**Hive assignment-2 Questions**

Scenario Based questions:

Will the reducer work or not if you use “Limit 1” in any HiveQL query?

Suppose I have installed Apache Hive on top of my Hadoop cluster using default metastore configuration. Then, what will happen if we have multiple clients trying to access Hive at the same time?

Suppose, I create a table that contains details of all the transactions done by the customers: CREATE TABLE transaction\_details (cust\_id INT, amount FLOAT, month STRING, country STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY ‘,’ ;

Now, after inserting 50,000 records in this table, I want to know the total revenue generated for each month. But, Hive is taking too much time in processing this query. How will you solve this problem and list the steps that I will be taking in order to do so?

How can you add a new partition for the month December in the above partitioned table?

I am inserting data into a table based on partitions dynamically. But, I received an error – FAILED ERROR IN SEMANTIC ANALYSIS: Dynamic partition strict mode requires at least one static partition column. How will you remove this error?

Suppose, I have a CSV file – ‘sample.csv’ present in ‘/temp’ directory with the following entries:

id first\_name last\_name email gender ip\_address

How will you consume this CSV file into the Hive warehouse using built-in SerDe?

Suppose, I have a lot of small CSV files present in the input directory in HDFS and I want to create a single Hive table corresponding to these files. The data in these files are in the format: {id, name, e-mail, country}. Now, as we know, Hadoop performance degrades when we use lots of small files.

So, how will you solve this problem where we want to create a single Hive table for lots of small files without degrading the performance of the system?

LOAD DATA LOCAL INPATH ‘Home/country/state/’

OVERWRITE INTO TABLE address;

The following statement failed to execute. What can be the cause?

Is it possible to add 100 nodes when we already have 100 nodes in Hive? If yes, how?

Hive Practical questions:

Hive Join operations

Create a table named CUSTOMERS(ID | NAME | AGE | ADDRESS | SALARY)

Create a Second table ORDER(OID | DATE | CUSTOMER\_ID | AMOUNT

)

Now perform different joins operations on top of these tables

(Inner JOIN, LEFT OUTER JOIN ,RIGHT OUTER JOIN ,FULL OUTER JOIN)

BUILD A DATA PIPELINE WITH HIVE

Download a data from the given location -

https://archive.ics.uci.edu/ml/machine-learning-databases/00360/

1. Create a hive table as per given schema in your dataset

2. try to place a data into table location

3. Perform a select operation .

4. Fetch the result of the select operation in your local as a csv file .

5. Perform group by operation .

7. Perform filter operation at least 5 kinds of filter examples .

8. show and example of regex operation

9. alter table operation

10 . drop table operation

12 . order by operation .

13 . where clause operations you have to perform .

14 . sorting operation you have to perform .

15 . distinct operation you have to perform .

16 . like an operation you have to perform .

17 . union operation you have to perform .

18 . table view operation you have to perform .

hive operation with python

Create a python application that connects to the Hive database for extracting data, creating sub tables for data processing, drops temporary tables.fetch rows to python itself into a list of tuples and mimic the join or filter operations