Session 14: Hive Introduction

Assignment 2

# Problem Statement

**Give brief answers to the below questions:**

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

**Ans:**

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

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

**Ans:**

A SerDe is a short name for a Serializer Deserializer. Hive uses SerDe (and FileFormat) to read and write data from tables. An important concept behind Hive is that it DOES NOT own the Hadoop File System (HDFS) format that data is stored in. Users are able to write files to HDFS with whatever tools/mechanism takes their fancy("CREATE EXTERNAL TABLE" or "LOAD DATA INPATH," ) 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 the functionality of query processor in Apache Hive?**

**Ans:**

This component 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?**

**Ans:**

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

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

**Ans:**

There are two types of tables in Hive.

1. **Managed Table:**

* In Hive, Managed tables are stored in warehouse folder of hive.
* When you drop an internal table, it drops the data, and it also drops the metadata.
* So you can create manage table when **data is temporary.**

1. **External Table Benefits:**

* When you want to store data permanently in HDFS, you can use External Table.
* You can use External Table, when data is also used outside of Hive. AR Network uses same data file for **filtration** and **transformation**.
* In case of External Table, data remains in the underlying location even after a DROP TABLE. This helps you when you want to point multiple schemas (tables or views) at a single data set.

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

**Ans:**

Hive does not support record level insert, delete or update.

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

**Ans:**

Data is stored in File formats llike text file ,CSv file,RC file, Sequence file etc.

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

**Ans:**

Managed Table: Schema stored in the local system, but external table Metastore separate from the node and stored in a secure database.

Internal Table, hive reads and loads entire file as it is to process, but in External simply loads depends on the query logic.

If we drop the table, Hive drop original data and Metastore, but in external table just drop Metastore, but not original table.

Hive by default store in internal table, but it’s not recommendable. Store the data in external table.