

..

Data

Date published: 2020-10-30

Date modified: 2022-09-21

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

Datasets in Data Visualization.....	4
Data modeling in Data Visualization.....	4
Data joins in Data Visualization.....	5

Datasets in Data Visualization

Datasets are the foundation and starting point for visualizing your data. They are defined on the connections to your data and provide access to the specific tables in the data store.

A dataset is the logical representation of the data you want to use to build visuals. It is a logical pointer to a physical table or a defined structure in your data source. Datasets may represent the contents of a single data table or a data matrix from several tables that may be in different data stores on the same connection.

Other than providing access to data, datasets enhance data access and use in many ways, including (but not limited to):

- Table joins allow you to supplement the primary data with information from various other data sources. For more information, see *Data modeling*.
- Derived fields/attributes support flexible expressions, both for dimensions and for aggregates. For more information, see *Creating calculated fields*.
- Hiding fields enables you to eliminate the fields that are unnecessary to the business use case or to obscure sensitive data without affecting the base tables. For more information, see *Hiding dataset fields from applications*.
- Changing data types of the field attributes often helps you to deal with data types, or to ensure that numeric codes (like event ids) are processed correctly. For more information, see *Changing data type*.
- Changing the default aggregation of fields at the dataset level prevents common mistakes when building visuals. For more information, see *Changing field aggregation*.
- Providing user-friendly names for columns or derived attributes often makes the visuals more accessible and saves some of the efforts of applying aliases to each field of the visual. For more information, see *Automatically renaming dataset fields* and *Custom renaming dataset fields*.

Related Information

[Data modeling](#)

[Creating calculated fields](#)

[Hiding dataset fields from applications](#)

[Changing data type](#)

[Changing field aggregation](#)

[Automatically renaming dataset fields](#)

[Custom renaming dataset fields](#)

Data modeling in Data Visualization

With CDP Data Visualization, you can create logical datasets that model semantic relationships across multiple data sources. These datasets enable you to build blended visuals, dashboards, and applications.

You can expand a basic, one-table dataset by creating joins with other relevant tables from the same source or from other data stores. The joins created in CDP Data Visualization are not materialized; they are calculated during run-time.

Combining data across multiple tables enriches the dataset considerably; it enables you to conduct much more meaningful research and produce insightful visual analytics.

There are two distinct approaches to create data joins for visualization:

- Defining in UI is ideal for datasets that include star-type schemas.
- Defining on back-end ETL is preferable for fact-fact joins, so they may be pre-materialized.

See *Working with data models* for instructions on how to create and manage joins.

See topics in *How To: Advanced Analytics* for more advanced data modeling details, such as dimension hierarchies, segments, and events.

See *Setting filter associations* for information about filter associations related to data modeling and management.

Related Information

[Working with data models](#)

[How To: Advanced Analytics](#)

[Setting filter associations](#)

Data joins in Data Visualization

With CDP Data Visualization, you can create joins between different table columns.

Data Visualization supports five types of column connections, also known as joins.



Note: The types of joins available depend on the underlying database. For example: MySQL connections do not support FULL OUTER JOIN operations. SQLite connections do not support FULL OUTER JOIN and RIGHT OUTER JOIN operations.

Inner join

It is the most common join type. Rows in the result set contain the requested columns from both tables for all combinations of rows where the selected columns of the tables have identical values.

Left join

The result set contains all rows from the left table and the matching data from the right table. Whenever no matching data is available on the right side of the join, the corresponding columns in the result set have the value NULL.

Right join

The result set contains all rows from the right table and the matching data from the left table. Whenever no matching data is available on the left side of the join, the corresponding columns in the result set have the value NULL.

Outer join

An outer join returns all rows from both tables. Whenever no matching data is available on one of the sides of the join, the corresponding columns in the result set have the value NULL.

Left outer join

It supports join elimination in datasets that contain more than one table. By default, this setting is on, but it can be disabled.

Related Information

[Working with data models](#)