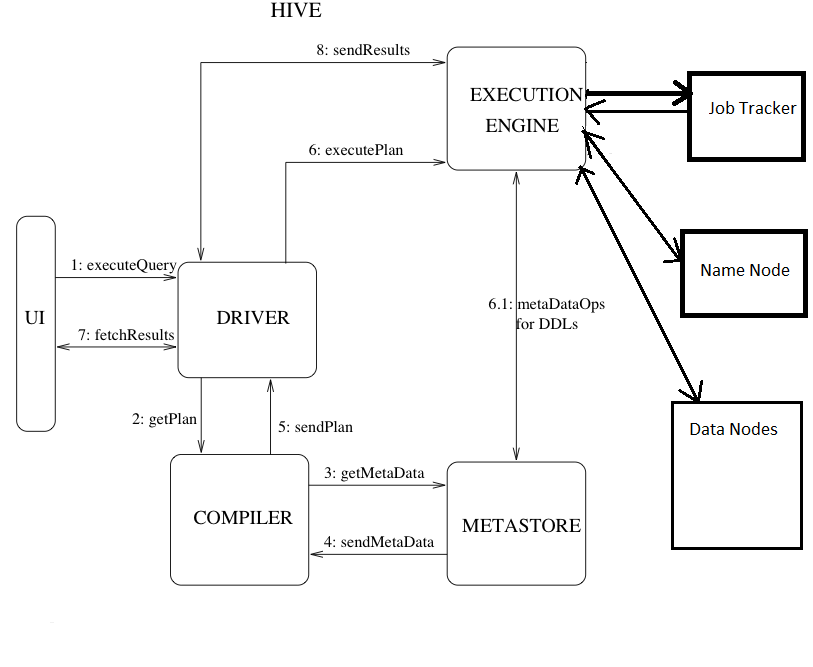
**ASSIGNMENT 23.3**

**QUESTION 1 : Explain Hive Architecture in Brief.**

There are 3 major components in Hive as shown in the architecture diagram. They are hive clients, hive services and Meta Store. Under hive client, we can have different ways to connect to HIVE SERVER in hive services.

These are Thrift client, ODBC driver and JDBC driver. Coming to thrift client, it provides an easy environment to execute the hive commands from a vast range of programming languages. Thrift client bindings for Hive are available for C++, Java, PHP scripts, python scripts and Ruby. Similarly, JDBC and ODBC drivers can be used for communication between hive client and hive servers for compatible options.



* HIVESERVER is an API that allows the clients (JDBC) to execute the queries on hive data warehouse and get the desired results. Under hive services driver, compiler and execution engine interact with each other and process the query.
* The client submits the query via a GUI. The driver receives the queries in the first instance from GUI and it will define session handlers, which will fetch required APIs that is designed with different interfaces like JDBC or ODBC. The compiler creates the plan for the job to be executed. Compiler in turn is in contact with matter and its gets metadata from Meta Store.
* Execution Engine (EE) is the key component here to execute a query by directly communicating with Job Tracker, *Name Node* and *Data nodes*. As discussed earlier, by running hive query at the backend, it will generate a series of MR (Map Reduce) Jobs. In this scenario, the execution engine plays like a bridge between hive and Hadoop to process the query. For DFS operations, EE contacts Name Node.
* At the end, EE is going to fetch desired results from Data Nodes. EE will be having bi-directional communication with Metastore. In hive, side is a framework to serialize and de-serialize input and output data from HDFS to local or vice versa.
* Metastore is used for collection of all the Hive metadata and it’s having back up services to backup meta store info. The service runs on the same JVM as the services of hive running on. The structural information of tables, their columns, column types and similarly the partition structure information will also be stored in this.

**QUESTION 2 : Explain Hive Components in brief :-**

**The following are the components of HIVE :**

* UI :

The user interface for users to submit queries and other operations to the system. As of 2011 the system had a command line interface and a web based GUI was being developed.

* Driver :-

The component which receives the queries. This component implements the notion of session handles and provides execute and fetch APIs modeled on JDBC/ODBC interfaces.

* 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.

* 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.

* 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.