

..

Data services tools

Date published: 2022-03-15

Date modified: 2022-03-15

CLOUDERA

Legal Notice

© Cloudera Inc. 2024. 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

| | |
|--|----------|
| Apache Iceberg in Cloudera Data Platform..... | 4 |
|--|----------|

Apache Iceberg in Cloudera Data Platform

Apache Iceberg is a cloud-native, high-performance open table format for organizing petabyte-scale analytic datasets on a file system or object store. Combined with Cloudera Data Platform (CDP), users can build an open data lakehouse architecture for multi-function analytics and to deploy large scale end-to-end pipelines.

Open data lakehouse on CDP simplifies advanced analytics on all data with a unified platform for structured and unstructured data and integrated data services to enable any analytics use case from ML, BI to stream analytics and real-time analytics. Apache Iceberg is the secret sauce of the open lakehouse.

The following table shows the support for Iceberg in CDP and below the table Iceberg versions v1 and v2 are defined:

Table 1: Iceberg Support Matrix

| Release | Iceberg support level | SQL Engine | | | | |
|--|--|--|--|--|--------------------------------------|--------------------------------------|
| | | Impala | Hive | Spark | NiFi | Flink |
| Public Cloud Data Services | GA | v1, v2: read, insert, delete, update | v1, v2: read, insert, update, and delete | v1, v2: read, insert, update, and delete | v1, v2: read and insert | N/A |
| Data Hub 7.2.16.2 | GA | v1, v2: read | v1: read, insert, update, delete | v1, v2: read, insert, update, and delete | v1, v2: read and insert | v1: read and insert |
| Data Hub 7.2.17 | GA | v1, v2: read, insert, | v1, v2: read, insert, update, delete | v1, v2: read, insert, update, and delete | v1, v2: read and insert | v1, v2: read, append, overwrite *** |
| Data Hub 7.2.18 | GA | v1, v2: read, insert, update, and delete | v1, v2: read, insert, update, delete | v1, v2: read, insert, update, and delete | v1, v2: read and insert | v1, v2: read, append, overwrite *** |
| Private Cloud Data Services 1.5.1 2023.0.13.0-20 | Technical Preview (7.1.7 Base, 7.1.8 Base) | v1, v2: read | v1, v2: read, insert, update, and delete | v1, v2: read, insert, update, and delete | No Private Cloud support | No Private Cloud support |
| Private Cloud Data Services 1.5.2 | GA (7.1.9 Base) Technical Preview (7.1.7 Base, 7.1.8 Base) | v1, v2: read, insert, and delete | v1, v2: read, insert, update, and delete | v1, v2: read, insert, update, and delete | v1, v2: read and insert (7.1.9 Base) | v1, v2: read and insert (7.1.9 Base) |
| Private Cloud Data Services 1.5.3 | GA (7.1.9 Base) Technical Preview (7.1.7 Base, 7.1.8 Base) | v1, v2: read, insert, update, and delete | v1, v2: read, insert, update, and delete | v1, v2: read, insert, update, and delete | v1, v2: read and insert (7.1.9 Base) | v1, v2: read and insert (7.1.9 Base) |
| Base 7.1.7 SP2, 7.1.8 | No Iceberg support | | | | | |
| Base 7.1.9 | GA | v1, v2: read, insert | No Iceberg support | v1, v2: read, insert, update, and delete | v1, v2: read and insert | v1, v2: read and insert |

** The support for delete operations, except from Flink, shown in this table is limited to position deletes. Equality deletes are not supported in these releases except from Flink.

*** Iceberg v2 updates and deletes from Flink are a technical preview in CDP Public Cloud 7.2.17.

The [Apache Iceberg format specification](#) describes the following versions of tables:

- v1
Defines large analytic data tables using open format files.

- v2

Specifies ACID compliant tables including row-level deletes and updates.

Table 2: Iceberg Docs and Availability Matrix

| Release | Docs | Iceberg Support Level |
|---|---|---|
| Open Data Lakehouse (Cloudera Private Cloud Base 7.1.9) | Iceberg in Open Data Lakehouse | GA |
| | Iceberg support for Atlas | GA |
| | SQL Stream Builder with Iceberg (CSA 1.11) and Flink with Iceberg (CSA 1.11) Iceberg replication policies | GA |
| Data Engineering (CDE) Public Cloud | Using Iceberg | GA |
| Data Warehouse (CDW) Public Cloud | Iceberg features | GA |
| Data Engineering (CDE) Private Cloud | Using Iceberg | Technical Preview |
| Data Warehouse (CDW) Private Cloud | Iceberg introduction Moving data into Iceberg tables on CDW | GA (7.1.9 Base), Technical Preview (7.1.7-7.1.8 Base) |
| Public Cloud Data Hub 7.2.16 and later | Iceberg features | Technical Preview |
| Public Cloud Data Hub 7.2.17 and later | Iceberg in Apache Atlas | Technical Preview |
| | Streaming Analytics Iceberg support in Flink | GA |
| | Flink/Iceberg connector | GA |
| | Using NiFi to ingest Iceberg data | GA |
| Public Cloud Data Hub 7.2.18 | Iceberg in Apache Atlas | GA |
| Flow Management for CDP Private Cloud | Technical preview features | Technical Preview |
| DataFlow (CDF) Public Cloud | Using the PutIceberg processor | GA |
| Flow Management 2.1.5 and later for CDP Private Cloud | Using NiFi to ingest Iceberg data | GA |
| Cloudera Machine Learning (CML) Public Cloud | Connection to Iceberg | GA |