**Assignment 6.3**

1. **HIVE ARCHITECTURE:**

HIVE is a data warehouse infrastructure tool to process structured data in Hadoop. It resides on top of Hadoop to summarize Big Data , and makes querying and analyzing easy.

HIVE is not:

* A relational DATABESE
* A design for real-time queries and row-level updates

Important characteristics of HIVE:

1. In HIVE, tables and database are created FIRST and then data is loaded into these tables.
2. HIVE as data warehouse is used for managing and querying only structured data that is stored in tables.
3. HIVE’s SQL-Inspired Language separates the user from the complexity of MAP REDUCE programming.
4. Hadoop’s programming works on flat files. So, HIVE can use directory structures to “partition” data to improve performance on certain queries.
5. A new and important component of HIVE is METASTORE used for storing schema information.
6. HIVE is designed for OLAP.
7. HIVE provides SQL type language for querying called HiveQL or HQL.
8. It is familiar, fast, scalable and extensible.

**ARCHITECTURE OF HIVE**:

Sjjs

**HDFS AND HBASE Data Storage**

**USER INTERFACE**

**MAP REDUCE**

**HIVE COMMAND LINE**

**HD Insight**

**WEB UI**

**HIVE QL PROCESS ENGINE**

**EXECUTION ENGINE**

**META STORE**

This component diagram contains different units. The following table describes each unit:

|  |  |
| --- | --- |
| Unit Name | Operation |
| 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. |

1. **Major COMPONENTS of HIVE:**
2. **UI :-**UI means User Interface, The user interface for users to submit queries and other operations to the system.
3. **Driver :-**The Driver is used for receives the quires from UI .This component implements the notion of session handles and provides execute and fetch APIs modeled on JDBC/ODBC interfaces.
4. **Compiler :-**The component that parses the query, does semantic analysis on the different query blocks and query expressions and eventually generates an execution plan with the help of the table and partition metadata looked up from the metastore.
5. **MetaStore :-**The component that stores all the structure information of the various tables and partitions in the warehouse including column and column type information, the **serializers** and **deserializers** necessary to read and write data and the corresponding HDFS files where the data is stored.
6. **Execution Engine :-**The component which executes the execution plan created by the compiler. The plan is a DAG of stages. The execution engine manages the dependencies between these different stages of the plan and executes these stages on the appropriate system components.

**SHELL**

Interactive Queries Connections to DATABASE. Support Web & JDBC Clients.

**EXECUTION ENGINE**

Session handling

Other imp. activities

Compiled Query plan execution

Query Parsing

Query Plan

Plan optimization

HDFS Location Schema

**HIVE COMPONENTS**

**META STORE**

**DRIVER**

**COMPILER**