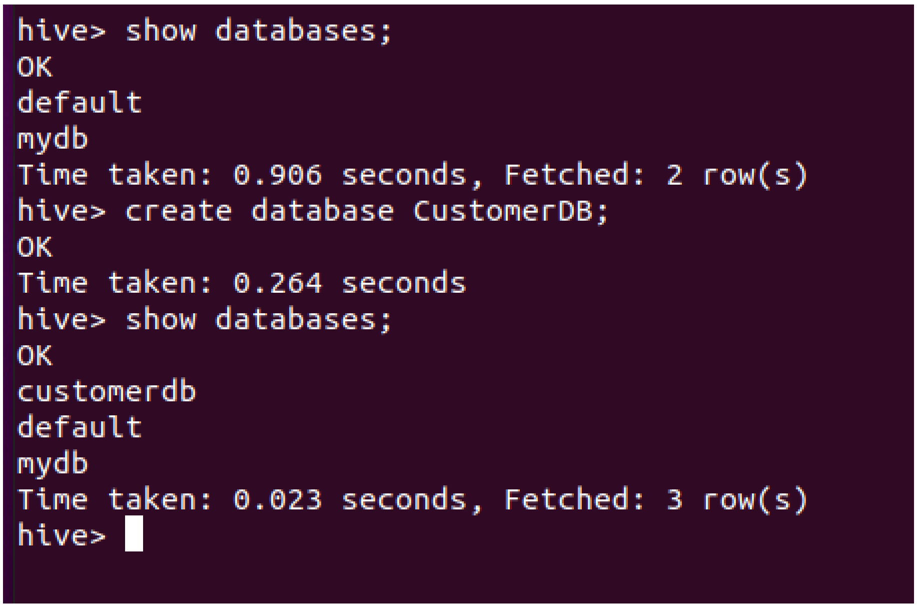
Hive Assignment

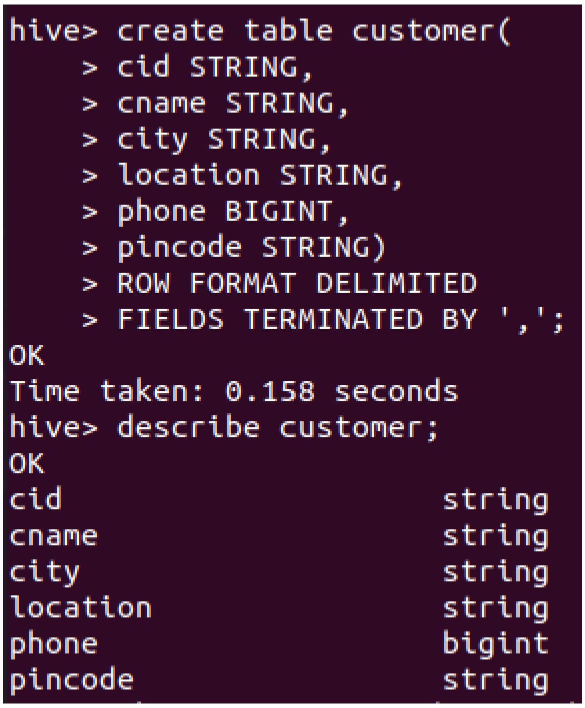
Q1) Perform the following in Hive

1. **Create a Customer Database**



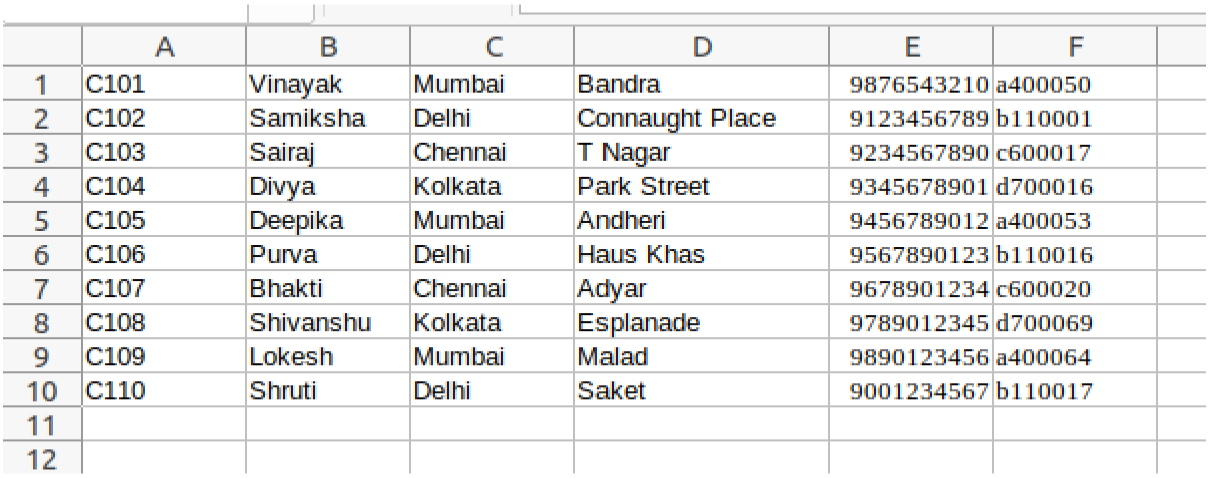
```
hive> show databases;
OK
default
mydb
Time taken: 0.906 seconds, Fetched: 2 row(s)
hive> create database CustomerDB;
OK
Time taken: 0.264 seconds
hive> show databases;
OK
customerdb
default
mydb
Time taken: 0.023 seconds, Fetched: 3 row(s)
hive> 
```

2. **Create a table called Customer_Details (cid, cname, city, location, phone, pincode)**

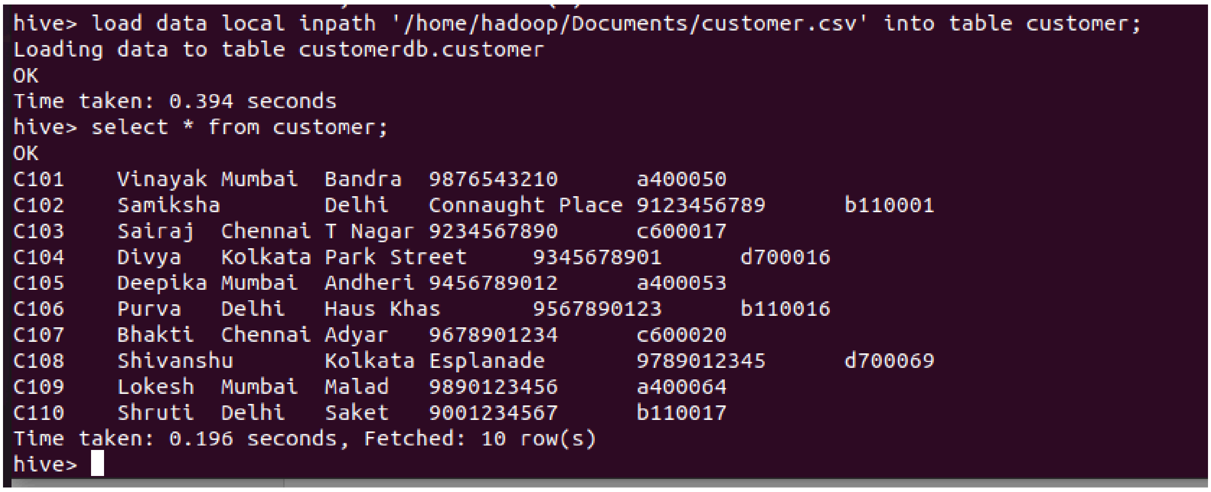


```
hive> create table customer(
    > cid STRING,
    > cname STRING,
    > city STRING,
    > location STRING,
    > phone BIGINT,
    > pincode STRING)
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY ',';
OK
Time taken: 0.158 seconds
hive> describe customer;
OK
cid                     string
cname                   string
city                    string
location                string
phone                   bigint
pincode                 string
```

3. **Insert data from a .csv file (10 records)**

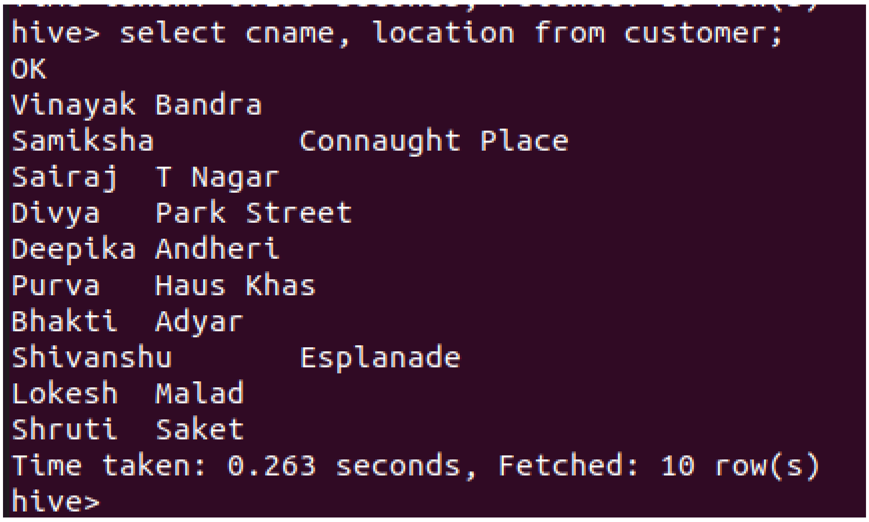


	A	B	C	D	E	F
1	C101	Vinayak	Mumbai	Bandra	9876543210	a400050
2	C102	Samiksha	Delhi	Connaught Place	9123456789	b110001
3	C103	Sairaj	Chennai	T Nagar	9234567890	c600017
4	C104	Divya	Kolkata	Park Street	9345678901	d700016
5	C105	Deepika	Mumbai	Andheri	9456789012	a400053
6	C106	Purva	Delhi	Haus Khas	9567890123	b110016
7	C107	Bhakti	Chennai	Adyar	9678901234	c600020
8	C108	Shivanshu	Kolkata	Esplanade	9789012345	d700069
9	C109	Lokesh	Mumbai	Malad	9890123456	a400064
10	C110	Shruti	Delhi	Saket	9001234567	b110017
11						
12						



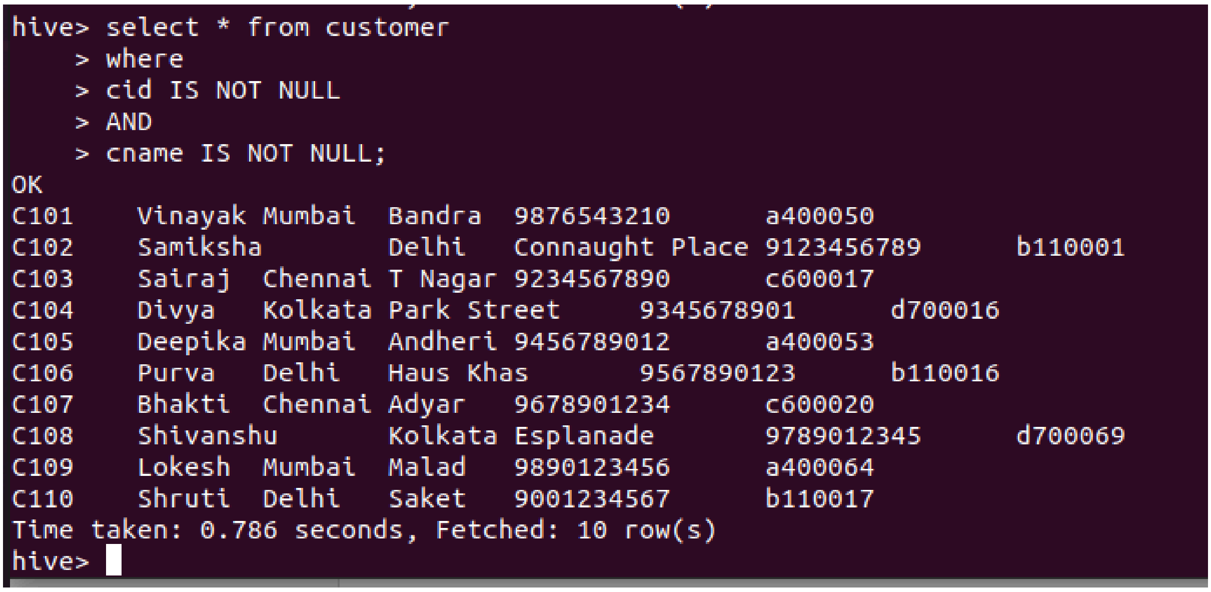
```
hive> load data local inpath '/home/hadoop/Documents/customer.csv' into table customer;
Loading data to table customerdb.customer
OK
Time taken: 0.394 seconds
hive> select * from customer;
OK
C101    Vinayak Mumbai  Bandra  9876543210      a400050
C102    Samiksha        Delhi   Connaught Place 9123456789      b110001
C103    Sairaj  Chennai T Nagar 9234567890      c600017
C104    Divya   Kolkata Park Street     9345678901      d700016
C105    Deepika Mumbai  Andheri 9456789012      a400053
C106    Purva   Delhi   Haus Khas       9567890123      b110016
C107    Bhakti  Chennai Adyar   9678901234      c600020
C108    Shivanshu       Kolkata Esplanade       9789012345      d700069
C109    Lokesh  Mumbai  Malad   9890123456      a400064
C110    Shruti  Delhi   Saket   9001234567      b110017
Time taken: 0.196 seconds, Fetched: 10 row(s)
hive> 
```

4. **Write a query to display all the customer name and location**



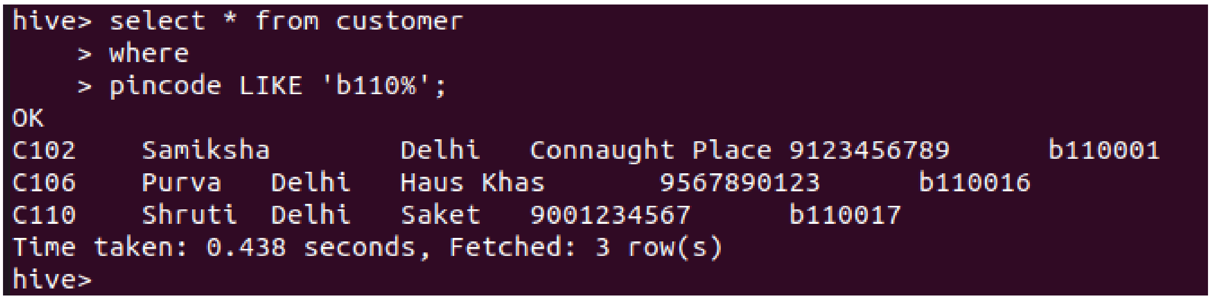
```
hive> select cname, location from customer;
OK
Vinayak Bandra
Samiksha        Connaught Place
Sairaj  T Nagar
Divya   Park Street
Deepika Andheri
Purva   Haus Khas
Bhakti  Adyar
Shivanshu       Esplanade
Lokesh  Malad
Shruti  Saket
Time taken: 0.263 seconds, Fetched: 10 row(s)
hive>
```

5. **Display all information where customer cust_code and cust_name**



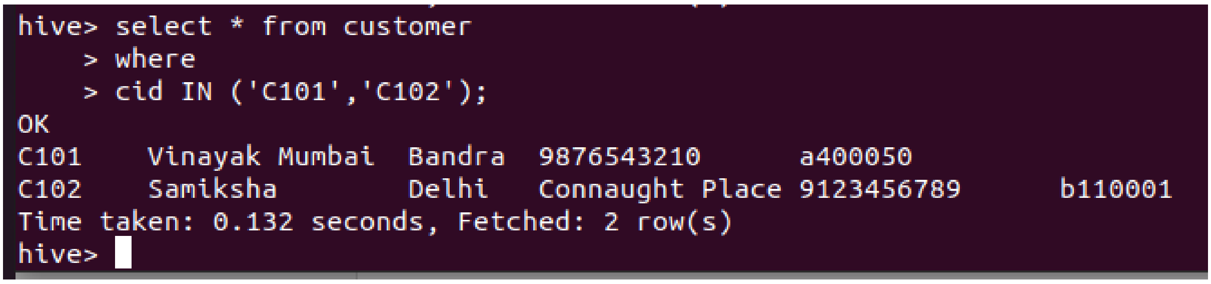
```
hive> select * from customer
    > where
    > cid IS NOT NULL
    > AND
    > cname IS NOT NULL;
OK
C101    Vinayak Mumbai  Bandra  9876543210      a400050
C102    Samiksha        Delhi   Connaught Place 9123456789      b110001
C103    Sairaj  Chennai T Nagar 9234567890      c600017
C104    Divya   Kolkata Park Street     9345678901      d700016
C105    Deepika Mumbai  Andheri 9456789012      a400053
C106    Purva   Delhi   Haus Khas       9567890123      b110016
C107    Bhakti  Chennai Adyar   9678901234      c600020
C108    Shivanshu       Kolkata Esplanade       9789012345      d700069
C109    Lokesh  Mumbai  Malad   9890123456      a400064
C110    Shruti  Delhi   Saket   9001234567      b110017
Time taken: 0.786 seconds, Fetched: 10 row(s)
hive>
```

6. **Display the customer information for area_code=a101**



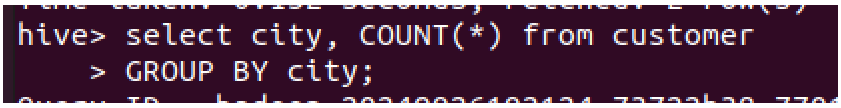
```
hive> select * from customer
    > where
    > pincode LIKE 'b110%';
OK
C102    Samiksha        Delhi   Connaught Place 9123456789      b110001
C106    Purva   Delhi   Haus Khas       9567890123      b110016
C110    Shruti  Delhi   Saket   9001234567      b110017
Time taken: 0.438 seconds, Fetched: 3 row(s)
hive>
```

7. **Display customer details where cid is ‘C101’ or ‘C201’**

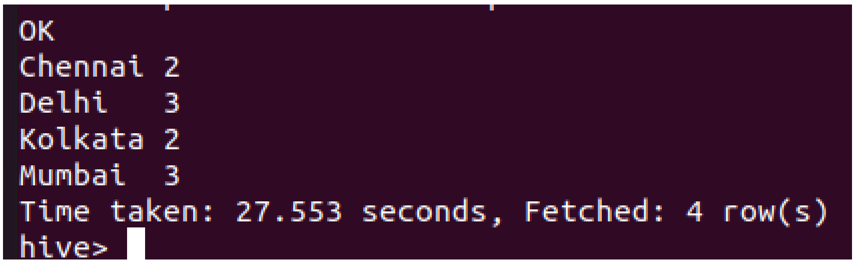


```
hive> select * from customer
    > where
    > cid IN ('C101','C102');
OK
C101    Vinayak Mumbai  Bandra  9876543210      a400050
C102    Samiksha        Delhi   Connaught Place 9123456789      b110001
Time taken: 0.132 seconds, Fetched: 2 row(s)
hive>
```

8. **Display the city wise customer count.**

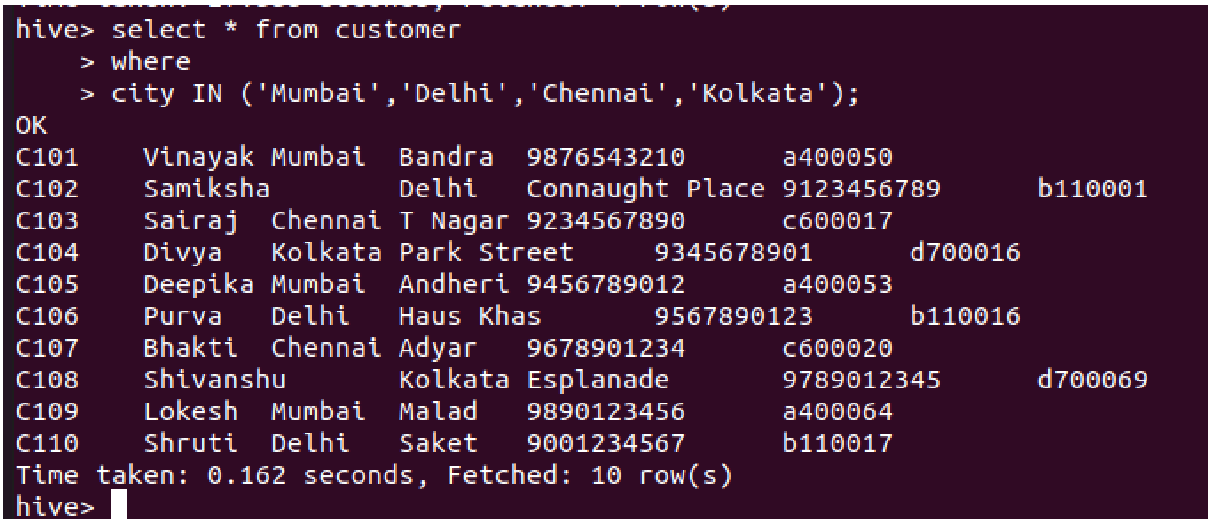


```
hive> select city, COUNT(*) from customer
    > GROUP BY city;
```



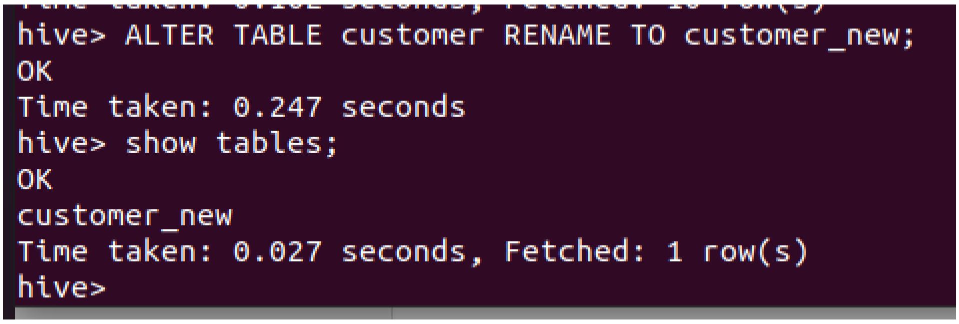
```
OK
Chennai 2
Delhi   3
Kolkata 2
Mumbai  3
Time taken: 27.553 seconds, Fetched: 4 row(s)
hive>
```

9. **Display the customers from city Mumbai, Delhi , Chennai or Kolkata**



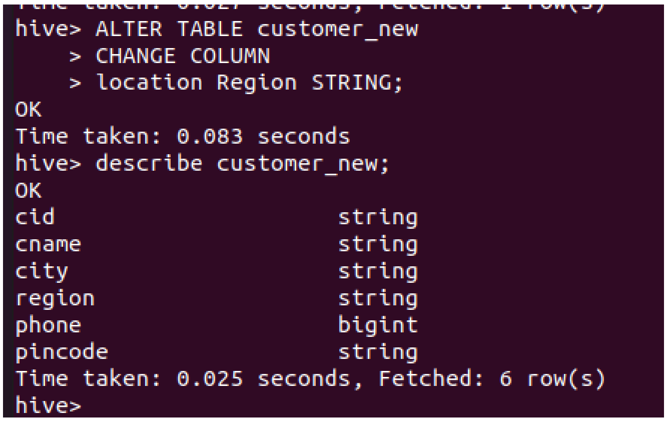
```
hive> select * from customer
    > where
    > city IN ('Mumbai','Delhi','Chennai','Kolkata');
OK
C101    Vinayak Mumbai  Bandra  9876543210      a400050
C102    Samiksha        Delhi   Connaught Place 9123456789      b110001
C103    Sairaj  Chennai T Nagar 9234567890      c600017
C104    Divya   Kolkata Park Street     9345678901      d700016
C105    Deepika Mumbai  Andheri 9456789012      a400053
C106    Purva   Delhi   Haus Khas       9567890123      b110016
C107    Bhakti  Chennai Adyar   9678901234      c600020
C108    Shivanshu       Kolkata Esplanade       9789012345      d700069
C109    Lokesh  Mumbai  Malad   9890123456      a400064
C110    Shruti  Delhi   Saket   9001234567      b110017
Time taken: 0.162 seconds, Fetched: 10 row(s)
hive>
```

10. **Rename the table to customer_New**



```
hive> ALTER TABLE customer RENAME TO customer_new;
OK
Time taken: 0.247 seconds
hive> show tables;
OK
customer_new
Time taken: 0.027 seconds, Fetched: 1 row(s)
hive>
```

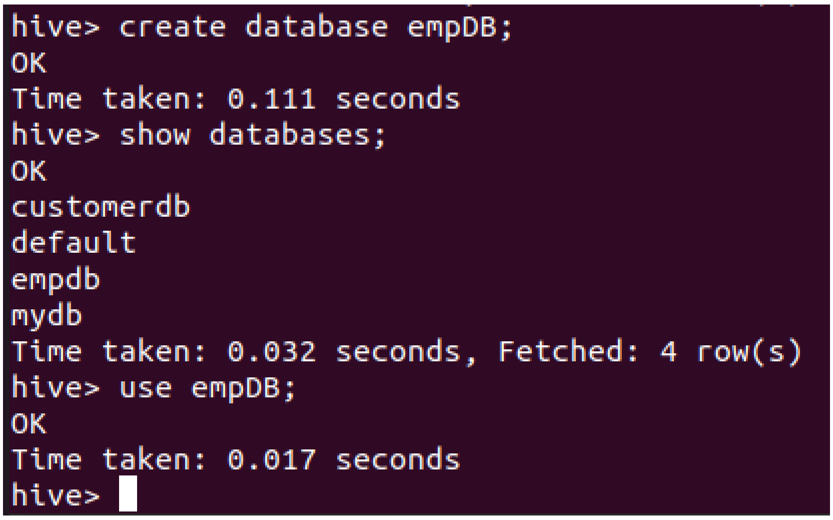
11. **Rename the column ‘location’ to ‘Region’**



```
hive> ALTER TABLE customer_new
    > CHANGE COLUMN
    > location Region STRING;
OK
Time taken: 0.083 seconds
hive> describe customer_new;
OK
cid                     string
cname                   string
city                    string
region                  string
phone                   bigint
pincode                 string
Time taken: 0.025 seconds, Fetched: 6 row(s)
hive>
```

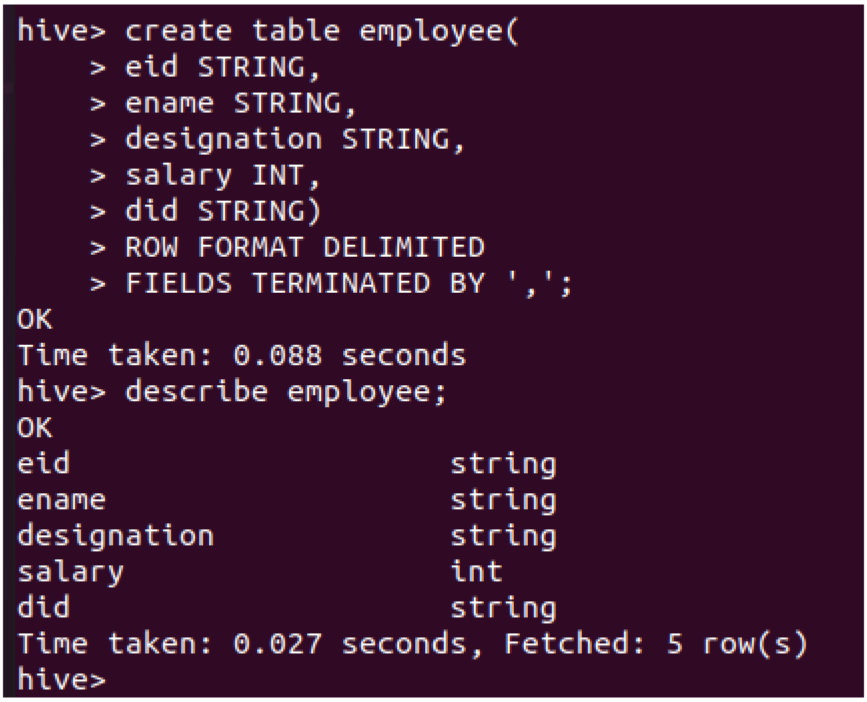
Q2) Perform the following in HIVE:

1. **Create a Emp Database**



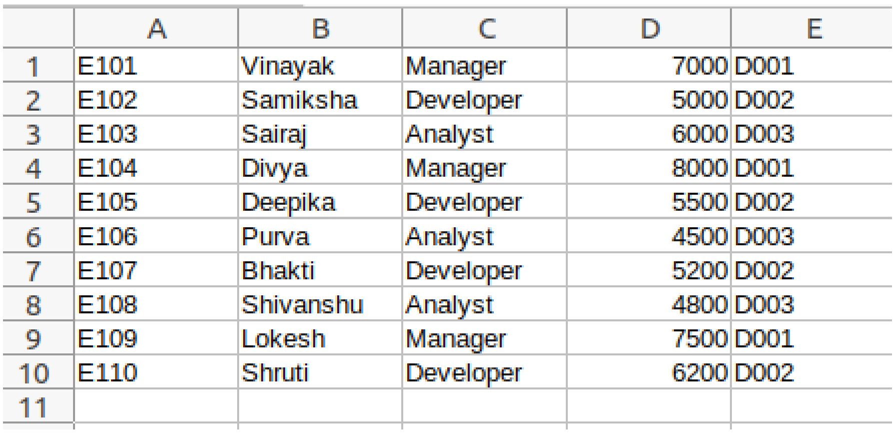
```
hive> create database empDB;
OK
Time taken: 0.111 seconds
hive> show databases;
OK
customerdb
default
empdb
mydb
Time taken: 0.032 seconds, Fetched: 4 row(s)
hive> use empDB;
OK
Time taken: 0.017 seconds
hive> 
```

2. **Create a table called ‘Employee’**

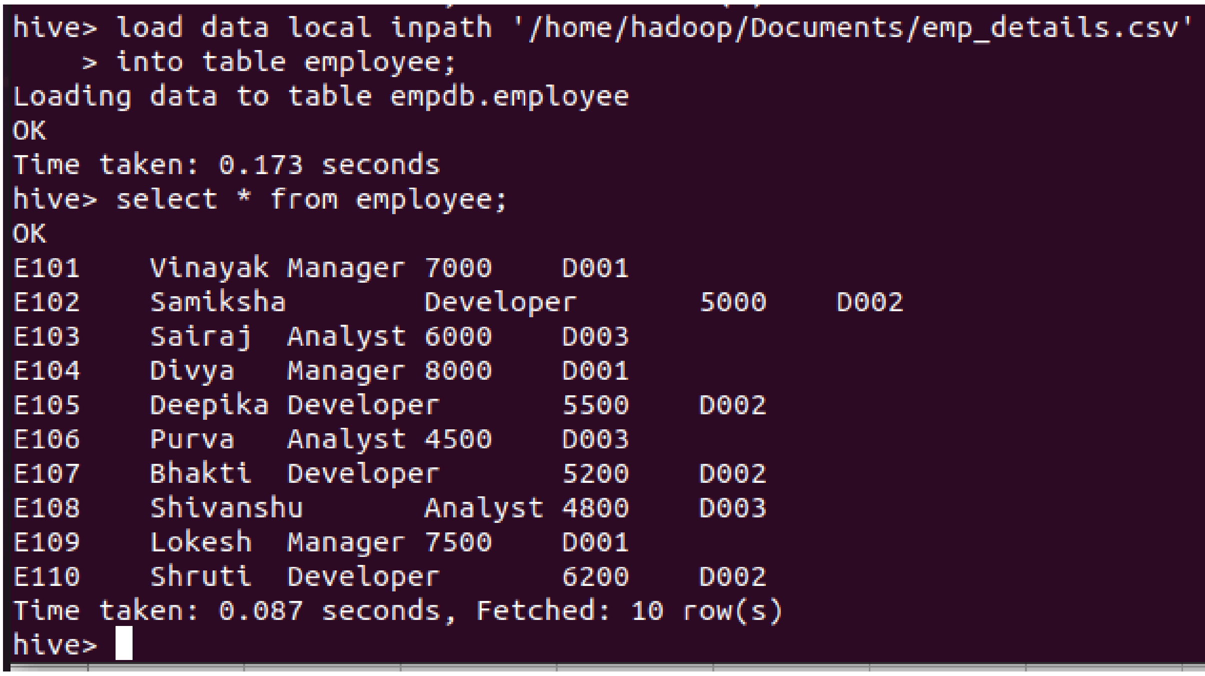


```
hive> create table employee(
    > eid STRING,
    > ename STRING,
    > designation STRING,
    > salary INT,
    > did STRING)
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY ',';
OK
Time taken: 0.088 seconds
hive> describe employee;
OK
eid                     string
ename                   string
designation             string
salary                  int
did                     string
Time taken: 0.027 seconds, Fetched: 5 row(s)
hive>
```

3. **Add 10 employees who have joined the company to the database. (eid,ename,designation,salary,did)**

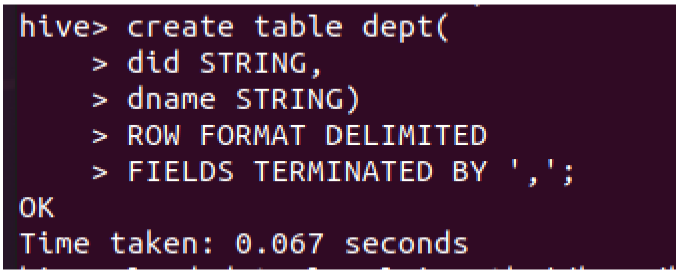


	A	B	C	D	E
1	E101	Vinayak	Manager	7000	D001
2	E102	Samiksha	Developer	5000	D002
3	E103	Sairaj	Analyst	6000	D003
4	E104	Divya	Manager	8000	D001
5	E105	Deepika	Developer	5500	D002
6	E106	Purva	Analyst	4500	D003
7	E107	Bhakti	Developer	5200	D002
8	E108	Shivanshu	Analyst	4800	D003
9	E109	Lokesh	Manager	7500	D001
10	E110	Shruti	Developer	6200	D002
11					



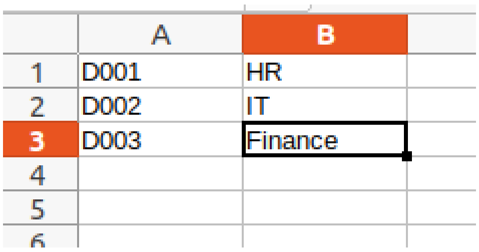
```
hive> load data local inpath '/home/hadoop/Documents/emp_details.csv'
    > into table employee;
Loading data to table empdb.employee
OK
Time taken: 0.173 seconds
hive> select * from employee;
OK
E101    Vinayak Manager 7000    D001
E102    Samiksha        Developer       5000    D002
E103    Sairaj  Analyst 6000    D003
E104    Divya   Manager 8000    D001
E105    Deepika Developer       5500    D002
E106    Purva   Analyst 4500    D003
E107    Bhakti  Developer       5200    D002
E108    Shivanshu       Analyst 4800    D003
E109    Lokesh  Manager 7500    D001
E110    Shruti  Developer       6200    D002
Time taken: 0.087 seconds, Fetched: 10 row(s)
hive> 
```

4. **Create Dept employee (did, dname)**

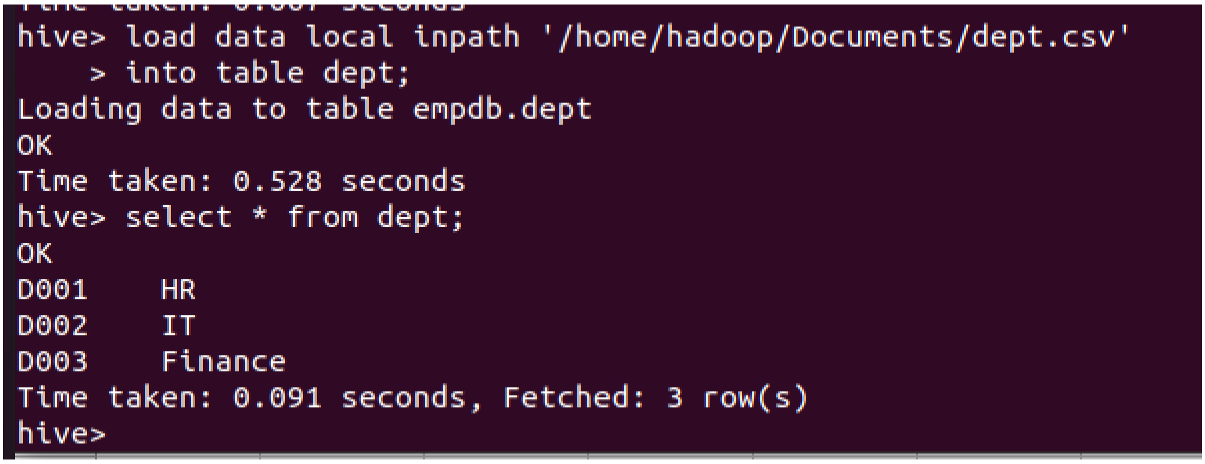


```
hive> create table dept(
    > did STRING,
    > dname STRING)
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY ',';
OK
Time taken: 0.067 seconds
```

5. Insert details of 3 departments

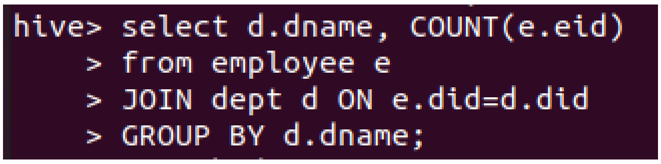


	A	B
1	D001	HR
2	D002	IT
3	D003	Finance
4		
5		
6		

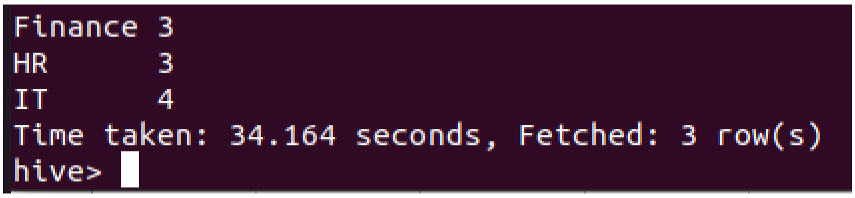


```
hive> load data local inpath '/home/hadoop/Documents/dept.csv'
    > into table dept;
Loading data to table empdb.dept
OK
Time taken: 0.528 seconds
hive> select * from dept;
OK
D001    HR
D002    IT
D003    Finance
Time taken: 0.091 seconds, Fetched: 3 row(s)
hive>
```

6. Display the department wise employee count.

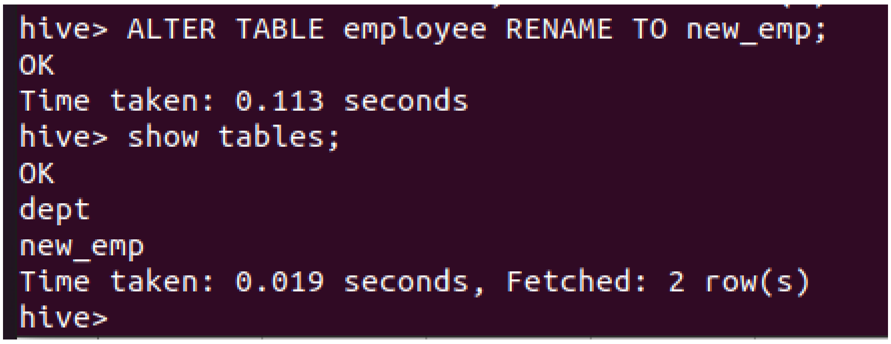


```
hive> select d.dname, COUNT(e.eid)
    > from employee e
    > JOIN dept d ON e.did=d.did
    > GROUP BY d.dname;
```



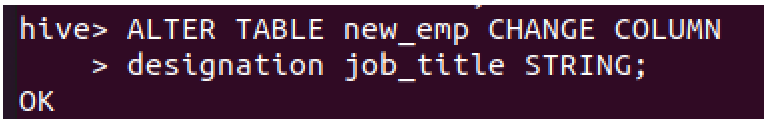
```
Finance 3
HR      3
IT      4
Time taken: 34.164 seconds, Fetched: 3 row(s)
hive> 
```

7. Rename the table Employee to new_emp

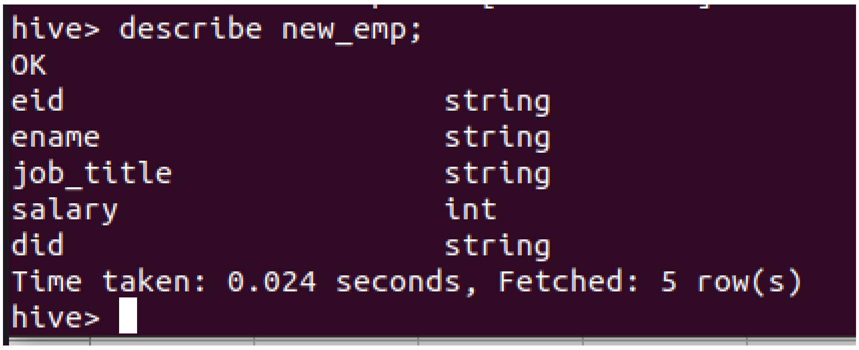


```
hive> ALTER TABLE employee RENAME TO new_emp;
OK
Time taken: 0.113 seconds
hive> show tables;
OK
dept
new_emp
Time taken: 0.019 seconds, Fetched: 2 row(s)
hive>
```

8. **Rename the column ‘designation to ‘job_title’**

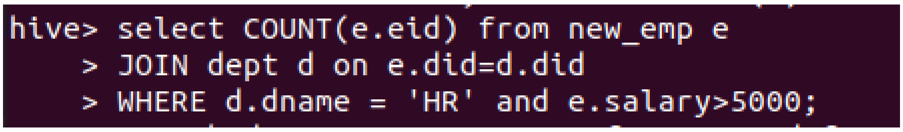


```
hive> ALTER TABLE new_emp CHANGE COLUMN
    > designation job_title STRING;
OK
```

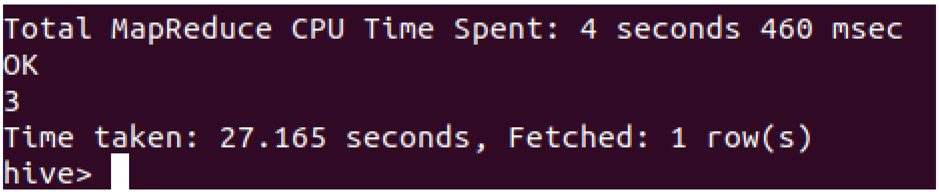


```
hive> describe new_emp;
OK
eid                     string
ename                   string
job_title               string
salary                  int
did                     string
Time taken: 0.024 seconds, Fetched: 5 row(s)
hive>
```

9. **Display the number of employees present in “HR” dept and salary greater than 5000.**



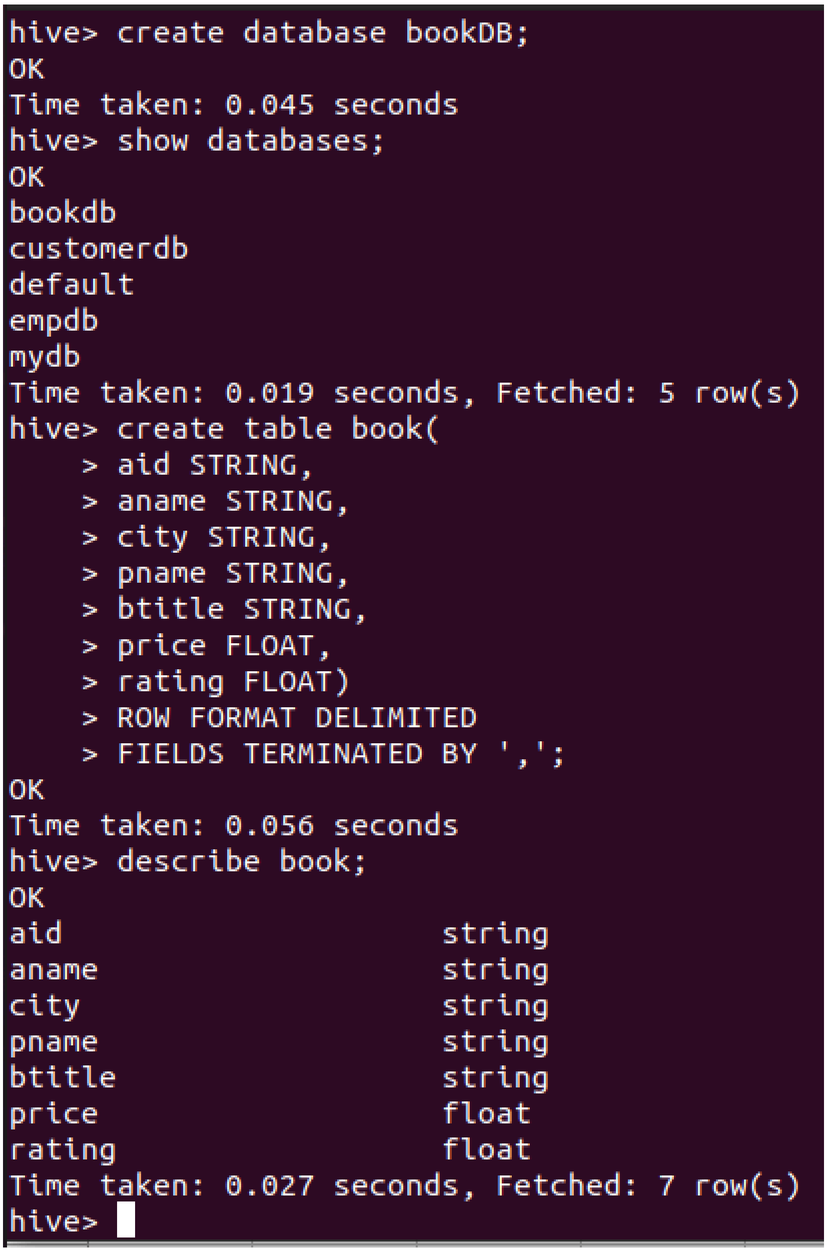
```
hive> select COUNT(e.eid) from new_emp e
    > JOIN dept d on e.did=d.did
    > WHERE d.dname = 'HR' and e.salary>5000;
```



```
Total MapReduce CPU Time Spent: 4 seconds 460 msec
OK
3
Time taken: 27.165 seconds, Fetched: 1 row(s)
hive>
```

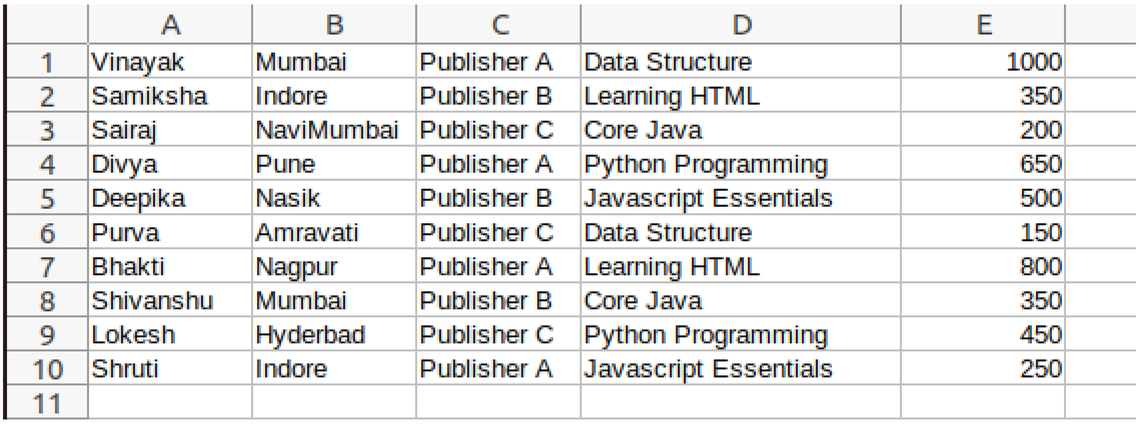
Q3) Implement in HIVE

1. **Create a table Book (aid,aname,city,pname,btitle,price,rating)**

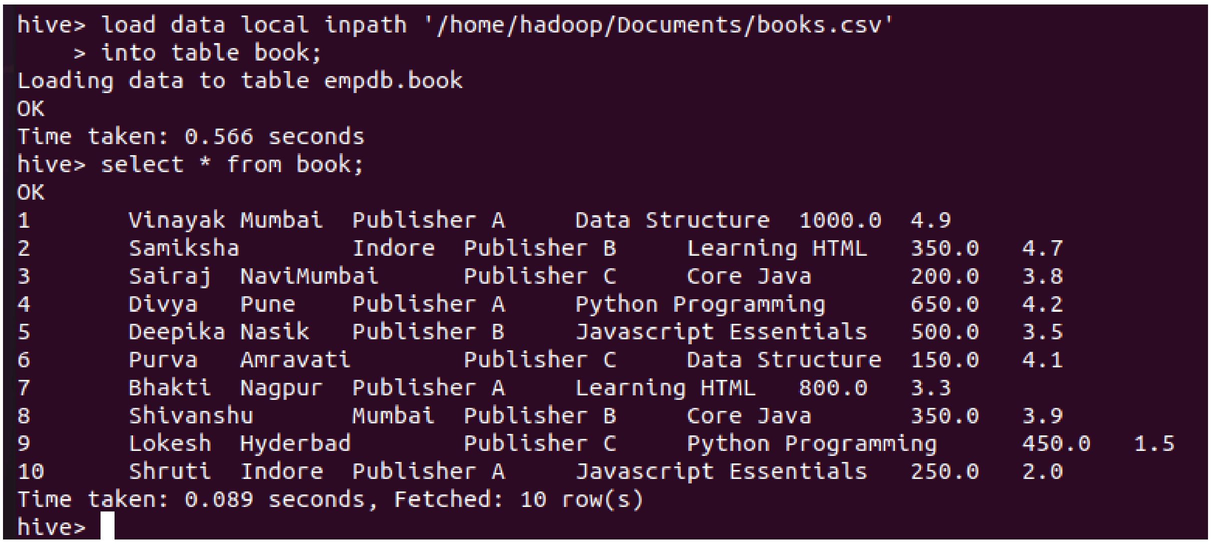


```
hive> create database bookDB;
OK
Time taken: 0.045 seconds
hive> show databases;
OK
bookdb
customerdb
default
empdb
mydb
Time taken: 0.019 seconds, Fetched: 5 row(s)
hive> create table book(
    > aid STRING,
    > aname STRING,
    > city STRING,
    > pname STRING,
    > btitle STRING,
    > price FLOAT,
    > rating FLOAT)
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY ',';
OK
Time taken: 0.056 seconds
hive> describe book;
OK
aid                     string
aname                   string
city                    string
pname                   string
btitle                  string
price                   float
rating                  float
Time taken: 0.027 seconds, Fetched: 7 row(s)
hive> 
```

2. Load the data from a .csv file

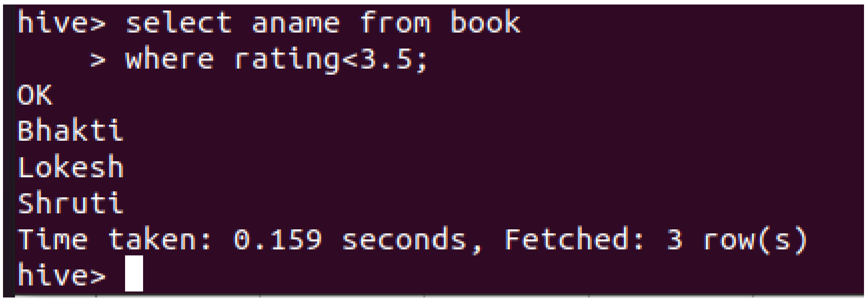


	A	B	C	D	E	
1	Vinayak	Mumbai	Publisher A	Data Structure	1000	
2	Samiksha	Indore	Publisher B	Learning HTML	350	
3	Sairaj	NaviMumbai	Publisher C	Core Java	200	
4	Divya	Pune	Publisher A	Python Programming	650	
5	Deepika	Nasik	Publisher B	Javascript Essentials	500	
6	Purva	Amravati	Publisher C	Data Structure	150	
7	Bhakti	Nagpur	Publisher A	Learning HTML	800	
8	Shivanshu	Mumbai	Publisher B	Core Java	350	
9	Lokesh	Hyderbad	Publisher C	Python Programming	450	
10	Shruti	Indore	Publisher A	Javascript Essentials	250	
11						



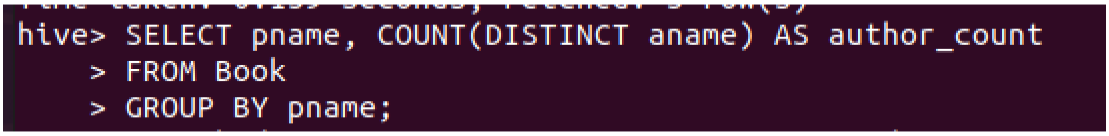
```
hive> load data local inpath '/home/hadoop/Documents/books.csv'
    > into table book;
Loading data to table empdb.book
OK
Time taken: 0.566 seconds
hive> select * from book;
OK
1       Vinayak Mumbai  Publisher A     Data Structure  1000.0  4.9
2       Samiksha        Indore  Publisher B     Learning HTML   350.0   4.7
3       Sairaj  NaviMumbai      Publisher C     Core Java       200.0   3.8
4       Divya   Pune    Publisher A     Python Programming      650.0   4.2
5       Deepika Nasik   Publisher B     Javascript Essentials   500.0   3.5
6       Purva   Amravati        Publisher C     Data Structure  150.0   4.1
7       Bhakti  Nagpur  Publisher A     Learning HTML   800.0   3.3
8       Shivanshu       Mumbai  Publisher B     Core Java       350.0   3.9
9       Lokesh  Hyderbad        Publisher C     Python Programming      450.0   1.5
10      Shruti  Indore  Publisher A     Javascript Essentials   250.0   2.0
Time taken: 0.089 seconds, Fetched: 10 row(s)
hive> 
```

3. Display the name of author whose Rating is less than 2

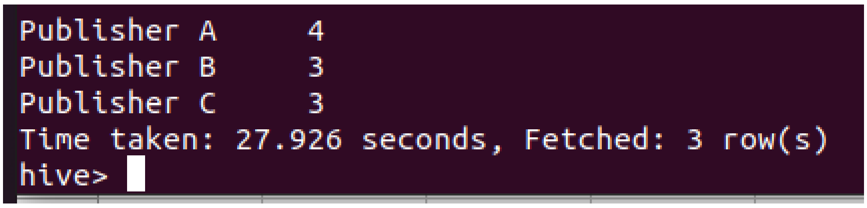


```
hive> select aname from book
    > where rating<3.5;
OK
Bhakti
Lokesh
Shruti
Time taken: 0.159 seconds, Fetched: 3 row(s)
hive> 
```

4. Display the publisher wise count of authors

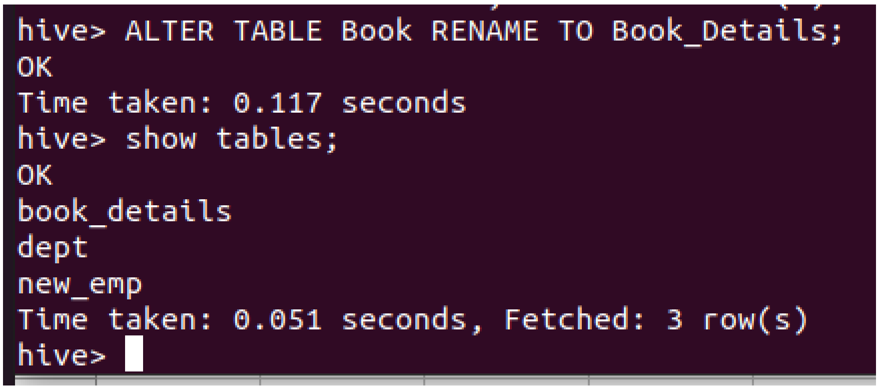


```
Time taken: 0.159 seconds, Fetched: 3 row(s)
hive> SELECT pname, COUNT(DISTINCT aname) AS author_count
    > FROM Book
    > GROUP BY pname;
```



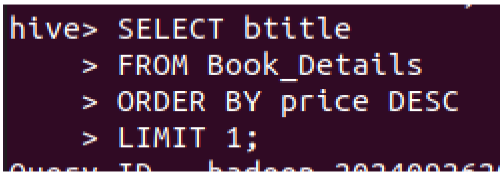
```
Publisher A     4
Publisher B     3
Publisher C     3
Time taken: 27.926 seconds, Fetched: 3 row(s)
hive>
```

5. **Rename the table to Book_Details**

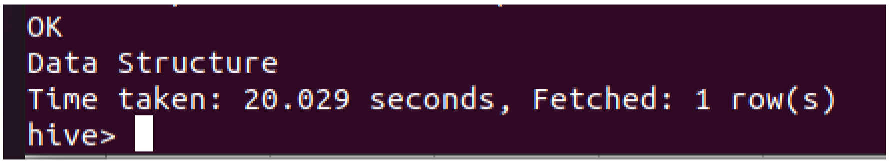


```
hive> ALTER TABLE Book RENAME TO Book_Details;
OK
Time taken: 0.117 seconds
hive> show tables;
OK
book_details
dept
new_emp
Time taken: 0.051 seconds, Fetched: 3 row(s)
hive>
```

6. **Display the name of the book having the highest price.**

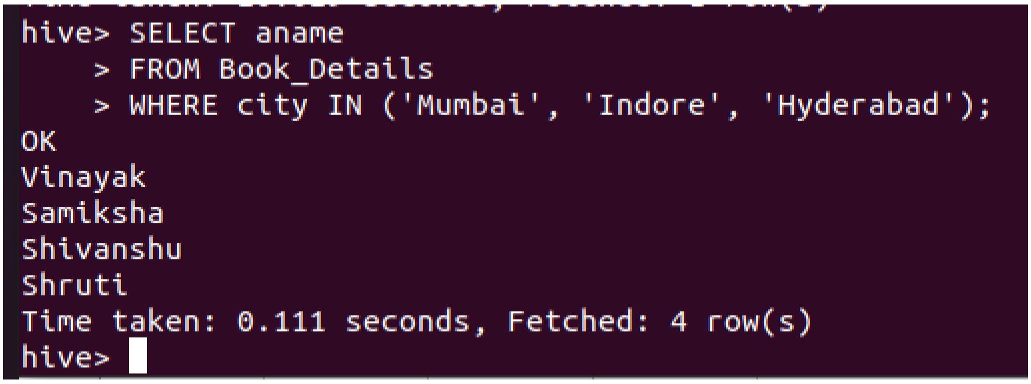


```
hive> SELECT btitle
    > FROM Book_Details
    > ORDER BY price DESC
    > LIMIT 1;
```



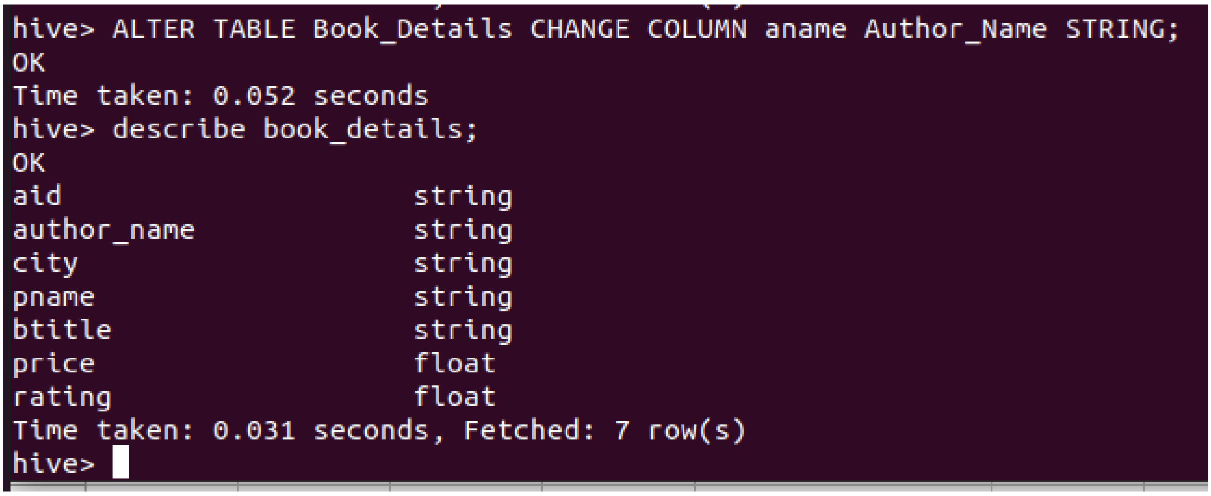
```
OK
Data Structure
Time taken: 20.029 seconds, Fetched: 1 row(s)
hive>
```

7. **Display the authors from city Mumbai, Delhi or Chennai**



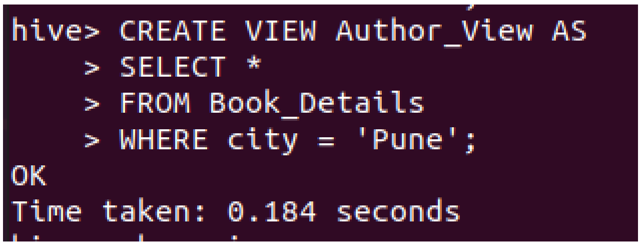
```
hive> SELECT aname
    > FROM Book_Details
    > WHERE city IN ('Mumbai', 'Indore', 'Hyderabad');
OK
Vinayak
Samiksha
Shivanshu
Shruti
Time taken: 0.111 seconds, Fetched: 4 row(s)
hive>
```

8. **Rename the column ‘aname’ to ‘Author_Name’**



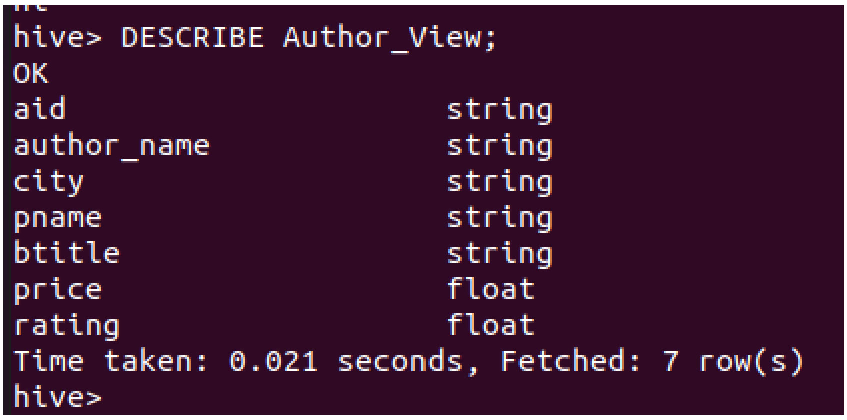
```
hive> ALTER TABLE Book_Details CHANGE COLUMN aname Author_Name STRING;
OK
Time taken: 0.052 seconds
hive> describe book_details;
OK
aid                     string
author_name             string
city                    string
pname                   string
btitle                  string
price                   float
rating                  float
Time taken: 0.031 seconds, Fetched: 7 row(s)
hive> 
```

9. **Create a view Author_View for all the authors in the city Pune**



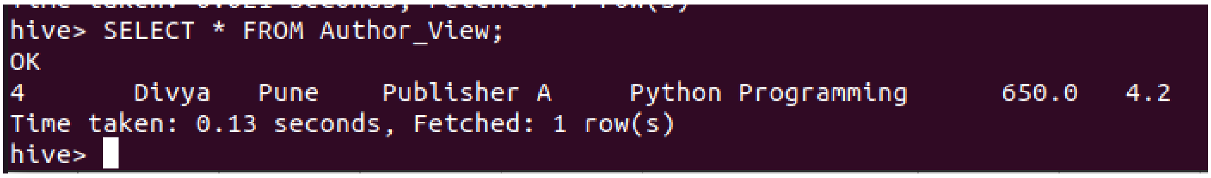
```
hive> CREATE VIEW Author_View AS
    > SELECT *
    > FROM Book_Details
    > WHERE city = 'Pune';
OK
Time taken: 0.184 seconds
```

10. **Describe the view.**



```
hive> DESCRIBE Author_View;
OK
aid                     string
author_name             string
city                    string
pname                   string
btitle                  string
price                   float
rating                  float
Time taken: 0.021 seconds, Fetched: 7 row(s)
hive>
```

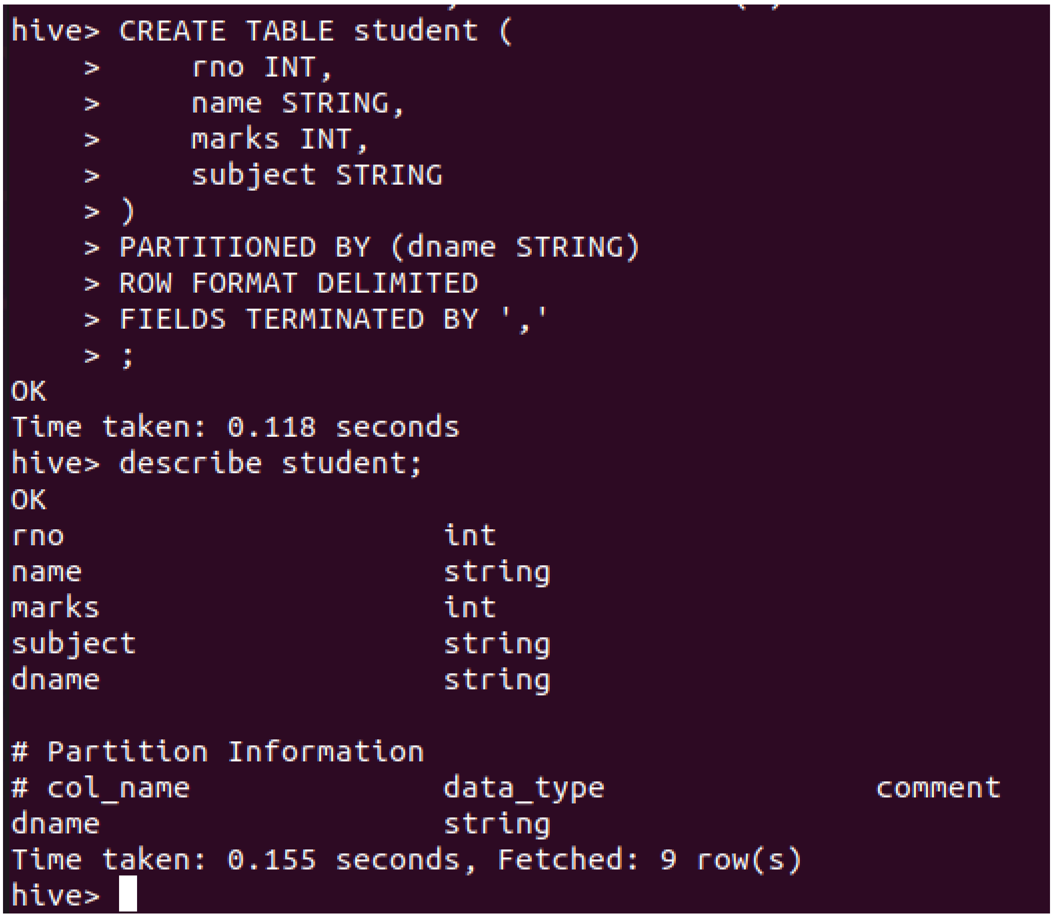
11. **Display the contents of the view.**



```
hive> SELECT * FROM Author_View;
OK
4       Divya   Pune    Publisher A     Python Programming      650.0   4.2
Time taken: 0.13 seconds, Fetched: 1 row(s)
hive> 
```

Q4) Implement in HIVE Partition

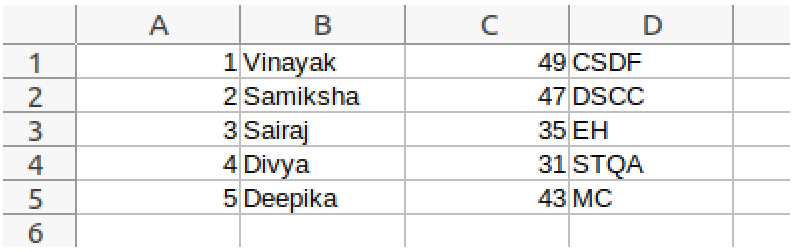
1. **Create a table student with a partition on dname with the details (rno,name,marks,subject,dname)**



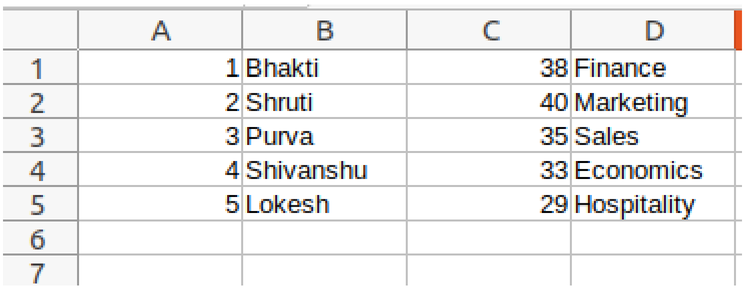
```
hive> CREATE TABLE student (
    >     rno INT,
    >     name STRING,
    >     marks INT,
    >     subject STRING
    > )
    > PARTITIONED BY (dname STRING)
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY ','
    > ;
OK
Time taken: 0.118 seconds
hive> describe student;
OK
rno                     int
name                    string
marks                   int
subject                 string
dname                   string

# Partition Information
# col_name              data_type               comment
dname                   string
Time taken: 0.155 seconds, Fetched: 9 row(s)
hive> 
```

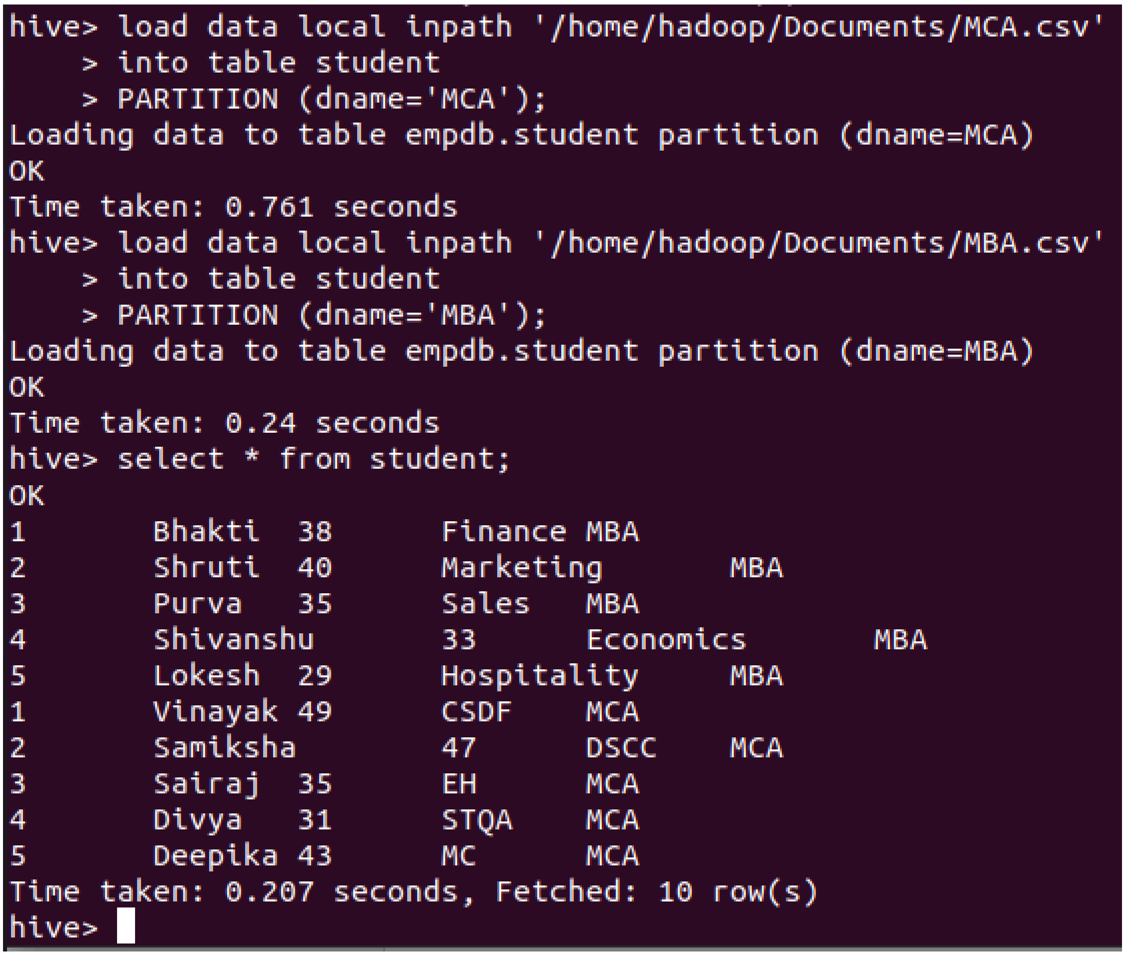
2. **Insert data for 5 students for two departments MCA and MBA**



	A	B	C	D
1	1	Vinayak	49	CSDF
2	2	Samiksha	47	DSCC
3	3	Sairaj	35	EH
4	4	Divya	31	STQA
5	5	Deepika	43	MC
6				

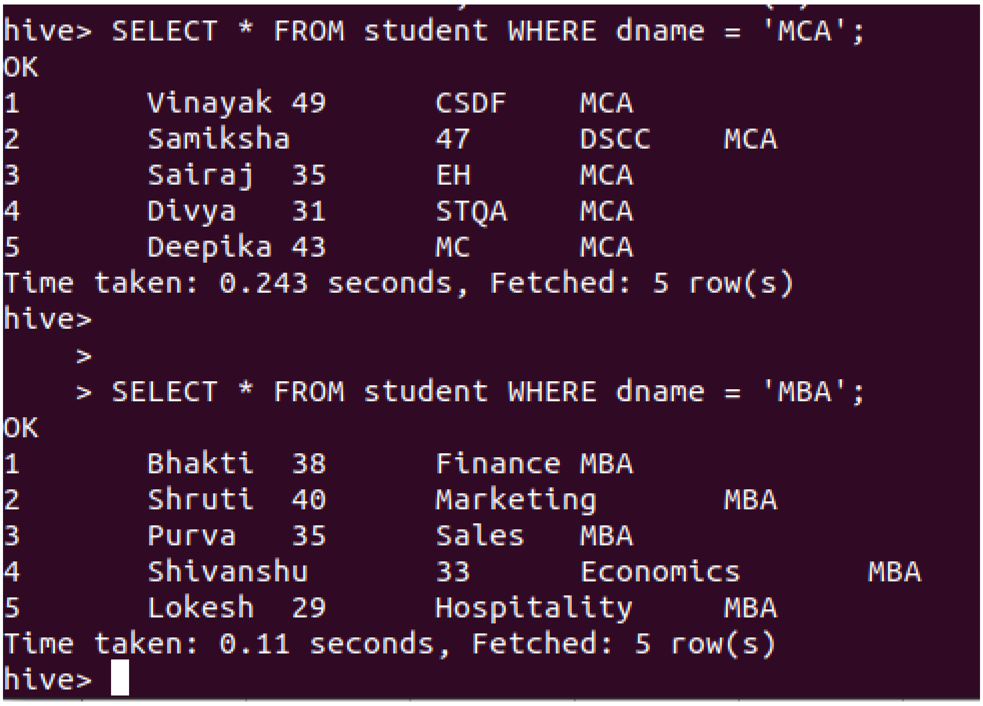


	A	B	C	D
1	1	Bhakti	38	Finance
2	2	Shruti	40	Marketing
3	3	Purva	35	Sales
4	4	Shivanshu	33	Economics
5	5	Lokesh	29	Hospitality
6				
7				



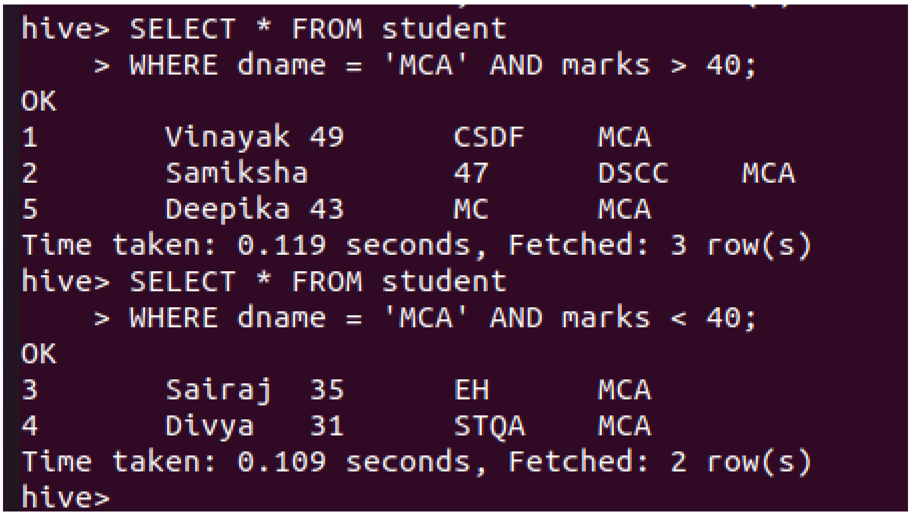
```
hive> load data local inpath '/home/hadoop/Documents/MCA.csv'
    > into table student
    > PARTITION (dname='MCA');
Loading data to table empdb.student partition (dname=MCA)
OK
Time taken: 0.761 seconds
hive> load data local inpath '/home/hadoop/Documents/MBA.csv'
    > into table student
    > PARTITION (dname='MBA');
Loading data to table empdb.student partition (dname=MBA)
OK
Time taken: 0.24 seconds
hive> select * from student;
OK
1       Bhakti  38      Finance MBA
2       Shruti  40      Marketing       MBA
3       Purva   35      Sales   MBA
4       Shivanshu       33      Economics       MBA
5       Lokesh  29      Hospitality     MBA
1       Vinayak 49      CSDF    MCA
2       Samiksha        47      DSCC    MCA
3       Sairaj  35      EH      MCA
4       Divya   31      STQA    MCA
5       Deepika 43      MC      MCA
Time taken: 0.207 seconds, Fetched: 10 row(s)
hive> 
```

3. **Display the contents of each partition.**



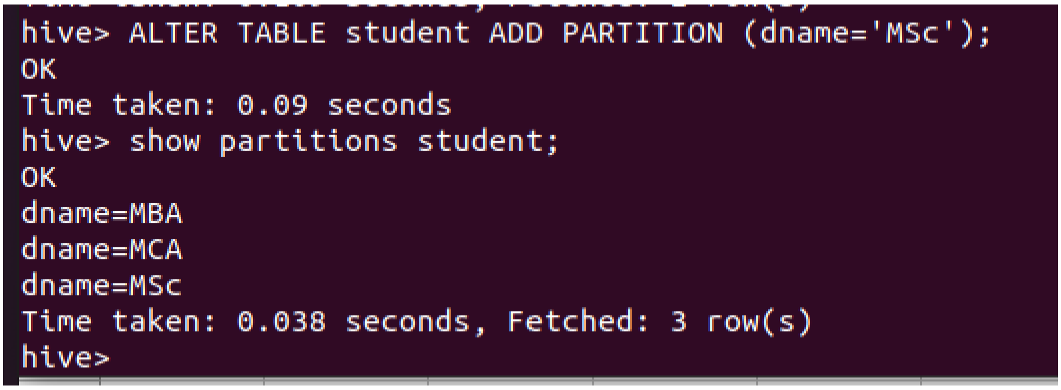
```
hive> SELECT * FROM student WHERE dname = 'MCA';
OK
1       Vinayak 49      CSDF    MCA
2       Samiksha        47      DSCC    MCA
3       Sairaj  35      EH      MCA
4       Divya   31      STQA    MCA
5       Deepika 43      MC      MCA
Time taken: 0.243 seconds, Fetched: 5 row(s)
hive>
    >
    > SELECT * FROM student WHERE dname = 'MBA';
OK
1       Bhakti  38      Finance MBA
2       Shruti  40      Marketing       MBA
3       Purva   35      Sales   MBA
4       Shivanshu       33      Economics       MBA
5       Lokesh  29      Hospitality     MBA
Time taken: 0.11 seconds, Fetched: 5 row(s)
hive> 
```

4. **Display students having marks more than 60 in MCA department.**



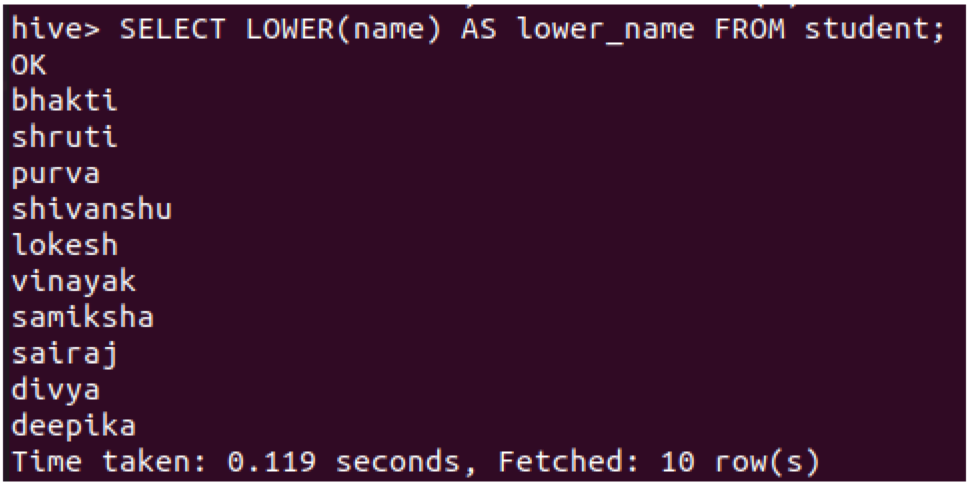
```
hive> SELECT * FROM student
    > WHERE dname = 'MCA' AND marks > 40;
OK
1       Vinayak 49      CSDF    MCA
2       Samiksha        47      DSCC    MCA
5       Deepika 43      MC      MCA
Time taken: 0.119 seconds, Fetched: 3 row(s)
hive> SELECT * FROM student
    > WHERE dname = 'MCA' AND marks < 40;
OK
3       Sairaj  35      EH      MCA
4       Divya   31      STQA    MCA
Time taken: 0.109 seconds, Fetched: 2 row(s)
hive>
```

5. **Add a new partition for the department MSc**

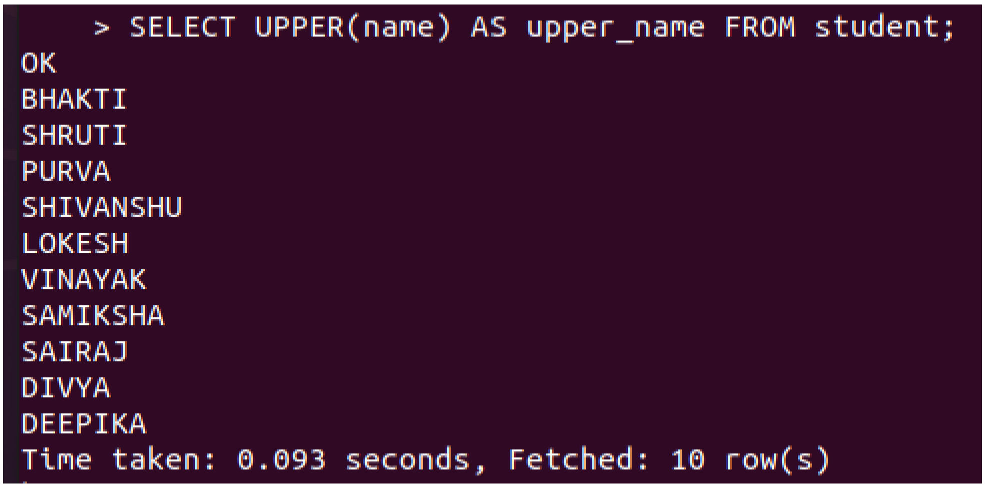


```
hive> ALTER TABLE student ADD PARTITION (dname='MSc');
OK
Time taken: 0.09 seconds
hive> show partitions student;
OK
dname=MBA
dname=MCA
dname=MSc
Time taken: 0.038 seconds, Fetched: 3 row(s)
hive>
```

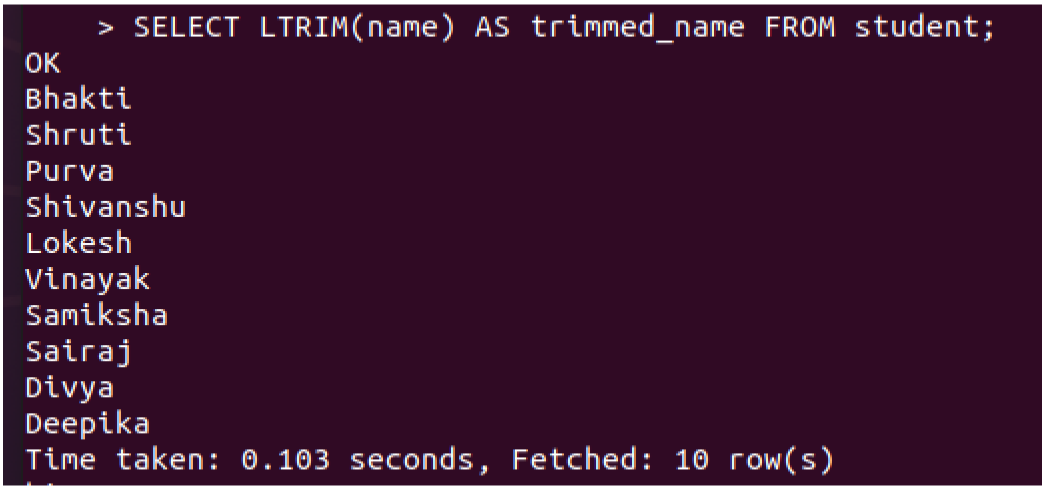
6. **Perform the following built in functions – lower, upper, ltrim,reverse**



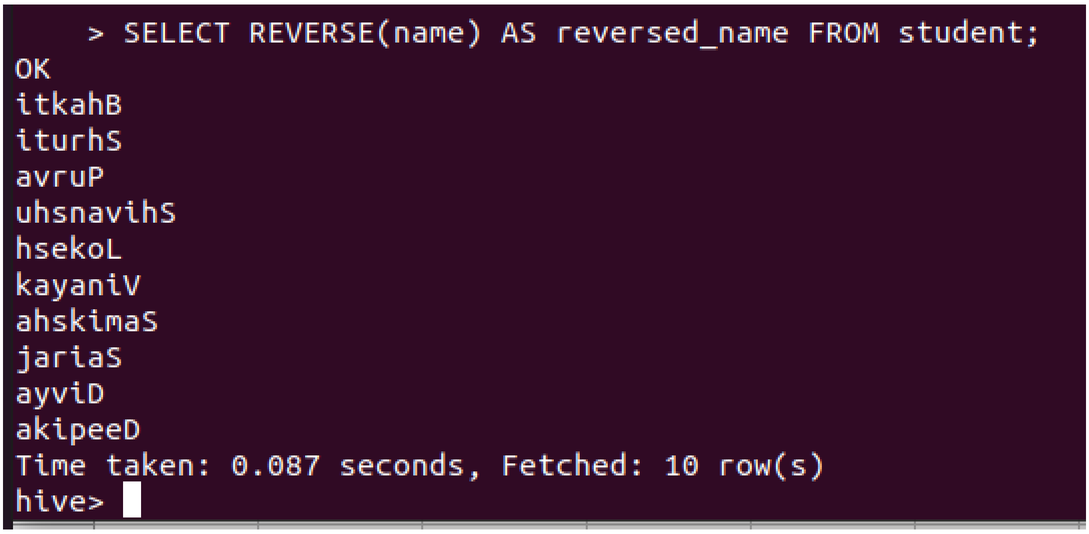
```
hive> SELECT LOWER(name) AS lower_name FROM student;
OK
bhakti
shruti
purva
shivanshu
lokesh
vinayak
samiksha
sairaj
divya
deepika
Time taken: 0.119 seconds, Fetched: 10 row(s)
```



```
     > SELECT UPPER(name) AS upper_name FROM student;
OK
BHAKTI
SHRUTI
PURVA
SHIVANSHU
LOKESH
VINAYAK
SAMIKSHA
SAIRAJ
DIVYA
DEEPIKA
Time taken: 0.093 seconds, Fetched: 10 row(s)
```



```
     > SELECT LTRIM(name) AS trimmed_name FROM student;
OK
Bhakti
Shruti
Purva
Shivanshu
Lokesh
Vinayak
Samiksha
Sairaj
Divya
Deepika
Time taken: 0.103 seconds, Fetched: 10 row(s)
```



```
    > SELECT REVERSE(name) AS reversed_name FROM student;
OK
itkahB
iturhS
avruP
uhsnavihS
hsekoL
kayaniV
ahskimaS
jariaS
ayviD
akipeeD
Time taken: 0.087 seconds, Fetched: 10 row(s)
hive>
```