

Hive Data Definition Language

Hive Data Definition Language (DDL) is a subset of Hive SQL statements that describe the data structure in Hive by creating, deleting, or altering schema objects such as databases, tables, views, partitions, and buckets. Most Hive DDL statements start with the keywords CREATE, DROP, or ALTER. The syntax of Hive DDL is very similar to the DDL in SQL. The comments in Hive start from –

HiveQL Data Manipulation

It is important to note that HiveQL data manipulation doesn't offer any row-level insert, update or delete operation. Therefore, data can be inserted into hive tables using either "bulk" load operations or writing the files into correct directories by other methods.

HiveQL Load Data into Managed Tables

Loading data from input file (Schema on Read)

```
Hive>LOAD DATA LOCAL INPATH '/home/hduser/sampleddata/users.txt'  
OVERWRITE INTO TABLE users;
```

'LOCAL' indicates the source data is on local file system

Local data will be copied into the final destination (HDFS file system) by Hive

If 'Local' is not specified, the file is assumed to be on HDFS

Hive does not do any data transformation while loading the data

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 using the DDL commands listed above.

Data can be loaded into Hive tables using –

LOAD command

Insert command