Cloudera Runtime 7.2.10

Apache Hive metastore overview

Date published: 2021-02-22
Date modified: 2021-06-08

https://docs.cloudera.com/

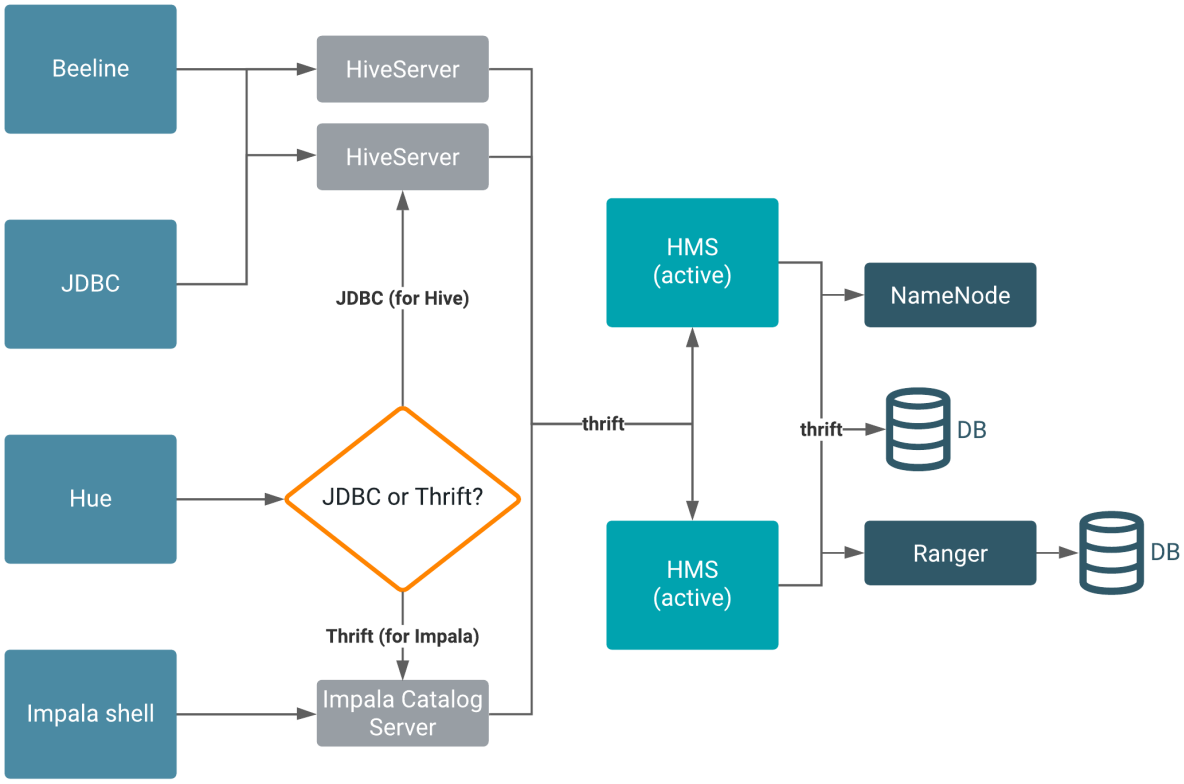
Legal Notice

Contents

Introduction to Hive metastore

Hive metastore (HMS) is a service that stores metadata related to Apache Hive and other services, in a backend RDBMS, such as MySQL or PostgreSQL. Impala, Spark, Hive, and other services share the metastore. The connections to and from HMS include HiveServer, Ranger, and the NameNode that represents HDFS.

Beeline, Hue, JDBC, and Impala shell clients make requests through thrift or JDBC to HiveServer. The HiveServer instance reads/writes data to HMS. By default, redundant HMS operate in active/active mode. The physical data resides in a backend RDBMS, one for HMS. All HMS instances use the same backend database. A separate RDBMS supports the security service, Ranger for example. All connections are routed to a single RDBMS service at any given time. HMS talks to the NameNode over thrift and functions as a client to HDFS.



HMS connects directly to Ranger and the NameNode (HDFS), and so does HiveServer, but this is not shown in the diagram for simplicity. One or more HMS instances on the backend can talk to other services, such as Ranger.

Related Information

Working with Apache Hive Metastore

Apache Hive storage in public clouds

Knowing the storage locations of Apache Hive data and metadata helps you troubleshoot problems and replicate data.

On public clouds, the Hive warehouse that stores Hive data is located in an object store, such as S3, by default. In the cloud, Hive uses HDFS merely for storing temporary files. The Hive Metastore (HMS) stores the schema for Hive tables. HMS uses a pre-installed MySQL database. You perform little, or no, configuration of HMS in the cloud.