Hive Data Definitions

Hive query language – closest to MySQL’s. the so-called *data definition language* parts of HiveQL, which are used for creating, altering, and dropping databases, tables, views, functions, and indexes. The Hive concept of a database is essentially just a *catalog* or *namespace* of tables.

DDL commands are the statements that are responsible for defining and changing the structure of a database or table in Hive by creating, deleting, or altering schema objects such as databases, tables, views, partitions, and buckets.

Example:

create table if not exists emp\_details

(

emp\_name string,

unit string,

exp int,

location string

)

row format delimited

fields terminated by ',';

CREATE EXTERNAL TABLE IF NOT EXISTS stocks (

exchange STRING,

symbol STRING,

ymd STRING,

price\_open FLOAT,

price\_high FLOAT,

price\_low FLOAT,

price\_close FLOAT,

volume INT,

price\_adj\_close FLOAT)

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

LOCATION '/data/stocks';

Hive Data Manipulations

DML (Data Manipulation Language) commands in Hive are used for inserting and querying the data from hive tables once the structure and architecture of the database has been defined such as load and insert.

Example:

LOAD DATA LOCAL INPATH '/home/acadgild/hive/dataset.txt'

INTO TABLE temperature\_data;

INSERT INTO TABLE students

  VALUES ('fred flintstone', 35, 1.28), ('barney rubble', 32, 2.32);

HiveQL Manipulations

HiveQL  is a query language for Hive to process and analyze structured data in a Metastore.

Example:

SELECT location, COUNT(\*) FROM emp\_details

GROUP BY location

HAVING AVG(exp) > 1.5;