# <u>HIVE  CODES:</u>

Create database db
Show databases present in hive
Use database db to work on

```
hive> create database  if not exists db;
OK
Time taken: 0.227 seconds
hive> show databases;
OK
bda01
db
default
Time taken: 0.062 seconds, Fetched: 3 row(s)
hive> use db;
OK
Time taken: 0.044 seconds
hive>
```

Create an internal table student with the column names student id, student name ,marks ,department number

```
hive> create table if not exists student(sid int,sname string,marks int,dpno int) row format delimited fields terminated by ',';
OK
Time taken: 0.572 seconds
hive>
```

Structure of the table

```
hive> describe student;
OK
sid                     int
sname                   string
marks                   int
dpno                    int
Time taken: 0.263 seconds, Fetched: 4 row(s)
hive>
```

Inserting the data from local disk

```
hive> load data local inpath "/home/vboxuser/hadoopdata/data.csv" into table student;
Loading data to table db.student
OK
Time taken: 0.624 seconds
hive>
```

```
vboxuser@vanditha:~$ cd hadoopdata
vboxuser@vanditha:~/hadoopdata$ nano data.csv
vboxuser@vanditha:~/hadoopdata$ pwd
/home/vboxuser/hadoopdata
vboxuser@vanditha:~/hadoopdata$
```

Displays the content of the table with column names

```
Time taken: 0.624 seconds
hive> SET hive.cli.print.header=true;
hive> select * from student;
OK
student.sid     student.sname   student.marks   student.dpno
1         raj     80      12
2         ram     81      11
3         sam     88      12
4         rita    70      10
NULL      NULL    NULL    NULL
Time taken: 2.165 seconds, Fetched: 5 row(s)
hive>
```

Create external table

```
NULL    NULL    NULL    NULL
Time taken: 2.165 seconds, Fetched: 5 row(s)
hive> create external table if not exists student1(sid int,sname string,marks int,dpno int) row format delimited fields terminated by ',';
OK
Time taken: 0.086 seconds
hive> describe student1;
OK
col_name        data_type       comment
sid                  int
sname                string
marks                int
dpno                 int
Time taken: 0.076 seconds, Fetched: 4 row(s)
```

Moving the data from local disk to hdfs

```
vboxuser@vanditha:~/hadoopdata$ hdfs dfs -ls /user/data
Found 1 items
-rw-r--r--   1 vboxuser supergroup        72 2023-04-21 13:59 /user/data/empdata_pig.csv
vboxuser@vanditha:~/hadoopdata$ hdfs dfs -put "/home/vboxuser/hadoopdata/data.csv" "/user/data/data.csv"
vboxuser@vanditha:~/hadoopdata$ hdfs dfs -ls /user/data
Found 2 items
-rw-r--r--   1 vboxuser supergroup        50 2023-04-28 23:58 /user/data/data.csv
-rw-r--r--   1 vboxuser supergroup        72 2023-04-21 13:59 /user/data/empdata_pig.csv
vboxuser@vanditha:~/hadoopdata$
```

Inserting data into external table student1 from the HDFS

```
Time taken: 0.076 seconds, Fetched: 4 row(s)
hive> load data inpath "/user/data/data.csv" into table student1;
Loading data to table db.student1
OK
Time taken: 0.219 seconds
hive>
```

DEPARTMENT OF BDA

```
Time taken: 0.219 seconds
hive> select * from student1;
OK
student1.sid    student1.sname   student1.marks   student1.dpno
1       raj      80        12
2       ram      81        11
3       sam      88        12
4       rita     70        10
NULL    NULL     NULL      NULL
Time taken: 0.149 seconds, Fetched: 5 row(s)
hive>
```

Partion:

```
Time taken: 0.149 seconds, Fetched: 5 row(s)
hive> create external table if not exists ext_student(sid int,sname string,marks int) partitioned by (dpno int) row format delimited fields terminated by ',';
OK
Time tselect * from ext_student;
```

```
Time tselect * from ext_student;
OKve>
ext_student.sid ext_student.sname        ext_student.marks        ext_student.dpno
```

```
hive> set hive.exec.dynamic.partition.mode;
hive.exec.dynamic.partition.mode=strict
hive> set hive.exec.dynamic.partition.mode=unstrict;
hive>
```

Insert data to the table from existing one

```
hive> insert into table ext_student partition(dpno) select * from student;
Query ID = vboxuser_20230429001356_01a5c84a-6dd8-4b7f-92cb-0372a3334539
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
```