**Assignment 14\_ 1**

**By Shabnam**

1. When Hive is best suited and when is it not?

* Hive is best suited for data summarazition , querying and analysis. For example data warehouses where the data is static and doesn’t change during operation or process.
* Hive is not suitable for dynamic data. Hive queries have higher latency, due to the start-up overhead for MapReduce jobs. Hive doesn’t support transaction processing.

1. When should one use Hive over MapReduce?

When the user is not familiar with map reduce programs he /she can opt for hive which provides SQL as frontend and mapreduce as backend.

1. What is Hive metastore?

The metadata for Hive tables and partitions are stored in the Hive Metastore.

1. How can Hive improve performance with orc file format tables?

By using orc file formats, each file with the columnar layout is optimised for compression and skipping of data/columns to reduce read and decompression load.

1. What is thrift server and client, jdbc and odbc driver importance in hive?

Thrift server provides clean abstractions for data transport, data serialization, and application level processing. Thrift is interface definition language used in RPC communication, it allows for cross language services development. For example it can be used in data visualization and integration tools.

A JDBC driver is a software component enabling a java application to interact with the database. Where as ODBC driver accomplishes dbms independence by using a odbc driver as a translation layer between the application and database.

1. What is the importance of partition in hive?

The importance of hive partition is that we save a considerable amount of space on disk and it can be very fast to perform partition elimination. The downside of the partitioning is that it’s necessary to tell Hive which partition we’re loading in a query.

1. What is the use of bucketing in hive?

Hive provides Bucketing, which allows user to divide table data sets into more manageable parts. Bucketing concept also provides the flexibility to keep the records in each bucket to be sorted by one or more columns.

1. What is the difference between static partitioning and dynamic partitioning in hive?

|  |  |
| --- | --- |
| **Static partitioning** | **Dynamic partitioning** |
| 1. Less time consuming | 1. More time consuming |
| 1. Should insert files into each and every partition manually | 1. Single insert to partition table is dynamic partition |
| 1. Static partition is in Strict Mode | 1. To use Dynamic partition in hive then mode is in nonstrict mode |
| 1. In static partitioning you are well aware of the no of partition columns | 1. In dynamic partitioning we don’t know how many columns are there and choose dynamic way. |
| 1. property set hive.mapred.mode = strict | 1. property set hive.mapred.mode = nonstrict |