Cloudera Data Visualization 8.0.9

Configuring Natural Language Search

Date published: 2020-10-30
Date modified: 2025-12-16

https://docs.cloudera.com/

Legal Notice

Contents

Enabling Natural Language Search in Site Settings

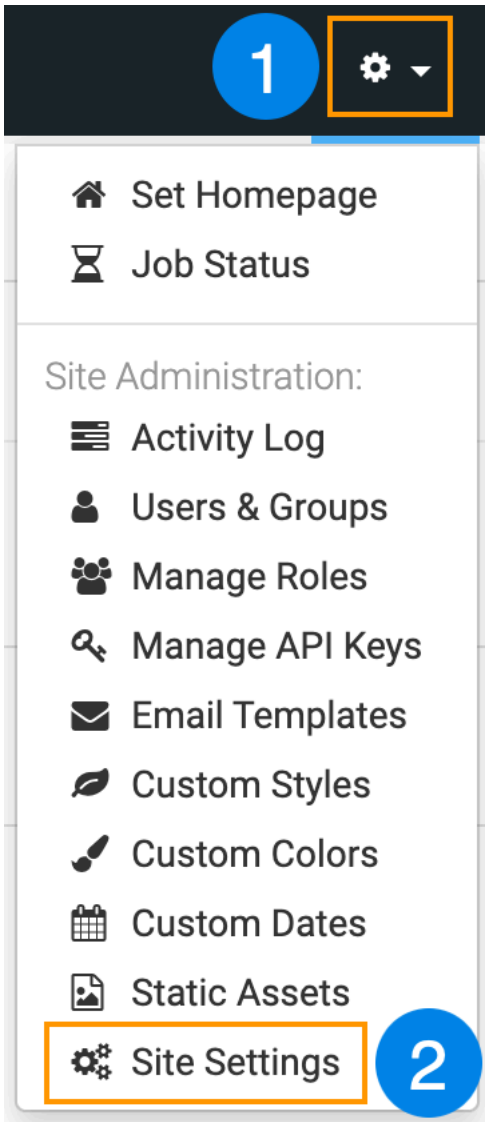
Before using Natural Language Search (NLS) in Cloudera Data Visualization, you must enable the feature in the Site Settings.

About this task

Note: This setting is only available for users with administrative privileges.

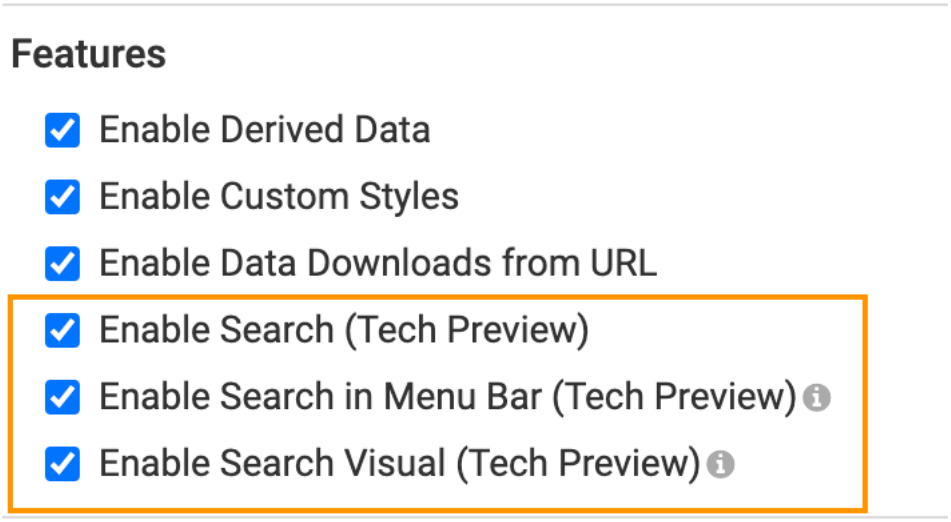
Procedure

1. Click the Gear icon in the upper right corner to open the Site Administration menu.
2. Click Site Settings.



The settings page appears.

3. Select the following options under Features:
   - Enable Search - You have to enable this option for the other two to take effect.
   - Enable Search in Menu Bar - This option allows you to open the search modal on any screen from the icon from the Search button in the top right corner.
   - Enable Search Visual - This option allows you to add search visuals to a dashboard.



4. Click SAVE in the upper left corner.

Enabling a dataset for NLS

Before configuring Natural Language Search (NLS) in Cloudera Data Visualization, you must enable and configure the datasets.

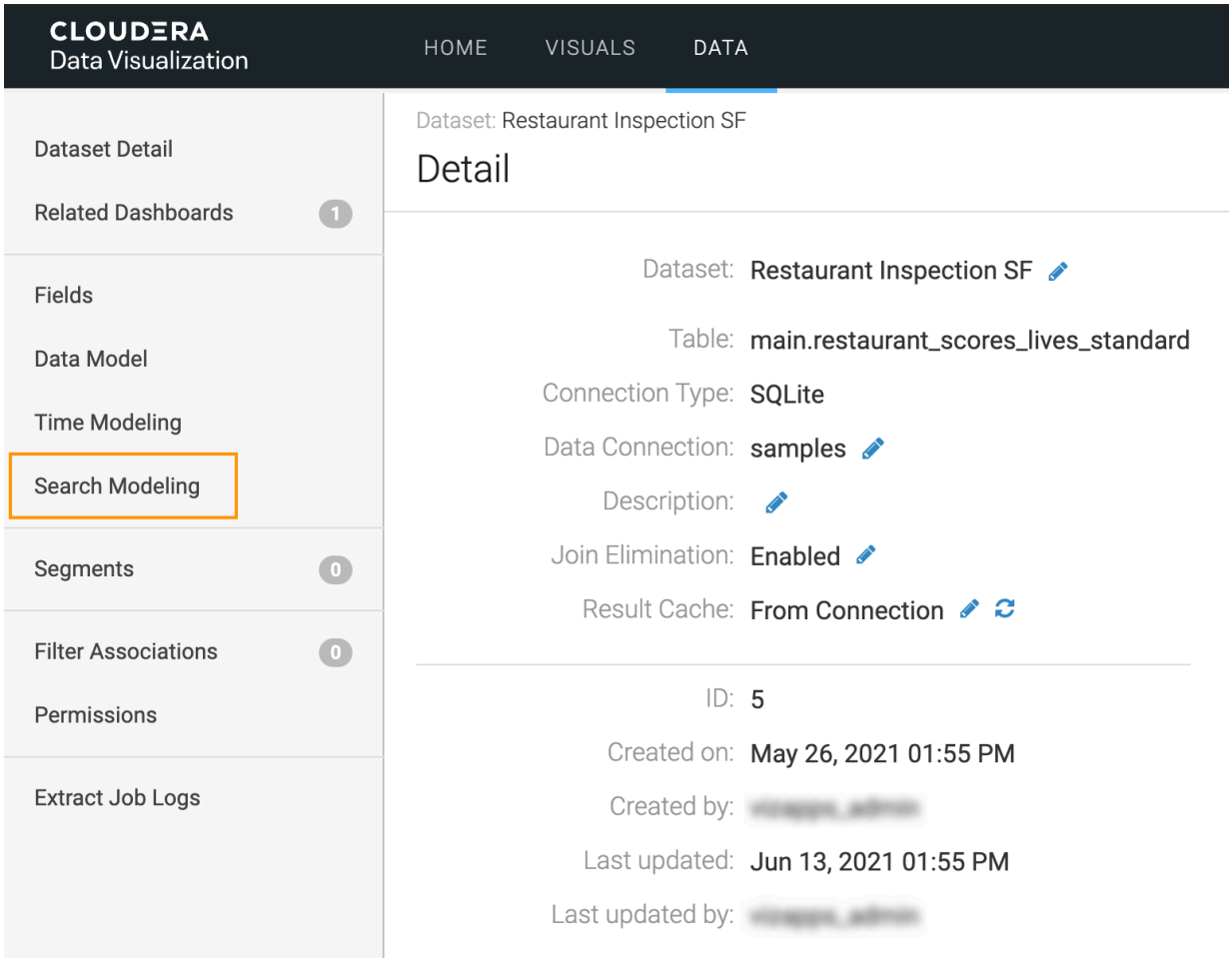
Procedure

1. On the main navigation bar, click DATA.

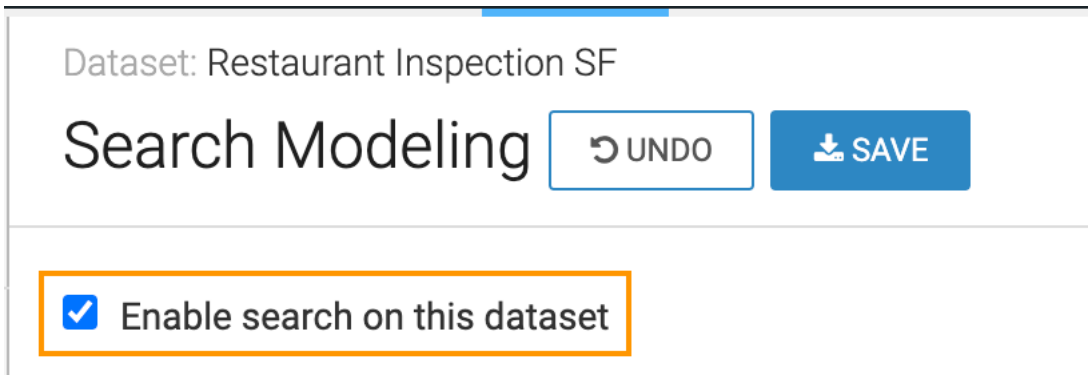
2. In the list of datasets, click the one you want to enable for NLS.

   The Dataset Detail page appears.

   In the figure below, the Restaurant Inspection SF dataset is selected.



3. Click Search Modeling.
4. In the Search Modeling page, select Enable search on this dataset.



5. Click SAVE.

Note: On the list of datasets, a search-enabled dataset has an active Search icon.

Note: To disable the search on a dataset, clear the Enable search on this dataset checkbox.

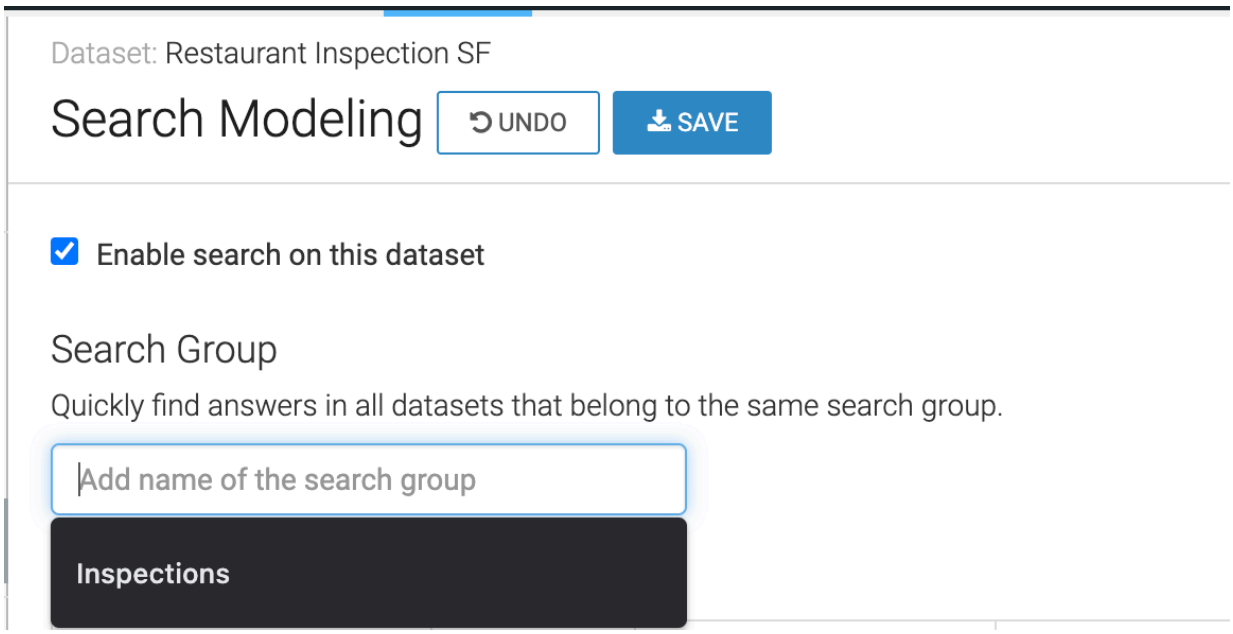
Specifying a group for a dataset in NLS

Cloudera Data Visualization enables you to group datasets for Natural Language Search (NLS). You can use a group to more easily combine datasets during search operations.

Procedure

1. Under Search Group, add a new name for the group or choose one from the already existing ones.

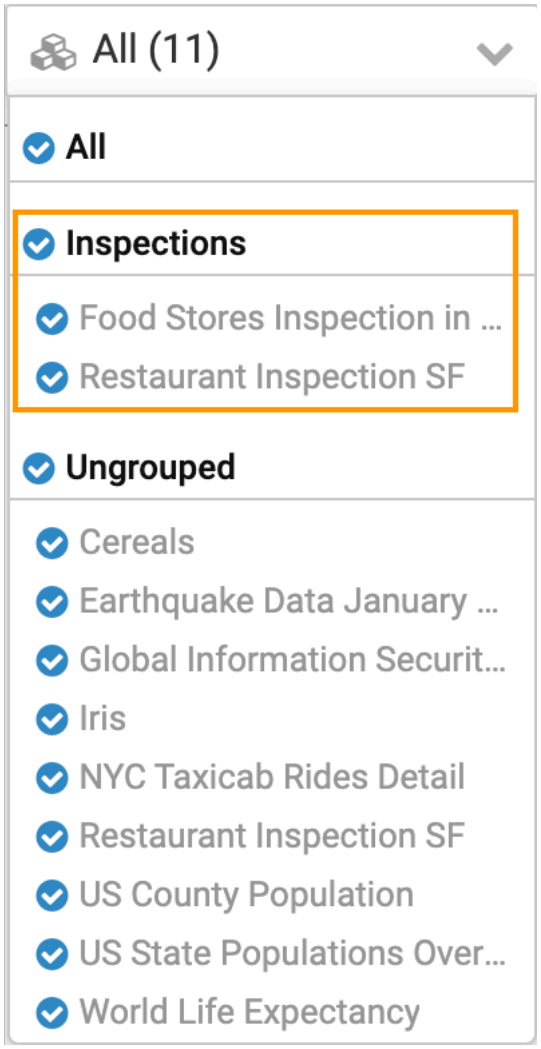
   In the figure below, the dataset is added to the group 'Inspections'.



2. Click SAVE.

Results

You can see the datasets grouped in the drop-down list in the Search modal window.



Specifying fields for a dataset in NLS

Cloudera Data Visualization makes it possible to limit Natural Language Search (NLS) to values of specified fields of the dataset.

About this task

You can restrict the search for a number of reasons, either from operational considerations, or for security reasons.

Procedure

1. Under Fields, specify which fields may be searched through NLS.

   In the example below, most of the fields are selected. If you want to disable search for a specific field in a dataset, simply remove all words and phrases from Matching search terms of that field. In the example below, the terms from Record Count were removed.

2. Choose an option from the drop-down list under Auto Search Priority.

   Higher priority columns are searched first, then medium, and finally low.

   Note: Enabling Auto Search Priority will result in additional queries even if the column is not specifically mentioned in the search and can affect performance. Only enable it for columns from which users want to pull information.

   Fields

Field	Data Type	Dimension or Aggregate	Matching search terms	Auto Search Priority
[Record Count]	BIGINT	Aggregate		NONE
[business_id]	BIGINT	Dimension	business_id	HIGH
[business_name]	STRING	Dimension	business establishment restaurant	LOW
[business_address]	STRING	Dimension	business_address	NONE
[business_city]	STRING	Dimension	business_city city	HIGH
[business_state]	STRING	Dimension	business_state state	MEDIUM
[business_postal_code]	STRING	Dimension	"postal code" zip	LOW
[business_latitude]	DOUBLE	Dimension	"business latitude" lat	LOW

3. Click SAVE.

Specifying synonyms for a dataset in NLS

With Cloudera Data Visualization, you can specify synonyms for dataset fields; these words can be used in Natural Language Search (NLS).

Procedure

1. To specify alternate words for the dataset fields, you can add them to the column Matching search terms. Separate synonyms by including a single space, and use quotations for phrases.
2. Under Fields, add synonymous terms that can be searched through NLS.

   In the example below the terms 'business name', 'establishment', 'business', 'restaurant', and 'cafe' are added for the column ‘Business Name’.

   Fields

Field	Data Type	Dimension or Aggregate	Matching search terms	Auto Search Priority
[Record Count]	BIGINT	Aggregate		NONE
[business_id]	BIGINT	Dimension	business_id	NONE
[business_name]	STRING	Dimension	“business name” establishment business restaurant cafe	LOW

3. Click SAVE.

Specifying default aggregation field for a dataset in NLS

In Cloudera Data Visualization, you can specify which field's values report the default aggregation.

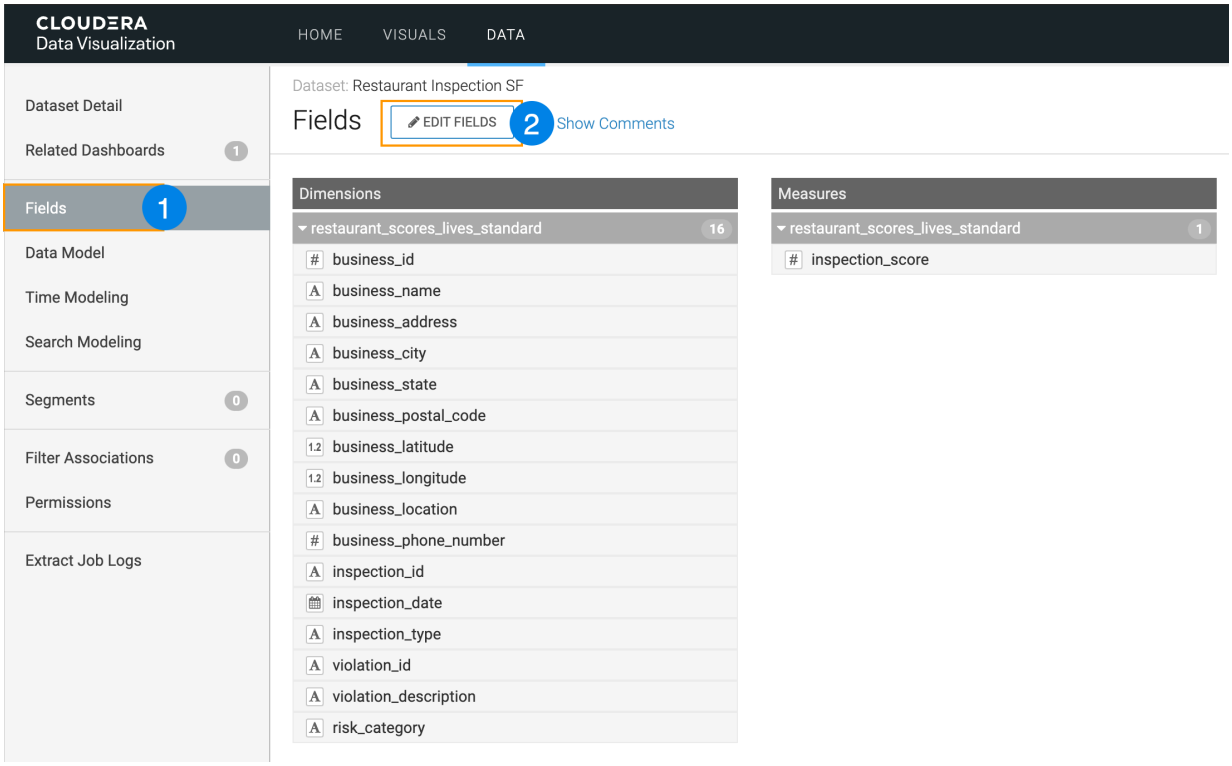
About this task

When a search phrase does not specify an aggregate measure, Cloudera Data Visualization reports results of a default column as an aggregate. In the majority of cases, this default column is Record Count, a default aggregation measurement generated by the system.

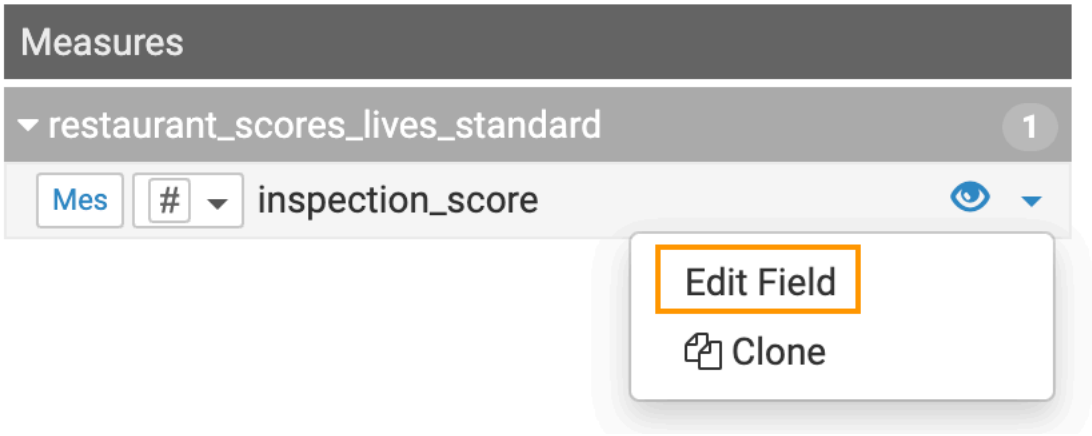
In this example, inspection_score is selected to illustrate this functionality.

Procedure

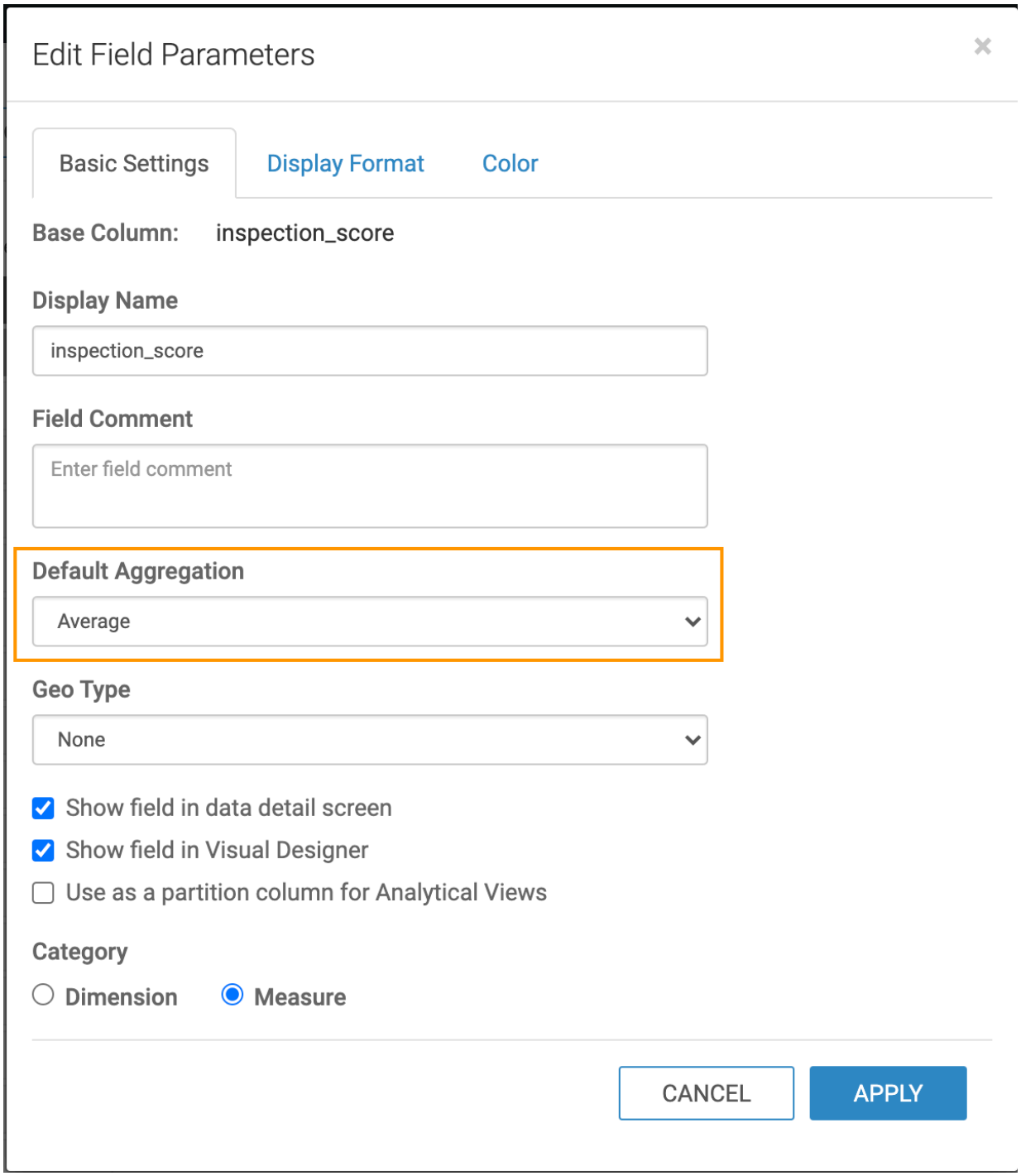
1. Navigate to the Fields page on the dataset page.
2. Click Edit Fields.



3. Click Edit Field for the field that you want to use as an aggregate.

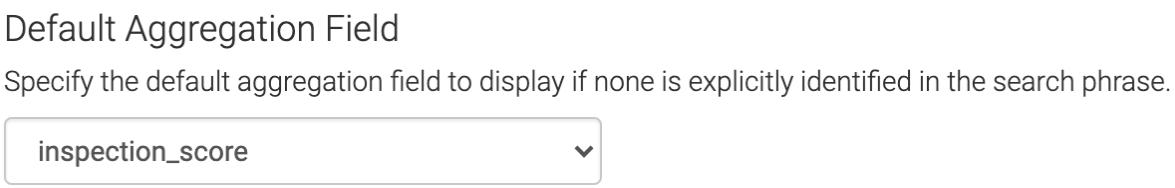


4. Choose a Default Aggregation type from the drop-down list.



5. Click APPLY.
6. Navigate to the Search Modeling page.

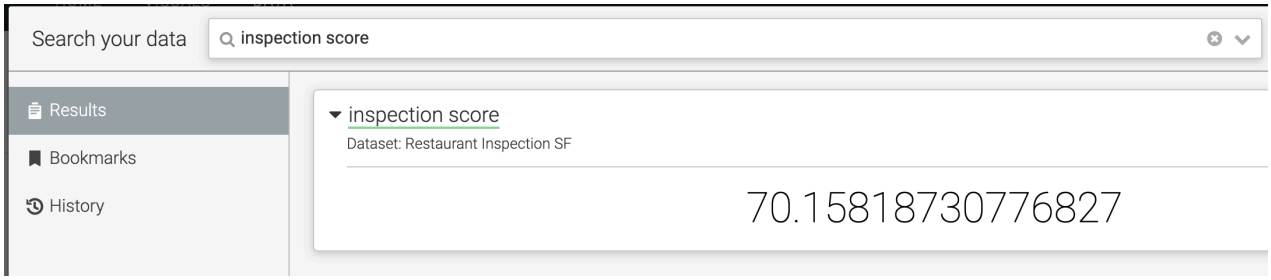
7. Select the Default Aggregation Field from the drop-down list.



8. Click SAVE.

Results

If you type 'inspection score' in the search bar you get an average score as a result.



Specifying suggested questions for a dataset in NLS

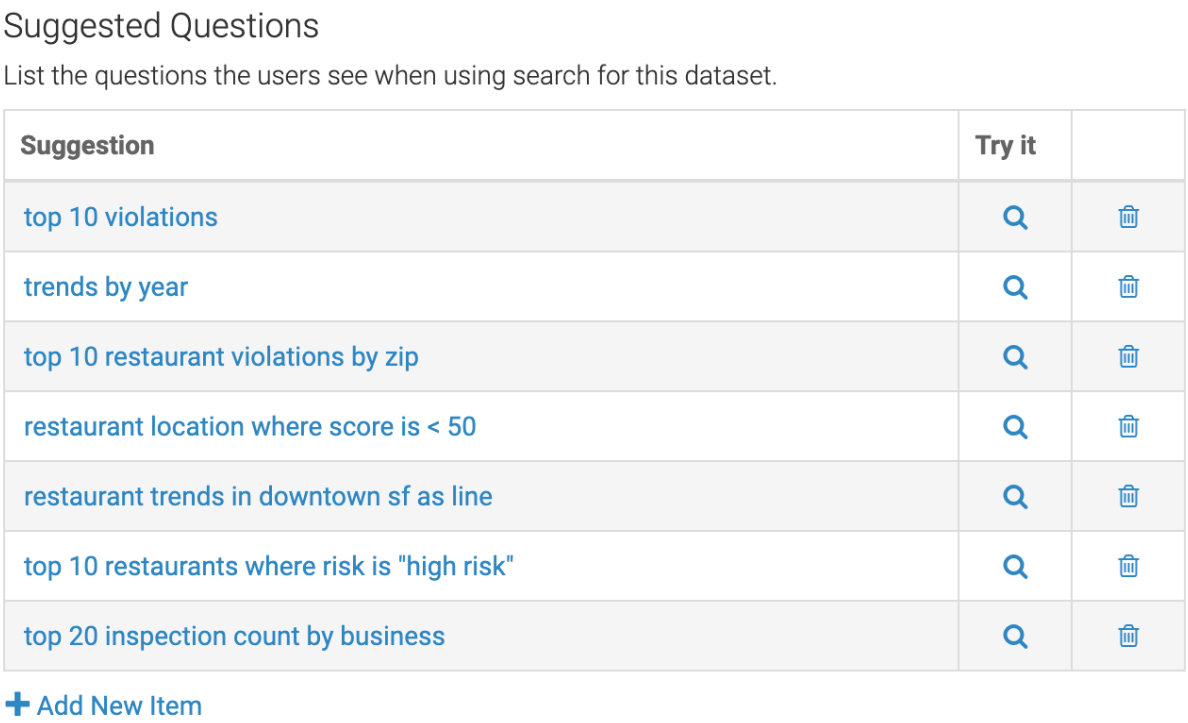
You can help users explore the dataset by pre-populating the dropdown with some search words and phrases in the dataset. You can also preview the results of the searches.

Procedure

1. Below Suggested Questions, click Add New Item.

2. In the text box, enter the search phrase.

   In this example, several suggestions have been added.



3. Click the Search icon to preview the visual results.

   A separate modal window appears, showing your results.
4. Click SAVE.

Note: To remove a search phrase, click the Trash icon next to it.

Specifying word substitutions for a dataset in NLS

Word substitutions enable you to connect common words with valid data specifications when using Natural Language Search (NLS) on Cloudera Data Visualization datasets.

About this task

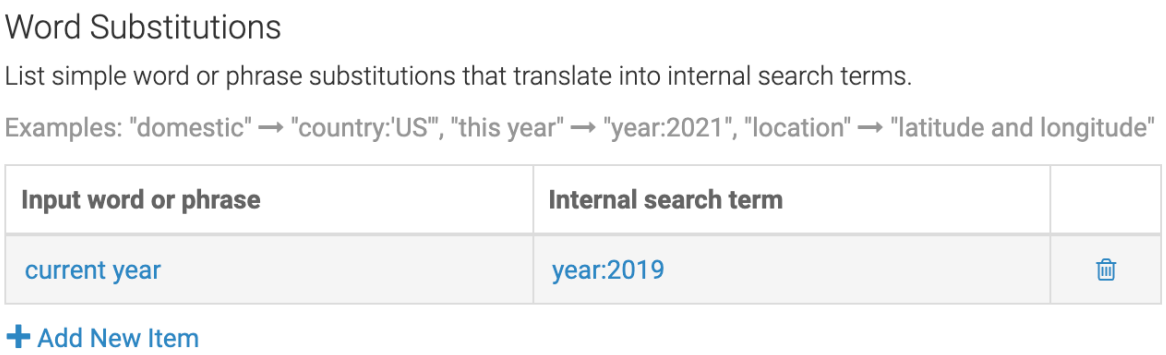
Common words and phrases have specific meaning in each dataset. In this example, you can see how to specify the meaning of the phrase 'current year' and the word 'domestic' into internal search terms.

Procedure

1. Under Word Substitutions, click Add New Item.
2. In the new row, add a common speech component in the Input word or phrase column.

3. In the same row, under Internal search term, specify the equivalent term in the dataset.

   In this example, the phrase 'current year' was added with the clause 'year:2019', setting the current year to 2019.

4. Click Add New Item again, if you want to add more word substitutions.

Configuring date and time for search requirements

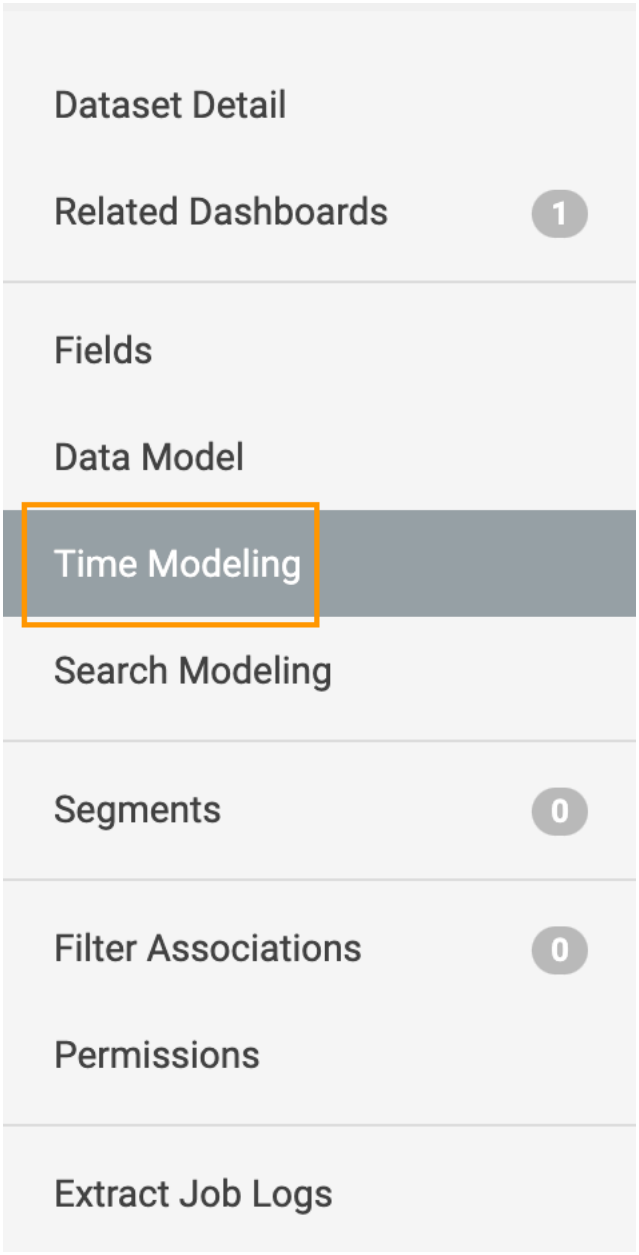
To configure date and time settings on datasets in Cloudera Data Visualization, follow the steps below.

Procedure

1. On the main navigation bar, click DATA.
2. In the list of datasets, click the one you want to enable for Natural Language Search (NLS).

   The Dataset Detail page appears.

3. On the left navigation menu of the Dataset Detail page, click Time Modeling.



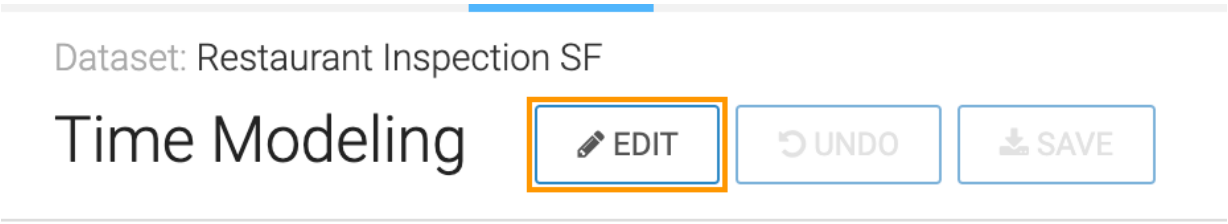
Specifying date/time field format for a dataset

About this task

On the Time Modeling page, specify the Date/Time Field of the dataset to ensure that Cloudera Data Visualization can correctly interpret and utilize the date/time information for search and analysis.

Procedure

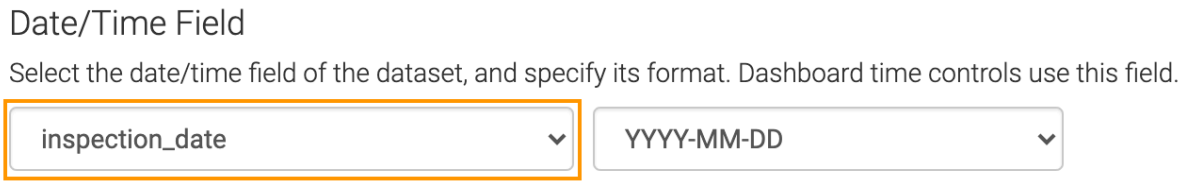
1. At the top of the page, click EDIT.



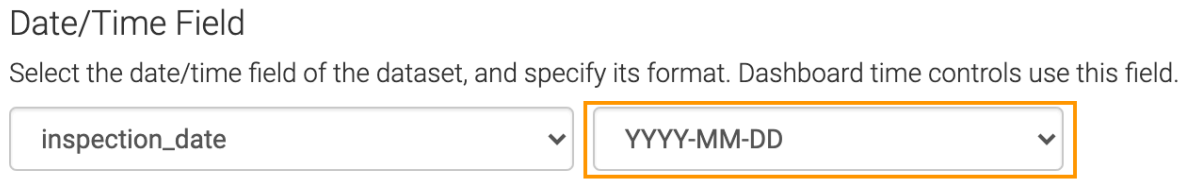
2. In the Date/Time Field section, select the name of the correct field that contains the date information from the first drop-down list.

   The system groups dataset fields in timestamp or other time formats at the top, under the heading Date Fields.

   In this example, the 'inspection_date' field has been selected.



3. In the Date/Time Field section, select the appropriate date format mask from the second drop-down list.



   Depending on your location, the masks may have a different format. For example, the two options in this example are YYYY-MM-DD and YYYY-MM-DD HH:mm:ss. In this example, the dataset does not contain the hour, minute, and second information, so the default format YYYY-MM-DD is kept.

   Note: If the dataset's date format differs from the options provided, you must create a computed column to parse the date into the desired format.

4. Click SAVE to apply your changes..

Specifying default time interval in dataset

About this task

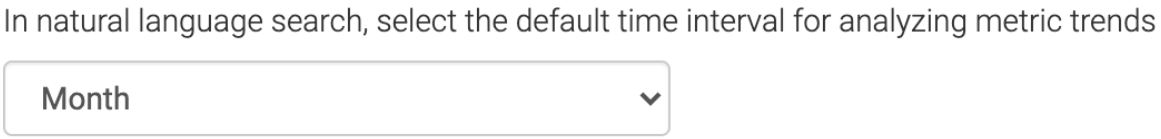
On the Time Modeling page, specify the default time interval for data aggregation and analysis to ensure that time-based filters, visualizations, and searches are applied consistently across the dataset, based on the selected interval.

Procedure

1. In the Date/Time Field section, click the third drop-down list for Default Interval.

2. From the options provided, select the appropriate interval for analyzing metric trends.

   Depending on your location, you may have different options. For example: Month, Quarter, or Year.

3. Click SAVE to confirm your selection.