Cloudera Runtime 7.2.6

Atlas Use Case: Extending Metadata

Date published: 2019-09-23
Date modified: 2020-12-11

https://docs.cloudera.com/

Legal Notice

Contents

Top-down process for adding a new metadata source

You can create entity types in Atlas to represent data assets, operations, or other types of artifacts from sources other than those supported by default.

Atlas' data model is designed to be flexible enough to represent a wide variety of data assets and the processes that generate them. You can design your own entities to collect metadata from sources beyond what comes predefined in Atlas. The high-level process for creating entities is as follows:

1	Think about how you want to model the core assets from your source. The predefined entity types inherit basic attributes from the DataSet and Process entity types. Using these general types as your starting point ensures that your entities can display lineage in the Atlas Dashboard and take advantage of other predefined rules and hierarchies built into the models. That said, new entity types are not limited to this data/process model.
2	Define the relationships you want to track among the data assets. To have your entities appear in lineage graphs in the Atlas Dashboard, include relationship attributes for "inputToProcesses" and "outputFromProcesses".
3	List all the metadata you want to track for each entity type. If necessary or convenient, define enumerations or structures to describe metadata specifically for your entities.
4	Use REST API or create an Atlas client to build the entity type definitions in Atlas. (build JSON)
5	Validate the model by manually filling in metadata in the types to validate that they behave the way you expect. You can do this using the Atlas Dashboard or REST API and then test that lineage, search, and Ranger tag-based-policies do what you expect.
6	Write a hook, bridge, or both to collect metadata automatically from your source.