



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 LABORATORY

Report on,

Learning Activity II - Programming Assignment

Submitted by,

Aakash N.S

1NT18IS002

HADOOP PROBLEM STATEMENT

Exercise-I

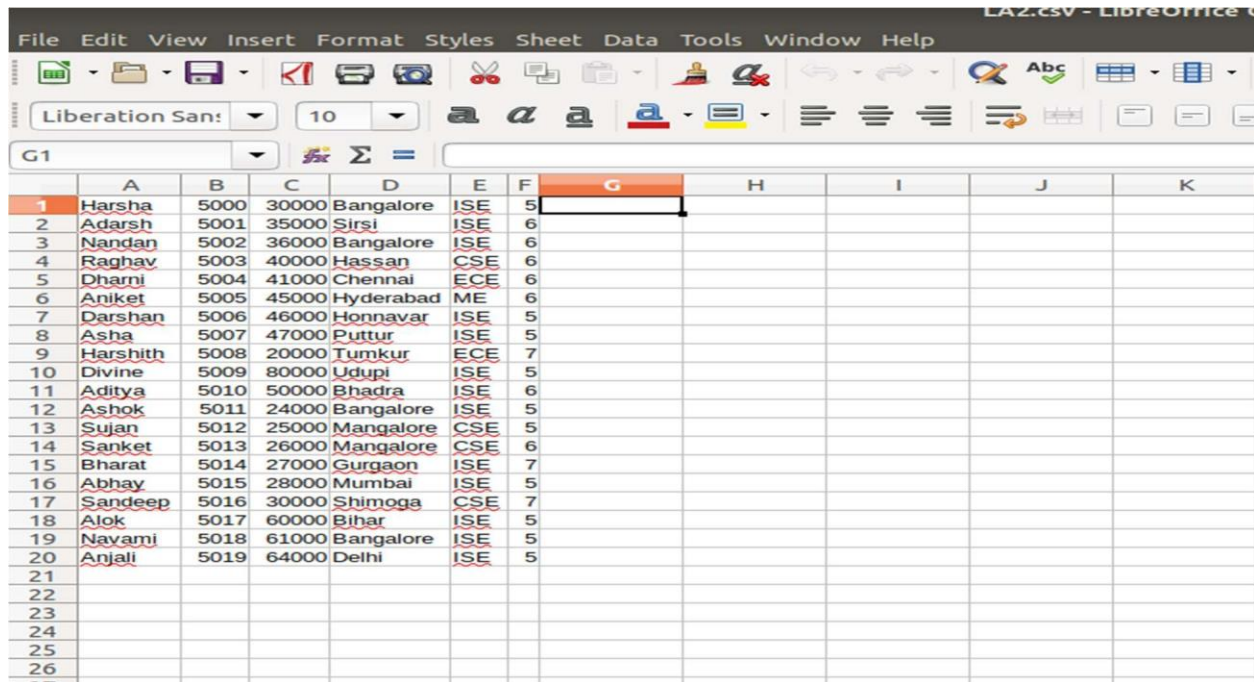
Create a dataset in excel as .csv file and it should contain the following fields with at least 20 sample datasets in it.

Name	SSN	Salary	Address	Dname	Experience
Harsha	5000	30000	Bangalore	ISE	5

Use the Hadoop MapReduce programming framework to come up with a Program which will take the data from this .csv file and computes the following.

1. Total number of employees who works in ISE Department.
2. Total number of employees with experience = 5 years.
3. Count the number of employees who lives in Bangalore

DATASET DESCRIPTION

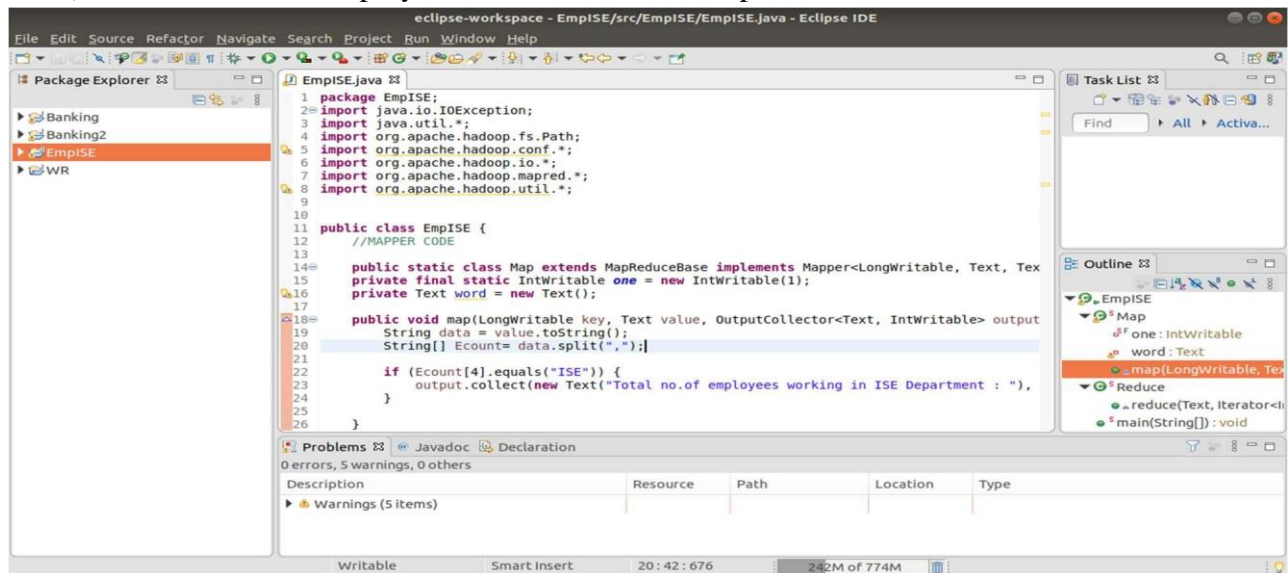


The screenshot shows a LibreOffice spreadsheet with a dataset of 20 employees. The columns are labeled A through K. The data is as follows:

	A	B	C	D	E	F	G	H	I	J	K
1	Harsha	5000	30000	Bangalore	ISE	5					
2	Adarsh	5001	35000	Sirsi	ISE	6					
3	Nandan	5002	36000	Bangalore	ISE	6					
4	Raghav	5003	40000	Hassan	CSE	6					
5	Dharmi	5004	41000	Chennai	ECE	6					
6	Aniket	5005	45000	Hyderabad	ME	6					
7	Darshan	5006	46000	Honnavar	ISE	5					
8	Asha	5007	47000	Puttur	ISE	5					
9	Harshith	5008	20000	Tumkur	ECE	7					
10	Divine	5009	80000	Udupi	ISE	5					
11	Aditya	5010	50000	Bhadra	ISE	6					
12	Ashok	5011	24000	Bangalore	ISE	5					
13	Sujan	5012	25000	Mangalore	CSE	5					
14	Sanket	5013	26000	Mangalore	CSE	6					
15	Bharat	5014	27000	Gurgaon	ISE	7					
16	Abhay	5015	28000	Mumbai	ISE	5					
17	Sandeep	5016	30000	Shimoga	CSE	7					
18	Alok	5017	60000	Bihar	ISE	5					
19	Navami	5018	61000	Bangalore	ISE	5					
20	Anjali	5019	64000	Delhi	ISE	5					
21											
22											
23											
24											
25											
26											

RESULTS AND SNAPSHOTS OF HADOOP PROGRAM

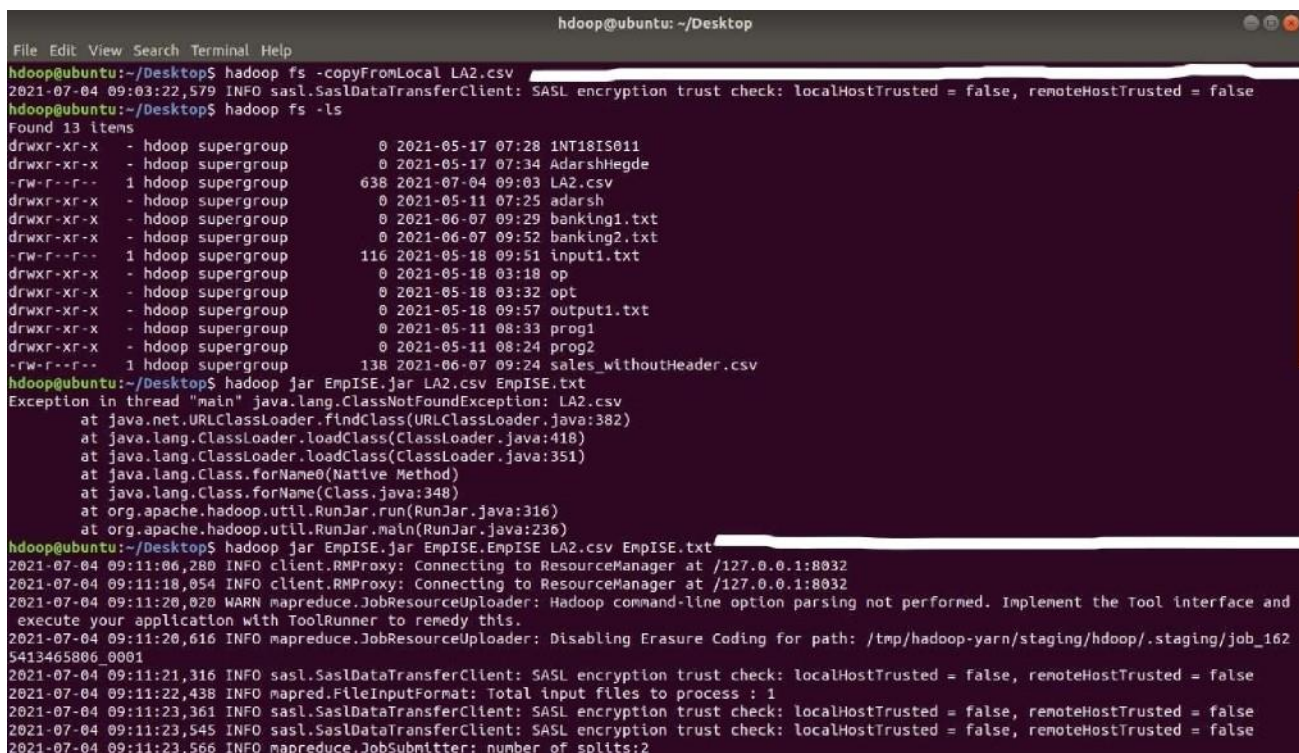
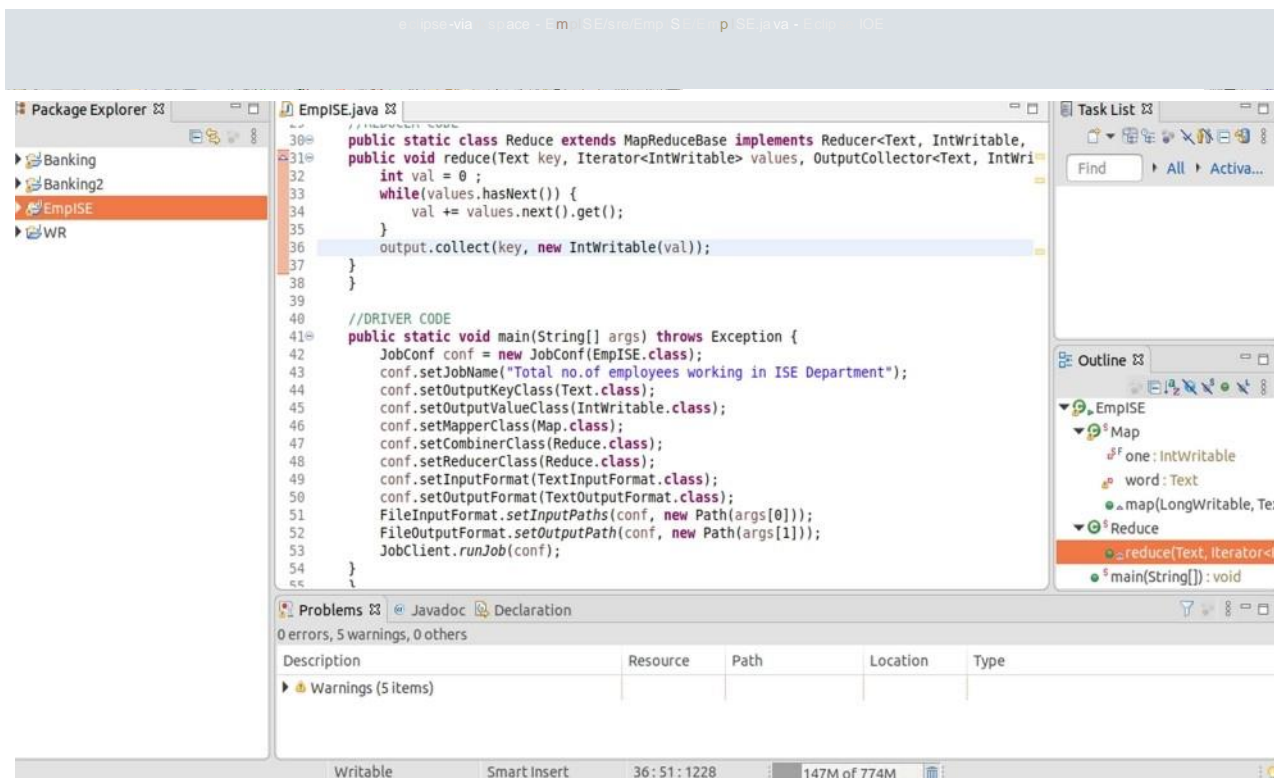
- 1) Total number of employees who works in ISE Department.



The screenshot shows the Eclipse IDE with the EmpISE.java file open. The code is as follows:

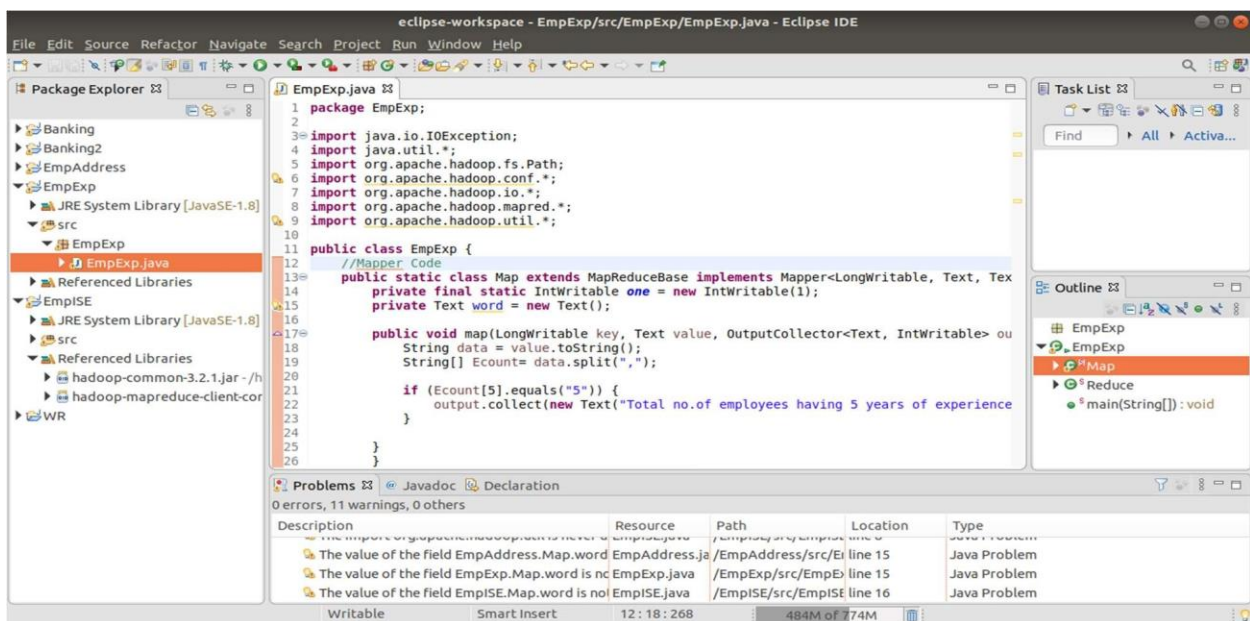
```
1 package EmpISE;
2 import java.io.IOException;
3 import java.util.*;
4 import org.apache.hadoop.fs.Path;
5 import org.apache.hadoop.conf.*;
6 import org.apache.hadoop.io.*;
7 import org.apache.hadoop.mapred.*;
8 import org.apache.hadoop.util.*;
9
10
11 public class EmpISE {
12     //MAPPER CODE
13
14     public static class Map extends MapReduceBase implements Mapper<LongWritable, Text, Text> {
15         private final static IntWritable one = new IntWritable(1);
16         private Text word = new Text();
17
18         public void map(LongWritable key, Text value, OutputCollector<Text, IntWritable> output
19             throws IOException, InterruptedException {
20             String data = value.toString();
21             String[] Ecount = data.split(",");
22
23             if (Ecount[4].equals("ISE")) {
24                 output.collect(new Text("Total no.of employees working in ISE Department : "),
25                     one);
26             }
27         }
28     }
29 }
```

The Problems window shows 5 warnings, 0 errors, and 0 others.



```
hadoop@ubuntu: ~/Desktop
File Edit View Search Terminal Help
Combine input records=13
Combine output records=2
Reduce input groups=1
Reduce shuffle bytes=126
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)=28715
CPU time spent (ms)=360840
Physical memory (bytes) snapshot=790487040
Virtual memory (bytes) snapshot=7789518848
Total committed heap usage (bytes)=644874240
Peak Map Physical memory (bytes)=304123904
Peak Map Virtual memory (bytes)=2595966976
Peak Reduce Physical memory (bytes)=183201792
Peak Reduce Virtual memory (bytes)=2604199936
Shuffle Errors
BAD_ID=0
CONNECTION=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=54
hadoop@ubuntu:~/Desktop$ hadoop fs -ls EmpISE.txt
Found 2 items
-rw-r--r-- 1 hadoop supergroup 0 2021-07-04 09:19 EmpISE.txt/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 54 2021-07-04 09:19 EmpISE.txt/part-00000
hadoop@ubuntu:~/Desktop$ hadoop fs -cat EmpISE.txt/part-00000
2021-07-04 09:20:31,372 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
Total no.of employees working in ISE Department : 13
hadoop@ubuntu:~/Desktop$
```

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



eclipse-workspace - EmpExp/src/EmpExp/EmpExp.java - Eclipse IDE

File Edit Source Refactor Navigate Search Project Run Window Help

Package Explorer B 0,2 EmpExp.java zz - a \$)TaskList

Banking
Banking2
EmpAddress
MJRE System Library [JavaSE-1.8]
src
EmpExp
EmpExp.java
Referenced Libraries
EmpISE
MJRE System Library [JavaSE-1.8]
Referenced Libraries
hadoop-common-3.2.1.jar - /h
hadoop-mapreduce-client-core
wR

```

29- public static class Reduce extends MapReduceBase implements ReducersText, IntWritable
30- public void reduce(Text key, List<IntWritable> values, OutputCollector<Text>,
31-                    Configuration conf) throws IOException {
32-     while (values.hasNext()) {
33-         IntWritable value = values.next().get();
34-         output.collect(key, new IntWritable(value));
35-     }
36- }
37-
38- //DRIVER CODE
39- public static void main(String[] args) throws Exception {
40-     JobConf conf = new JobConf(EmpExp.class);
41-     conf.setJobName("Total no. of employees having 5 years of experience");
42-     conf.setOutputKeyClass(Text.class);
43-     conf.setOutputValueClass(IntWritable.class);
44-     conf.setMapperClass(EMPMap.class);
45-     conf.setReducerClass(Reduce.class);
46-     conf.setCombinerClass(Reduce.class);
47-     conf.setInputFormat(TextInputFormat.class);
48-     conf.setOutputFormat(TextOutputFormat.class);
49-     FileInputFormat.setInputPaths(conf, new Path(args[0]));
50-     FileOutputFormat.setOutputPath(conf, new Path(args[1]));
51-     JobClient.runJob(conf);
52- }
53-
54- Problem's Et e Javadoc @ declaration
55- java.lang.IllegalArgumentException: warnin@s..0 éthers

```

Outline

- EMPMap
- EMPMap
- Map
- Reduce
- main(Strings) :void

Description	Resource	Path	Location	Type
The value of the field EmpAddress.Map.word is not	EmpAddress.java	/EmpAddress/src/EmpAddress	line 15	Java Problem
The value of the field EmpExp.Map.word is not	EmpExp.java	/EmpExp/src/EmpExp	line 15	Java Problem
The value of the field EmpISE.Map.word is not	EmpISE.java	/EmpISE/src/EmpISE	line 16	Java Problem

Writable Smart Insert 12:18:268 79M of 774M

hadoop@ubuntu: ~/Desktop

File Edit View Search Terminal Help

```

hadoop@ubuntu:~/Desktop$ hadoop fs -cat EmpISE.txt/part-00000
2021-07-04 09:20:31,372 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
Total no. of employees working in ISE Department : 13
hadoop@ubuntu:~/Desktop$ hadoop jar EmpExp.jar EmpExp,EmpExp LA2.csv EmpExp.txt
2021-07-04 09:56:04,465 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-04 09:56:09,024 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-04 09:56:10,405 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and
execute your application with ToolRunner to remedy this.
2021-07-04 09:56:10,709 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_162
5413465806_0002
2021-07-04 09:56:11,286 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-04 09:56:12,643 INFO mapred.FileInputFormat: Total input files to process : 1
2021-07-04 09:56:12,836 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-04 09:56:12,862 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-04 09:56:12,989 INFO mapreduce.JobSubmitter: number of splits:2
2021-07-04 09:56:13,318 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-04 09:56:13,366 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1625413465806_0002
2021-07-04 09:56:13,366 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-07-04 09:56:13,903 INFO conf.Configuration: resource-types.xml not found
2021-07-04 09:56:13,904 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-07-04 09:56:16,566 INFO impl.YarnClientImpl: Submitted application application_1625413465806_0002
2021-07-04 09:56:18,577 INFO mapreduce.Job: The url to track the job: http://ubuntu:8088/proxy/application_1625413465806_0002/
2021-07-04 09:56:18,712 INFO mapreduce.Job: Running job: job_1625413465806_0002
2021-07-04 09:57:23,814 INFO mapreduce.Job: Job job_1625413465806_0002 running in uber mode : false
2021-07-04 09:57:23,818 INFO mapreduce.Job: map 0% reduce 0%
2021-07-04 09:58:21,633 INFO mapreduce.Job: map 100% reduce 0%
2021-07-04 09:58:26,679 INFO mapreduce.Job: map 100% reduce 100%
2021-07-04 09:58:27,708 INFO mapreduce.Job: Job job_1625413465806_0002 completed successfully
2021-07-04 09:58:27,841 INFO mapreduce.Job: Counters: 55
File System Counters
FILE: Number of bytes read=126
FILE: Number of bytes written=677738
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=1141
HDFS: Number of bytes written=57
HDFS: Number of read operations=11

```

```
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)=2600173568

Shuffle Errors
BAD_ID=0
CONNECTION=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 09:58:55,206 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> ou
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 : "),
22
23             }
24         }
25     }
26 }
```

Task List

Find All Active...

Outline

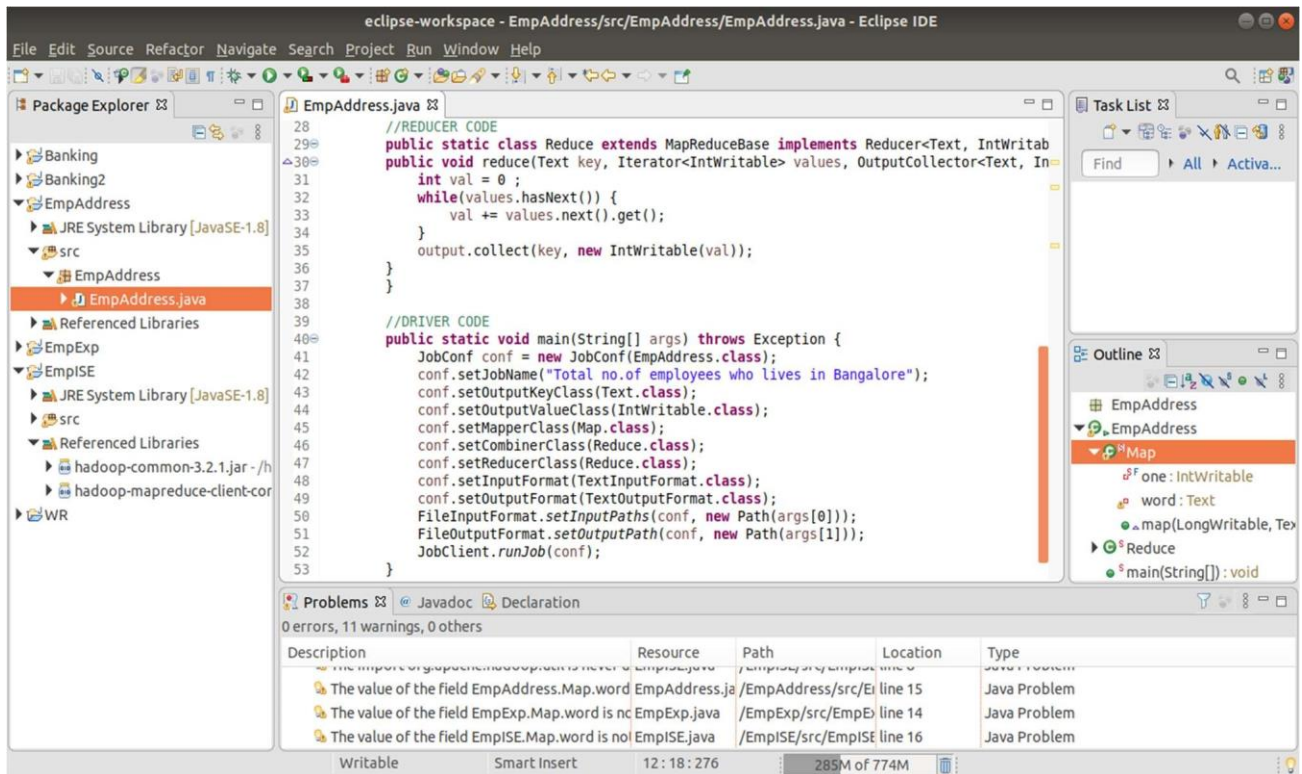
- EmpAddress
- EmpAddress
 - Map
 - one: IntWritable
 - word: Text
 - map(LongWritable, Text)
 - Reduce
 - main(String[]): void

Problems

0 errors, 11 warnings, 0 others

Description	Resource	Path	Location	Type
The value of the field EmpAddress.Map.word is not	EmpAddress.java	/EmpAddress/src/Emp	line 14	Java Problem
The value of the field EmpExp.Map.word is not	EmpExp.java	/EmpExp/src/EmpE	line 14	Java Problem
The value of the field EmpISE.Map.word is not	EmpISE.java	/EmpISE/src/EmpISE	line 16	Java Problem

Writable Smart Insert 9:33:234 162M of 774M



```

hadoop@ubuntu: ~/Desktop
File Edit View Search Terminal Help
-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 10:00:36,206 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
Total no. of employees having 5 years of experience : 10
hadoop@ubuntu:~/Desktop$ hadoop jar EmpAddress.jar EmpAddress.EmpAddress LA2.csv EmpAddress.txt
2021-07-04 10:00:36,404 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-04 10:00:36,566 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-04 10:00:36,727 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and
execute your application with ToolRunner to remedy this.
2021-07-04 10:00:36,781 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_162
5413465806_0003
2021-07-04 10:00:36,873 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2021-07-04 10:00:37,008 INFO mapred.FileInputFormat: Total input files to process : 1
2021-07-04 10:00:37,032 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2021-07-04 10:00:37,069 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2021-07-04 10:00:37,077 INFO mapreduce.JobSubmitter: number of splits:2
2021-07-04 10:00:37,177 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2021-07-04 10:00:37,657 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1625413465806_0003
2021-07-04 10:00:37,658 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-07-04 10:00:37,833 INFO conf.Configuration: resource-types.xml not found
2021-07-04 10:00:37,834 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-07-04 10:00:37,935 INFO impl.YarnClientImpl: Submitted application application_1625413465806_0003
2021-07-04 10:00:38,102 INFO mapreduce.Job: The url to track the job: http://ubuntu:8088/proxy/application_1625413465806_0003/
2021-07-04 10:00:38,103 INFO mapreduce.Job: Running job: job_1625413465806_0003
2021-07-04 10:00:43,215 INFO mapreduce.Job: Job job_1625413465806_0003 running in uber mode : false
2021-07-04 10:00:43,218 INFO mapreduce.Job: map 0% reduce 0%
2021-07-04 10:00:48,291 INFO mapreduce.Job: map 100% reduce 0%
2021-07-04 10:00:52,327 INFO mapreduce.Job: map 100% reduce 100%
2021-07-04 10:00:53,361 INFO mapreduce.Job: Job job_1625413465806_0003 completed successfully
2021-07-04 10:00:53,449 INFO mapreduce.Job: Counters: 54
File System Counters
  FILE: Number of bytes read=114
  FILE: Number of bytes written=677780
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=1141
  HDFS: Number of bytes written=50

```



```

hadoop@ubuntu: ~/Desktop
File Edit View Search Terminal Help
Peak Reduce Virtual memory (bytes)=2602868736
Shuffle Errors
BAD_ID=0
CONNECTION=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=50
hadoop@ubuntu:~/Desktop$ hadoop fs -ls EmpAddress.txt
Found 2 items
-rw-r--r-- 1 hadoop supergroup 0 2021-07-04 10:00 EmpAddress.txt/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 50 2021-07-04 10:00 EmpAddress.txt/part-00000
hadoop@ubuntu:~/Desktop$ hadoop fs -cat EmpAddress.txt/part-00000
2021-07-04 10:01:18,780 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
Total no.of employees who lives in Bangalore : 4
hadoop@ubuntu:~/Desktop$ hadoop fs -ls
Found 16 items
drwxr-xr-x - hadoop supergroup 0 2021-05-17 07:28 INT18IS011
drwxr-xr-x - hadoop supergroup 0 2021-05-17 07:34 AdarshHegde
drwxr-xr-x - hadoop supergroup 0 2021-07-04 10:00 EmpAddress.txt
drwxr-xr-x - hadoop supergroup 0 2021-07-04 09:58 EmpExp.txt
drwxr-xr-x - hadoop supergroup 0 2021-07-04 09:19 EmpISE.txt
-rw-r--r-- 1 hadoop supergroup 638 2021-07-04 09:03 LA2.csv
drwxr-xr-x - hadoop supergroup 0 2021-05-11 07:25 adarsh
drwxr-xr-x - hadoop supergroup 0 2021-06-07 09:29 banking1.txt
drwxr-xr-x - hadoop supergroup 0 2021-06-07 09:52 banking2.txt
-rw-r--r-- 1 hadoop supergroup 116 2021-05-18 09:51 input1.txt
drwxr-xr-x - hadoop supergroup 0 2021-05-18 03:18 op
drwxr-xr-x - hadoop supergroup 0 2021-05-18 03:32 opt
drwxr-xr-x - hadoop supergroup 0 2021-05-18 09:57 output1.txt
drwxr-xr-x - hadoop supergroup 0 2021-05-11 08:33 prog1
drwxr-xr-x - hadoop supergroup 0 2021-05-11 08:24 prog2
-rw-r--r-- 1 hadoop supergroup 138 2021-06-07 09:24 sales_withoutHeader.csv
hadoop@ubuntu:~/Desktop$

```

QUERIES OF THE USE-CASES

Exercise-II – HIVE:

Use the above dataset in .csv file and create a database called as EmployeeDB. Create a table under the database called as Employee using HIVEQL. The table fields are same, that is,

Name	SSN	Salary	Address	Dname	Experience
Harsha	5000	30000	Bangalore	ISE	5

Use the HiveQL language to perform the following Query based Map-reduce operations,

1. Insert 5 records using INSERT command.
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”.

3. Retrieve all the employees who's salary is not less than 50000.
4. Extract all employees who live in Bangalore but having less than 5 years of experience.
5. Create separate view containing Name, Dept_name of employees
6. Display Name and SSN and use group by SSN and order by Name
7. Retrieve Maximum salary, minimum salary and Average salary of the employees.
8. Create Another table called Department with the following fields (Dname = Dept_name and perform the following joins (outer, left outer, right outer) over Dname

Dno	Dname
6	ISE

HIVE QUERIES RESULTS AND SNAPSHOTS

- 1) Insert 5 records using INSERT command.

```

hadoop@ubuntu: ~/apache-hive-3.1.2-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 07:15:18,608 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 Sirsi ISE 6
Nandan 5002 36000.0 Bangalore ISE 6
Raghav 5003 40000.0 Hassan CSE 6
Dharni 5004 41000.0 Chennai ECE 6
Aniket 5005 45000.0 Hyderabad ME 6
Darshan 5006 46000.0 Honnavar ISE 5
Asha 5007 47000.0 Puttur ISE 5
Harshith 5008 20000.0 Tumkur ECE 7
Divine 5009 80000.0 Udupi 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 Shimoga CSE 7
Alok 5017 60000.0 Bihar ISE 5
Navami 5018 61000.0 Bangalore ISE 5
Anjali 5019 64000.0 Delhi ISE 5
Aayesha 5020 15000.0 Dehradun ISE 7
Aishwarya 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: 19.088 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-3.1.2-bin
File Edit View Search Terminal Help
Gagan 5023 80000.0 Mandya ECE 4
Guru 5024 75000.0 Dharwad AE 6
Time taken: 19.088 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-3.1.2-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 80000.0
Aditya 5010 50000.0
Alok 5017 60000.0
Navami 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 10001]: Line 1:12 Table not found 'Employee'
hive> insert into Emp values("Akash",5025,25000.0,"Bangalore","CSE",3);
Query ID = hadoop_20210703073349_b4b9eacc-9778-4379-80a5-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:8088/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 07:35:34,930 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
Aadithya 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
Navami  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
Aishwarya    ME
Ajay    CSE
Gagan   ECE
Guru    AE
Akash   CSE
Harsha  ISE
Aadithya 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
Navami  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.

```
hadoop@ubuntu: ~/apache-hive-3.1.2-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 name,ssn order by name;
Query ID = hdoop_20210703084449_b69f2eca-0a4c-4f0b-a74c-6d6c8fc9d8b8
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/hadoop/hadoop-3.2.1/bin/mapred job -kill 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/hadoop/hadoop-3.2.1/bin/mapred job -kill 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-3.1.2-bin
File Edit View Search Terminal Help
Harsha 5000
Harshith 5008
Nandan 5002
Navami 5018
Raghav 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 = hdoop_20210703084736_dfc5874b-032d-437a-b46e-3a2ef96cba99
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_0006, Tracking URL = http://ubuntu:8088/proxy/application_1625326304682_0006/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1625326304682_0006
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-03 08:47:42,349 Stage-1 map = 0%, reduce = 0%
2021-07-03 08:47:47,497 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.66 sec
2021-07-03 08:47:53,658 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.19 sec
MapReduce Total cumulative CPU time: 5 seconds 190 msec
Ended Job = job_1625326304682_0006
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.19 sec HDFS Read: 18503 HDFS Write: 133 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 190 msec
OK
80000.0 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))

```

hadoop@ubuntu: ~/apache-hive-3.1.2-bin
File Edit View Search Terminal Help
Time taken: 18.57 seconds, Fetched: 1 row(s)
hive> create table department(dno int,dname string)row format delimited fields terminated by ",";
OK
Time taken: 0.544 seconds
hive> insert into department values(6,"ISE"),(1,"CSE"),(2,"ECE"),(5,"EEE"),(3,"AE"),(4,"ME");
Query ID = hdoop_20210703085517_2da6bcf8-1ad9-4f45-b834-6fe8cc690592
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_1625326304682_0007, Tracking URL = http://ubuntu:8088/proxy/application_1625326304682_0007/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1625326304682_0007
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-03 08:55:24,308 Stage-1 map = 0%, reduce = 0%
2021-07-03 08:55:30,595 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 7.03 sec
2021-07-03 08:55:35,727 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 8.51 sec
MapReduce Total cumulative CPU time: 8 seconds 510 msec
Ended Job = job_1625326304682_0007
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://127.0.0.1:9000/user/hive/warehouse/employee.db/department/.hive-staging_hive_2021-07-03_08-55-17_267_5975454
517276939290-1/-ext-10000
Loading data to table employee.db.department
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 8.51 sec HDFS Read: 15866 HDFS Write: 342 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 510 msec
OK
Time taken: 20.311 seconds
hive>

```

- 9) Performing outer join over Dname.

```

hadoop@ubuntu: ~/apache-hive-3.1.2-bin
File Edit View Search Terminal Help
Aayesha 5020 ISE 6
Aishwarya 5021 ME 4
Aniket 5005 ME 4
Time taken: 93.298 seconds, Fetched: 27 row(s)
hive> select name,ssn,d.deptname,dno from emp e full outer join department d on e.deptname=d.deptname;
Query ID = hdoop_20210703090948_f15491bd-c455-463c-8ced-4b370c5d86cb
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>
Starting Job = job_1625326304682_0010, Tracking URL = http://ubuntu:8088/proxy/application_1625326304682_0010/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1625326304682_0010
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2021-07-03 09:09:57,071 Stage-1 map = 0%, reduce = 0%
2021-07-03 09:10:52,913 Stage-1 map = 50%, reduce = 0%, Cumulative CPU 125.87 sec
2021-07-03 09:11:09,193 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 166.57 sec
2021-07-03 09:11:17,724 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 169.24 sec
MapReduce Total cumulative CPU time: 2 minutes 49 seconds 240 msec
Ended Job = job_1625326304682_0010
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2 Reduce: 1 Cumulative CPU: 169.24 sec HDFS Read: 18268 HDFS Write: 883 SUCCESS
Total MapReduce CPU Time Spent: 2 minutes 49 seconds 240 msec
OK
Guru 5024 AE 3
Akash 5025 CSE 1
Sanket 5013 CSE 1
Ajay 5022 CSE 1
Raghav 5003 CSE 1
Sujan 5012 CSE 1
Sandeep 5016 CSE 1

```



```
hadoop@ubuntu: ~/apache-hive-3.1.2-bin
File Edit View Search Terminal Help
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2 Reduce: 1 Cumulative CPU: 169.24 sec HDFS Read: 18268 HDFS Write: 883 SUCCESS
Total MapReduce CPU Time Spent: 2 minutes 49 seconds 240 msec
OK
Guru 5024 AE 3
Akash 5025 CSE 1
Sanket 5013 CSE 1
Ajay 5022 CSE 1
Raghav 5003 CSE 1
Sujan 5012 CSE 1
Sandeep 5016 CSE 1
Gagan 5023 ECE 2
Dharni 5004 ECE 2
Harshith 5008 ECE 2
NULL NULL EEE 5
Anjali 5019 ISE 6
Navami 5018 ISE 6
Alok 5017 ISE 6
Abhay 5015 ISE 6
Bharat 5014 ISE 6
Ashok 5011 ISE 6
Aditya 5010 ISE 6
Divine 5009 ISE 6
Asha 5007 ISE 6
Darshan 5006 ISE 6
Nandan 5002 ISE 6
Adarsh 5001 ISE 6
Harsha 5000 ISE 6
Aaysha 5020 ISE 6
Aishwarya 5021 ME 4
Aniket 5005 ME 4
Time taken: 90.669 seconds, Fetched: 27 row(s)
hive> select name,ssn,d.deptname,dno from emp e left outer join department d on e.deptname=d.deptname;
Query ID = hdoop_20210703091523_1917b41e-438b-4837-805e-567ff1197abe
Total jobs = 1
```

10) Performing on left outer join over Dname

```
hadoop@ubuntu: ~/apache-hive-3.1.2-bin
File Edit View Search Terminal Help
Aaysha 5020 ISE 6
Aishwarya 5021 ME 4
Aniket 5005 ME 4
Time taken: 90.669 seconds, Fetched: 27 row(s)
hive> select name,ssn,d.deptname,dno from emp e left outer join department d on e.deptname=d.deptname;
Query ID = hdoop_20210703091523_1917b41e-438b-4837-805e-567ff1197abe
Total jobs = 1
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1625326304682_0011, Tracking URL = http://ubuntu:8088/proxy/application_1625326304682_0011/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1625326304682_0011
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2021-07-03 09:15:41,893 Stage-3 map = 0%, reduce = 0%
2021-07-03 09:15:45,997 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 2.12 sec
MapReduce Total cumulative CPU time: 2 seconds 120 msec
Ended Job = job_1625326304682_0011
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 2.12 sec HDFS Read: 10624 HDFS Write: 859 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 120 msec
OK
Aaysha 5020 ISE 6
Aishwarya 5021 ME 4
Ajay 5022 CSE 1
Gagan 5023 ECE 2
Guru 5024 AE 3
Akash 5025 CSE 1
Harsha 5000 ISE 6
Adarsh 5001 ISE 6
Nandan 5002 ISE 6
Raghav 5003 CSE 1
Dharni 5004 ECE 2
Aniket 5005 ME 4
Darshan 5006 ISE 6
```

```
hadoop@ubuntu: ~/apache-hive-3.1.2-bin
File Edit View Search Terminal Help
MapReduce Total cumulative CPU time: 2 seconds 120 msec
Ended Job = job_1625326304682_0011
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 2.12 sec HDFS Read: 10624 HDFS Write: 859 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 120 msec
OK
Aayesha 5020 ISE 6
Aishwarya 5021 ME 4
Ajay 5022 CSE 1
Gagan 5023 ECE 2
Guru 5024 AE 3
Akash 5025 CSE 1
Harsha 5000 ISE 6
Adarsh 5001 ISE 6
Nandan 5002 ISE 6
Raghav 5003 CSE 1
Dharni 5004 ECE 2
Aniket 5005 ME 4
Darshan 5006 ISE 6
Asha 5007 ISE 6
Harshith 5008 ECE 2
Divine 5009 ISE 6
Aditya 5010 ISE 6
Ashok 5011 ISE 6
Sujan 5012 CSE 1
Sanket 5013 CSE 1
Bharat 5014 ISE 6
Abhay 5015 ISE 6
Sandeep 5016 CSE 1
Alok 5017 ISE 6
Navami 5018 ISE 6
Anjali 5019 ISE 6
Time taken: 23.698 seconds, Fetched: 26 row(s)
hive> select name,ssn,e.deptname,dno from emp e left outer join department d on e.deptname=d.deptname;
Query ID = hadoop_20210703091651_81fa7de7-81f8-443c-82f9-61f39cea4596
```

11) Performing on right outer join over Dname.

```
hadoop@ubuntu: ~/apache-hive-3.1.2-bin
File Edit View Search Terminal Help
Aditya 5010 ISE 6
Ashok 5011 ISE 6
Sujan 5012 CSE 1
Sanket 5013 CSE 1
Bharat 5014 ISE 6
Abhay 5015 ISE 6
Sandeep 5016 CSE 1
Alok 5017 ISE 6
Navami 5018 ISE 6
Anjali 5019 ISE 6
Time taken: 18.687 seconds, Fetched: 26 row(s)
hive> select name,ssn,d.deptname,dno from emp e right outer join department d on e.deptname=d.deptname;
Query ID = hadoop_20210703091746_bddd2031-e2a2-47ad-a39b-dfd7ae18be39
Total jobs = 1
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1625326304682_0013, Tracking URL = http://ubuntu:8088/proxy/application_1625326304682_0013/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1625326304682_0013
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2021-07-03 09:18:00,763 Stage-3 map = 0%, reduce = 0%
2021-07-03 09:18:04,861 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 1.92 sec
MapReduce Total cumulative CPU time: 1 seconds 920 msec
Ended Job = job_1625326304682_0013
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 1.92 sec HDFS Read: 9150 HDFS Write: 883 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 920 msec
OK
Aayesha 5020 ISE 6
Harsha 5000 ISE 6
Adarsh 5001 ISE 6
Nandan 5002 ISE 6
Darshan 5006 ISE 6
Asha 5007 ISE 6
```

```
hadoop@ubuntu: ~/apache-hive-3.1.2-bin
File Edit View Search Terminal Help
Ended Job = job_1625326304682_0013
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 1.92 sec HDFS Read: 9150 HDFS Write: 883 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 920 msec
OK
#ayesha 5626 ISE 6
larsha 5066 ISE 6
xdar sh seel ISE 6
Jandan 5662 ISE 6
>arshan 5666 ISE 6
Asha 5667 ISE 6
Divine 5669 ISE 6
dttya 5616 ISE 6
Ashok 5611 ISE 6
Bharat se14 ISE 6
#bhay sels ISE 6
stok 5617 ISE 6
Havant 5618 ISE 6
snjatt 5819 ISE 6
^1.y 5622 CSE 1
Akash 5e25 CSE 1
Taghav 5663 CSE 1
Susan 5612 CSE 1
»anket 5613 CSE 1
Sandeep 5616 CSE 1
Gagan 5623 ECE 2
>harnt 5664 ECE 2
ñarshth 5668 ECE 2
FULL NULL EEE S
Guru 5624 AE 3
stshxarya 5621 hE 4
#ntket sees HE 4
Time taken: 20.427 seconds, Fetched: 27 row(s)
hive> select name,ssn,e.deptname,dno from emp e right outer join department d on e.deptname=d.deptname;
Query ID = hadoop 20210703091843 33dc1fc4-5c8b-4672-9d0f-0608900bf4d5
```