**Session 2**

**Assignment 2.9**

Student Name: Karthik K

Course: Big Data Hadoop & Spark Training

# What is HIVE?

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.

# Architecture of HIVE:

The following component diagram depicts the architecture of Hive:



Hive is an ETL and Data warehousing tool developed on top of Hadoop Distributed File System (HDFS). Hive makes job easy for performing operations like

* Data encapsulation
* Ad-hoc queries
* Analysis of huge datasets

# HIVE Components:

**UI**:- UI means User Interface, The user interface for users to submit queries and other operations to the system.

**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.

**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.

## Working of Hive

The following diagram depicts the workflow between Hive and Hadoop

