**SESSION 14: HIVE INTRODUCTION**

**ASSIGNMENT 2**

1. **Is it possible to use same metastore by multiple users in case of embedded Hive, if no then why?**

Embedded mode means that the Hive driver, metastore and Apache Derby are all running in one Java Virtual Machine (JVM). This mode can only support a single Hive session, and thus can’t be used by multiple users.

It is recommended to use standalone databases like MySQL.

1. **What is SerDe in Hive?**

SerDe stands for Serializer Deserializer. Hive uses SerDe to read and write data from tables. Hive doesn’t own the HDFS format that data is stored in. Users are able to write files to HDFS with certain commands, and use Hive to correctly “parse” that file format in a way that can be used by Hive. A SerDe is a powerful (and customizable) mechanism that Hive uses to “parse” data stored in HDFS to be used by Hive.

1. **What is functionality of query processor in Apache Hive?**

Query processor implements the processing framework for converting SQL to a graph of map/reduce jobs and the execution time framework to run those jobs in the order of dependencies.

1. **How can Hive avoid MapReduce?**

The SELECT statement doesn’t involve any mapreduce job. It is normal query which fetches records from the table. Once, we use the WHERE statement it produces a mapreduce job.

Thus, one can avoid MapReduce in Hive by avoiding the use of WHERE clause.

1. **What are the types of tables in Hive?**

There are two types of tables in Hive – External and Managed (or Internal) tables.

1. **Does Hive support record level insert, delete or update?**

Transactions were introduced in Hive 0.13, but they only partially fulfilled the ACID properties like atomicity, consistency, durability at the partition level. Isolation could be provided by turning on one of the locking mechanisms available with zookeeper or in memory.

But in Hive 0.14, new APIs have been added to completely fulfill the ACID properties while performing any transaction. Any Insert, Delete, or Update transaction can now be performed at the row level but with a number of limitations. It is essential to have the ORC file format for performing transactions, and the tables must be bucketed.

1. **What are the binary storage formats supported in Hive?**

SEQUENCEFILE, RCFILE, ORCFILE, PARQUETS, etc. are the binary storage formats supported in Hive.

1. **What is the difference between external table and internal table in Hive?**

The difference between external and internal tables is that when an internal table is dropped, hive deletes both data and the metadata of the table but when an external tables is dropped, then only the metadata is deleted.