Assignment 14.2

By Shabnam

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

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

1. What is SerDe in Hive?

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?

It’s a core processing unit in Hive framework, it converts SQL to map reduce jobs . 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?

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?

There are two types of tables in HIVE. That is

* External table: In external table only the schema is under control of Hive.
* Internal table: In managed table both the data an schema in under control of hive

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

No , by default it does not support record level insert, delete, or update. But in newer versions of HIVE we have mechanism to do so.

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

There are some specific file formats which Hive can handle such as:

• TEXTFILE

• SEQUENCEFILE

• RCFILE

• ORCFILE

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

The main difference is that when you drop an external table, the underlying data files stay intact. This is because the user is expected to manage the data files and directories. With a managed table, the underlying directories and data get wiped out when the table is dropped.