Hive

Apache Hive™ is a data warehouse system for Hadoop that facilitates easy data summarization, ad-hoc queries, and the analysis of large datasets stored in Hadoop-compatible file systems, such as the MapR Data Platform (MDP). Hive provides a mechanism to project structure onto this data and query the data using a SQL-like language called HiveQL. At the same time this language also allows traditional map/reduce programmers to plug in their custom mappers and reducers when it is inconvenient or inefficient to express this logic in HiveQL.

You can refer also to documentation available from the Apache Hive project.

Hive components include the following:

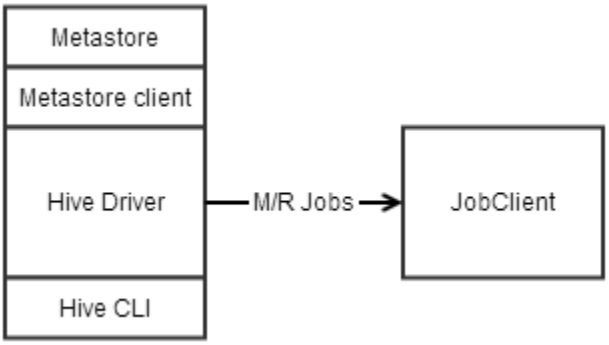
- Hive Metastore
- HiveServer2
- HCatalog
- WebHCat
- Hive CLI
- Beeline

The following examples show how these components communicate with each other and when you might want to configure security features such as authentication and encryption:

- Case 1: Jobs Submitted by the Hive CLI, Embedded Metastore
- Case 2: Jobs Submitted by the Hive CLI, Remote Metastore
- Case 3: Jobs Submitted by HiveServer2, Embedded Metastore
- Case 4: Jobs Submitted by HiveServer2, Remote Metastore

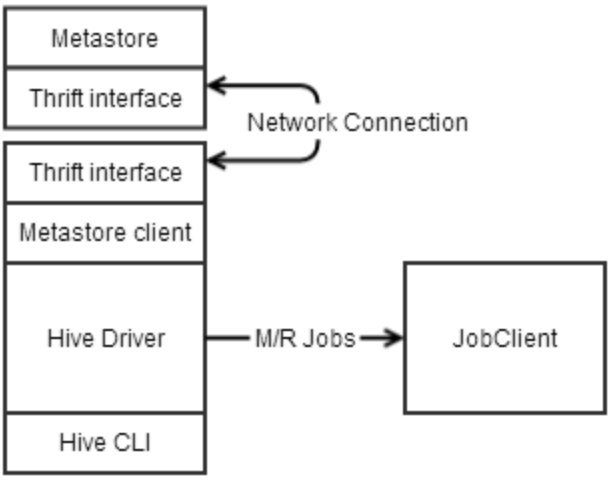
Case 1: Jobs Submitted by the Hive CLI, Embedded Metastore

In this case, all the information needed by Hive is contained within a single process, and no security is needed beyond that already provided by the JobClient's communications.



Case 2: Jobs Submitted by the Hive CLI, Remote Metastore

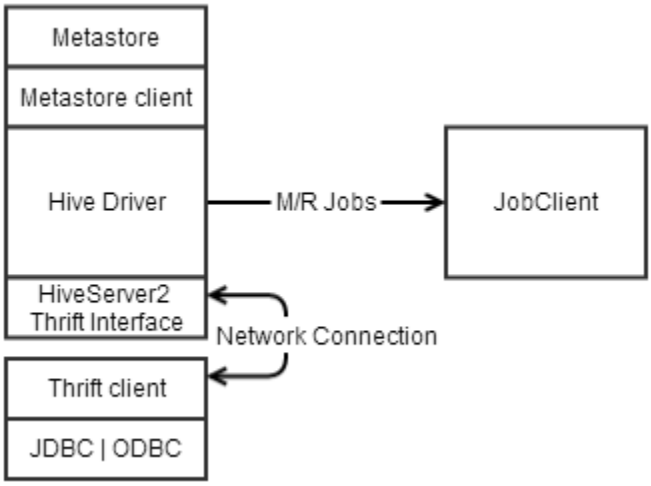
In this case, Hive needs to access a metastore remote to Hive's process using a Thrift interface. This communication can be left unsecured, secured with Kerberos, or secured with MapR-SASL.



Case 3: Jobs Submitted by HiveServer2, Embedded Metastore

In this case, JDBC or ODBC on a user's machine sends queries to HiveServer2, which submits the queries to the driver for parsing. The communication between JDBC and HIveServer2 can be secured with username and password with SSL, MapR-SASL, or with Kerberos. Either approach offers authentication and encryption. JDBC/ODBC can also be configured to use username and password without SSL, which offers authentication *only*.

To use SSL from a client machine the `ssl_truststore` file must be copied from the cluster to the client.



Case 4: Jobs Submitted by HiveServer2, Remote Metastore

In this case, JDBC or ODBC on a user's machine sends queries to HiveServer2, which submits the queries to the driver, which runs the query and returns the results. The metastore is remote. In this case, there are two communications links to secure: the Thrift interface between the metastore client and server, and the Thrift interface between the client JDBC/ODBC and HiveServer2. The security arrangements for these links are identical to Cases 2 and 3.