

Working with Data

Date published: 2020-10-30

Date modified: 2025-09-30

CLOUDERA

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

Working with datasets in Cloudera Data Visualization.....	5
Creating a dataset.....	5
Creating a dataset from a table.....	5
Creating a dataset from a query.....	7
Finding a dataset.....	9
Cloning a dataset.....	10
Deleting a dataset.....	10
Exploring dataset details.....	13
Tables.....	16
Connection.....	16
Description.....	17
Join elimination.....	17
Result cache.....	17
SQL for first table.....	18
Information on creation date and updates.....	18
Checking related dashboards.....	18
Tracking dataset versions [Technical Preview].....	19
Changing dataset fields.....	21
Editing dataset fields.....	21
Hiding dataset fields from applications.....	25
Changing field aggregation.....	29
Creating calculated fields.....	33
Changing the field display format.....	39
Changing data type.....	49
Specifying geographic fields.....	50
Adding field comments.....	52
Automatically renaming dataset fields.....	56
Custom renaming dataset fields.....	58
Refreshing a dataset.....	60
Editing datasets based on a SQL query.....	62
Restricting columns in datasets based on SQL query.....	63
Restricting rows in datasets based on SQL query.....	66
 Working with data models in Cloudera Data Visualization.....	 68
Creating a join.....	69
Changing join types.....	73
Editing join details.....	75
Deleting a join.....	79
Applying field display format on sample data.....	80
 Working with data profiling.....	 82
Using profiling.....	82
Using correlation.....	85
 Working with data extracts in Cloudera Data Visualization.....	 85
Creating a data extract.....	86

Exploring data extract details.....	90
Running a data extract.....	90
Creating a dataset from a data extract.....	92
Editing a data extract.....	92
Deleting a data extract.....	94

Working with datasets in Cloudera Data Visualization

Datasets are defined on the connections to your data, and provide access to the specific tables in the data store.

In Cloudera Data Visualization, visualizations are built from datasets. These datasets provide access to data and enhance data access and usage.

Related Information

[Datasets](#)

Creating a dataset

Cloudera Data Visualization allows you to create datasets based on your data available through a data connection.

About this task

There are two options to create a new dataset:

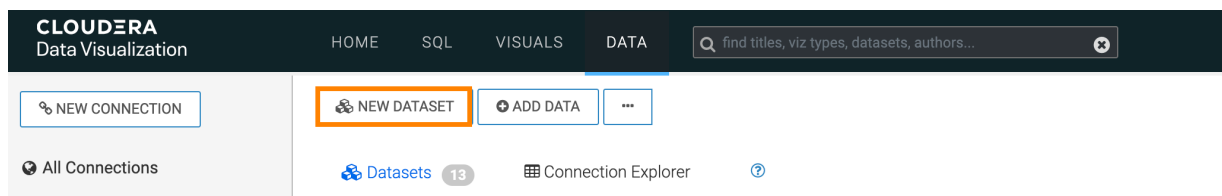
- [Creating a dataset from a table](#) on page 5
- [Creating a dataset from a query](#) on page 7

Creating a dataset from a table

Cloudera Data Visualization allows you to define a new dataset based on an existing table.

Procedure

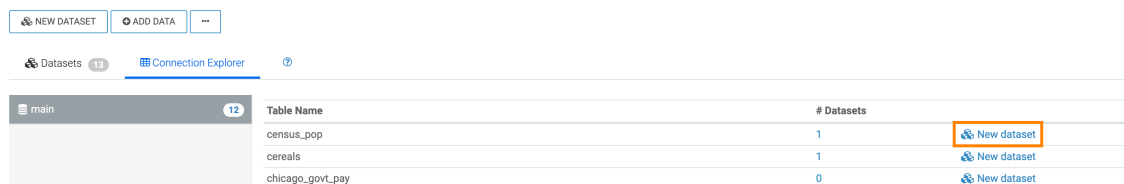
1. On the main navigation bar, click DATA.
The Data view opens, displaying the Datasets tab.
2. Click NEW DATASET near the top of the screen.



The New Dataset modal window opens.



Note: You can also open this modal window from the Connection Explorer, by clicking New dataset next to the table's name.



3. In the New Dataset modal window, specify the following values:

Dataset title

Provide a name for the new dataset.

Dataset Source

Open the menu and select the From Table option.

Select Database

Scroll down the list of connected databases to select the correct database.

Select Table

Scroll down the list of tables to select the correct table.

New Dataset

Create a dataset from data on this connection. You need to create a dataset before you can create dashboards or apps.

Dataset title *

Dataset Source

From Table ▾

Select Database

main ▾

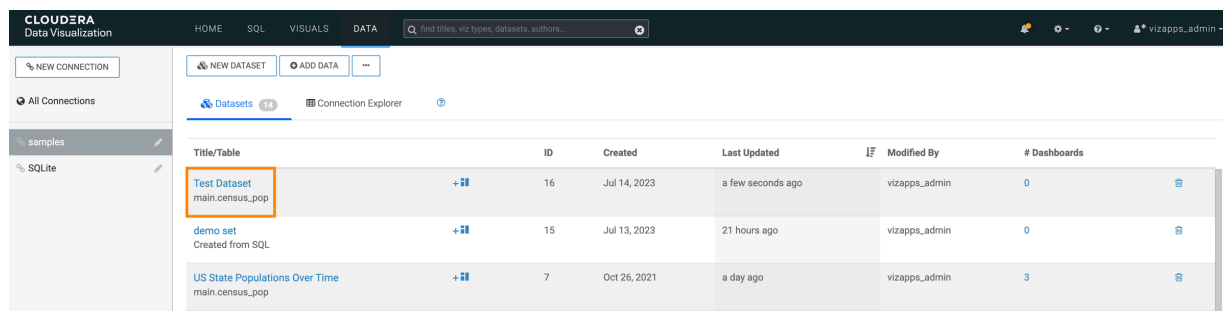
Select Table

census_pop ▾

CANCELCREATE

4. Click CREATE.

You can now see the new dataset on the Datasets tab.



Title/Table	ID	Created	Last Updated	Modified By	# Dashboards
Test Dataset main.census_pop	16	Jul 14, 2023	a few seconds ago	vizapps_admin	0
demo set Created from SQL	15	Jul 13, 2023	21 hours ago	vizapps_admin	0
US State Populations Over Time main.census_pop	7	Oct 26, 2021	a day ago	vizapps_admin	3



Tip: To find the dataset in the future, you can scroll through the list of datasets available on the connection, or use Search at the top of the page.

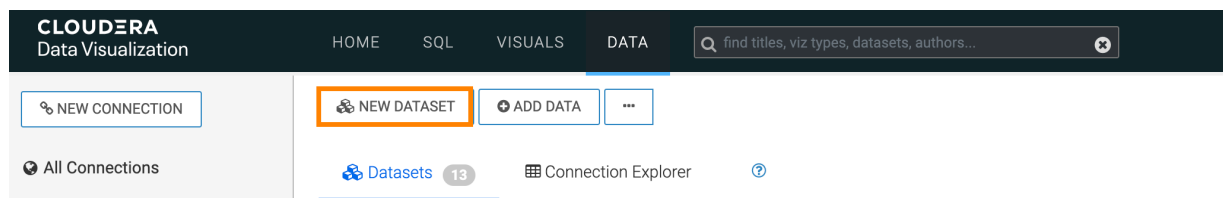
Creating a dataset from a query

Cloudera Data Visualization allows you to define a new dataset using a SQL query. This feature makes it easy to restrict access to specific table columns or rows for all users, either for security or for relevancy reasons. It also enables you to specify complex joins and analytic functions at the dataset level.

Using the DATA tab

Procedure

1. On the main navigation bar, click DATA.
The Data view opens, displaying the Datasets tab.
2. Click NEW DATASET near the top of the screen.



NEW CONNECTION	NEW DATASET	ADD DATA	...
All Connections	Datasets 13	Connection Explorer	?

The New Dataset modal window opens.



Note: You can also open this modal window from the Connection Explorer, by clicking New dataset next to the table's name.

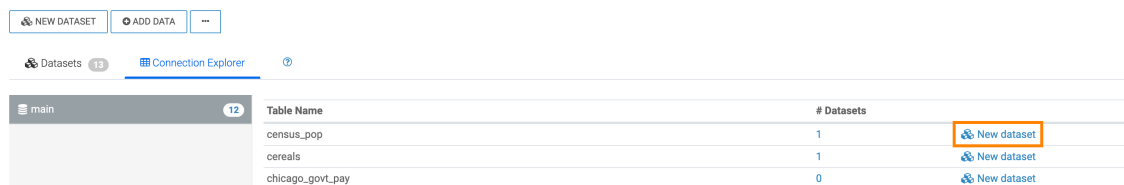


Table Name	# Datasets	
census_pop	1	New dataset
cereals	1	New dataset
chicago_govt_pay	0	New dataset

3. In the New Dataset modal window, specify the following values:

Dataset title

Provide a name for the new dataset.

Dataset Source

Open the menu and select the From SQL option.

Enter SQL below

- a. Enter the SQL query you want to use to define the new dataset.
- b. Run the query.

New Dataset

Create a dataset from data on this connection. You need to create a dataset before you can create dashboards or apps.

Dataset title *

Test Dataset from SQL

Dataset Source

From SQL

Enter SQL below

```
select * from main.us_counties
```

☒ Autocomplete on

CANCEL CREATE

4. Click CREATE.

You can now see the new dataset on the Datasets tab.



Tip: To find the dataset in the future, you can scroll through the list of datasets available on the connection, or use Search at the top of the page.

Using the SQL tab

Procedure

1. On the main navigation bar, click SQL.
The SQL view opens.
2. Select the database and the table.
3. Add the SQL query to the SQL text field.

4. Click RUN.

The screenshot shows the Cloudera Data Visualization interface. At the top, there's a navigation bar with 'HOME', 'SQL', 'VISUALS', and 'DATA'. Below this, the 'Data Connection' is set to 'samples'. The 'Connection Explorer' is open, showing a list of tables under the 'main' database. The 'SQL' tab is selected, and the query editor contains the following SQL: `1 SELECT * FROM main.census_pop`. Below the query editor, there are buttons for 'RUN', 'SAVE QUERY', 'SAVE AS DATASET', and '+ NEW DASHBOARD'. The 'RUN' button is highlighted. Below the buttons, the 'Results' tab is selected, showing the query results in a table format. The table has columns 'year', 'state', and 'population'. The results are as follows:

year	state	population
1960	AL	3266740
1960	AK	226167
1960	AZ	1302161
1960	AR	1786272
1960	CA	15717204
1960	CO	1753947
1960	CT	2535234

5. Click SAVE AS DATASET.

The New Dataset modal window opens.

6. Add a title for the dataset.

7. Click CREATE.

You can now see the new dataset on the Datasets tab.



Tip: To find the dataset in the future, you can scroll through the list of datasets available on the connection, or use Search at the top of the page.

Finding a dataset

Cloudera Data Visualization makes it easy to find your datasets. You can browse the list of Datasets, check out the Connection Explorer, or use Search.

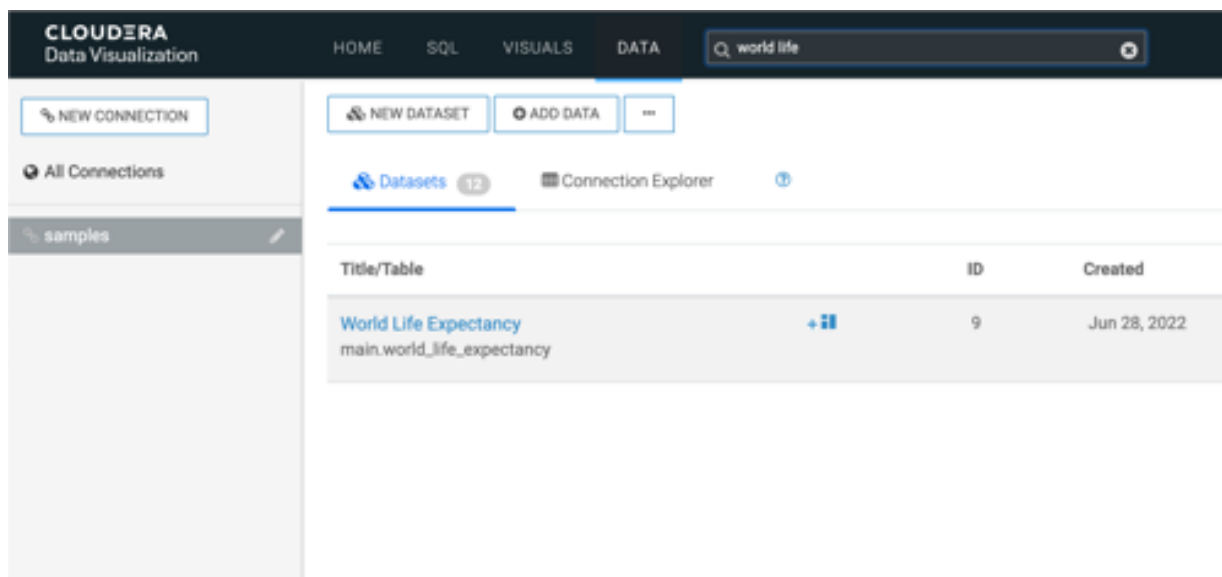
Procedure

1. On the main navigation bar, click DATA.

The Data view opens, displaying the Datasets tab.

2. On the left-side panel, select the connection on which the dataset is defined.

3. Use Search on the main navigation bar. Type the search string that matches all or part of the dataset you are looking for.



Some examples of search criteria are for example data connection names, data table names, dataset names, visualization names, or application names.

4. Select the correct dataset from the abbreviated list of datasets.

Cloning a dataset

Cloning a dataset in Cloudera Data Visualization enables you to quickly create a duplicate dataset to use as a foundation for new analyses, testing, or experimentation. This is especially useful for exploring different transformations, filters, or display options without affecting the original dataset.

Procedure

1. On the main navigation bar, click DATA.
The Data view opens, displaying the Datasets tab.
2. Find the dataset you want to clone, by either browsing through the connection and the dataset lists, or by using the Search function.
3. Click the name of the dataset.
The dataset side navigation panel appears, open on the Dataset Detail view.
4. Click Actions Clone Dataset in the top-right corner.
The original dataset is cloned. The cloned dataset will appear with the name "Clone of [original dataset name]" and will have a new ID, inheriting all other settings and attributes of the original dataset.

Deleting a dataset

In Cloudera Data Visualization, you can delete a dataset without removing the underlying data in the database.

About this task

You can delete a dataset from either the Datasets view or the Dataset Detail view.




Note: Deleting a dataset removes all associated dashboards and visuals, both linked and unlinked. However, it does not delete the data from the database, and any analytical views defined on the dataset will be preserved.

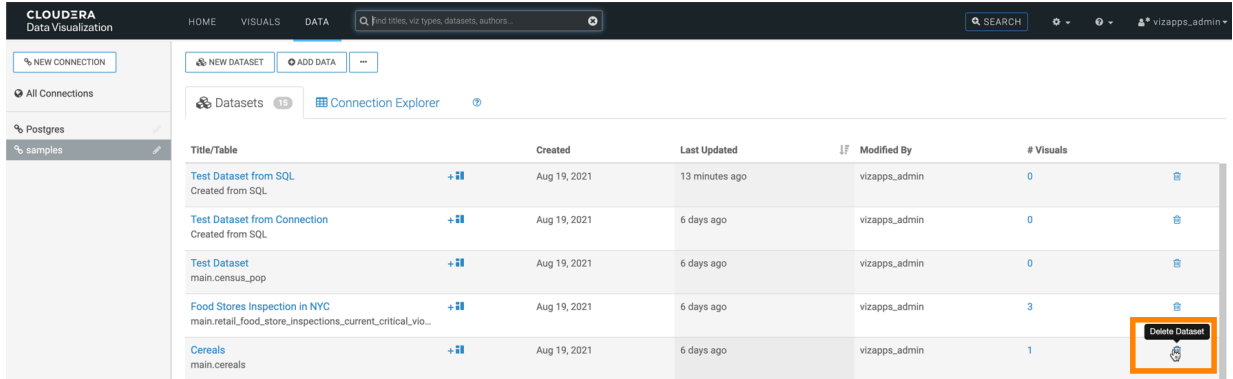
Procedure

1. On the main navigation bar, click DATA.

The Data view opens, displaying the Datasets tab.

2. Find the dataset that you want to delete by either scrolling through the list or using the search function.

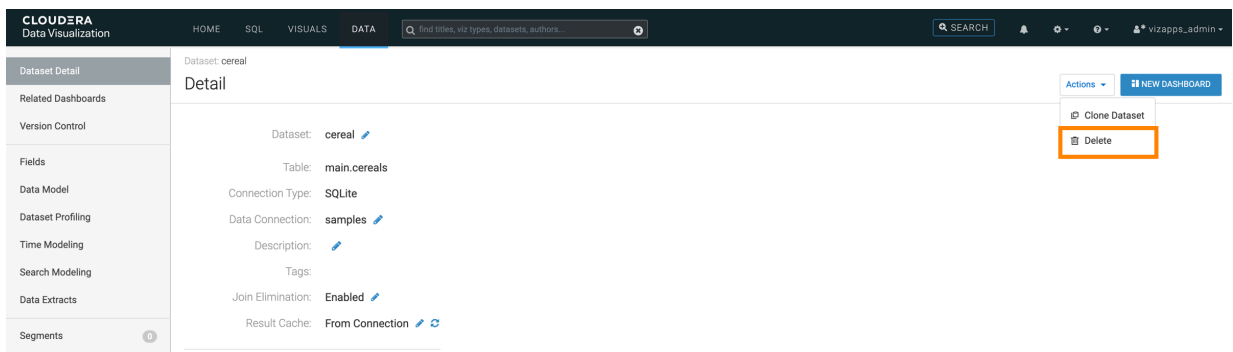
3. On the row of the dataset, click the  icon to delete the dataset.



The screenshot shows the Cloudera Data Visualization interface. On the left, there's a sidebar with 'NEW CONNECTION', 'All Connections', 'Postgres', and 'samples'. The main area has tabs for 'NEW DATASET', 'ADD DATA', and 'Connection Explorer'. Below these is a table of datasets. The 'Cereals' dataset is highlighted, and a 'Delete Dataset' button is visible in the bottom right corner of its row.

Title/Table	Created	Last Updated	Modified By	# Visuals
Test Dataset from SQL Created from SQL	Aug 19, 2021	13 minutes ago	vizapps_admin	0
Test Dataset from Connection Created from SQL	Aug 19, 2021	6 days ago	vizapps_admin	0
Test Dataset main.census_pop	Aug 19, 2021	6 days ago	vizapps_admin	0
Food Stores Inspection in NYC main.retail_food_store_inspections_current_critical_vio...	Aug 19, 2021	6 days ago	vizapps_admin	3
Cereals main.cereals	Aug 19, 2021	6 days ago	vizapps_admin	1

Alternatively, you can click the dataset to open the Dataset Detail view, then click **Actions Delete**.






The screenshot shows the Cloudera Data Visualization interface with the 'Dataset Detail' view for 'cereal'. The left sidebar has 'Dataset Detail', 'Related Dashboards', 'Version Control', 'Fields', 'Data Model', 'Dataset Profiling', 'Time Modeling', 'Search Modeling', 'Data Extracts', and 'Segments'. The main area shows details for the 'cereal' dataset, including 'Table: main.cereals', 'Connection Type: SQLite', 'Data Connection: samples', 'Description:', 'Tags:', 'Join Elimination: Enabled', and 'Result Cache: From Connection'. An 'Actions' menu is open, showing 'Clone Dataset' and 'Delete' (highlighted).

The Delete Dataset modal window opens displaying information about the dashboards and visuals linked to the dataset that will also be deleted with the dataset.

Delete Dataset
×




Dataset **census** will be deleted

Dashboards
The following dashboards related to this dataset will also be deleted:

	Title ↑	ID	Related Visuals	Created	Last Updated	Modified by	Workspace
		1075	1074	Apr 05, 2022	2 years ago	vizapps_admin	Private
		3747	3746	Aug 25, 2023	9 months ago	vizapps_admin	Private
	1130	1311	1310	Jun 28, 2022	2 years ago	vizapps_admin	Private

Showing 1 to 3 of 10 entries Previous [1](#) [2](#) [3](#) [4](#) Next

Visuals
The following visuals related to this dataset will also be deleted:

	Title ↑	ID	Parent Dashboard(s)	Created	Last Updated	Modified by	Workspace
	Racial profile for <<state>>	40	43	May 03, 2021	3 years ago	vizapps_admin	Public
	Racial profile for <<state>>	2309	2313	Feb 21, 2023	a year ago	vizapps_admin	Private
	Single state with counties	36	43	May 03, 2021	3 years ago	vizapps_admin	Public

Showing 1 to 3 of 8 entries Previous [1](#) [2](#) [3](#) Next

To confirm, type the word DELETE in the field below and then click the Delete button.

Cancel
Delete

- *Snapshot* shows an icon indicating whether the artifact is a visual or a dashboard.
 - *Title* shows the name of the artifact that will also be deleted.
 - *ID* shows the identification number of the artifact that will also be deleted.
 - *Related Visuals* / *Parent Dashboard(s)* list the number of related artifacts and their IDs.
 - For dashboards, *Related Visuals* include the IDs of visuals contained within the dashboard.
 - For visuals, *Parent Dashboard(s)* include the IDs of dashboards where the visuals appear.
 - *Created* shows the creation date of the artifact.
 - *Last Updated* shows the time of the last modification.
 - *Modified By* shows the username of the last person who modified the artifact.
 - *Workspace* indicates where the artifact is located.
4. After you reviewed the listed dashboards and visuals, type DELETE in the confirmation text field and click Delete to permanently remove the dataset and its associated artifacts.

Exploring dataset details

Cloudera Data Visualization makes it easy to check out and examine the information available about your datasets.

Procedure

1. On the main navigation bar, click DATA.

The Data view opens, displaying the Datasets tab.

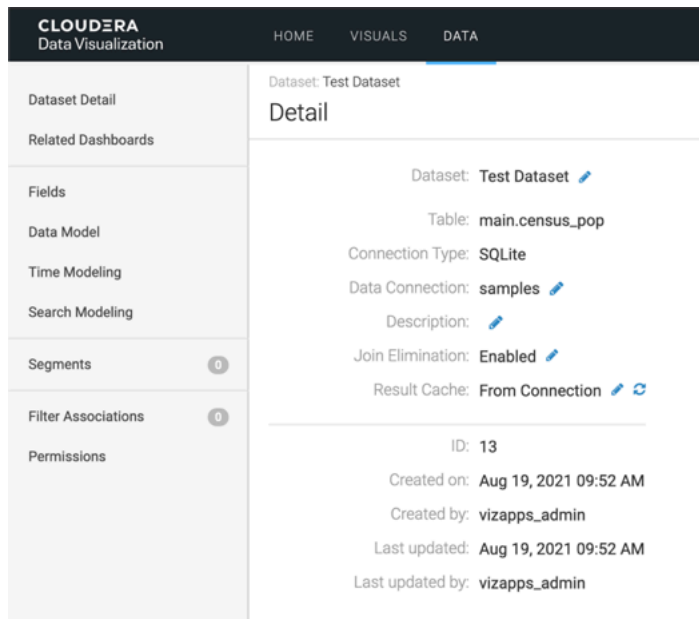
2. Find the dataset you are interested in, either by browsing to a known connection and scrolling, or by using Search.

3. Click the dataset name.

The dataset side navigation panel appears, open on Dataset Detail view.

Depending on the dataset definition approach, there are two alternatives:

- Defined on Table



- Defined on SQL

The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'VISUALS', and 'DATA'. The left sidebar contains a menu with 'Dataset Detail' (selected), 'Related Dashboards', 'Fields', 'Data Model', 'Time Modeling', 'Search Modeling', 'Segments' (0), 'Filter Associations' (0), and 'Permissions'. The main content area is titled 'Dataset: Test Dataset from SQL' and 'Detail'. It displays the following information:

- Dataset: Test Dataset from SQL
- Connection Type: SQLite
- Data Connection: samples
- Description:
- Join Elimination: Enabled
- Result Cache: From Connection
- SQL for first table: `select * from main.us_counties`
- ID: 14
- Created on: Aug 19, 2021 09:55 AM
- Created by: vizapps_admin
- Last updated: Aug 19, 2021 09:55 AM
- Last updated by: vizapps_admin

The following information is available on Dataset Detail view, under Detail:

Tables

This is the qualified name of the data source. It appears in the form `DatabaseName.DatabaseTable`.


Connection

This is the name of the database that hosts the data that appears in the form `DataConnection`. This feature is ideal for enterprise environments, with dashboards developed on test clusters, and then deployed to a production environment.

About this task

This assumes that the new connection has, at the minimum, the relevant base tables with metadata definitions that match those on the original connection. Users with appropriate permissions can switch the data connection of a dataset by following these steps.


Procedure

1. Click the  icon.
2. Choose a different data connection from the menu.
3. Click Save.

Description

This is an optional field. You can add a description of the dataset in the available textbox.

Procedure

1. Click the  icon.
2. Enter the description in the text box.
3. Click Save.

Join elimination

Join elimination improves query execution and visual rendering in Cloudera Data Visualization.

About this task

Join elimination is available both for left outer and inner joins. It is turned on by default.

When a visual uses fields and expressions that reference only a subset of the joined tables that form the dataset, this feature eliminates the unnecessary joins and access only the necessary subset of the joined tables. This improves query execution speeds, and renders the visuals faster.


Procedure


1. Click the Edit (pencil) icon.
2. You can enable or disable join elimination:
 - Disabling: select the Disabled option.
 - Enabling: select the Enabled option.
3. Click Save.

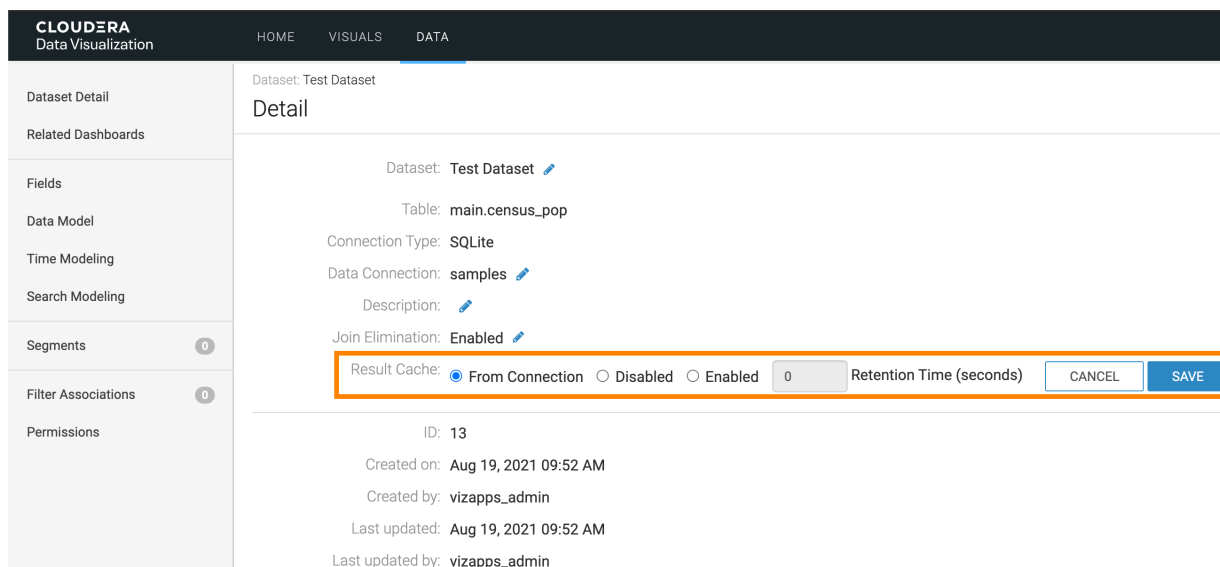
Result cache

Each dataset inherits the result caching preferences configured for its connection. The value for this field is From Connection by default, but this may be changed at the level of the dataset.

Procedure

1. Click the  icon.
2. You can enable or disable the result cache:
 - Disabling: select the Disabled option.
 - Enabling: select the Enabled option, and specify the Retention Time, in seconds.
3. Click SAVE.


4. To clear the result cache, click the  icon.



SQL for first table

For datasets created from a SQL query, you can modify the query at any time to adjust fields, update the WHERE clause, change the ORDER BY sequence, and more.

Procedure

1. Click the  icon.
2. Make the desired changes to the SQL statement.
3. Click SAVE to apply the changes.

To test your updates, click RUN before saving the updated SQL query.

Information on creation date and updates

The Dataset Detail interface provides information about the dataset's creation and its most recent updates.

ID

The unique identification number assigned to the dataset.

Created on

The exact date and time (in timestamp format) when the dataset was created.

Created by

The username of the user who initially created the dataset.

Last updated

The date and time (in timestamp format) of the most recent update made to the dataset.

Last updated by

The username of the user who last modified or updated the dataset.

Checking related dashboards

In Cloudera Data Visualization, you can easily determine which visuals use a particular dataset.

Procedure

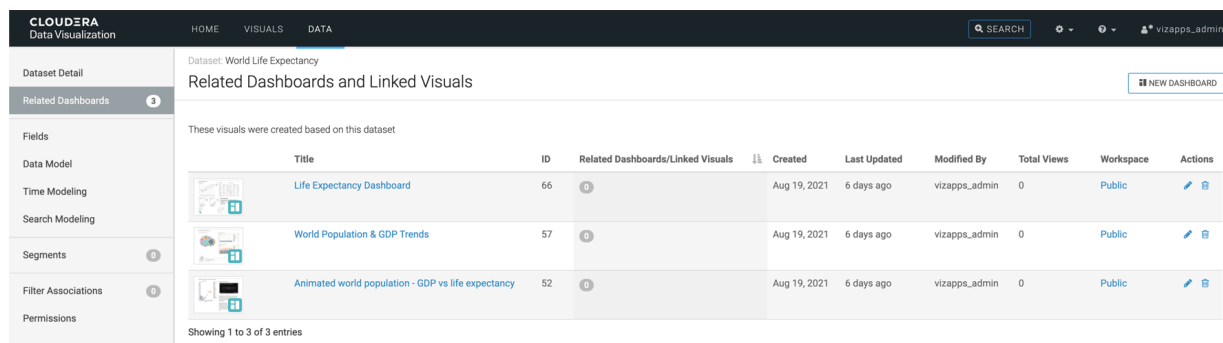
1. On the main navigation bar, click Data.

The Data view opens, displaying the Datasets tab.

2. Find the dataset that you want to examine in the list of datasets, either by scrolling or by using search.
3. Click the dataset.

The dataset side navigation panel appears, open on the Dataset Detail view.

4. In the side navigation menu, click Related Dashboards.



The Related Dashboards view opens, showing a list of visuals that use this dataset.

The following information is available for each app in this list, under Related Dashboards:

- Type icon represents the style of the dashboard, or a snapshot icon of the visual (if this feature is on).
- Title is the name of the visual.
- ID
- Related Dashboards/Linked Visuals
- Created is the date when the visual was created.
- Last Updated is the time interval after the last update of the visual. It is expressed in minutes, hours, days, or months, as appropriate.
- Modified by is the name of the user who modified the app most recently.
- Total Views is the number of times the app was viewed.
- Workspace
- Actions are the permissions available to you They may include the following:
 - Clicking Edit (pencil) icon edits the visual.
 - Clicking Delete (trash) icon deletes the visual.

Tracking dataset versions [Technical Preview]

Cloudera Data Visualization supports dataset versioning, a feature that helps track changes, enable collaborative dataset development, and ensure the integrity of your datasets over time.

About this task

Dataset versioning allows you to manage multiple versions of a dataset as it evolves, which is particularly useful when dealing with frequent updates or iterative changes. Every time a dataset is modified, a new version is automatically created, preserving the previous state. Dataset versioning provides a collaborative environment by ensuring that changes made by different team members can be tracked, and previous states can be restored if necessary.

Key features of dataset versioning:

- **Track changes:** Monitor how data structures and configurations evolve over time.

- **Restore previous versions:** Revert to earlier versions to correct errors or undo unwanted changes.
- **Compare versions:** Review structural differences between dataset versions to understand the impact of changes before switching.



Note: Dataset versioning is currently available as a technical preview.

Follow these steps to dataset versioning in Cloudera Data Visualization.

Before you begin

To use this feature, you must first enable dataset version control in the site settings. For instructions on how to enable and configure dataset version control, see [Managing version control site settings](#).

Accessing dataset versioning

1. On the main navigation bar, click DATA.

The Data view opens, displaying the Datasets tab.

2. Find the dataset that you want to review, by scrolling through the list or using the search function.
3. Click the dataset name to view its details.

The dataset side navigation pane opens for the selected dataset, displaying the Dataset Detail page.

4. Click Version Control in the side navigation to view version history.

Viewing version history

The Version Control page shows the details of the current dataset version along with any previous versions, if available. Metadata includes version name, dataset name, IDs, creation and update timestamps, and user info.


- **Current version:** The first version of a dataset is always the active or current version, and it remains the current version, even if unnamed, until it is replaced by a new version. It is marked with a green Current Version tag.
- **Previous versions:** Each dataset modification (for example editing a dataset field, the data model or the time model, or adding new segments) triggers the creation of a new version of the dataset. A snapshot of the dataset's state prior to the change is saved as a previous version, and the new modification becomes the current version.



Note: Changes to filter associations or dataset permissions do not trigger a new dataset version.

Managing versions


- **Sorting:** Click any column header to sort the version list.
- **Filtering:** Use the search bar to find versions by name, ID, or creator.
- **Naming:** Versions are unnamed by default and display a timestamp as the name. Unnamed versions may be

automatically deleted based on version control site settings. To preserve a version, click the  icon next to its name and assign a custom name.


You can also manage versions using the action icons located at the end of each version row.

Reverting to a previous version


To make a previous version the current one:

1. Click the  icon next to the version you want to reinstate.
2. Review the version details in the Change Dataset Versions modal window.
3. Click Resume to apply the change or Cancel to discard the action.

Deleting dataset versions

Click the  icon to remove a previous dataset version. For bulk deletion, select multiple dataset versions and click Delete to clean up version history.

Comparing dataset versions

To view the differences between the current and a previous version of a dataset, click the  icon next to the previous version you want to compare with the current one.

The Compare Datasets page shows the changes using color-coded sections to help you assess the impact of modifications before deciding to restore a previous version.

- **Added** (green): Entries that exist in the selected version but not in the current one.
- **Removed** (red): Entries that are present in the current version but not in the selected one.
- **Modified** (yellow): Entries that exist in both versions but have differences.

Each section can be expanded to view a detailed list of the changes.

Changing dataset fields

It is easy to make changes to the fields of a dataset in Cloudera Data Visualization.

Editing dataset fields

Editing dataset fields allows you to tailor your data to better suit specific analytical needs and visualization requirements. By modifying field parameters, you can enhance the accuracy, clarity, and relevance of your datasets.

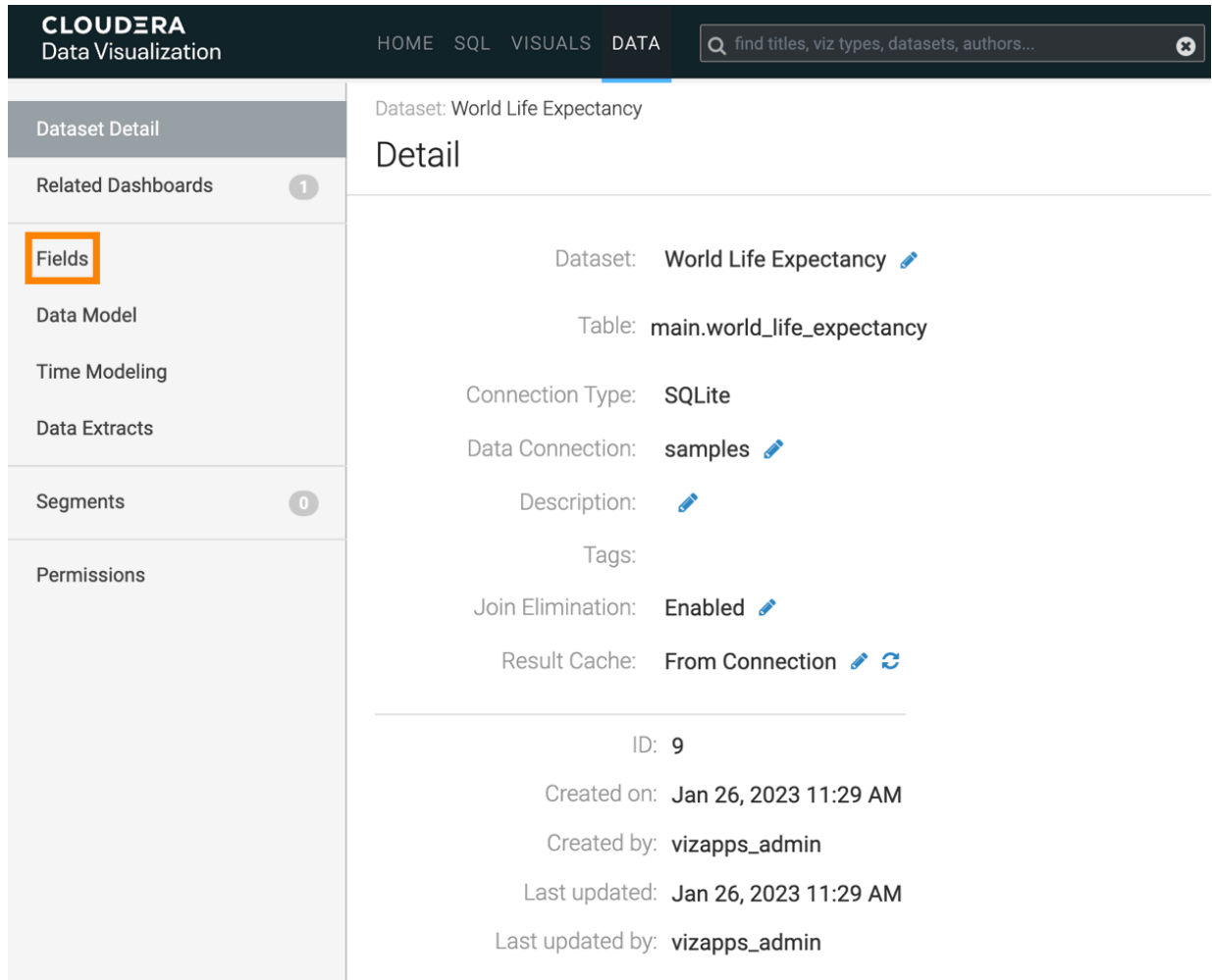
Procedure

1. Navigate to the Data view by clicking DATA on the main navigation bar.

The Data view opens, displaying the Datasets tab.

2. Select a connection from the left navigation menu.
3. In the Datasets area, select the dataset you want to modify.

4. In the Dataset Detail menu, select Fields.



The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'SQL', 'VISUALS', and 'DATA'. A search bar is present on the right. The left sidebar contains a 'Dataset Detail' section with the following items: 'Related Dashboards' (1), 'Fields' (highlighted with an orange box), 'Data Model', 'Time Modeling', 'Data Extracts', 'Segments' (0), and 'Permissions'. The main content area displays the 'Detail' view for the 'World Life Expectancy' dataset. The details include: Dataset: World Life Expectancy, Table: main.world_life_expectancy, Connection Type: SQLite, Data Connection: samples, Description: (empty), Tags: (empty), Join Elimination: Enabled, and Result Cache: From Connection. Below these details, the ID is 9, and the creation and update information is as follows: Created on: Jan 26, 2023 11:29 AM, Created by: vizapps_admin, Last updated: Jan 26, 2023 11:29 AM, and Last updated by: vizapps_admin.

CLOUDERA
Data Visualization

HOME SQL VISUALS DATA

find titles, viz types, datasets, authors...

Dataset Detail

Related Dashboards 1

Fields

Data Model

Time Modeling

Data Extracts

Segments 0

Permissions

Dataset: World Life Expectancy

Detail

Dataset: World Life Expectancy

Table: main.world_life_expectancy

Connection Type: SQLite

Data Connection: samples

Description:

Tags:

Join Elimination: Enabled

Result Cache: From Connection

ID: 9

Created on: Jan 26, 2023 11:29 AM

Created by: vizapps_admin

Last updated: Jan 26, 2023 11:29 AM

Last updated by: vizapps_admin

- Click EDIT FIELDS to access the editing interface.

The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'SQL', 'VISUALS', and 'DATA'. A search bar is present with the text 'find titles, viz types, datasets, authors...'. The left sidebar contains a menu with 'Dataset Detail', 'Related Dashboards' (1), 'Fields' (selected), 'Data Model', 'Time Modeling', 'Data Extracts', 'Segments' (0), and 'Permissions'. The main content area shows the 'Fields' tab for the 'World Life Expectancy' dataset. The 'EDIT FIELDS' button is highlighted with an orange box. Below it, a list of fields is shown under the 'Dimensions' section, including 'country', 'year', 'country_5', 'alt_names', 'code2', 'code3', 'fips_code', 'fips_country_name', 'un_region', 'un_subregion', and 'comments'.


This action adds an edit button (pencil) to each dataset field row.

The screenshot shows the 'Fields' tab for the 'World Life Expectancy' dataset. Each field row now includes an edit button (pencil icon) on the right side. The fields listed are: 'country', 'year', 'country_5', 'alt_names', 'code2', 'code3', 'fips_code', 'fips_country_name', 'un_region', 'un_subregion', and 'comments'. Each row also has a 'Dim' button and a dropdown menu on the left side.

- Click the edit button for the specific field you want to update.
The Edit Field Parameters modal window opens.

7. Make your changes as needed.

- You can adjust basic field settings such as name, default aggregation, or geo type, and add comments to the field.
- You can mark a field as 'Sensitive' to restrict its usage in specific visual shelves that are designated as 'Sensitive'. For example the Embedding Context shelf of the Artificial Intelligence (AI) Assistant visual is a sensitive shelf.

When you mark a field as 'Sensitive', the  icon is displayed next to it.

- You can switch to the Display Format tab to assign a category to the field or to the Color tab to enable custom colors for the field.

Edit Field Parameters

Basic Settings

Display Format

Color

Base Column: code2

Display Name

code2

Field Comment

Enter field comment

Default Aggregation


Maximum

Geo Type

None

☒ Show field in data detail screen

☒ Show field in Visual Designer

☒ Sensitive field 

Category

☒ Dimension ☐ Measure

CANCEL

APPLY

8. Click APPLY.

Hiding dataset fields from applications

You may find it useful to hide dataset fields that are not typically used for visualizations to prevent unintended bias in BI and analytics, or even to obscure confidential data. In Cloudera Data Visualization, you can do this by turning off the default visibility option of a particular dataset field.

About this task

The following steps demonstrate how to prevent data fields from appearing in visualizations and applications of dataset World Life Expectancy [data source samples.world_life_expectancy]. The fields comments, lat, and lng are empty, so they are good candidates for this operation.

Procedure

1. On the main navigation bar, click DATA.
- The Data view opens, displaying the Datasets tab.
2. In the left navigation menu, click Samples.
3. In the Datasets area, select World Life Expectancy (main.world_life_expectancy).
4. In the Dataset Detail menu, select Fields.

The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'SQL', 'VISUALS', and 'DATA'. The left sidebar contains a 'Dataset Detail' section with options: 'Related Dashboards' (1), 'Fields' (highlighted with an orange box), 'Data Model', 'Time Modeling', and 'Data Extracts'. Below this are 'Segments' (0) and 'Permissions'. The main content area is titled 'Dataset: World Life Expectancy' and 'Detail'. It displays the following information:

- Dataset: World Life Expectancy (with an edit icon)
- Table: main.world_life_expectancy
- Connection Type: SQLite
- Data Connection: samples (with an edit icon)
- Description: (with an edit icon)
- Tags:
- Join Elimination: Enabled (with an edit icon)
- Result Cache: From Connection (with edit and refresh icons)

Below the horizontal line, the following metadata is shown:

- ID: 9
- Created on: Jan 26, 2023 11:29 AM
- Created by: vizapps_admin
- Last updated: Jan 26, 2023 11:29 AM
- Last updated by: vizapps_admin

5. In the Fields interface, select EDIT FIELDS.

CLUSTERA
Data Visualization

HOME SQL VISUALS DATA

find titles, viz types, datasets, authors...

Dataset Detail

Related Dashboards 1

Fields

Data Model

Time Modeling

Data Extracts

Segments 0

Permissions

Dataset: World Life Expectancy

Fields

EDIT FIELDS

Hide Comments

Dimensions

world_life_expectancy 11

country

year

country_5

alt_names

code2

code3

fips_code

fips_country_name

un_region

un_subregion

comments

6. Under Dimensions, find the field comments.

7. Click the (eye) icon on the comments line.

Dataset Detail

Related Dashboards 3

Fields

Data Model

Time Modeling

Segments 0

Filter Associations 0

Permissions

Dataset: World Life Expectancy

Fields

UNDO REFRESH TITLE CASE SAVE Show Comments

To add a new calculated field, use the down arrow to the right of a field to clone it, and then edit the expression of the cloned field.

Dimensions

world_life_expectancy 11

country year country_5 alt_names code2 code3 fips_code fips_country_name un_region un_subregion comments

Measures

world_life_expectancy 7

life_expectancy gdp_per_capita population iso_cc cdh_id lat lng

Clone

Hide

Create Hierarchy

The icon next to the comment field changes to (slashed eye).

Dimensions

world_life_expectancy 11

country year country_5 alt_names code2 code3 fips_code fips_country_name un_region un_subregion

comments

8. Under Measures, find the fields lat and lng, and hide them.



9. Click Save.

Results

In the updated Fields interface, Dimensions table has a total of 11 fields and Measures table lists a total of 7 fields, as before.

CLOUDERA Data Visualization HOME SQL VISUALS DATA

Dataset: World Life Expectancy

Fields [EDIT FIELDS](#) [Show Comments](#)

Dimensions	
▼ world_life_expectancy 11	
A	country
A	year
A	country_5
A	alt_names
A	code2
A	code3
A	fips_code
A	fips_country_name
A	un_region
A	un_subregion
A	comments

Measures	
▼ world_life_expectancy 7	
1.2	life_expectancy
1.2	gdp_per_capita
1.2	population
#	iso_cc
#	cdh_id
1.2	lat
1.2	lng

- The number of Dimensions is 11. This is calculated as All Dimensions (11) - Hidden Dimensions (1) + Segment (1).
- The number of Measures is 6. This is calculated as All Measures (7) - Hidden Measures (2) + Record Count (1).

However, when using Visual Designer, the hidden fields do not show.

The screenshot shows the Cloudera Data Visualization interface. The main area displays a table visualization titled 'enter title...' with data for Afghanistan from 1900 to 1905. The table has columns: country, year, life_expectancy, gdp_per_capita, population, country_5, and alt. The right sidebar shows the 'Dashboard Designer' with the 'World Life Expectancy' dataset selected. The 'Dimensions' section lists fields like country, year, country_5, alt_names, code2, code3, fps_code, fps_country_name, un_region, and un_subregion. The 'Measures' section lists fields like Record Count, life_expectancy, gdp_per_capita, population, iso_cc, and cdi_id.

**Tip:**

The field visibility can be changed in the Edit Field Parameters window modal.

Changing field aggregation

You can change the basic fields defaults in your dataset.

About this task

The following steps demonstrate how to change the default aggregation function from Sum to Average for the field life_expectancy in the dataset World Life Expectancy [data source: main.world_life_expectancy].

Procedure

1. On the main navigation bar, click DATA.
The Data view opens, displaying the Datasets tab.
2. In the left navigation menu, click samples.
3. In the Datasets area, select World Life Expectancy (samples.world_life_expectancy).

4. In the Dataset Detail menu, select Fields.

The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'SQL', 'VISUALS', and 'DATA'. A search bar on the right contains the text 'find titles, viz types, datasets, authors...'. The left sidebar contains a 'Dataset Detail' menu with the following items: 'Dataset Detail', 'Related Dashboards' (with a count of 1), 'Fields' (highlighted with an orange box), 'Data Model', 'Time Modeling', 'Data Extracts', 'Segments' (with a count of 0), and 'Permissions'. The main content area is titled 'Dataset: World Life Expectancy' and 'Detail'. It displays the following information:

- Dataset: World Life Expectancy
- Table: main.world_life_expectancy
- Connection Type: SQLite
- Data Connection: samples
- Description:
- Tags:
- Join Elimination: Enabled
- Result Cache: From Connection

ID: 9

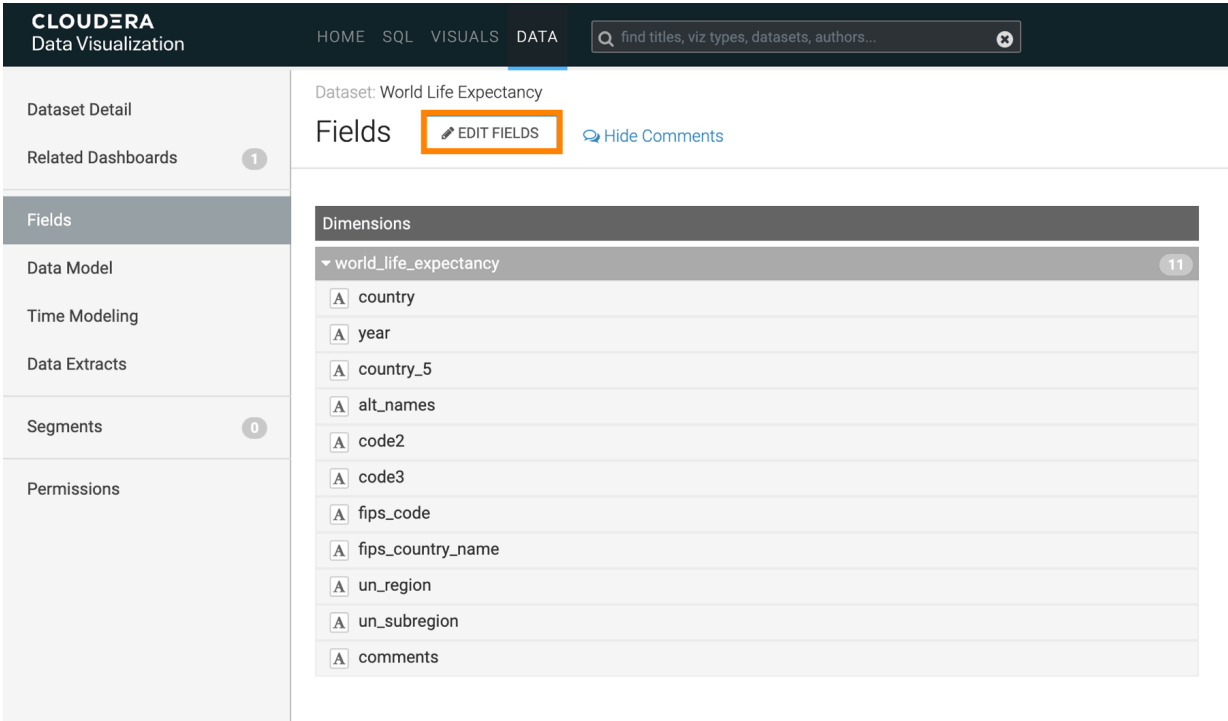
Created on: Jan 26, 2023 11:29 AM

Created by: vizapps_admin

Last updated: Jan 26, 2023 11:29 AM

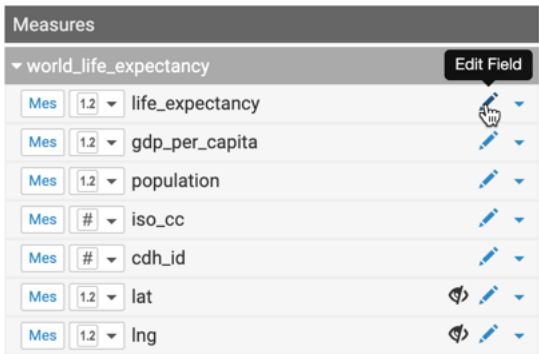
Last updated by: vizapps_admin

5. In the Fields interface, select EDIT FIELDS.



6. Under Measures, find the field life_expectancy and click the (down arrow) icon on its right side.

7. Click the Edit (pencil) icon.



The Edit Field Parameters window modal opens.

8. Change Default Aggregation from Sum to Average, and click APPLY.

Edit Field Parameters

Basic Settings Display Format Color

Base Column: life_expectancy

Display Name
life_expectancy

Field Comment
Enter field comment

Default Aggregation

- ✓ Sum
- Count
- Approx Distinct Count
- Exact Distinct Count
- Minimum
- Maximum
- Average
- String Concat

Category
☐ Dimension ☒ Measure

CANCEL APPLY

9. Under Dataset: World Life Expectancy, click SAVE.

Results

As a result of this change, all new visuals created from this dataset will automatically use the new aggregation.

Dashboard Designer

VISUALS

Table

1234

Dimensions

drag fields to add here

Measures

avg(life_expectancy)

Tooltips

drag fields to add here

Filters

drag fields to add here

Limit: 100

REFRESH VISUAL

DATA

World Life Expectancy

Search

Dimensions 10

world_life_expectancy

- country
- year
- country_5
- alt_names
- code2
- code3
- fips_code
- fips_country_name
- un_region
- un_subregion

Measures 6

world_life_expectancy

- # Record Count
- 1.2 life_expectancy
- 1.2 gdp_per_capita
- 1.2 population
- # iso_cc
- # cdh_id

Creating calculated fields

In Cloudera Data Visualization, you can easily create a new calculated field in the dataset, and subsequently use it in dashboards and visuals.

About this task

Sometimes the data in the base tables cannot be used directly, and you must use an expression to change or "correct" it. For other use cases, you can create a calculation based on one or more fields. Instead of adding these expression for every visual, you can create a new calculated field in the dataset.

The following steps demonstrate how to create a new field gdp (gross domenstic product) in the dataset World Life Expectancy [data source main.world_life_expectancy]. We define it by the following equation:

```
gdp = gdp_per_capita x population
```

Cloudera Data Visualization supports three primary methods of editing fields at the dataset level: Basic, Expression, and Display Format. In this example, we make changes both on the Basic and Expression tabs. For innformationn on how to use Display Format options, see *Changing the Field Display Format*.

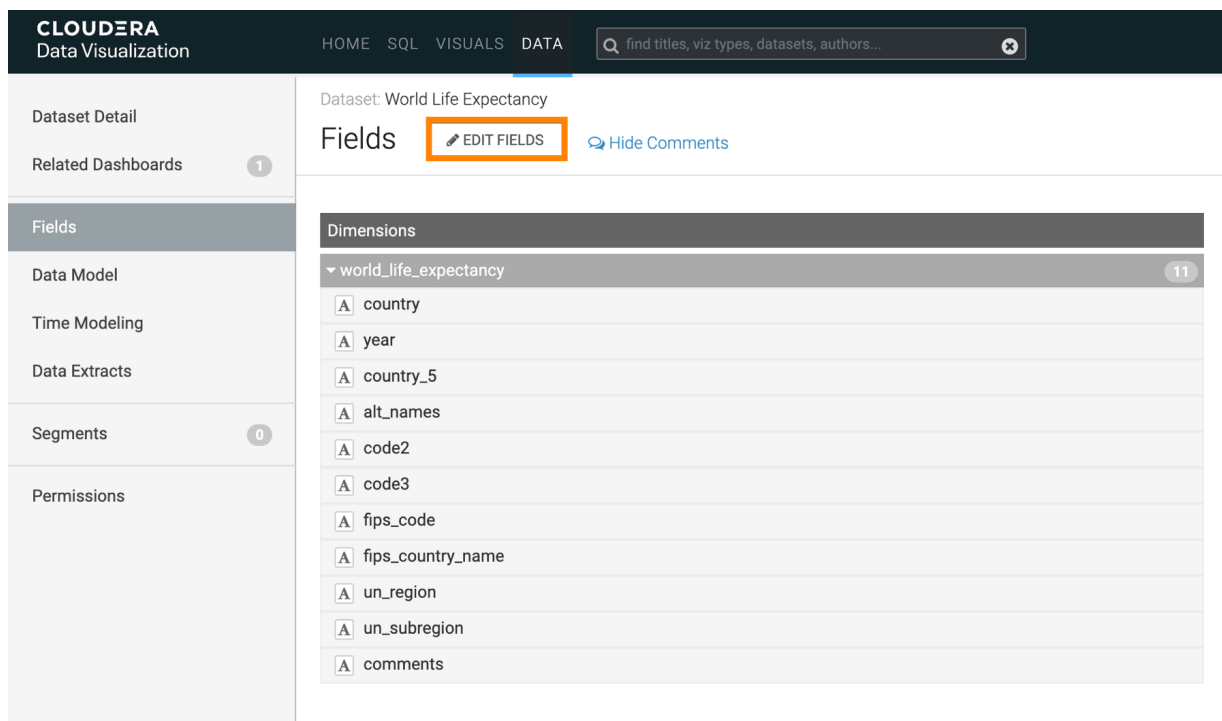
Procedure

1. On the main navigation bar, click DATA.
2. In the left navigation menu, click samples.
3. In the Datasets area, select World Life Expectancy (main.world_life_expectancy).

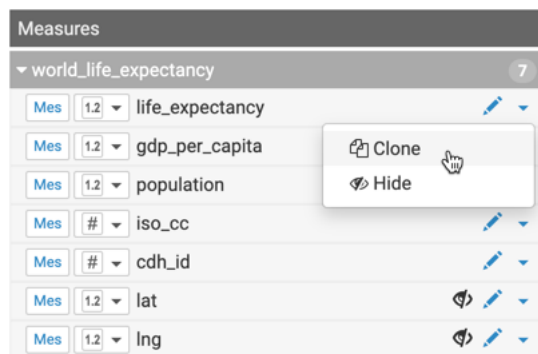
4. In the Dataset Detail menu, select Fields.

The screenshot displays the Cloudera Data Visualization interface. The top navigation bar includes the Cloudera logo, the text 'Data Visualization', and navigation links for HOME, SQL, VISUALS, and DATA. A search bar on the right prompts the user to 'find titles, viz types, datasets, authors...'. The left sidebar contains a menu with the following items: 'Dataset Detail', 'Related Dashboards' (with a count of 1), 'Fields' (highlighted with an orange box), 'Data Model', 'Time Modeling', 'Data Extracts', 'Segments' (with a count of 0), and 'Permissions'. The main content area is titled 'Dataset: World Life Expectancy' and 'Detail'. It lists the following information: 'Dataset: World Life Expectancy' with an edit icon, 'Table: main.world_life_expectancy', 'Connection Type: SQLite', 'Data Connection: samples' with an edit icon, 'Description:' with an edit icon, 'Tags:', 'Join Elimination: Enabled' with an edit icon, and 'Result Cache: From Connection' with edit and refresh icons. A horizontal line separates this information from the metadata section below, which includes: 'ID: 9', 'Created on: Jan 26, 2023 11:29 AM', 'Created by: vizapps_admin', 'Last updated: Jan 26, 2023 11:29 AM', and 'Last updated by: vizapps_admin'.

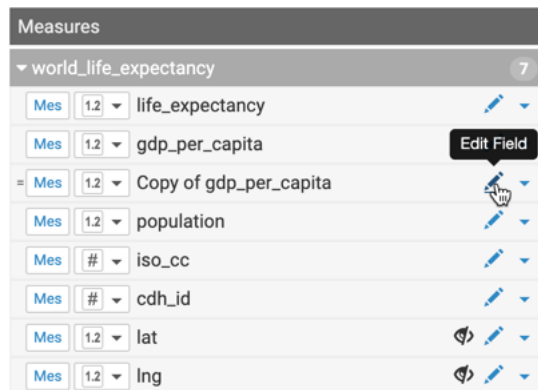
5. In the Fields interface, select EDIT FIELDS.



6. Under Measures, find the field `gdp_per_capita`, and click the (down arrow) icon on its right side.
7. From the menu, select Clone.



8. Under Measures, find the new cloned field Copy of `gdp_per_capita`, click the Edit (pencil) icon on its right side.



The Edit Field Parameters modal window opens, which supports three primary methods of editing fields. They match the three tabs of the modal: Basic, Expression, and Display Format.

9. Make the following changes on the Basic tab:
 - a) Change Display Name to gdp.
 - b) Add Field Comment gdp_per_capita * population.
 - c) Ensure that the Default Aggregation is Sum.

The screenshot shows the 'Edit Field Parameters' dialog box with the 'Basic Settings' tab selected. The 'Base Column' is 'gdp_per_capita'. The 'Display Name' is 'gdp'. The 'Field Comment' is 'gdp_per_capita * population'. The 'Default Aggregation' is 'Sum'. The 'Geo Type' is 'None'. The 'Show field in data detail screen' and 'Show field in Visual Designer' checkboxes are checked. The 'Use as a partition column for Analytical Views' checkbox is unchecked. The 'Category' is 'Measure'.

10. Click the Expression tab and make the following changes:
 - a) Change Expression to [gdp_per_capita] * [population].
 - b) Click VALIDATE EXPRESSION to ensure that the calculation works.
 - c) When the Validation Successful message appears on the modal, click APPLY.

The screenshot shows the 'Edit Field Parameters' dialog box with the 'Expression' tab selected. The 'Expression' field contains '[gdp_per_capita] * [population]'. The 'All Functions' dropdown is open, showing a list of functions including abs, AND, avg, BETWEEN, CASE, cast, char, COALESCE, and count. The 'All Fields' dropdown is also open, showing a list of fields including alt_names, cdh_id, code2, code3, Copy of gdp_per..., country, country_5, fips_code, and gdp_per_capita. The 'Expression contains an aggregation' checkbox is unchecked, and the 'Autocomplete on' checkbox is checked. The 'VALIDATE EXPRESSION' button is highlighted. Below the button, the 'Save expression only after validation succeeds' checkbox is checked. A green message box at the bottom says 'Validation Successful!'. The 'REMOVE', 'CANCEL', and 'APPLY' buttons are at the bottom right.

The new calculated field has an equal sign (=) notation.

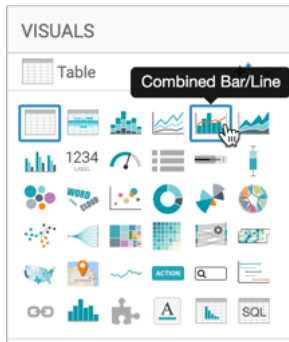
11. Under Dataset: World Life Expectancy, click SAVE.

Testing calculated fields

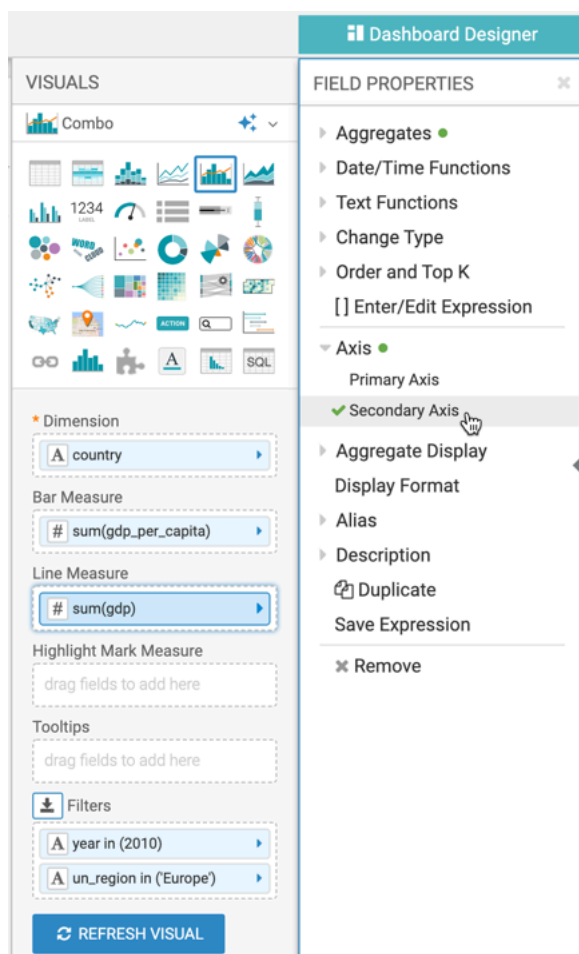
In Cloudera Data Visualization, you can easily test whether a newly calculated field works correctly or not.

Procedure

1. Click New Dashboard in the top right corner of this interface.
2. Select the Combined Bar/Line visual type.



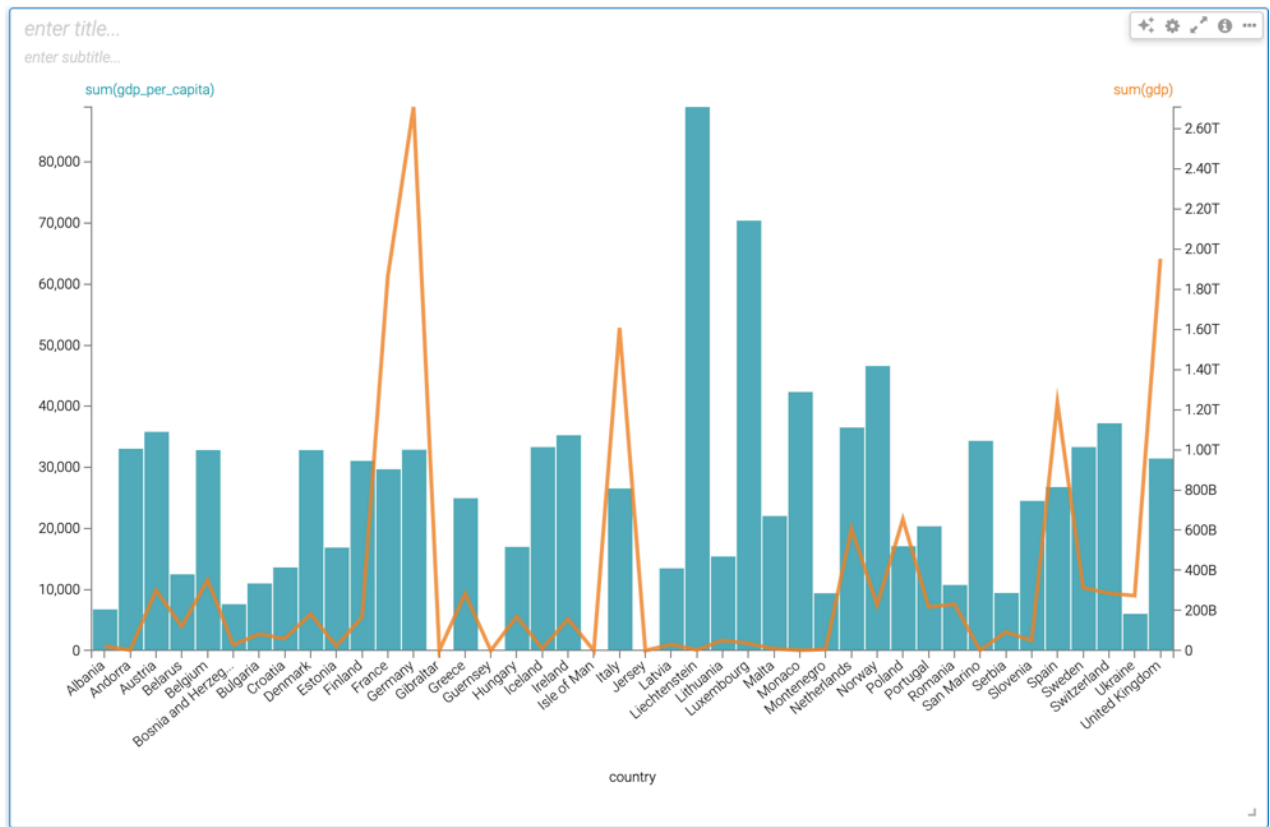
3. Populate the shelves from the available fields in the following way:
 - a) Under Dimensions, select country, and place it on the Dimensions shelf.
 - b) Under Dimensions, select year, and place it on the Filters shelf.
 - c) In the Filter for year modal window, under the Value tab, select the year 2010.
 - d) Under Dimensions, select un_region, and place it on the Filters shelf.
 - e) In the Filter for un_region modal window, under the Value tab, select Europe.
 - f) Under Measures, select gdp_per_capita, and place it on the Bar Measure shelf.
 - g) Under Measures, select gdp, and place it on the Line Measure shelf.
 - h) Click the arrow on gdp field.
 - i) In the Field Properties under Axis, select Secondary Axis.



4. Click REFRESH VISUAL.

Results

The two measurements appear on the graph, superimposed on each other: the original gdp_per_capita represented by the bars, and the calculated gdp, represented by the line.



Changing the field display format

Cloudera Data Visualization lets you configure the display format of each field at the dataset level.

Procedure

1. In the World Life Expectancy dataset, click EDIT FIELDS.

- Under Measures, find the field `gdp_per_capita` and click the (pencil) icon on its right side.

The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'VISUALS', and 'DATA'. The left sidebar contains a 'Dataset Detail' section with links to 'Related Dashboards', 'Fields', 'Data Model', 'Time Modeling', 'Segments', 'Filter Associations', and 'Permissions'. The main area is titled 'Dataset: World Life Expectancy' and 'Fields'. It includes buttons for 'UNDO', 'REFRESH', 'TITLE CASE', 'SAVE', and 'Show Comments'. Below these are two panels: 'Dimensions' and 'Measures'. The 'Measures' panel lists fields like 'life_expectancy', 'gdp_per_capita', 'population', 'iso_cc', 'cdh_id', 'lat', and 'lng'. The 'gdp_per_capita' field is highlighted with an orange square, and its edit icon (pencil) is visible.

The Edit Field Parameters modal window opens.

- Click the Display Format tab.

Under Display Format, you have several options in the Category menu:

- None
- Real Number
- Integer
- Percentage
- Scientific
- Currency
- Date/Time
- Custom Format
- Custom Javascript

For more information, see:

Changing currency field display format

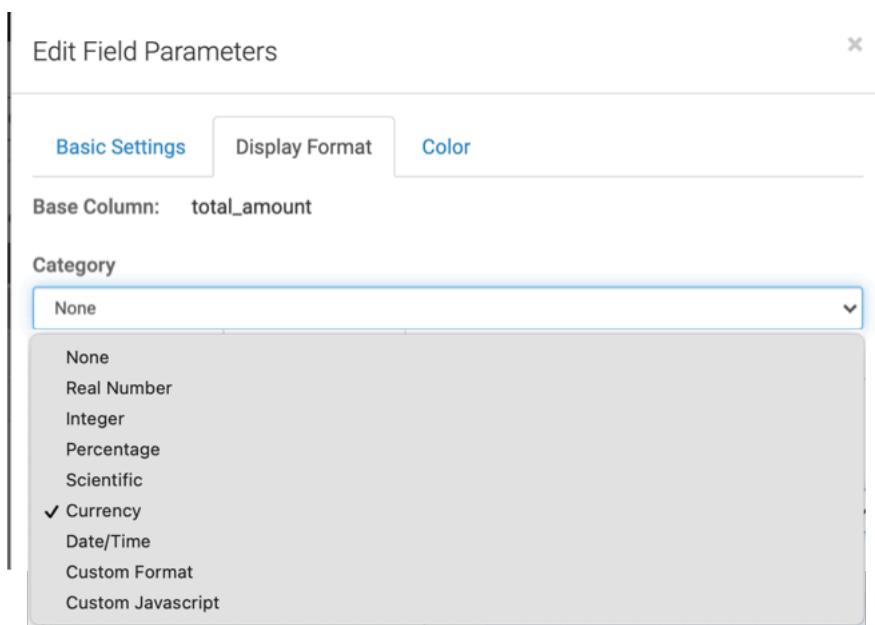
In Cloudera Data Visualization, you can set currency display options for numerical fields across all visuals of a dataset.

About this task

Follow these steps to continue configuring a field at the dataset level for currency format. See *Changing the field display format* for the initial navigation steps.

Procedure

1. In the Edit Field Parameters modal window, under the Display Format tab, select Currency from the Category menu.



2. In the Currency Symbols menu, select the appropriate currency symbol: \$ (Dollar), £ (Pound), ¥ (Yen/Yuan), # (Rupee), € (Euro), or # (Cedi).

Edit Field Parameters

Basic Settings

Display Format

Color

Base Column: gdp_per_capita

Category

Currency

Example:

12345.6789123

→

\$12,345.68

Currency Symbols

✓ \$ (Dollar)

£ (Pound)

¥ (Yen/Yuan)

₹ (Rupee)

€ (Euro)

¢ (Cedi)

No Currency Symbol

For more documentation, go [here](#)

CANCEL

APPLY

3. Select the Basic Format for your records.

You can also define and apply a custom format. Enter a valid format mask in the Customize Currency text box. For a full list of options, see *Display Format Reference*.

Display Format

Category

Currency

Example:

12345.6789123

→

¢12,345.68

Currency Symbols

¢ (Cedi)

Basic Formats

¢12,345.68

Customize Currency

¢,.2f

For more documentation, go [here](#)

None

Real Number

Integer

Percentage

Scientific

✓ Currency

Date/Time

Custom Format

Custom Javascript

CLOSE

SAVE

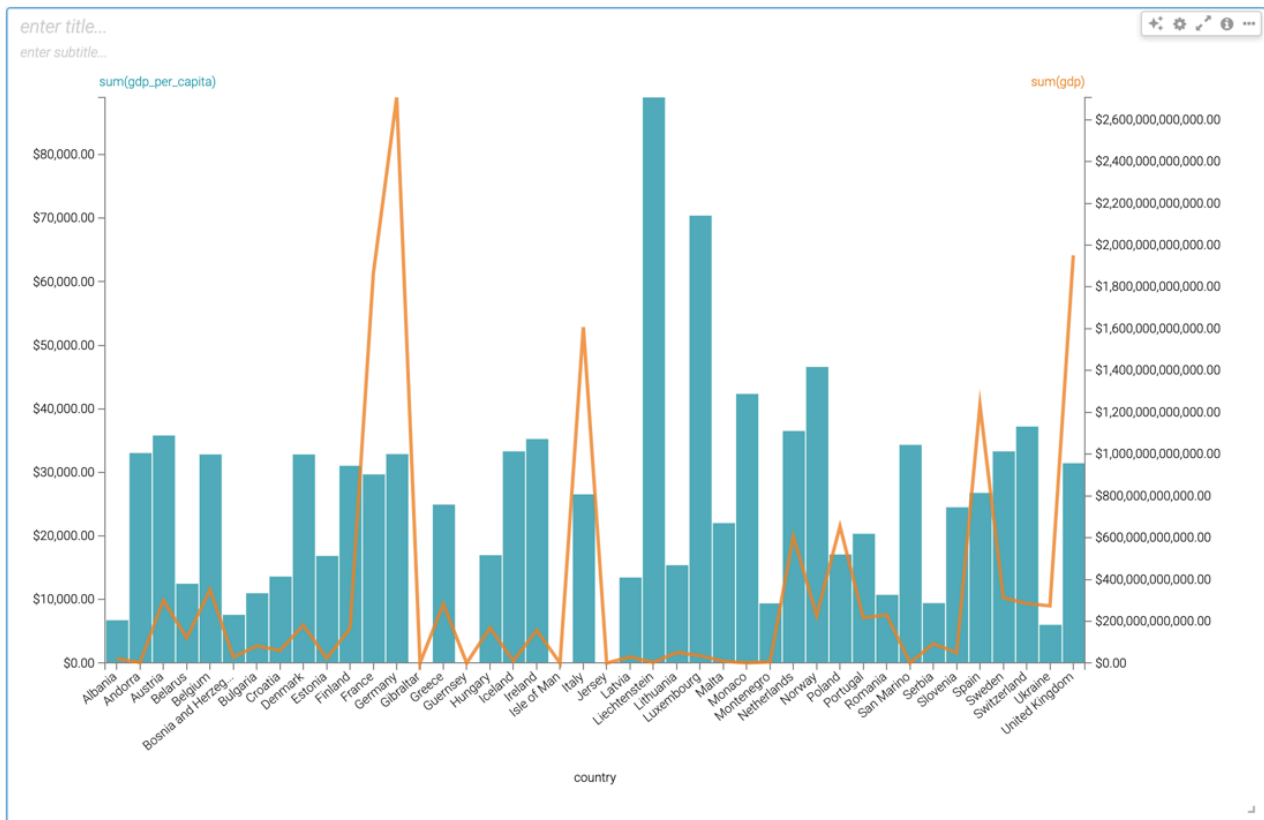
4. Click APPLY.

5. Click SAVE to save the changes to the dataset.

6. To verify that the format applies to all new visuals that use the field, create a new visual by repeating the steps in *Testing the Calculated Field*.

Results

The new visual displays the vertical axis numbers in currency format.



Related Information

[Changing the field display format](#)

[Display format reference](#)

Changing custom field display format

In Cloudera Data Visualization, you can Set currency display options for numerical fields across all visuals of a dataset.

About this task

Follow these steps to continue configuring a field at the dataset level for custom format. See *Changing the field display format* for the initial navigation steps.

Procedure

1. In the Edit Field Parameters window modal, under the Display Format tab, select Custom Format from the Category menu.

2. Under Customize Format, enter \$\$.

- \$ appends the dollar currency symbol to the left of the number.
- S simplifies the number by minimizing the significant numbers (which on the axes appear with a large number of trailing zeros) and appending the appropriate non-scientific (currency) suffix to the right of the number.



Note:

Display Format Examples demonstrate some of the available options for defining custom format.

Edit Field Parameters

Basic Settings Expression Display Format Color

Base Column: gdp_per_capita

Category: Custom Format

Example: 12345.6789123 → \$12.3456789123k

Customize Format: \$\$

Display Format Examples: For more documentation, go [here](#)

Enter	For Display	Enter	For Currencies	Mac Shortcut
f	12346	\$	\$12345	Shift+4
,f	12,346	£	£12345	Alt+3
.,2f	12,345.68	¥	¥12345	Alt+Y
\$.f	\$12,345	₹	₹12345	
.4s	12.35k (for SI notation)	€	€12345	Alt+Shift+2
.4S	12.35k (for currencies)			
.1s	10k			
.2s	12k			
\$.2S	\$12k			
%	1234568%			

REMOVE CANCEL APPLY

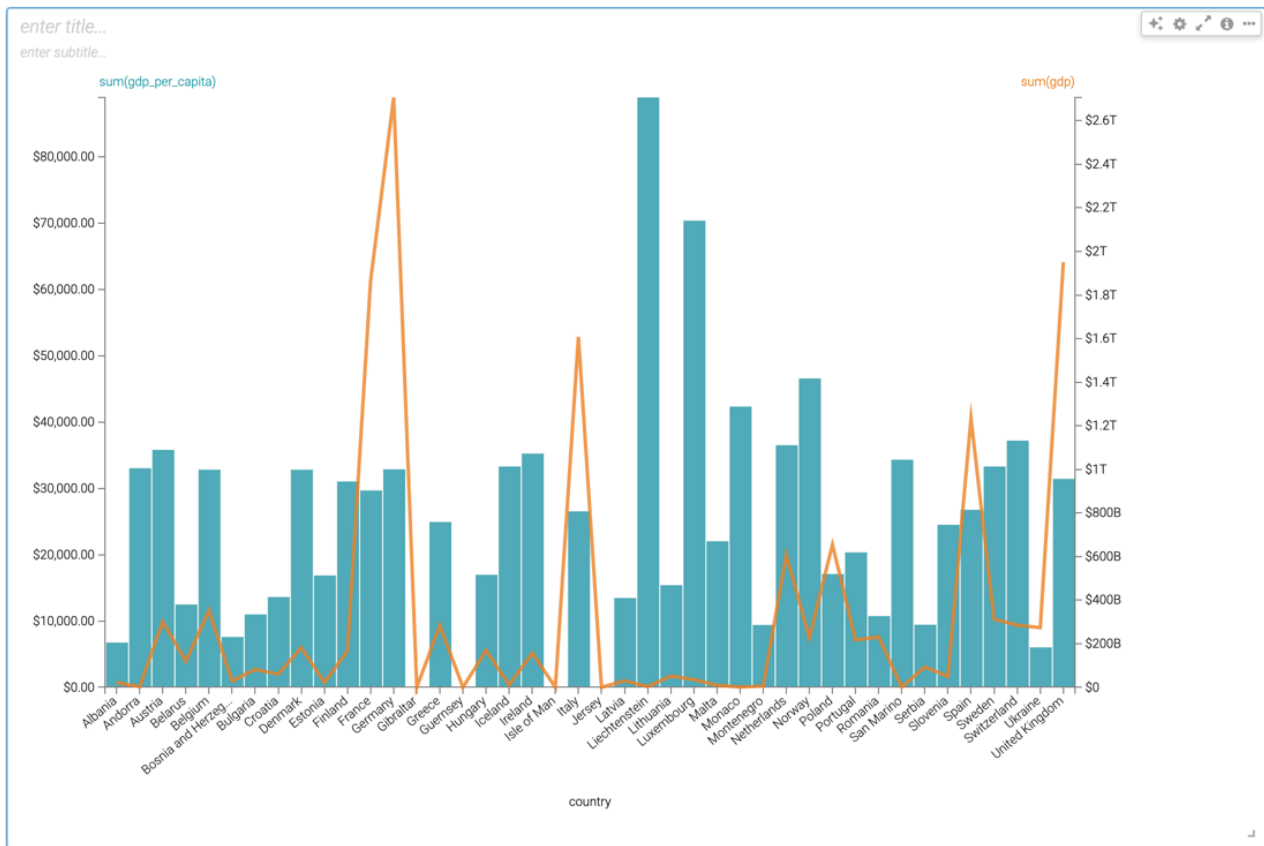
3. Click APPLY.

4. Save the changes to the dataset.

5. [Optional] To verify that the format applies to all new visuals that use the field, create a new visual by repeating the steps in *Testing the calculated field.*

Results

When we use the \$\$ custom format on the gdp field, the visual still uses the dollar currency sign, but also simplifies/abbreviates the number and shows the corresponding currency suffix. In the case of this visual, T for trillion; this custom format transforms the representation \$2,500,000,000,000.00 into \$2.5T.



Related Information

[Changing the field display format](#)

Changing custom Javascript field display format

About this task

Follow these steps to continue configuring a field at the dataset level using Javascript (js). See *Changing the field display format* for initial navigation steps.

Procedure

1. In the Edit Field Parameters window modal, under the Display Format tab, select Custom Javascript from the Category menu.
2. Under Custom JS Format Function, enter the following js code:

```
function myFunc(value) {
  // Show the number in trillions with a dollar sign return
  '$${value/1000000000000}';
}
```

}

Edit Field Parameters ×

Basic Settings
Expression
Display Format
Color

Base Column: gdp_per_capita

Category: Custom Javascript ▼

Example:

12345.6789123

→

12345.6789123

☒ Autocomplete on

Custom JS Format Function

Enter a custom function that takes a single value and returns the value with the desired modifications.

```

1 function myFunc(value) {
2   // Show the number in trillions with a dollar sign return
3   return '$${value/1000000000000}';
4 }

```

UPDATE EXAMPLE

REMOVE
CANCEL
APPLY

3. Click APPLY.

4. Save the changes to the dataset.

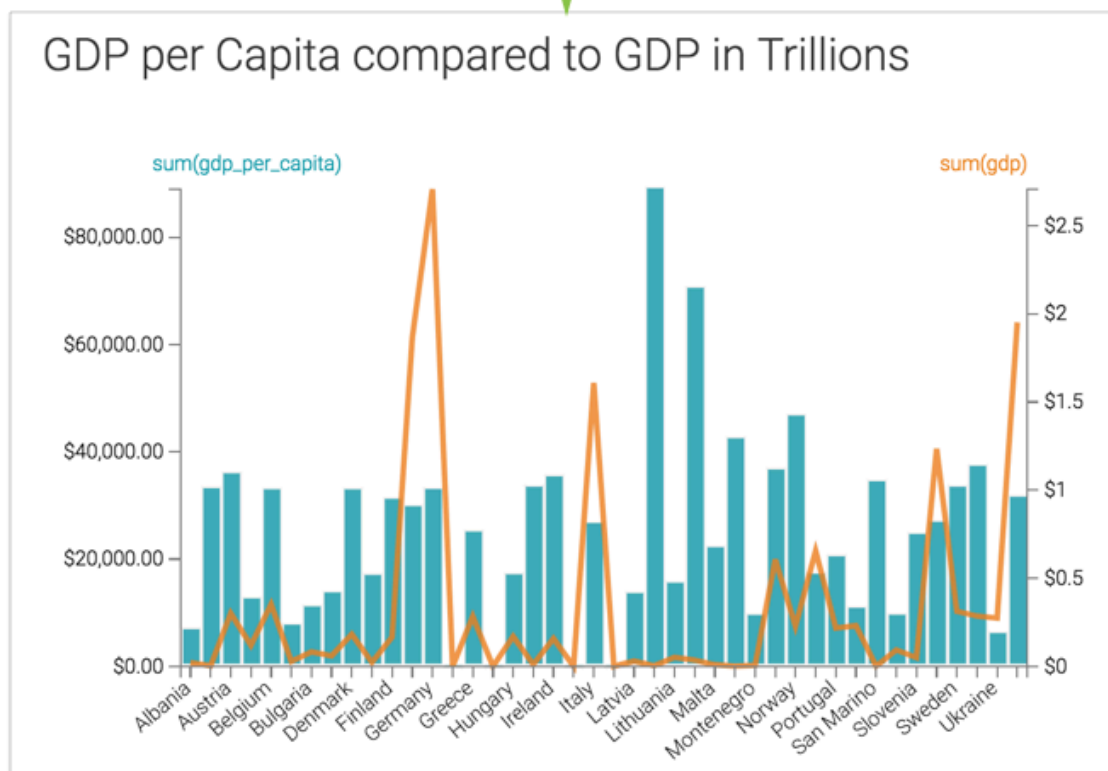
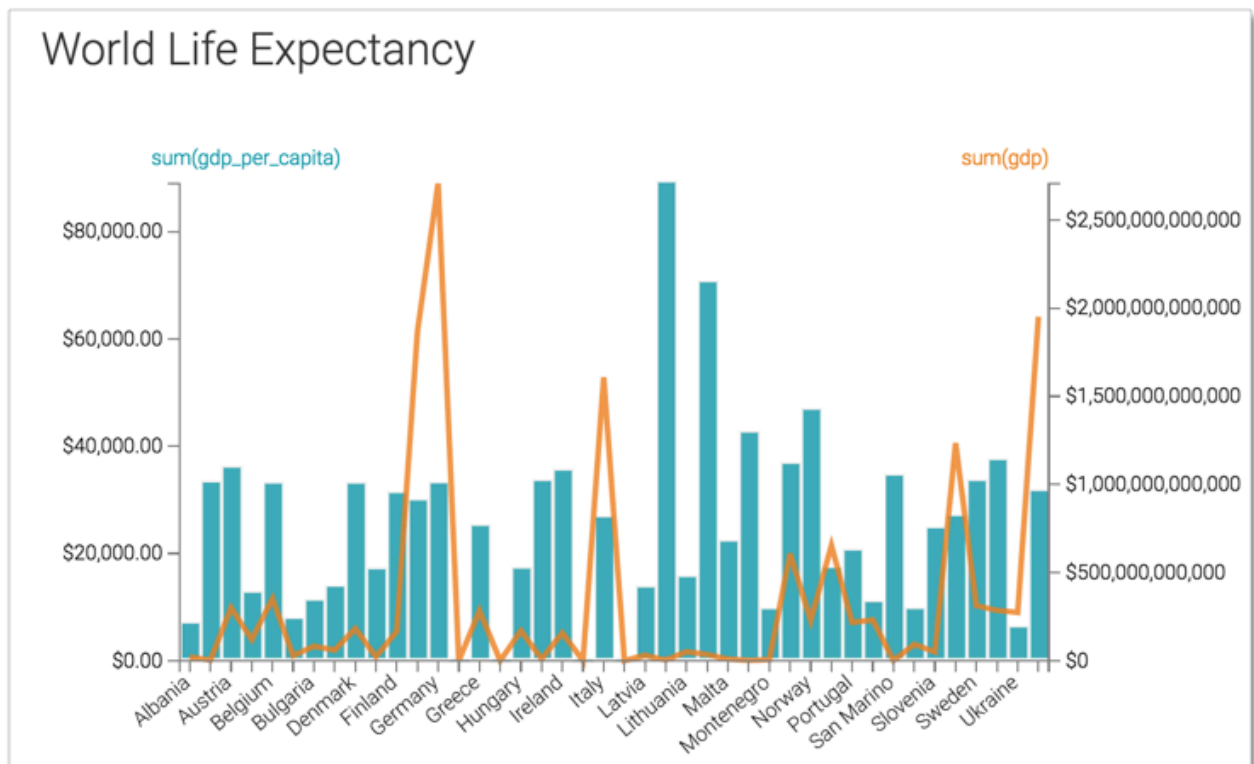
In this example, under Dataset: World Life Expectancy, click Save.

5. To verify that the format applies to all new visuals that use the field, repeat the steps in *Testing the calculated field*.

Results

Let's compare the visuals before and after we apply the js format on the gdp field.

- Before we apply the js format, the visual displays the currency with many trailing zeros.
- After applying the custom js format, notice that the trailing zeros no longer appear.



Related Information

[Changing the field display format](#)

Changing data type

Cloudera Data Visualization allows you to change the effective data type of a column in the dataset model without changing the source data. This is useful in many business environments to ensure correct processing of numerical codes, catalog numbers, event IDs, dates, and so on.

About this task

The following steps demonstrate how to change the type of a column. The column `iso_cc` (the ISO-compliant country code) in the dataset World Life Expectancy [data source `samples.world_life_expectancy`] is used.

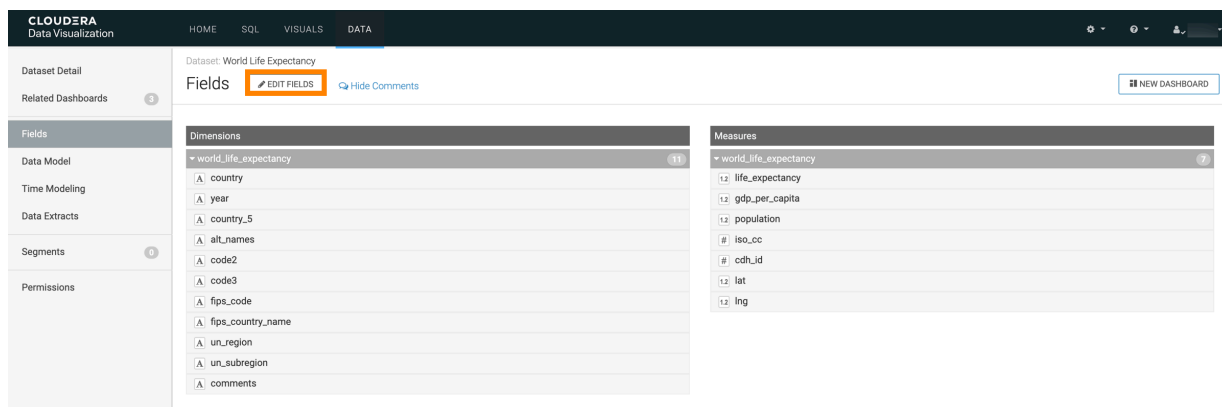
Procedure

1. On the main navigation bar, click DATA.
2. In the left navigation menu, click samples.
3. In the Datasets area, select World Life Expectancy (`samples.world_life_expectancy`).
4. In the Dataset Detail menu, select Fields.

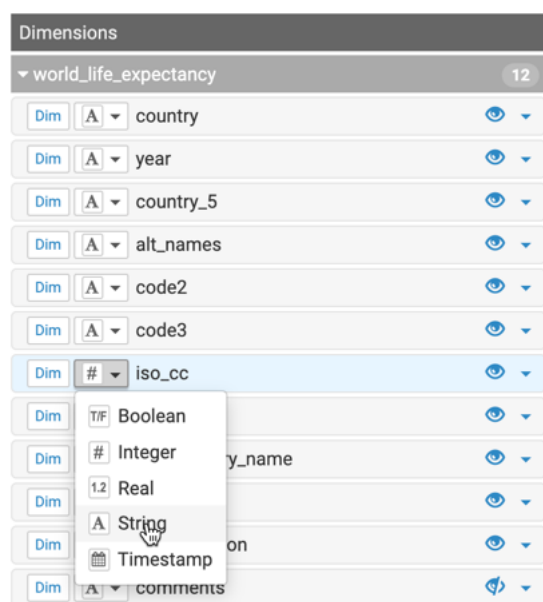
The screenshot displays the Cloudera Data Visualization interface. The top navigation bar includes 'HOME', 'SQL', 'VISUALS', and 'DATA'. The left sidebar shows a 'Dataset Detail' section with options: 'Related Dashboards' (1), 'Fields' (highlighted with an orange box), 'Data Model', 'Time Modeling', and 'Data Extracts'. Below this are 'Segments' (0) and 'Permissions'. The main content area is titled 'Dataset: World Life Expectancy' and 'Detail'. It lists the following information:

- Dataset: World Life Expectancy
- Table: `main.world_life_expectancy`
- Connection Type: SQLite
- Data Connection: `samples`
- Description:
- Tags:
- Join Elimination: Enabled
- Result Cache: From Connection
- ID: 9
- Created on: Jan 26, 2023 11:29 AM
- Created by: `vizapps_admin`
- Last updated: Jan 26, 2023 11:29 AM
- Last updated by: `vizapps_admin`

5. In the Fields interface, select EDIT FIELDS.



6. Under Dimensions, find the field fips_code, and click the down arrow icon, immediately following the # icon.
7. In the menu, select String.



8. Under Dataset: World Life Expectancy, click Save.

Results

All new visuals created from this dataset will automatically use the new type.



Note: Use this functionality carefully, as it may break the visuals that already use the column in aggregations or custom expressions.

Specifying geographic fields

In Cloudera Data Visualization, you can explicitly specify a dataset field as one of the many supported geographical types.

About this task

In this example the dataset Canadian Census is used, constructed from example datasets, and joined of the fields fsa and Postal Code, respectively.

The following steps demonstrate how to assign Geo Types to a dataset field. We are using the two fields from the join of the Canadian Census dataset: fsa from canada_census_population_dwellings, and Postal Code from ca_postal_codes.



Note: Before creating visuals that use geographic data, specify the appropriate information as Geo Types.

Procedure

1. Navigate to the dataset that you must modify. In this example, the Food Stores Inspection in NYC dataset is used.
2. In the Dataset Detail menu, select Fields.
3. In the Fields interface, select EDIT FIELDS.

CLOUDERA
Data Visualization

HOME SQL VISUALS **DATA**

find titles, viz types, datasets, authors...

Dataset Detail

Related Dashboards 1

Fields

Data Model

Time Modeling

Data Extracts

Segments 0

Permissions

Dataset: World Life Expectancy

Detail

Dataset: World Life Expectancy

Table: main.world_life_expectancy

Connection Type: SQLite

Data Connection: samples

Description:

Tags:

Join Elimination: Enabled

Result Cache: From Connection

ID: 9

Created on: Jan 26, 2023 11:29 AM

Created by: vizapps_admin

Last updated: Jan 26, 2023 11:29 AM

Last updated by: vizapps_admin

4. Click the Edit (pencil) icon on the right side of the field.

Measures

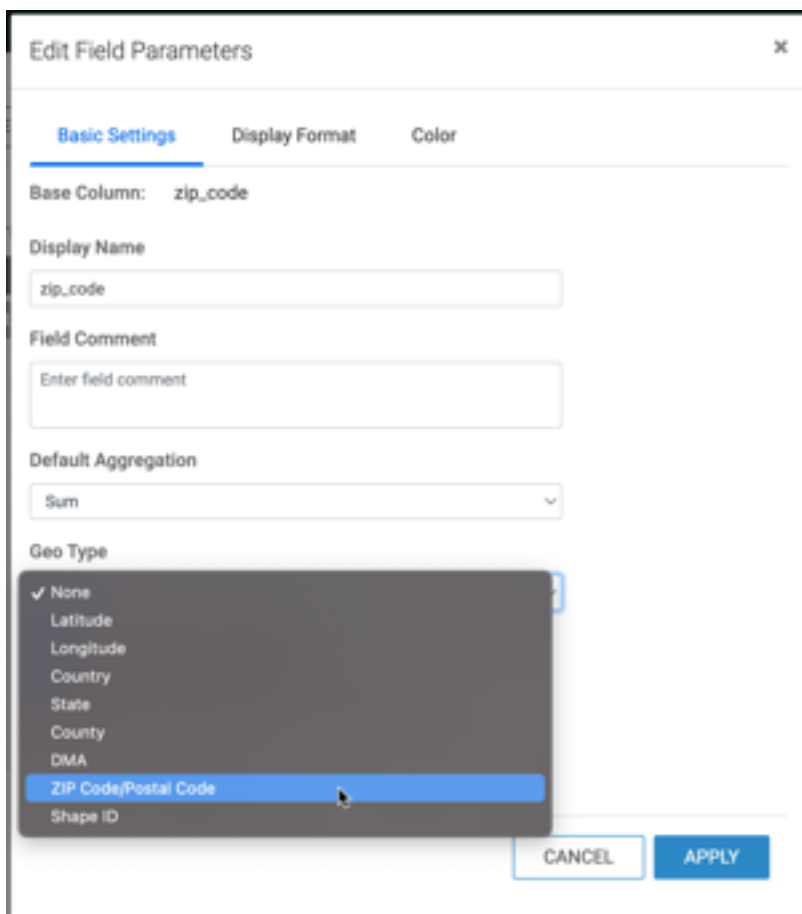
retail_food_store_inspections_current_critical_violations

Measures: # zip_code

Edit Field

5. In the Edit Field Parameters window modal, under Geo Type, select the appropriate option from the menu.

- Click APPLY. For zip_code, we also changed the Display Name to Postal Code in the example.



- Under Dataset: Food Stores Inspection in NYC.
- Click SAVE.

Results

The dataset can now be successfully used for map and interactive map visuals, without further adjustments at the level of the visual.

Adding field comments

When working with big data, it can be very helpful to have access to comprehensive field-level descriptions. In Cloudera Data Visualization, you can use field comments to provide the context and meaning of each dataset field.

Adding field comments in dataset

When working with large datasets data, it can be useful to have access to comprehensive field-level descriptions. In Cloudera Data Visualization, you can use field comments to provide the context and meaning of each dataset field.

About this task

The following steps demonstrate how to add description to a column of a dataset, as a 'comment'. In this example, the column iso_cc (the ISO-compliant country code) is used from the dataset World Life Expectancy [data source samples.world_life_expectancy].

Procedure

1. On the main navigation bar, click DATA.

The Data view opens, displaying the Datasets tab.

2. In the left navigation menu, click samples.
3. In the Datasets area, select World Life Expectancy (samples.world_life_expectancy).
4. In the Dataset Detail menu, select Fields.

CLOUDERA
Data Visualization

HOME SQL VISUALS **DATA**

find titles, viz types, datasets, authors...

Dataset Detail

Related Dashboards 1

Fields

Data Model

Time Modeling

Data Extracts

Segments 0

Permissions

Dataset: World Life Expectancy

Detail

Dataset: **World Life Expectancy**

Table: **main.world_life_expectancy**

Connection Type: **SQLite**

Data Connection: **samples**

Description:

Tags:

Join Elimination: **Enabled**

Result Cache: **From Connection**

ID: **9**

Created on: **Jan 26, 2023 11:29 AM**

Created by: **vizapps_admin**

Last updated: **Jan 26, 2023 11:29 AM**

Last updated by: **vizapps_admin**

5. In the Fields interface, select EDIT FIELDS.

CLOUDERA
Data Visualization

HOME SQL VISUALS **DATA**

Dataset: World Life Expectancy

Fields **EDIT FIELDS** Hide Comments

NEW DASHBOARD

Fields

Data Model

Time Modeling

Data Extracts

Segments 0

Permissions

Dimensions

world_life_expectancy 10

A	country
A	year
A	country_5
A	alt_names
A	code2
A	code3
A	fips_code
A	fips_country_name
A	un_region
A	un_subregion
A	comments

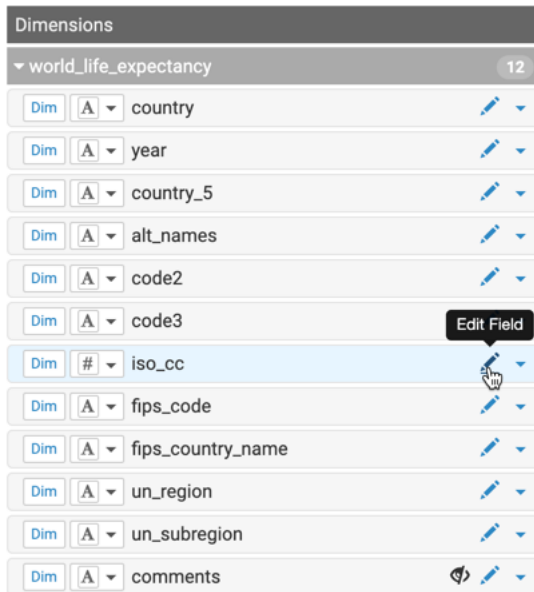
Measures

world_life_expectancy 7

1,2	life_expectancy
1,2	gdp_per_capita
1,2	population
#	iso_cc
#	cdh_id
1,2	lat
1,2	lng

6. Under Dimensions, find the field iso_cc, and click the down arrow icon on the right.

- Click the Edit (pencil) icon.



The Edit Field Parameters modal window opens.

- Under Field Comment, add the following text:

ISO 3166 Country Code Standard, Numeric, described in http://www.iso.org/iso/country_codes. Mappings are at <https://www.iso.org/obp/ui/#search>.

- Click APPLY.

The screenshot shows the 'Edit Field Parameters' modal window. The 'Field Comment' section is highlighted, showing the text: 'ISO 3166 Country Code Standard, Numeric, described in http://www.iso.org/iso/country_codes. Mappings are at <https://www.iso.org/obp/ui/#search>'. The 'Default Aggregation' is set to 'Sum', 'Geo Type' is 'None', and 'Category' is 'Dimension'. The 'APPLY' button is highlighted.

10. Click SAVE in the top row.

11. Click Show Comments to display the field comment.

Dataset: World Life Expectancy

Fields

EDIT FIELDS

Show Comments

Dimensions	Measures
<div>▼ world_life_expectancy 12</div> <div> <div>A country</div> <div>A year</div> <div>A country_5</div> <div>A alt_names</div> <div>A code2</div> <div>A code3</div> <div>A iso_cc</div> <div>A fips_code</div> <div>A fips_country_name</div> <div>A un_region</div> <div>A un_subregion</div> <div>A comments</div> </div>	<div>▼ world_life_expectancy 7</div> <div> <div>1.2 life_expectancy</div> <div>1.2 gdp_per_capita</div> <div>= 1.2 gdp</div> <div>1.2 population</div> <div># cdh_id</div> <div>1.2 lat</div> <div>1.2 lng</div> </div>

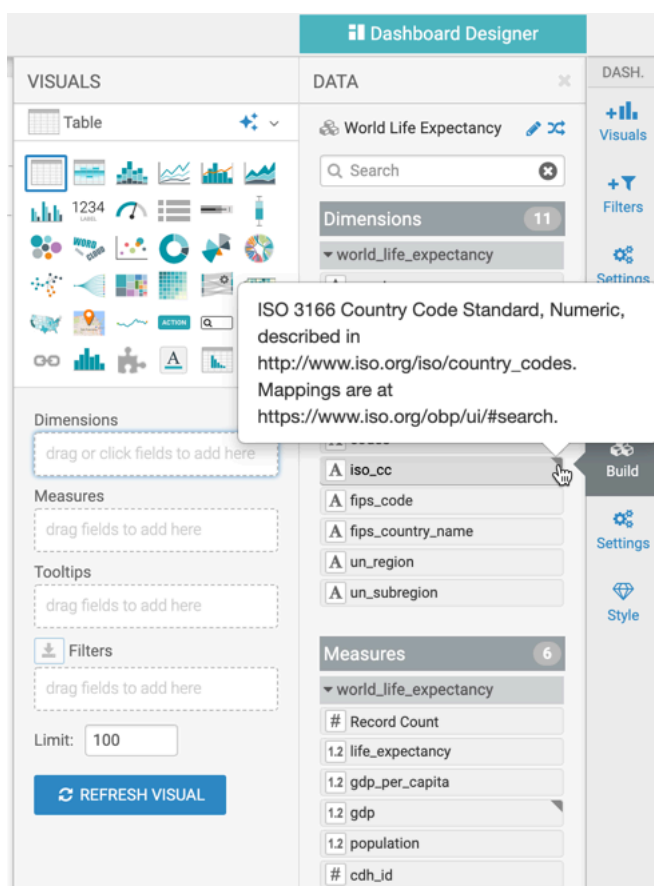
12. To hide field comments, click Hide Comments.

Viewing field comments in Visual Builder

When working with big data, it can be very helpful to have access to comprehensive field-level descriptions. In Cloudera Data Visualization, you can use field comments to provide the context and meaning of each dataset field.

To view field comments in the Visual Builder, create a new visual using the World Life Expectancy dataset [data source samples.world_life_expectancy]. For more information, see *Creating visuals*.

The fields iso-cc and gdp-per-capita display a gray triangle on the top right corner of the field. Hover over the triangle to view the field comments defined in the dataset.



Automatically renaming dataset fields

Quite often, the column names of the base data tables are not very human-friendly. Cloudera Data Visualization gives you the option to automatically adjust field names at the level of the dataset.

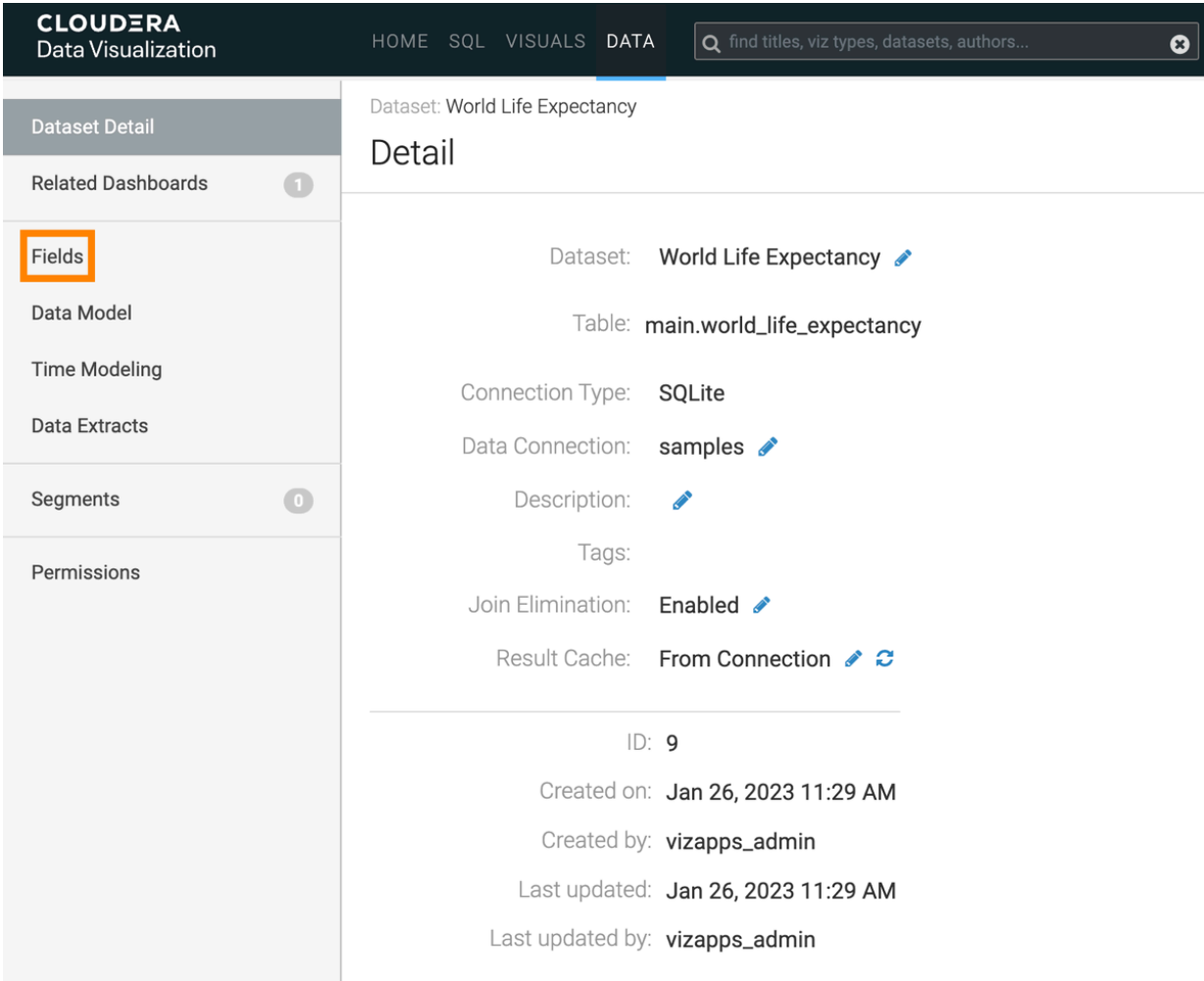
About this task

The following steps demonstrate how to prevent data fields from appearing in visualizations and applications of dataset World Life Expectancy [data source samples.world_life_expectancy]. The fields comments, lat, and lng are empty, so they are good candidates for this operation.

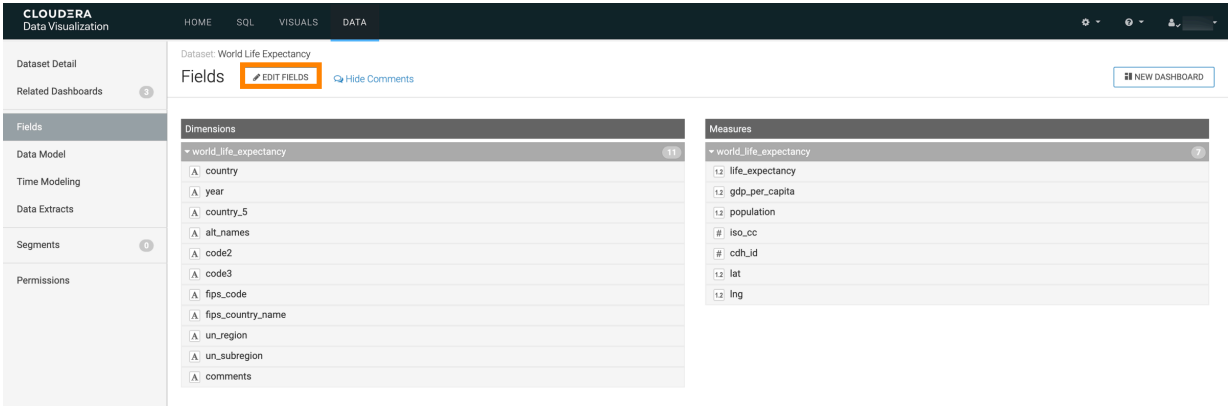
Procedure

1. On the main navigation bar, click DATA.
The Data view opens, displaying the Datasets tab.
2. In the left navigation menu, click samples.
3. In the Datasets area, select World Life Expectancy (samples.world_life_expectancy).

4. In the Dataset Detail menu, select Fields.



5. In the Fields interface, select EDIT FIELDS.



- Near the top of the page, click TITLE CASE.

Dataset: World Life Expectancy

Fields ↶ UNDO ↻ REFRESH T TITLE CASE SAVE Show Comments

To add a new calculated field, use the down arrow to the right of a field to clone it, and then edit the expression of the cloned field.

Dimensions		Measures	
▼ world_life_expectancy	12	▼ world_life_expectancy	7
Dim A country		Mes 1.2 life_expectancy	
Dim A year		Mes 1.2 gdp_per_capita	
Dim A country_5		= Mes 1.2 Copy of gdp_per_capita	
Dim A alt_names		Mes 1.2 population	
Dim A code2		Mes # cdh_id	
Dim A code3		Mes 1.2 lat	
Dim # iso_cc		Mes 1.2 lng	
Dim A fips_code			
Dim A fips_country_name			
Dim A un_region			
Dim A un_subregion			
Dim A comments			

The field titles change. For example, the Measure gdp_per_capita appears as GDP Per Capita.



Note: The Title Case option does not affect fields that are calculated over table fields, such as gdp.

- Click SAVE.

Custom renaming dataset fields

Often, we find it useful to rename a frequently-used field directly in the dataset, instead of using an alias in each visual. Cloudera Visualization makes it very easy to change the display name of a dataset field.

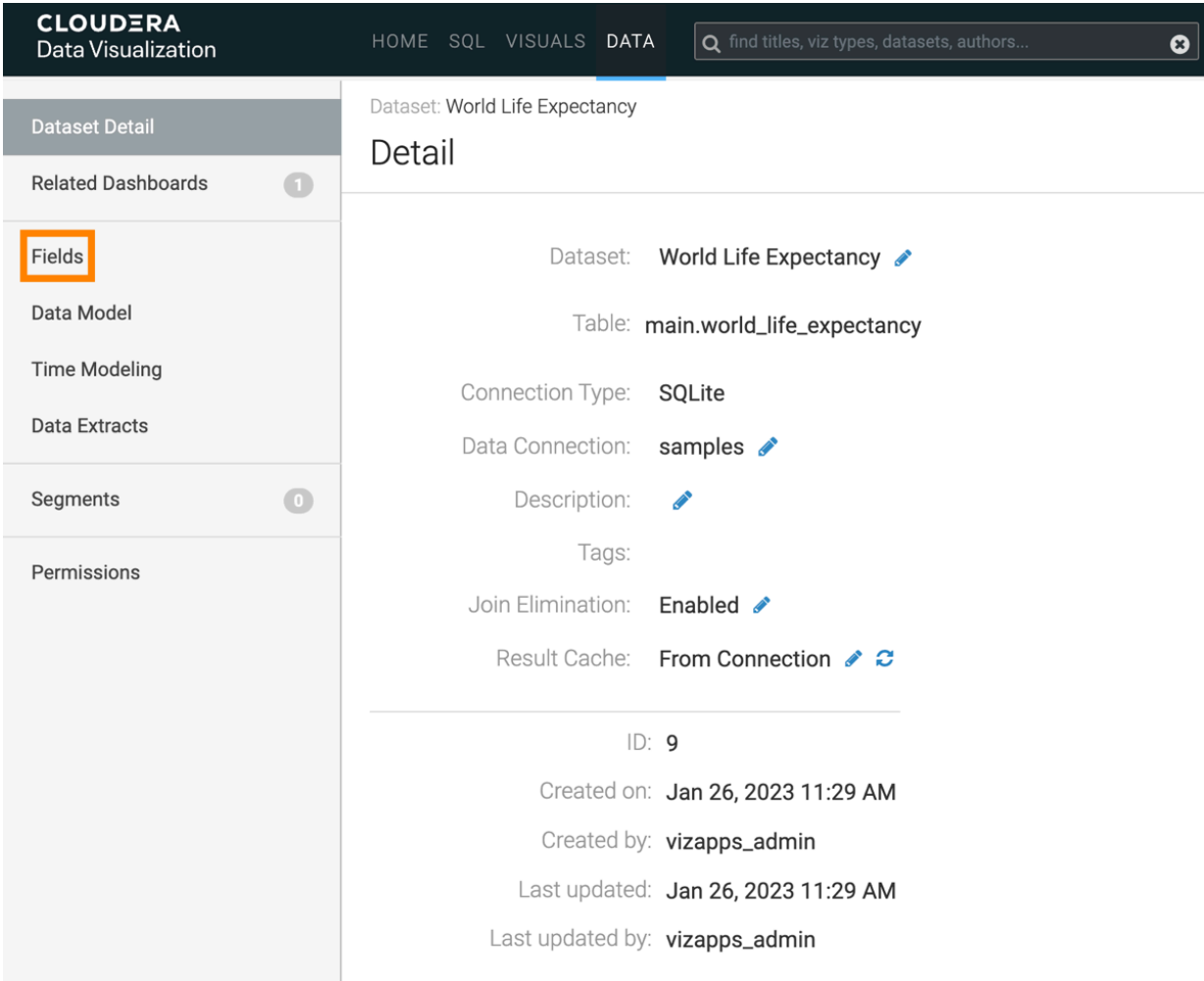
About this task

The following steps demonstrate how to rename a field. In this example, the column iso_cc (the ISO-compliant country code) is used from the dataset World Life Expectancy [data source: samples.world_life_expectancy].

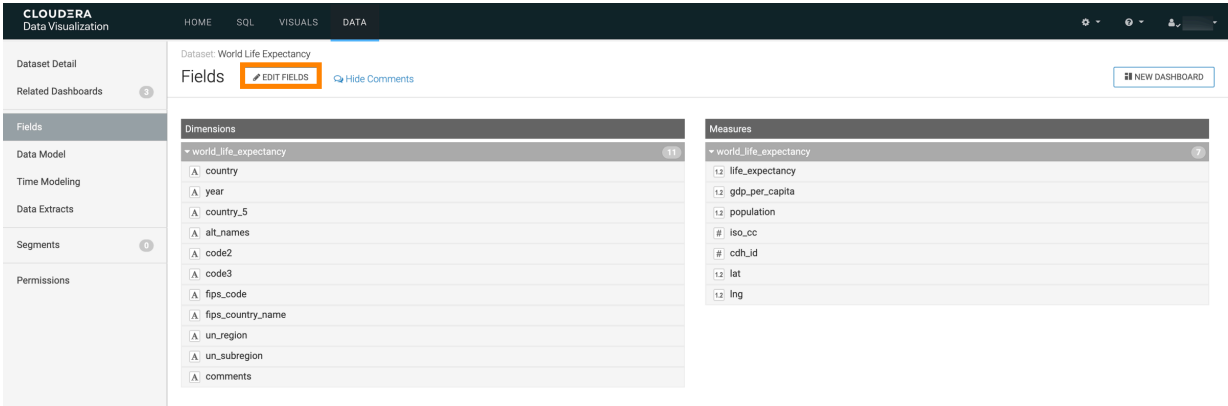
Procedure

- On the main navigation bar, click DATA.
The Data view opens, displaying the Datasets tab.
- In the left navigation menu, click samples.
- In the Datasets area, select World Life Expectancy (samples.world_life_expectancy).

4. In the Dataset Detail menu, select Fields.

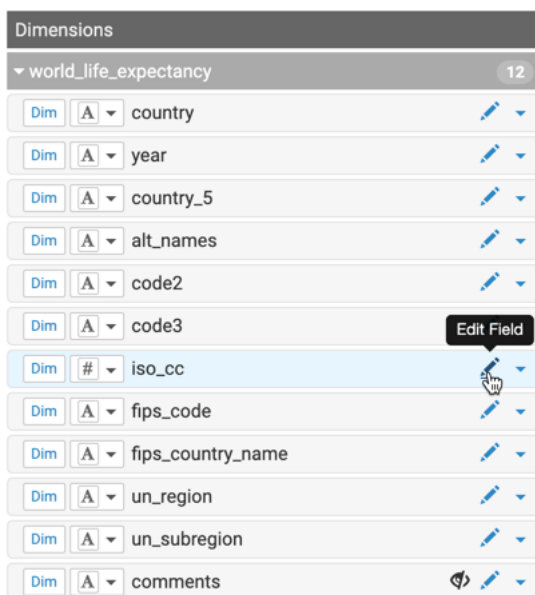


5. In the Fields interface, select EDIT FIELDS.



6. Under Dimensions, find the field iso_cc, and click the (down arrow) icon on the right.

- Click the Edit (pencil) icon.



The Edit Field Parameters modal window opens.



Note: The Base Column name cannot be edited, but you can change the Display Name of the column.

- Change the Display Name from iso_cc to ISO Country Code.
- Click APPLY.
- Under Dataset: World Life Expectancy, click SAVE.

Results

All new visuals created from this dataset use the new name automatically.

Refreshing a dataset

You can refresh an existing dataset if the underlying tables change or if you update the associated extract with new or changed fields.

Procedure

- Click DATA on the main navigation bar.
The Data view opens, displaying the Datasets tab.
- Click the dataset you want to refresh.
The Detail view opens.

3. Select Fields in the Dataset Detail menu.

CLUSTERA
Data Visualization

HOME SQL VISUALS DATA

find titles, viz types, datasets, authors...

Dataset Detail

Related Dashboards

Fields

Data Model

Time Modeling

Data Extracts 3

Segments 0

Filter Associations 0

Permissions

Dataset: doc-test-dataset2

Fields

EDIT FIELDS

Hide Comments

Dimensions

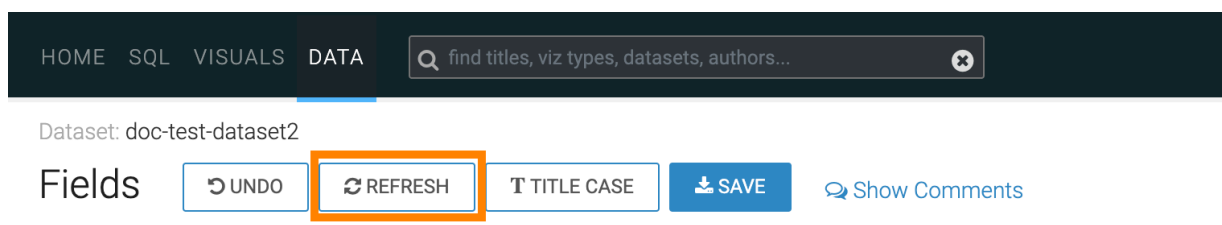
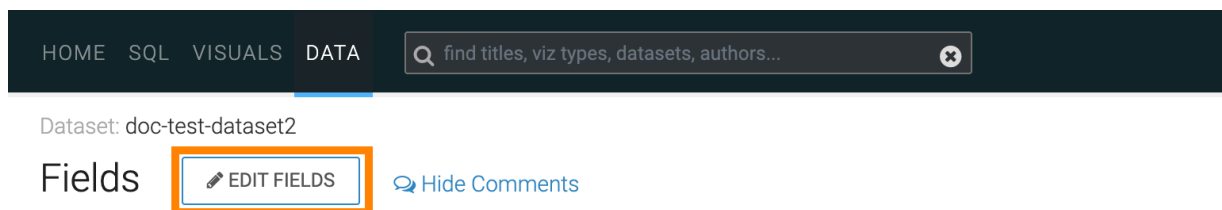
cdv_doc_test_data 3

country

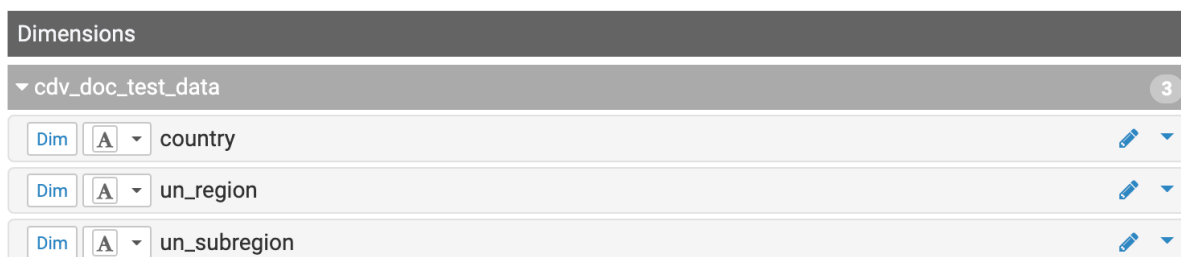
un_region

un_subregion

4. In the Fields interface, click **EDIT FIELDS REFRESH**



To add a new calculated field, use the down arrow to the right of a field to clone it, and then edit the expression of the cloned



Cloudera Data Visualization scans the table for changed fields. If no changes are detected, no refresh is done and a notification is displayed. If changes are detected, the table columns are updated and a notification is displayed about the change.

5. Close the notification.

Editing datasets based on a SQL query

One of the major advantages Cloudera Data Visualization provides is the option to edit the data selection that defines the dataset.

Procedure

1. Navigate to a dataset that you created based on a SQL query.

2. Click to open the dataset, on the Dataset Detail view.

There is a SQL text window, which you can edit.

The screenshot shows the 'Dataset Detail' view for a dataset named 'Test Dataset from SQL'. The left sidebar contains navigation links: Dataset Detail, Related Dashboards, Fields, Data Model, Time Modeling, Search Modeling, Segments (0), Filter Associations (0), and Permissions. The main content area displays the following details:

- Dataset: Test Dataset from SQL (with an edit icon)
- Connection Type: SQLite
- Data Connection: samples (with an edit icon)
- Description: (with an edit icon)
- Join Elimination: Enabled (with an edit icon)
- Result Cache: From Connection (with edit and refresh icons)

A text window labeled 'SQL for first table:' is highlighted with an orange border. It contains the SQL query: `select * from main.us_counties` (with an edit icon). Below this, the dataset's metadata is shown:

- ID: 14
- Created on: Aug 19, 2021 09:55 AM
- Created by: vizapps_admin
- Last updated: Aug 19, 2021 09:55 AM
- Last updated by: vizapps_admin

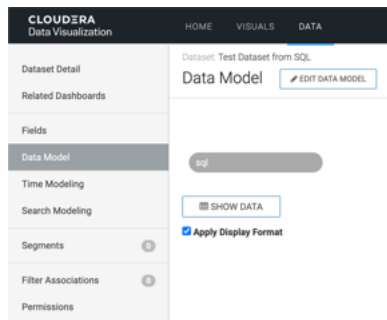
3. You can restrict rows and columns depending on what you need.

Restricting columns in datasets based on SQL query

In Cloudera Data Visualization, you can easily restrict the table columns in the dataset by changing the SQL definition of that dataset. SQL-defined datasets make it easy to limit their content to specific columns.

Procedure

1. Switch to Data Model interface, and click Show Data.



There is a large number of columns in the query result, and many of them are not necessary when it comes to answering most common questions.

sumlev	state	county	stname	ctname	year	agegrp	tot_pop	tot_male	tot_female	wa_male	wa_female	ba_male	ba_female	la_male	la_female	aa_male	aa_female	na_male	na_female	tom_male	tom_female	wac_r
50	51	149	Virginia	Prince George County	5	0	36941	20368	16573	12155	10721	7230	4763	171	98	254	410	61	55	497	526	1257
50	51	153	Virginia	Prince William County	5	0	430289	213820	216469	141918	138857	44291	47256	2453	2331	16249	18465	405	374	8504	9186	1495
50	51	155	Virginia	Pulaski County	5	0	34736	17284	17452	15915	16222	959	866	35	37	93	105	6	2	276	220	1617
50	51	157	Virginia	Rappahannock County	5	0	7456	3694	3762	3420	3496	181	171	5	13	19	28	1	2	68	52	3483
50	51	159	Virginia	Richmond County	5	0	9059	5066	3993	3138	2925	1799	961	24	15	31	17	1	2	73	73	3204
50	51	161	Virginia	Roanoke County	5	0	92901	44385	48516	39834	43608	2424	2605	85	82	1338	1465	12	21	692	735	4048
50	51	163	Virginia	Rockbridge County	5	0	22394	11071	11323	10482	10693	321	323	60	63	50	76	1	3	157	165	10631
50	51	165	Virginia	Rockingham County	5	0	77391	37840	39551	36012	37908	841	629	227	213	232	278	15	11	513	512	3649
50	51	167	Virginia	Russell County	5	0	28445	13914	14531	13631	14249	148	120	28	35	32	24	0	1	75	102	13701
50	51	169	Virginia	Scott County	5	0	22781	11413	11368	11197	11172	92	68	25	23	17	29	8	7	74	69	1126

2. Find the fields that you would like to keep in the dataset definition.
3. Switch back to Dataset Detail interface, and edit SQL text window by applying the following statement:

```
select county, stname, ctname, tot_pop, tot_male, tot_female from main.us_counties
```

In this example we keep the columns county, stname, ctname, tot_pop, tot_male, and tot_female.

4. Click Save.

The screenshot shows the 'Dataset Detail' page for 'Test Dataset from SQL'. The left sidebar contains navigation links: Dataset Detail, Related Dashboards, Fields, Data Model, Time Modeling, Search Modeling, Segments (0), Filter Associations (0), and Permissions. The main content area displays the following details:

- Dataset: Test Dataset from SQL
- Connection Type: SQLite
- Data Connection: samples
- Description:
- Join Elimination: Enabled
- Result Cache: From Connection

The 'SQL for first table:' section is highlighted with an orange border. It contains the following SQL query:

```
select county, stname, ctynome, tot_pop, tot_male, tot_female from  
main.us_counties
```

To the right of the SQL query are 'CANCEL' and 'SAVE' buttons. Below the SQL section, the following metadata is displayed:

- ID: 14
- Created on: Aug 19, 2021 09:55 AM
- Created by: vizapps_admin
- Last updated: Aug 19, 2021 09:55 AM
- Last updated by: vizapps_admin

5. In the Refresh dataset table column information modal window, click Close.

The screenshot shows a modal window titled 'Refresh dataset table column information'. The main text inside the modal reads 'Table columns updated'. At the bottom right of the modal is a 'CLOSE' button.

- Switch back to the Data Model interface, click Show Data, and check that the dataset only has the explicitly specified columns:

county	stname	ctyname	tot_pop	tot_male	tot_female
149	Virginia	Prince George County	36941	20368	16573
153	Virginia	Prince William County	430289	213820	216469
155	Virginia	Pulaski County	34736	17284	17452
157	Virginia	Rappahannock County	7456	3694	3762
159	Virginia	Richmond County	9059	5066	3993
161	Virginia	Roanoke County	92901	44385	48516
163	Virginia	Rockbridge County	22394	11071	11323
165	Virginia	Rockingham County	77391	37840	39551
167	Virginia	Russell County	28445	13914	14531
169	Virginia	Scott County	22781	11413	11368

In this example we have kept the columns `county`, `stname`, `ctyname`, `tot_pop`, `tot_male`, and `tot_female`.

Restricting rows in datasets based on SQL query

In Cloudera Data Visualization, you can easily restrict the table rows in the dataset by changing the SQL definition of that dataset. SQL-defined datasets make it easy to limit their content to specific rows.

Procedure

- Switch to Dataset Detail interface, and edit SQL text window by applying the following statement:

```
select county, stname, ctyname, tot_pop, tot_male, tot_female from main.us_counties
where stname in ('Arizona','New Mexico',
'California','Nevada','Colorado','Utah')
```

2. Click Save.

The screenshot shows the 'Dataset Detail' page for 'Test Dataset from SQL'. The left sidebar contains navigation links: Dataset Detail, Related Dashboards, Fields, Data Model, Time Modeling, Search Modeling, Segments (0), Filter Associations (0), and Permissions. The main content area displays the following details:

- Dataset: Test Dataset from SQL
- Connection Type: SQLite
- Data Connection: samples
- Description:
- Join Elimination: Enabled
- Result Cache: From Connection

The 'SQL for first table:' section is highlighted with an orange border and contains the following SQL query:

```
select county, stname, ctynome, tot_pop, tot_male, tot_female from  
main.us_counties  
where stname in ('Arizona', 'New Mexico', 'California', 'Nevada',  
'Colorado', 'Utah')
```

To the right of the SQL query are 'CANCEL' and 'SAVE' buttons. Below the SQL section, the following metadata is displayed:

- ID: 14
- Created on: Aug 19, 2021 09:55 AM
- Created by: vizapps_admin
- Last updated: Aug 25, 2021 01:35 PM
- Last updated by: vizapps_admin

3. In the Refresh dataset table column information modal window, click Close.

The screenshot shows a modal window titled 'Refresh dataset table column information'. The main text inside the modal reads 'Table columns updated'. At the bottom right of the modal is a 'CLOSE' button.

4. Switch back to the Data Model interface, click Show Data, and notice that the dataset is limited to the states specified in the SQL statement.

CLUSTERDATA

Data Visualization

HOME

VISUALS

DATA

SEARCH

vizapps_admin

Dataset Detail

Related Dashboards

Fields

Data Model

Time Modeling

Search Modeling

Segments

Filter Associations

Permissions

Dataset: Test Dataset from SQL

Data Model

EDIT DATA MODEL

NEW DASHBOARD

sql

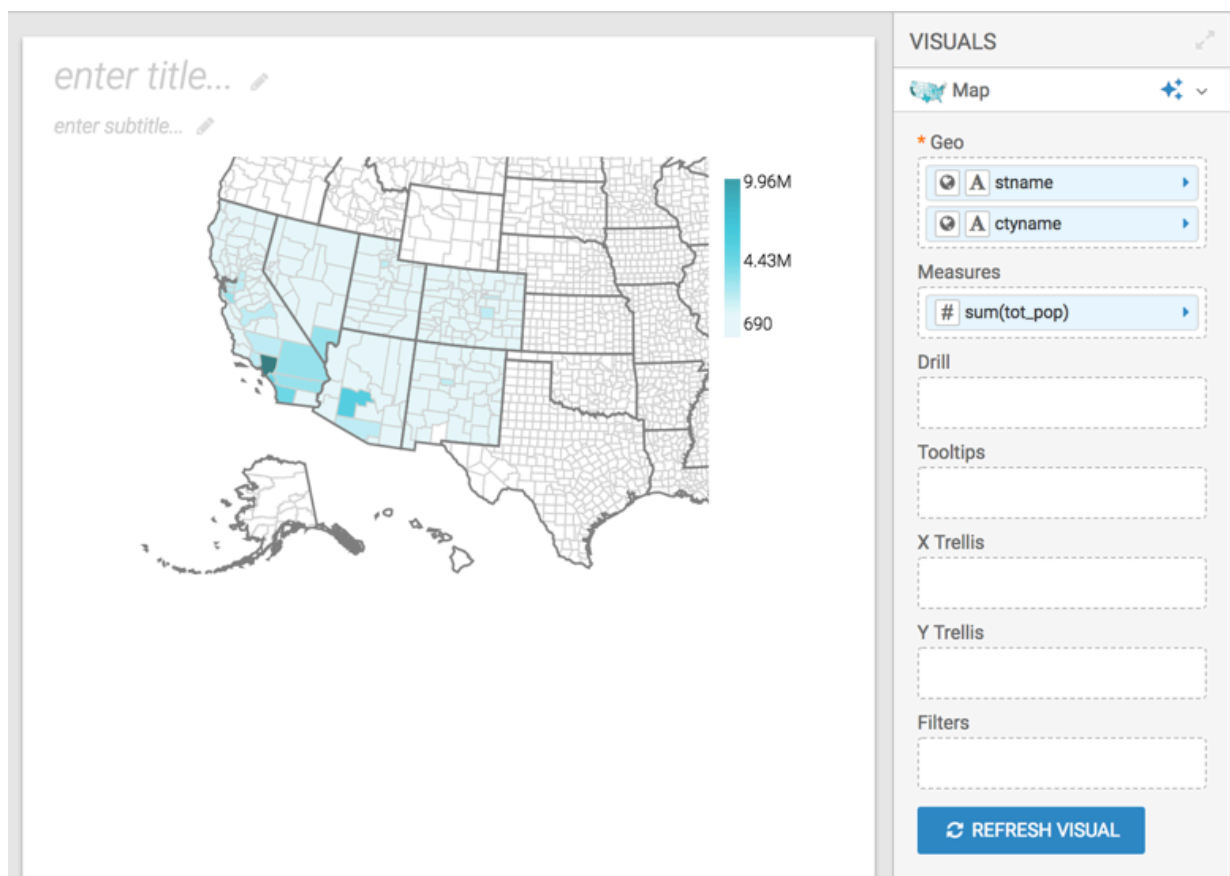
HIDE DATA

☒ Apply Display Format

sql

county	stname	ctyname	tot_pop	tot_male	tot_female
1	Arizona	Apache County	73195	36300	36895
3	Arizona	Cochise County	132088	67820	64268
5	Arizona	Coconino County	136011	67136	68875
7	Arizona	Gila County	53144	26434	26710
9	Arizona	Graham County	37416	20000	17416
11	Arizona	Greenlee County	8802	4611	4191
12	Arizona	La Paz County	20281	10336	9945
13	Arizona	Maricopa County	3942169	1950188	1991981
15	Arizona	Mohave County	203334	102329	101005
17	Arizona	Navajo County	107094	53545	53549

- 5.** If you were to test it by creating a simple map visual on the dataset, it would look something like this:



Working with data models in Cloudera Data Visualization

You can expand the basic one-table dataset by creating joins with other relevant tables from the same or other data stores. 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 for creating data joins for the purpose of 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.

Related Information

[Data modeling](#)

Creating a join

Learn how you can create new data joins in a dataset in Cloudera Data Visualization.

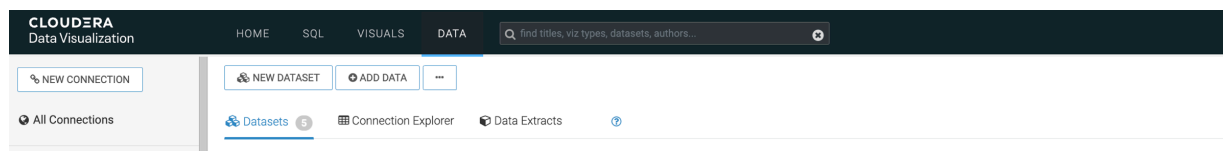
About this task

This example shows you how to create new data joins using the Flight Delays dataset.

Procedure

1. On the main navigation bar, click DATA.

The Data view opens, displaying the Datasets tab.



2. Create a new dataset based on the sample datafile.
3. Find the dataset in the list of datasets, either by scrolling or by using search, and click it.

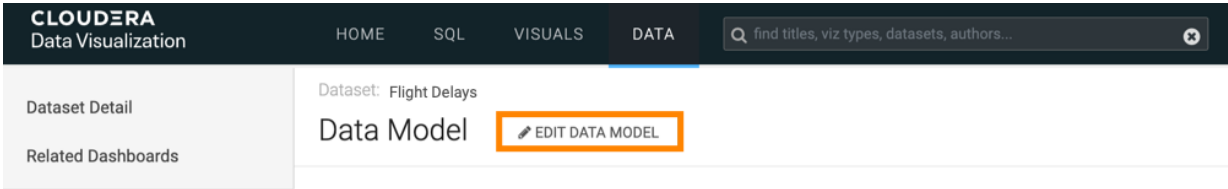
The dataset side navigation panel appears, open on Dataset Detail view.

4. In the side navigation menu, click Data Model.

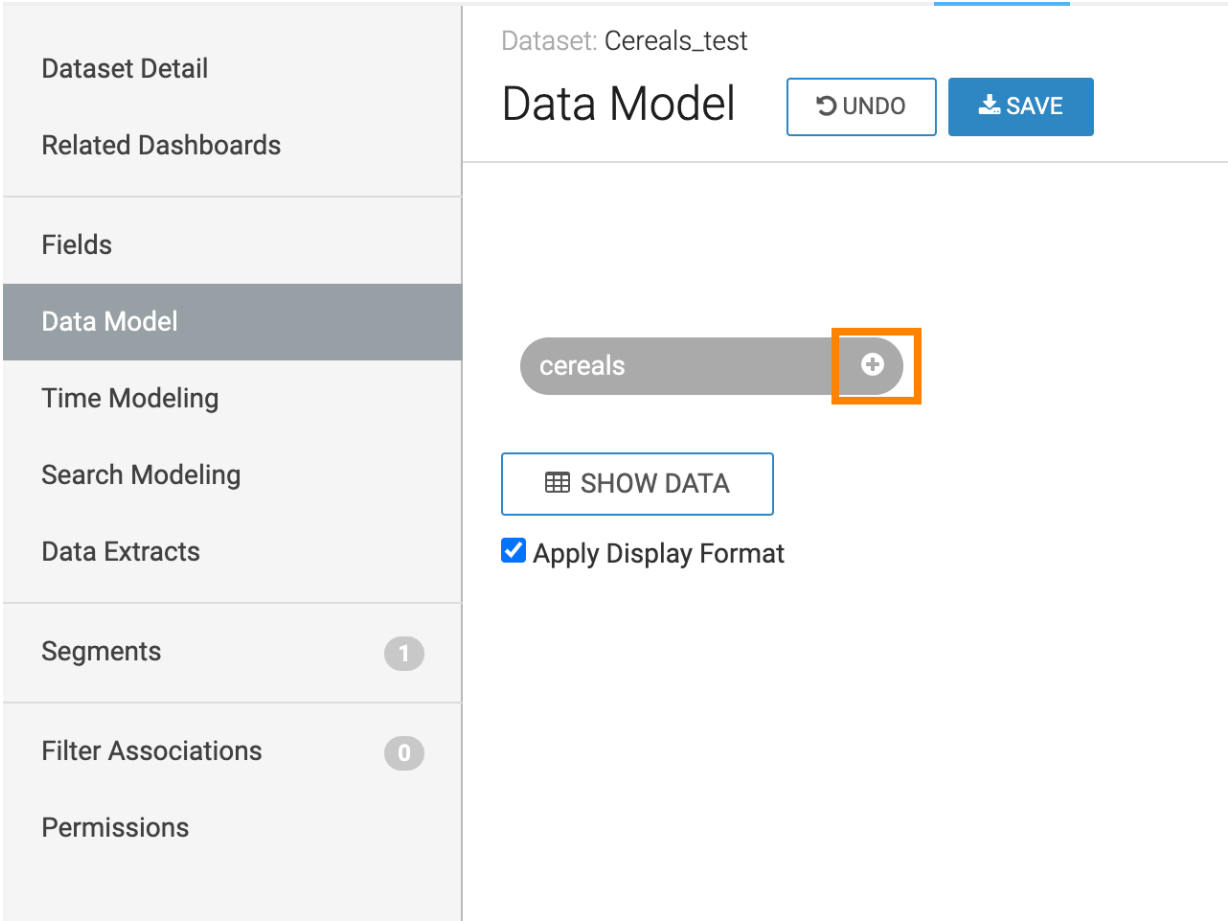
The Data Model view appears, and shows the name of the only table in the dataset. You may click SHOW DATA to display the data of that table.

cereal_name	manufacturer_code	cold_or_hot	calories	protein_grams	fat_grams	sodium_mg	dietary_fiber_grams	complex_carbohydrates_grams	sugars_grams	display_shelf	potassium_mg	vitamins_a
100%_Bran	N	C	70	4	1	130	10	5	6	3	280	25
100%_Natural_Bran	Q	C	120	3	5	15	2	8	8	3	135	0
All-Bran	K	C	70	4	1	260	9	7	5	3	320	25
All-Bran_with_Extra_Fiber	K	C	50	4	0	140	14	8	0	3	330	25
Almond_Delight	R	C	110	2	2	200	1	14	8	3	-1	25
Apple_Cinnamon_Cheerios	G	C	110	2	2	180			10	1	70	25
Apple_Jacks	K	C	110	2	0	125	1	11	14	2	30	25
Basic_4	G	C	130	3	2	210	2	18	8	3	100	25
Bran_Cheer	R	C	90	2	1	200	4	15	6	1	125	25
Bran_Flakes	P	C	90	3	0	210	5	13	5	3	190	25
Cap'nCrunch	Q	C	120	1	2	220	0	12	12	2	35	25
Cheerios	G	C	110	6	2	290	2	17	1	1	105	25
Cinnamon_Toast_Crunch	G	C	120	1	3	210	0	13	9	2	45	25
Clusters	G	C	110	3	2	140	2	13	7	3	105	25
Cocoa_Puffs	G	C	110	1	1	180	0	12	13	2	55	25
Corn_Cheer	R	C	110	2	0	280	0	22	3	1	25	25
Corn_Flakes	K	C	100	2	0	290	1	21	2	1	35	25

5. Click EDIT DATA MODEL.



6. Click the plus sign on the table representation.



The Table Browser modal window opens.

7. In the Table Browser modal window, make the following selections:

- In the Database Name selector, choose the database documentation.



Note: You can join tables from different databases. This value is pre-populated to match the dataset's existing table, but it may be changed.

- In the Table Name selector, choose the table name `airline_id`.

This value is pre-populated to match the existing table of the dataset, but it may be changed.

- Click SELECT.

Table Browser

Choose the table you want to join. You will be able to select the columns that are joined in the next step.

Database Name

main

Table Name

iris

CANCELSELECT

The Edit Join modal window opens.

8. In the Edit Join modal window, the following options are available:
- [Optional] Click Clear Fields to clear all already defined joins between the two tables.
 - [Optional] Click sample data to preview the data. Click again to hide sample data.
 - [Optional] Click Add Join Pair to add another column connection between the same two tables.
 - [Optional] Click Add Join Expression to add a join between the two tables based on a custom SQL expression.
 - [Optional] Click icon (minus) to remove an existing join pair or an existing join expression.
 - [Optional] Under Join Expressions, click the text box to open the Join Expression interface and specify or update a custom SQL expression that defines the join conditions.
 - Click APPLY to save the changes.

Edit Join

×

CLEAR FIELDS

main.cereals

main.iris

select column...

▼

=

select column...

▼

⊖

▶ sample data

▶ sample data

Join Expressions

If you enter multiple expressions they will automatically have an "AND" logic between them

Click to update in SQL expression editor

+ ADD JOIN PAIR

+ ADD JOIN EXPRESSION

CANCEL

APPLY

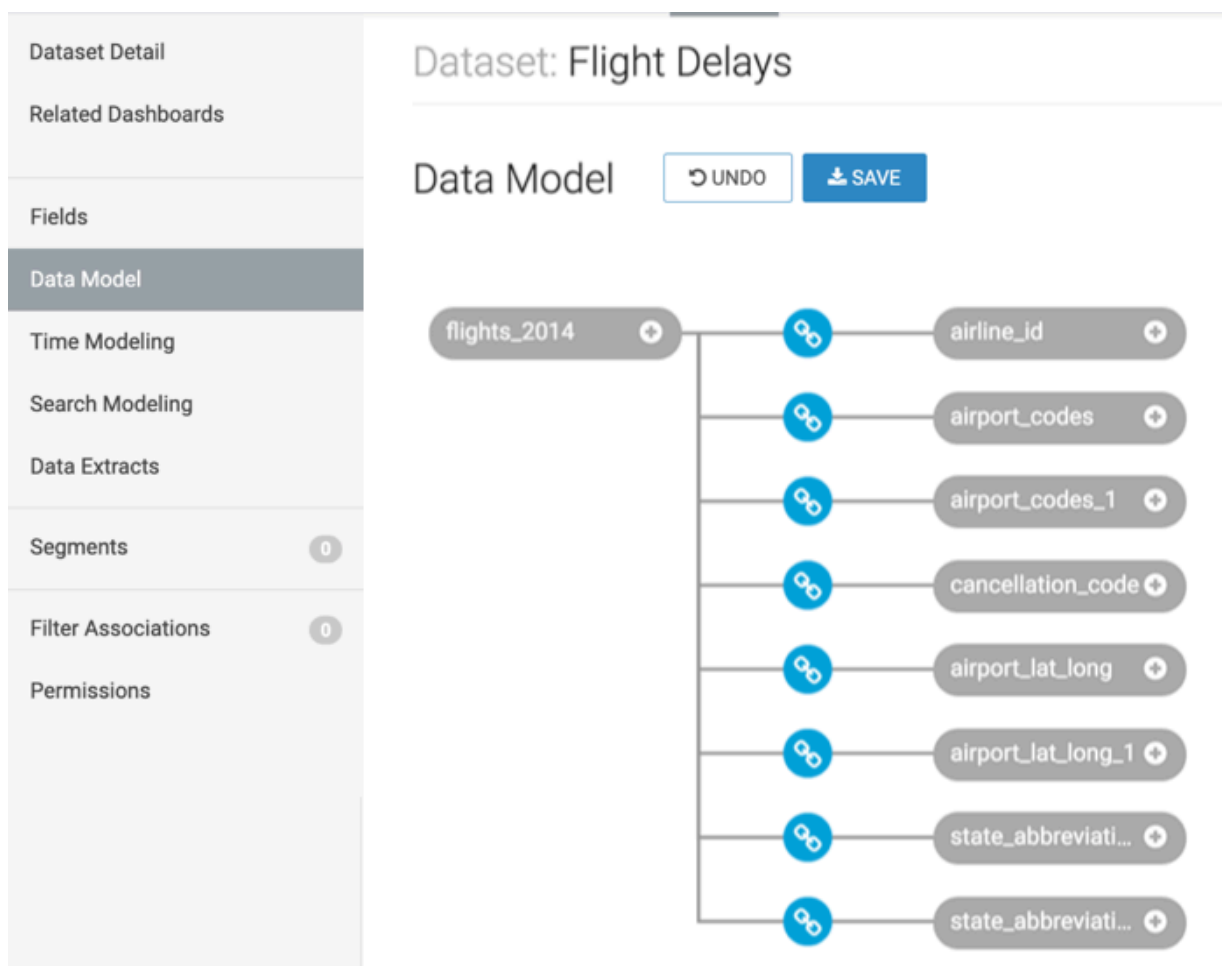
9. In the Edit Join modal window, do the following:
- Select the matching columns for both tables. On the left side, select the field airlineid. On the right side, select the field code.
 - Click Sample Data to view some data in both columns, and verify that the join makes sense. Click again to hide sample data.
10. Click APPLY.

11. Repeat the previous two steps to create seven more joins as follows:

- The table `airport_codes` has two joins to the main table, and you must create each join separately as follows:
 - Left column `origin` = right column `code`.
 - Left column `dest` = right column `code`.
- The table `cancellation_code` has a join for left column `cancellationcode` = right column `code`.
- The table `airport_lat_long` has two joins to the main table, and you must create each join separately as follows:
 - Left column `origin` = right column `locationid`.
 - Left column `dest` = right column `locationid`.
- The table `state_abbreviations` has two joins to the main table, and you must create each join separately:
 - Left column `deststate` = right column `abbreviation`.
 - Left column `originstate` = right column `abbreviation`.

This step is optional, and depends on whether your `flights_*` table has fully extended state names.

12. Click SAVE.



13. Click the (link) icon to edit joins or to change join type.

Changing join types

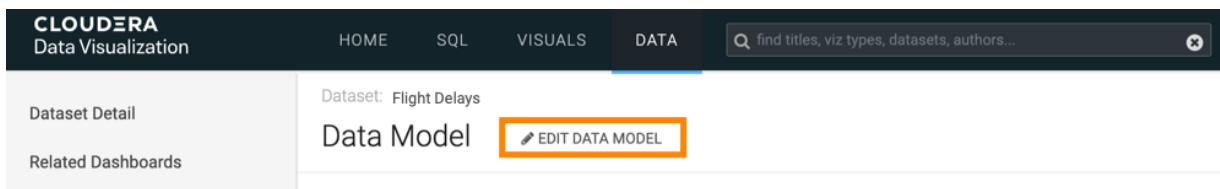
Learn how you can change the join type in a table in Cloudera Data Visualization.

About this task

The following steps demonstrate how to change the join type in the `airport_codes` table from the default Left join to the Right join.

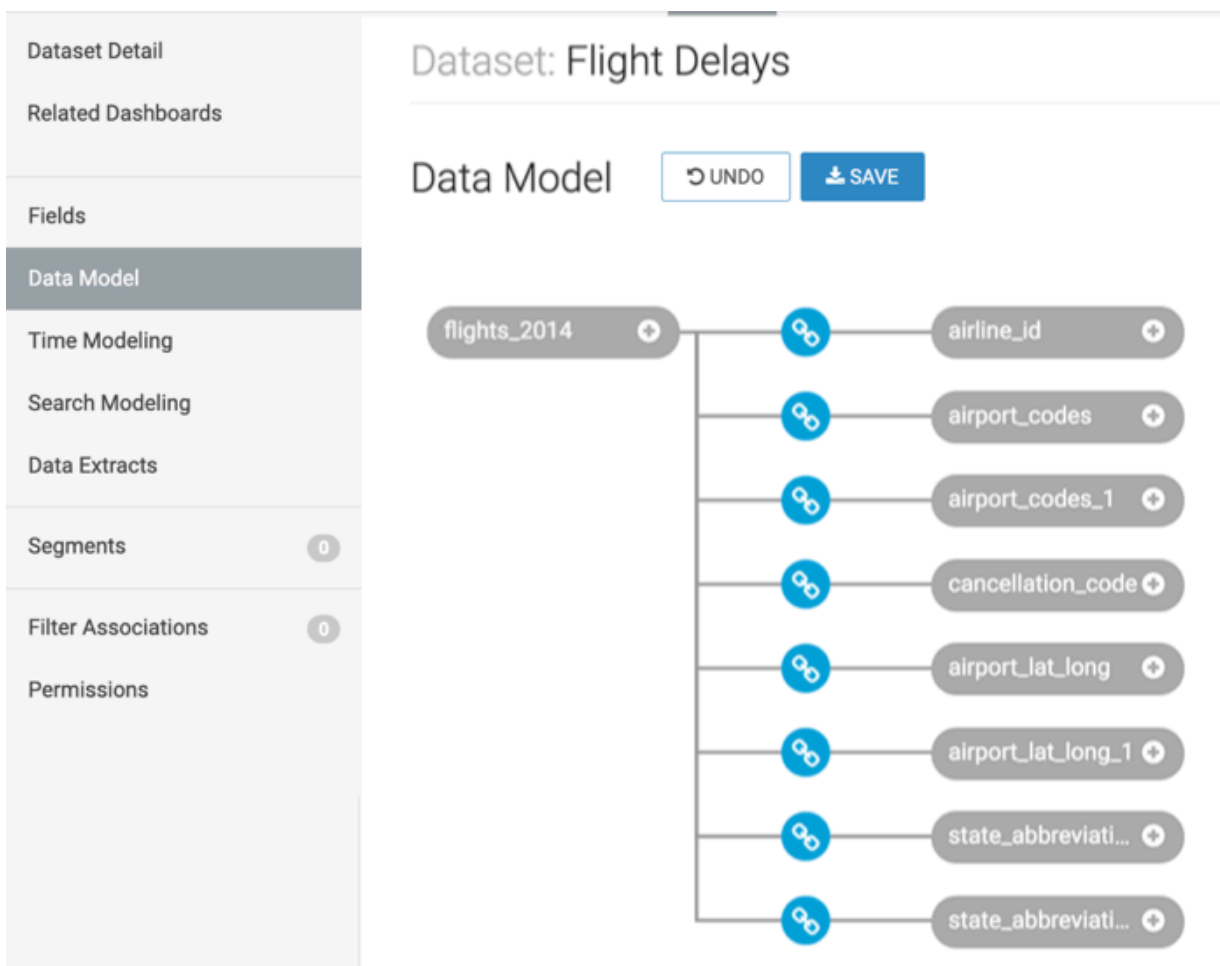
Procedure

1. Navigate to the Data Model page of the dataset.
2. Click Edit Data Model to edit the data model.



3. Click the (link) icon that represents the connection that must be changed.

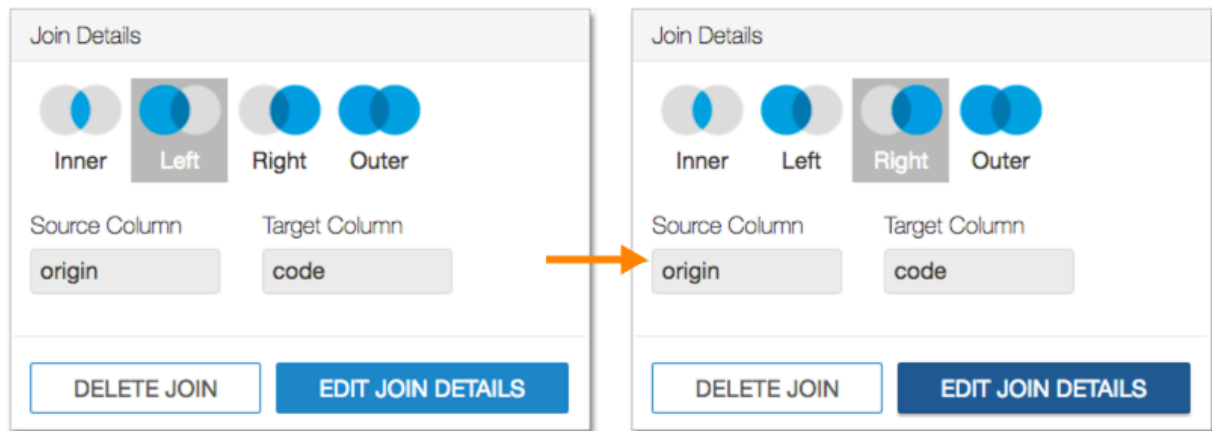
In our example, we clicked the join with the table `airport_codes`.



The Join Details modal window opens.

4. In the Join Details modal window, select an alternate join type.

For example, instead of the default Left join, select Right join.



5. Click outside the Join Details modal window, or click Edit Join Details.
6. Click Save.

Editing join details

Learn how you can change the specifications of existing table joins in Cloudera Data Visualization.

About this task

To demonstrate how to create new data joins, we used dataset Flight Delays, based on data previously imported from a sample datafile

Procedure

1. Navigate to the Data Model page of the dataset.
2. Click EDIT DATA MODEL to edit the model.

3. Click the (link) icon beside the connection you want to edit.

Dataset Detail

Related Dashboards

Fields

Data Model

Time Modeling

Search Modeling

Data Extracts

Segments 0

Filter Associations 0

Permissions

Dataset: test-cereals

Data Model

EDIT DATA MODEL

cereals

cereals_1

cereals_2

SHOW DATA

☒ Apply Display Format

The Join Details modal window opens.

4. Click EDIT JOIN.

Join Details

Inner

Left

Source Column

cereal_name

=

Target Column

manufacturer

manufacturer

=

manufacturer_c...

DELETE JOIN

EDIT JOIN

5. In the Edit Join modal window, the following options are available:

×

CLEAR FIELDS


main.cereals

cereal_name

▼

▶ sample data

=

cereals_2 

manufacturer

▼

▶ sample data

⊖

manufacturer

▼

▶ sample data

=

manufacturer_code

▼

▶ sample data

⊖

Join Expressions

If you enter multiple expressions they will automatically have an "AND" logic between them

Click to update in SQL expression editor

+ ADD JOIN PAIR

+ ADD JOIN EXPRESSION

CANCEL

APPLY

a. CLEAR FIELDS allows you to clear all already defined joins between the two tables.

b. The (pencil) icon allows you to edit the name of the joined table.



Important: This action is only available when a join is created to the same table. In these cases, the table name is duplicated with a number appended to distinguish it. This alias is applied to both dimension and measure names within the joined dataset.

If you wish to modify this duplicated alias for clarity or consistency, you can edit the duplicated table name.

When you provide a new alias, it is applied to the corresponding dimension and measure names of the dataset.

c. sample data allows you to preview the data from the tables. Click it again to hide the sample data.

d. The (minus) icon allows you to remove an existing join pair or an existing join expression.

e. Clicking the Join Expressions text box opens the Join Expression interface where you can specify or update a custom SQL expression that defines the join conditions.

f. ADD JOIN PAIR can be used to add another column connection between the same two tables.

g. ADD JOIN EXPRESSION can be used to add a join between the two tables based on a custom SQL expression.

Click APPLY to save the changes.

6. To add a join expression and replace the original field:field join, perform the following steps in the Edit Join modal window:
 - a. Remove the initial join between the two columns by clicking the (minus) icon.
 - b. Under Join Expressions, click the text box to open the Join Expression interface, where you can specify or update the custom SQL expression that defines the join conditions.

For example, you can enter the following expression to only show data where the amount of complex carbohydrates equals the amount of sugars, and the calorie content is greater than 250.

```
[Complex Carbohydrates Grams]=[Sugars Grams] AND [cereals_1 Calories] > 250
```

Join Expression ✕

1 [Complex Carbohydrates Grams]=[Sugars Grams] AND [cereals_1 Calories] > 250

VALIDATE EXPRESSION

☒ Autocomplete on

Validation Successful! ✕

All Functions

abs
AND
avg
BETWEEN
CASE
cast
char
COALESCE
count

All Fields

Calories
A Cereal Name
cereals_1 Cal...
A cereals_1 Cer...
A cereals_1 Col...
cereals_1 Co...
1.2 cereals_1 Cu...
cereals_1 Die...
cereals_1 Dis

CANCEL

APPLY

7. Click APPLY to save the expression and return to the Edit Join modal window.

8. In the Edit Join modal window, verify that the new join expression appears under Join Expressions and click APPLY.

×

Edit Join

CLEAR FIELDS


main.cereals

select column...

▼

▶ sample data

=

cereals_1 

select column...

▼

▶ sample data

+

Join Expressions

If you enter multiple expressions they will automatically have an "AND" logic between them

[Complex Carbohydrates Grams]=[Sugars Grams] AND [cereals_1 Calories] > 250

+ ADD JOIN PAIR

+ ADD JOIN EXPRESSION

CANCEL

APPLY

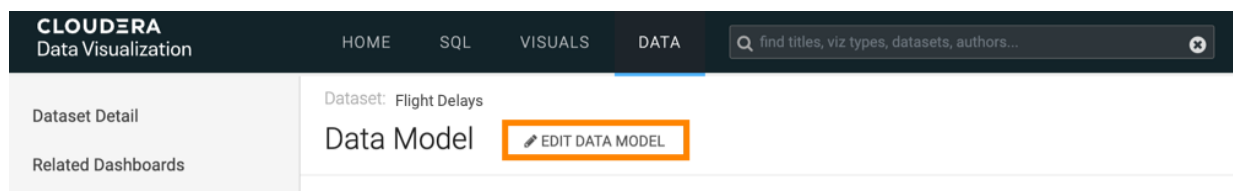
9. To revert this change prior to saving, click UNDO.
10. Click SAVE.

Deleting a join

Learn how you can remove a join that exists between tables in your dataset.

Procedure

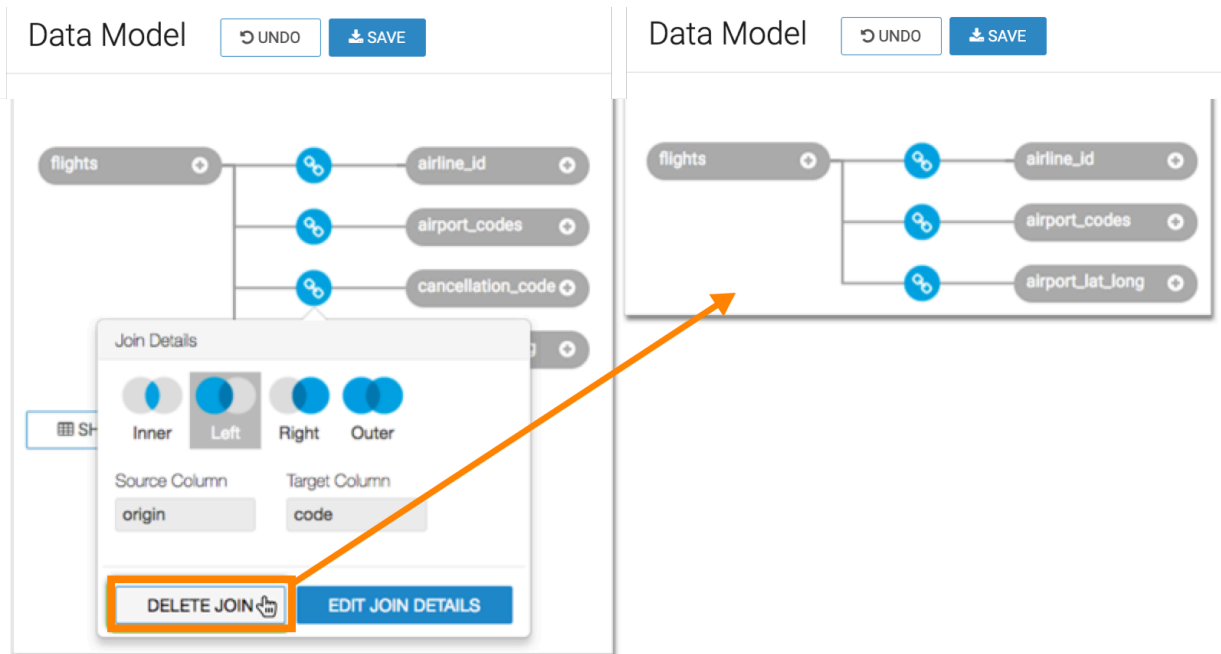
1. Navigate to the **Data Model** page of the dataset.
2. Click EDIT DATA MODEL to enable editing of the data model.



3. Click the (link) icon representing the connection you want to delete.
The **Join Details** modal window opens.

4. Click DELETE JOIN.

For example, if you delete the `cancellation_code` connection, you can see that after the deletion, this table no longer appears in the Data Model.

**5. Click UNDO if you want to revert the change before saving.****6. Click SAVE to apply your changes.**

Applying field display format on sample data

Learn how you can test field display formats you have configured.

Procedure

1. Navigate to the Data Model page of the dataset.

2. Select Apply Display Format.

Selecting/deselecting the checkbox applies or removes the formatting without refetching the data.



Note: The display format checkbox setting will be remembered.

For information on how to configure field display at the dataset level, see *Changing the field display format*.

3. You can click SHOW DATA to view a sample of your data model.

trips_detail								trips				
pickup_datetime	passenger_count	trip_distance	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_latitude	pickup_neighborhood	trips_pickup_neighborhood	pickup_boro	pickup_hour	ride_cnt	total_amo
2013-10-07 20:13:00	1	10	-73.8734512329	40.7741127014	-73.9803848267	40.7706794739	Airport	Airport	Queens	0	44	2138.2
2013-10-07 20:13:00	1	10	-73.8734512329	40.7741127014	-73.9803848267	40.7706794739	Airport	Airport	Queens	1	263	13743.07
2013-10-07 20:13:00	1	10	-73.8734512329	40.7741127014	-73.9803848267	40.7706794739	Airport	Airport	Queens	2	463	24303.97
2013-10-07 20:13:00	1	10	-73.8734512329	40.7741127014	-73.9803848267	40.7706794739	Airport	Airport	Queens	3	895	45841.37
2013-10-07 20:13:00	1	10	-73.8734512329	40.7741127014	-73.9803848267	40.7706794739	Airport	Airport	Queens	4	1076	52089.04
2013-10-07 20:13:00	1	10	-73.8734512329	40.7741127014	-73.9803848267	40.7706794739	Airport	Airport	Queens	5	1127	52571.34
2013-10-07 20:13:00	1	10	-73.8734512329	40.7741127014	-73.9803848267	40.7706794739	Airport	Airport	Queens	6	1138	50456.47
2013-10-07 20:13:00	1	10	-73.8734512329	40.7741127014	-73.9803848267	40.7706794739	Airport	Airport	Queens	7	1033	47973.92
2013-10-07 20:13:00	1	10	-73.8734512329	40.7741127014	-73.9803848267	40.7706794739	Airport	Airport	Queens	8	1050	47469.01

Related Information

[Changing the field display format](#)

Working with data profiling

Dataset profiling provides a comprehensive view of your dataset by combining field-level statistics with relationship analysis. It helps you understand the structure, content, and quality of your data, so you can prepare it for analysis or visualization.

About this task

Dataset profiling offers two complementary perspectives:

- **Profiling** provides detailed statistics for each dimension or measure in a dataset, such as data type distributions, missing or null values, unique counts, and duplication rates. Column-level charts and summary metrics highlight distributions, patterns, and anomalies.
- **Correlation** shows how strongly different fields are related, making it easier to identify meaningful patterns, key drivers, or unexpected relationships in your data.

By combining these views, dataset profiling helps you quickly evaluate data quality, detect anomalies, and determine whether a dataset is fit for your analytical goals before you start building visualizations or applying transformations.

Using profiling

The Profiling tab provides statistics and visualizations for selected fields in a dataset. You can explore dimensions and measures, view data types, refresh results, and manage cards to better understand the dataset.

Before you begin

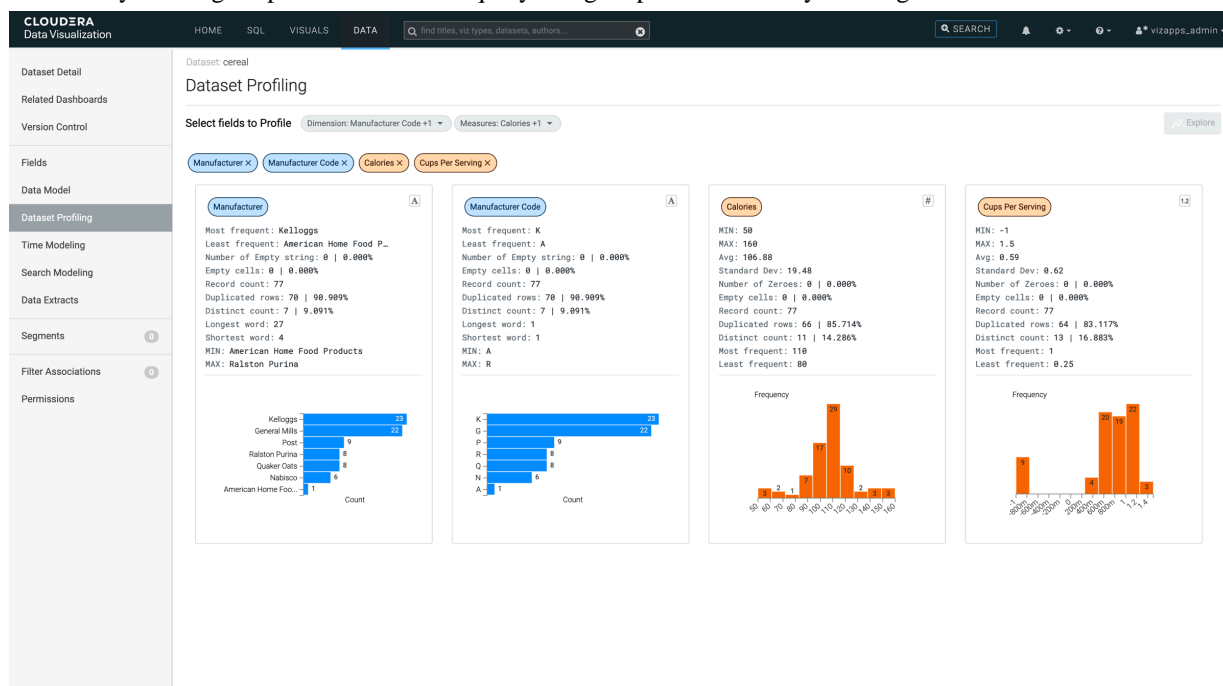
Enable data profiling in [Site Settings Data Enable Data Profiling](#) . For more information, see [Managing data-related site settings](#).

Procedure

1. Click DATA on the main navigation bar.
The Data view appears, open on the Datasets tab.
2. Find the dataset that you want to profile, either by scrolling through the list or using the search function.
3. Click the dataset you want to examine.
The dataset side navigation pane opens for the selected dataset, displaying the Dataset Detail page.
4. Click Dataset Profiling in the left navigation panel.
The Dataset Profiling view opens on the Profiling tab. This tab provides the option to select fields from the dimensions and measures of the dataset.

5. Select the fields that you want to examine and click Explore.

Cards appear with statistics and visualizations for each selected field. If the query is running too long, you can cancel it by clicking Stop. You can rerun a query and get updated results by clicking Refresh.



Tip:

- The data type of the specific dimension or measure is shown in the top-right corner of the card.
- If there are a lot of cards displayed, you can click a selected data field pill to jump directly to its card.
- To remove a card, click # located in the corresponding data field label pill in the top row. This also stops any running queries.

The following information is shown for the selected dataset fields.

Dimension values:

Most frequent

The string that appears the most often in the dataset, helping to identify the most common value.

Least frequent

The string that appears the least often in the dataset, helping to identify rare or unusual values.

Number of Empty string

The count of cells that contain an empty string, for example cells that are not null but contain no characters.

Empty cells

The number of cells that are completely empty, including both null values and empty strings.

Record count

The total number of records in the dataset.

Duplicated rows

The number of rows that are exactly identical, helping to identify redundant data. A high number of duplicates often signifies low variability in the dataset.

Distinct count

The number of unique values in the dataset, showing how varied the data is.

Longest word

The string with the highest number of characters in the dataset, providing insight into the potential size of values.

Shortest word

The string with the fewest number of characters, indicating the shortest data entry.

MIN

The minimum value in the dataset based on alphabetical order, useful for sorting or range analysis.

MAX

The maximum value in the dataset based on alphabetical order, helping to define the upper range of the data.

The count histogram for dimensions shows how many times each category appears, that is the number of occurrences (counts) of each category (dimension). This visualization is useful for understanding the quality, distribution, and structure of your dataset.

Measure values:**MIN**

The smallest numeric value in the dataset, showing the lower bound of your data.

MAX

The largest numeric value in the dataset, showing the upper bound of your data.

Avg (Average)

The sum of all numeric values divided by the total number of records, giving an overall sense of the central value.

Standard Dev

Standard deviation that shows how spread out the numbers are in the dataset. A higher value indicates more variability in the data and a lower value indicates that the data points are closer to the average.

Number of Zeroes

The count of records where the value is exactly zero, indicating gaps or null-equivalents in the data.

Empty cells

The number of cells that are completely empty, including both null values and missing values.

Record count

The total number of records in the dataset.

Duplicated rows

The number of rows where all values are identical, helping to identify redundant data. A high number of duplicates often signifies low variability in the dataset.

Distinct count

The number of unique values in the dataset, showing how varied the data is.

Most frequent

The numeric value that appears the most often in the dataset, helping to identify the most common value.

Least frequent

The numeric value that appears the least often in the dataset, helping to identify rare or unusual values.

The frequency histogram for measures shows the distribution of numeric values (measures) within specific ranges (buckets), giving insight into their spread and common ranges.

Using correlation

On the Correlation tab you can calculate correlations between fields in your dataset. A correlation is a statistical measure of how closely two variables move together (positive, negative, or no correlation).

About this task



Important: Valid correlations can only be calculated using numerical fields (for example: INT, DOUBLE, TINYINT, BIGINT).

Before you begin

Enable data profiling in [Site Settings Data Enable Data Profiling](#) . For more information, see [Managing data-related site settings](#).

Procedure

1. Click DATA on the main navigation bar.
The Data view appears, open on the Datasets tab.
2. Find the dataset that you want to profile, either by scrolling through the list or using the search function.
3. Click the dataset you want to examine.
The dataset side navigation pane opens for the selected dataset, displaying the Dataset Detail page.
4. Click Dataset Profiling in the left navigation panel.
The Dataset Profiling view opens on the Profiling tab.
5. Switch to the Correlation tab.
6. Select the fields that you want to calculate correlation for and click Correlate to run the calculation.
 - Dimension: You choose one categorical or grouping variable.
 - Measures: You choose one or more numeric metrics you want to check correlation against.

The correlation results are displayed.

If the query is running too long, you can cancel it by clicking Stop.

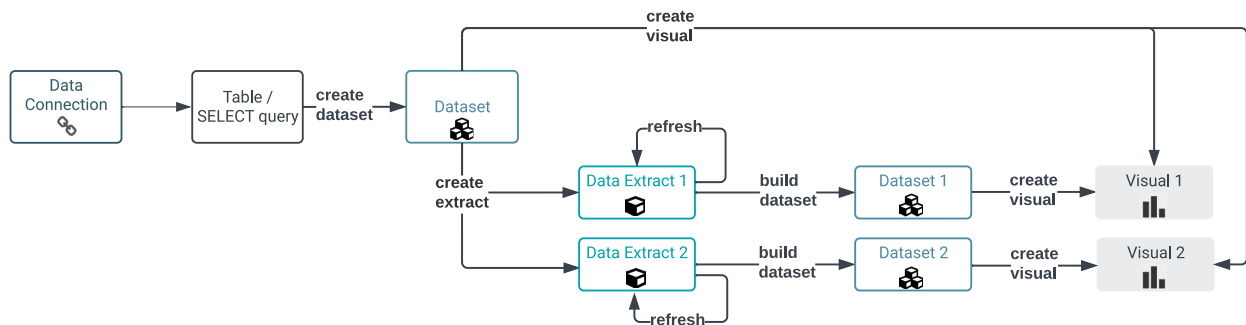
You can rerun a query and get updated results by clicking Refresh.

Working with data extracts in Cloudera Data Visualization

Data extracts are saved subsets of data that you can use for data discovery and analytics.

You create an extract to reduce the total amount of data you work with by selecting certain dimensions and measures. With the help of data extracts you can manage the analytical capabilities, performance, concurrency and security of data access in your system.

The following diagram shows you the workflow of building visualizations on top of data extracts:



Creating a data extract

Learn how you can extract data from a dataset to a table in the same or a different data connection.

Before you begin

- You need to enable the data extract feature in [Site Settings Data Enable Data Extracts](#) . For more information, see [Managing data related site settings](#).
- You need the following privileges to work with data extracts:
 - Manage dataset right for the source dataset
 - Manage AVs/extracts right for source and target data connections.

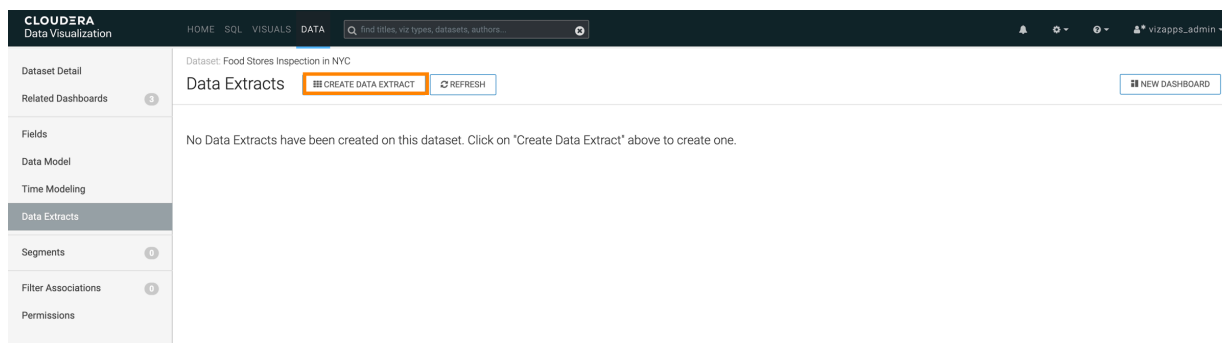
About this task

Follow these steps to create a data extract:

Procedure

- Click DATA on the main navigation bar.
The Data view opens, displaying the Datasets tab.
- In the left navigation panel, select the connection on which your source dataset is defined.
- Select the dataset you want to use as a source for the data extract.
- Click Data Extracts in the left navigation panel.
In the Data Extracts view, you can create data extracts and manage extracts that exist from this dataset.

5. Click CREATE DATA EXTRACT.



The CREATE DATA EXTRACT modal window opens.

Create Data Extract

×

A Data Extract is a copy of the data in this dataset to a table in the same or a different Data Connection. Once an extract is created, it can be refreshed manually or on a schedule.

Data Extract

Advanced

Refresh Schedule ⓘ

Manual

Target Data Connection * ⓘ

Select data connection

Target Database * ⓘ

Target Table Name (starting with cdv_) * ⓘ

* Please select at least one field below

Q Search

✖

■ Dimensions

☐ cereal_name
 ☐ manufacturer_code
 ☐ cold_or_hot
 ☐ manufacturer_code_15
 ☐ manufacturer

☒ Use aggregate functions below ⓘ

■ Measures

☐ # Record Count
 ☐ # sum(calories)
 ☐ # sum(protein_grams)
 ☐ # sum(fat_grams)
 ☐ # sum(sodium_mg)
 ☐ # sum(dietary_fiber_grams)
 ☐ # sum(complex_carbohydrates_grams)
 ☐ # sum(sugars_grams)
 ☐ # sum(display_shelf)
 ☐ # sum(potassium_mg)
 ☐ # sum(vitamins_and_minerals)
 ☐ # sum(weight_ounces)
 ☐ 1.2 sum(cups_per_serving)

CANCEL

CREATE DATA EXTRACT

6. Select the frequency for the Data Extract refreshes.

The default is Manual, but you can create custom schedules. For more information, see *Creating new schedules*.

7. Select the data connection where the Data Extract will be stored.

It can be the same connection where the source dataset is or you can also choose a different connection.



Note: In Cloudera Data Warehouse you cannot choose SQLite data connections as targets because those connections are read-only.

8. Select the database on the previously selected connection where you want to store the Data Extract.

9. Enter a target table name, starting with `cdv_`.



Note: `cdv_` is the default prefix for target table names. If you have admin privileges, you can change the prefix in the Advanced Site Settings by changing the value in `DATA_EXTRACTS_TARGET_TABLE_NAME_PREFIX = "cdv_"`.

10. Select the dimension and measure fields depending on what data you want to store in your extract.

If you want to use the selected measures as they are and you do not want the columns to be aggregated, deselect the Use aggregate functions below option.



Note: You must select at least one dimension or measure field.

When configuring a data extract, you can search for a column by name, as well as view the field types and complex data types available.

11. If you want to send an email confirmation about the extract's status, switch to the Advanced tab of the Create Data Extract modal window.



Important: The Advanced tab is only visible if you enabled sending visuals and receiving extract refresh status messages through email in Site Settings.

- Select what event(s) you want to be notified about - success, failure or both.
- Add one or more email addresses as recipients.
- You can edit the default subject line of the notification email.

Create Data Extract ✕

A Data Extract is a copy of the data in this dataset to a table in the same or a different Data Connection. Once an extract is created, it can be refreshed manually or on a schedule.

Data Extract

Advanced

Email Refresh Status Upon ☐ Success ☐ Fail

To *

Subject * ⓘ

CANCEL

CREATE DATA EXTRACT

12. Click CREATE DATA EXTRACT after all required fields are completed.

When the extract is created, you can refresh it manually or it will be refreshed on the schedule you defined in [Step 5](#).

Dataset: Food Stores Inspection in NYC

Data Extracts CREATE DATA EXTRACT REFRESH NEW DASHBOARD

ID	Target Connection	Target Table	Target Datasets	State	Last Successful Run	Schedule	Columns	
6 ⓘ	doc-test	main.cdv_test_extract	CREATE DATASET	Never run		Manual	Dimensions: county, owner...	EDIT RUN NOW DELETE



Note: After the source dataset becomes associated with the data extract, a new tag appears in the Datasets list showing Extract Source.

Results

The extract is created, and its state shows as Never run.

What to do next

You must run the extract to populate the target table and allow building a dataset or dashboard on the target table.

Related Information

[Creating new schedules](#)

Exploring data extract details

Learn how you can examine the available information about your data extracts.

Procedure

1. Click DATA on the main navigation bar.
The Data view opens, displaying the Datasets tab.
2. In the left navigation panel, select the connection where your source dataset is defined.
3. Select the dataset of your data extract.
4. Click Data Extracts in the left navigation panel to display the Data Extracts view.

You can see the list of available data extracts with the following information shown:

- ID
- Target Connection
- Target Table
- Target Datasets
- State
- Last Successful Run
- Schedule
- Columns

You can also access the following action buttons:

- Edit
- Run now
- Delete

Running a data extract

Learn how to run or refresh a data extract.

About this task

After creating a data extract, you must run it as an initial 'refresh' to populate the target table and allow building a dataset or dashboard on the target table. When the original data changes, becomes stale or invalid, you can manually refresh the extracted data by running your data extract.

Procedure

1. Click DATA on the main navigation bar.
The Data view opens, displaying the Datasets tab.
2. In the left navigation panel, select the connection where your source dataset is defined.
3. Select the dataset of your extract.

- Click Data Extracts in the left navigation panel.
- Select the data extract you want to run/refresh.
- Click RUN NOW.

A confirmation modal window opens where you have to confirm the action.

- Click RUN EXTRACT.

A green success message is displayed and while the refresh is in progress, the state of the extract shows as Running. When the refresh is successfully completed, the state of the extract changes to Success.

Dataset: Food Stores Inspection in NYC

Data Extracts CREATE DATA EXTRACT REFRESH NEW DASHBOARD

ID	Target Connection	Target Table	Target Datasets	State	Last Successful Run	Schedule	Columns	
6	doc-test	main.cdv_test_extract	CREATE DATASET	Success	2023-01-30 16:09:25	Manual	Dimensions: county, owner...	EDIT RUN NOW DELETE

You can check out the extract refresh job logs on the Job Logs tab of the Jobs page. For more information, see *Job log*.

CLoudERA Data Visualization HOME SQL VISUALS DATA vizapps_admin

Type: All Adhoc Email Scheduled Email Threshold Email Data Extract

Status: All Error Finished Running Cancelled Creator: All

Status	Job ID	Log ID	Type	Name	Details	Initiated by	Start Time	Total Run Time	
Finished	11	71	Data Extract	Adhoc Data Extract: 6	Source Dataset: Food Stores Inspection in NYC Target Data Connection: doc-test Target Table: main.cdv_test_extract	vizapps_admin	2023-01-30 16:09	a few seconds	DETAILS RUN NOW CANCEL
Finished	7	70	Data Extract	Data Extract: 4	Source Dataset: doc-test-dataset2 Target Data Connection: doc-test Target Table: main.cdv_world_misc_data Custom refresh schedule schedule	vizapps_admin	2023-01-30 15:31	a few seconds	DETAILS RUN NOW CANCEL
Finished	10	57	Data Extract	Adhoc Data Extract: 2	Source Dataset: doc-test-dataset2 Target Data Connection: doc-test Target Table: main.cdv_world_population	vizapps_admin	2023-01-27 14:16	a few seconds	DETAILS RUN NOW CANCEL
Finished	9	47	Data Extract	Adhoc Data Extract: 3	Source Dataset: doc-test-dataset2 Target Data Connection: doc-test Target Table: main.cdv_world-life-expectancy	vizapps_admin	2023-01-27 11:08	a few seconds	DETAILS RUN NOW CANCEL
Finished	8	46	Data Extract	Adhoc Data Extract: 2	Source Dataset: doc-test-dataset2 Target Data Connection: doc-test Target Table: main.cdv_world_population	vizapps_admin	2023-01-27 11:08	a few seconds	DETAILS RUN NOW CANCEL
Finished	6	6	Data Extract	Adhoc Data Extract: 4	Source Dataset: doc-test-dataset2 Target Data Connection: doc-test Target Table: main.cdv_world_misc_data	vizapps_admin	2023-01-26 14:11	a few seconds	DETAILS RUN NOW CANCEL

You can configure extracts to be automatically refreshed, and you can define a schedule for the automatic refreshes to update the data included in the extract.

If you set a refresh schedule for the extract to run, you can see the job logs on the Scheduled Jobs tab of the Jobs page. For more information, see *Scheduled jobs*.

Jobs REFRESH

Job Logs (last few) Scheduled Jobs Manage Schedule Intervals DELETE SELECTED RUN NOW RESUME

Type: All Scheduled Email Threshold Email Data Extract

Status: All Error Finished Running Cancelled Paused Creator: All

Job ID	Type	Name	Details	Created By	Schedule	Last Run	Next Run	Last Status	
7	Data Extract	Data Extract: 4	Source Dataset: doc-test-dataset2 Target Data Connection: doc-test Target Table: main.cdv_world_misc_data Custom refresh schedule schedule	vizapps_admin	Custom refresh schedule	2023-01-30 15:31	2023-01-30 16:15	✓	DETAILS EDIT RUN NOW PAUSE

Showing 1 to 1 of 1 entries

Previous Next

You can edit an existing schedule in the Edit Data Extract modal window from the Data Extracts view. For more information, see *Editing a data extract*. Or you can also update an existing schedule in the Manage Schedule Intervals interface. For more information, see *Changing schedule intervals*.

Related Information

[Job log](#)

[Scheduled jobs](#)

[Editing a data extract](#)

[Changing schedule intervals](#)

Creating a dataset from a data extract

Learn how to create a second dataset built on the extracted data, which is used as the dataset for building visuals.

Procedure

1. Click DATA on the main navigation bar.
The Data view opens, displaying the Datasets tab.
2. Click Data Extracts.
3. Find the data extract that you want to use for the new dataset.
4. Click CREATE DATASET.

Dataset: doc-test-dataset2

Data Extracts CREATE DATA EXTRACT REFRESH NEW DASHBOARD

ID	Target Connection	Target Table	Target Datasets	State	Last Successful Run	Schedule	Columns	
2	doc-test	main.cdv_world_population	doc-test-dataset-population CREATE DATASET	Success	2023-01-27 14:16:49	Manual	Dimensions: country, un_re...	EDIT RUN NOW DELETE
3	doc-test	main.cdv_world-life-expect...	doc-test-dataset-life-expec... main-world-life-expectancy CREATE DATASET	Success	2023-01-27 11:08:59	Manual	Dimensions: country, year, l...	EDIT RUN NOW DELETE
4	doc-test	main.cdv_world_misc_data	doc-test-dataset-misc CREATE DATASET	Success	2023-01-30 16:16:18	Custom refresh schedule	Dimensions: country	EDIT RUN NOW DELETE

5. Add a name in the Dataset title field and click CREATE.



Note: After the dataset becomes associated with a data extract, a new tag appears in the Datasets list showing Extract Target.

Editing a data extract

Learn about the option to edit the data extract after it has been created.

Procedure

1. Click DATA on the main navigation bar.
The Data view opens, displaying the Datasets tab.
2. Click Data Extracts.

3. Select the data extract you want to edit and click EDIT.

The Edit Data Extract modal window opens.

Edit Data Extract

A Data Extract is a copy of the data in this dataset to a table in the same or a different Data Connection. Once an extract is created, it can be refreshed manually or on a schedule.

Data Extract

Advanced

Refresh Schedule ⓘ

Manual

Target Data Connection * ⓘ

doc-test

Target Database * ⓘ

main

Target Table Name (starting with cdv_) * ⓘ

cdv_world_population

* Please select at least one field below

Dimensions

☒ country
 ☒ un_region

Use aggregate functions below ⓘ

Measures

☐ Record Count
 ☐ sum(year)
 ☐ sum(life_expectancy)
 ☐ sum(population)

CANCEL

EDIT DATA EXTRACT

4. Make changes in the available fields.

- You can set a Data Extract refresh schedule.

Note: Editing the refresh schedule of a data extract does not delete previous jobs or logs. Instead, the schedule for the existing job is simply updated, ensuring that all historical data remains intact.

- You can change the target data connection and database.
- You can update the name of the target table.
- You can add or remove dimension and measure columns.

Important: If you change the target data connection, target database, target table name or target table fields the target table will be deleted and a new one will be created.

5. If you want to send an email confirmation about the extract's status, switch to the Advanced tab of the Edit Data Extract modal window.

Important: The Advanced tab is only visible if you enabled sending visuals and receiving extract refresh status messages through email in Site Settings.

- Select what event(s) you want to be notified about - success, failure or both.
- Add one or more email addresses as recipients.
- You can edit the default subject line of the notification email.

6. Click EDIT DATA EXTRACT to save your changes.

93

Deleting a data extract

Learn how you can delete a data extract.

About this task

For various business reasons or because the extract has become invalid, you may choose to delete an existing data extract.

Procedure

1. Click DATA on the main navigation bar.
The Data view opens, displaying the Datasets tab.
2. Click Data Extracts.
3. Select the data extract you want to edit and click DELETE.

A confirmation modal window opens where you have to confirm the action.



Important: Deleting an extract does not automatically remove the associated target table. The confirmation modal offers you an option to drop the target table when deleting the extract. If you want to delete the target table, mark the checkbox.

Confirmation

Confirm that you want to delete this extract by clicking the "Delete Extract" button.

Note that deleting this extract will NOT remove the target table. If you want to also delete the extract target table, select the checkbox below.

☐ Delete extract table

CANCEL

DELETE EXTRACT

4. Click DELETE EXTRACT.



Note: Deleting an extract does not delete the target table or any datasets that were created on the extract target.