

Cloudera Runtime 7.0.3

Atlas Classifications

Date published: 2019-11-22

Date modified:

CLOUdera

<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.

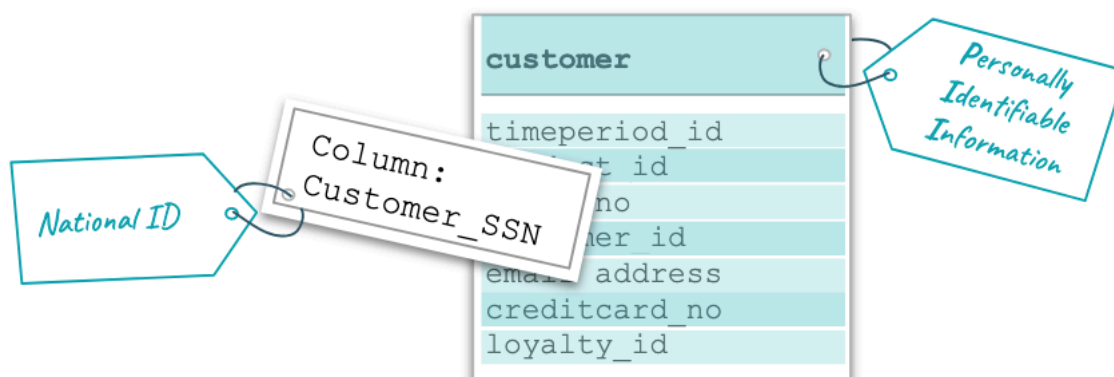
Contents

Working with Atlas classifications and labels.....	4
Creating classifications.....	5
Creating labels.....	6
Adding attributes to classifications.....	7
Associating classifications with entities.....	8
Propagating classifications through lineage.....	10
Searching for entities using classifications.....	12

Working with Atlas classifications and labels

Add metadata to Atlas entities using labels and classifications.

You can add metadata to Atlas entities to help your organization find, organize, and share your understanding of the data assets that drive business processes. Atlas provides two mechanisms for adding metadata to entities: labels and classifications. Both labels and classifications can be applied to entities to help describe the entity's content, status, or other business-driven value.



Labels

Labels are words or phrases (strings) that you can associate with an entity and reuse for other entities. They are a light-weight way to add information to an entity so you can find it easily and share your knowledge about the entity with others.

Anyone can create labels and associate labels with entities.

Classifications

Classifications are strings like labels, with added complexity and structure:

- Atlas includes precise search tools for finding entities using classifications.
- Classifications can automatically propagate to additional entities through lineage relationships.
- You can use classifications to drive access control policies in Ranger.
- You can enrich a classification with attributes in the form of key-value pairs and set the value to describe a particular entity.

Administrators can control who can create and assign classifications through Atlas policies in Ranger.

In short, use labels for annotating entities; use classifications to involve entities in processes inside and outside Atlas.

More about classifications: attributes

You can add key-value pairs, or *attributes* to the definition of a classification. A typical use for classification 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.” A Ranger policy based on the classification can use the attribute values to identify masks for the phone and credit card numbers and to block columns tagged as “PII” with attribute “Other.”



Planning for classifications and labels

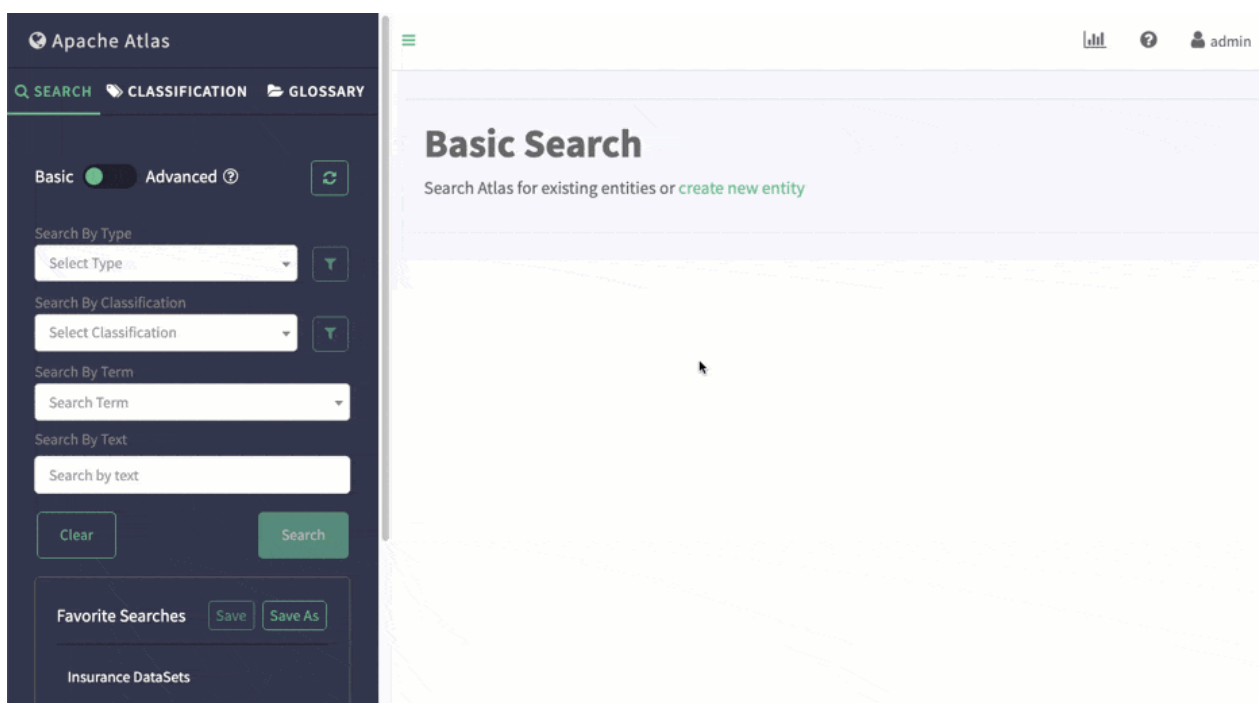
Here are some questions to help you think about how you define classifications and labels in your system:

- Will you use the metadata to drive a workflow outside of Atlas? If so, use a classification. If not, a label may work fine.
- Do you need the text of the metadata to be exact values? You can limit who has the ability to create classifications: you can manage a single list for an organization. The ability to apply a classification to an entity can be controlled separately from defining classifications, which lets you potentially allow more users to applying the "official" classifications. Any users can create and add labels to entities, so it may be more difficult to standardize the content of labels.
- What types of entities do you expect the metadata to apply to? Labels can be added to any entity type; classifications can be defined to apply to a specific entity type. If you define a classification to only to apply to certain entity types, such as table columns, make sure that the name and description helps data stewards use the classification correctly.
- Are you adding metadata to make Atlas searches easier? Consider using classifications with attributes to allow 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.) For both labels and classifications, consider setting guidelines so everyone creating metadata uses the same conventions: underscores or hyphens? detail first or last? table-level metadata or column-level?
- Do you want a classification to follow lineage relationships and be assigned automatically as data is used in other entities? 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. Labels do not propagate.

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.

Creating labels

Create and assign a label from the detail page for an entity.

From the detail page for an entity, the Properties tab includes a section for labels:

Results for: **Type: hive_table**
If you do not find the entity in search result below then you can [create new entity](#)

Showing 4 records From 1 - 25 Exclude sub-classifications Show historical entities Exclude sub-types Columns ▾

<input type="checkbox"/>	Name ↕	Owner ↕	Description ↕	Type ↕	Classifications	Term
<input type="checkbox"/>	custom...	admin		hive_table	<input data-bbox="1081 611 1117 653" type="button" value="+"/>	<input data-bbox="1268 611 1304 653" type="button" value="+"/>
<input type="checkbox"/>	web_logs	admin		hive_table	<input data-bbox="1081 688 1117 730" type="button" value="+"/>	<input data-bbox="1268 688 1304 730" type="button" value="+"/>
<input type="checkbox"/>	sample...	admin		hive_table	<input data-bbox="1081 766 1117 808" type="button" value="+"/>	<input data-bbox="1268 766 1304 808" type="button" value="+"/>
<input type="checkbox"/>	sample...	admin		hive_table	<input data-bbox="1081 844 1117 886" type="button" value="+"/>	<input data-bbox="1268 844 1304 886" type="button" value="+"/>

Labels can be 50 characters and contain ASCII characters, hyphens (-) and underscores (_). You can search for entities using the text from labels assigned to the entities.

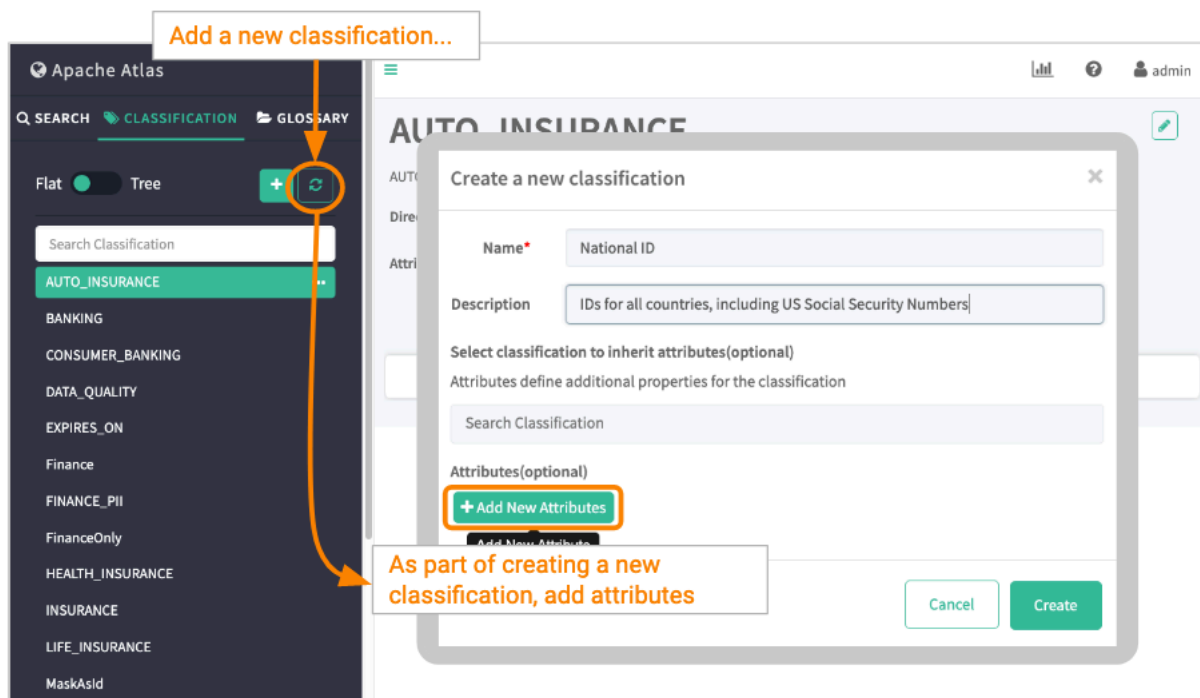
Related Information

[Using Free-text Search](#)

Adding attributes to classifications

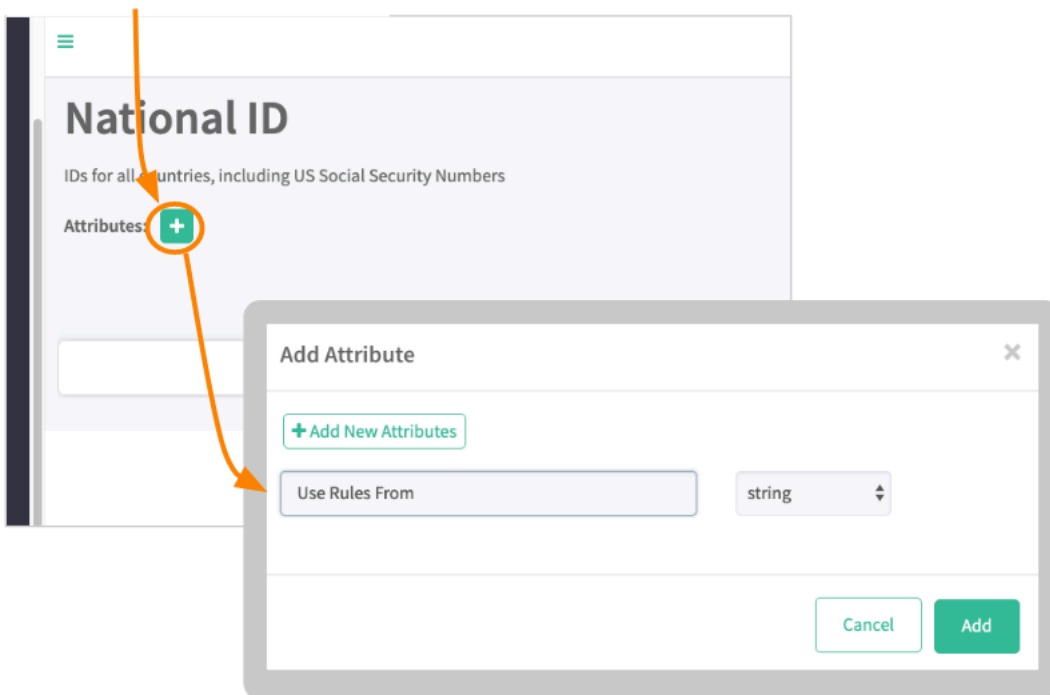
Attributes are key-value pairs defined for a classification that you can set to a value when you assign the classification to an entity

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.

From an existing classification...

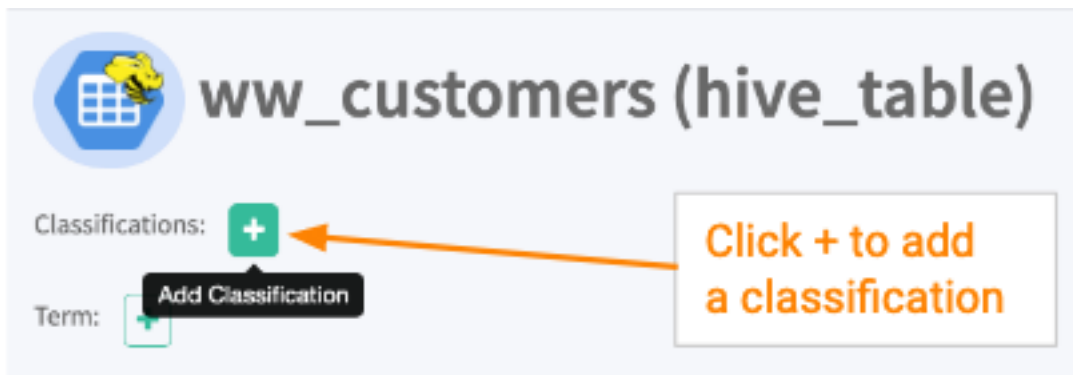


Associating classifications with entities

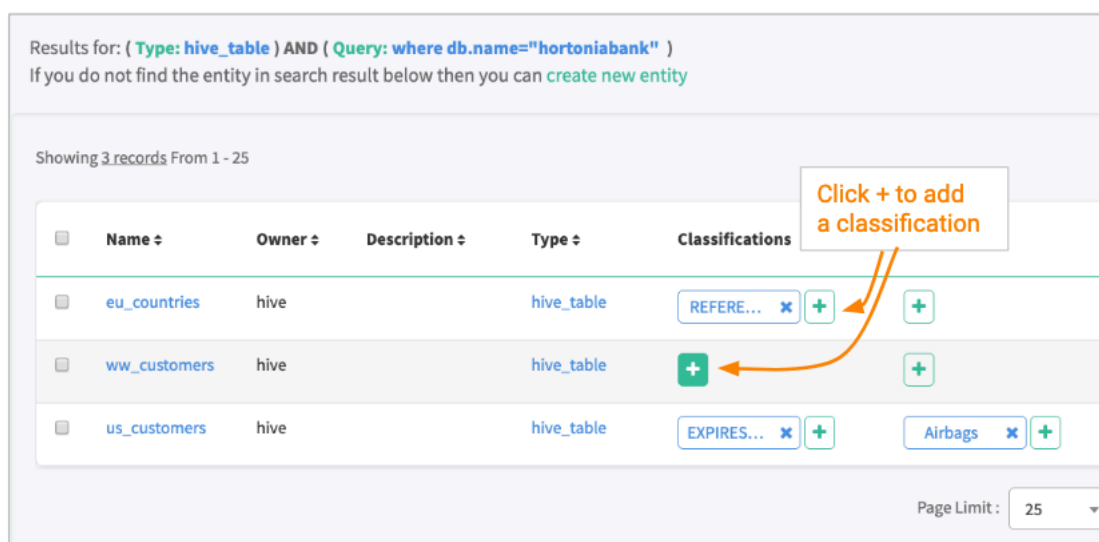
Assign classifications to entities in the entity detail page or in the search results.

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

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



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



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

Add Classification ✕

-- Select a Classification from the dropdown list --

Propagate
 Apply Validity Period

Remove propagation on entity delete

Cancel
Add

- **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

Showing 1 - 2

Show Propagated Classifications

Classification	Attributes	Action				
DATA_QUALITY	<table border="1"> <thead> <tr> <th>Name</th> <th>Value</th> </tr> </thead> <tbody> <tr> <td>score</td> <td>0.51</td> </tr> </tbody> </table>	Name	Value	score	0.51	Propagated From
Name	Value					
score	0.51					
INSURANCE	NA	Propagated From				

Toggle the view of classifications associated with this entity by propagation

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.

The screenshot shows the Cloudera interface for a table named 'claims_view (hive_table)'. Under the 'Classifications' section, there is a green '+' button. An orange arrow points to this button with the text 'Click to add a classification'. Below this, the 'Term' is set to 'Claim'. Under 'Propagated Classifications', 'DATA_QUALITY' and 'INSURANCE' are listed. An 'Add Classification' dialog box is open, showing a dropdown menu with 'CONSUMER_BANKING' selected. Below the dropdown, there are two checked checkboxes: 'Propagate' and 'Remove propagation on entity delete'. There is also an unchecked checkbox for 'Apply Validity Period'. A green 'Add' button is at the bottom right of the dialog. An orange arrow points from the 'Propagate' checkbox to the text 'This classification is applied to entities in the lineage impact of this entity'. Another orange arrow points from the 'Remove propagation on entity delete' checkbox to the text 'When the current entity is deleted, this classification is removed from the entities where it propagated in the lineage impact.'

- Set propagation at a specific point in the lineage.

claim_savings (hive_table)

Classifications: DATA_QUALITY

Term: +

Properties Lineage Relationships Classifications Audits Schema

Current Entity → Lineage → Impact

create external t... claim_savings create view if no... claims_view

Click on a lineage impact line

Enable/Disable Propagation

Enable/Disable Propagation Select Classifications to Block Propagation

claim_savings → create view if not from cost_savings.claim_savings

hive_table → hive_process

none

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.

Cancel Update

Enable/Disable Propagation

Enable/Disable Propagation Select Classifications to Block Propagation

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

Choose the classifications to propagate to entities in the lineage impact.

Update

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.

