1. The SQL interface to hadoop from Hive perspective. Give the available options in Hive and explain the process what happens when the Hive query is executed?

2. List the differences between Hive and RDBMS?

3. What are data abstractions in Hive?

    Create 2 data bases and 3 tables in each and show the tables list in each.

4. What are Hive Join Strategies?

    As given in hive tuning document create the data, create the tables, impose join queries and then capture the response times.

   Paste all the screenshots from table creation to data load and other sql queries.

   Please write your observations after executing the join queries.

5. How data locality will be achieved in hive?

6. Create the tables with help of different file formats.

    i. Parquet

    ii. ORC

    iii. Avro

    The data source is from previous assignment. Create the tables on top of that data.

After storing the data into files, Check for the data size in those tables.

     Execute the same queries which you have done in last assignment and capture the response times.

     The create queries should be submitted and responses in excel sheet should be submitted.

7. Take any file of your choice and load it in normal format and other formats as given below.

    a. Create the table with partitions and load the data.

    b. Create the table with sort by option and load the data

    c.  Create the table with cluster by option and load the data

    d. create the table with bucketing option and load the data.

  Execute one simple select query on the tables.