Cloudera Runtime 7.1.1

Atlas Glossaries

Date published: 2019-09-23
Date modified: 2020-05-30

https://docs.cloudera.com/

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

Glossaries overview.....4

Creating glossaries.....4

Creating terms.....5

Associating terms with entities.....6

Defining related terms.....7

Creating categories.....8

Assigning terms to categories.....8

Searching using terms.....8

Importing Glossary terms in bulk.....9

Glossaries overview

Use glossaries to define a common set of search terms that data users across your organization use to describe their data.

“Data” can describe a wide variety of content: lists of names or text or columns full of numbers. You can use algorithms to describe data as having a specific pattern, of being within a range or having wide variation, but what’s missing from these descriptions is what does the data mean in a given business context and what is it used for? Is this column of integers the count of pallets that entered a warehouse on a given day or number of visitors for each room in a conference center? An Atlas glossary is a way to organize the context information that your business uses to make sense of your data beyond what can be figured out just by looking at the content. The glossary holds the terms you’ve agreed upon across your organization so business users can use familiar terms to find what they are looking for.

Apache Atlas glossaries enable you to define a hierarchical set of business terms that represents your business domain. You can then associate these glossary terms with the metadata entities that Atlas manages.

Glossaries

You can create any number of separate glossaries to group terms used in different domains. Define glossaries to correspond to how your users think about their data. For example, you might want to put terms specific to Sales in one glossary, while your Manufacturing terms might go in a separate glossary. Terms from multiple glossaries can still be associated with the same data; the same term (“Shipping Address City”) can appear in more than one glossary.

Terms

Glossary terms can be thought of as of a flat (but searchable) list of business terms organized by glossaries. Unlike classifications, terms are not propagated through lineage relationships: the context of the term is what’s important, so propagation may or may not make sense.

Categories

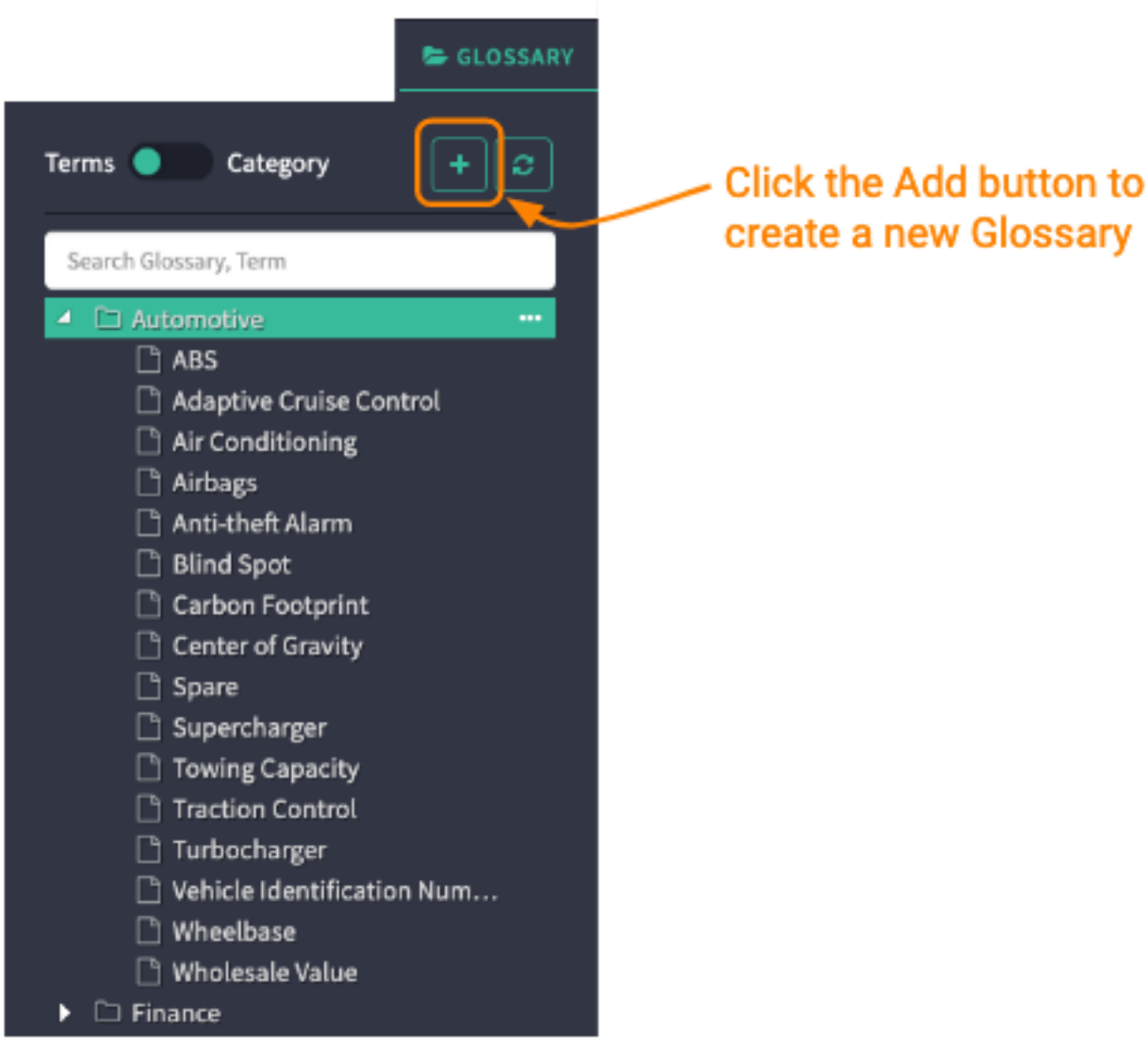
Glossary categories enable you to create a hierarchy of categories and subcategories within each glossary. You can then associate terms with these categories, thereby establishing an organizational hierarchy of business terms.

Creating glossaries

Create a new glossary when you need to create a new context for terms.

Typically you would want to create new glossaries when terms are used by significantly different audiences or when the same term has significantly different meanings depending on the context. You can create any number of glossaries; glossary names can include letters, numbers, spaces, and underscores.

Create a new glossary from the Glossary tab in the left navigation pane:

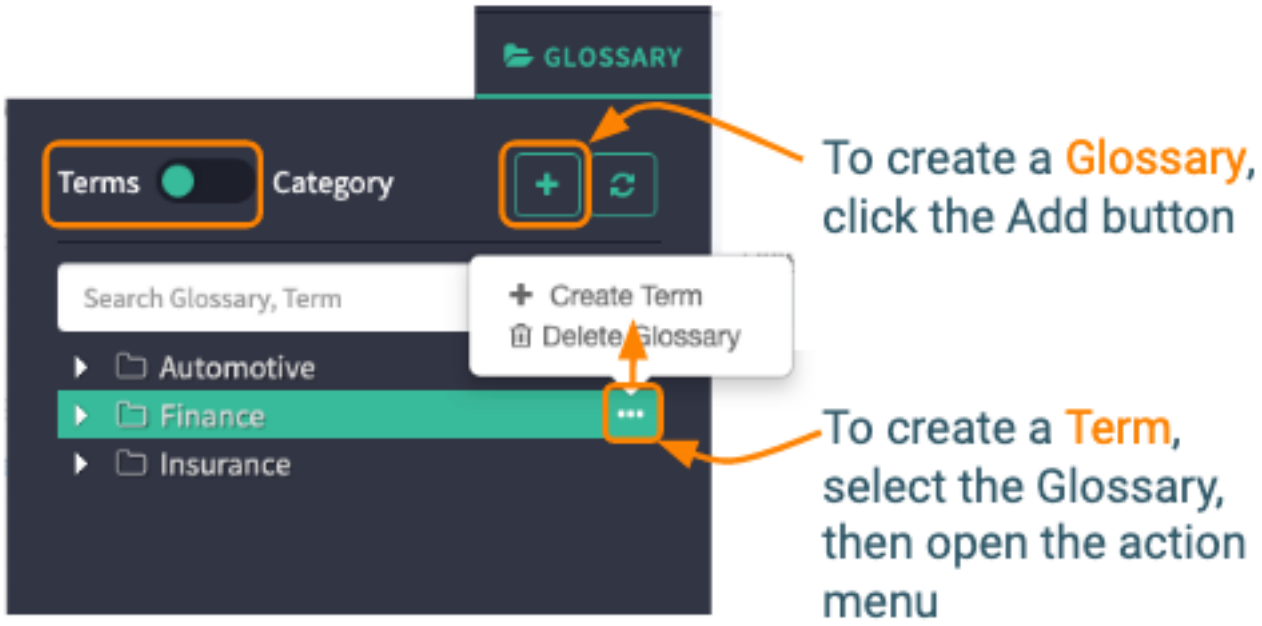


Creating terms

After you've created a glossary, you can add terms to it.

Glossary terms describe business concepts or information.

To create a glossary term, go to the Glossary tab in the left navigation panel. With the toggle switch set to "Terms", click the plus button.



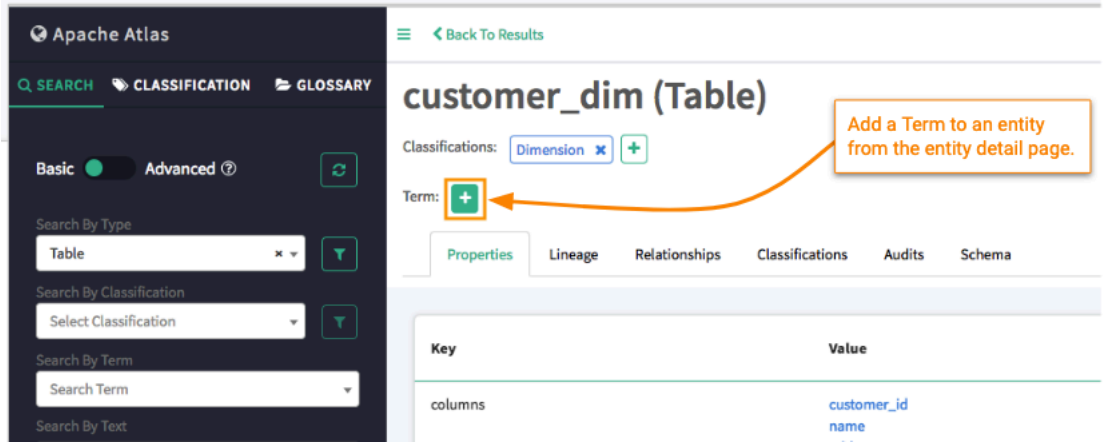
Associating terms with entities

The value of terms comes with using them to identify data and data sets. You can associate terms with entities from search results and within an entity detail page.

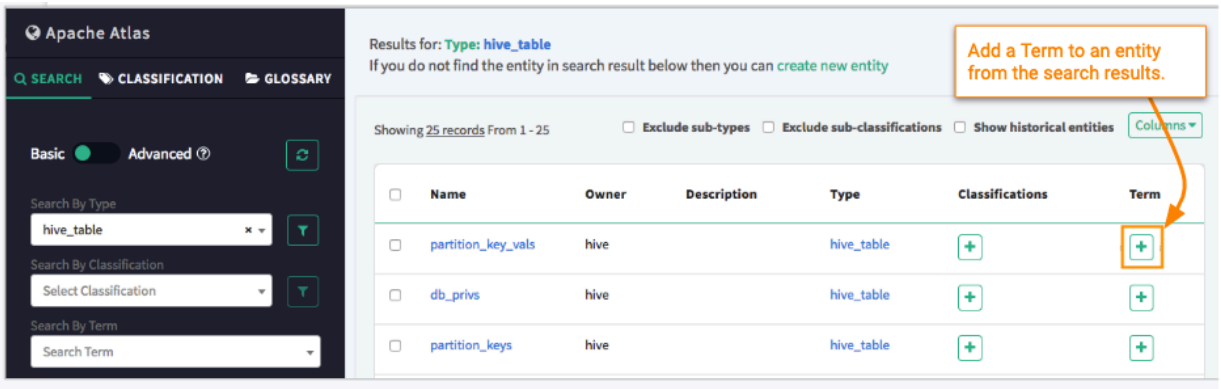
When you associate a term with an entity, that entity is part of the group of entities returned when you search on the term.

You can associate a term with an entity in the dashboard:

- From an entity detail page:



- From search results:



Both methods open a dialog box where you can search for the term or browse through the glossary hierarchy.

Related Information

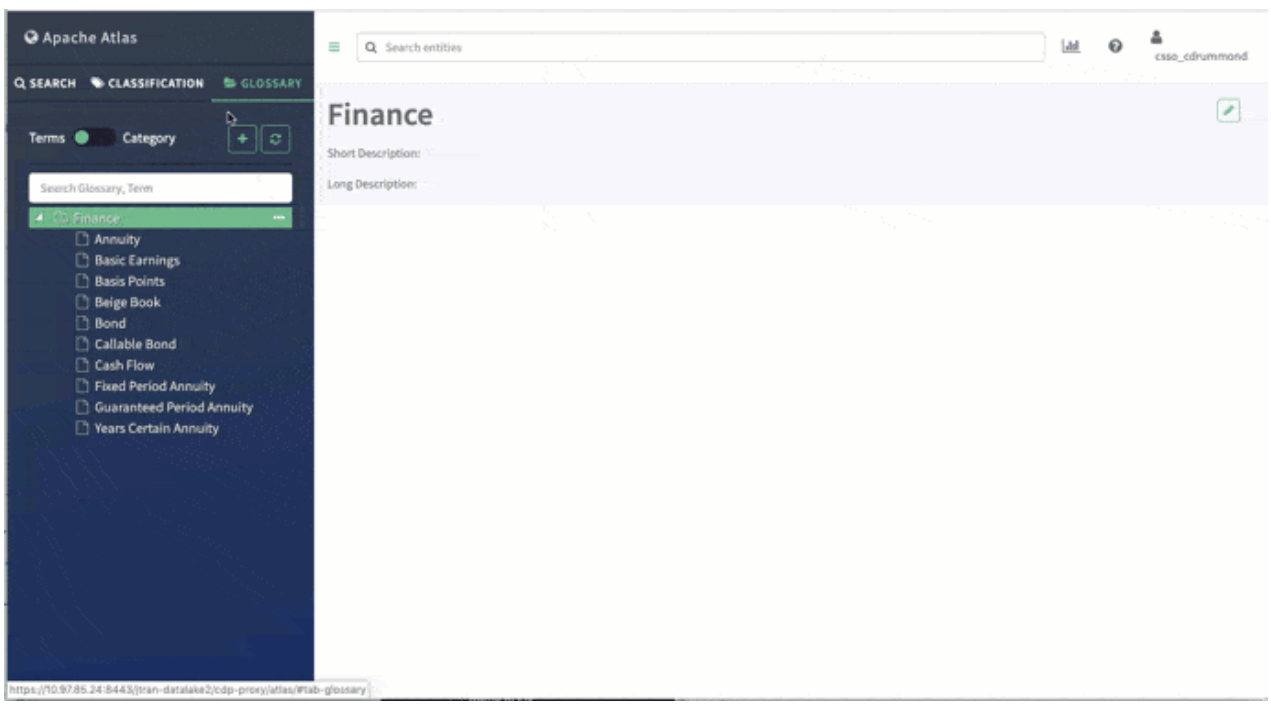
Searching using terms

Defining related terms

Can't agree on a single term among multiple organizations? Define two terms as synonyms so you can use either term to search for entities. Atlas provides many types of relationships to help data stewards build useful glossaries.

The glossary allows you to indicate relationships among terms. When you indicate a relationship, it is automatically added to the definition of the related term.

To create a relationship between two glossary terms, go to the detail page for one term, open the Related Terms tab, and add a term to the appropriate relationship type.



The relationship types are as follows:

Antonym

Glossary terms that have the opposite (or near opposite) meaning, in the same language.

Classifies

The current term is a generic name for the indicated more specific term or terms. "Has Types"

isA

The current term is an example of or type of the identified term. "Is A Type Of"

Preferred Terms

The identified term is the preferred synonym and should be used instead of the current term. "Preferred Synonym"

Preferred To Terms

The current term is the preferred term among a group of synonyms.

Replaced By

The current term is replaced by the indicated term.

Replacement Terms

The indicated term or terms replace the current term.

See Also

The indicated term or terms provide additional information related to the current term.

Synonyms

A relationship that indicates that the current term is the same or similar to the indicated term.

Translated Term

The current term is translated from the indicated term.

Translation Term

The current term is translated into one or more languages in the indicated terms.

Valid Values

Values that are meaningful for this term, such as a list of types that could apply to the term.

Valid Values For

The current term is a valid value for the indicated term.

Creating categories

Categories form the layers in a hierarchy so you can organize glossary terms to make them intuitive to find.

Assigning terms to categories

Group terms into categories to create meaningful contexts in a glossary.

Searching using terms

Now that you have terms defined and associated with entities, you have a search index that lets you quickly find data assets based on their value in your organization.

Importing Glossary terms in bulk

You can import a comma-separated list of Glossary terms into Atlas.

Create the Glossary or Glossaries before you begin the import. Then, download a template CSV file to organize your term list so Atlas can consume it.

Note: The mechanism to import Glossary terms can be used to import new terms only. If you want to update existing terms, use the Atlas UI or API.

To import terms:

1. Log into Atlas.
2. From the user menu, chose Bulk Import Glossary Download Import template to download a template for the bulk import file.
3. Fill in the template with the Glossary term details.

   Consider the following guidelines for the import file:

   - Each line in the file produces a term in a pre-defined Glossary.
   - Only the GlossaryName and TermName are required.
   - Only new terms can be added; if your list contains an existing term, it will fail to import any terms.
   - If you provide other metadata for a term, be sure to include the correct separators so the value is applied to the correct relationship type.
4. When you have a term list to import, from the user menu, chose Bulk Import Glossary Import Business Metadata and select the file to upload.

   If a row in the input file does not import successfully, you will see an error in the UI; additional errors in the same import are not shown. Rows that are correct import even if you see an error for a row that does not import. If you see an error, fix the error, remove the rows previous to the error row, and reimport the file. Note that if you provide a single entry for an array, the new value replaces all the values of the array. To see all the errors for a given import, look in the Atlas Server log.

Related Information

Defining related terms

Accessing Atlas logs