Hive Metastore

Hive Metastore

Hive Metastore is a central repository for Hive metadata. It has 2 components

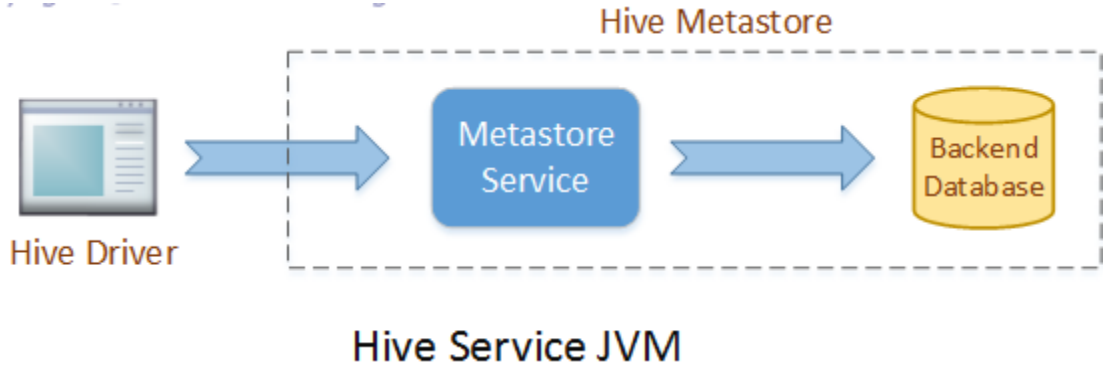
1. A Metastore Service to which the Hive Driver connects to and queries for the database schema.
2. A backing database to store the metadata. Currently Hive supports 5 backend databases: Derby, MySQL, MySQL Server, Oracle & Postgress.

Hive Metastore Modes

There are three modes to configure Hive Metastore:

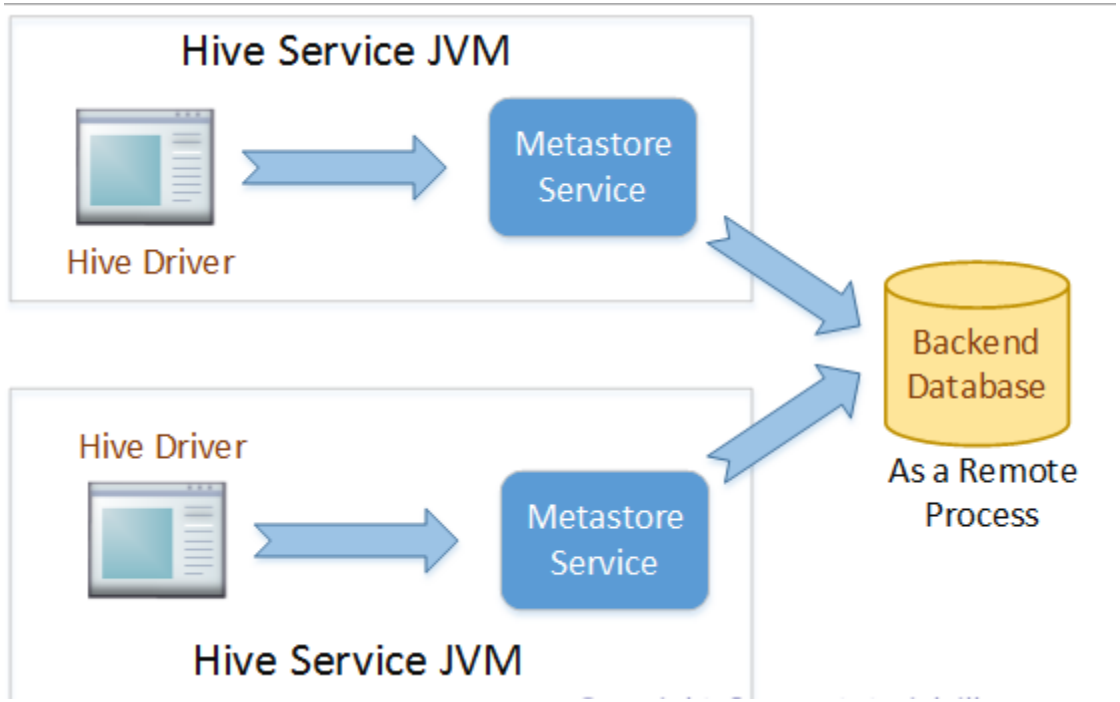
1. Embedded Metastore
2. Local Metastore
3. Remote Metastore

Embedded Metastore



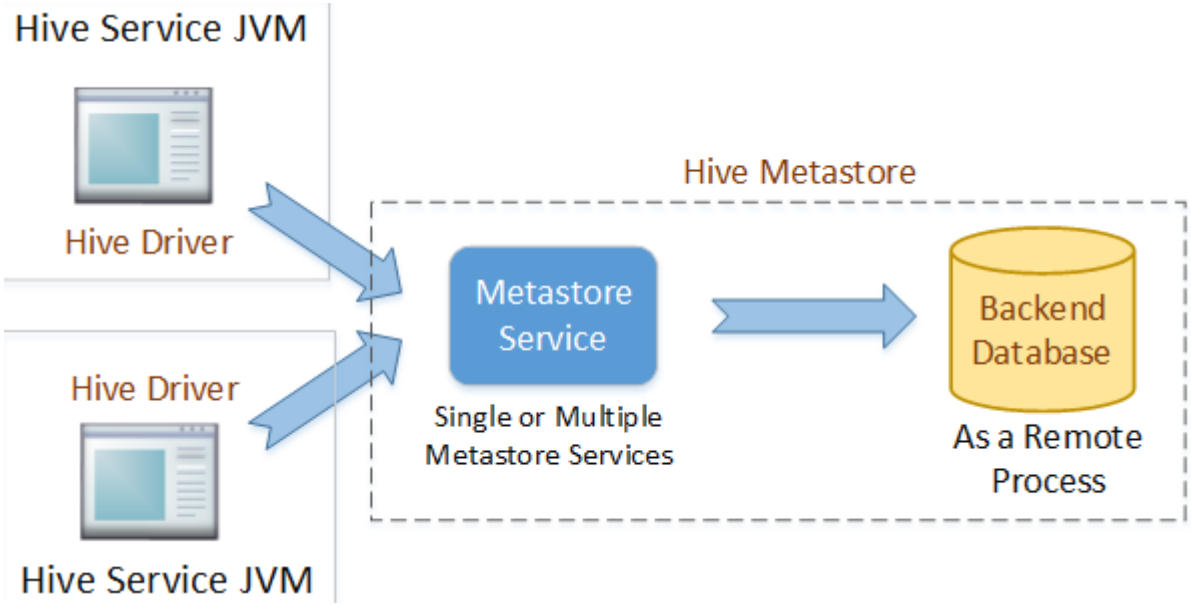
- In this mode the metastore uses a Derby database, and both the database and the Metastore service run embedded in the main HiveServer process.
- Both are started for you start the HiveServer process.
- This mode of Hive has a limitation that only one session could be opened at a time as only one embedded Derby database can access the database files on disk.

Local Metastore



- In this mode the Hive metastore service runs in the same process as the main HiveServer process, but the metastore database runs in a separate process, and can be on a separate host.
- The embedded metastore service communicates with the metastore database over JDBC.

Remote Metastore



- In this mode the Hive metastore services runs in its own JVM process; other process communicate with it via the Thrift network API (configured via the hive.metastore.uri property)
- The metastore service communicates with the metastore database over JDBC (configured via the java.jdo.option.ConnectionURL property)
- The main advantage of Remote mode over Local mode is that Remote mode does not require the administrator to share JDBC login information for the metastore database with each Hive user.