Structured Data Analysis using Hive

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

```
hive> CREATE TABLE employee_details( id INT, tenure INT, designation STRING, salary BIGINT ) ROW FORMAT DELIMITED FIELDS TERMINATED BY '|' S
TORED AS TextFile TBLPROPERTIES( "skip.header.line.count"="1", "skip.footer.line.count"="1" );
OK
Time taken: 0.294 seconds
hive> LOAD DATA LOCAL INPATH '/home/march8lab23/damini_file/Files/Files/user.dat' into table employee_details;
Loading data to table damini_assignment.employee_details
OK
Time taken: 1.119 seconds
```

```
hive> select *, case when tenure<2 then 20 when tenure between 2 and 10 then 30 when tenure>10 then 40 end as extra_vacation from employee_d
etails;
Query ID = march8lab23_20230727120440_b5582416-d08a-4185-aa0d-07d719d1d545
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
23/07/27 12:04:41 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/07/27 12:04:42 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1685754149182_7735, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1685754149182_
7735/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job  -kill job_1685754149182_7735
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2023-07-27 12:04:50,932 Stage-1 map = 0%,  reduce = 0%
2023-07-27 12:04:59,198 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.76 sec
MapReduce Total cumulative CPU time: 3 seconds 760 msec
Ended Job = job_1685754149182_7735
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 3.76 sec   HDFS Read: 5881 HDFS Write: 443 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 760 msec
OK
1       2       technician      200000  30
2       5       other   1000000 30
3       2       writer  1600000 30
4       5       technician      100000  30
5       2       other   100000  30
6       2       executive       98101   30
7       21      administrator   91344   40
8       16      administrator   91344   40
9       12      student 123230  40
10      5       lawyer  90703   30
```

2) Create a table “temperature” to store the dataset as mentioned in the schema and load the data
   Write a query to calculate the maximum temperature of each state.

```
hive> CREATE TABLE temperature( Name STRING,state STRING,temperature array<double>) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'
N ITEMS TERMINATED BY ':' STORED AS TextFile TBLPROPERTIES( "skip.header.line.count"="1", "skip.footer.line.count"="0" );
OK
Time taken: 0.089 seconds
hive> select * from temperature;
OK
Time taken: 0.083 seconds
hive> LOAD DATA LOCAL INPATH '/home/march8lab23/damini_file/Files/Files/temperature.csv.dat' into table temperature;
Loading data to table damini_assignment.temperature
OK
Time taken: 0.71 seconds
hive> select * from temperature;
OK
1517581354      Goa     [23.3,25.6,34.7,19.8,41.7,32.9,22.4,19.8,24.1,22.1,23.5,23.9]
1523050092      Delhi   [13.3,22.6,24.7,109.18,41.2,32.9,24.4,19.8,24.1,21.1,23.5,22.9]
1526749245      Kerala  [13.3,22.6,24.7,109.18,41.2,32.9,24.4,19.8,24.1,21.1,23.5,22.9]
1518351770      Tamil Nadu      [13.3,22.6,24.7,109.18,41.2,32.9,24.4,19.8,24.1,21.1,23.5,22.9]
1469755036      Uttar Pradesh   [13.3,22.6,24.7,109.18,41.2,32.9,24.4,19.8,24.1,21.1,23.5,22.9]
1477582469      Rajasthan       [13.3,22.6,24.7,109.18,41.2,32.9,24.4,19.8,24.1,21.1,23.5,22.9]
1508991065      Punjab  [23.3,25.6,34.7,19.8,41.7,32.9,22.4,19.8,24.1,22.1,23.5,23.9]
1499217916      Gujarat [13.3,22.6,24.7,109.18,41.2,32.9,24.4,19.8,24.1,21.1,23.5,22.9]
1492684452      Haryana [23.3,25.6,34.7,19.8,41.7,32.9,22.4,19.8,24.1,22.1,23.5,23.9]
1525740700      Karnataka       [13.3,22.6,24.7,109.18,41.2,32.9,24.4,19.8,24.1,21.1,23.5,22.9]
1481609997      Assam   [13.3,22.6,24.7,109.18,41.2,32.9,24.4,19.8,24.1,21.1,23.5,22.9]
Time taken: 0.086 seconds, Fetched: 11 row(s)
```

```
hive> select state,max(temp) as max_temp from temperature lateral view explode(temperature) explode_table as temp group by state;
Query ID = march8lab23_20230727125114_07ba4a81-3b69-43d1-8c6f-7f739984b343
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>
23/07/27 12:51:15 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/07/27 12:51:15 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1685754149182_7742, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1685754149182_
7742/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job  -kill job_1685754149182_7742
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-07-27 12:51:27,282 Stage-1 map = 0%,  reduce = 0%
2023-07-27 12:51:35,552 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.13 sec
2023-07-27 12:51:43,802 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.85 sec
MapReduce Total cumulative CPU time: 5 seconds 850 msec
Ended Job = job_1685754149182_7742
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.85 sec   HDFS Read: 11921 HDFS Write: 381 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 850 msec
OK
Assam   109.18
Delhi   109.18
Goa     41.7
Gujarat 109.18
```

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

```
hive> CREATE TABLE student_marks( Name STRING, Marks Map<STRING, INT>) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' COLLECTION ITEMS
TERMINATED BY '$' MAP KEYS TERMINATED BY ':' STORED AS TextFile TBLPROPERTIES( "skip.header.line.count"="1", "skip.footer.line.count"=
"0" );
OK
Time tselect * from student_marks limit 5;
OK
Time taken: 0.077 seconds
hive> LOAD DATA LOCAL INPATH '/home/march8lab26/numan_shaikh/hadoop_assignment/Files/student-struct-dataset.csv' into table student_ma
rks;>
Loading data to table numan_assignment.student_marks
OK
Time taken: 0.715 seconds
```

a) Write a query to perform below mentioned tasks: 1. Display NAME who have scored more than 90 in subject Maths subject

```
hive> select name,marks_value from student_marks lateral view explode(marks) scored as subject, marks_value where subject='maths' and mar
value>90 limit 10;
Query ID = march8lab23_20230727130722_7479a30a-f4c5-4abb-9bbe-fb67b0346955
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
23/07/27 13:07:23 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/07/27 13:07:23 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1685754149182_7747, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_16857541491
7747/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job  -kill job_1685754149182_7747
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2023-07-27 13:07:33,629 Stage-1 map = 0%,  reduce = 0%
2023-07-27 13:07:40,896 Stage-1 map = 100%,  reduce = 0%
MapReduce Total cumulative CPU time: 3 seconds 190 msec
Ended Job = job_1685754149182_7747
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 3.19 sec   HDFS Read: 72054 HDFS Write: 315 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 190 msec
OK
Nagesh  92
Najma   92
Rajani  92
Adarsh  92
Suhrid  92
Harigopal       92
Purandar        92
Urvashi 92
Panchanan       92
Sunasi  92
Time taken: 20.968 seconds, Fetched: 10 row(s)
```

b) Display NAME and marks scored in physics subject.

```
hive> select name,marks_value,subject from student_marks lateral view explode(marks) scored as subject, marks_value where subject='physics
Query ID = march8lab23_20230727151247_9a5c7a30-eeb3-4818-b598-3a16c5873bc3
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
23/07/27 15:12:47 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/07/27 15:12:47 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1685754149182_7760, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_16857541491
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job  -kill job_1685754149182_7760
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2023-07-27 15:12:55,277 Stage-1 map = 0%,  reduce = 0%
2023-07-27 15:13:03,484 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.95 sec
MapReduce Total cumulative CPU time: 3 seconds 950 msec
Ended Job = job_1685754149182_7760
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 3.95 sec   HDFS Read: 72108 HDFS Write: 395 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 950 msec
OK
Kiran   98      physics
Nagesh  76      physics
Kusumanjali     98      physics
Majma   76      physics
Majani  76      physics
Akshar  98      physics
Swetha  98      physics
Punyasloka      98      physics
Adarsh  76      physics
Vasudev 98      physics
Time taken: 17.247 seconds, Fetched: 10 row(s)
```

C) Display NAME, and <maximum-subject-marks>

```
hive> select name,max(max_marks) from (select name, map_values(marks) as subject_marks from s
arks) scored as max_marks group by name order by name limit 10;
Query ID = march8lab23_20230727130809_8bed607b-2134-4a44-92fd-ce373e9b19ed
Total jobs = 2
Launching Job 1 out of 2
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>
23/07/27 13:08:10 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-sout
23/07/27 13:08:10 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-sout
Starting Job = job_1685754149182_7748, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute
7748/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop jo
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-07-27 13:08:20,083 Stage-1 map = 0%,  reduce = 0%
2023-07-27 13:08:29,370 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 5.91 sec
2023-07-27 13:08:34,550 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 8.58 sec
MapReduce Total cumulative CPU time: 8 seconds 580 msec
Ended Job = job_1685754149182_7748
Launching Job 2 out of 2
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>
23/07/27 13:08:36 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-sout
23/07/27 13:08:36 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-sout
Starting Job = job_1685754149182_7749, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute
7749/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop jo
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
```

```
2023-07-27 13:08:45,480 Stage-2 map = 0%,  reduce = 0%
2023-07-27 13:08:53,675 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 2.46 sec
2023-07-27 13:09:01,866 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 5.83 sec
MapReduce Total cumulative CPU time: 5 seconds 830 msec
Ended Job = job_1685754149182_7749
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 8.58 sec   HDFS Read: 7320804 HDFS Write:
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 5.83 sec   HDFS Read: 5960 HDFS Write: 309
Total MapReduce CPU Time Spent: 14 seconds 410 msec
OK
Aaarti  98
Aachman 98
Aadesh  98
Aadi    98
Aafreen 98
Aakar   98
Aakash  98
Aalap   98
Aandaleeb       98
Aashika 98
Time taken: 53.327 seconds, Fetched: 10 row(s)
```

c) Display NAME, and <average -Subject-Marks>

```
ive> select name,avg(m_marks) from (select name, map_values(marks) as subject_marks from student_marks) t1 lateral view explode(subject_
order by name limit 10;
uery ID = march8lab23_20230727152043_8987675e-cbe6-432a-8602-b38cc35280ad
otal jobs = 2
aunching Job 1 out of 2
umber of reduce tasks not specified. Estimated from input data size: 1
n order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
n order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
n order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
3/07/27 15:20:44 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
3/07/27 15:20:44 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
tarting Job = job_1685754149182_7761, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1685754149
ill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job  -kill job_1685754149182_7761
adoop job information for Stage-1: number of mappers: 1; number of reducers: 1
023-07-27 15:20:54,949 Stage-1 map = 0%,  reduce = 0%
023-07-27 15:21:05,216 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 6.11 sec
023-07-27 15:21:14,456 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 8.88 sec
apReduce Total cumulative CPU time: 8 seconds 880 msec
nded Job = job_1685754149182_7761
aunching Job 2 out of 2
umber of reduce tasks determined at compile time: 1
n order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
n order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
n order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
3/07/27 15:21:15 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
3/07/27 15:21:15 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
ill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job  -kill job_1685754149182_7762n_1685754149
adoop job information for Stage-2: number of mappers: 1; number of reducers: 1
023-07-27 15:21:26,437 Stage-2 map = 0%,  reduce = 0%
023-07-27 15:21:34,954 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 1.47 sec
023-07-27 15:21:44,186 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 5.01 sec
nded Job = job_1685754149182_7762
apReduce Jobs Launched:
tage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 8.88 sec   HDFS Read: 7321407 HDFS Write: 418 HDFS EC Read: 0 SU
tage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 5.01 sec   HDFS Read: 6060 HDFS Write: 423 HDFS EC Read: 0 SUCCE
otal MapReduce CPU Time Spent: 13 seconds 890 msec
OK
aarti  79.06818181818181
achman 78.01315789473684
adesh  71.38157894736842
adi    73.99
afreen 76.75
akar   74.08333333333333
akash  73.75
alap   74.08333333333333
andaleeb       76.90789473684211
ashika 79.06818181818181
ime taken: 61.257 seconds, Fetched: 10 row(s)
```

d) Display NAME and <percentage of marks>

```
hive> select name,map("physics",cast(marks["physics"] as double)/100,"chemistry",cast(marks["chemistry"] as double)/100,"maths",cast(marks["m
st(marks["biology"] as double)/100)as percentage_marks from student_marks limit 10;
Query ID = march8lab23_20230727154244_539e93ed-ad39-46b3-b1ab-fa26e11c1b0b
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
23/07/27 15:42:44 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/07/27 15:42:44 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1685754149182_7763, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1685754149182_7
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job  -kill job_1685754149182_7763
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2023-07-27 15:42:52,433 Stage-1 map = 0%,  reduce = 0%
2023-07-27 15:43:00,631 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.69 sec
MapReduce Total cumulative CPU time: 3 seconds 690 msec
Ended Job = job_1685754149182_7763
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 3.69 sec   HDFS Read: 71677 HDFS Write: 805 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 690 msec
OK
Kiran   {"physics":0.98,"chemistry":0.95,"maths":0.83,"biology":0.67}
Nagesh  {"physics":0.76,"chemistry":0.34,"maths":0.92,"biology":0.57}
Kusumanjali     {"physics":0.98,"chemistry":0.95,"maths":0.83,"biology":0.67}
Rajma   {"physics":0.76,"chemistry":0.34,"maths":0.92,"biology":0.57}
Rajani  {"physics":0.76,"chemistry":0.34,"maths":0.92,"biology":0.57}
Akshar  {"physics":0.98,"chemistry":0.95,"maths":0.83,"biology":0.67}
Swetha  {"physics":0.98,"chemistry":0.95,"maths":0.83,"biology":0.67}
Punyasloka      {"physics":0.98,"chemistry":0.95,"maths":0.83,"biology":0.67}
Adarsh  {"physics":0.76,"chemistry":0.34,"maths":0.92,"biology":0.57}
Vasudev {"physics":0.98,"chemistry":0.95,"maths":0.83,"biology":0.67}
Time taken: 17.294 seconds, Fetched: 10 row(s)
```

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

```
hive> CREATE TABLE student_info(Name STRING, Marks Map<STRING, INT>, Address Struct<doorNo: INT,Location: String,Pincode:
OLLECTION ITEMS TERMINATED BY '$' MAP KEYS TERMINATED BY ':' STRUCT KEYS TERMINATED BY '$' STORED AS TextFile TBLPROPERTIE
unt"= "0");
FAILED: ParseException line 1:230 missing EOF at 'STRUCT' near '':''
hive> CREATE TABLE student_info(Name STRING, Marks Map<STRING, INT>, Address Struct<doorNo: INT,Location: String,Pincode:
OLLECTION ITEMS TERMINATED BY '$' MAP KEYS TERMINATED BY ':' STORED AS TextFile TBLPROPERTIES( "skip.header.line.count"="1
OK
Time taken: 0.503 seconds
hive> LOAD DATA LOCAL INPATH '/home/march8lab23/damini_file/Files/Files/student-struct-dataset.csv' into table student_in
Loading data to table damini_assignment.student_info
OK
Time taken: 1.24 seconds
```

a) Display all “NAME” who is located in Banashankari

```
hive> select name, address.location from student_info where address.location='Banashankari'
Query ID = march8lab23_20230727155025_0389f3c2-bc23-4802-8245-af24dd93e3dd
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
23/07/27 15:50:26 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-so
23/07/27 15:50:26 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-so
Starting Job = job_1685754149182_7764, Tracking URL = http://ip-10-1-1-204.ap-south-1.compu
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2023-07-27 15:50:37,523 Stage-1 map = 0%,  reduce = 0%
2023-07-27 15:50:45,061 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.06 sec
MapReduce Total cumulative CPU time: 3 seconds 60 msec
Ended Job = job_1685754149182_7764
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 3.06 sec   HDFS Read: 70872 HDFS Write: 417 HDFS EC
Total MapReduce CPU Time Spent: 3 seconds 60 msec
OK
Rajani  Banashankari
Punyasloka      Banashankari
Panchanan       Banashankari
Kundan  Banashankari
Sindhu  Banashankari
Maharth Banashankari
Rasul   Banashankari
Vadunath        Banashankari
Keshi   Banashankari
Anarghya        Banashankari
Time taken: 21.613 seconds, Fetched: 10 row(s)
hive>
```

b) Calculate the total count who is staying in pin code 560001

```
hive> select count(*) as total_count from student_info where address.pincode=560001;
Query ID = march8lab23_20230727155350_91f23dce-bd8e-4225-a0a8-b96b39fe26aa
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>
23/07/27 15:53:50 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-s
23/07/27 15:53:50 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-s
Starting Job = job_1685754149182_7765, Tracking URL = http://ip-10-1-1-204.ap-south-1.comp
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoo
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-07-27 15:53:59,737 Stage-1 map = 0%,  reduce = 0%
2023-07-27 15:54:11,015 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 4.02 sec
2023-07-27 15:54:20,248 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 7.06 sec
MapReduce Total cumulative CPU time: 7 seconds 60 msec
Ended Job = job_1685754149182_7765
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 7.06 sec   HDFS Read: 7320375 HDFS Writ
Total MapReduce CPU Time Spent: 7 seconds 60 msec
OK
24890
Time taken: 30.63 seconds, Fetched: 1 row(s)
```