

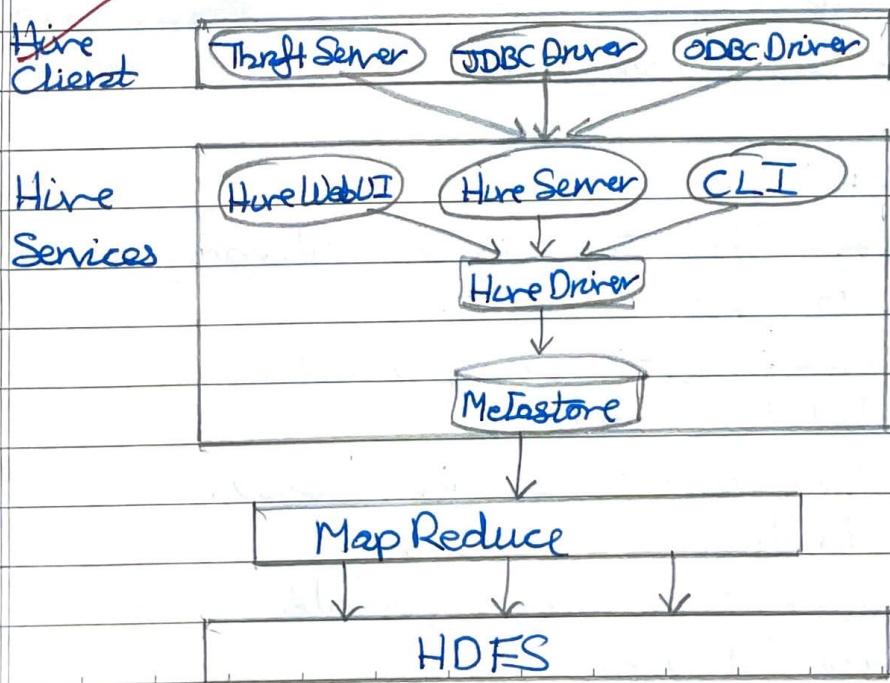
Experiment - 4

Aim: To execute HIVE commands in Hadoop Hortonworks Sandbox.

Theory:

Apache HIVE is an open source data warehouse software for reading, writing & managing large data set files that are stored directly in either HDFS or other data storage systems such as apache HBASE. Hive enables SQL developers to write Hive Query Language (HQL) statements that are similar to SQL statements for data query and analysis. It is designed to make MapReduce programming easier as you do not have to know & write lengthy Java code. Instead, you can write queries more simply in HQL & HIVE can then create the map & reduce functions.

Architecture:



i) Hive Client: Hive allows writing applications in various language including JAVA, Python & C++. It supports different types of clients such as:

- Thrift Server - Thrift Server is a cross language service provider ~~platform~~ that serves the request from all these programming languages that support Thrift.
- JDBC Server - JDBC driver is used to establish a connection between Hive and Java application.
- ODBC Server - ODBC driver allows the applications that support the ODBC protocol to connect the HIVE.

ii) Hive Services:

- Hive CLT (Command Line interface)
- Hive Metastore
- Hive Web User Interface
- Hive Server
- Hive Driver
- Hive Compiler
- Hive Execution Engine

Hive Commands

i) Show databases: Show the databases already existing in HIVE
Eg: show database;

ii) Create database: Create a new database
Eg: create database studentdb;

iii) Use database: Use the database for executing queries
Eg: use studentdb;

iv) Select: To display the contents of a table
Eg: select * from student;

v) Insert: Insert value in the table

Eg: insert into student partition (load_date = '2023-03-09')
values (101, 'Manor', 'Computer', 123456);

vi) Order By: Order the result in descending or ascending order

Eg: select id, name from student order by id desc;

vii) Group By: Group the result wrt an attribute

Eg: select branch, count(*) from student group by branch;

viii) Delete: Delete the entry from the table

Eg: delete from student where name = 'Raj'

ix) ALTER: Used to alter a table

Eg: remove: alter table student rename to stud;

x) Aggregate functions: These are sum, max, min, avg, count
functions on table

Eg: select sum(cgpa) from stud;

Conclusion: In conclusion we have performed several
HiveQL queries like groupby, join, insert, alter etc. We
also learnt about basics of Apache HIVE and the architecture
of Apache HIVE.

(contd.)