

Cloudera Runtime 7.3.1

Cloudera Data Sharing Overview

Date published: 2020-07-28

Date modified: 2024-12-10

CLOUDERA

<https://docs.cloudera.com/>

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

Cloudera Data Sharing overview.....	4
Cloudera Iceberg REST Catalog overview.....	4

Cloudera Data Sharing overview

Cloudera Data Sharing allows secure sharing of Iceberg table data from Cloudera on cloud with external clients using third-party engines that support the Iceberg REST catalog.

Cloudera Data Sharing lets you share Cloudera on cloud data, specifically Iceberg tables, with external users (clients) who are outside of Cloudera environments. As a Cloudera user, you can share Iceberg table data with your clients. These clients can then access the data using third-party engines like Databricks or Snowflake that support the Iceberg REST catalog.

The REST Catalog service is implemented based on the [Iceberg REST Catalog API specification](#). You can make Cloudera Data Sharing available to your clients by using the OAuth authentication mechanism defined by the KNOX Token management system and using Apache Ranger policies for defining the data shares for the clients.

Benefits of Cloudera Data Sharing

- Enables cross-platform data sharing. For example, you can share data from a Cloudera data lake with clients who use other data platforms like Databricks, Amazon EMR, Snowflake, Splunk, and more.
- Provides zero-copy Cloudera live data, meaning no replication is needed, and there is no latency.
- Offers centralized governance of data using Apache Ranger and Atlas
- Secure collaboration between customers and their clients
- Creates a foundation for building a marketplace to share datasets, notebooks, and Cloudera AI models.

Scope and assumptions

- In this release, only read access to Iceberg tables is supported.
- Clients accessing the data can use non-Cloudera engines that understand the Iceberg table format.
- Supports table access audit in Ranger.
- Currently, this feature is supported only on AWS S3 storage in Cloudera on cloud environments.
- Ranger column masking and row-level filtering is not supported. You can restrict the access only to complete tables in Data Share.
- The High Availability feature is not supported for Cloudera Data Sharing powered by REST Catalog.

Related Information

[Cloudera Iceberg REST Catalog](#)

[How Cloudera Data Sharing works](#)

Cloudera Iceberg REST Catalog overview

Cloudera Iceberg REST Catalog is a server-side implementation based on Apache Iceberg REST Catalog Open API, enabling REST-enabled third-party tools to manage Iceberg table metadata.

Apache Iceberg is a table format for huge analytics datasets in the cloud that defines how metadata is stored and data files are organized. Iceberg is also a library that compute engines can use to read or write a table.

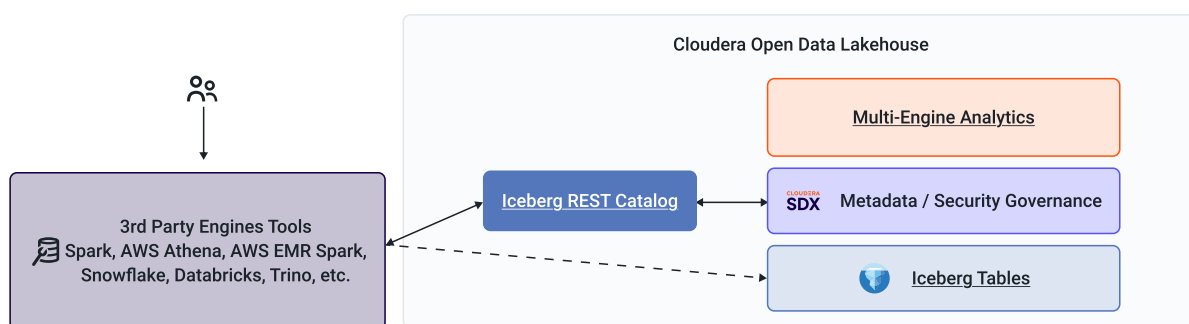
Cloudera supports a data lakehouse architecture by pre-integrating and unifying the capabilities of Data Warehouses and Data Lakes to support data engineering, business intelligence, and machine learning—all on a single platform. Cloudera's support for an open data lakehouse brings high-performance, self-service reporting and analytics to your business, simplifying data management for both data practitioners and administrators. Cloudera's open data lakehouse is built on Apache Iceberg, which makes it easy to manage operational metadata.

Table operations like creating, dropping, or renaming tables are handled by a catalog. An Iceberg catalog helps query engines to manage and organize collections of tables, which are usually grouped into namespaces. Iceberg catalog

implementations can be of various types, such as REST, HiveCatalog, JDBC, HadoopCatalog, Nessie, and so on. For more information about Iceberg catalogs, see the [Apache Iceberg Catalogs documentation](#).

Cloudera supports the Iceberg REST catalog server implementation, which enables the server to expose Iceberg table metadata to REST-enabled third-party compute engines and tools like Spark, Trino, Snowflake, AWS Athena, AWS EMR Spark, and Databricks.

Cloudera Iceberg REST Catalog is a server-side catalog based on the [Apache Iceberg REST Catalog Open API specification](#) and is exposed through REST APIs to manage Iceberg tables. The REST catalog provides API endpoints to perform table management tasks, such as creating, listing, updating, or deleting tables. It also allows users to access and manage Iceberg table metadata.



As part of the current offering, the Iceberg REST Catalog service is provided as an embedded service that runs within the same Java Virtual Machine (JVM) as the Hive Metastore (HMS). The scope of the metadata that the catalog can serve is limited to what the host HMS can serve. The REST catalog can be made available to third-party engines by using the OAuth authentication mechanism defined by the Knox Token management system, and you can use Apache Ranger policies to govern the Iceberg data that is accessed.

The REST service is deployed within the HMS through an embedded jetty engine. The configuration for the REST Catalog server is within the host HMS, and any changes made to the configurations will affect the REST service.

Related Information

[Iceberg REST Catalog API specification](#)

[How Cloudera Data Sharing works](#)