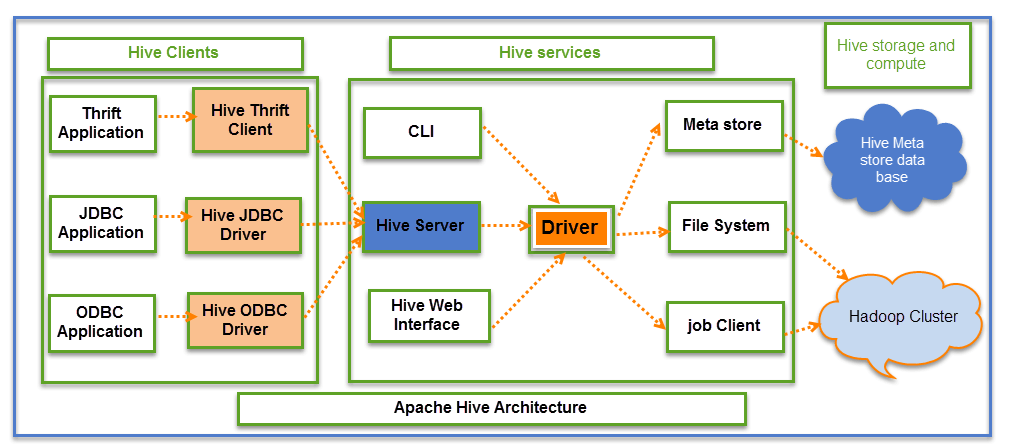
ACD\_BDD2.3\_Session\_6\_Assignment\_3\_a

## Hive Architecture

[](https://cdn.guru99.com/images/Hive/120415_1251_Introductio1.png)

Hive Consists of Mainly 3 core parts

1. **Hive Clients**
2. **Hive Services**
3. **Hive Storage and Computing**

**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](https://www.guru99.com/java-tutorial.html)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.

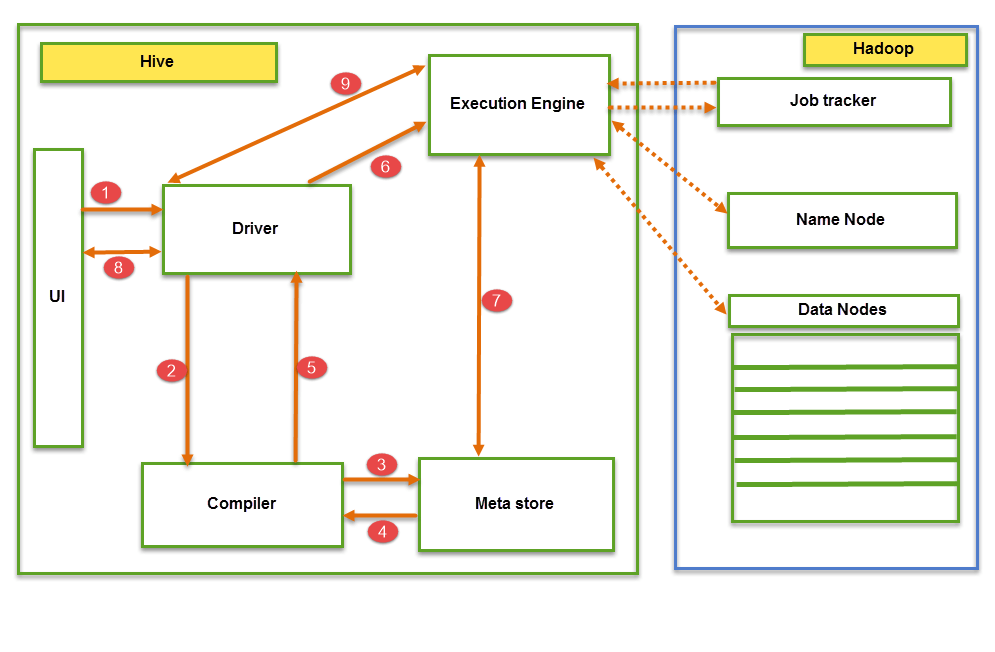
ACD\_BDD2.3\_Session\_6\_Assignment\_3\_b

## Hive Components

The main components of Hive are:

* 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.

**Job exectution flow:**

[](https://cdn.guru99.com/images/Hive/120415_1251_Introductio2.png)

The data flow in Hive behaves in the following pattern;

1. Executing Query from the UI( User Interface)
2. The driver interacts with Compiler for getting the plan, process and its related metadata information gathering.
3. The compiler creates the plan for a job to be executed. Compiler communicates with Meta store for getting metadata request.
4. Meta store sends metadata information back to compiler
5. Compiler communicates with Driver with the proposed plan to execute the query
6. Driver sends execution plans to Execution engine
7. Execution Engine (EE) acts as a bridge between Hive and Hadoop to process the query. For DFS operations.

* EE should first contacts Name Node and then to Data nodes to get the values stored in tables.
* EE is going to fetch desired records from Data Nodes. The actual data of tables resides in data node only. While from Name Node it only fetches the metadata information for the query.
* It collects actual data from data nodes related to mentioned query
* Execution Engine (EE) communicates bi-directionally with Meta store present in Hive to perform DDL (Data Definition Language) operations. Here DDL operations like CREATE, DROP and ALTERING tables and databases are done. Meta store will store information about database name, table names and column names only. It will fetch data related to query mentioned.
* Execution Engine (EE) in turn communicates with Hadoop daemons such as Name node, Data nodes, and job tracker to execute the query on top of Hadoop file system

1. Fetching results from driver
2. Sending results to Execution engine. Once the results fetched from data nodes to the EE, it will send results back to driver and to UI ( front end)