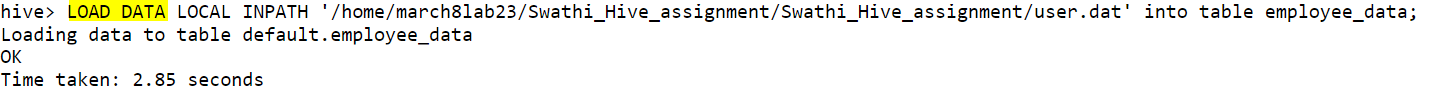


A computer code with black text

Description automatically generated

A screenshot of a computer code

Description automatically generated



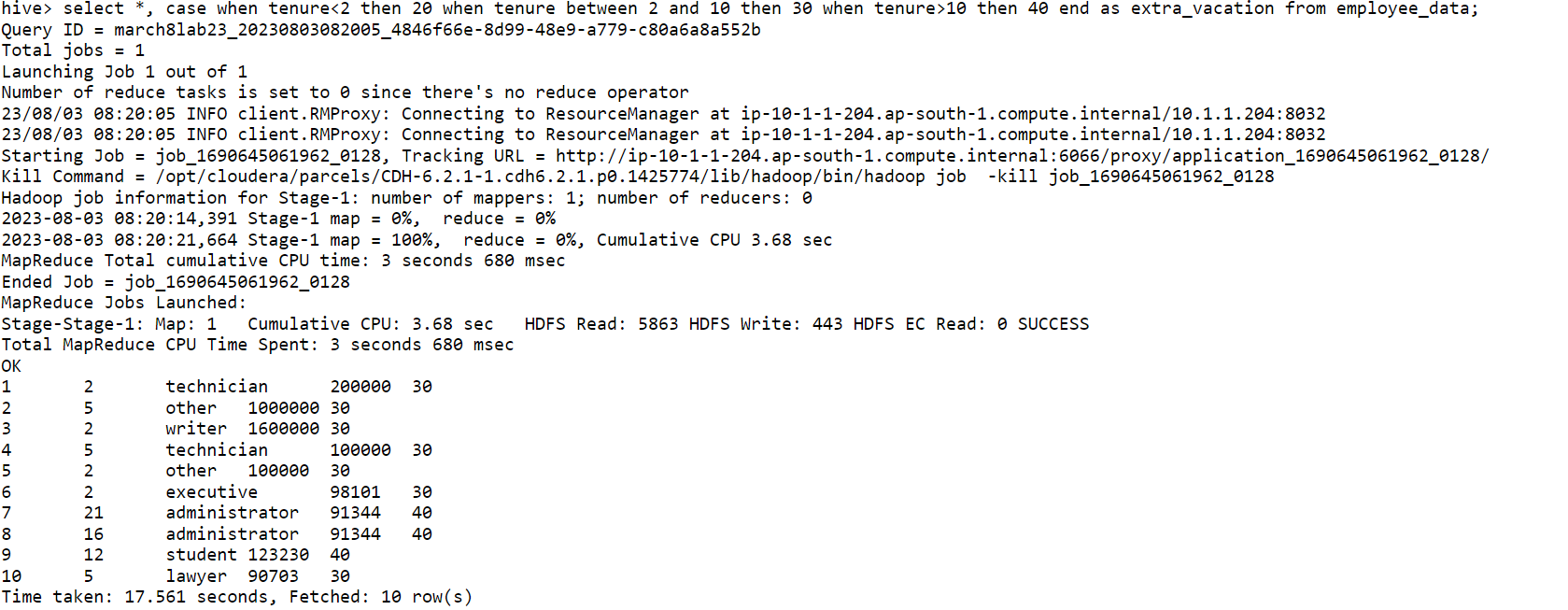
**Problem Statement 01**

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

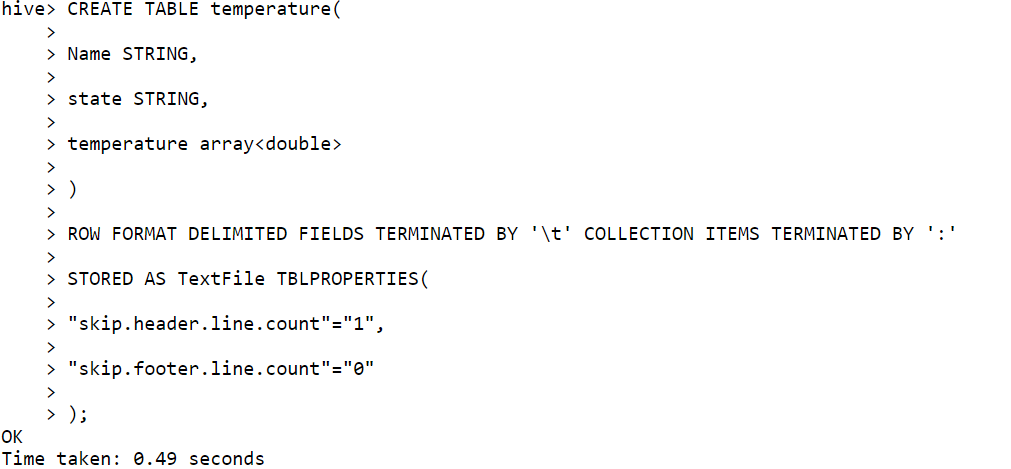


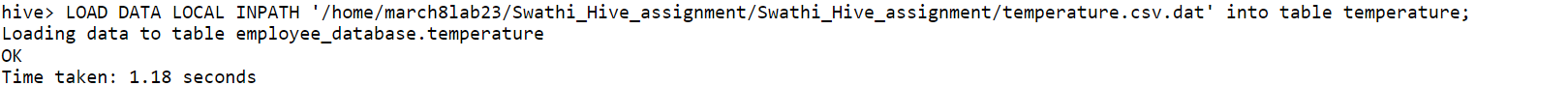
\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Problem Statement 02**

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.





A screenshot of a computer

Description automatically generated

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Problem Statement 03**

Create a table 'student\_marks' with schema as shown above and load the data into the 'student\_marks' table.

A close-up of a computer screen

Description automatically generated

Write a query to perform below mentioned tasks:

1. Display NAME who have scored more than 90 in subject Maths subject

A close-up of a white background

Description automatically generated

1. Display NAME and marks scored in physics subject.

A close-up of a white background

Description automatically generated

1. Display NAME, and <Maximum-Subject-Marks>

A close up of a white page

Description automatically generated

1. Display NAME, and <average -Subject-Marks>
2. Display NAME and <percentage of marks>

A white background with black text

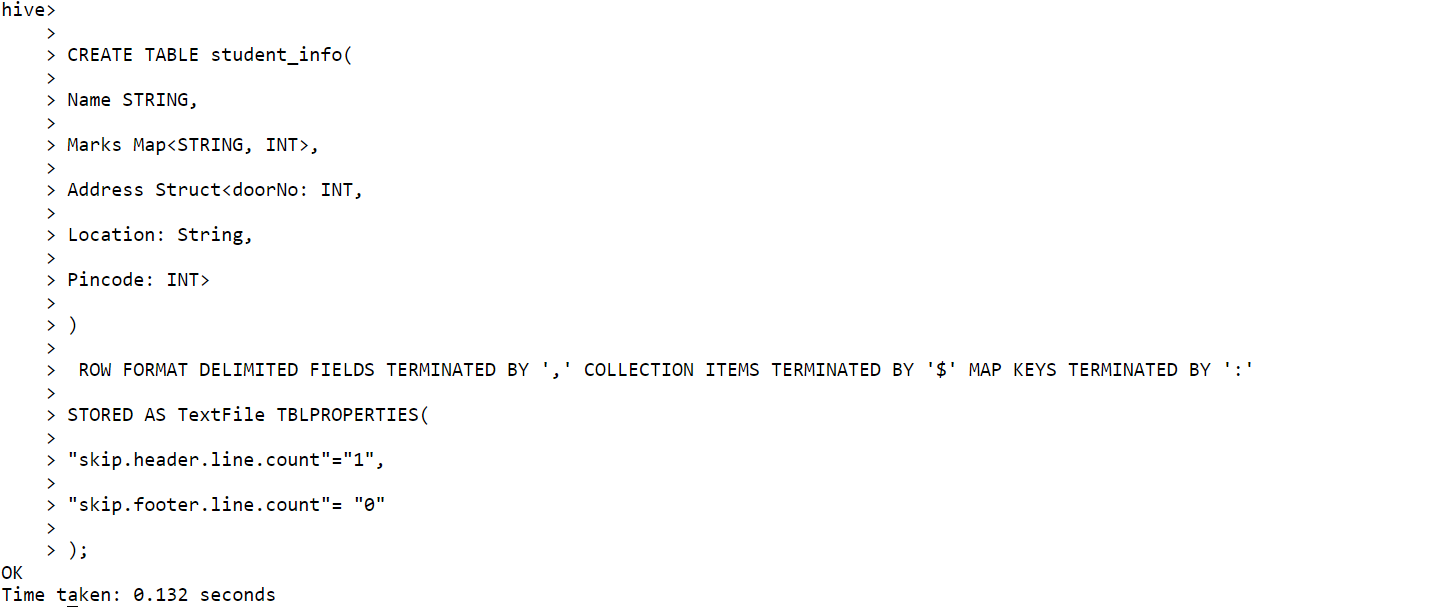
Description automatically generated

Write a query to perform below mentioned tasks:

1. Display NAME who have scored more than 90 in subject Maths subject

**Problem Statement 04**

Create a table “student\_info” with schema as show below and load the data

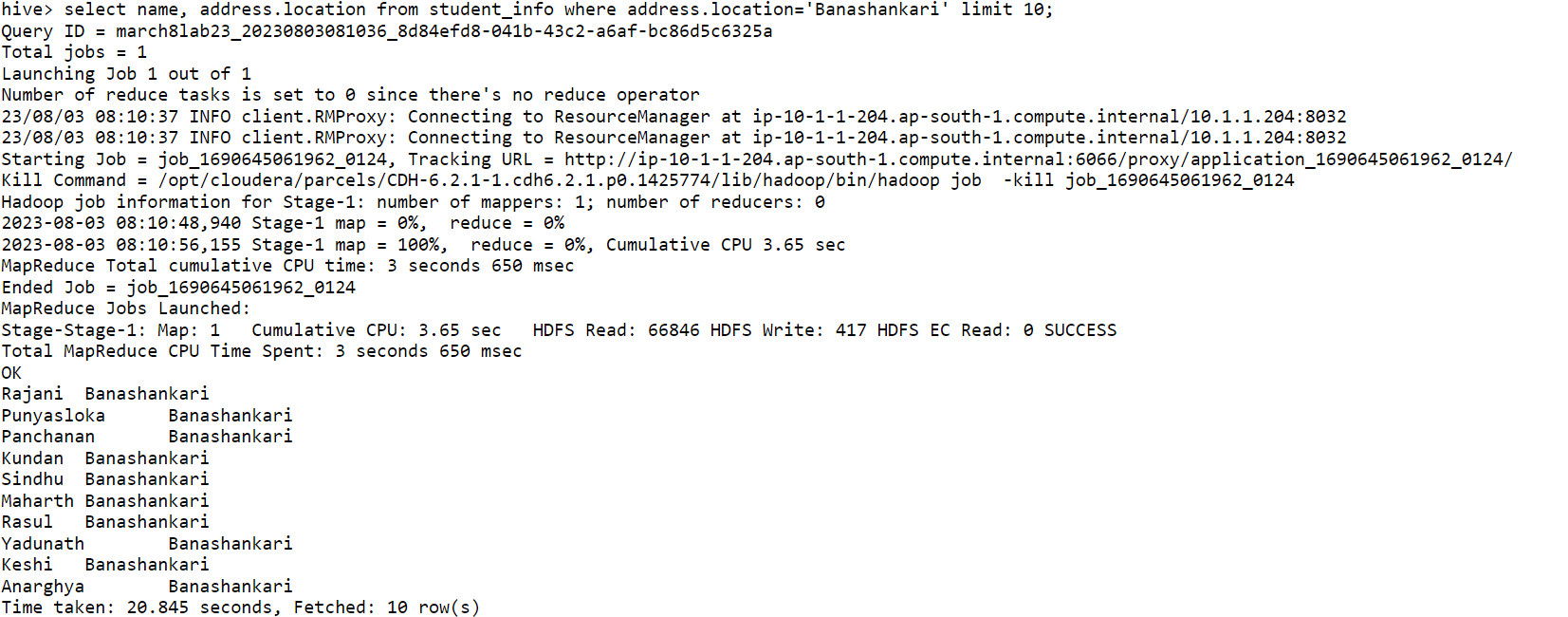


hive> LOAD DATA LOCAL INPATH '/home/march8lab23/Swathi\_Hive\_assignment/Swathi\_Hive\_assignment/student\_struct\_dataset.csv' into table student\_info;

Loading data to table employee\_database.student\_info

OK

Time taken: 1.224 seconds



hive> select count(\*) from student\_info where address.pincode=560001;

Query ID = march8lab23\_20230803085742\_aeea4387-ecbc-43f7-bc5d-7181b4cb89c2

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>

23/08/03 08:57:42 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032

23/08/03 08:57:42 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\_1690645061962\_0145, Tracking URL = <http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1690645061962_0145/>

Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job\_1690645061962\_0145

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2023-08-03 08:57:52,700 Stage-1 map = 0%, reduce = 0%

2023-08-03 08:58:01,888 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.84 sec

2023-08-03 08:58:10,061 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 9.16 sec

MapReduce Total cumulative CPU time: 9 seconds 160 msec

Ended Job = job\_1690645061962\_0145

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 9.16 sec HDFS Read: 7382011 HDFS Write: 105 HDFS EC Read: 0 SUCCESS

Total MapReduce CPU Time Spent: 9 seconds 160 msec

OK

24890

Time taken: 28.415 seconds, Fetched: 1 row(s)

hive>