**Problem Statement**

**Explain with an example in brief.**

**● Hive Data Definitions**

Hive Data Definition commands are used to perform basic database operations. Some DDL commands are-

1. CREATE DATABASE- This command is used to create a database. The simplest syntax for creating a database is shown below

CREATE DATABASE custom;

Hive will throw an error if financials already exists. You can suppress these warnings with this variation:

CREATE DATABASE IF NOT EXISTS custom;

1. DESCRIBE DATABASE – Describe command describes the database. It shows the location of database.

DESCRIBE DATABASE custom;

1. USE DATABASE- The USE command sets a database as your working database, analogous to changing working directories in a filesystem.

USE custom;

1. DROP DATABASE- This command is used to delete the database. Cascade keyword is used to drop he tables as well from the database.

DROP DATABASE custom CASCADE;

1. CREATE TABLE- It is used to create a table in database. Complete schema is defined while giving the syntax.

CREATE TABLE IF EXISTS emp\_details

(emp\_name string, unit string, exp int, location string ) row format delimited fields terminated by ',’;

1. DROP TABLE- It is used to delete table.

DROP TABLE IF EXISTS emp\_details;

1. ALTER TABLE- Used to change or modify content and schema of table.
   1. RENAME-

ALTER TABLE emp\_details RENAME TO employee\_details;

* 1. Changing Columns-

ALTER TABLE emp\_details CHANGE COLUMN emp\_name emp\_name STRING COMMENT 'Employee Name' AFTER UNIT;

**● Hive Data Manipulations**

1. LOAD DATA- Load operations are currently pure copy/move operations that move data files into locations corresponding to Hive tables.

LOAD DATA [LOCAL] INPATH 'filepath' [OVERWRITE] INTO TABLE tablename [PARTITION (partcol1=val1, partcol2=val2 ...)] •

* file path can refer to a file (in which case Hive will move the file into the table) or it can be a directory (in which case Hive will move all the files within that directory into the table).
* If the keyword LOCAL is specified, then:
* the load command will look for filepath in the local file system.
* If a relative path is specified, it will be interpreted relative to the user's current working directory.
* The user can specify a full URI for local files as well, for example: file:///user/hive/project/data1
* the load command will try to copy all the files addressed by filepath to the target filesystem.
* If the OVERWRITE keyword is used then the contents of the target table (or partition) will be deleted and replaced by the files referred to by filepath, otherwise the files referred by filepath will be added to the table.
* Query results can be inserted into filesystem directories.
* INSERT OVERWRITE [LOCAL] DIRECTORY directory1 [ROW FORMAT row\_format] [STORED AS file\_format] (Note: Only available starting with Hive 0.11.0) SELECT ... FROM ...

**● HiveQL Manipulations-**

The Hive Query Language (HiveQL) is a query language for Hive to process and analyze structured data in a Metastore.

1. SELECT- SELECT statement is used to retrieve the data from a table.

SELECT \* FROM emp\_details;

This statement will give all the data stored in table. Some clauses are used with SELECT statement to select data based on some condition.

1. WHERE- WHERE clause works similar to a condition. It filters the data using the condition and gives you a finite result.
2. ORDER BY- ORDER BY clause is used to retrieve the details based on one column and sort the result set by ascending or descending order.
3. GROUP BY- GROUP BY clause is used to group all the records in a result set using a particular collection column. It is used to query a group of records.