**Session 14: Hive Introduction**

**Assignment 2**

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

**Answer:** No, it is not possible to use metastore in sharing mode. It is recommended to use standalone "real" database like MySQL or PostGresSQL.

Whenever we run the hive in embedded mode, it creates the local metastore. And before creating the metastore it looks whether metastore already exist or not. This property is defined in configuration file hive-site.xml. Property is "javax.jdo.option.ConnectionURL" with default value "jdbc:derby:;databaseName=metastore\_db;create=true". So to change the behavior change the location to absolute path, so metastore will be used from that location

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

**Answer:** A SerDe is a combination of a Serializer and a Deserializer (hence, Ser-De). Apache **Hive** uses **SerDe** to read and write data from tables while selecting the data from Apache Hive SerDe.deserialize() method is called and while inserting the data SerDe.serialize() method is called.

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

**Answer:** 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.

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

**Answer:** If we set the property hive.exec.mode.local.auto to true then hive will avoid mapreduce to fetch query results.

**5. What are the types of table in Hive?**

**Answer:** There are two types of tables.  
a). Managed tables.  
b). External tables.

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

**Answer:** Hive does not provide record-level update, insert, or delete. Users can go with CASE statements and built in functions of Hive to satisfy the above DML operations. Thus, a complex update query in a RDBMS may need many lines of code in Hive.

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

**Answer:**

1. Sequence File.
2. RCFile (Record Columnar File)

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

**Answer:**

**External Table**

1. External table stores files on the HDFS server but tables are not linked to the source file completely.
2. If we delete an external table the file still remains on the HDFS server.
3. The file and the table link is there but read only.As an example if you create an external table called “**example\_test**” in HIVE using HIVE-QL and link the table to file “**flat\_file.txt**”, then deleting “**example\_test**” from HIVE will not delete “**flat\_file.txt**” from HDFS.
4. External table files are accessible to anyone who has access to HDFS file structure and therefore security needs to be managed at the HDFS file/folder level.
5. Meta data is maintained on master node and deleting an external table from HIVE, only deletes the metadata not the data/file.

**Internal Table:**

1. **Stores data on HDFS but in a kind of restricted area.**
2. Stored in a directory based on settings in the following file: **hive.metastore.warehouse.dir** .By default internal tables are stored in the following directory “ **/user/hive/warehouse**” you can change it by updating the location in the config file
3. Deleting the table deletes the metadata & data from masternode and HDFS respectively
4. Security needs to be managed within HIVE, probably at the schema level (depends on organisation to organisation). HDFS security is out of scope in this case.
5. It is used when we want to store the data temporary or want to use HIVE to manage the lifecycle of tables and data.