

Cloudera Data Catalog Overview

Date published: 2023-12-16

Date modified: 2025-11-10

The Cloudera logo is displayed in a bold, orange, sans-serif font. The word "CLOUDERA" is written in all caps, with the letter "E" stylized as a horizontal bar with a small triangle on its right side.

Legal Notice

© Cloudera Inc. 2025. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 (“ASLv2”), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER’S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

Contents

About Cloudera Data Catalog.....	4
Cloudera Data Catalog Dashboard.....	5
Cloudera Data Catalog terminology.....	7
Datasets overview.....	8
Search overview.....	9
Bookmarks overview.....	11
Cloudera Data Catalog Profilers.....	15
The Cluster Sensitivity Profiler.....	15
The Ranger Audit Profiler.....	43
The Hive Column Profiler.....	44
Atlas tags overview.....	47
 Before you start.....	 47
Cloudera Base on premises cluster requirements.....	47
Prerequisites to access Cloudera Data Catalog.....	48
Providing role access.....	50

About Cloudera Data Catalog

Cloudera Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across the enterprise.

Cloudera Data Catalog helps you understand data in your data lake. You can locate relevant data of interest based on various parameters. Using Cloudera Data Catalog, you can understand how data is interpreted for use, how it is created and modified, and how data access is secured and protected.



Note: Cloudera Data Catalog is supported in the EU and APAC regional Control Plane. You can ensure that you manage, secure, collaborate, and govern data assets across multiple clusters and environments within the EU region or APAC region where your organization operates as per the data protection regulatory requirements.



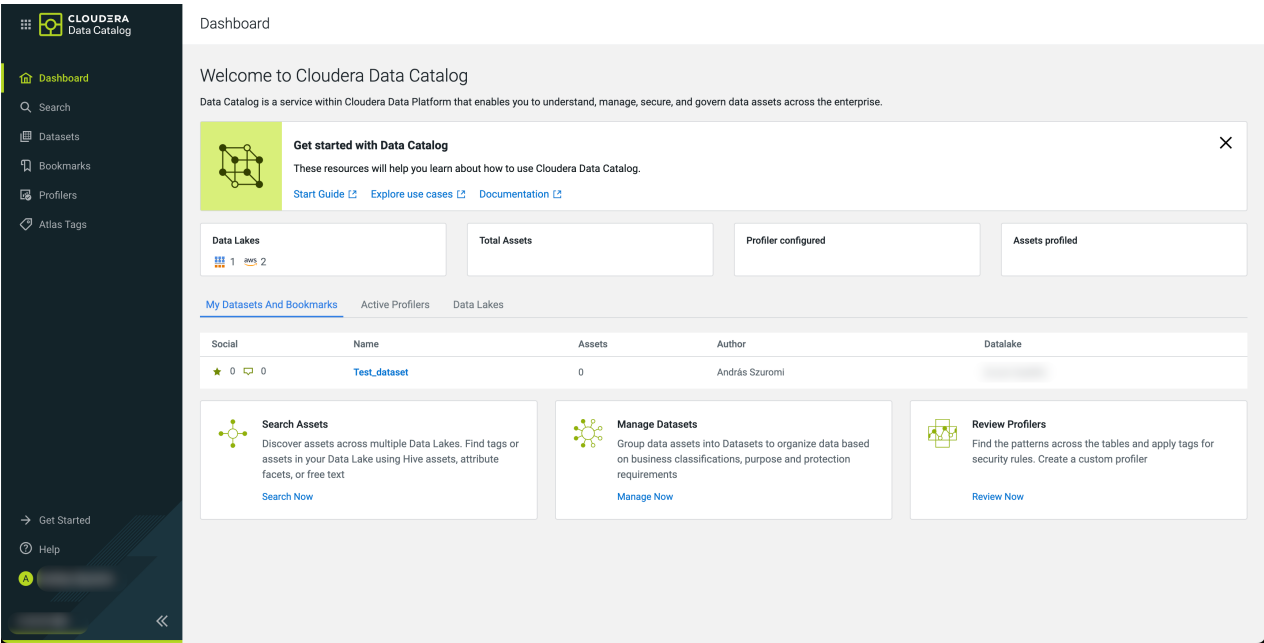
Cloudera Data Catalog enables data stewards across the enterprise to work with data assets in the following ways:

- Organize and curate data globally:
 - Organize data based on business classifications, purpose, protections needed, etc. For more information, see:
 - [Creating a classification from Asset Details](#)
 - [Creating a classification from Search page](#)
 - Promote responsible collaboration across enterprise data workers. For more information, see [Bookmarks overview](#).
- Understand where relevant data is located:
 - Catalog and search to locate relevant data of interest (sensitive data, commonly used, high risk data, etc.).
 - Understand what types of sensitive personal data exists and where it is located
 - For more information, see:
 - [The Cluster Sensitivity Profiler](#)
- Understand how data is interpreted for use:
 - View basic descriptions: schema, classifications (business cataloging), and encodings
 - View statistical models and parameters
 - View user annotations, wrangling scripts, view definitions etc.
 - For more information, see [Viewing Data Asset details](#).
 - For more information, see [Navigation in Asset Details](#).
 - [The Hive Column Profiler](#)
- Understand how data is created and modified:
 - Visualize upstream lineage and downstream impact
 - Understand how schema or data evolve
 - View and understand data supply chain (pipelines, versioning, and evolution)
 - For more information, see [Navigation support for hive entities within Lineage](#).
- Understand how data access is secured, protected, and audited:
 - Understand who can see which data and metadata (for example, based on business classifications) and under what conditions (security policies, data protection, anonymization)
 - View who has accessed what data from a forensic audit or compliance perspective
 - Visualize access patterns and identify anomalies
 - For more information, see
 - [The Ranger Audit Profiler](#)
 - [Viewing Ranger access audits](#)
 - [Viewing Atlas entity audits](#)
 - [Viewing Ranger policies](#)

Cloudera Data Catalog Dashboard

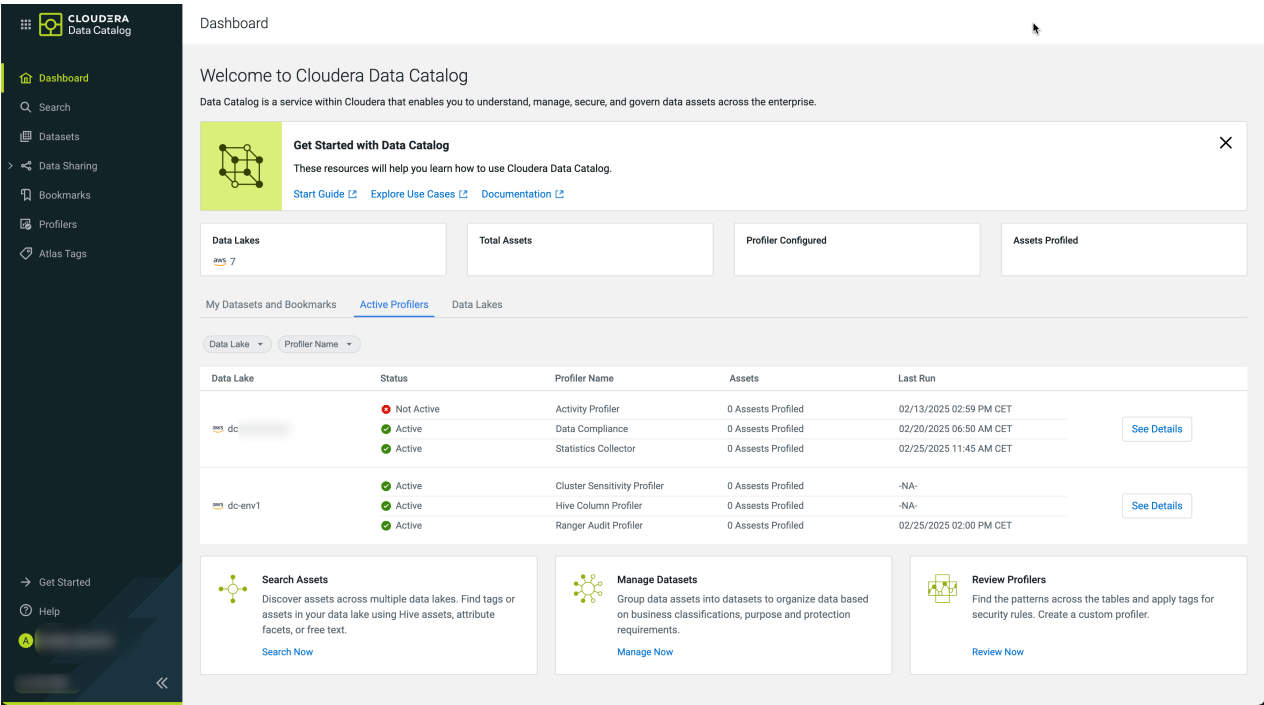
The Dashboard provides quick access to vital service information at a glance, in the form of visual, actionable navigation for multiple operations. The user-friendly navigation enables viewing, filtering, and acting upon data quickly and in a simple manner.

Data Stewards can view the **Dashboard** at a glance, and also focus on the most important tasks, enabling faster decision making as well as immediate action. The application lets you perform multiple actions for different types of content that helps in visualizing information with ease.



The displayed cards and tabs are fully interactive, with clickable areas for easy navigation to relevant parts of applications. Users can access individual sections and narrow down the information displayed. For example, you can manage the datasets created and bookmarked by you. For more information, see:

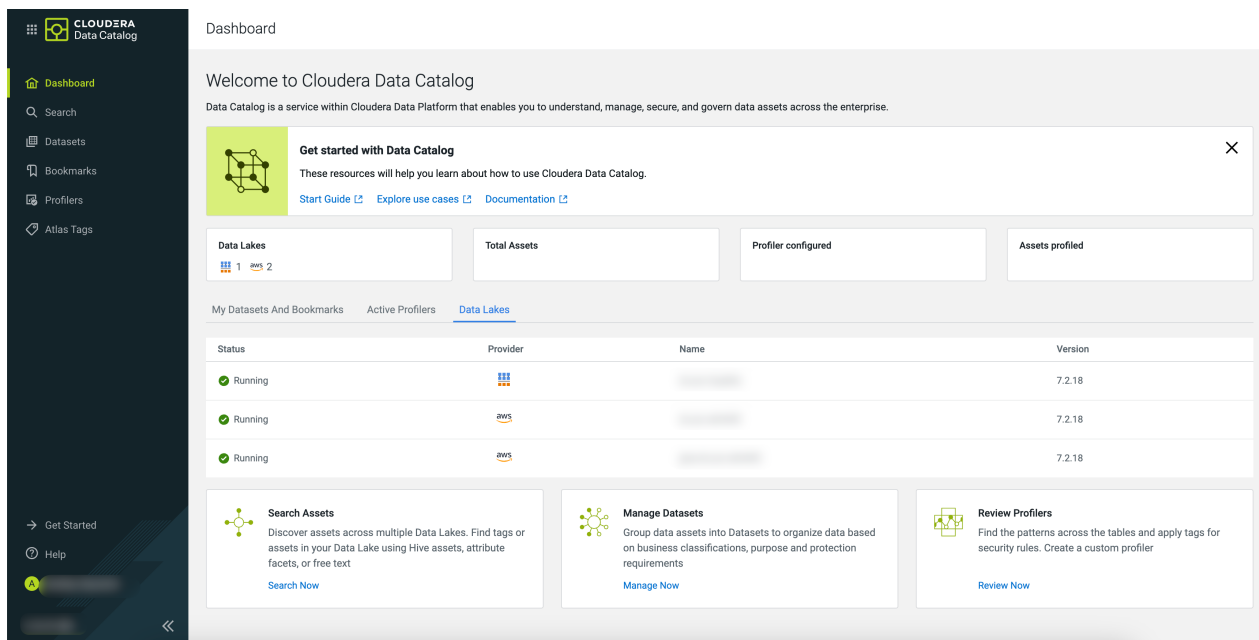
- [Creating datasets](#)
- [Managing datasets](#)
- [Bookmarks overview](#)



The **Dashboard** page contains information on your profilers, the total number of assets that are profiled, along with the assets that are scanned for data. Clicking See Details leads you to your individual profilers:

- [The Cluster Sensitivity Profiler](#)
- [The Ranger Audit Profiler](#)

- [The Hive Column Profiler](#)



Cloudera Data Catalog terminology

An overview of terminology used in Cloudera Data Catalog.

Profiler

Enables the Cloudera Data Catalog service to gather and view information about different relevant characteristics of data such as shape, distribution, quality, and sensitivity which are important to understand and use the data effectively. For example, view the distribution between males and females in column Gender, or min/max/mean/null values in a column named avg_income. Profiled data is generated on a periodic basis from the profilers, which run at regularly scheduled intervals. Works with data sourced from Apache Ranger Audit Logs, Apache Atlas Metadata Store, and Hive.

Data Lake

A trusted and governed data repository that stores, processes, and provides access to many kinds of enterprise data to support data discovery, data preparation, analytics, insights, and predictive analytics. In the context of Cloudera, a Data Lake can be realized in practice with an Cloudera Manager enabled Cloudera cluster that runs Apache Atlas for metadata and governance services, and Apache Ranger for security services.

ECS

The Embedded Container Service (ECS) service enables you to run Cloudera Data Services on premises by creating container-based clusters in your data center. In addition to the option to use OpenShift, which requires that you deploy and manage the Kubernetes infrastructure, you can also deploy a Embedded Container Service cluster, which creates and manages an embedded Kubernetes infrastructure for use with Cloudera Data Services on premises.

OpenShift Container (OCP)

OpenShift is an enterprise platform for container orchestration.

Data Asset

A data asset is a physical asset located in the Cloudera ecosystem such as a Hive table which contains business or technical data. A data asset could include a specific instance of an Apache Hive database, table, or column. An asset can belong to multiple asset collections. Data assets are equivalent to “entities” in Apache Atlas.

Datasets

Datasets allow users of Cloudera Data Catalog to manage and govern various kinds of data objects as a single unit through a unified interface. Asset collections help organize and curate information about many assets based on many facets including data content and metadata, such as size/schema/tags/alterations, lineage, and impact on processes and downstream objects in addition to the display of security and governance policies.

Related Information

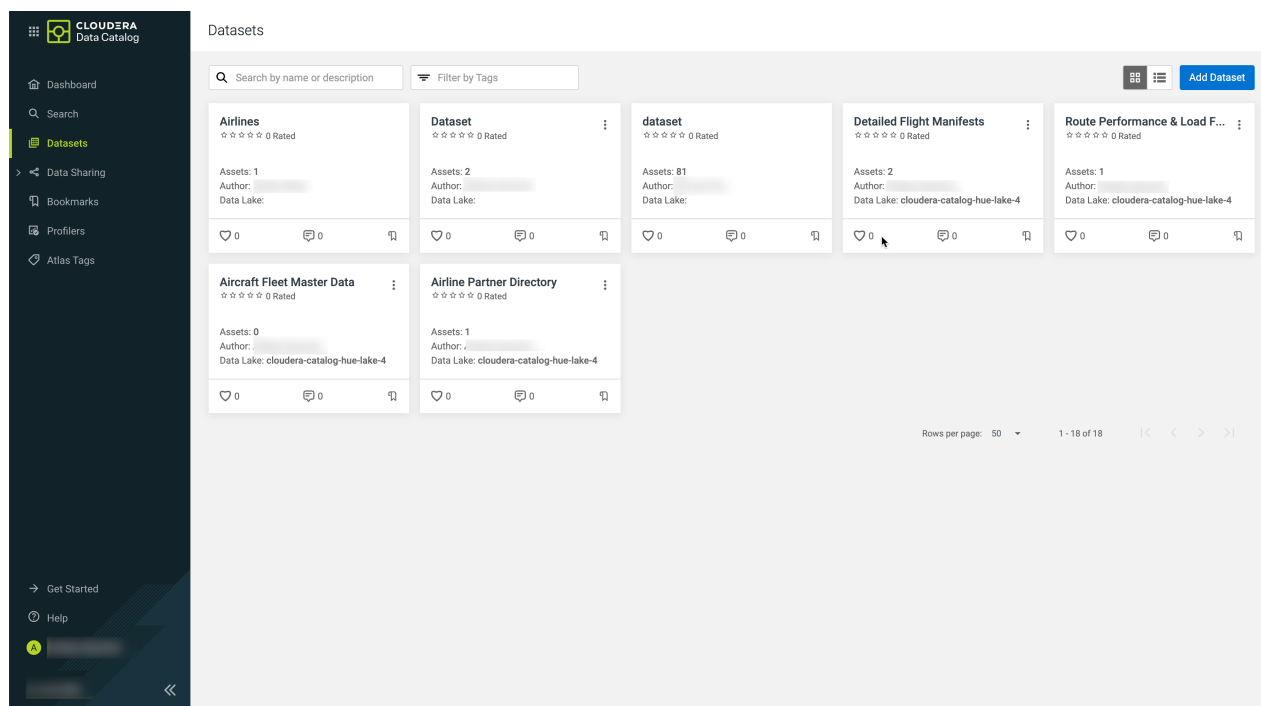
[About Cloudera Data Catalog](#)

[Cloudera Management Console](#)

Datasets overview

A dataset is a group of assets that fit a set of search criteria so that you can manage and administer them collectively for specific business purposes.

Figure 1: Datasets



Asset collections enable you to perform the following tasks when working with your data:

- **Organize**

Group data assets into datasets based on business classifications, purpose, protections, relevance, ownership etc.

Example use cases:

- A company may mark all data that needs to be compliant with GDPR, for example, with the tag "PII" and use Ranger policies to control access to these assets.
- A data steward may mark all data related to sales with tags as sales_transactions, sales_targets, customer_segments and include them into a common dataset for easier discovery.
- Ownership by different teams can be signified by datasets.

- Search

Find tags or assets in your data lake using Hive assets, attribute facets, or free text.

Advanced asset search uses facets of technical and business metadata about the assets, such as those captured in Apache Atlas, to help users define and build collections of interest.

- Understand

Audit data asset security and use for anomaly detection, forensic audit and compliance, and proper control mechanisms.

You can edit datasets after you create them and the assets contained within the collection will be updated. CRUD (Create, Read, Update, Delete) is supported for datasets.

Related Information

[Managing datasets](#)

[Bookmarks overview](#)

Search overview

On the Cloudera Data Catalog **Search** page, and enter a search string in the search box to view all the assets with details that contain the search string.

When you enter the search terms in **Search**, you are looking up names, types, descriptions, and other metadata collected by Cloudera Data Catalog. The search index includes metadata (not data) about your environment and cluster data assets and operations. You can make the search more powerful by associating your own information (business metadata) to the stored assets.

Accessing data lakes

Users have access to the lakes based on the permissions that are granted. You can choose the available lake by selecting the appropriate radio button.

The screenshot shows the Cloudera Data Catalog Search interface. On the left, a sidebar contains navigation links: Dashboard, Search (highlighted), Datasets, Bookmarks, Profilers, and Atlas Tags. The main area is titled 'Search' and includes a search input field. Below the search bar, there's a 'Data Lakes' section with two radio buttons: 'pvcdc1013-1-datalake' (selected) and 'pvcdc1015-1-datalake'. To the right of the radio buttons is a table listing various assets. The table has columns: Type, Name, Qualified Name, Created On, Owner, and Source. The assets listed include 'stg_flight_manifests', 'dim_aircraft', 'raw_bookings', 'airlines_new', 'flight_summary_complex', 'agg_route_performance', 'enriched_flight_data', 'route_performance_archive_hive', 'default', 'airline_operations', and 'atlas_janus'.

**Note:**

- You can search the assets of one data lake at a time.
- For the selected data lake, click the Atlas and Ranger links to go to the respective base cluster services in a new browser tab.

Using search filters

Use the search filter to fine-tune your results. You can specify further filters for the give categories (**Type**, **Owners**, **Entity Tag**, **Glossary Terms**) in by starting to type the filter name.

Search

The screenshot shows the Cloudera Data Catalog search interface. On the left, there are filters for Data Lakes, Type, Entity Tag, and Glossary Terms. The 'Type' filter is expanded, showing options like Hive Table, HBase Table, HBase Column Family, and HBase Namespace. The 'Entity Tag' and 'Glossary Terms' filters have 'Add New Value' buttons. The main area displays search results for the data lake 'pvcdcaszuromi1013-1-datalake | 90'. The results table has columns: Type, Name, Qualified Name, Created On, Owner, and Source. The results include various Hive Tables and HBase Column Families.

Type	Name	Qualified Name	Created On	Owner	Source
Hive Table	stg_flight_manifests	airline_operations.stg_flight_manifests@cm	Wed Oct 15 2025	admin	hive
Hive Table	dim_aircraft	airline_operations.dim_aircraft@cm	Wed Oct 15 2025	admin	hive
Hive Table	raw_bookings	airline_operations.raw_bookings@cm	Wed Oct 15 2025	admin	hive
Hive Table	airlines_new	airline_operations.airlines_new@cm	Wed Oct 15 2025	admin	hive
Hive Table	flight_summary_complex	airline_operations.flight_summary_complex...	Mon Oct 20 2025	admin	hive
Hive Table	agg_route_performance	airline_operations.agg_route_performance...	Wed Oct 15 2025	admin	hive
Hive Table	enriched_flight_data	airline_operations.enriched_flight_data@cm	Wed Oct 15 2025	admin	hive
Hive Table	route_performance_archive_hive	airline_operations.route_performance_archi...	Wed Oct 15 2025	admin	hive
HBase Namespace	default	default@cm	-NA-	atlas	hbase
Hive DB	airline_operations	airline_operations@cm	-NA-	admin	hive
Hive DB	default	default@cm	-NA-	public	hive
HBase Table	atlas_janus	default:atlas_janus@cm	-NA-	atlas	hbase
HBase Column Family	m	default:atlas_janus.m@cm	-NA-	atlas	hbase
HBase Column Family	f	default:atlas_janus.f@cm	-NA-	atlas	hbase
HBase Column Family	g	default:atlas_janus.g@cm	-NA-	atlas	hbase
HBase Column Family	h	default:atlas_janus.h@cm	-NA-	atlas	hbase
HBase Column Family	i	default:atlas_janus.i@cm	-NA-	atlas	hbase
HBase Column Family	s	default:atlas_janus.s@cm	-NA-	atlas	hbase

Downloading search results as CSV files

You can also download the search result for the current query with the selected data lake. The feature allows you to download up to 10000 rows for the current search query.

The CSV file format does not conform to any specific order or continuation in the downloaded results. For example, a user can download 10000 assets and later downloads the results for the same query again, then the downloaded CSV files may not contain the search results in the same order as it was downloaded previously.

Click Download CSV to start your download:

The screenshot shows the Cloudera Data Catalog search interface. On the left, there are filters for Data Lakes, Type, Entity Tag, and Glossary Terms. The main area displays search results for the data lake 'n9def9-env-datalake | 829'. The results table has columns: Type, Name, Qualified Name, Created On, Owner, and Source. The results include various Hive Tables. A 'Download CSV' button is highlighted in the bottom right corner of the results table.

Type	Name	Qualified Name	Created On	Owner	Source
Hive Table	eu_countries	hortoniabank.eu_countries@cm	Thu Aug 29 2024	hive	
Hive Table	vw_customers	hortoniabank.vw_customers@cm	Thu Aug 29 2024	hive	
Hive Table	provider_summary	claim.provider_summary@cm	Thu Aug 29 2024	hive	

Related Information

[Search filters](#)

[Searching for assets using Atlas glossaries](#)

[Additional search options for asset types](#)

[Accessing tables based on Ranger policies](#)

Bookmarks overview

Using the tools in Datasets, you can collaborate and share insights with other users in the enterprise.

You can rate datasets and view the average rating of a dataset. This can help other users to find datasets with higher ratings easily. You can also add your knowledge and insights about the asset collection by adding comments. Other users can respond to your comments or add their comments about each data asset collection.

Dashboard

Welcome to Cloudera Data Catalog

Data Catalog is a service within Cloudera that enables you to understand, manage, secure, and govern data assets across the enterprise.

Get Started with Data Catalog
 These resources will help you learn how to use Cloudera Data Catalog.
[Start Guide](#) [Top Tasks](#) [Documentation](#)

Data Lakes

Total Assets
 1680
 Realtime

Profiler Configured
 6/6 in all data lakes

Assets Profiled
 0 in last 7 days

[My Datasets and Bookmarks](#) [Active Profiles](#) [Data Lakes](#)

Social	Name	Assets	Author	Data Lake
★ 1 🗨 2	Route Performance & Load Factor Summary	1		cloudera-catalog-hue-lake-4
★ 0 🗨 0	Legacy Airline Reference Data	1		cloudera-catalog-hue-lake-4
★ 0 🗨 0	Aircraft Fleet Master Data	0		cloudera-catalog-hue-lake-4
★ 0 🗨 0	Airline Partner Directory	1		cloudera-catalog-hue-lake-4
★ 0 🗨 0	Detailed Flight Manifests	2		cloudera-catalog-hue-lake-4
★ 0 🗨 0	Dataset	2		-NA-

Search Assets
 Discover assets across multiple data lakes. Find tags or assets in your data lake using Hive assets, attribute facets, or free text.
[Search Now](#)

Manage Datasets
 Group data assets into datasets to organize data based on business classifications, purpose and protection requirements.
[Manage Now](#)

Review Profilers
 Find the patterns across the tables and apply tags for security rules. Create a custom profiler.
[Review Now](#)

On the right hand side of each dataset **Details** page, you can see additional details about the dataset. The collaboration details are also displayed in this tab.

Datasets Details

[Home](#) / [Datasets](#) / Datasets Details

Route Performance & Load Factor Summary

This dataset contains high-level, aggregated performance metrics for every flight route, updated daily. Its primary Key Performance Indicator (KPI) is the Load Factor, which measures the operational efficiency of each route.

[Atlas](#) [Ranger](#)

Data Lake
 cloudera-catalog-hue-lake-4

Tags
[Flight Operations](#)
[Revenue Management](#)
[Business Performance](#)
[Aggregated](#) [Summary](#)
[KPI](#) [Analytics](#)
[Anonymized](#) [Internal](#)
[Certified](#)

Created By
 [User]

Created On
 09/05/2025 12:43 PM CEST

Last Modified On
 09/05/2025 01:08 PM CEST

The tab displays the following details:

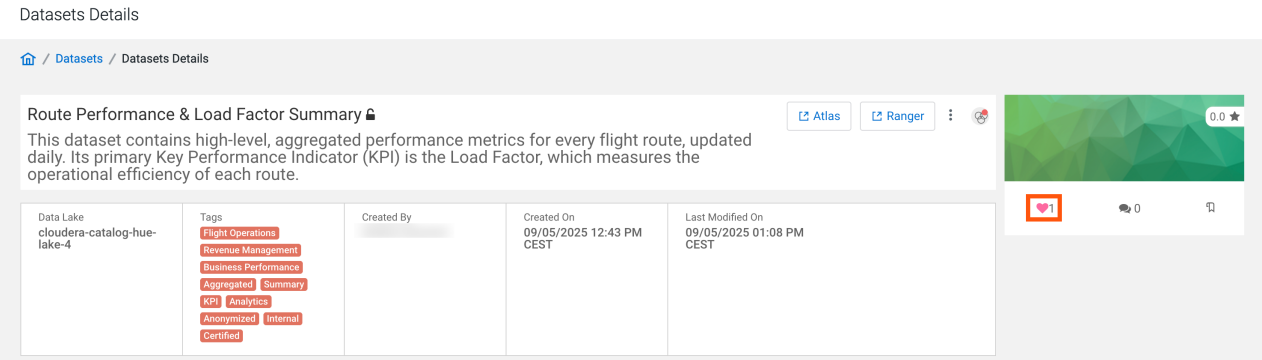
- The average rating for the asset collection
- The number of likes

- The number of comments
- The bookmark icon indicating if the dataset is bookmarked by the current user or not

You can perform the following collaboration actions for each dataset.

Like a dataset

You can let other users know that you like a dataset. The  icon on the dataset **Details** page displays the total number of likes received by this dataset helping you to find it faster.

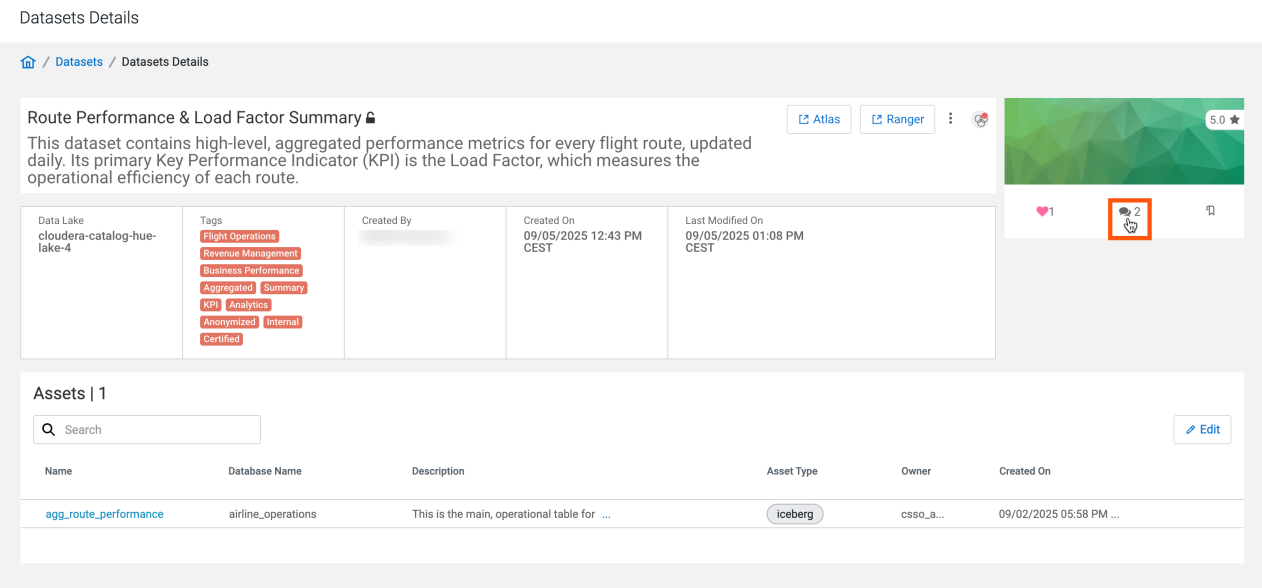


Click the  icon to add the dataset to your list of liked collections.

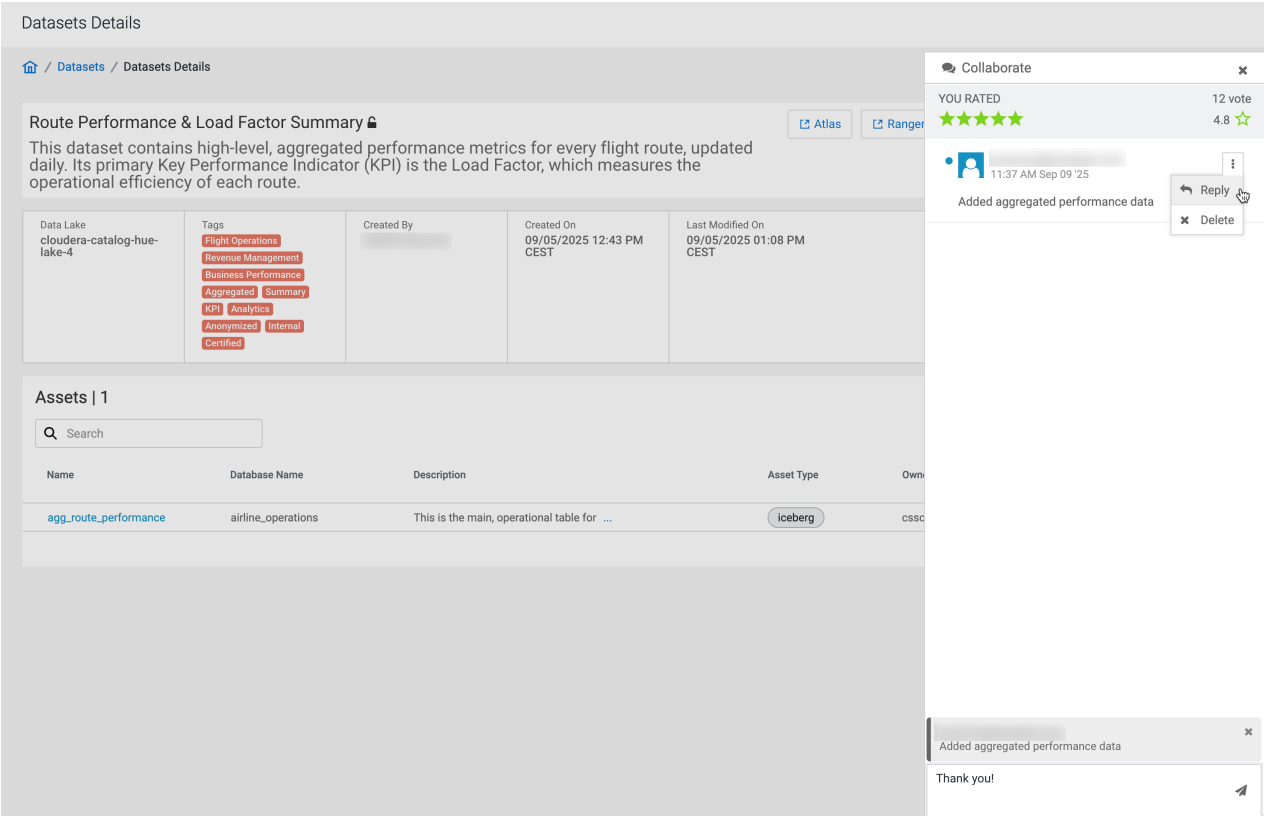
Comment and discuss about a dataset

You might want to share your knowledge or insights about this dataset with other users. Cloudera Data Catalog allows you to collaborate with other users by adding comments.

Click the  icon to add a comment about this dataset.



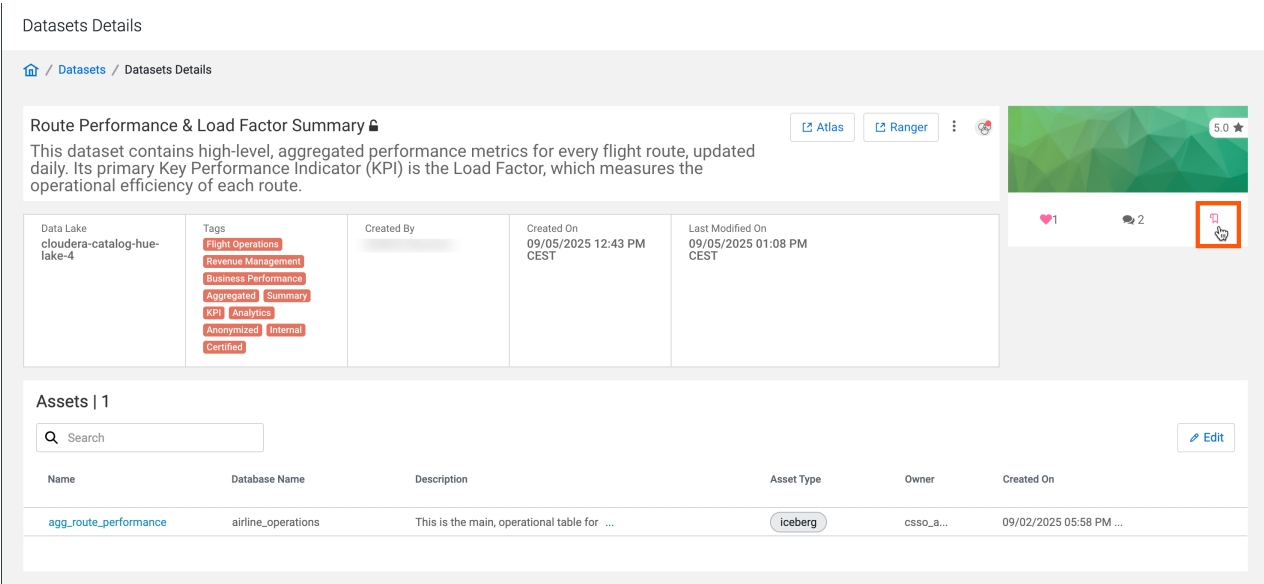
The **Collaborate** tab expands. Click  icon to reply to an existing comment. You can continue to add comments for each dataset.



Bookmark the dataset

In addition to sharing with other users, you can also bookmark datasets for easy access in the future.

Click the  icon to add the dataset to your list of bookmarks.



This dataset will appear in the list of bookmarks when you click the Bookmarks link in the left navigation menu.

Rate the dataset

You can also rate the datasets on a scale of one to five. Click the  icon to enable rating, then click the  icons to rate the open dataset. The **Collaborate** tab expands.

Click the stars to provide your own rating.

Datasets Details

[Home](#) / [Datasets](#) / Datasets Details

Route Performance & Load Factor Summary 🔒

This dataset contains high-level, aggregated performance metrics for every flight route, updated daily. Its primary Key Performance Indicator (KPI) is the Load Factor, which measures the operational efficiency of each route.

[Atlas](#)
[Range](#)

Data Lake	Tags	Created By	Created On	Last Modified On
cloudera-catalog-hue-lake-4	Flight Operations Revenue Management		09/05/2025 12:43 PM CEST	09/05/2025 01:08 PM CEST

Collaborate
✕

RATE THIS COLLECTION
12 votes

★ ★ ★ ★ ☆
3.0 ☆

@cloudera.com
11:37 AM Sep 09 '25

⋮

Added aggregated performance data

[View all 1 reply](#)

The rating on the **Datasets** page shows the average of the rating provided by various users. The **Rating** section also displays the number of votes given for this dataset.

View the tags of an dataset

You can add tags while creating a dataset. You can filter your datasets based on these tags.

Datasets

The screenshot shows the Cloudera Data Lake interface. At the top, there is a search bar with the placeholder text "Search by name or description". Below the search bar, there is a card titled "Detailed Flight Manifests" with a subtitle "☆☆☆☆ 0 Rated". The card also displays "Assets: 2" and "Author: [redacted]". The "Data Lake" is identified as "cloudera-catalog-hue-lake-4". A dropdown menu is open, showing a list of "Flight Operations" with counts: "Revenue Management" (1), "Summary" (1), "Flight Operations" (2), "Aggregated" (1), "Anonymized" (1), and "test_dss_tag_0" (1). The "Flight Operations" item is highlighted with a mouse cursor. At the bottom of the card, there are icons for a heart (0), a comment bubble (0), and a bookmark.



Note: These tags do not synchronize to Atlas.

Related Information

Managing datasets

Cloudera Data Catalog Profilers

Profilers create metadata annotations that summarize the content and shape characteristics of the data assets (such as distribution of values in a box plot or histogram).

The Cloudera Data Catalog profiler employs Kubernetes enabled job scheduling and runs profilers jobs on-demand.

Profiler Name	Description
Cluster Sensitivity Profiler	The profiler automatically classifies your data with preconfigured tags, such as, PII, PCI, HIPAA and others.
Ranger Audit Profiler	A Ranger audit log summarizer.
Hive Column Profiler	Provides summary statistics like Maximum, Minimum, Mean, Unique, and Null values at the Hive column level.

The Cluster Sensitivity Profiler

The Cluster Sensitivity Profiler automatically performs context and content inspection to detect various types of sensitive data and suggest suitable classifications or tags based on the type of sensitive content detected or discovered.

Tag Rules

The Cluster Sensitivity Profiler have system-defined and user-defined tags that identify data types based on regular expressions.

Profilers / Tag Rules

pvc155-sp1-4-datalake

JobsConfigsTag Rules

Rule Groups

System Deployed77

Custom Deployed0

Custom Draft0

Q Type to search

+ New

Name	Description	Associated Tags	Created By	Status
AUT_Passport_Detection	AUT_Passport_Detection	AUT_Passport_Detection		Deployed
SVK_NationalID_Detection	SVK_NationalID_Detection	SVK_NationalID_Detection		Deployed
LVA_IBAN_Detection	LVA_IBAN_Detection	LVA_IBAN_Detection		Deployed
ROU_IBAN_Detection	ROU_IBAN_Detection	ROU_IBAN_Detection		Deployed
NOR_NationalID_Detection	NOR_NationalID_Detection	NOR_NationalID_Detection		Deployed
FRA_IBAN_Detection	FRA_IBAN_Detection	FRA_IBAN_Detection		Deployed
DEU_IBAN_Detection	DEU_IBAN_Detection	DEU_IBAN_Detection		Deployed
FIN_NationalID_Detection	FIN_NationalID_Detection	FIN_NationalID_Detection		Deployed
ESP_Passport_Detection	ESP_Passport_Detection	ESP_Passport_Detection		Deployed
DEU_Passport_Detection	DEU_Passport_Detection	DEU_Passport_Detection		Deployed
CYP_IBAN_Detection	CYP_IBAN_Detection	CYP_IBAN_Detection		Deployed
FIN_Passport_Detection	FIN_Passport_Detection	FIN_Passport_Detection		Deployed
email	email	email		Deployed
AUT_IBAN_Detection	AUT_IBAN_Detection	AUT_IBAN_Detection		Deployed
GRC_NationalID_Detection	GRC_NationalID_Detection	GRC_NationalID_Detection		Deployed
BEL_IBAN_Detection	BEL_IBAN_Detection	BEL_IBAN_Detection		Deployed

Auto-detected data types and system-defined regular expressions

The system-defined tag rules identify the following data types:

Bank account, credit card, email

- Bank account
- Credit card
- Driver number (UK)
- Email
- Name: **bankaccount**
 - Match Type: **Column value**

- Confidence: 85

- Regular expression:

```

regex("\b(AD)\d{2}\d{8}[a-zA-Z0-9]{12}|(AE|IL)\d{2}\d{19}|(AL|CY)\d{2}\d{8}[a-zA-Z0-9]{16}|(PL)\d{2}\d{24}|(AT|BA|EE|LT)\d{2}\d{16}|(BE)\d{2}\d{12}|(BG)\d{2}[A-Z]{4}\d{6}[a-zA-Z0-9]{8}|(BH)\d{2}[A-Z]{4}[a-zA-Z0-9]{14}|(CH)\d{2}\d{5}[a-zA-Z0-9]{12}|(CZ|ES|SE|SK|TN)\d{2}\d{20}|(DE|ME|RS)\d{2}\d{18}|(DK|FI|FO|GL)\d{2}\d{14}|(DO)\d{2}[A-Z]{4}\d{20}|(FR|MC)\d{2}\d{10}[a-zA-Z0-9]{11}\d\d\d(GB)\d{2}[A-Z]{4}\d{14}|(IE)\d{2}[a-zA-Z0-9]{4}\d{14}|(GE)\d{2}[a-zA-Z0-9]{2}\d{16}|(GI)\d{2}[A-Z]{4}[a-zA-Z0-9]{15}|(GR)\d{2}\d{7}[a-zA-Z0-9]{16}|(HR)\d{2}\d{17}|(HU)\d{2}\d{24}|(IS)\d{2}\d{22}|(IT|SM)\d{2}[A-Z]\d{10}[a-zA-Z0-9]{12}|(KW)\d{2}[A-Z]{4}[a-zA-Z0-9]{22}|(KZ)\d{2}\d{3}[a-zA-Z0-9]{3}\d{10}|(LU)\d{2}\d{3}[a-zA-Z0-9]{13}|(LB)\d{2}\d{4}[a-zA-Z0-9]{20}|(LI)\d{2}\d{5}[a-zA-Z0-9]{12}|(LV)\d{2}[A-Z]{4}[a-zA-Z0-9]{13}|(MK)\d{2}\d{3}[a-zA-Z0-9]{10}\d\d\d(MR)\d{2}\d{23}|(MT)\d{2}[A-Z]{4}\d{5}[a-zA-Z0-9]{18}|(MU)\d{2}[A-Z]{4}\d{19}[A-Z]{3}|(NL)\d{2}[A-Z]{4}\d{10}|(NO)\d{2}\d{11}|(PT)\d{2}\d{21}|(RO)\d{2}[A-Z]{4}[a-zA-Z0-9]{16}|(SA)\d{2}\d{2}[a-zA-Z0-9]{18}|(SI)\d{2}\d{15}|(TR)\d{2}\d{5}[a-zA-Z0-9]{17}\b")

```

- Match Type: **Column name**

- Confidence: 15

- Regular expression:

```
is in ci("bankaccount","account","accountnumber")
```

- Name: **creditcard**

- Match Type: **Column value**

- Confidence: 85

- Regular expression:

```
regex("\\b((5[1-5][0-9]{2}[0-9]{4}[0-9]{4}[0-9]{4})|((633(3|4)|67(67|59)|49(36|11|05|03)))[0-9]{4}[0-9]{4}[0-9]{4}(( [0-9]{3}|[0-9]{2}))?)|(4[0-9]{3}[0-9]{4}[0-9]{4}[0-9]{4}|4[0-9]{3}[0-9]{4}[0-9]{4}[0-9]{1})|((6011)[0-9]{4}[0-9]{4}[0-9]{4})|((622)[0-9]{1}[0-9]{4}[0-9]{4}[0-9]{4})|((65)[0-9]{2}[0-9]{4}[0-9]{4}[0-9]{4})|((64)[0-9]{2}[0-9]{4}[0-9]{4}[0-9]{4})|(((3[47][0-9]{2})|2014|2149)[0-9]{4}[0-9]{4}[0-9]{3})|(3(0[0-5]|[68][0-9])[0-9][0-9]{4}[0-9]{4}[0-9]{2})|((3528)[0-9]{4}[0-9]{4}[0-9]{4})|((3529)[0-9]{4}[0-9]{4}[0-9]{4})|((35)[3-8][0-9][0-9]{4}[0-9]{4}[0-9]{4})|(63[789][0-9][0-9]{4}[0-9]{4}[0-9]{4})|(5610[0-9]{4}[0-9]{4}[0-9]{4})|(56022[1-5][0-9]{2}[0-9]{4}[0-9]{4}))\\b|\\b((5893|0604|6304|6759|676[1-3]|50(18|20|38))[0-9]{4}[0-9]{4}([0-9]{0,4})?([0-9]{0,3})?)\\b|\\b(67(06|71|09)[0-9]{4}[0-9]{4}[0-9]{4}([0-9]{0,3})?)\\b|\\b((633110|564182)[0-9]{2}[0-9]{4}[0-9]{4}([0-9]{3}|[0-9]{2})?)\\b", "\\b((5[1-5][0-9]{2}[\\ ][0-9]{4}[\\ ] [0-9]{4})|((633(3|4)|67(67|59)|49(36|11|05|03))[\\ ][0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4}([\\ ][0-9]{3}|[\\ ][0-9]{2}))?)|(4[0-9]{3}[\\ ][0-9]{4}[\\ ][0-9]{1})|((6011)[0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|((622)[0-9]{1}[0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|((65)[0-9]{2}[0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|((64)[0-9]{2}[0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|(((3[47][0-9]{2})|2014|2149)[0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{3})|(3(0[0-5]|[68][0-9])[0-9][0-9]{4}[0-9]{4}[0-9]{2})|((3528)[0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|((3529)[0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|((35)[3-8][0-9][\\ ][0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|(63[789][0-9][0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|(5610[0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|(56022[1-5][0-9]{2}[0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4}))\\b|\\b((5893|0604|6304|6759|676[1-3]|50(18|20|38))[0-9]{4}[0-9]{4}([0-9]{0,4})?([0-9]{0,3})?)\\b|\\b(67(06|71|09)[0-9]{4}[0-9]{4}[0-9]{4}([0-9]{0,3})?)\\b|\\b((633110|564182)[0-9]{2}[0-9]{4}[0-9]{4}([0-9]{3}|[0-9]{2})?)\\b", "\\b((5[1-5][0-9]{2}[\\ ][0-9]{4}[\\ ] [0-9]{4})|((633(3|4)|67(67|59)|49(36|11|05|03))[\\ ][0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4}([\\ ][0-9]{3}|[\\ ][0-9]{2}))?)|(4[0-9]{3}[\\ ][0-9]{4}[\\ ][0-9]{1})|((6011)[0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|((622)[0-9]{1}[0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|((65)[0-9]{2}[0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|((64)[0-9]{2}[0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|(((3[47][0-9]{2})|2014|2149)[0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{3})|(3(0[0-5]|[68][0-9])[0-9][0-9]{4}[0-9]{4}[0-9]{2})|((3528)[0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|((3529)[0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|((35)[3-8][0-9][\\ ][0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|(63[789][0-9][0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|(5610[0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|(56022[1-5][0-9]{2}[0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4}))\\b|\\b((5893|0604|6304|6759|676[1-3]|50(18|20|38))[0-9]{4}[0-9]{4}([0-9]{0,4})?([0-9]{0,3})?)\\b|\\b(67(06|71|09)[0-9]{4}[0-9]{4}[0-9]{4}([0-9]{0,3})?)\\b|\\b((633110|564182)[0-9]{2}[0-9]{4}[0-9]{4}([0-9]{3}|[0-9]{2})?)\\b", "\\b((5[1-5][0-9]{2}[\\ ][0-9]{4}[\\ ] [0-9]{4})|((633(3|4)|67(67|59)|49(36|11|05|03))[\\ ][0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4}([\\ ][0-9]{3}|[\\ ][0-9]{2}))?)|(4[0-9]{3}[\\ ][0-9]{4}[\\ ][0-9]{1})|((6011)[0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|((622)[0-9]{1}[0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|((65)[0-9]{2}[0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|((64)[0-9]{2}[0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|(((3[47][0-9]{2})|2014|2149)[0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{3})|(3(0[0-5]|[68][0-9])[0-9][0-9]{4}[0-9]{4}[0-9]{2})|((3528)[0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|((3529)[0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|((35)[3-8][0-9][\\ ][0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|(63[789][0-9][0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|(5610[0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|(56022[1-5][0-9]{2}[0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4}))\\b|\\b((5893|0604|6304|6759|676[1-3]|50(18|20|38))[0-9]{4}[0-9]{4}([0-9]{0,4})?([0-9]{0,3})?)\\b|\\b(67(06|71|09)[0-9]{4}[0-9]{4}[0-9]{4}([0-9]{0,3})?)\\b|\\b((633110|564182)[0-9]{2}[0-9]{4}[0-9]{4}([0-9]{3}|[0-9]{2})?)\\b"
```

```
9]{1})|((6011)[\\][0-9]{4}[\\][0-9]{4}[\\ ][0-9]{4})|((622)[0-9]{1}[\\ ][0-9]{4}[\\][0-9]{4}[\\][0-9]{4})|((65)[0-9]{2}[\\ ][0-9]{4}[\\ ][0-9]{4}[\\][0-9]{4})|((64)[0-9]{2}[\\][0-9]{4}[\\ ][0-9]{4}[\\][0-9]{4})|((3[47][0-9]{2})|2014|2149)[\\ ][0-9]{4}[\\][0-9]{4}[\\ ][0-9]{3})|((3[47][0-9]{2})[\\ ][0-9]{6}[\\][0-9]{5})|(3(0[0-5]|[68][0-9])[0-9][\\ ][0-9]{4}[\\ ][0-9]{4}[\\][0-9]{2})|(3(0[0-5]|[68][0-9])[0-9][\\ ][0-9]{6}[\\ ][0-9]{4})|((3528)[\\][0-9]{4}[\\][0-9]{4}[\\ ][0-9]{4})|((3529)[\\ ][0-9]{4}[\\ ][0-9]{4}[\\][0-9]{4})|((35)[3-8][0-9][\\][0-9]{4}[\\ ][0-9]{4}[\\ ][0-9]{4})|(63[789][0-9][\\][0-9]{4}[\\ ][0-9]{4}[\\][0-9]{4})|(5610[\\ ][0-9]{4}[\\ ][0-9]{4}[\\][0-9]{4})|(5602[\\ ][2[1-5][0-9]{2}[\\][0-9]{4}[\\][0-9]{4})|((5893|0604|6304|6759|676[1-3]|50(18|20|38))[\\ ][0-9]{4}[\\][0-9]{4}[\\ ][0-9]{0,4})?((\\ ][0-9]{0,3})?)|(67(06|71|09)[\\ ][0-9]{4}[\\][0-9]{4}[\\ ][0-9]{4})((\\ ][0-9]{0,3})?)|((6331[\\ ][10|5641][\\ ][82][0-9]{2}[\\ ][0-9]{4}[\\][0-9]{4})((\\ ][0-9]{3}|[\\][0-9]{2})?)\\b","\\b(5[1-5][0-9]{2}-[0-9]{4}-[0-9]{4}-[0-9]{4})|((633(3|4)|67(67|59)|49(36|11|05|03))-[0-9]{4}-[0-9]{4}-[0-9]{4}(-( [0-9]{3}|[0-9]{2}))?)|(4[0-9]{3}-[0-9]{4}-[0-9]{4}-[0-9]{4}|4[0-9]{3}-[0-9]{4}-[0-9]{4}-[0-9]{1})|((6011)-[0-9]{4}-[0-9]{4}-[0-9]{4})|((622)[0-9]{1}-[0-9]{4}-[0-9]{4}-[0-9]{4})|((65)[0-9]{2}-[0-9]{4}-[0-9]{4}-[0-9]{4})|((64)[0-9]{2}-[0-9]{4}-[0-9]{4}-[0-9]{4})|((3[47][0-9]{2})|2014|2149)-[0-9]{4}-[0-9]{4}-[0-9]{3})|((3[47][0-9]{2})-[0-9]{6}-[0-9]{5})|(3(0[0-5]|[68][0-9])[0-9]-[0-9]{4}-[0-9]{4}-[0-9]{2})|(3(0[0-5]|[68][0-9])[0-9]-[0-9]{6}-[0-9]{4})|((3528)-[0-9]{4}-[0-9]{4}-[0-9]{4})|((3529)-[0-9]{4}-[0-9]{4}-[0-9]{4})|((35)[3-8][0-9]-[0-9]{4}-[0-9]{4}-[0-9]{4})|(63[789][0-9]-[0-9]{4}-[0-9]{4}-[0-9]{4})|(5610-[0-9]{4}-[0-9]{4}-[0-9]{4})|((5602-2[1-5][0-9]{2}-[0-9]{4}-[0-9]{4})|((5893|0604|6304|6759|676[1-3]|50(18|20|38))-[0-9]{4}-[0-9]{4}(-[0-9]{0,4})?(-[0-9]{0,3})?)|(67(06|71|09)-[0-9]{4}-[0-9]{4}-[0-9]{4}(-[0-9]{0,3})?)|((6331-10|5641-82)[0-9]{2}-[0-9]{4}-[0-9]{4}-[0-9]{3})|(-[0-9]{3})|(-[0-9]{2})?)\\b")
```

- Match Type: **Column name**

- Confidence: 15
- Regular expression:

```
is_in_ci("creditcard","ccnumber")
```

- Name: **ukdrivernumber**

- Match Type: **Column value**

- Confidence: 85
- Regular expression:

```
regex("\\b([A-Z0-9]{5}[0-9][0156][0-9]([0][1-9]|[1-2][0-9]|[3][01])[0-9][A-Z0-9]{3}[A-Z]{2})\\b")
```

- Name: **email**
 - Match Type: **Column value**
 - Confidence: 85
 - Regular expression:

```
regex("\\b([a-zA-Z0-9_\\-\\.]+)@((\\[[0-9]{1,3}\\.[0-9]{1,3}\\.[0-9]{1,3}\\.|)((([a-zA-Z0-9\\-]+\\.)+)|([a-zA-Z]{2,4}|[0-9]{1,3})(\\[\\]?))\\b")
```

- Match Type: **Column name**
 - Confidence: 15
 - Regular expression:

```
is_in_ci("emailaddress","emailaddr","email")
```

- Match Type: **Table name**
 - Confidence: 100
 - Regular expression:

```
(?i)\\w*email\\w*
```

IBAN numbers

- Austria (AUT)
- Belgium (BEL)
- Bulgaria (BGR)
- Switzerland (CHE)
- Cyprus (CYP)
- Czech Republic (CZE)
- Germany (DEU)
- Denmark (DNK)
- Spain (ESP)
- Estonia (EST)
- Finland (FIN)
- France (FRA)
- United Kingdom (GBR)
- Greece (GRC)
- Croatia (HRV)
- Hungary (HUN)
- Ireland (IRL)
- Iceland (ISL)
- Italy (ITA)
- Liechtenstein (LIE)
- Lithuania (LTU)
- Latvia (LVA)
- Luxembourg (LUX)
- Malta (MLT)
- Netherlands (NLD)
- Norway (NOR)
- Poland (POL)
- Portugal (PRT)

- Romania (ROU)
- Slovakia (SVK)
- Slovenia (SVN)
- Sweden (SWE)
- Name: **LVA_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
- Regular expression:

```
regex("(\\s|,|;|\\A)(LV\\d{2}([ ]?) [A-Z]{4}\\3[A-Za-z0-9]{4}\\3[A-Za-z0-9]{4}\\3[A-Za-z0-9]{4}\\3[A-Za-z0-9]{4}\\3[A-Za-z0-9]{4})(\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
- Regular expression:

```
is_in_ci("Latvia IBAN","Latvia International Bank Account Number")
```

- Name: **ROU_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
- Regular expression:

```
regex("(\\s|,|;|\\A)(RO\\d{2}([ ]?) [A-Z]{4}\\3[A-Za-z0-9]{4}\\3[A-Za-z0-9]{4}\\3[A-Za-z0-9]{4}\\3[A-Za-z0-9]{4}\\3[A-Za-z0-9]{4})(\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
- Regular expression:

```
is_in_ci("Romania IBAN","Romania International Bank Account Number")
```

- Name: **FRA_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
- Regular expression:

```
regex("(\\s|,|;|\\A)(FR\\d{2}([ ]?)\\d{4}\\3\\d{4}\\3\\d{2}[0-9A-Z]{2}\\3[0-9A-Z]{4}\\3[0-9A-Z]{4}\\3[0-9A-Z]{9}[0-7]|0[1-9]|[1-8]\\d)(\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
- Regular expression:

```
is_in_ci("France IBAN","France International Bank Account Number")
```

- Name: **DEU_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) (DE\\d{2} ([ ]?) \\d{4} \\3 \\d{4} \\3 \\d{4} \\3 \\d{4} \\3 \\d{2}) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Germany IBAN", "Germany International Bank Account Number")
```

- Name: **CYP_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) (CY\\d{2} ([ ]?) \\d{4} \\3 \\d{4} \\3 \\d{4} \\3 \\d{4} \\3 \\d{4} \\3 \\d{4}) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Cyprus IBAN", "Cyprus International Bank Account Number")
```

- Name: **AUT_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) (AT\\d{2} ([ ]?) \\d{4} \\3 \\d{4} \\3 \\d{4} \\3 \\d{4}) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Austria IBAN", "Austria International Bank Account Number")
```

- Name: **BEL_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A)(BE\\d{2}([ ]?)\\d{4}\\3\\d{4}\\3\\d{4})(\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Belgium IBAN","Belgium International Bank Account Number")
```

- Name: **EST_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A)(EE\\d{2}([ ]?)\\d{4}\\3\\d{4}\\3\\d{4}\\3\\d{4})(\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Estonia IBAN","Estonia International Bank Account Number")
```

- Name: **CHE_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A)(CH\\d{2}([ ]?)\\d{4}\\3\\d{4}\\3\\d{4}\\3\\d{4}\\3\\d{4})(\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Switzerland IBAN","Switzerland International Bank Account Number")
```

- Name: **ESP_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) (ES\\d{2} ([ ]?) \\d{4} \\3 \\d{4} \\3 \\d{4} \\3 \\d{4} \\3 \\d{4}) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Spain IBAN","Spain International Bank Account Number")
```

- Name: **ITA_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) (IT\\d{2} ([ ]?) [A-Z]{1} \\3 \\d{10} \\3 [A-Z0-9]{12}) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Italy IBAN","Italy International Bank Account Number")
```

- Name: **SVK_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) (SK\\d{2} ([ ]?) \\d{4} \\3 \\d{4} \\3 \\d{4} \\3 \\d{4} \\3 \\d{4}) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Slovakia IBAN","Slovakia International Bank Account Number")
```

- Name: **SWE_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) (SE\\d{2} ([ ]?) \\d{4} \\3 \\d{4} \\3 \\d{4} \\3 \\d{4} \\3 \\d{4}) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Sweden IBAN", "Sweden International Bank Account Number")
```

- Name: **NOR_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) (NO\\d{2} ([ ]?) \\d{4} \\3 \\d{4} \\3 \\d{3}) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Norway IBAN", "Norway International Bank Account Number")
```

- Name: **GRC_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) (GR\\d{2} ([ ]?) \\d{3} \\3 \\d{4} \\3 [A-Z0-9]{4} \\3 [A-Z0-9]{4} \\3 [A-Z0-9]{4} \\3 [A-Z0-9]{3}) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Greece IBAN", "Greece International Bank Account Number")
```

- Name: **IBAN_Detection**

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("IBAN", "International Bank Account Number")
```

- Name: **GBR_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) (GB\\d{2} ([ ]?) [A-Z]{4}\\\\3\\\\d{4}\\\\3\\\\d{4}\\\\3\\\\d{4}) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("United Kingdom IBAN","United Kingdom International Bank Account Number")
```

- Name: **IRL_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) (IE\\d{2} ([ ]?) [A-Z]{4}\\\\3\\\\d{4}\\\\3\\\\d{4}\\\\3\\\\d{4}) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Ireland IBAN","Ireland International Bank Account Number")
```

- Name: **FIN_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) (FI\\d{2} ([ ]?) \\d{4}\\\\3\\\\d{4}\\\\3\\\\d{4}\\\\3\\\\d{2}) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Finland IBAN","Finland International Bank Account Number")
```

- Name: **NLD_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) (NL\\d{2} ([ ]?) [A-Z]{4} \\3\\d{4} \\3\\d{4} \\3\\d{2}) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Netherlands IBAN", "Netherlands International Bank Account Number")
```

- Name: **BGR_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) (BG\\d{2} ([ ]?) [A-Z]{4} \\3\\d{4} \\3[A-Z0-9]{4} \\3[A-Z0-9]{4} \\3[A-Z0-9]) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Bulgaria IBAN", "Bulgaria International Bank Account Number")
```

- Name: **SVN_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) (SI\\d{2} ([ ]?) \\d{4} \\3\\d{4} \\3\\d{4} \\3\\d{3}) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Slovenia IBAN", "Slovenia International Bank Account Number")
```

- Name: **LIE_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A)(LI\\d{2}([ ]?)\\d{4}\\3\\d{1}[A-Z0-9]{3}\\3[A-Z0-9]{4}\\3[A-Z0-9]{4}\\3[A-Z0-9]{1})(\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Liechtenstein IBAN","Liechtenstein International Bank Account Number")
```

- Name: **POL_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A)(PL\\d{2}([ ]?)\\d{4}\\3\\d{4}\\3\\d{4}\\3\\d{4}\\3\\d{4}\\3\\d{4})(\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Poland IBAN","Poland International Bank Account Number")
```

- Name: **PRT_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A)(PT\\d{2}([n]?)\\d{4}\\3\\d{4}\\3\\d{4}\\3\\d{4}\\3\\d{4}\\3\\d{1})(\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Portugal IBAN","Portugal International Bank Account Number")
```

- Name: **LTU_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
- Regular expression:

```
regex("(\\s|,|;|\\A) (LT\\d{2} ([ ]?) \\d{4} \\3 \\d{4} \\3 \\d{4} \\3 \\d{4}) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
- Regular expression:

```
is_in_ci("Lithuanian IBAN", "Lithuania International Bank Account Number")
```

- Name: **HRV_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
- Regular expression:

```
regex("(\\s|,|;|\\A) (HR\\d{2} ([ ]?) \\d{4} \\3 \\d{4} \\3 \\d{4} \\3 \\d{4} \\3 \\d{1}) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
- Regular expression:

```
is_in_ci("Croatian IBAN", "Croatia International Bank Account Number")
```

- Name: **LUX_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
- Regular expression:

```
regex("(\\s|,|;|\\A) (LU\\d{2} ([ ]?) \\d{3} [A-Z0-9]{1} \\3 [A-Z0-9]{4} \\3 [A-Z0-9]{4} \\3 [A-Z0-9]{4}) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
- Regular expression:

```
is_in_ci("Luxembourg IBAN", "Luxembourg International Bank Account Number")
```

- Name: **DNK_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) (DK\\d{2} ([ ]?) \\d{4} \\3 \\d{4} \\3 \\d{4} \\3 \\d{2}) (\\s|,|;|\\Z) ")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Denmark IBAN","Denmark International Bank Account Number")
```

- Name: **HUN_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) (HU\\d{2} ([ ]?) \\d{4} \\3 \\d{4} \\3 \\d{4} \\3 \\d{4} \\3 \\d{4} \\3 \\d{4}) (\\s|,|;|\\Z) ")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Hungary IBAN","Hungary International Bank Account Number")
```

- Name: **CZE_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) (CZ\\d{2} ([ ]?) \\d{4} \\3 \\d{4} \\3 \\d{4} \\3 \\d{4} \\3 \\d{4}) (\\s|,|;|\\Z) ")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Czech Republic IBAN","Czechia IBAN","Czech Republic International Bank Account Number","Czechia International Bank Account Number")
```

- Name: **ISL_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A)(IS\\d{2}([ ]?)\\d{4}\\3\\d{4}\\3\\d{4}\\3\\d{4}\\3\\d{4}\\3\\d{2})(\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Iceland IBAN","Iceland International Bank Account Number")
```

- Name: **MLT_IBAN_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("\\A(MT\\d{2}([ ]?) [A-Z]{4}\\3\\d{4}\\3\\d{1} [A-Z0-9]{3}\\3[A-Z0-9]{4}\\3[A-Z0-9]{4}\\3[A-Z0-9]{4}\\3[A-Z0-9]{3})\\Z")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Malta IBAN","Malta International Bank Account Number")
```

IP addresses, NPI, names

- IP address
- NPI
- Name

Regular expressions:

- Name: **ipaddress**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A)((25[0-5]|2[0-4][0-9]|[01]?[0-9][0-9]?)\\. (25[0-5]|2[0-4][0-9]|[01]?[0-9][0-9]?)\\. (25[0-5]|2[0-4][0-9]|[01]?[0-9][0-9]?)\\. (25[0-5]|2[0-4][0-9]|[01]?[0-9][0-9]?)\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("ipaddress","ipaddr","ip")
```

- Name: **npi**
 - Match Type: **Column value**
 - Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A)(([12]\\d{9})|((80840)[12]\\d{9})) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**
 - Confidence: 15
 - Regular expression:

```
is_in_ci("npi")
```

National ID numbers

- Bulgaria (BGR)
- Switzerland (CHE)
- Czech Republic (CZE)
- Denmark (DNK)
- Spain (ESP)
- Estonia (EST)
- Finland (FIN)
- Greece (GRC)
- Ireland (IRL)
- Iceland (ISL)
- Italy (ITA)
- Lithuania (LTU)
- Latvia (LVA)
- Norway (NOR)
- Poland (POL)
- Portugal (PRT)
- Romania (ROU)
- Slovakia (SVK)
- Sweden (SWE)

Regular expressions:

- Name: **NOR_NationalID_Detection**
 - Match Type: **Column value**
 - Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A)((0[1-9]|12[0-9]|3[01]) ([0][1-9]|[1][012])\\d{2}\\d{3}\\d{2}) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**
 - Confidence: 15
 - Regular expression:

```
is_in_ci("Norway National ID","Norway National identification number")
```

- Name: **FIN_NationalID_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A)((0[1-9]|[12][0-9]|3[01])((0[1-9]|[1][012])\\d{2}[+\\-A](00[2-9]|[0][1-9]\\d|[1-9]\\d{2})[A-Za-z0-9])(\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Finland National ID","Finland National identification number")
```

- Name: **GRC_NationalID_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A)([ABEZHIKMNOPTYX]{2}\\-\\d{6})(\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Greece National ID","Greece National identification number")
```

- Name: **CHE_NationalID_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A)(756([ -.]?)\\d{4}\\3\\d{4}\\3\\d{2})(\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Switzerland National ID","Switzerland National identification number")
```

- Name: **PRT_NationalID_Detection**

- Match Type: **Column value**

- Confidence: 85
- Regular expression:

```
regex("(\\s|,|;|\\A) (\\d{8} ([ -]?) \\d[0-9a-zA-Z]{2} \\d) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
- Regular expression:

```
is_in_ci("Portugal  
National ID","Portugal National identification number")
```

- Name: **LTU_NationalID_Detection**

- Match Type: **Column value**

- Confidence: 85
- Regular expression:

```
regex("(\\s|,|;|\\A) ([3-6] \\d{2} ([0][1-9] | [1][012]) (0[1-9] | [12][0-9] | 3[01]) \\d{3} \\d) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
- Regular expression:

```
is_in_ci("Lithuania  
National ID","Lithuania National identification number")
```

- Name: **POL_NationalID_Detection**

- Match Type: **Column value**

- Confidence: 85
- Regular expression:

```
regex("(\\s|,|;|\\A) (\\d{2} ([0][1-9] | [1][012]) (0[1-9] | [12][0-9] | 3[01]) \\d{3} \\d{1} \\d{1}) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
- Regular expression:

```
is_in_ci("Poland National ID","Poland National identification number")
```

- Name: **LVA_NationalID_Detection**

- Match Type: **Column value**

- Confidence: 85
- Regular expression:

```
regex("(\\s|,|;|\\A)((0[1-9]|[12][0-9]|3[01])([0][1-9]|[1][012])\\d{2}\\-?[012]\\d{3}\\d)(\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
- Regular expression:

```
is_in_ci("Latvia National ID","Latvia National identification number")
```

- Name: **DNK_NationalID_Detection**

- Match Type: **Column value**

- Confidence: 85
- Regular expression:

```
regex("(\\s|,|;|\\A)(([0][1-9]|[12][0-9]|[3][01])([0][1-9]|[1][012])\\d{2}[ -]?\\d{4})(\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
- Regular expression:

```
is_in_ci("Denmark National ID","Denmark National identification number")
```

- Name: **CZE_NationalID_Detection**

- Match Type: **Column value**

- Confidence: 85
- Regular expression:

```
regex("(\\s|,|;|\\A)(\\d{4}(0[1-9]|[12][0-9]|3[01])\\/?\\d{4})(\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
- Regular expression:

```
is_in_ci("Czech Republic National ID","Czechia National ID","Czech Republic National identification number","Czechia National identification number")
```

- Name: **SWE_NationalID_Detection**

- Match Type: **Column value**

- Confidence: 85
- Regular expression:

```
regex("(\\s|,|;|\\A)((19|20)?\\d{2}(0[1-9]|1[012])([06][1-9]|[12]\\d|3[01])[+-]\\d{4})(\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
- Regular expression:

```
is_in_ci("Sweden  
National ID","Sweden National identification number")
```

- Name: **ISL_NationalID_Detection**

- Match Type: **Column value**

- Confidence: 85
- Regular expression:

```
regex("(\\s|,|;|\\A)((0[1-9]|[12][0-9]|3[01])([0][1-9]|[1][012])\\d{2}\\-\\d{2}\\d{09})(\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
- Regular expression:

```
is_in_ci("Iceland National ID","Iceland National identification number")
```

- Name: **ESP_NationalID_Detection**

- Match Type: **Column value**

- Confidence: 85
- Regular expression:

```
regex("(\\s|,|;|\\A)(\\d{8}([ -]?[A-Z])(\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
- Regular expression:

```
is_in_ci("Spain National ID","Spain National identification number")
```

- Name: **EST_NationalID_Detection**

- Match Type: **Column value**

- Confidence: 85
- Regular expression:

```
regex("(\\s|,|;|\\A) ([1-6]\\d{2} ([0] [1-9] | [1] [012]) (0 [1-9] | [12] [0-9] | 3 [01]) \\d{3} \\d) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
- Regular expression:

```
is_in_ci("Estonia National ID", "Estonia National identification number")
```

- Name: **ITA_NationalID_Detection**

- Match Type: **Column value**

- Confidence: 85
- Regular expression:

```
regex("(\\s|,|;|\\A) ([A-Z]{3} [A-Z]{3} \\d{2} [A-EHLMPRST] ([04] [1-9] | [1256] \\d | [37] [01]) [A-Z] \\d{3} [A-Z]) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
- Regular expression:

```
is_in_ci("Italy National ID", "Italy National identification number")
```

- Name: **NationalID_Detection**

- Match Type: **Column name**

- Confidence: 15
- Regular expression:

```
is_in_ci("National ID", "National identification number")
```

- Name: **ROU_NationalID_Detection**

- Match Type: **Column value**

- Confidence: 85
- Regular expression:

```
regex("(\\s|,|;|\\A) ([1-9] \\d{2} ([0] [1-9] | [1] [012]) (0 [1-9] | [12] [0-9] | 3 [01]) ([01234] \\d | 5 [012] | 99) \\d{4}) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
- Regular expression:

```
is_in_ci("Romania National ID", "Romania National identification number")
```

- Name: **BGR_NationalID_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) (\\d{2}([024][1-9]|[135][012]) (0[1-9]|[12][0-9]|3[01])\\d{3}\\d{1}) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Bulgaria National ID","Bulgaria National identification number")
```

- Name: **IRL_NationalID_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) (\\d{7}[A-Za-z]{1,2}) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Ireland National ID","Ireland National identification number")
```

- Name: **SVK_NationalID_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) (\\d{4}(0[1-9]|[12][0-9]|3[01])\\d{4}) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Slovakia National ID","Slovakia National identification number")
```

National insurance number (UK)

Regular expressions:

- Name: **uknationalinsurancenum**
 - Match Type: **Column value**
 - Confidence: 85
 - Regular expression:

```
regex("\\b[A-CEGHJ-NOPR-TW-Z]{1}[A-CEGHJ-NPR-TW-Z]{1}[0-9]{6}[A-D\\s]{1}\\b")
```

Passport numbers

- Austria (AUT)
- Belgium (BEL)
- Switzerland (CHE)
- Germany (DEU)
- Spain (ESP)
- Finland (FIN)
- France (FRA)
- Greece (GRC)
- Ireland (IRL)
- Italy (ITA)
- Poland (POL)
- United Kingdom (UK)

Regular expressions:

- Name: **AUT_Passport_Detection**
 - Match Type: **Column value**
 - Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) ([A-Z] [ ]?\\d{7}) (\\s|,|;|\\Z) ")
```

- Match Type: **Column name**
 - Confidence: 15
 - Regular expression:

```
is_in_ci("Austria Passport Number")
```

- Name: **ESP_Passport_Detection**
 - Match Type: **Column value**
 - Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) ([A-Z]{2} [ -]?\\d{6}) (\\s|,|;|\\Z) ")
```

- Match Type: **Column name**
 - Confidence: 15
 - Regular expression:

```
is_in_ci("Spain Passport Number")
```

- Name: **DEU_Passport_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) ([0-9CF-HJ-NPRTV-Z]{9}) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Germany Passport Number")
```

- Name: **FIN_Passport_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) ([A-Za-z]{2}\\d{7}) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Finland Passport Number")
```

- Name: **POL_Passport_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) ([A-Za-z]{2}\\d{7}) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Poland Passport Number")
```

- Name: **CHE_Passport_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) ([A-Z]\\d{7}) (\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Switzerland Passport Number")
```

- Name: **Passport_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) ([A-Z] [A-Z0-9] [0-9] {7}) (\\s|,|;|\\A)", "(\\s|,|;|\\A) ([A-Za-z] {2} \\d {7}) (\\s|,|;|\\Z)", "(\\s|,|;|\\A) ([A-Z] [ ]? \\d {7}) (\\s|,|;|\\Z)", "(\\s|,|;|\\A) ([A-Z] {2} [0-9] {7}) (\\s|,|;|\\Z)", "(\\s|,|;|\\A) ([A-Za-z] {2} \\d {7}) (\\s|,|;|\\Z)", "(\\s|,|;|\\A) ([A-Z] {2} [0-9] {7}) (\\s|,|;|\\Z)", "(\\s|,|;|\\A) ([0-9CF-HJ-NPRTV-Z] {9}) (\\s|,|;|\\Z)", "(\\s|,|;|\\A) ([A-Za-z] {2} \\d {6}) (\\s|,|;|\\Z)", "(\\s|,|;|\\A) ([A-Z] {2} [ -]? \\d {6}) (\\s|,|;|\\Z)", "(\\s|,|;|\\A) ([A-Z] \\d {7}) (\\s|,|;|\\Z)", "(\\s|,|;|\\A) (\\d {2} [A-Z] {2} \\d {5}) (\\s|,|;|\\Z) ")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Passport Number")
```

- Name: **GRC_Passport_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) ([A-Z] {2} [0-9] {7}) (\\s|,|;|\\Z) ")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Greece Passport Number")
```

- Name: **BEL_Passport_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A) ([A-Za-z] {2} \\d {6}) (\\s|,|;|\\Z) ")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Belgium Passport Number")
```

- Name: **FRA_Passport_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A)(\\d{2}[A-Z]{2}\\d{5})(\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("France Passport Number")
```

- Name: **ITA_Passport_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A)([A-Z]{2}[0-9]{7})(\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Italy Passport Number")
```

- Name: **IRL_Passport_Detection**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("(\\s|,|;|\\A)([A-Z][A-Z0-9][0-9]{7})(\\s|,|;|\\Z)")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("Ireland Passport Number")
```

Miscellaneous identifications: bank, passport, social security

- SWIFT code
- UK passport number
- Bank Routing Number
- US Social Security Number
- Society for Worldwide Interbank Financial Telecommunication (SWIFT)
- Telephone

Regular expressions:

- Name: **swiftcode**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("\\b[a-zA-Z]{4}(AF|AX|AL|DZ|AS|AD|AO|AI|AQ|AG|AR|AM|AW|AU|AT|AZ|BS|BH|BD|BB|BY|BE|BZ|BJ|BM|BT|BO|BQ|BA|BW|BV|BR|IO|BN|BG|BF|BI|KH|CM|CA|CV|KY|CF|TD|CL|CN|CX|CC|CO|KM|CG|CD|CK|CR|CI|HR|CU|CW|CY|CZ|DK|DJ|DM|DO|EC|EG|SV|GQ|ER|EE|ET|FK|FO|FJ|FI|FR|GF|PF|TF|GA|GM|GE|DE|GH|GI|GR|GL|GD|GP|GU|GT|GG|GN|GW|GY|HT|HM|VA|HN|HK|HU|IS|IN|ID|IR|IQ|IE|IM|IL|IT|JM|JP|JE|JO|KZ|KE|KI|KP|KR|KW|KG|LA|LV|LB|LS|LR|LY|LI|LT|LU|MO|MK|MG|MW|MY|MV|ML|MT|MH|MQ|MR|MU|YT|MX|FM|MD|MC|MN|ME|MS|MA|MZ|MM|NA|NR|NP|NL|NC|NZ|NI|NE|NG|NU|NF|MP|NO|OM|PK|PW|PS|PA|PG|PY|PE|PH|PN|PL|PT|PR|QA|RE|RO|RU|RW|BL|SH|KN|LC|MF|PM|VC|WS|SM|ST|SA|SN|RS|SC|SL|SG|SX|SK|SI|SB|SO|ZA|GS|ES|LK|SD|SR|SJ|SZ|SE|CH|SY|TW|TJ|TZ|TH|TL|TG|TK|TO|TT|TN|TR|TM|TC|TV|UG|UA|AE|GB|US|UM|UY|UZ|VU|VE|VN|VG|VI|WF|EH|YE|ZM|ZW)[a-zA-Z0-9]{2}([a-zA-Z0-9]{3})?\\b")
```

- Match Type: **Column name**

- Confidence: 15
 - Regular expression:

```
is_in_ci("swiftcode")
```

- Name: **ukpassportnumber**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("\\b([A-Z0-9]{1}[0-9]{4}[A-Z0-9]{1})\\b")
```

- Name: **telephone**

- Match Type: **Column value**

- Confidence: 85
 - Regular expression:

```
regex("\\b(((\\+)?(1|7|2([07]|1[12368]|2-6[0-9]|9[01789])|3([57][0-9]|8[0125679]|0[123469])|4([013456789]|2[013])|5([1-8]|39(9)?|0[9][0-9])|6([0-6]|7[8][0-9]|9[0-2])|(758)|8([1246]|0[089]|5[02356]|7(8|[0-4])|8(1[01236789]|21[36]|0[16]))|9(6[0-8]|7[0-7]|0[123458]|9(8|[2-6]))|(1-(68|2(4[16]|6[48])|(34[05])|4(41|73)|6(49|64|7[01]|84)|7(58|67|8[47])|8(0|6[89]|24[9|76]|(939)))|(7-7))\\[ ])?([2-9][0-9]{9}))\\b","\\b(((\\+)?(1|7|2([07]|1[12368]|2-6[0-9]|9[01789])|3([57][0-9]|8[0125679]|0[123469])|4([013456789]|2[013])|5([1-8]|39(9)?|0[9][0-9])|6([0-6]|7[8][0-9]|9[0-2])|(758)|8([1246]|0[089]|5[02356]|7(8|[0-4])|8(1[01236789]|21[36]|0[16]))|9(6[0-8]|7[0-7]|0[123458]|9(8|[2-6]))|(1-(68|2(4[16]|6[48])|(34[05])|4(41|73)|6(49|64|7[01]|84)|7(58|67|8[47])|8(0|6[89]|24[9|76]|(939)))|(7-7))\\[ ])?([2-9][0-9]{2})\\[ ]?([0-9]{3}\\[\\[ ]?([2-9][0-9]{4}))\\b","\\b(((\\+)?(1|7|2([07]|1[12368]|2-6[0-9]|9[01789])|3([57][0-9]|8[0125679]|0[123469])|4([013456789]|2[013])|5([1-8]|39(9)?|0[9][0-9])|6([0-6]|7[8][0-9]|9[0-2])|(758)|8([1246]|0[089]|5[02356]|7(8|[0-4])|8(1[01236789]|21[36]|0[16]))|9(6[0-8]|7[0-7]|0[123458]|9(8|[2-6]))|(1-(68|2(4[16]|6[48])|(34[05])|4(41|73)|6(49|64|7[01]|84)|7(58|67|8[47])|8(0|6[89]|24[9|76]|(939)))|(7-7))\\[ ])?([2-9][0-9]{2})-([0-9]{3}-[0-9]{4}))\\b","\\b(((\\+)?(1|7|2([07]|1[12368]|2-6[0-9]|9[01789])|3([57][0-9]|8[0125679]|0[123469])|4([013456789]|2[013])|5([1-8]|39(9)?|0[9][0-9])|6([0-6]|7[8][0-9]|9[0-2])|(758)|8([1246]|0[089]|5[02356]|7(8|[0-4])|8(1[01236789]|21[36]|0[16]))|9(6[0-8]|7[0-7]|0[123458]|9(8|[2-6]))|(1-(68|2(4[16]|6[48])|(34[05])|4(41|73)|6(49|64|7[01]|84)|7(58|67|8[47])|8(0|6[89]|24[9|76]|(939)))|(7-7))\\[ ])?([2-9][0-9]{2})-([0-9]{3})\\b")
```

```

])|(758)|8([1246]|0[089]|5[02356]|7(8|[0-4])|8(1[01236789]|21[36]|0
16))|9(6[0-8]|7[0-7]|0123458)|9(8|[2-6])|(1-(68|2(4[16]|6[48])|(3
4[05])|4(41|73)|6(49|64|7[01]|84)|7(58|67|8[47])|8(0|6[89]|24[9|76]
|(939)))|(7-7))[\.\[ ]?([2-9][0-9]{2}\\. [0-9]{3}\\. [0-9]{4}))\\b",
"((\\+)?(1|7|2([07]|1[12368]|2-6)[0-9]|9[01789])|3([57][0-9]|8[0125
679]|0123469))|4([013456789]|2[013])|5([1-8]|39(9)?|09[0-9])|6([0
-6]|78[0-9]|9[0-2])|(758)|8([1246]|0[089]|5[02356]|7(8|[0-4])|8(1[
01236789]|21[36]|016)))|9(6[0-8]|7[0-7]|0123458)|9(8|[2-6])|(1-(6
8|2(4[16]|6[48])|(34[05])|4(41|73)|6(49|64|7[01]|84)|7(58|67|8[47])|
8(0|6[89]|24[9|76]|(939)))|(7-7))\\[ ]??(\\([2-9][0-9]{2}\\. [\\ ]
0-9]{3}-[0-9]{4})","((\\+)?(1|7|2([07]|1[12368]|2-6)[0-9]|9[01789])
|3([0123469]|57[0-9]|8[0125679])|4(2[013]|013456789))|5([1-8]|399
|[09][0-9]|39)|6([0-6]|78[0-9]|9[0-2])|(758)|8(0[089]|5[02356]|7([
0-4]|8)|8([016]|1[01236789]|21[36])|[1246])|9(6[0-8]|7[0-7]|9([2-6]|
8)|0123458))|(1-(68|2(4[16]|6[48]|84)|(34[05])|4(41|73)|6(49|64|7[0
1]|84)|7(58|67|8[47])|8(0[89]|24[9|6[89]|76)|(939)))|(7-7))\\. )?((\\
([2-9][0-9]{2}\\. [\\ ]0-9]{3}\\. [0-9]{4})","((\\+)?(1|7|2([07]|1[12
368]|2-6)[0-9]|9[01789])|3([57][0-9]|8[0125679]|0123469))|4([01345
6789]|2[013])|5([1-8]|39(9)?|09[0-9])|6([0-6]|78[0-9]|9[0-2])|(7
58)|8([1246]|0[089]|5[02356]|7(8|[0-4])|8(1[01236789]|21[36])|[016])
)|9(6[0-8]|7[0-7]|0123458)|9(8|[2-6])|(1-(68|2(4[16]|6[48])|(34[05
])|4(41|73)|6(49|64|7[01]|84)|7(58|67|8[47])|8(0|6[89]|24[9|76]|(93
9)))|(7-7)))?((\\([2-9][0-9]{2}\\. [0-9]{3}-[0-9]{4})","((\\+)?(1|7|2(
[07]|1[12368]|2-6)[0-9]|9[01789])|3([0123469]|57[0-9]|8[0125679])
|4(2[013]|013456789))|5([1-8]|399|[09][0-9]|39)|6([0-6]|78[0-9]|9
[0-2])|(758)|8(0[089]|5[02356]|7([0-4]|8)|8([016]|1[01236789]|21[36]
)|[1246])|9(6[0-8]|7[0-7]|9([2-6]|8)|0123458))|(1-(68|2(4[16]|6[48]
|84)|(34[05])|4(41|73)|6(49|64|7[01]|84)|7(58|67|8[47])|8(0[89]|24
9|6[89]|76)|(939)))|(7-7))\\. )?((\\([2-9][0-9]{2}\\. [\\ ]0-9]{3}\\. [
0-9]{4})")

```

- Match Type: **Column name**

- Confidence: 15
- Regular expression:

```
is_in_ci("telephone","tel")
```

- Name: **routingnumber**

- Match Type: **Column value**

- Confidence: 85
- Regular expression:

```
regex("\\b^[459][0-9]{8}\\b")
```

- Name: **ssn**

- Match Type: **Column value**

- Confidence: 85
- Regular expression:

```

regex("\\b(((?!000|666)([0-8][0-9][0-9]))((?!00)([0-9]{2}))((?!00
00)([0-9]{4}))|(98765432[0-9]))\\b","\\b(((?!000|666)([0-8][0-9]
[0-9]))[\\]((?!00)([0-9]{2}))[\\]((?!0000)([0-9]{4})))|(987)-(65)-(

```

```
432) [0-9]))\\b", "\\b(((?! (000|666)) ([0-8] [0-9] [0-9])) - (?!00) ([0-9] {2})) - (?!0000) ([0-9] {4})) | ((987) - (65) - (432) [0-9]))\\b")
```

- Match Type: **Column name**

- Confidence: 15
- Regular expression:

```
is_in_ci("ssn", "ssnnumber", "ssnno")
```

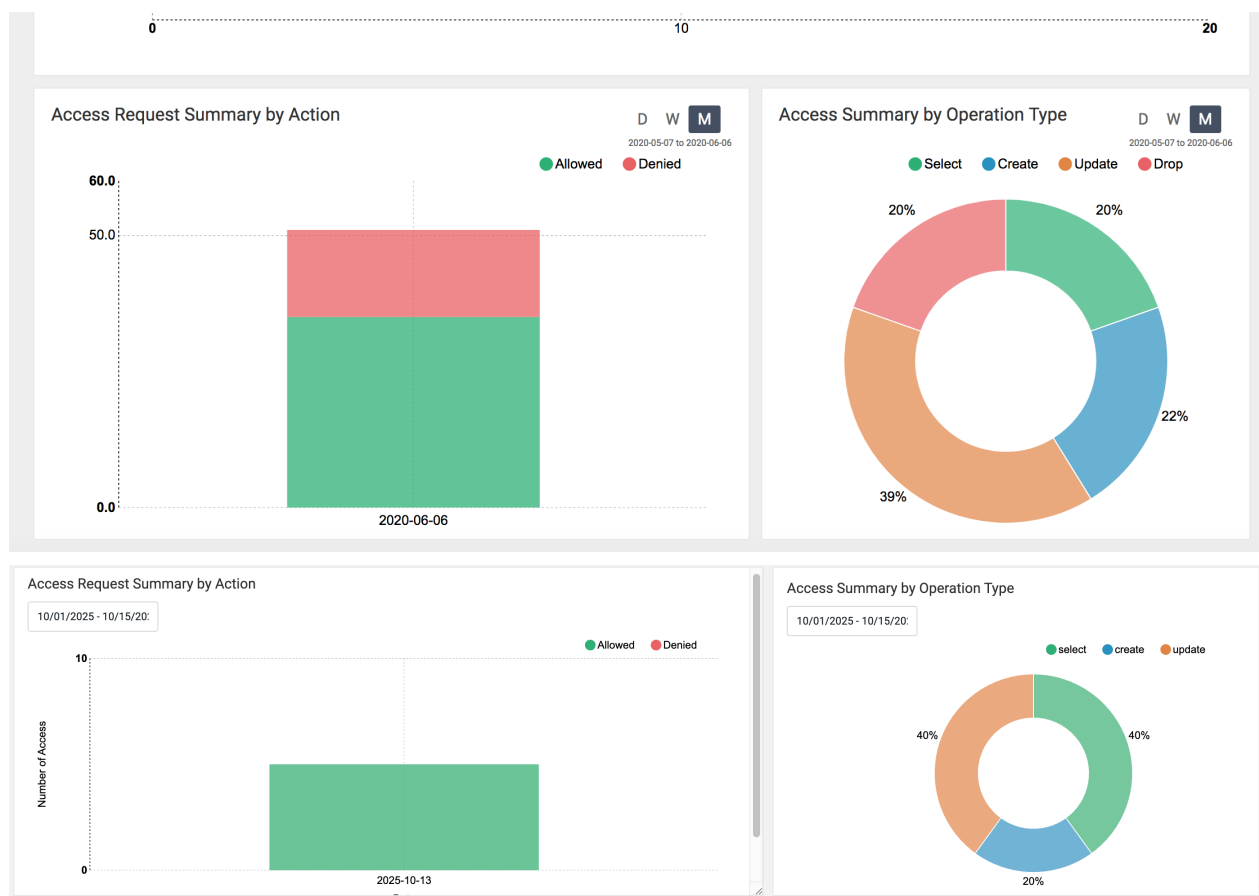
Related Information

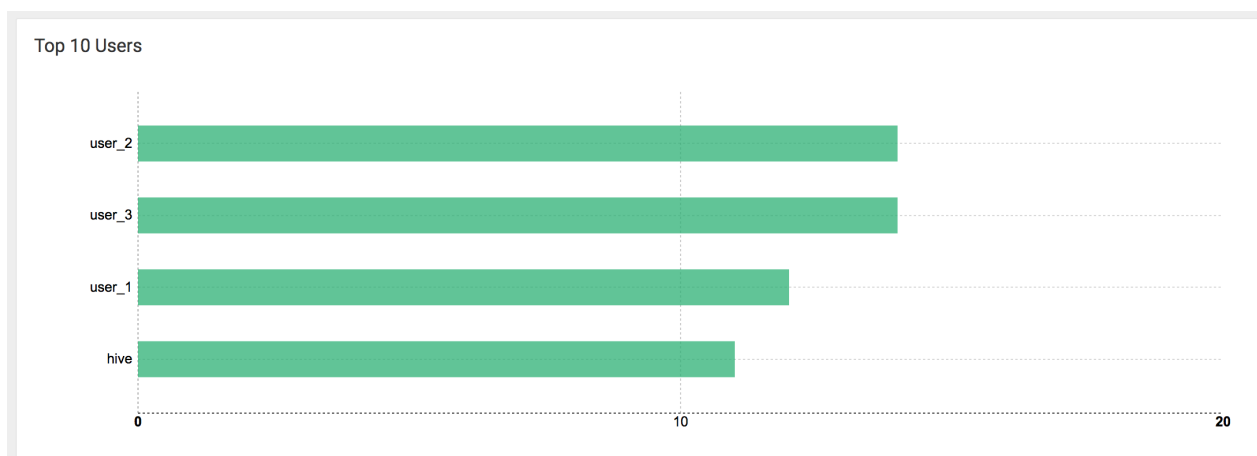
[Cluster Sensitivity Profiler configuration](#)

The Ranger Audit Profiler

You can view who has accessed which data from a forensic audit or compliance perspective, visualize access patterns, and identify anomalies in access patterns using the Ranger Audit Profiler.

The Ranger Audit Profiler uses the Apache Ranger logs to show the most recent raw event data as well as summarized views of event by type of access and access outcomes (allowed/denied). These views are obtained by profiling the records in the data lake.





Related Information

[Ranger Audit Profiler configuration](#)

The Hive Column Profiler

You can view the shape or distribution characteristics of the columnar data within a Hive table based on the .

Categorical Data

The Hive Column Profiler can categorize your Hive tables according to their data type. Datatypes are received from Atlas. These include the following:

- array
- bigint
- binary
- boolean
- char
- date
- decimal
- double
- float
- int¹
- map
- smallint
- string
- struct
- timestamp
- tinyint
- varchar

Numerical Data



Note: Non-primitive data types are not supported. Only basic data types such as numbers, string, boolean, floating point numbers and so on are supported

There are different charts available to help visualize the shape and distribution of the data within the column as well as summary statistics for a column with numerical data. These include the following:

- Unique Values

¹ These rows will receive the boxplot graphs and the min, max mean values after profiling.

- Null Values
- Max
- Min
- Mean

OverviewSchemaMetadata AuditsPolicyAccess Audits

Q Search column

Edit

Chart Type	Name	Type	Unique Values *	Null Values	Max	Min	Mean	Comment	Classifications	Terms
▼	flight_id_raw	string	15	0						
▼	passenger_name_raw	string	22	0						
▼	seat_assignment_raw	string	21	0						
▼	aircraft_model_raw	string	7	0						
▼	origin_airport_raw	string	7	0						
▼	destination_airport_raw	string	12	0						
▼	age_raw	int	20	1	62	25	38.55	Passenger age. PII data subject to ...	+ 1	+ 1
▼	ticket_price_usd_raw	decimal(10,2)	17	0	1,500	85.5	667.35	Total ticket fare in USD. Confident ...	+ 2	+ 1

* Approximate values as being computed using HLL algorithm.

The profiler computes column univariate statistics that are displayed using an appropriate chart in the **Schema** tab.

Pie charts are presented for categorical data with limited number of categories or classes. Examples include data such as eye colors that only have a fixed list of values (categories or labels).

Asset Details

Q Search column

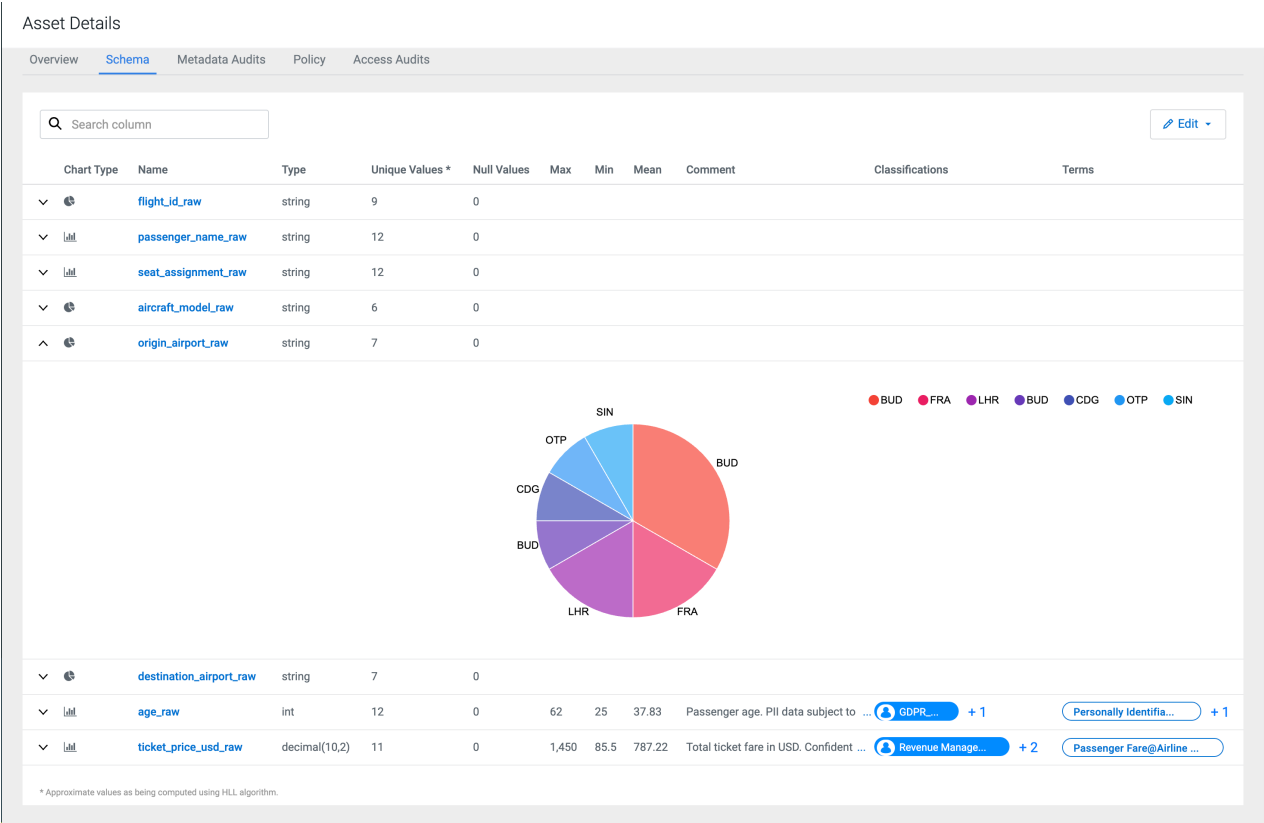
Edit

Chart Type	Name	Type	Unique Values *	Null Values	Max	Min	Mean	Comment	Classifications	Terms
▼	flight_id_raw	string	15	0						
▼	passenger_name_raw	string	24	1						
▼	seat_assignment_raw	string	25	0						
▼	aircraft_model_raw	string	8	0						
▲	origin_airport_raw	string	8	0						

Profiled : 100.0% rows, 52 minutes ago

▼	destination_airport_raw	string	12	0						
▼	age_raw	int	21	1	62	25	39.29	Passenger age. PII data subject to ...	+ 1	+ 1
▼	ticket_price_usd_raw	decimal(10,2)	18	0	1,500	85.5	671.99	Total ticket fare in USD. Confident ...	+ 2	+ 1

* Approximate values as being computed using HLL algorithm.



When the data within columns is numeric, a histogram of the distribution of values organized into 10 groups (decile frequency histogram) and a box plot with a five-number summary (mean, median, quartiles, maximum, and minimum values) are shown for the column, if there are more than 10 elements available.

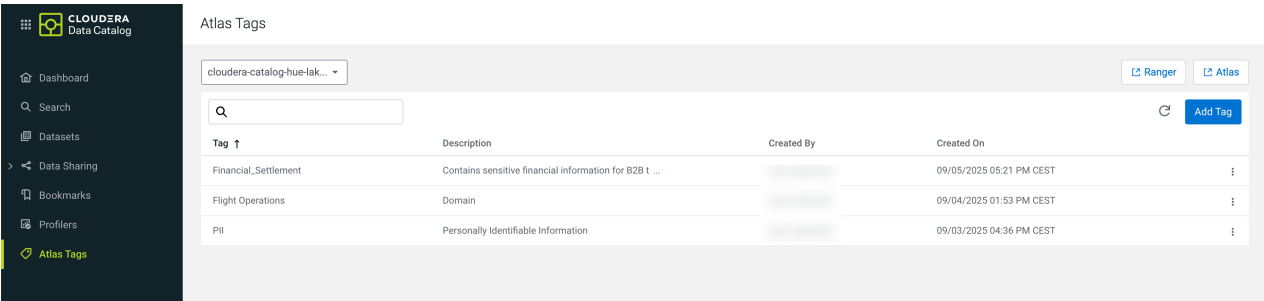




Related Information
[Hive Column Profiler configuration](#)

Atlas tags overview

Atlas tags can created directly from Cloudera Data Catalog which represent Apache Atlas classifications. You can use these as metadata labels to improve searching assets. These tags or classifications can also be automatically applied by using the Data Compliance Profiler or the Cluster Sensitivity Profiler.



Related Information
[Working with Atlas classifications and labels](#)
[Adding attributes to classifications](#)

Before you start

Before you access Cloudera Data Catalog on premises, you must deploy the service.

Before deploying Cloudera Data Catalog, make sure you have reviewed and compiled with the requirements in the installation guide for your environment.

Cloudera Data Catalog is supported on both Openshift and ECS clusters. For more information, see the related documents.

Cloudera Base on premises cluster requirements

The Cloudera Base on premises cluster is used for Cloudera Data Catalog.

As Apache Atlas and Apache Ranger services are installed in the Base cluster, Cloudera Data Catalog communicates with these services to fetch metadata and security policies. Profilers connect with Hive and HDFS services running in the Base cluster for data profiling operations.

For more information, check the following links:

- [Supported base clusters for Data Services and supported Cloudera Manager versions](#)
- [Supported Cloudera Private Cloud Data Services versions](#)

Prerequisites to access Cloudera Data Catalog

To access the Cloudera Data Catalog, you must have the required credentials.

Follow these instructions to provide the required access to the Cloudera Data Catalog users.

The PowerUser must provide the requisite access to subscribers who plan to use Cloudera Data Catalog as per the requirement.



Note: You must be a PowerUser to launch and delete profilers. However, users without this rule can still use the rest of the features of Cloudera Data Catalog.

You need the following additional roles to use Cloudera Data Catalog:

- DataCatalogCspRuleManager

Using this role, you can create, deploy new Cluster Sensitivity Profiler rules, create new regex, and run validations on newly created rules.

- DataCatalogCspRuleViewer

Using this role you can view Cluster Sensitivity Profiler rules.

Additionally, using Cloudera Manager, you must configure Apache Atlas and Apache Ranger services. Use the following instructions to complete the process.

1. Cloudera Manager Clusters Atlas Configuration .
2. Enable the Kerberos Authentication for the Apache Atlas service.

Enable Kerberos Authentication

☒ ATLAS-1 (Service-Wide) ↩

atlas.authentication.method.kerberos

⚙️ [kerberos.auth.enable](#)

- In the search bar, look up the following property: `conf/atlas-application.properties_role_safety_valve` and enter the values for the Atlas service:

`atlas.proxyuser.dpprofiler.hosts=*`

`atlas.proxyuser.dpprofiler.users=*`

`atlas.proxyuser.dpprofiler.groups=*`

Cluster 1

CDEP Deployment from 2023-Oct-22 00:23

Oct 26, 5:26 AM UTC

Search:

Clusters Hosts Diagnostics Audits Charts Replication Administration Data Services **New**

Parcels Running Commands Support admin 7.11.3

Cluster 1

ATLAS-1 Actions

Status Instances Configuration Commands Charts Library Audits Web UI Quick Links

Q conf/atlas-application.properties

Filters

SCOPE

ATLAS-1 (Service-Wide)	0
Atlas Server	1
Gateway	0

CATEGORY

Main	0
Advanced	1
Logs	0
Monitoring	0
Performance	0
Ports and Addresses	0
Resource Management	0
Security	0
Stacks Collection	0

STATUS

Error	0
Warning	0
Edited	1
Non-Default	1
Include Overrides	0

Atlas Server Advanced Configuration Snippet (Safety Valve) for conf/atlas-application.properties

conf/atlas-application.properties_role_safety_valve

Atlas Server Default Group Undo

atlas.proxyuser.dpprofiler.hosts=*

atlas.proxyuser.dpprofiler.users=*

atlas.proxyuser.dpprofiler.groups=*

1 - 1 of 1

1 Edited Value Reason for change: Modified Atlas Server Advanced Configuration Snippet (Safety Valve) for conf/atlas-application.properties Save Changes(CTRL+S)

- For the Apache Ranger configuration updates, go to Cloudera Manager Clusters Ranger Configuration .
- In the search bar, look up the following property: `conf/ranger-admin-site.xml` and enter the values for the Ranger service:

Name: `ranger.proxyuser.dpprofiler.hosts`

Value: `*`

Name: `ranger.proxyuser.dpprofiler.users`

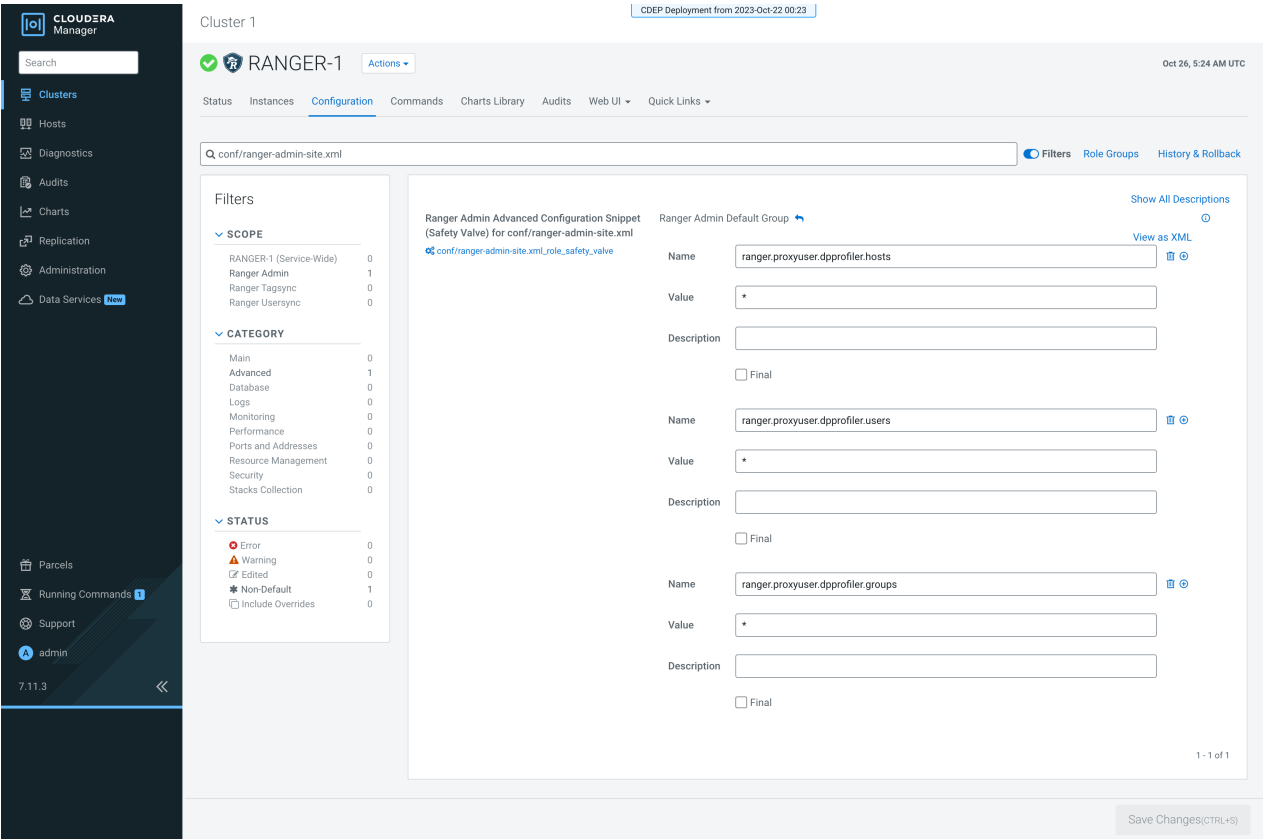
Value: `*`

Name: `ranger.proxyuser.dpprofiler.groups`

Value: `*`



Note: You can add new rows using the + icon.



Later, restart Atlas and Ranger services respectively.

Related Information

[Understanding roles](#)

[Installing on OpenShift](#)

[The Embedded Container Service \(ECS\)](#)

Providing role access

You must have either the EnvironmentAdmin or the EnvironmentUser role access to use Cloudera Data Catalog.

About this task

Compared to EnvironmentUsers, EnvironmentAdmins can manage environments and data lakes. For detailed differences between EnvironmentAdmin and EnvironmentUser roles, see [Understanding roles](#). For more information, see the related documents. The following procedure shows how to assign the EnvironmentAdmin role as a environment resource role. This means that the role will be valid only for a specific environment.

Procedure

1. From Cloudera Management Console Environments > Select an environment > select the Actions drop-down > Manage Access.

2. Search for the user who requires Cloudera Data Catalog access > Select the check-box, either EnvironmentAdmin or EnvironmentUser > Click Update Roles.

Update Roles

×

☒	Role	Description	Policies
☐	DataCatalogCspRuleManager	Grants permission to perform all task on CSP rules in Data Catalog.	Policies
☐	DataCatalogCspRuleViewer	Grants permission to list and view CSP rules in Data Catalog.	Policies
☒	EnvironmentAdmin	Deprecated - Grants permission to perform all task on environments, Data Lake and Data Hub clusters.	Policies
☐	EnvironmentUser	Deprecated - Grants permission to perform read tasks on environments and Data Lake clusters and perform all tasks on Data Hub clusters.	Policies
☐	IamUser	Grants permission to create access keys for the user, view assigned roles, and view all users in the account.	Policies
☐	IamViewer	Grants permission to view access keys, view assigned roles and view all users in the account.	Policies
☒	PowerUser	Grants permission to perform all tasks on all resources in the account.	Policies

Cancel

Update

3. Navigate back to the **Clusters** page and select Actions > select Synchronize Users
Allow the sync operation to complete and the changes to take effect.

Related Information

[Missing authorization for viewing assets](#)

[Understanding roles](#)

[Understanding Cloudera on premises user accounts](#)