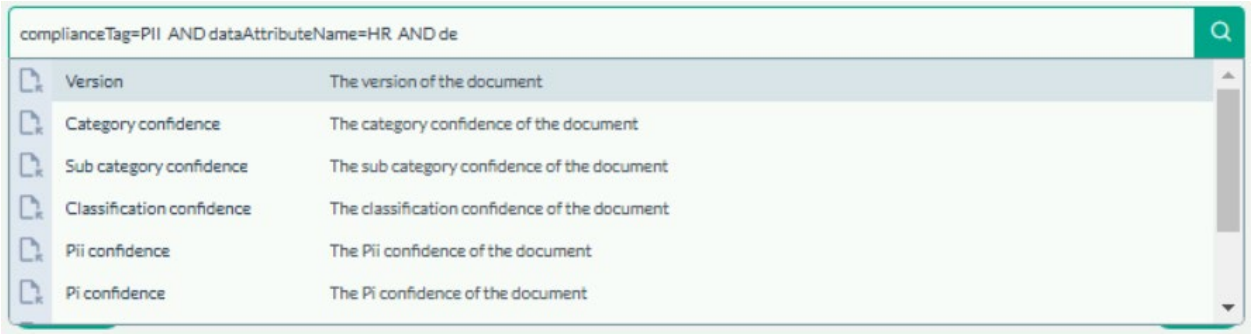

Figure 5.

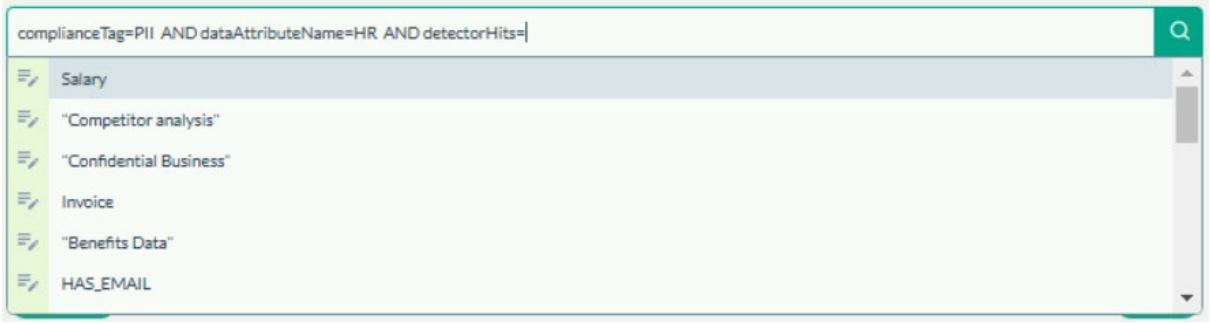

Figure 6.

You should experiment with GQL queries across various platform interfaces. See what works and what doesn't. Get creative and let the real-time suggestions assist you. Remember, you can save the queries you create as bookmarks for future use.

Figure 7.

1. Select the star

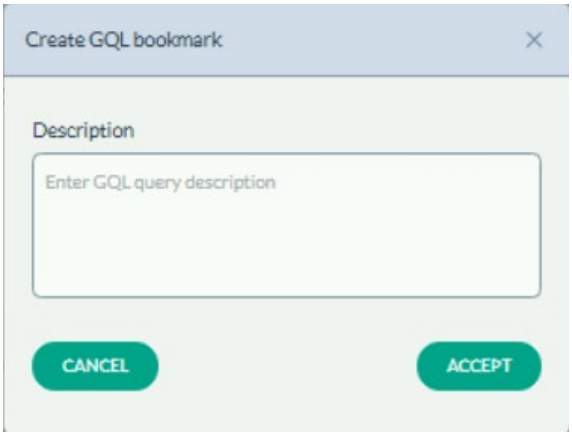

Figure 8.

2. Enter a description, select Accept

Figure 9.

The bookmark is saved

Figure 10.

3. Scroll down to see all your saved bookmarks

Dates

Queries can be created that incorporate dates. These can include exact dates and times or ranges. Date types include: `createdAt, lastModifiedAt`, and `ingestedAt.`

Figure 11.

GQL will provide suggestions for common time intervals such as minutes, days, months, and years.

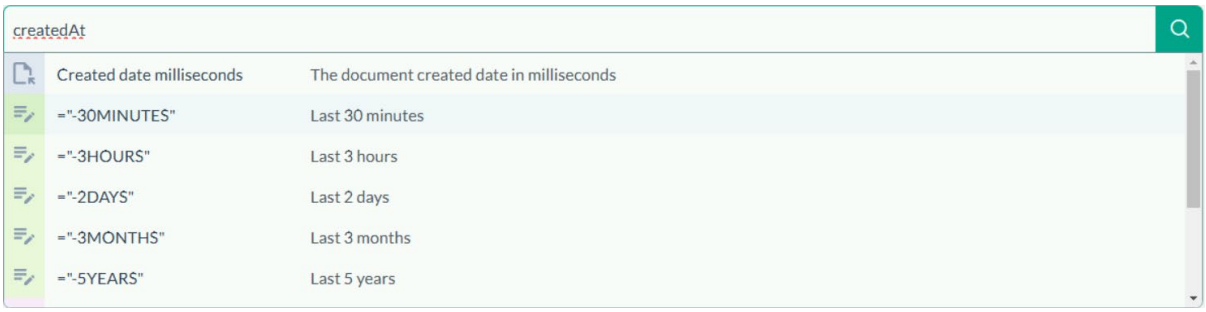

Figure 12.

Once a date type has been selected and an operation associated with it, a date interface will be presented to the user. Simply search for and select the appropriate date to create the query.

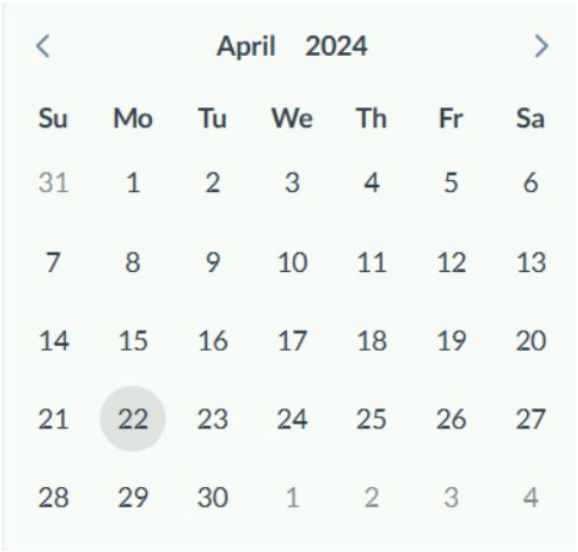

Figure 13.

Date Ranges

If a specific range of dates are needed, for example, all files created in May 2022, the following method should be used.

This method will search for files whose creation dates are greater than or equal to midnight on the 1st of May 2022 and less than midnight on the 1st of June 2022.

1. Type `createdAt>=` and select the first date

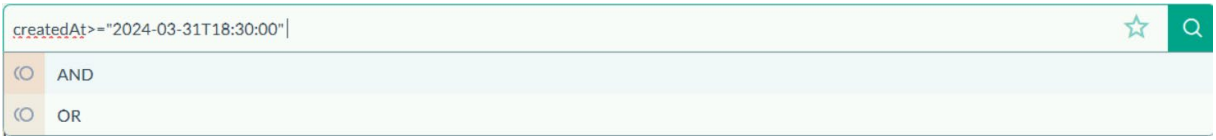

Figure 14.

2. Select **AND**

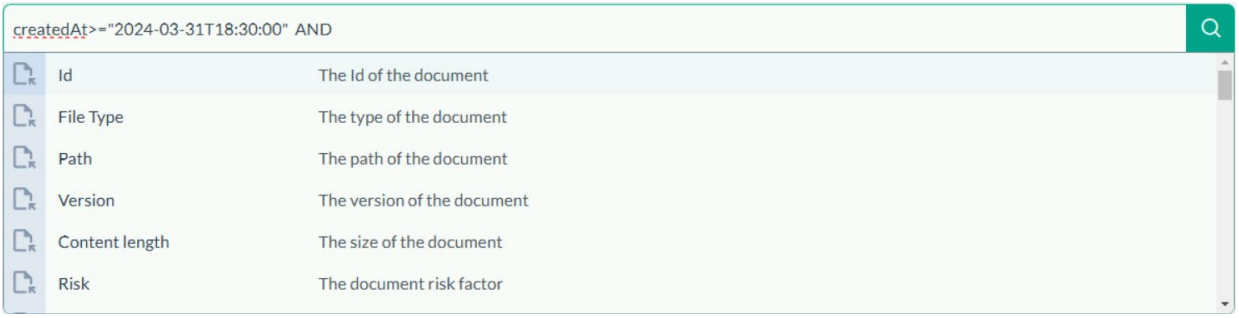

Figure 15.

3. Type `createdAt<` and select the closing date

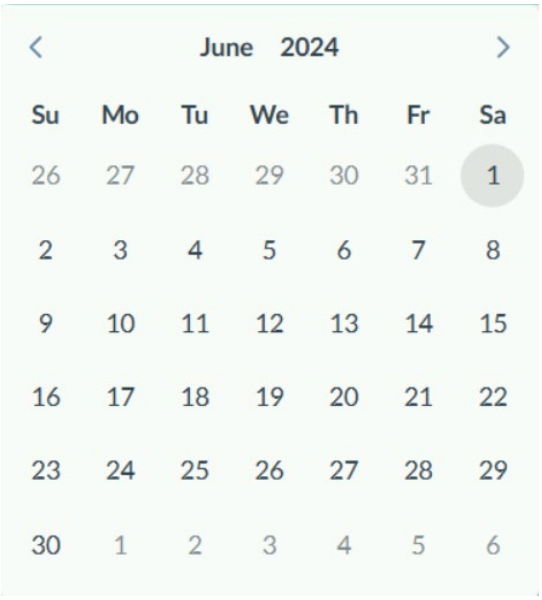

Figure 16.

Figure 17.

4. Hit enter or the search icon and the query will the filter the results

   This method can be used with any date data type. It can be as granular as seconds or as broad as years.

Aggregation

When creating or editing widgets such as counters, charts, or maps in the Analytics boards you will have the ability to aggregate some of the terms in the datasets. For example: you can use counts to show critical shared files, group by file type when displaying classification results, or use multiple groupings to create more complex visualisations.

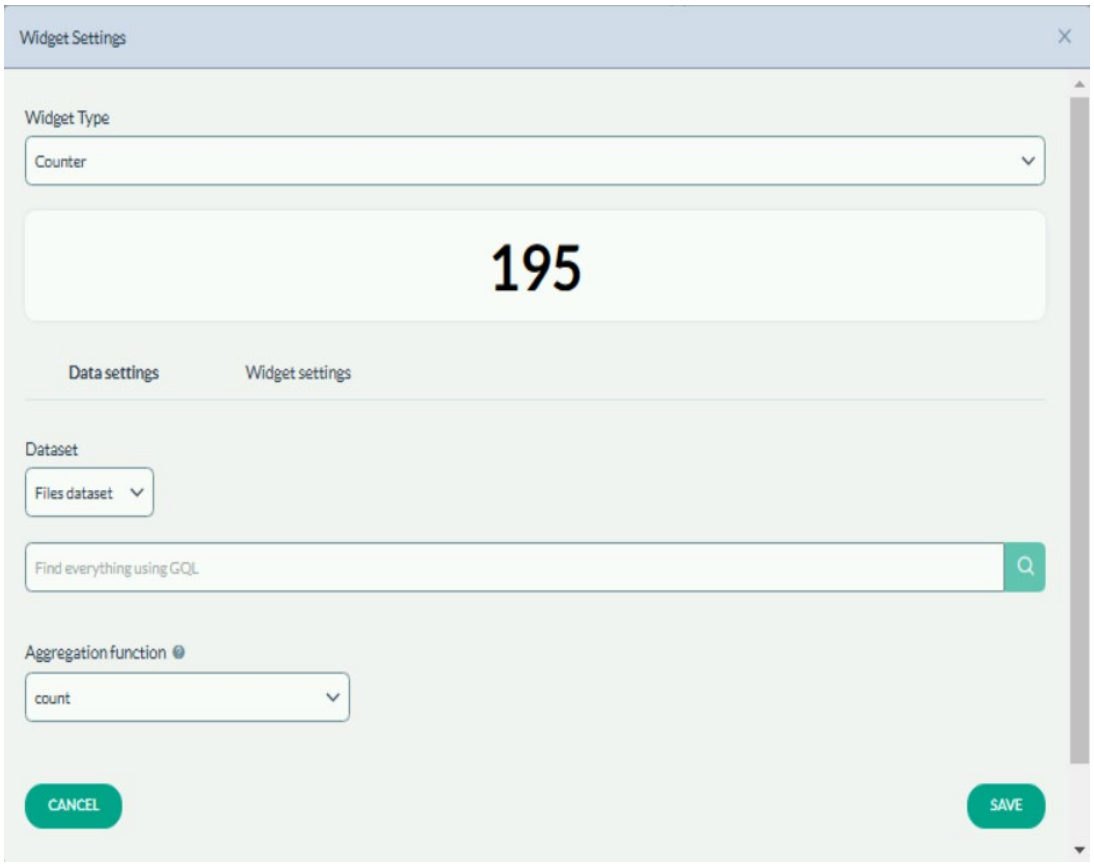

Figure 18.

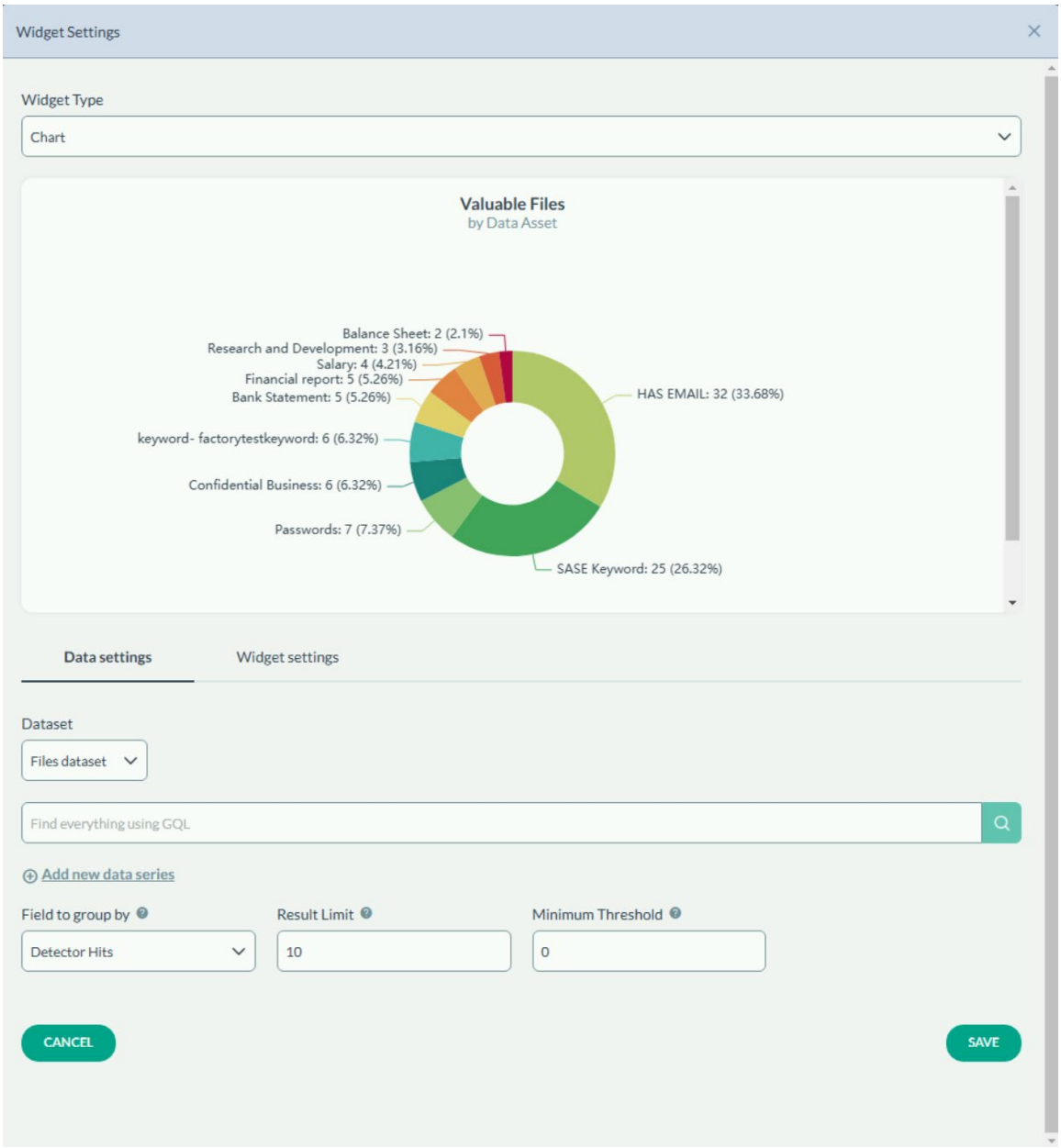
Widget Settings
Widget Type
Chart
Valuable Files
by Data Asset
Balance Sheet: 2 (2.1%)
Research and Development: 3 (3.16%)
Salary: 4 (4.21%)
Financial report: 5 (5.26%)
Bank Statement: 5 (5.26%)
keyword- factorytestkeyword: 6 (6.32%)
Confidential Business: 6 (6.32%)
Passwords: 7 (7.37%)
HAS EMAIL: 32 (33.68%)
SASE Keyword: 25 (26.32%)
Data settings
Widget settings
Dataset
Files dataset
Find everything using GQL
Add new data series
Field to group by
Result Limit
Minimum Threshold
Detector Hits
10
0
CANCEL
SAVE

Figure 19.

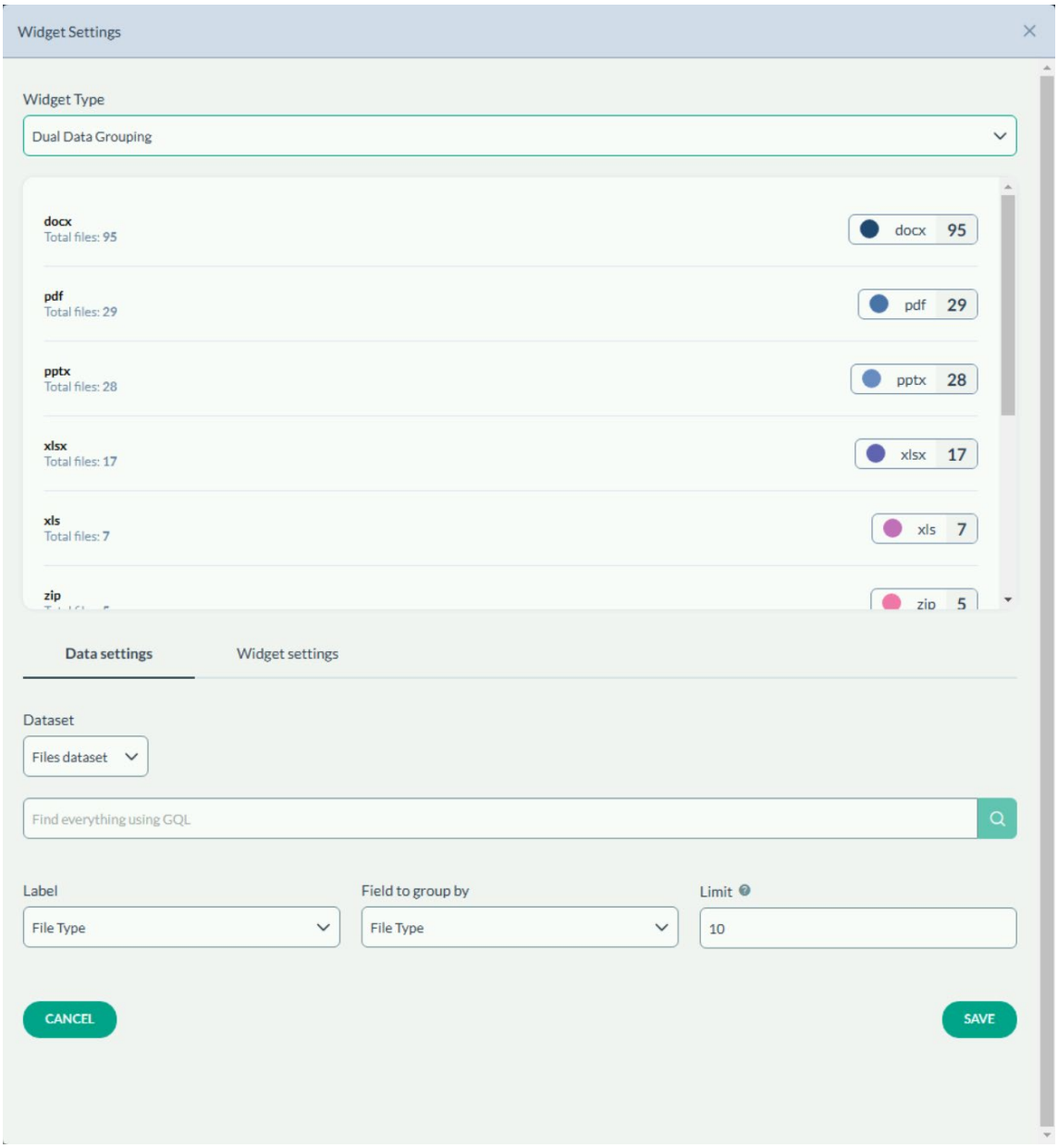

Figure 20.

While not strictly part of GQL yet, they are useful to know as it will help in constructing more descriptive visualisations

GQL Glossaries

- **GQL Term** - Used in the query
- **Label** - Displayed in the interface
- **Type** - Data type of the term
- **Aggregation** - Grouping types that are available to that term, only in the Analytics boards

Files Dataset

Unstructured data discovered and classified on-prem and in the cloud file storage locations.

GQL Term	Label	Type	Description	Aggregation
`fileId`	Id	STRING	The internal Id of the document	
`fileType`	File Type	STRING	The type of the document	Can be grouped
`path`	Path	STRING	The path of the document	
`modelVersion`	Version	STRING	The ML version used on the document	
`contentLength`	Content length	BYTES	The size of the document in bytes	count, sum, average, min, max, median, Can be grouped
`risk`	Risk	NUMBER	The document risk factor. low=0, medium=1, high=2	
`category`	Category	STRING	The ML category of the document	Can be grouped
`categoryConfidence`	Category confidence	DOUBLE	The ML category confidence of the document	
`subCategory`	Sub category	STRING	The ML sub category of the document	
`subCategoryConfidence`	Sub category confidence	DOUBLE	The ML sub category confidence of the document	
`source`	Source	STRING	The source of the document	Can be grouped

createdAt	Created at	DATE	The document creation date	min, max, Can be grouped
lastModifiedAt	Last modified at	DATE	The document last modified date	min, max, Can be grouped
ingestedAt	Ingested at	DATE	Data document passed through the ML pipeline	min, max, Can be grouped
flow	Flow	STRING	The document current flow stage in the ML pipeline. Classified, Catalogued, etc…	Can be grouped
classification	Classification	STRING	The ML classification of the document	Can be grouped
classificationConfidence	Classification confidence	DOUBLE	The ML classification confidence of the document	
configurationIds	Configuration Id	STRING	The scan configuration id of the document	
connectorId	Connector name	STRING	Name of the scan connector	Can be grouped
classifierResult	Classifier result	NUMBER	The classifier results of the document	
pii	Pii	BOOLEAN	The document Pii flag	
piiConfidence	Pii confidence	DOUBLE	The Pii confidence of the document	
pi	Pi	BOOLEAN	The document Pi flag	
piConfidence	Pi confidence	DOUBLE	The Pi confidence of	

			the document	
sensitive	Sensitive	BOOLEAN	The document sensitive flag	
manual	Manual Classification	BOOLEAN	The flag for manually classified files	
critical	Critical	BOOLEAN	The document critical flag	
modifiedAtMilli	Last modified date milliseconds	DATE	The document last modified date in milliseconds	
createdAtMilli	Created date milliseconds	DATE	The document created date in milliseconds	
md5	Document hash	STRING	The hash value of the document	Can be grouped
keywordHits	Keyword Hits	STRING	The keyword hits of the document	Can be grouped
detectorHits	Detector Hits	STRING	The detector hits of the document	Can be grouped
trusteeName	Trustee Name	STRING	The name of an owner of the document	Can be grouped
trusteeLoginName	Trustee Login Name	STRING	The login name of the owner of the document	
signature	Signature	STRING	The signature of the document	
signatureConfidence	Signature Confidence	DOUBLE	The signature confidence of the document	
dataAttributeName	Data Attribute Name	STRING	The data attribute or ML Model hits of the document.	Can be grouped

distributionTag	Distribution Tag Name	STRING	The distribution tag of the document	Can be grouped
keyword	Keyword	STRING	Keyword of the document	Can be grouped
complianceTag	Compliance Tag	STRING	Compliance Tag of the document	Can be grouped
location	Location	STRING	To get Documents by connection location	Can be grouped
language	Language	STRING	The document language	Can be grouped
externalSharedLink	External Shared Link	BOOLEAN	The document sharing status	Can be grouped
sourceSpecificLabelsAttributes	Source Specific Labels	STRING	The document source specific labels	

Trustees Dataset

Users and groups that are discovered in on-prem and in cloud IAM systems.

GQL Term	Label	Type	Description	Aggregation
type	type	STRING	User/Group	Can be grouped
source	source	STRING	The type of the connector	Can be grouped
name	name	STRING	Login name of the trustee	Can be grouped
displayName	displayName	STRING	Name of the trustee	Can be grouped
isEnabled	isEnabled	BOOLEAN	if the trustee is enabled	
isAdmin	isAdmin	BOOLEAN	if trustee is an admin	
outdatedPassword	outdatedPassword	BOOLEAN	The trustee has outdated password	

lastLoginAt	lastLoginAt	DATE	The last time trustee logged in	min, max, Can be grouped, median, average
lastModifiedAt	lastModifiedAt	DATE	The last time trustee was modified	min, max, median, average
createdAt	createdAt	DATE	The time trustee was created	min, max, median, average
connectorId	Configuration Id	STRING	Configuration Id of the trustee	
isActive	isActive	BOOLEAN	if trustee is active	

Activity Dataset

User activities tracked by the endpoint classification platform.

GQL Term	Label	Type	Description	Aggregation
recipients	Email Recipients	STRING	The recipients of the email	
senderEmail	Email Sender	STRING	The sender of the email	
operation	Operation Type	STRING	The type of the operation performed	Can be grouped
eventTime	Event Time	DATE	The time when the event occurred	min, max, Can be grouped
ipAddress	IP Address	STRING	The IP address of the machine where the activity was performed	Can be grouped
hostName	Host Name	STRING	The identification of the agent who performed the activity	Can be grouped
department	Department	STRING	The department of the user who performed the activity	Can be grouped
agentId	Agent	STRING	Unique identifier of the machine	

user	User	STRING	The username of the individual who performed the activity	Can be grouped
contentLength	File Size	BYTES	The size of the file involved in the activity	sum, average, min, max, median, Can be grouped
mimeType	File Type	STRING	The MIME type of the file	Can be grouped
fileName	File Name	STRING	The name of the file	Can be grouped
creationTime	Created At	DATE	The time when the file involved in the activity was created	min, max, Can be grouped
lastModificationTime	Last Modified At	DATE	The last time the file involved in the activity was changed	min, max, Can be grouped
tags	Tags	STRING	Classification tags	Can be grouped

Management dataset

Administrative data from individual classification endpoints.

GQL Term	Label	Type	Description	Aggregation
lastSeen	Last Seen	DATE	The last time the device was observed to be online	min, max, Can be grouped
hostName	Host Name	STRING	The identification of the agent who performed the activity	Can be grouped
domain	Domain	STRING	Shows the Active Directory domain name, if applicable	Can be grouped
ipAddress	IP Address	STRING	Shows the IP address last	Can be grouped

			recorded when the device was active	
`status`	Online Status	STRING	Shows whether the device is currently online or offline	
`user`	User Name	STRING	Displays the name of the last user who logged into the device	Can be grouped
`version`	Agent Version	STRING	The version of the agent software currently installed on the device	Can be grouped
`os`	OS	STRING	Indicates the operating system of the device, either Windows or Mac	Can be grouped
`deviceId`	Device ID	STRING	Displays the ID of the device	
`department`	Department	STRING	Displays the department the agent belongs to	Can be grouped

Forcepoint

forcepoint.com/contact

About Forcepoint

Forcepoint simplifies security for global businesses and governments. Forcepoint's all-in-one, truly cloud-native platform makes it easy to adopt Zero Trust and prevent the theft or loss of sensitive data and intellectual property no matter where people are working. Based in Austin, Texas, Forcepoint creates safe, trusted environments for customers and their employees in more than 150 countries. Engage with Forcepoint on www.forcepoint.com, Twitter and LinkedIn.

[Forcepoint Data Security Posture Management Powered by GetVisibility GQL Reference Guide] [Status]23 Apr. 24