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 or PostGresSQL.

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. 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?

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?

Hive tables can be created as EMBEDDED, EXTERNAL or INTERNAL. This is a choice that affects how data is loaded, controlled, and managed.

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

Hive does not provide record-level update, insert, or delete. Henceforth, Hive does not provide transactions too. However, 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.

Hive uses Hadoop to provide highly scaleable bandwidth to the data, but until recently did not support updates, deletes, or transaction isolation. This has prevented many desirable use cases, such as updating of dimension tables or doing data clean up. We have implemented the standard SQL commands insert, update, and delete allowing users to insert new records as they become available, update changing dimension tables, repair incorrect data, and remove individual records.

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

Text file

Sequence file

Rc file

ORC file

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.