HIVE

Hive is a data warehouse system which is used to analyse structured data. It is built on the top of Hadoop. Hive provides the functionality of reading, writing, and managing large datasets residing in distributed storage. It runs SQL like queries called HQL (Hive query language) which gets internally converted to MapReduce jobs.

Features of Hive

These are the following features of Hive:

- Hive is fast and scalable.
- It provides SQL-like queries (i.e., HQL) that are implicitly transformed to MapReduce or Spark jobs.
- It is capable of analysing large datasets stored in HDFS.
- It allows different storage types such as plain text, RCFile, and HBase.
- It uses indexing to accelerate queries.
- It can operate on compressed data stored in the Hadoop ecosystem.
- It supports user-defined functions (UDFs) where user can provide its functionality.

Operations on Big Data Using Hive

Create Database

Create Database is a statement used to create a database in Hive.
Syntax: CREATE DATABASE <database name>

Example:

```
hive> create database demo;
OK
Time taken: 1.436 seconds
hive> show databases;
OK
default
demo
Time taken: 0.103 seconds, Fetched: 2 row(s)
```

Create Table

Create Table is a statement used to create a table in Hive.
Syntax: CREATE [TEMPORARY] [EXTERNAL] TABLE [IF NOT EXISTS] [db_name.]
table_name
[(col_name data_type [COMMENT col_comment], ...)]
[COMMENT table_comment]
[ROW FORMAT row_format]
[STORED AS file_format]

Example:

```
hive> create table iris_tab_managed(length float, width float, length1
 float, width1 float, typeof string)
    > row format delimited
    > fields terminated by ',';
OK
Time taken: 0.145 seconds
hive> show tables;
OK
emp
iris_tab_managed
Time taken: 0.042 seconds, Fetched: 2 row(s)
hive>
```

Load Data in Table

Load Data is a statement used to create a table in Hive.
Syntax: load data local inpath '<file_path>' into table <table_name>;

Example:

```
hive> load data local inpath '/home/piyushpp/Desktop/iris.csv' into ta
ble iris_tab_managed;
Loading data to table demo.iris_tab_managed
OK
Time taken: 2.029 seconds
```

Read Data from Table

We can use SQL Queries to fetch data from Table.

Example:

```
hive> select * from iris_tab_managed;
OK
NULL    NULL    NULL    NULL    typeof
5.1     3.5     1.4     0.2     setosa
4.9     3.0     1.4     0.2     setosa
4.7     3.2     1.3     0.2     setosa
4.6     3.1     1.5     0.2     setosa
5.0     3.6     1.4     0.2     setosa
5.4     3.9     1.7     0.4     setosa
4.6     3.4     1.4     0.3     setosa
```

Delete Table

Drop Table is a statement that drops the table.
Syntax: DROP TABLE [IF EXISTS] table_name;

Example:

```
hive> drop table iris_tab_managed;
OK
Time taken: 2.329 seconds
hive> show tables;
OK
emp
iris_tab_ext
Time taken: 0.05 seconds, Fetched: 2 row(s)
```

Delete Database

Drop Database is a statement that drops all the tables and deletes the database.
Syntax: DROP DATABASE database_name;

Example:

```
hive> drop database demo;
OK
Time taken: 0.515 seconds
hive> show databases;
OK
default
Time taken: 0.11 seconds, Fetched: 1 row(s)
```