**Hive Assignment:**

**Problem statement 1:**

Create a table with the schema as specified below and load the data.

Download Dataset

| id | tenure | designation | salary |
|---|---|---|---|
| INT | INT | STRING | BIGINT |

Write a query to derive a new column extra_vacation based on the tenure served, the logic is as given below.

1. If tenure < 2, Then 20
2. If tenure is 2-10 then 30 days
3. If tenure > 10 then 40 days

**Sol:**

```
hive> create table employees (id INT, tenure INT, designation STRING, salary BIGINT)
    > ROW FORMAT DELIMITED FIELDS TERMINATED BY '|'
    > STORED AS TextFile
    > TBLPROPERTIES(
    > "skip.header.line.count"="1",
    > "skip.footer.line.count"="1"
    > );
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask. AlreadyExistsException
(message:Table employees already exists)
hive> create table employee (id INT, tenure INT, designation STRING, salary BIGINT)
    > ROW FORMAT DELIMITED FIELDS TERMINATED BY '|'
    > STORED AS TextFile
    > TBLPROPERTIES(
    > "skip.header.line.count"="1",
    > "skip.footer.line.count"="1"
    > );
OK
Time taken: 0.107 seconds
hive> show tables;
OK
employee
employees
```

```
hive> LOAD DATA LOCAL INPATH
    > '/home/march8lab23/soumya_banerjee/user.txt' into table employee;
Loading data to table sban.employee
OK
Time taken: 0.998 seconds
```

```
hive> SELECT
    > id,
    > tenure,
    > designation,
    > salary,
    > CASE
    > WHEN tenure < 2 THEN 20
    > WHEN tenure BETWEEN 2 AND 10 THEN 30
    > ELSE 40
    > END AS extra_vacation
    > FROM employee;
Query ID = march8lab23_20230724190109_01837fc1-1415-425e-b058-ace0336e841d
Total jobs = 1
```

```
Stage-Stage-1: Map: 1   Cumulative CPU: 3.72 sec   HDFS Read: 5790 HDFS Write: 740 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 720 msec
OK
NULL    NULL    NULL    NULL    40
1       2       technician      200000  30
NULL    NULL    NULL    NULL    40
2       5       other   1000000 30
NULL    NULL    NULL    NULL    40
3       2       writer  1600000 30
NULL    NULL    NULL    NULL    40
4       5       technician      100000  30
NULL    NULL    NULL    NULL    40
5       2       other   100000  30
NULL    NULL    NULL    NULL    40
6       2       executive       98101   30
NULL    NULL    NULL    NULL    40
7       21      administrator   91344   40
NULL    NULL    NULL    NULL    40
8       16      administrator   91344   40
NULL    NULL    NULL    NULL    40
9       12      student 123230  40
NULL    NULL    NULL    NULL    40
10      5       lawyer  90703   30
NULL    NULL    NULL    NULL    40
Time taken: 17.618 seconds, Fetched: 21 row(s)
```

**Problem statement 2:**

## Problem Statement 02

**Prerequisite :**

Create a table “temperature” to store the dataset as mentioned in the schema and load the data.

Download Dataset

| Date | State | Temperature |
|---|---|---|

**Trainer : Naveen Pn /**
www.linkedin.com/in/naveen-pn

| STRING | STRING | Array<DOUBLE> |
|---|---|---|

Write a query to calculate the maximum temperature of each state.

**Sol:**

```
hive> use sban;
OK
Time taken: 1.541 seconds
hive> show tables;
OK
employee
employees
temperature
Time taken: 0.191 seconds, Fetched: 3 row(s)
hive> LOAD DATA LOCAL INPATH
    > '/home/march8lab23/soumya_banerjee/temperature.txt' into table temperature;
Loading data to table sban.temperature
OK
Time taken: 1.117 seconds
hive> SELECT State,
    > MAX(temp) AS max_temperature
    > FROM temperaturey
    > LATERAL VIEW EXPLODE(Temperature) temp_table AS temp
    > GROUP BY State;
FAILED: SemanticException [Error 10001]: Line 3:5 Table not found 'temperaturey'
hive> SELECT State,
    > MAX(temp) AS max_temperature
    > FROM temperature
    > LATERAL VIEW EXPLODE(Temperature) temp_table AS temp
    > GROUP BY State;
Query ID = march8lab23_20230724193515_24fa5ca9-4d2b-443b-9974-f5c2b81ef395
Total jobs = 1
Launching Job 1 out of 1
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:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
23/07/24 19:35:16 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.i
nternal/10.1.1.204:8032
23/07/24 19:35:17 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.i
nternal/10.1.1.204:8032
Starting Job = job_1685754149182_7433, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:606
6/proxy/application_1685754149182_7433/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job  -kill job_
1685754149182_7433
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-07-24 19:35:26,972 Stage-1 map = 0%,  reduce = 0%
2023-07-24 19:35:34,204 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.85 sec
2023-07-24 19:35:45,516 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 6.41 sec
MapReduce Total cumulative CPU time: 6 seconds 410 msec
Ended Job = job_1685754149182_7433
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 6.41 sec   HDFS Read: 11534 HDFS Write: 87 HDFS EC Read
: 0 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 410 msec
OK
Time taken: 30.763 seconds
```

**Problem statement 3:**

## Problem Statement 03

**Prerequisite :**

Create a table 'student_marks' with schema as shown above and load the data into the 'student_marks' table.

Download Dataset

| Name | Marks |
|---|---|
| STRING | Map<STRING, INT> |

Write a query to perform below mentioned tasks:

1. Display NAME who have scored more than 90 in subject Maths subject
2. Display NAME and <Marks Scored in Physics subject>
3. Display NAME, and <Maximum-Subject-Marks>
4. Display NAME and <Average Marks Scored>
5. Display NAME and <Percentage of marks>

**Sol:**

**Display NAME who has scored more than 90 in subject Maths subject**

```
hive> CREATE TABLE student_marks (Name STRING,
    > Marks MAP<STRING, INT>
    > );
OK
Time taken: 0.111 seconds
hive> show tables;
OK
employee
employees
student_marks
temperature
Time taken: 0.05 seconds, Fetched: 4 row(s)

hive> LOAD DATA LOCAL INPATH
    > '/home/march8lab23/soumya_banerjee/studentstruct.txt' into table student_marks;
Loading data to table sban.student_marks
OK
Time taken: 0.786 seconds
hive> show tables
    > ;
OK
employee
employees
student_marks
temperature
Time taken: 0.038 seconds, Fetched: 4 row(s)
```

```
hive> SELECT Name
    > FROM student_marks
    > WHERE Marks['Maths'] > 90;
Query ID = march8lab23_20230724200025_a2663fb8-61b7-44b5-8bd7-77ef1718774a
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
23/07/24 20:00:26 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.i
nternal/10.1.1.204:8032
23/07/24 20:00:26 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.i
nternal/10.1.1.204:8032
Starting Job = job_1685754149182_7434, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:606
6/proxy/application_1685754149182_7434/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job  -kill job_
1685754149182_7434
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2023-07-24 20:00:39,148 Stage-1 map = 0%,  reduce = 0%
2023-07-24 20:00:46,364 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.65 sec
MapReduce Total cumulative CPU time: 3 seconds 650 msec
Ended Job = job_1685754149182_7434
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 3.65 sec   HDFS Read: 35355 HDFS Write: 87 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 650 msec
OK
Time taken: 22.572 seconds
```

## Display NAME and Marks scored in physics

```
hive> SELECT Name, Marks['Physics'] AS Marks_Scored_in_Physics
    > FROM student_marks;
Query ID = march8lab23_20230724200244_4c9f8697-f79f-46a7-aef8-3258d40dcde9
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
23/07/24 20:02:44 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.i
nternal/10.1.1.204:8032
23/07/24 20:02:44 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.i
nternal/10.1.1.204:8032
Starting Job = job_1685754149182_7435, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:606
6/proxy/application_1685754149182_7435/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job  -kill job_
1685754149182_7435
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2023-07-24 20:02:53,098 Stage-1 map = 0%,  reduce = 0%
2023-07-24 20:03:00,387 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.76 sec
MapReduce Total cumulative CPU time: 3 seconds 760 msec
Ended Job = job_1685754149182_7435
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 3.76 sec   HDFS Read: 35241 HDFS Write: 43490 HDFS EC Read: 0 SUCC
ESS
Total MapReduce CPU Time Spent: 3 seconds 760 msec
OK
Name,Marks,Address      NULL
        NULL
Kiran,physics:98$chemistry:95$maths:83$biology:67,14$Anekal$560072      NULL
        NULL
Nagesh,physics:76$chemistry:34$maths:92$biology:57,14$Anekal$560072     NULL
        NULL
Kusumanjali,physics:98$chemistry:95$maths:83$biology:67,12$Agara$560034 NULL
        NULL
Najma,physics:76$chemistry:34$maths:92$biology:57,14$Anekal$560072      NULL
        NULL
Rajani,physics:76$chemistry:34$maths:92$biology:57,25$Banashankari$560050       NULL
        NULL
Akshar,physics:98$chemistry:95$maths:83$biology:67,100$Brigade Road$560001      NULL
        NULL
Swetha,physics:98$chemistry:95$maths:83$biology:67,100$Brigade Road$560001      NULL
```

```
        NULL
Nischal,physics:98$chemistry:95$maths:83$biology:67,100$Brigade Road$560001     NULL
        NULL
Urvashi,physics:76$chemistry:34$maths:92$biology:57,12$Agara$560034     NULL
        NULL
Panchanan,physics:76$chemistry:34$maths:92$biology:57,25$Banashankari$560050    NULL
        NULL
Sunasi,physics:76$chemistry:34$maths:92$biology:57,14$Anekal$560072     NULL
        NULL
Santayani,physics:98$chemistry:95$maths:83$biology:67,14$Anekal$560072  NULL
        NULL
Rupa,physics:98$chemistry:95$maths:83$biology:67,100$Brigade Road$560001        NULL
        NULL
Kundan,physics:98$chemistry:95$maths:83$biology:67,25$Banashankari$560050       NULL
        NULL
Chinmayananda,physics:98$chemistry:95$maths:83$biology:67,14$Anekal$560072      NULL
        NULL
Grihith,physics:76$chemistry:34$maths:92$biology:57,100$Brigade Road$560001     NULL
        NULL
Chandrakirthi,physics:98$chemistry:95$maths:83$biology:67,14$Anekal$560072      NULL
        NULL
Hitendra,physics:98$chemistry:95$maths:83$biology:67,100$Brigade Road$560001    NULL
        NULL
Saudamini,physics:98$chemistry:95$maths:83$biology:67,12$Agara$560034   NULL
        NULL
Shatrunjay,physics:76$chemistry:34$maths:92$biology:57,100$Brigade Road$560001  NULL
        NULL
Rudra,physics:76$chemistry:34$maths:92$biology:57,12$Agara$560034       NULL
        NULL
Mohini,physics:98$chemistry:95$maths:83$biology:67,12$Agara$560034      NULL
        NULL
Shravan,physics:76$chemistry:34$maths:92$biology:57,100$Brigade Road$560001     NULL
        NULL
Druthi,physics:98$chemistry:95$maths:83$biology:67,12$Agara$560034      NULL
        NULL
Priyabrata,physics:76$chemistry:34$maths:92$biology:57,100$Brigade Road$560001  NULL
        NULL
Pranati,physics:98$chemistry:95$maths:83$biology:67,12$Agara$560034     NULL
        NULL
Neepa,physics:76$chemistry:34$maths:92$biology:57,100$Brigade Road$560001       NULL
        NULL
Devalekha,physics:98$chemistry:95$maths:83$biology:67,100$Brigade Road$560001   NULL
        NULL
Samgram,physics:98$chemistry:95$maths:83$biology:67,100$Brigade Road$560001     NULL
        NULL
Prasata,physics:98$chemistry:95$maths:83$biology:67,12$Agara$560034     NULL
        NULL
Indraneel,physics:98$chemistry:95$maths:83$biology:67,100$Brigade Road$560001   NULL
        NULL
Susita,physics:76$chemistry:34$maths:92$biology:57,100$Brigade Road$560001      NULL
        NULL
```

## Display NAME, and Maximum subject marks

```
hive> SELECT
    >
    >   Name,
    >
    >   max(mark) AS Maximum_Subject_Marks
    >
    > FROM student_marks
    >
    > LATERAL VIEW explode(Marks) marks_table AS subject, mark
    >
    > GROUP BY Name;
Query ID = march8lab23_20230724200557_d30a095d-9e23-4eec-af5a-2980d73875ef
Total jobs = 1
Launching Job 1 out of 1
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:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
23/07/24 20:05:57 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.i
nternal/10.1.1.204:8032
23/07/24 20:05:57 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.i
nternal/10.1.1.204:8032
Starting Job = job_1685754149182_7436, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:606
6/proxy/application_1685754149182_7436/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job  -kill job_
1685754149182_7436
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-07-24 20:06:14,743 Stage-1 map = 0%,  reduce = 0%
2023-07-24 20:06:23,224 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.2 sec
2023-07-24 20:06:30,424 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.34 sec
MapReduce Total cumulative CPU time: 5 seconds 340 msec
Ended Job = job_1685754149182_7436
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.34 sec   HDFS Read: 41757 HDFS Write: 87 HDFS EC Read
: 0 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 340 msec
OK
Time taken: 35.134 seconds
```

## Display NAME and Average marks scored

```
hive> SELECT
    >
    >   Name,
    >
    >   avg(mark) AS Average_Marks_Scored
    >
    > FROM student_marks
    >
    > LATERAL VIEW explode(Marks) marks_table AS subject, mark
    >
    > GROUP BY Name;
Query ID = march8lab23_20230724200745_15103fc2-4b68-4a98-b9c8-7ab26d15b031
Total jobs = 1
Launching Job 1 out of 1
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:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
23/07/24 20:07:45 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.i
nternal/10.1.1.204:8032
23/07/24 20:07:45 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.i
nternal/10.1.1.204:8032
Starting Job = job_1685754149182_7437, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:606
6/proxy/application_1685754149182_7437/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job  -kill job_
1685754149182_7437
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-07-24 20:07:58,437 Stage-1 map = 0%,  reduce = 0%
2023-07-24 20:08:06,774 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.39 sec
2023-07-24 20:08:12,048 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.27 sec
MapReduce Total cumulative CPU time: 5 seconds 270 msec
Ended Job = job_1685754149182_7437
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.27 sec   HDFS Read: 42219 HDFS Write: 87 HDFS EC Read
: 0 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 270 msec
OK
Time taken: 29.073 seconds
```

## Display Name and Percentage of marks

```
hive> SELECT
    >
    >   Name,
    >
    >   (sum(mark) * 100) / (count(subject) * 100) AS Percentage_of_Marks
    >
    > FROM student_marks
    >
    > LATERAL VIEW explode(Marks) marks_table AS subject, mark
    >
    > GROUP BY Name;
Query ID = march8lab23_20230724200915_24ab30ab-77ec-4cc4-8869-f78cf154c48b
Total jobs = 1
Launching Job 1 out of 1
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:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
23/07/24 20:09:15 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.i
nternal/10.1.1.204:8032
23/07/24 20:09:15 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.i
nternal/10.1.1.204:8032
Starting Job = job_1685754149182_7438, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:606
6/proxy/application_1685754149182_7438/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job  -kill job_
1685754149182_7438
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-07-24 20:09:25,184 Stage-1 map = 0%,  reduce = 0%
2023-07-24 20:09:35,443 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.34 sec
2023-07-24 20:09:45,687 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 9.11 sec
MapReduce Total cumulative CPU time: 9 seconds 110 msec
Ended Job = job_1685754149182_7438
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 9.11 sec   HDFS Read: 43039 HDFS Write: 87 HDFS EC Read
: 0 SUCCESS
Total MapReduce CPU Time Spent: 9 seconds 110 msec
OK
Time taken: 31.671 seconds
```

## Problem statement 4:

Create a table “student_info” with schema as show below and load the data

Download Dataset

| Name | Marks | Address |
|---|---|---|

**Trainer : Naveen Pn /**
www.linkedin.com/in/naveen-pn

| STRING | Map<STRING, INT> | Struct<doorNo INT,Location String,Pincode INT> |
|---|---|---|

Write a query for the below mentioned tasks

1. Display all “NAME” who is located in Banashankari
2. Calculate the total count who is staying in pin code 560001

**Sol:**

```
hive> CREATE TABLE student_info (
    >
    >   Name STRING,
    >
    >   Marks MAP<STRING, INT>,
    >
    >   Address STRUCT<doorNo:INT, Location:STRING, Pincode:INT>
    >
    > );
OK
Time taken: 0.103 seconds
hive> show tables;
OK
employee
employees
student_info
student_marks
temperature
Time taken: 0.042 seconds, Fetched: 5 row(s)

hive> LOAD DATA LOCAL INPATH
    > '/home/march8lab23/soumya_banerjee/studentstruct.txt' into table student_info;
Loading data to table sban.student_info
OK
Time taken: 0.676 seconds
```

## All the names located in banashakari

```
hive> SELECT Name
    > FROM student_info
    > WHERE Address.Location = 'Banashankari';
Query ID = march8lab23_20230724202831_8cc9b96a-81e2-430c-9ad2-5b79e2fd46e3
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
23/07/24 20:28:31 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.i
nternal/10.1.1.204:8032
23/07/24 20:28:31 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.i
nternal/10.1.1.204:8032
Starting Job = job_1685754149182_7439, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:606
6/proxy/application_1685754149182_7439/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job  -kill job_
1685754149182_7439
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2023-07-24 20:28:41,681 Stage-1 map = 0%,  reduce = 0%
2023-07-24 20:28:49,999 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.15 sec
MapReduce Total cumulative CPU time: 3 seconds 150 msec
Ended Job = job_1685754149182_7439
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 3.15 sec   HDFS Read: 35637 HDFS Write: 87 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 150 msec
OK
Time taken: 19.29 seconds
```

## Total count of individuals residing in pin code ‘560001’

```
hive> SELECT COUNT(*) AS total_count
    > FROM student_info
    > WHERE Address.Pincode = 560001;
Query ID = march8lab23_20230724203035_3f28e39d-b208-4e7f-b81f-692992afba88
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 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:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
23/07/24 20:30:35 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.i
nternal/10.1.1.204:8032
23/07/24 20:30:35 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.i
nternal/10.1.1.204:8032
Starting Job = job_1685754149182_7440, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:606
6/proxy/application_1685754149182_7440/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job  -kill job_
1685754149182_7440
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-07-24 20:30:44,033 Stage-1 map = 0%,  reduce = 0%
2023-07-24 20:30:51,199 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.98 sec
2023-07-24 20:30:57,347 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 7.1 sec
MapReduce Total cumulative CPU time: 7 seconds 100 msec
Ended Job = job_1685754149182_7440
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 7.1 sec   HDFS Read: 40454 HDFS Write: 101 HDFS EC Read
: 0 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 100 msec
OK
0
Time taken: 23.305 seconds, Fetched: 1 row(s)
```