

Cloudera Runtime 7.0.1

Atlas Classifications

Date published: 2019-09-23

Date modified:

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

Working with Atlas classifications.....	4
Creating classifications.....	5
Adding attributes to classifications.....	5
Associating classifications with entities.....	6
Propagating classifications through lineage.....	8
Searching for entities using classifications.....	10

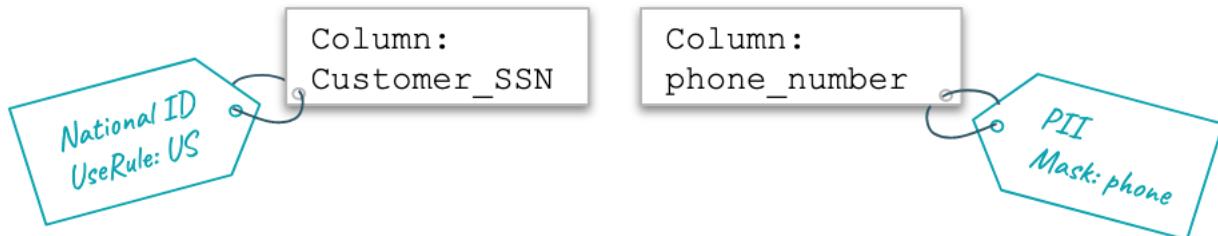
Working with Atlas classifications

Add metadata labels to Atlas entities using classifications

In Atlas, classifications are labels that can be assigned to entities. The flexibility of classifications makes them useful for many applications. For example, you can define classifications to describe the phases in your data prep processes and assign the classifications to specific assets to mark where they are in the process. You can define classifications to identify data to block or mask and use the classifications in access control policies in Ranger.



Classifications can be simple labels but they can also be defined with attributes that allow you to assign values to further describe the entity where the classification is assigned. A typical use for attributes would be to refine the meaning of a general category. Data assets identified with a classification of “PII” or Personally Identifiable Information can have classification attributes that indicate the nature of the information to drive data masking or expiration policies. Columns tagged with “PII” might be further separated into phone numbers, credit card numbers, and “other” where a Ranger policy identifies masks for the phone and credit card numbers and simply blocks columns tagged as “PII” with attribute “Other”.



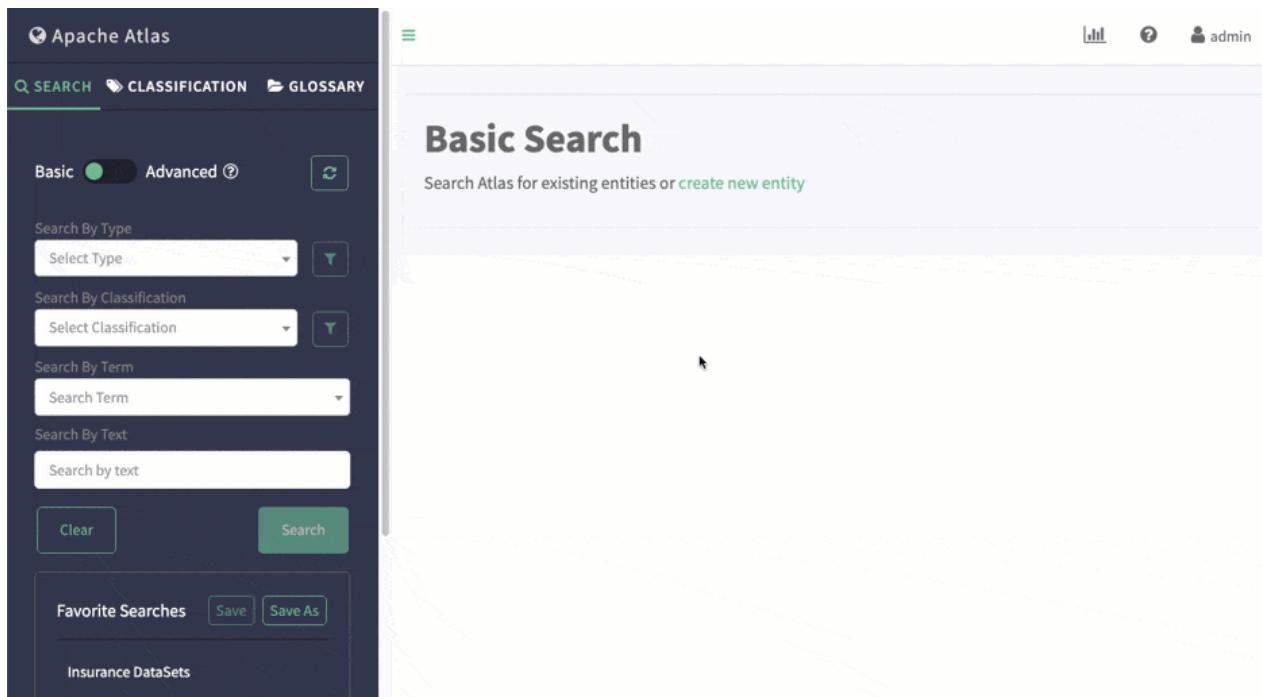
Here are some questions to ask yourself about your classification choices:

- What entities do you expect the classification to apply to? If you mean them only to apply to table columns, make sure that the name and description helps data stewards use the classification correctly.
- Will the classification be used for Atlas searches? Consider including attributes to allow further refinement of search results. If your Atlas users are more likely to search using terms, you might consider connecting the classification to a glossary term so searches from either mechanism return the correct results. (Assigning the term to an entity automatically assigns the classification to the same entity)
- Do you want the classification to follow lineage relationships and be assigned automatically to data that is created from the entities assigned this classification? If not, you can use the classification description to help data stewards understand that they should turn off propagation when assigning this classification to entities or terms.

Creating classifications

Create classifications from the tab in the left navigation bar.

Create classifications from the Classifications tab in the left navigation pane.



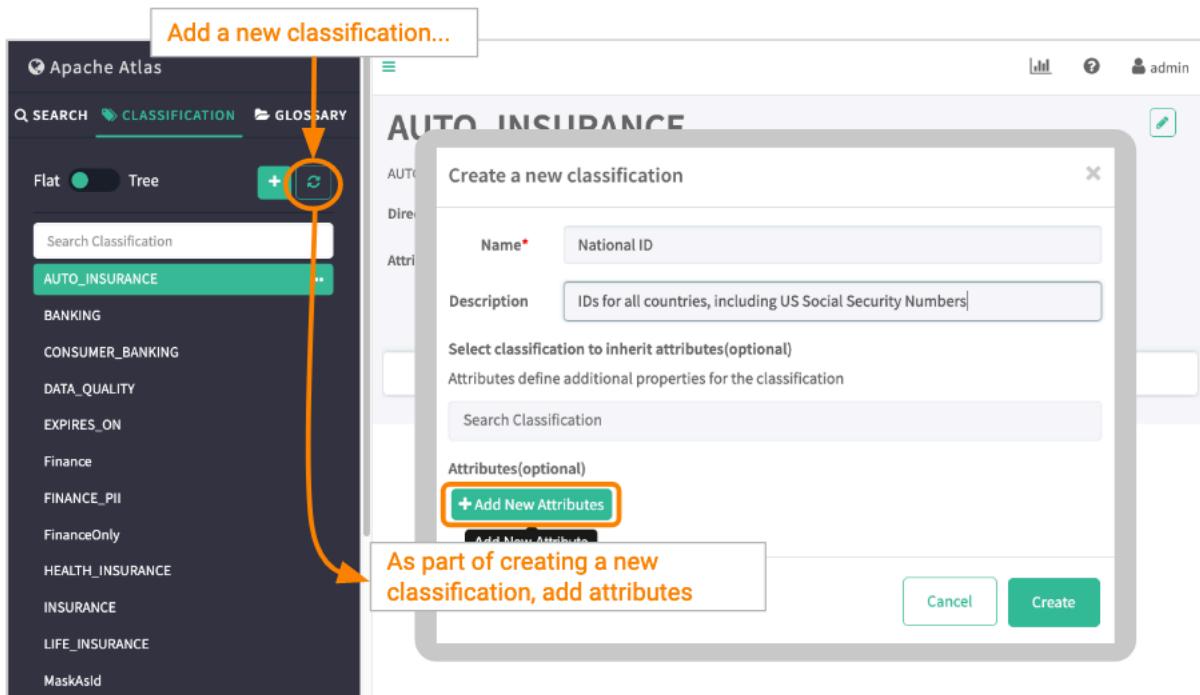
Names can be 255 characters and contain ASCII characters and spaces. Descriptions can be 1024 characters and contain any characters.

When creating a classification, you can choose an existing classification to reuse its attribute definitions.

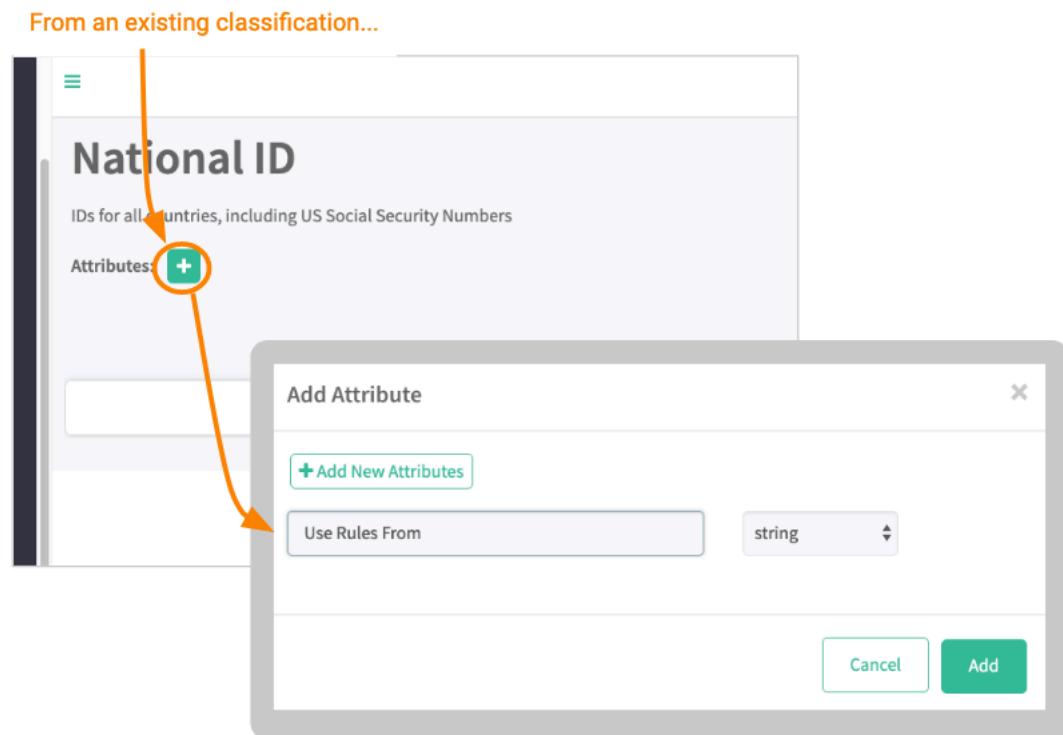
Adding attributes to classifications

Attributes are key-value pairs defined for a classification that you can set to a value when you assign the classification to an entity.

For a new classification, you can add attributes as part of the classification definition.



You can also add attributes to an existing classification. Select the classification in the list, then click the plus button next to Attributes.



Associating classifications with entities

Assign classifications to entities in the entity detail page or in the search results.

There are two ways to associate a classification with an entity:

- Add the classification to the entity in the entity detail page.

The screenshot shows the entity detail page for 'ww_customers (hive_table)'. At the top left is a blue circular icon with a yellow elephant logo. To its right is the entity name. Below the name is a section labeled 'Classifications:' with a green '+' button. Underneath is a section labeled 'Term:' with a green '+' button. A large orange arrow points from a callout box containing the text 'Click + to add a classification' towards the green '+' button in the 'Classifications:' section.

- Add the classification to the entity in the search results.

The screenshot shows the search results page with the query 'Type: hive_table AND Query: where db.name="hortoniabank"'. It displays three records: 'eu_countries', 'ww_customers', and 'us_customers', all of which are 'hive_table' type. In the 'Classifications' column for each row, there is a green '+' button. An orange arrow points from a callout box containing the text 'Click + to add a classification' towards one of these green '+' buttons.

When you associate a classification with an entity, you can set the classification's behavior with these controls:

The screenshot shows the 'Add Classification' dialog box. It has a title bar 'Add Classification' with a close button. Below is a dropdown menu with the placeholder '-- Select a Classification from the dropdown list --'. At the bottom are three checkboxes: 'Propagate' (checked), 'Apply Validity Period' (unchecked), and 'Remove propagation on entity delete' (unchecked). At the very bottom are 'Cancel' and 'Add' buttons.

- Propagate: Any entities in the impact chain of this entity will automatically be tagged with this classification. The classification can't be removed from the "downstream" entities without being removed from this source entity. See more at Propagate classifications to derived entities.
- Apply validity period: You can set a time period after which this classification is removed from the entity. Set the start time or end time or both. If you do not choose a time zone, the local time of the Atlas server is used.
- Remove propagation on entity delete: When an entity is deleted in the source and marked as deleted in Atlas, you can choose to maintain the propagation path, or to break propagation at the deleted entity.

You can change the behavior of classifications after they are assigned: go to the Classifications tab for an entity.

You can remove classifications from an entity using the X button on search results or entity detail page. Don't see an X on the classification name? That classification was assigned through propagation; to remove it, you have to turn off propagation for the source entity or term.

Related Information

[Propagating classifications through lineage](#)

Propagating classifications through lineage

Classifications on table columns can automatically apply to other tables where the same data is found when the data transformations are captured in lineage.

Viewing propagated classifications

When you look at the classifications assign to an entity, you can see which classifications were originally assigned to this entity and which were propagated here through lineage. From the Classifications tab, you can:

- Show or hide propagated classifications (Show Propagated Classifications button)
- Navigate to the original entity where a classification was assigned (Propagated From button)

Navigate to the source of the classification

claims_view (hive_table)

Classifications: [+](#)

Term: [Claim](#) [X](#) [+](#)

Propagated Classifications: [DATA_QUALITY](#) [INSURANCE](#)

Properties Lineage Relationships **Classifications** Audits Schema

Show 1 - 2

Classification Attributes Action

Classification	Attributes	Action
DATA_QUALITY Propagated From	Name Value	
	score 0.51	
INSURANCE Propagated From	NA	

Toggle the view of classifications associated with this entity by propagation

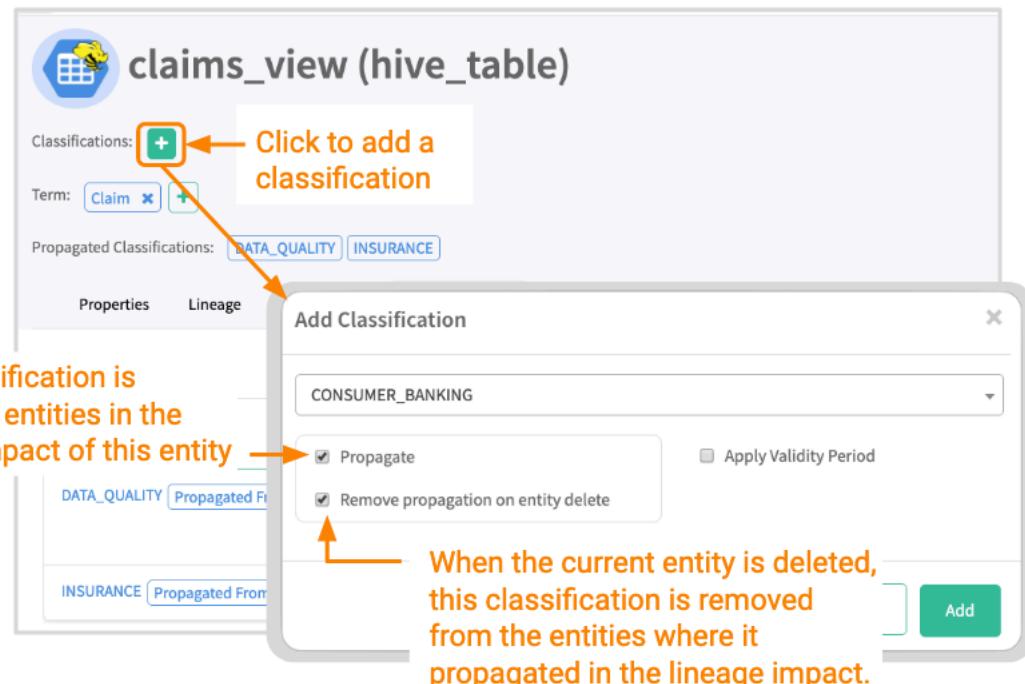
Show Propagated Classifications

Click to go to the entity that was first marked with this classification

Setting propagation behavior

You can set propagation for a classification when you create the classification or in the lineage for an entity where the classification is assigned:

- Set propagation at the time when the classification is associated with an entity or term.



- Set propagation at a specific point in the lineage.

The screenshot shows the Cloudera Runtime Atlas interface for managing entity lineage. At the top, there's a navigation bar with tabs for Properties, Lineage, Relationships, Classifications, Audits, and Schema. The Lineage tab is selected. Below the navigation bar, there's a classification section with a 'DATA_QUALITY' button and a '+' icon. A 'Term:' input field with a '+' icon is also present. The main area displays a lineage graph for the entity 'claim_savings (hive_table)'. The graph shows nodes for 'create external t...', 'claim_savings', 'create view if no...', and 'claims_view', connected by arrows indicating the flow of lineage. A callout bubble with orange text says 'Click on a lineage impact line' with an arrow pointing to one of the impact lines in the graph. A modal window titled 'Enable/Disable Propagation' is open, containing a toggle switch labeled 'Enable/Disable Propagation' and a button 'Select Classifications to Block Propagation'. It lists lineage steps: 'claim_savings → create view if not from cost_savings.claim_savings' and 'hive_table → hive_process'. A radio button for 'hive_table → hive_process' is selected. Another callout bubble with orange text says 'Turn off propagation at this point in the lineage impact: no classifications assigned to the entity on the left will propagate to entities on the right.' An arrow points from this text to the 'hive_table → hive_process' step. At the bottom of the modal are 'Cancel' and 'Update' buttons. Below the modal, another smaller 'Enable/Disable Propagation' window is shown with a table:

Classification	Entity Name	Block Propagation
DATA_QUALITY	claim_savings (hive_table)	<input checked="" type="checkbox"/>

A callout bubble with orange text says 'Choose the classifications to propagate to entities in the lineage impact.' with an arrow pointing to the 'Block Propagation' checkbox. To the right of the table is an 'Update' button.

Related Information

- Associating classifications with entities
- Controlling data access using tags

Searching for entities using classifications

After you tag entities with classifications, it's very easy to search for the entities.

The Atlas dashboard makes it easy to find entities that you've tagged with classifications: in the left navigation pane, go to the Classification tab and select the classification you want. You can refine your search results by adding criteria on the Search tab.

The screenshot shows the Apache Atlas search interface. On the left, there's a sidebar with a 'Basic' toggle switch (which is turned on), an 'Advanced' link, and a refresh button. Below these are four dropdown menus: 'Search By Type' (with 'Select Type'), 'Search By Classification' (with 'Select Classification'), 'Search By Term' (with 'Search Term'), and 'Search By Text' (with a text input field). At the bottom of the sidebar are 'Clear' and 'Search' buttons, and sections for 'Favorite Searches' (with 'Save' and 'Save As' buttons) and 'Insurance DataSets'. The main area is titled 'Basic Search' and contains the instruction 'Search Atlas for existing entities or [create new entity](#)'. A large, semi-transparent watermark of a map of the world is overlaid on the right side of the main search area.