



NITTE MEENAKSHI INSTITUTE OF TECHNOLOGY

(An autonomous institution with A+ Grade by NAAC /UGC, Affiliated to Visvesvaraya Technological University, Belgaum, Approved by UGC/AICTE/Govt. of Karnataka)
Yelahanka, Bengaluru-560064

DEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING

BIGDATA LAB

LA II - Programming Assignment

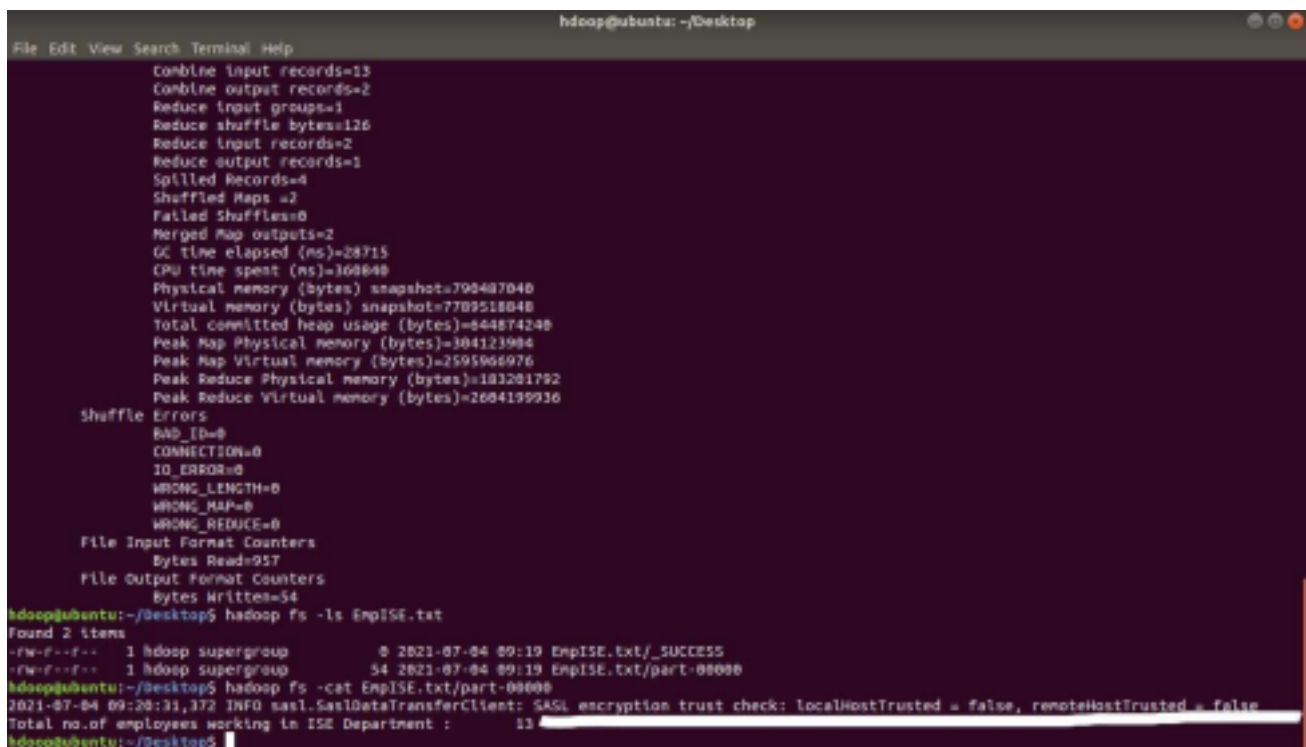
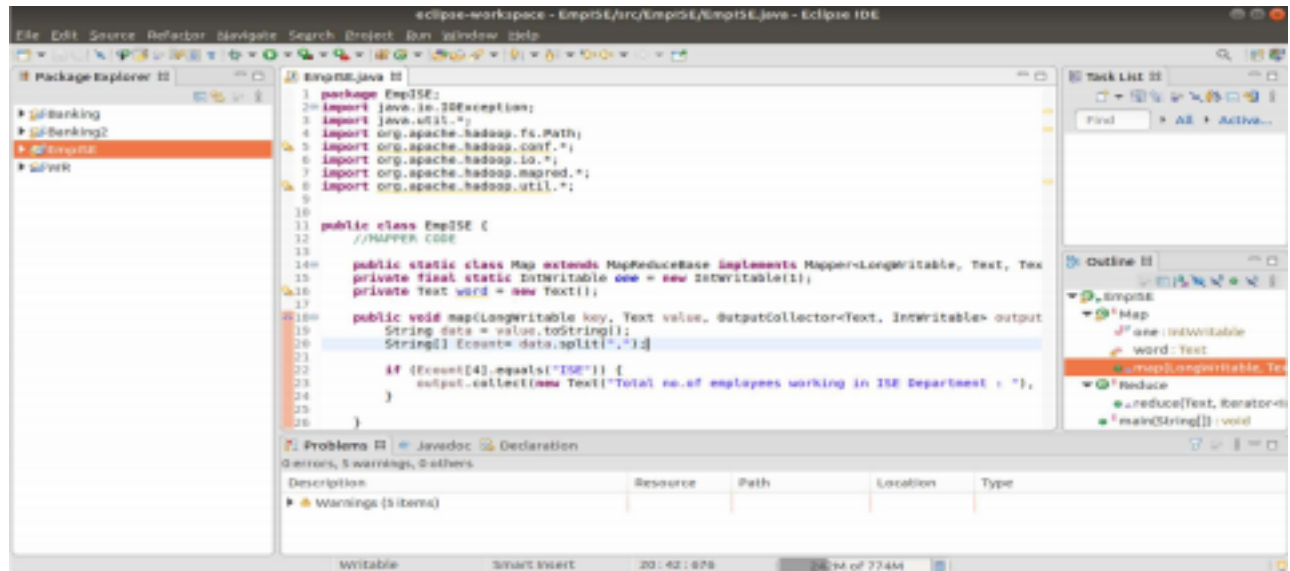
Navneeth Bhat

1NT17IS106

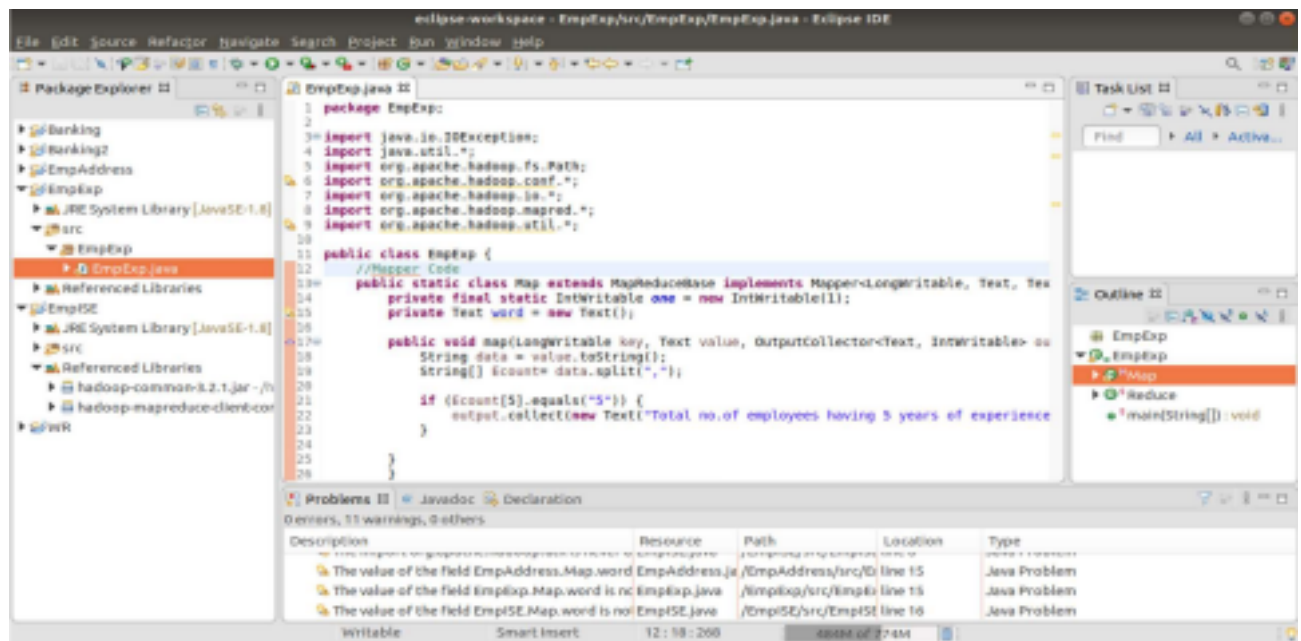
Github Link: <https://github.com/1nt17is106/Big-Data-New.git>

Exercise-I

1) Total number of employees who works in ISE Department.



2) Total number of employees with experience = 5 years.



```

hadoop@ubuntu: ~/Desktop
File Edit View Search Terminal Help
Combine input records=10
Combine output records=2
Reduce input groups=1
Reduce shuffle bytes=132
Reduce input records=2
Reduce output records=1
Spilled Records=4
Shuffled Maps =2
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=4467
CPU time spent (ms)=45330
Physical memory (bytes) snapshot=812605440
Virtual memory (bytes) snapshot=7793565696
Total committed heap usage (bytes)=626524160
Peak Map Physical memory (bytes)=314843136
Peak Map Virtual memory (bytes)=2597437440
Peak Reduce Physical memory (bytes)=182988800
Peak Reduce Virtual memory (bytes)=2608173568

Shuffle Errors
BAD_ID=0
CONNECTIO=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0

File Input Format Counters
  Bytes Read=957
File Output Format Counters
  Bytes Written=57

hadoop@ubuntu:~/Desktop$ hadoop fs -ls EmpExp.txt
Found 2 items
-rw-r--r-- 1 hadoop supergroup 0 2021-07-04 09:58 EmpExp.txt/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 57 2021-07-04 09:58 EmpExp.txt/part-00000

hadoop@ubuntu:~/Desktop$ hadoop fs -cat EmpExp.txt/part-00000
2021-07-04 09:58:55,286 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
Total no.of employees having 5 years of experience : 10

hadoop@ubuntu:~/Desktop$

```

3) Count the number of employees who lives in Bangalore.

```

eclipse-workspace - EmpAddress/src/EmpAddress/EmpAddress.java - Eclipse IDE
File Edit Source Refactor Navigate Search Project Run Window Help

Package Explorer
  Banking
  Banking2
  EmpAddress
    JRE System Library [JavaSE-1.8]
    src
      EmpAddress
        EmpAddress.java
      Referenced Libraries
    EmpExp
      JRE System Library [JavaSE-1.8]
      src
      Referenced Libraries
    EmpISE
      JRE System Library [JavaSE-1.8]
      src
      Referenced Libraries
    WR

EmpAddress.java
1 package EmpAddress;
2
3 import java.io.IOException;
4 import java.util.*;
5 import org.apache.hadoop.fs.Path;
6 import org.apache.hadoop.conf.*;
7 import org.apache.hadoop.io.*;
8 import org.apache.hadoop.mapred.*;
9 import org.apache.hadoop.util.*;
10
11 public class EmpAddress {
12     public static class Map extends MapReduceBase implements Mapper<LongWritable, Text, Text> {
13         private final static IntWritable one = new IntWritable(1);
14         private Text word = new Text();
15
16         public void map(LongWritable key, Text value, OutputCollector<Text, IntWritable> output, Context context) throws IOException, InterruptedException {
17             String data = value.toString();
18             String[] Ecount= data.split(",");
19
20             if (Ecount[3].equals("Bangalore")) {
21                 output.collect(new Text("Total no.of employees who lives in Bangalore : "), one);
22             }
23         }
24     }
25 }
26

Problems
0 errors, 11 warnings, 0 others
Description Resource Path Location Type
The value of the field EmpAddress.Map.word EmpAddress.java /EmpAddress/src/Emp line 14 Java Problem
The value of the field EmpExp.Map.word is no EmpExp.java /EmpExp/src/EmpExp line 14 Java Problem
The value of the field EmpISE.Map.word is no EmpISE.java /EmpISE/src/EmpISE line 16 Java Problem
Writable Smart Insert 9:33:234 762M of 774M

```


Exercise-II – HIVE:

HIVE QUERIES RESULTS AND SNAPSHOTS

- 1) Insert 5 records using INSERT command.


```

hadoop@ubuntu: ~/apache-hive-3.12-bin
File Edit View Search Terminal Help
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1625316400333_0006
Hadoop Job Information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-03 07:15:18,808 Stage-1 map = 0%, reduce = 0%
2021-07-03 07:15:22,709 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.36 sec
2021-07-03 07:15:27,828 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.91 sec
MapReduce Total cumulative CPU time: 2 seconds 910 msec
Ended Job = job_1625316400333_0006
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.91 sec HDFS Read: 12974 HDFS Write: 1228 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 910 msec
OK
Harsha 5000 30000.0 Bangalore ISE 5
Adarsh 5001 35000.0 Sirci ISE 6
Nandan 5002 36000.0 Bangalore ISE 6
Raghav 5003 48000.0 Hassan CSE 6
Dharni 5004 41000.0 Chennai ECE 6
Aniket 5005 45000.0 Hyderabad ME 6
Darshan 5006 46000.0 Hosnagar ISE 5
Asha 5007 47000.0 Puttur ISE 5
Harshith 5008 20000.0 Tunkur ECE 7
Divine 5009 80000.0 Udipi ISE 5
Aditya 5010 50000.0 Bhadra ISE 6
Ashok 5011 24000.0 Bangalore ISE 5
Sujan 5012 25000.0 Mangalore CSE 5
Sanket 5013 26000.0 Mangalore CSE 6
Bharat 5014 27000.0 Gurgaon ISE 7
Abhay 5015 28000.0 Mumbai ISE 5
Sandeep 5016 30000.0 Shlnaga CSE 7
Alok 5017 60000.0 Bihar ISE 5
Navani 5018 61000.0 Bangalore ISE 5
Anjali 5019 64000.0 Delhi ISE 5
Ayesha 5020 15000.0 Dehradun ISE 7
Alshwarya 5021 20000.0 Mysore ME 4
Ajay 5022 25000.0 KGF CSE 7
Gagan 5023 80000.0 Mandya ECE 4
Guru 5024 75000.0 Dharwad AE 6
Time taken: 19.008 seconds, Fetched: 25 row(s)
hive>

```

- 2) Demonstrate the Alter command for the following cases,
 - a. Rename the table name to "Emp".
 - b. Rename the column name "Dname" to "Dept_name".

```

hadoop@ubuntu: ~/apache-hive-3.12-bin
File Edit View Search Terminal Help
Gagan 5023 80000.0 Mandya ECE 4
Guru 5024 75000.0 Dharwad AE 6
Time taken: 19.008 seconds, Fetched: 25 row(s)
hive> show tables;
OK
employee
Time taken: 0.2 seconds, Fetched: 1 row(s)
hive> alter table Employee rename to Emp;
OK
Time taken: 0.224 seconds
hive> show tables;
OK
emp
Time taken: 0.029 seconds, Fetched: 1 row(s)
hive> desc emp;
OK
name                string
ssn                  int
salary               float
address              string
dname                string
experience            int
Time taken: 0.041 seconds, Fetched: 6 row(s)
hive> alter table Employee change dname Deptname string;
FAILED: SemanticException [Error 10001]: Table not found Employee
hive> alter table Emp change dname Deptname string;
OK
Time taken: 0.127 seconds
hive> desc emp;
OK
name                string
ssn                  int
salary               float
address              string
deptname             string
experience            int
Time taken: 0.031 seconds, Fetched: 6 row(s)
hive>

```

- 3) Retrieve all the employees who's salary is not less than 50000.

```

hadoop@ubuntu: ~/apache-hive-3.12-bin
File Edit View Search Terminal Help
name          string
ssn           int
salary       float
address       string
deptname      string
experience     int
Time taken: 0.031 seconds, Fetched: 6 row(s)
hive> select Name,SSN,salary from emp where salary>=50000;
OK
Gagan  5023  80000.0
Guru   5024  75000.0
Divine 5009  90000.0
Aditya 5010  50000.0
Alok   5017  60000.0
Navani 5018  61000.0
Anjali 5019  64000.0
Time taken: 1.343 seconds, Fetched: 7 row(s)
hive> select Name,address,experience from emp where address="Bangalore" and experience<5;
OK
Time taken: 0.897 seconds
hive> insert into Employee values("Akash",5025,25000.0,"Bangalore","CSE",3);
FAILED: SemanticException [Error 10801]: Line 1:12 Table not found 'Employee'
hive> insert into Emp values("Akash",5025,25000.0,"Bangalore","CSE",3);
Query ID = hadoop_20210703073349_b409eacc-9776-4379-80e5-c32bd0b1d9cd
Total jobs = 3
Launching Job 1 out of 3
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>
Starting Job = job_1625316400333_0007, Tracking URL = http://ubuntu:8086/proxy/application_1625316400333_0007/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1625316400333_0007
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-03 07:34:34,550 Stage-1 map = 0%, reduce = 0%
2021-07-03 07:35:14,910 Stage-1 map = 0%, reduce = 0%

```

4) Extract all employees who live in Bangalore but having less than 5 years of experience.

```

hive> select Name,address,experience from emp where address="Bangalore" and experience<5;
OK
Akash  Bangalore  3
Time taken: 0.337 seconds, Fetched: 1 row(s)
hive> create view Emp_Details as select Name,Deptname from emp;
OK
Time taken: 1.712 seconds
hive> select * from Emp_Details;
OK
Aayesha ISE
Aishwarya ME
Ajay CSE
Gagan ECE
Guru AE
Akash CSE
Harsha ISE
AdTerminal SE
Nandan ISE
Raghav CSE
Dharni ECE
Aniket ME
Darshan ISE
Asha ISE
Harshith ECE
Divine ISE
Aditya ISE
Ashok ISE
Sujan CSE
Sanket CSE
Bharat ISE
Abhay ISE
Sandeep CSE
Alok ISE
Navani ISE
Anjali ISE
Time taken: 0.812 seconds, Fetched: 26 row(s)

```

5) Create separate view containing Name, Dept_name of employees.


```

hive> select Name,address,experience from emp where address="Bangalore" and experience<5;
OK
Akash    Bangalore    3
Time taken: 0.337 seconds, Fetched: 1 row(s)
hive> create view Emp_Details as select Name,Deptname from emp;
OK
Time taken: 1.712 seconds
hive> select * from Emp_Details;
OK
Aayesha  ISE
Alshwarya ME
Ajay     CSE
Gagan    ECE
Guru     AE
Akash    CSE
Harsha   ISE
Adarsh   ISE
Nandan   ISE
Raghav   CSE
Dharni   ECE
Aniket   ME
Darshan  ISE
Asha     ISE
Harshith ECE
Divine   ISE
Aditya   ISE
Ashok    ISE
Sujan    CSE
Sanket   CSE
Bharat   ISE
Abhay    ISE
Sandeep  CSE
Alok     ISE
Navani   ISE
Anjali   ISE
Time taken: 0.812 seconds, Fetched: 26 row(s)

```

6) Display Name and SSN and use group by SSN and order by Name.

```

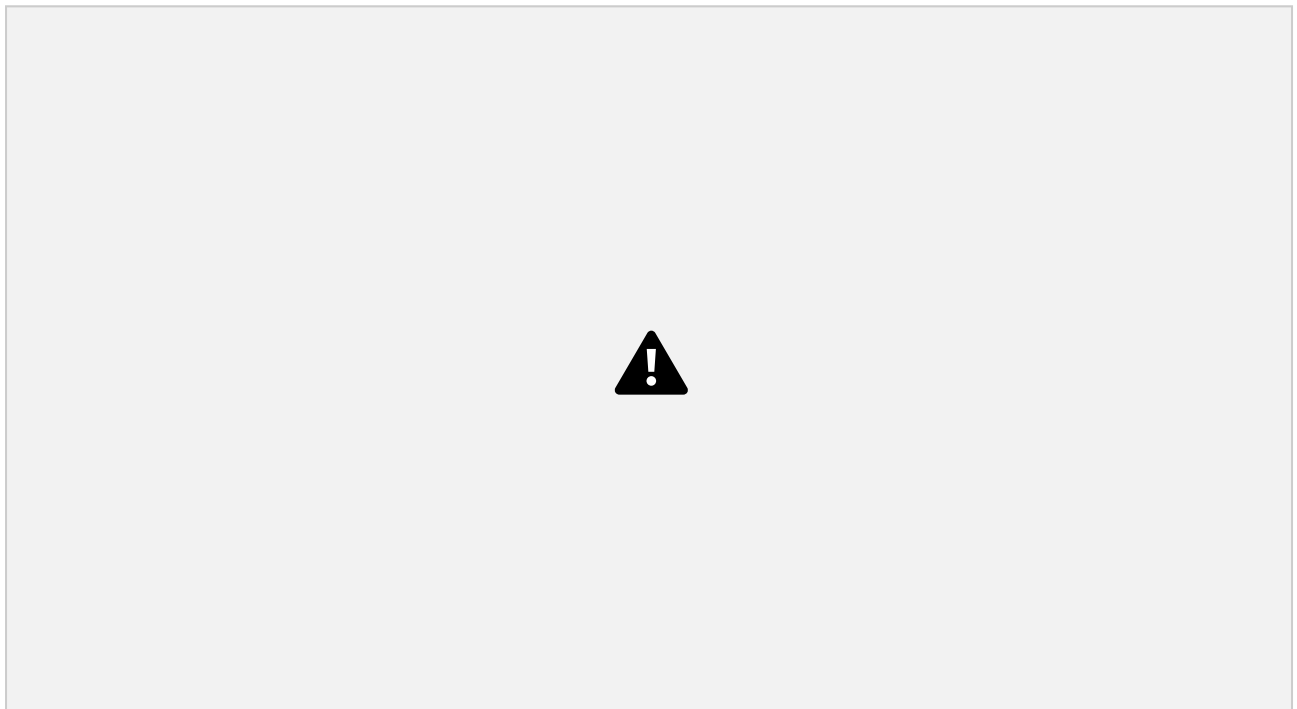
hdop@ubuntu: ~/hadoop-hive-3.12.0-bin
File Edit View Search Terminal Help
Sanket 5013
Sujan 5012
Time taken: 41.413 seconds, Fetched: 26 row(s)
hive> select name,ssn from emp group by ssn order by name;
Query ID = hdop_20210703084449_b69f2eca-0a4c-4f0b-a74c-6d8c8fc9db08
Total jobs = 2
Launching Job 1 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>
Starting Job = job_1625326304682_0004, Tracking URL = http://ubuntu:8088/proxy/application_1625326304682_0004/
Kill Command = /home/hdop/hadoop-3.2.1/bin/killall job_1625326304682_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-03 08:44:55,213 Stage-1 map = 0%, reduce = 0%
2021-07-03 08:44:59,312 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.47 sec
2021-07-03 08:45:04,445 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.78 sec
MapReduce Total cumulative CPU time: 2 seconds 780 msec
Ended Job = job_1625326304682_0004
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>
Starting Job = job_1625326304682_0005, Tracking URL = http://ubuntu:8088/proxy/application_1625326304682_0005/
Kill Command = /home/hdop/hadoop-3.2.1/bin/killall job_1625326304682_0005
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2021-07-03 08:45:16,443 Stage-2 map = 0%, reduce = 0%
2021-07-03 08:45:20,576 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 1.2 sec

```

7) Retrieve Maximum salary, minimum salary and Average salary of the employees.

```
hadoop@ubuntu: ~/apache-hive-3.12.0-bin
File Edit View Search Terminal Help
Harsha 5000
Harshith 5000
Nandan 5002
Navami 5010
Raghar 5003
Sandeep 5016
Sanket 5013
Sujan 5012
Time taken: 37.243 seconds, Fetched: 26 row(s)
hive> select max(salary),min(salary),avg(salary) from emp;
Query ID = hadoop_20210703004736_dfc5874b-032d-437a-b46e-3a2ef96cbe99
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>
Starting Job = job_1625326304682_0000, Tracking URL = http://ubuntu:8080/proxy/application_1625326304682_0000/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1625326304682_0000
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-03 00:47:42,349 Stage-1 map = 0%, reduce = 0%
2021-07-03 00:47:47,497 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.06 sec
2021-07-03 00:47:53,058 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.19 sec
MapReduce Total cumulative CPU time: 5 seconds 190 msec
Ended Job = job_1625326304682_0000
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.19 sec HDFS Read: 10503 HDFS Write: 133 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 190 msec
OK
00000.0 15000.0 40576.92307692308
Time taken: 18.57 seconds, Fetched: 1 row(s)
hive>
```

8) Create Another table called Department and Inserting values.
(Dname = Deptname(employee))



9) Performing outer join over Dname.



10) Performing on left outer join over Dname



11) Performing on right outer join over Dname.

