**Hive Basics**

**Hive table**

Meta Data - Schema(stored in RDBMs) => by default is derby but mysql is better choice)

Actual Data (stored in hdfs => user/hive/warehouse)

When we query the table then schema is imposed (from meta data) on the data, So, we see data in tabular view=.

**Accessing hive by beeline**

|  |  |
| --- | --- |
| **beeline -u jdbc:hive2://** | Best way to work with hive interface |
| **!q (exiting from beeline)** |  |
| **beeline -u jdbc:hive2:// -e "select \* from trial.names"** | Running directly from linux console. Without entering into beeline shell. |
| **source <hive script> (might not work)** | execute hive script inside beeline shell |
| Running and creating hive script by beeline | => crete a file (hive\_script.hql) : for convention give it hql extention  => write hive commands in the :  show databases;  show tables;  select \* from names;  => use this beeline to execute this hive script:  **beeline -u jdbc:hive2:// -f /data/hive\_script.hql**  local file path |

**Important hive commands**

|  |  |
| --- | --- |
| show databases; | To see all the existing databases in hive |
| create database <database\_name>; | To create a database; |
| use <database\_name>; | To go inside in the database or using a database |
| show tables; | Too see list of tables in a database; |
| describe <table name>; | To know basic info about the table |
| describe extended <table name>; | To get more that basic info in a summarized manner |
| describe formatted <table name>; | To know detailed info about the table |

**Hive tables**

**External Tables**

* Data will not be loaded in hive warehouse. Hive will create a link with data which is stored in hdfs.
* Data is independent from hive or not managed by hive. Only meta data is managed by hive
* If you will drop/delete the hive table only meta data will be delete

**Managed/Internal tables**

* Both data and meta data managed by hive. if you will drop/delete the hive table then “meta data” and “file data” both will be deleted.

**CREATING hive tables**

**Manage / internal tables**

|  |  |
| --- | --- |
| **CREATING TABLE** | CREATE TALBE If not exists <table\_name> (  field\_name\_1 **int**,  field\_name\_2 **VARCHAR(3),**  field\_name\_3 **int**  )  row format delimited  fields terminated by ‘,' ; |

**Load data in hive table**

| load data local inpath 'file:///localfilepath' into table <table\_name>; | To copy data from local file path to hive table. This will append the new data into existing file. |
| --- | --- |
| load data local inpath 'file:///localfilepath' **overwrite** into table <table\_name>; | This will overwrite new data into existing table. (old data will be removed new data will be added) |

**Load data from local file path to hive table**

**Load data from hadoop file path to hive table**

|  |  |  |
| --- | --- | --- |
| load data inpath '/tmp/hive\_class\_2/' into table <table\_name>; | load data from hdfs to hive table warehouse.  Imp Points:   1. Whenever you will load data from hdfs path to table in managed tables. In that case data will be moved form hdfs to file. that means now that directory 'tmp/hive\_class\_2/' is empty after loading the table.      1. It will append the data in hive table.   3) To overwrite this, you would need to use "overwrite" key word.  **load data inpath 'tmp/hive\_class\_2/' overwrite into table**  **<table\_name>;**     |  | | --- | |  | |

**External Tables**

|  |  |  |
| --- | --- | --- |
| CREATE External Table:  - In external table you don't need to load data.  as we already specify the location where data is  stored in hdfs.     |  | | --- | |  | | CREATE EXTERNAL TABLE <table\_name>  (  dept\_id int,  dept\_name varchar(40),  manger\_id int,  salary int  )  row format delimited  fields terminated by ','  location 'tmp/hive\_data\_class\_2' |

**Load data from one table to another table when they have exactly same schema:**

INSERT INTO <table name> SELECT \* <another\_table\_name>