**Assignment 7.3**

**Problem Statement**

**Explain the below concepts with an example in brief.**

**● Hive Data Definitions**

**● Hive Data Manipulations**

**● HiveQL Manipulations**

**HIVE DATA DEFINATIONS**

Hive data definations simple means commands *, which are used for creating, altering, and dropping databases, tables, views, functions, and indexes.*

**COMMANDS IN HIVE DATA DEFINATIONS**

Databases In Hive:-

* The Hive concept of a database is essentially just a catalog or namespace of tables.
* Databases are very useful for larger clusters with multiple teams and users, as a way of avoiding table name collisions.
* If we don’t specify a database, the default database is used.

**Syntax to create a Database:-**

hive> CREATE DATABASE financials;

**Creating Tables**

CREATE TABLE EmplSalary (

Id int,

Name string,

Salary int,

Unit string)

ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'

STORED AS TEXTFILE;

**HIVE DATA MANUPILATIONS**

HIVE data Manupilations ,DML refers to "Data Manipulation Language", a subset of SQL statements that modify the data stored in tables. Because Impala focuses on query performance and leverages the append-only nature of HDFS storage, currently Impala only supports a small set of DML statements:

* [**INSERT Statement**](http://www.cloudera.com/documentation/cdh/5-1-x/Impala/Installing-and-Using-Impala/ciiu_insert.html#insert)
* [**LOAD DATA Statement**](http://www.cloudera.com/documentation/cdh/5-1-x/Impala/Installing-and-Using-Impala/ciiu_load_data.html#load_data)

**INSERT:** It is used to insert rows in the Hive Table

INSERT INTO table college values(1,'Hmr','nlr'),(2,'TPU','vlr'),(3,'GreatLakes','chen'),(4,'IIT','del'),(5,'oxford','uk'),(6,'JNU','atp'),(7,'cambridge','us');

**LOAD DATA :** LOAD DATA moves existing data files into the directory for a table, making them immediately available for Hive queries.

LOAD DATA LOCAL INPATH '/home/cloudera/Desktop/Acadgild/Hive/Assignment\_8\_1/Emp\_Sal.txt' INTO TABLE EmplSalary ;

**HiveQL Manipulations**

## What is HiveQL(Hive Query Language)?

Hive provides a CLI to write Hive queries using Hive Query Language (HiveQL). Generally HQL syntax is similar to the[SQL](https://www.guru99.com/sql.html)syntax that most data analysts are familiar with.

Hive's SQL-inspired language separates the user from the complexity of Map Reduce programming. It reuses familiar concepts from the relational database world, such as tables, rows, columns and schema, to ease learning.

Most interactions tend to take place over a command line interface (CLI). Hive provides a CLI to write Hive queries using Hive Query Language (Hive-QL).

Generally, HiveQL syntax is similar to the[SQL](https://www.guru99.com/sql.html)syntax that most data analysts are familiar with. Hive supports four file formats those are TEXTFILE, SEQUENCEFILE, ORC and RCFILE (Record Columnar File).

* For single user metadata storage Hive uses derby database and
* For multiple user Metadata or shared Metadata case Hive uses MYSQL

## Built-in operators

Hive provides Built-in operators for Data operations to be implemented on the tables present inside Hive warehouse.

These operators are used for mathematical operations on operands, and it will return specific value as per the logic applied.

Types of Built-in Operators in HIVE are:

* Relational Operators
* Arithmetic Operators
* Logical Operators
* Operators on Complex types
* Complex type Constructors