

Natural Language Search References

Date published: 2020-10-30

Date modified: 2024-05-30

The Cloudera logo is displayed in a bold, orange, sans-serif font. The word "CLOUDERA" is written in all caps, with a stylized 'E' that has a horizontal bar extending to the right.

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

Icons for search types.....	4
NLS terms for basic filters.....	4
NLS terms for ranking.....	5
NLS terms for aggregation.....	6
NLS terms for sorting.....	6
NLS terms for comparing data points.....	7
NLS terms for visual types.....	7
NLS terms for dimension field matching.....	8
NLS terms for time series.....	8

Icons for search types

To identify the type of search, CDP Data Visualization displays the following icons in the Natural language Search (NLS) results.

Field name: string



Field name: integer



Field name: decimal



Field name: geo



Field name: timestamp



Field value



Field name: boolean



Suggested question



Suggested field



Bookmarked search



Historical search



NLS terms for basic filters

Data Visualization provides Natural Language Search (NLS) support for basic filter operations.

To enable search using the terms such as 'including', 'like', 'excluding', 'matching', and so on, refer to the following functions:

Function	Example
all	Count all employees with sales < 10000

Function	Example
between; and	Revenue between 0 and 1000 cereal_name by manufacturer where calories between 50 and 100
vs; versus	Revenue east versus west
greater than; >	Sum sale amount, by visitor and product, for last year sale amount > 2000 cereal_names where calories greater than 100
less than; <	Unique visitor count, by product and store, for sale amount < 20 cereal_name by manufacturer where calories less than 100
>=	Count calls by last name of the employee >= m cereal_name with calories >= 100
<=	Count shipments by latitude of the city <= 0
=	Unique visitor count by store purchased products = 3 for last 5 days
!=	Sum of sale amount by region != Canada and region != Mexico
begins with	Product name begins with "pro"
contains	Product name contains "alpine" and description contains "snow shoe"
ends with	Product name ends with "deluxe"
not begins with	Product name not begins with "tom's"
not contains	Product color not contains "tan" and product color not contains "red"
not ends with	Product name not ends with "trial"
similar to	Course name similar to "hand"
not similar to	Course name not similar to "hand"

NLS terms for ranking

Data Visualization provides Natural Language Search (NLS) support for ranking operations.

The ranking operations are the following functions:

Function	Example
top	Top sales representative with sales > 10000 Sales representative by average revenue of top regions
bottom	Bottom average revenue

Function	Example
	Bottom revenue by state Customer by revenue, for each bottom sales representative
n (Top Numbers)	Top 10 sales representative by revenue
n (Bottom Numbers)	Bottom 25 customers by revenue, for each sales representative

NLS terms for aggregation

Data Visualization provides Natural Language Search (NLS) support for aggregation concepts.

The aggregate support includes the following functions:

Function	Example
sum	Sum revenue
average	average revenue by store
count	Visitor count by site
max	maximum sales by visitor and by site
min	minimum revenue by store, by campaign, for cost > 5000
standard deviation	standard deviation revenue by product, by month, for date after 10/31/2010
unique count	unique count visitors last week, by product page
variance	variance sale amount last year, by visitor and by product

NLS terms for sorting

Data Visualization provides Natural Language Search (NLS) support for sorting.

You can sort the search results by using the following functions:

Function	Example
order by	Viewership by channel order by viewership descending Viewership by channel order by channel
sort by	Sales order_date sort by order_date ascending

NLS terms for comparing data points

Data Visualization provides Natural Language Search (NLS) support for comparing data points.

The user can request search results that compare two values, such as actual versus target.

Function	Example
versus; vs. (Metric Comparison)	Compare sales 'vs.' headcount
versus; vs. (Field Value Comparison)	Compare sales for region A 'vs.' region B
versus; vs. (Time Value Comparison)	Compare sales this week 'vs.' sales same week last year
versus; vs. (Contribution, or Window, Comparison)	Compare sales for region A 'vs.' all regions

NLS terms for visual types

In Data Visualization, you can specify a visual type that displays the results of Natural Language Search (NLS).

The support for visual types includes the following function:

Function	Example
as	Revenue and quantity sold by product name as scatter

The Natural Language Search supports the following visual types:

Visual Type	Keyword
Table	table, data
Cross Tabulation	crosstab
Bar	bar
Lines	line
Combo	combo, combined
Areas	area
Grouped Bars	grouped bar, groupedbar
KPI	kpi
Gauge	gauge
Bullet	bullet

Visual Type	Keyword
Packed Bubbles	packed bubble, bubble
Scatter	scatter
Pie	pie
Radial	radial
Chord	chord
Network	network
Dendogram	dendogram
Treemap	tree map, treemap
Correlation Heatmap	correlation heatmap, heatmap
Map	map
Interactive Map	interactive map, leaflet
Sparklines	sparkline

NLS terms for dimension field matching

Data Visualization provides Natural Language Search (NLS) support for matching a field in the dimension shelf.

The user can request search results to match a field (base or derived), in the Dimension shelf, defined on any search-enabled dataset.

Function	Example
net	Net Profit
by	By Customer
for	For Products

NLS terms for time series

Data Visualization provides Natural Language Search (NLS) support for time series concepts.

When the user enters a time-series term in the search box, such as Trends or Over Time, Data Visualization automatically recognizes that the user is requesting a time-series based visual, and renders a visual with time field on the X-axis.

Function	Example
trends	Sales 'trends'
over time	headcount 'over time'