

Working with Derived Data

Date published: 2020-10-30

Date modified: 2022-09-21

CLOUDERA

Legal Notice

© Cloudera Inc. 2024. All rights reserved.

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

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

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

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

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

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

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

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

Contents

| | |
|---|-----------|
| Defining Derived Data..... | 4 |
| Defining Additional Derived Data..... | 6 |
| Edit derived data..... | 10 |
| Using Derived Data..... | 11 |
| Viewing Derived Data Definitions..... | 14 |
| Saving derived data with full context..... | 16 |
| Saving derived data expression only..... | 16 |
| Deleting Derived Data Definitions..... | 16 |

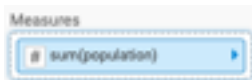
Defining Derived Data

Procedure

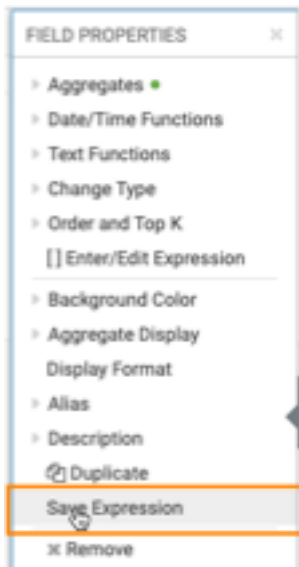
1. Inside a visual that generates a subset of data, select a measure field on a shelf.

In this case, we are using sum(population).

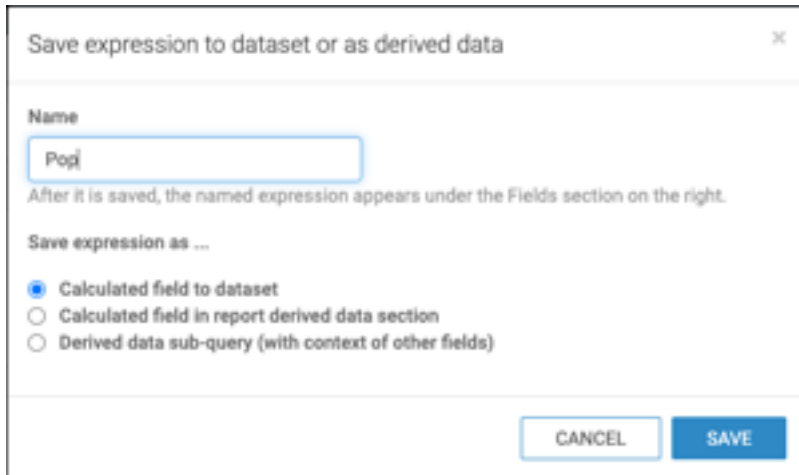
2. Click the Right Arrow icon to the right of the field.



3. In the dropdown, click Save Expression.



4. In the Save expression to dataset or as derived data modal window, under Name, type Pop, and click Save.



Save expression to dataset or as derived data

Name

Pop

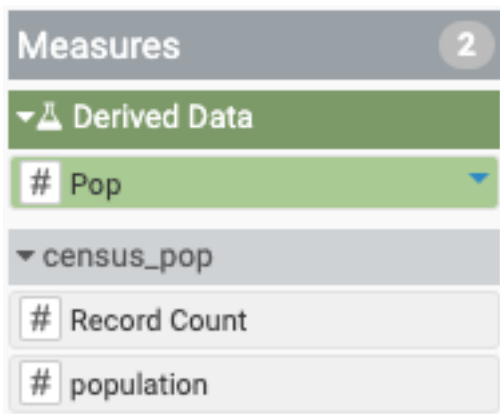
After it is saved, the named expression appears under the Fields section on the right.

Save expression as ...

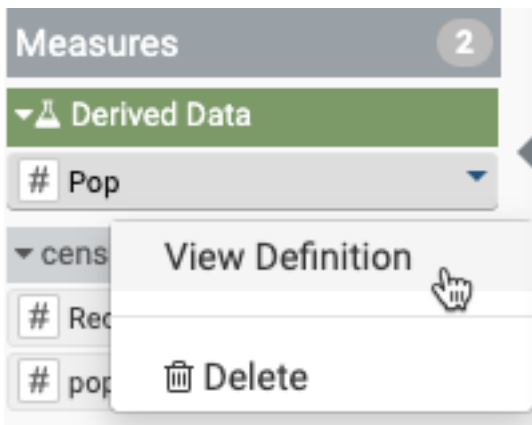
- Calculated field to dataset
- Calculated field in report derived data section
- Derived data sub-query (with context of other fields)

CANCEL SAVE

You can see that in the left navigation under Measures, you have a new category Derived Data that contains Pop.



- To check the definition of the saved derived data Pop, click the down arrow to the right of its name, and click View Definition.



The Derived Data Definition modal window appears. You can see that the definition includes a parameterized filter for year.



- Click Save to save both the visual and the new derived data definitions.

Results

You can now use Pop as any other measure.

Defining Additional Derived Data

About this task

Next, define other derived data in the same shelf set-up, and save it as derived data LifeExpXPop

Procedure

- On the Dimensions shelf, place the following fields: un_region, un_subregion, and country.
- Add the derived data Pop to the Dimensions shelf. We do this because fields on the Measure shelf must be aggregated, but the definition of Pop already includes the summation over the population field.
- Place the field life_expectancy on the Measures shelf, and change the aggregate to avg(life_expectancy).

4. Place the field year on the Filters shelf , and change the expression to [year]=<<year_param:2000>>.

The screenshot shows a dashboard configuration interface with three shelves:

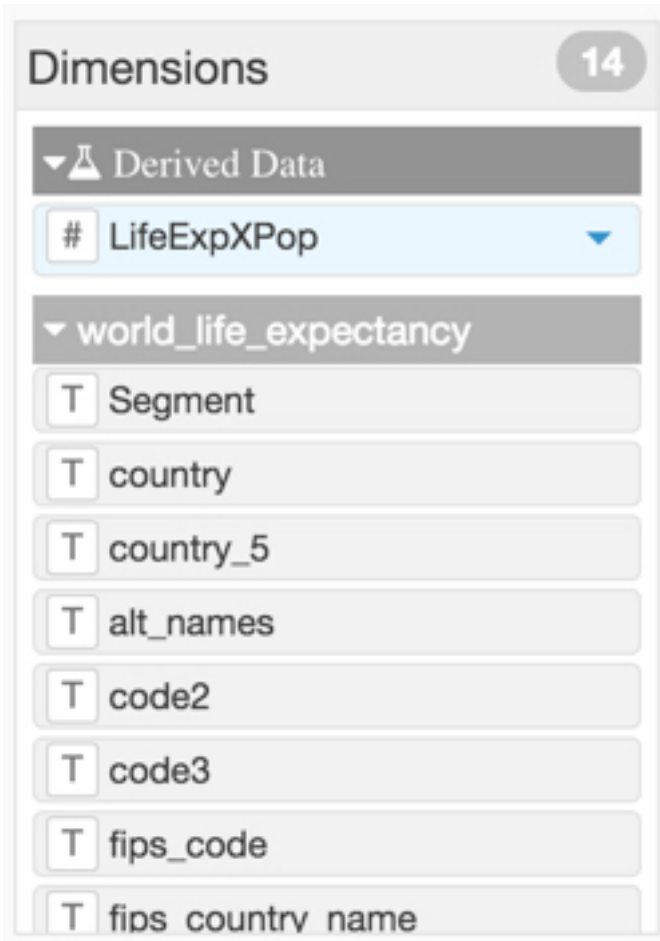
- Dimensions:** Contains three fields: 'T un_region', 'T un_subregion', and 'T country' (with a sort icon and '1'). Below them is a measure field '# Pop'.
- Measures:** Contains one field: '# avg(life_expectancy)'.
- Filters:** Contains one field: 'year=<<year_param:2000...'

5. Duplicate Derived data Pop in place, on the Dimensions shelf.
 6. Click the Down Arrow icon to the right of the cloned field, and in the dropdown, click Enter/Edit Expression.
 7. In the Enter/Edit Expression modal window, change the expression to [life_expectancy]*[Pop], and Save.

The screenshot shows the 'Enter/Edit Expression' modal window. The expression field contains '[life_expectancy]*[Pop]'. Below the field are buttons for 'Validate Expression' and a checked checkbox for 'Autocomplete on'. To the right, there are two dropdown menus: 'All Functions' and 'All Fields'. The 'All Fields' dropdown is open, showing a list of fields including 'iso_cc', 'lat', 'life_expectancy', 'lng', 'population', 'Record Count', 'un_region', 'un_subregion', and 'year'. At the bottom right, there are 'Cancel' and 'Save' buttons, with a mouse cursor clicking on the 'Save' button.

8. Save the same field as derived data, named LifeExpXPop.

9. Notice that in the Data menu, under Dimensions, you have a new category Derived Data that contains LifeExpXPop.



10. To check the definition of the saved derived data LifeExpXPop, click the down arrow to the right of its name, and click View Definition.

11. The Derived Data Definition modal window appears. Note that the definition includes a parametrized filter for year, and the Pop derived data field.

Derived Data Definition

Label: LifeExpXPop

Dimensions: [un_region] as '__un_region'
[un_subregion] as
 '__un_subregion'
[country] as '__country'
[ascending]
[Pop] as '_Pop'
[life_expectancy]*[Pop] as
 'LifeExpXPop'

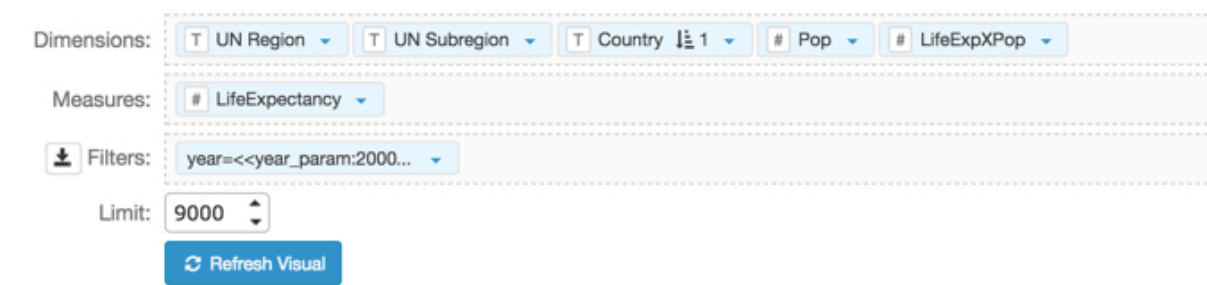
Measures: avg([life_expectancy]) as
 'avg(life_expectancy)'

Filters: [year]=<<year_param:2000>>

Limit: 100

12. Replace the field on the Dimensions shelf with the new derived data field, click Refresh Visual, and Save the application.

The visual should look similar to this one. Note that we aliased the fields and made some changes to display formats:



| UN Region | UN Subregion | Country | Pop | LifeExpXPop | LifeExpectancy |
|-----------|---------------------------|---------------------|------------|---------------|----------------|
| Asia | Southern Asia | Afghanistan | 22,856,302 | 1,254,811,040 | 54.90 |
| Europe | Southern Europe | Albania | 3,071,856 | 227,931,710 | 74.20 |
| Africa | Northern Africa | Algeria | 30,533,828 | 2,106,834,175 | 69.00 |
| Africa | Middle Africa | Angola | 13,926,373 | 629,472,083 | 45.20 |
| Americas | Caribbean | Antigua and Barbuda | 77,656 | 5,707,716 | 73.50 |
| Americas | South America | Argentina | 36,930,708 | 2,725,486,419 | 73.80 |
| Asia | Western Asia | Armenia | 3,076,098 | 219,941,012 | 71.50 |
| Americas | Caribbean | Aruba | 90,271 | 6,662,000 | 73.80 |
| Oceania | Australia and New Zealand | Australia | 19,164,352 | 1,531,231,785 | 79.90 |

< 1 2 3 4 5 >

Edit derived data

Editing derived data definitions in

About this task

Procedure

1. Delete the existing derived data definition, see [Deleting derived data definitions](#).
2. Define a new derived data with the same name, see [Defining derived data](#).

Related Information

[Defining Derived Data](#)

[Deleting Derived Data Definitions](#)

Using Derived Data

About this task

To use the derived data measurements in a calculation, treat it as any other measurement. In the example used here, we use the derived data LifeExpXPop to accurately calculate the weighted life expectancy for an aggregation dimension, such as UN Region or UN Subregion. We do this by dividing the sum of Life Expectancy X Population by sum(population) at the appropriate aggregate level.

Procedure

1. Place the following fields on the Dimensions shelf: un_region and un_subregion.
2. Add the derived data field Pop.
3. Place the field year on the Filters shelf, and change the expression to [year]=<<year_param:2000>>.
4. Place the derived data field LifeExpXPop onto the Measures shelf. Notice that it appears as an aggregation, sum(LifeExpXPop).
5. Using the Enter/Edit Expression window modal, change the definition of the field as follows, and click Save. Notice that we aliased the field.

```
(sum([LifeExpXPop])/sum([population])) as 'Weighted Life Expectancy'
```

6. Use the Alias option to change the column names to UN Region, UN Subregion

7. Click Refresh Visual.

The visual should look similar to this one:

Dimensions: T UN Region 1 T UN Subregion 2

Measures: # Weighted Life Expectancy # Population

Filters: year=<<year_param:200...

Limit:

[Refresh Visual](#)

| UN Region | UN Subregion | Weighted Life Expectancy | Population |
|-----------|---------------------------|--------------------------|---------------|
| Africa | Eastern Africa | 52.37 | 108,167,918 |
| Africa | Middle Africa | 47.42 | 89,208,823 |
| Africa | Northern Africa | 68.72 | 98,182,244 |
| Africa | Southern Africa | 50.60 | 1,757,925 |
| Africa | Western Africa | 54.14 | 49,296,775 |
| Americas | Caribbean | 68.85 | 23,593,157 |
| Americas | Central America | 70.25 | 27,565,249 |
| Americas | Northern America | 79.38 | 30,723,560 |
| Americas | South America | 71.35 | 279,783,195 |
| Asia | Central Asia | 64.17 | 19,911,544 |
| Asia | Eastern Asia | 72.97 | 1,401,620,301 |
| Asia | South-Eastern Asia | 67.01 | 225,842,357 |
| Asia | Southern Asia | 62.31 | 1,206,917,948 |
| Asia | Western Asia | 71.56 | 54,154,366 |
| Europe | Eastern Europe | 70.78 | 28,274,513 |
| Europe | Northern Europe | 75.83 | 18,353,582 |
| Europe | Southern Europe | 78.82 | 79,244,298 |
| Europe | Western Europe | 78.50 | 159,577,216 |
| Oceania | Australia and New Zealand | 79.90 | 19,164,352 |
| Oceania | Melanesia | 67.60 | 811,718 |
| Oceania | Micronesia | 71.35 | 239,157 |

8. [Optional] To see the correctly calculated Life Expectancy for a UN Region, simply remove UN Subregion from the Dimensions shelf.

Dimensions:

Measures:

Filters:

Limit:

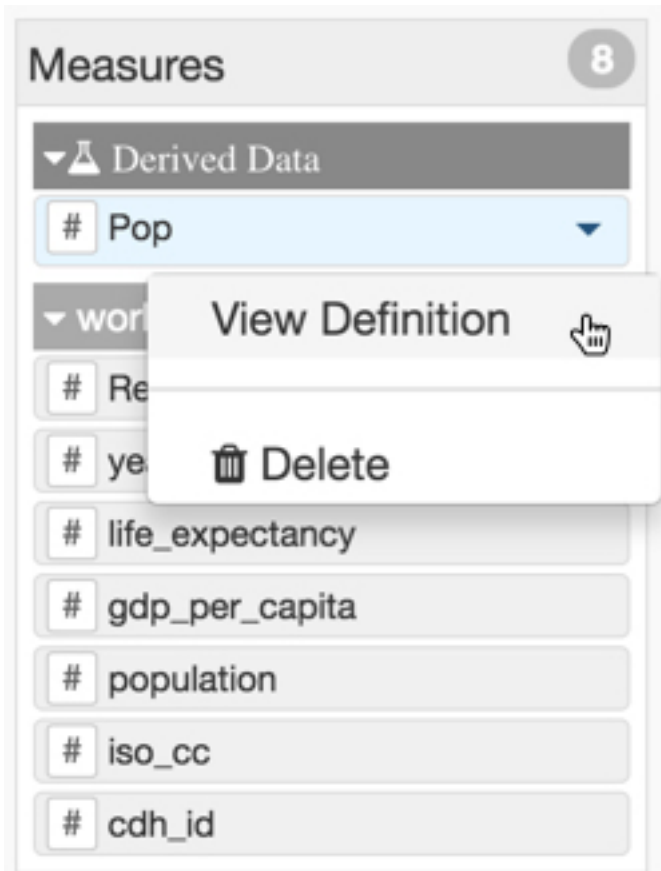
| UN Region | Weighted Life Expectancy | Population |
|-----------|--------------------------|---------------|
| Africa | 55.97 | 346,613,685 |
| Americas | 71.79 | 361,665,161 |
| Asia | 68.00 | 2,908,446,516 |
| Europe | 77.65 | 285,449,609 |
| Oceania | 79.22 | 20,452,865 |

Viewing Derived Data Definitions

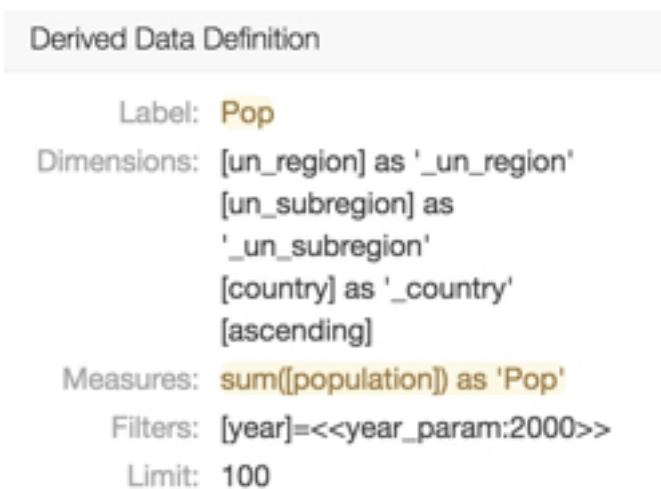
Procedure

- In the left navigation, under Measures, under Derived Data, select the derived measure you wish to check.
In this exercise, this is Pop.
- Click the Down Arrow icon to the right of the field.

- In the drop-down, select View Definition.



- The Derived Data Definition modal window appears.



Note that you cannot edit the definition of derived data. If you want to change it, you must first delete it, and then create and save a new definition. See, [Deleting Derived Data Definitions](#) and [Defining Derived Data](#)

Related Information

[Deleting Derived Data Definitions](#)

[Defining Derived Data](#)

Saving derived data with full context

Procedure

Saving derived data expression only

Procedure

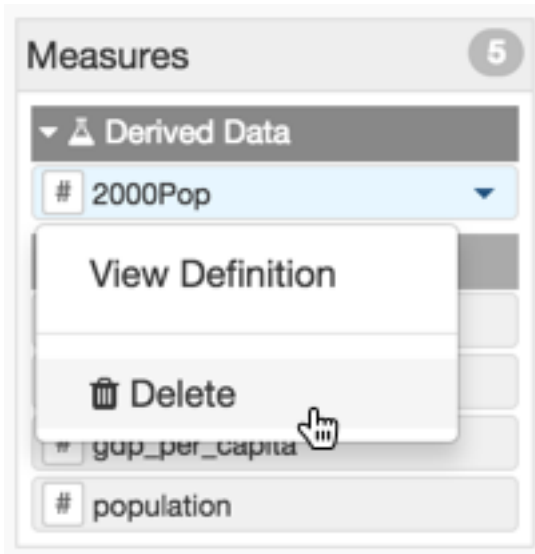
Deleting Derived Data Definitions

Procedure

1. In the left navigation under Measures, under Derived Data, identify the derived measure you wish to delete.

In this exercise, this is 2000Pop.

2. Click the Down Arrow icon to the right of the field.
3. In the drop-down, select Delete.



4. Click Save to save both the visual and the changes to derived definitions.