**ASSIGNMENT 6.3**

PROBLEM STATEMENT:

**1) Explain Hive Architecture in brief.**

Solution:



Hive Consists of Mainly 3 core parts:

1) Hive Clients

2) Hive services

3) Hive storage and compute.

**Hive Clients:**

Hive provides different drivers for communication with a different type of applications. For Thrift based applications, it will provide Thrift client for communication.

For Java related applications, it provides JDBC Drivers. Other than any type of applications provided ODBC drivers. These Clients and drivers in turn again communicate with Hive server in the Hive services.

**Hive Services:**

Client interactions with Hive can be performed through Hive Services. If the client wants to perform any query related operations in Hive, it has to communicate through Hive Services.

CLI is the command line interface acts as Hive service for DDL (Data definition Language) operations. All drivers communicate with Hive server and to the main driver in Hive services as shown in above architecture diagram.

Driver present in the Hive services represents the main driver, and it communicates all type of JDBC, ODBC, and other client specific applications. Driver will process those requests from different applications to meta store and field systems for further processing.

**Hive Storage and Computing:**

Hive services such as Meta store, File system, and Job Client in turn communicates with Hive storage and performs the following actions

* Metadata information of tables created in Hive is stored in Hive "Meta storage database".
* Query results and data loaded in the tables are going to be stored in Hadoop cluster on HDFS.

**2) Explain Hive components in brief.**

Solutions:

* **User Interface**

Hive is a data warehouse infrastructure software that can create interaction between user and HDFS. The user interfaces that Hive supports are Hive Web UI, Hive command line, and Hive HD Insight (In Windows server).

* **Meta Store**

Hive chooses respective database servers to store the schema or Metadata of tables, databases, columns in a table, their data types, and HDFS mapping.

* **HiveQL Process Engine**

HiveQL is similar to SQL for querying on schema info on the Metastore. It is one of the replacements of traditional approach for MapReduce program. Instead of writing MapReduce program in Java, we can write a query for MapReduce job and process it.

* **Execution Engine**

The conjunction part of HiveQL process Engine and MapReduce is Hive Execution Engine. Execution engine processes the query and generates results as same as MapReduce results. It uses the flavor of MapReduce.

* **HDFS or HBASE**

Hadoop distributed file system or HBASE are the data storage techniques to store data into file system.