



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,

Chavi Singh

1NT18IS043

Github link:

Hadoop Map-reduce Programming:

Hadoop MapReduce is a software framework for easily writing applications which process vast amounts of data (multi-terabyte data-sets) in-parallel on large clusters (thousands of nodes) of commodity hardware in a reliable, fault-tolerant manner.

A MapReduce *job* usually splits the input data-set into independent chunks which are processed by the *map tasks* in a completely parallel manner. The framework sorts the outputs of the maps, which are then input to the *reduce tasks*. Typically both the input and the output of the job are stored in a file-system. The framework takes care of scheduling tasks, monitoring them and re-executes the failed tasks.

Typically the compute nodes and the storage nodes are the same, that is, the MapReduce framework and the Hadoop Distributed File System are running on the same set of nodes. This configuration allows the framework to effectively schedule tasks on the nodes where data is already present, resulting in very high aggregate bandwidth across the cluster.

The MapReduce framework consists of a single master JobTracker and one slave TaskTracker per cluster-node. The master is responsible for scheduling the jobs' component tasks on the slaves, monitoring them and re-executing the failed tasks. The slaves execute the tasks as directed by the master.

Minimally, applications specify the input/output locations and supply *map* and *reduce* functions via implementations of appropriate interfaces and/or abstract-classes. These, and other job parameters, comprise the *job configuration*. The Hadoop *job client* then submits the job (jar/executable etc.) and configuration to the JobTracker which then assumes the responsibility of distributing the software/configuration to the slaves, scheduling tasks and monitoring them, providing status and diagnostic information to the job-client.

Hive:

Hive is a data warehouse infrastructure tool to process structured data in Hadoop. It resides on top of Hadoop to summarize Big Data, and makes querying and analyzing easy.

Initially Hive was developed by Facebook, later the Apache Software Foundation took it up and developed it further as an open source under the name Apache Hive. It is used by different companies. For example, Amazon uses it in Amazon Elastic MapReduce.

Hive is not:

- A relational database
- A design for OnLine Transaction Processing (OLTP)
- A language for real-time queries and row-level updates

Features of Hive:

- It stores schema in a database and processed data into HDFS.
- It is designed for OLAP.
- It provides SQL type language for querying called HiveQL or HQL.
- It is familiar, fast, scalable, and extensible

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.

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 work in ISE department
 2. Total number of employees with experience=5 years
 3. Count the number of employees who lives in Bangalore

Dataset Snapshot:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
9	Anjali	5231	65000	Bangalore	ISE	5												
10	Namitha	1177	54000	Pune	ECE	4												
11	Ajay	3662	73000	Chennai	EEE	6												
12	Akash	9815	20000	Bangalore	ISE	5												
13	Darshan	8000	45000	Kolkata	CSE	8												
14	Anish	3712	35000	Patna	AE	2												
15	Nandan	8868	15000	Bhopal	TC	3												
16	Purnima	6734	67000	Bangalore	ISE	5												
17	Sharada	4599	34000	Indore	CSE	2												
18	Pratik	2001	75000	Bangalore	ISE	5												
19	Pooja	6021	28000	Shimla	EEE	4												
20	Vikas	4378	32000	Agra	CSE	2												
21	Keerthi	1008	65000	Bangalore	ISE	5												
22	Sahana	4289	26000	Mangalore	EEE	4												
23	Deepak	6710	77000	Coimbatore	ECE	6												
24	Ujwal	3011	37000	Kochi	CSE	4												
25	Nithin	7888	32000	Goa	EEE	3												
26	Karthik	1671	80000	Bangalore	ISE	5												
27																		
28																		

JAVA program screenshots :-

1. Employee Project -> EmpISE class

```

import java.io.IOException;
import java.util.*;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapred.*;

public class EmpISE {
    //MAPPER CODE

    public static class Map extends MapReduceBase implements Mapper<LongWritable, Text, Text, IntWritable> {
        private final static IntWritable one = new IntWritable(1);
        private Text word = new Text();

        public void map(LongWritable key, Text value, OutputCollector<Text, IntWritable> output, Reporter reporter) throws IOException {
    }
}

```

```
String data = value.toString();
String[] Ecount= data.split(",");
if (Ecount[4].equals("ISE")) {
    output.collect(new Text("Total no.of employees working in ISE Department : "), one);
}
```

```
}
```

//REDUCER CODE

```
public static class Reduce extends MapReduceBase implements Reducer<Text, IntWritable, Text,
IntWritable> {
    public void reduce(Text key, Iterator<IntWritable> values, OutputCollector<Text, IntWritable>
output, Reporter reporter) throws IOException { //{{little: {1,1}}
        int val = 0 ;
        while(values.hasNext()) {
            val += values.next().get();
        }
        output.collect(key, new IntWritable(val));
    }
}
```

//DRIVER CODE

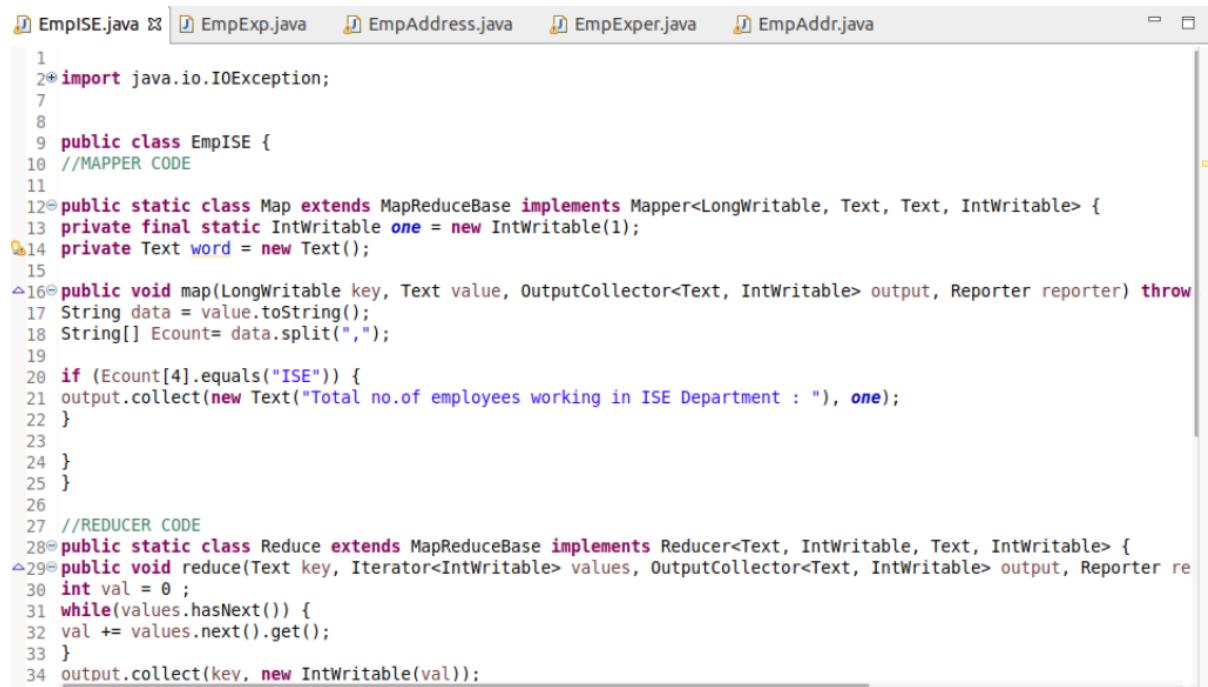
```
public static void main(String[] args) throws Exception {
    JobConf conf = new JobConf(EmpISE.class);
    conf.setJobName("Total no.of employees working in ISE Department");
    conf.setOutputKeyClass(Text.class);
    conf.setOutputValueClass(IntWritable.class);
    conf.setMapperClass(Map.class);
    conf.setCombinerClass(Reduce.class);
    conf.setReducerClass(Reduce.class);
```

```

        conf.setInputFormat(TextInputFormat.class);
        conf.setOutputFormat(TextOutputFormat.class);
        FileInputFormat.setInputPaths(conf, new Path(args[0]));
        FileOutputFormat.setOutputPath(conf, new Path(args[1]));
        JobClient.runJob(conf);
    }
}

```

Program:



```

1  import java.io.IOException;
2
3
4  public class EmpISE {
5      //MAPPER CODE
6
7      public static class Map extends MapReduceBase implements Mapper<LongWritable, Text, Text, IntWritable> {
8          private final static IntWritable one = new IntWritable(1);
9          private Text word = new Text();
10
11         public void map(LongWritable key, Text value, OutputCollector<Text, IntWritable> output, Reporter reporter) throws IOException {
12             String data = value.toString();
13             String[] Ecount= data.split(",");
14             if (Ecount[4].equals("ISE")) {
15                 output.collect(new Text("Total no.of employees working in ISE Department : "), one);
16             }
17         }
18
19         public void reduce(Text key, Iterator<IntWritable> values, OutputCollector<Text, IntWritable> output, Reporter reporter) throws IOException {
20             int val = 0 ;
21             while(values.hasNext()) {
22                 val += values.next().get();
23             }
24             output.collect(key, new IntWritable(val));
25         }
26     } //REDUCER CODE
27     public static class Reduce extends MapReduceBase implements Reducer<Text, IntWritable, Text, IntWritable> {
28         public void reduce(Text key, Iterator<IntWritable> values, OutputCollector<Text, IntWritable> output, Reporter reporter) throws IOException {
29             int val = 0 ;
30             while(values.hasNext()) {
31                 val += values.next().get();
32             }
33             output.collect(key, new IntWritable(val));
34         }
35     }
36 }

```

```
24 }
25 }
26
27 //REDUCER CODE
28 public static class Reduce extends MapReduceBase implements Reducer<Text, IntWritable, Text, IntWritable> {
29     public void reduce(Text key, Iterator<IntWritable> values, OutputCollector<Text, IntWritable> output, Reporter reporter) throws IOException {
30         int val = 0;
31         while(values.hasNext()) {
32             val += values.next().get();
33         }
34         output.collect(key, new IntWritable(val));
35     }
36 }
37
38 //DRIVER CODE
39 public static void main(String[] args) throws Exception {
40     JobConf conf = new JobConf(EmpISE.class);
41     conf.setJobName("Total no.of employees working in ISE Department");
42     conf.setOutputKeyClass(Text.class);
43     conf.setOutputValueClass(IntWritable.class);
44     conf.setMapperClass(Map.class);
45     conf.setCombinerClass(Reduce.class);
46     conf.setReducerClass(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 }
```

Number of employees working in ISE: 8

Execution:

hadoop@ubuntu:~\$ cd hadoop-3.2.1/sbin

hadoop@ubuntu:~/hadoop-3.2.1/sbin\$./start-all.sh

WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.

WARNING: This is not a recommended production deployment configuration.

WARNING: Use CTRL-C to abort.

Starting namenodes on [localhost]

Starting datanodes

Starting secondary namenodes [ubuntu]

Starting resourcemanager

Starting nodemanagers

hadoop@ubuntu:~/hadoop-3.2.1/sbin\$ jps

5202 SecondaryNameNode

6612 Jps

5700 NodeManager

5525 ResourceManager

4982 DataNode

3386 org.eclipse.equinox.launcher_1.5.600.v20191014-2022.jar

4831 NameNode

hadoop@ubuntu:~/hadoop-3.2.1/sbin\$ cd ~

hadoop@ubuntu:~\$ cd Desktop

hadoop@ubuntu:~/Desktop\$ ls

Banking Banking.zip dataset.csv

Banking2 EmpISE.jar sales.csv

Banking2.zip IIGNISER1 sales_withoutHeader.csv

BankingTransaction2.jar IIGNISER1.zip WR.jar

BankingTransaction.jar input1.txt

hadoop@ubuntu:~/Desktop\$ hadoop fs -copyFromLocal dataset.csv

2021-07-04 09:03:22,579 INFO sasl.SaslDataTransferClient: SASL encryption trust check:
localHostTrusted = false, remoteHostTrusted = false

hadoop@ubuntu:~/Desktop\$ hadoop fs -ls

Found 13 items

-rw-r--r-- 1 hadoop supergroup 638 2021-07-04 09:03 dataset.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\$ hadoop jar EmpISE.jar EmpISE.EmpISE dataset.csv EmpISE.txt

2021-07-04 09:11:06,280 INFO client.RMProxy: Connecting to ResourceManager at
/127.0.0.1:8032

2021-07-04 09:11:18,054 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032

2021-07-04 09:11:20,020 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 09:11:20,616 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1625413465806_0001

2021-07-04 09:11:21,316 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false

2021-07-04 09:11:22,438 INFO mapred.FileInputFormat: Total input files to process : 1

2021-07-04 09:11:23,361 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false

2021-07-04 09:11:23,545 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false

2021-07-04 09:11:23,566 INFO mapreduce.JobSubmitter: number of splits:2

2021-07-04 09:11:26,831 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false

2021-07-04 09:11:28,217 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1625413465806_0001

2021-07-04 09:11:28,218 INFO mapreduce.JobSubmitter: Executing with tokens: []

2021-07-04 09:11:32,114 INFO conf.Configuration: resource-types.xml not found

2021-07-04 09:11:32,114 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.

2021-07-04 09:11:38,975 INFO impl.YarnClientImpl: Submitted application application_1625413465806_0001

2021-07-04 09:11:40,184 INFO mapreduce.Job: The url to track the job: http://ubuntu:8088/proxy/application_1625413465806_0001/

2021-07-04 09:11:40,361 INFO mapreduce.Job: Running job: job_1625413465806_0001

2021-07-04 09:13:46,673 INFO mapreduce.Job: Job job_1625413465806_0001 running in uber mode : false

2021-07-04 09:13:46,674 INFO mapreduce.Job: map 0% reduce 0%

2021-07-04 09:18:31,564 INFO mapreduce.Job: map 33% reduce 0%

2021-07-04 09:18:32,587 INFO mapreduce.Job: map 67% reduce 0%

2021-07-04 09:18:34,657 INFO mapreduce.Job: map 100% reduce 0%

2021-07-04 09:19:23,241 INFO mapreduce.Job: map 100% reduce 100%

2021-07-04 09:19:25,278 INFO mapreduce.Job: Job job_1625413465806_0001 completed successfully

2021-07-04 09:19:26,524 INFO mapreduce.Job: Counters: 54

File System Counters

FILE: Number of bytes read=120

FILE: Number of bytes written=677717

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=54

HDFS: Number of read operations=11

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=2

Launched reduce tasks=1

Data-local map tasks=2

Total time spent by all maps in occupied slots (ms)=568879

Total time spent by all reduces in occupied slots (ms)=45707

Total time spent by all map tasks (ms)=568879

Total time spent by all reduce tasks (ms)=45707

Total vcore-milliseconds taken by all map tasks=568879

Total vcore-milliseconds taken by all reduce tasks=45707

Total megabyte-milliseconds taken by all map tasks=582532096

Total megabyte-milliseconds taken by all reduce tasks=46803968

Map-Reduce Framework

Map input records=20

Map output records=13
Map output bytes=715
Map output materialized bytes=126
Input split bytes=184
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 : 8

2.EmpExperience Project ->EmpExper class

```
import java.io.IOException;
import java.util.*;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapred.*;

public class EmpExper {
    //Mapper Code
    public static class Map extends MapReduceBase implements Mapper<LongWritable, Text,
    Text, IntWritable> {
        private final static IntWritable one = new IntWritable(1);
```

```

private Text word = new Text();

public void map(LongWritable key, Text value, OutputCollector<Text, IntWritable> output,
Reporter reporter) throws IOException {
    String data = value.toString();
    String[] Ecount= data.split(",");
    if (Ecount[5].equals("5")) {
        output.collect(new Text("Total no.of employees having 5 years of experience : "), one);
    }
}

//REDUCER CODE
public static class Reduce extends MapReduceBase implements Reducer<Text, IntWritable,
Text, IntWritable> {
    public void reduce(Text key, Iterator<IntWritable> values, OutputCollector<Text,
IntWritable> output, Reporter reporter) throws IOException { //{{little: {1,1}}
        int val = 0 ;
        while(values.hasNext()) {
            val += values.next().get();
        }
        output.collect(key, new IntWritable(val));
    }
}

//DRIVER CODE
public static void main(String[] args) throws Exception {
    JobConf conf = new JobConf(EmpExper.class);
    conf.setJobName("Total no.of employees having 5 years of experience");
    conf.setOutputKeyClass(Text.class);
}

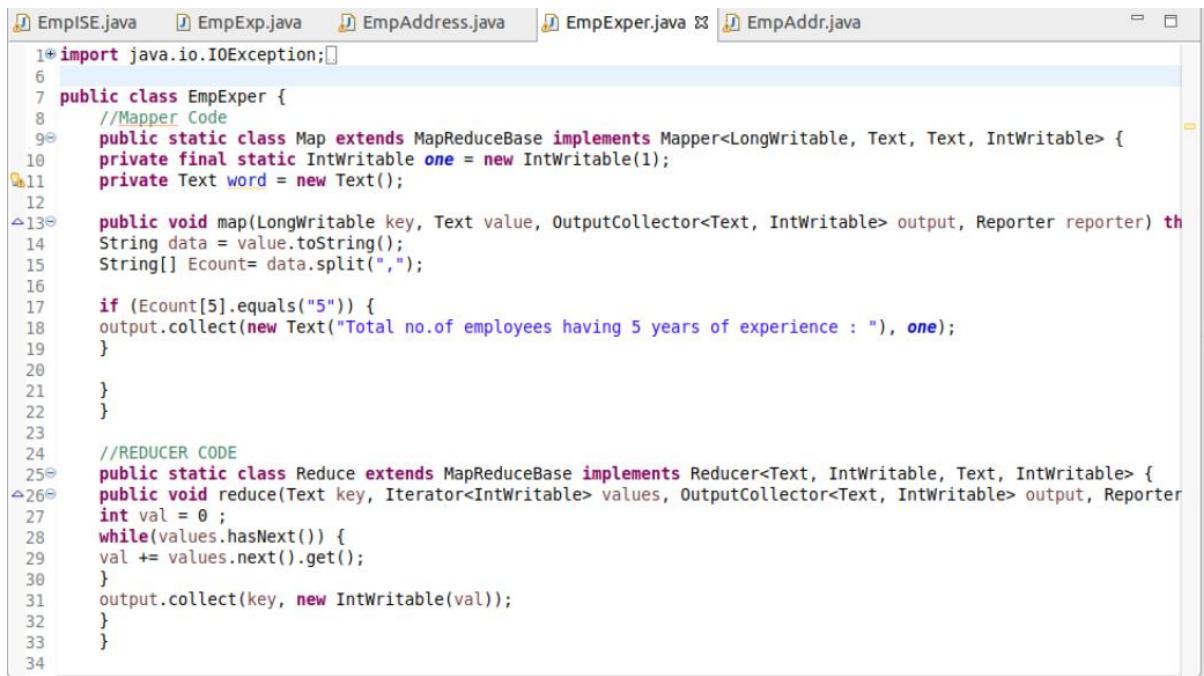
```

```

conf.setOutputValueClass(IntWritable.class);
conf.setMapperClass(Map.class);
conf.setCombinerClass(Reduce.class);
conf.setReducerClass(Reduce.class);
conf.setInputFormat(TextInputFormat.class);
conf.setOutputFormat(TextOutputFormat.class);
FileInputFormat.setInputPaths(conf, new Path(args[0]));
FileOutputFormat.setOutputPath(conf, new Path(args[1]));
JobClient.runJob(conf);
}
}

```

Program:



```

EmpISE.java  EmpExp.java  EmpAddress.java  EmpExper.java  EmpAddr.java
1* import java.io.IOException;
6
7 public class EmpExper {
8     //Mapper Code
9     public static class Map extends MapReduceBase implements Mapper<LongWritable, Text, Text, IntWritable> {
10        private final static IntWritable one = new IntWritable(1);
11        private Text word = new Text();
12
13        public void map(LongWritable key, Text value, OutputCollector<Text, IntWritable> output, Reporter reporter) th
14            String data = value.toString();
15            String[] Ecount= data.split(",");
16
17            if (Ecount[5].equals("5")) {
18                output.collect(new Text("Total no.of employees having 5 years of experience : "), one);
19            }
20        }
21    }
22
23
24    //REDUCER CODE
25    public static class Reduce extends MapReduceBase implements Reducer<Text, IntWritable, Text, IntWritable> {
26        public void reduce(Text key, Iterator<IntWritable> values, OutputCollector<Text, IntWritable> output, Reporter
27            int val = 0 ;
28            while(values.hasNext()) {
29                val += values.next().get();
30            }
31            output.collect(key, new IntWritable(val));
32        }
33    }
34

```

```

21     }
22 }
23
24 //REDUCER CODE
25 public static class Reduce extends MapReduceBase implements Reducer<Text, IntWritable, Text, IntWritable> {
26     public void reduce(Text key, Iterator<IntWritable> values, OutputCollector<Text, IntWritable> output, Reporter reporter) throws IOException {
27         int val = 0 ;
28         while(values.hasNext()) {
29             val += values.next().get();
30         }
31         output.collect(key, new IntWritable(val));
32     }
33 }
34
35 //DRIVER CODE
36 public static void main(String[] args) throws Exception {
37     JobConf conf = new JobConf(EmpExper.class);
38     conf.setJobName("Total no.of employees having 5 years of experience");
39     conf.setOutputKeyClass(Text.class);
40     conf.setOutputValueClass(IntWritable.class);
41     conf.setMapperClass(Map.class);
42     conf.setCombinerClass(Reduce.class);
43     conf.setReducerClass(Reduce.class);
44     conf.setInputFormat(TextInputFormat.class);
45     conf.setOutputFormat(TextOutputFormat.class);
46     TextInputFormat.setInputPaths(conf, new Path(args[0]));
47     TextOutputFormat.setOutputPath(conf, new Path(args[1]));
48     JobClient.runJob(conf);
49 }
50

```

Output: Total number of employees with experience=5 years : 8

Execution:

hadoop@ubuntu:~/Desktop\$ hadoop jar EmpExp.jar EmpExp.EmpExp dataset.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_1625413465806_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

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

HDFS: Number of bytes read erasure-coded=0

Job Counters

Killed map tasks=1

Launched map tasks=2

Launched reduce tasks=1

Data-local map tasks=2

Total time spent by all maps in occupied slots (ms)=107735

Total time spent by all reduces in occupied slots (ms)=2063

Total time spent by all map tasks (ms)=107735

Total time spent by all reduce tasks (ms)=2063

Total vcore-milliseconds taken by all map tasks=107735

Total vcore-milliseconds taken by all reduce tasks=2063

Total megabyte-milliseconds taken by all map tasks=110320640

Total megabyte-milliseconds taken by all reduce tasks=2112512

Map-Reduce Framework

Map input records=20

Map output records=10

Map output bytes=580

Map output materialized bytes=132

Input split bytes=184

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 : 8

3.EmpAddressBangalore Project -> EmpAddr

```
import java.io.IOException;
import java.util.*;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.conf.*;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapred.*;
import org.apache.hadoop.util.*;

public class EmpAddr {
    //Mapper Code
    public static class Map extends MapReduceBase implements Mapper<LongWritable,
    Text, Text, IntWritable> {
        private final static IntWritable one = new IntWritable(1);
        private Text word = new Text();

        public void map(LongWritable key, Text value, OutputCollector<Text,
        IntWritable> output, Reporter reporter) throws IOException {
            String data = value.toString();
            String[] Ecount= data.split(",");
            if (Ecount[3].equals("Bangalore")) {
                output.collect(new Text("Total no.of employees who stays
in Bangalore : "), one);
            }
        }
    }
}
```

```

//REDUCER CODE

public static class Reduce extends MapReduceBase implements
Reducer<Text, IntWritable, Text, IntWritable> {

    public void reduce(Text key, Iterator<IntWritable> values,
OutputCollector<Text, IntWritable> output, Reporter reporter) throws IOException { //{{little: {1,1}}
        int val = 0 ;
        while(values.hasNext()) {
            val += values.next().get();
        }
        output.collect(key, new IntWritable(val));
    }
}

//DRIVER CODE

public static void main(String[] args) throws Exception {
    JobConf conf = new JobConf(EmpAddr.class);
    conf.setJobName("Total no.of employees having 5 years of
experience");
    conf.setOutputKeyClass(Text.class);
    conf.setOutputValueClass(IntWritable.class);
    conf.setMapperClass(Map.class);
    conf.setCombinerClass(Reduce.class);
    conf.setReducerClass(Reduce.class);
    conf.setInputFormat(TextInputFormat.class);
    conf.setOutputFormat(TextOutputFormat.class);
    FileInputFormat.setInputPaths(conf, new Path(args[0]));
    FileOutputFormat.setOutputPath(conf, new Path(args[1]));
    JobClient.runJob(conf);
}
}

```

Program:

```
1 import java.io.IOException;
2 import java.util.*;
3 import org.apache.hadoop.fs.Path;
4 import org.apache.hadoop.conf.*;
5 import org.apache.hadoop.io.*;
6 import org.apache.hadoop.mapred.*;
7 import org.apache.hadoop.util.*;
8
9 public class EmpAddr {
10     //Mapper Code
11     public static class Map extends MapReduceBase implements Mapper<LongWritable, Text, Text, IntWritable> {
12         private final static IntWritable one = new IntWritable(1);
13         private Text word = new Text();
14
15         public void map(LongWritable key, Text value, OutputCollector<Text, IntWritable> output, Reporter reporter) {
16             String data = value.toString();
17             String[] Ecount= data.split(",");
18
19             if (Ecount[3].equals("Bangalore")) {
20                 output.collect(new Text("Total no.of employees who stays in Bangalore : "), one);
21             }
22         }
23     }
24
25     //REDUCER CODE
26     public static class Reduce extends MapReduceBase implements Reducer<Text, IntWritable, Text, IntWritable> {
27         public void reduce(Text key, Iterator<IntWritable> values, OutputCollector<Text, IntWritable> output,
28             int val = 0 ;
29             while(values.hasNext()) {
30
31
32
33
34
35
36
37         //DRIVER CODE
38         public static void main(String[] args) throws Exception {
39             JobConf conf = new JobConf(EmpAddr.class);
40             conf.setJobName("Total no. of employees having 5 years of experience");
41             conf.setOutputKeyClass(Text.class);
42             conf.setOutputValueClass(IntWritable.class);
43             conf.setMapperClass(Map.class);
44             conf.setCombinerClass(Reduce.class);
45             conf.setReducerClass(Reduce.class);
46             conf.setInputFormat(TextInputFormat.class);
47             conf.setOutputFormat(TextOutputFormat.class);
48             FileInputFormat.setInputPaths(conf, new Path(args[0]));
49             FileOutputFormat.setOutputPath(conf, new Path(args[1]));
50             JobClient.runJob(conf);
51         }
52     }
53 }
```

```
25
26     //REDUCER CODE
27     public static class Reduce extends MapReduceBase implements Reducer<Text, IntWritable, Text, IntWritable> {
28         public void reduce(Text key, Iterator<IntWritable> values, OutputCollector<Text, IntWritable> output,
29             int val = 0 ;
30             while(values.hasNext()) {
31                 val += values.next().get();
32             }
33             output.collect(key, new IntWritable(val));
34         }
35     }
36
37     //DRIVER CODE
38     public static void main(String[] args) throws Exception {
39         JobConf conf = new JobConf(EmpAddr.class);
40         conf.setJobName("Total no. of employees having 5 years of experience");
41         conf.setOutputKeyClass(Text.class);
42         conf.setOutputValueClass(IntWritable.class);
43         conf.setMapperClass(Map.class);
44         conf.setCombinerClass(Reduce.class);
45         conf.setReducerClass(Reduce.class);
46         conf.setInputFormat(TextInputFormat.class);
47         conf.setOutputFormat(TextOutputFormat.class);
48         FileInputFormat.setInputPaths(conf, new Path(args[0]));
49         FileOutputFormat.setOutputPath(conf, new Path(args[1]));
50         JobClient.runJob(conf);
51     }
52 }
53 }
```

Output: Count the number of employees who lives in Bangalore: 7

Execution:

```
hadoop@ubuntu:~/Desktop$ hadoop jar EmpAddress.jar EmpAddress.EmpAddress
dataset.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_1625413465806_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
HDFS: Number of read operations=11
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=2
Launched reduce tasks=1
Data-local map tasks=2
Total time spent by all maps in occupied slots (ms)=4934
Total time spent by all reduces in occupied slots (ms)=1767
Total time spent by all map tasks (ms)=4934
Total time spent by all reduce tasks (ms)=1767
Total vcore-milliseconds taken by all map tasks=4934
Total vcore-milliseconds taken by all reduce tasks=1767
Total megabyte-milliseconds taken by all map tasks=5052416
Total megabyte-milliseconds taken by all reduce tasks=1809408

Map-Reduce Framework

Map input records=20
Map output records=4

Map output bytes=208
Map output materialized bytes=120
Input split bytes=184
Combine input records=4
Combine output records=2
Reduce input groups=1
Reduce shuffle bytes=120
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)=171
CPU time spent (ms)=1660
Physical memory (bytes) snapshot=729804800
Virtual memory (bytes) snapshot=7796326400
Total committed heap usage (bytes)=606601216
Peak Map Physical memory (bytes)=276217856
Peak Map Virtual memory (bytes)=2597658624
Peak Reduce Physical memory (bytes)=181227520
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 hdoop supergroup 0 2021-07-04 10:00 EmpAddress.txt/_SUCCESS
```

```
-rw-r--r-- 1 hdoop 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 : 7

HADOOP MAP REDUCE SCREENSHOTS :

The screenshot shows a terminal window with three tabs:

- Tab 1: hdoop@hadoop-VirtualBox: ~/hadoop-3.2.1/sbin
- Tab 2: hdoop@hadoop-VirtualBox: ~/hadoop... (active)
- Tab 3: hdoop@hadoop-VirtualBox: ~

In the active tab, the user runs the following commands:

```
hdoop@hadoop-VirtualBox:~$ su - hdoop  
Password:  
hdoop@hadoop-VirtualBox:~$ cd hadoop-3.2.1/sbin  
hdoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ ./start-all.sh  
WARNING: Attempting to start all Apache Hadoop daemons as hdoop in 10 seconds.  
WARNING: This is not a recommended production deployment configuration.  
WARNING: Use CTRL-C to abort.  
Starting namenodes on [localhost]  
Starting datanodes  
Starting secondary namenodes [hadoop-VirtualBox]  
Starting resourcemanager  
Starting nodemanagers  
hdoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ jps  
8192 DataNode  
8609 ResourceManager  
8737 NodeManager  
8364 SecondaryNameNode  
8062 NameNode  
8894 Jps  
hdoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ cd ~  
hdoop@hadoop-VirtualBox:~$ cd Desktop  
-bash: cd: Desktop: No such file or directory  
hdoop@hadoop-VirtualBox:~$ cd /home/hadoop/Desktop  
hdoop@hadoop-VirtualBox:/home/hadoop/Desktop$ ls  
dataset.csv  
'Eclipse IDE for Java Developers - 2021-03.desktop'  
EmpExpclass.jar  
EmpExp.jar  
EmpISE.jar
```

```
hadoop@hadoop-VirtualBox: ~/hadoop-3.2.1/sbin$ ./start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hdoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [hadoop-VirtualBox]
Starting resourcemanager
Starting nodemanagers
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ jps
8192 DataNode
8609 ResourceManager
8737 NodeManager
8364 SecondaryNameNode
8062 NameNode
8894 Jps
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ cd ~
hadoop@hadoop-VirtualBox:~$ cd Desktop
-bash: cd: Desktop: No such file or directory
hadoop@hadoop-VirtualBox:~$ cd /home/hadoop/Desktop
hadoop@hadoop-VirtualBox:/home/hadoop/Desktop$ ls
dataset.csv          EmpExpclass.jar
'Eclipse IDE for Java Developers - 2021-03.desktop'  EmpExp.jar
EmpAddressclass.jar  EmpISE.jar
EmpAddress.jar        Employee.jar
```

```
hadoop@hadoop-VirtualBox: ~/hadoop-3.2.1/sbin$ ls: '.': No such file or directory
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ pwd
/home/hadoop/hadoop-3.2.1/sbin
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop fs -mkdir -p /user/hdoop
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop fs -ls /user
Found 2 items
drwxr-xr-x  - hdoop supergroup      0 2021-07-10 12:09 /user/hdoop
drwxr-xr-x  - hdoop supergroup      0 2021-07-05 21:50 /user/hive
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop fs -ls
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop fs -mkdir -p /user/hdoop/faizah
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop fs -ls
Found 1 items
drwxr-xr-x  - hdoop supergroup      0 2021-07-10 12:11 faizah
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop fs -copyFromLocal /home/hadoop/Desktop/dataset.csv
2021-07-10 12:11:47,556 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
```

```
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop jar Employee.jar Employee.E
mpISE dataset.csv EmpISE.txt
JAR does not exist or is not a normal file: /home/hadoop/hadoop-3.2.1/sbin/Employ
ee.jar
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop jar Employee.jar EmpISE dat
aset.csv EmpISE.txt
JAR does not exist or is not a normal file: /home/hadoop/hadoop-3.2.1/sbin/Employ
ee.jar
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop jar Employee.jar EmpISE dat
aset.csv EmpISE.txt
JAR does not exist or is not a normal file: /home/hadoop/hadoop-3.2.1/sbin/Employ
ee.jar
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop jar /home/hadoop/Desktop/E
mployee.jar EmpISE dataset.csv EmpISE.txt
2021-07-10 12:47:43,263 INFO client.RMProxy: Connecting to ResourceManager at /
127.0.0.1:8032
2021-07-10 12:47:43,623 INFO client.RMProxy: Connecting to ResourceManager at /
127.0.0.1:8032
2021-07-10 12:47:43,933 WARN mapreduce.JobResourceUploader: Hadoop command-line
option parsing not performed. Implement the Tool interface and execute your ap
plication with ToolRunner to remedy this.
2021-07-10 12:47:43,992 INFO mapreduce.JobResourceUploader: Disabling Erasure C
oding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1625897933179_0001
2021-07-10 12:47:44,243 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
```

```
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop jar Employee.jar Employee.E
mpISE dataset.csv EmpISE.txt
option parsing not performed. Implement the Tool interface and execute your ap
plication with ToolRunner to remedy this.
2021-07-10 12:47:43,992 INFO mapreduce.JobResourceUploader: Disabling Erasure C
oding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1625897933179_0001
2021-07-10 12:47:44,243 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
2021-07-10 12:47:44,468 INFO mapred.FileInputFormat: Total input files to proce
ss : 1
2021-07-10 12:47:44,553 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
2021-07-10 12:47:44,571 WARN hdfs.DataStreamer: Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1252)
        at java.lang.Thread.join(Thread.java:1326)
        at org.apache.hadoop.hdfs.DataStreamer.closeResponder(DataStreamer.java
:986)
        at org.apache.hadoop.hdfs.DataStreamer.endBlock(DataStreamer.java:640)
        at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:810)
2021-07-10 12:47:44,614 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
2021-07-10 12:47:44,639 INFO mapreduce.JobSubmitter: number of splits:2
2021-07-10 12:47:44,926 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
2021-07-10 12:47:44,998 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_1625897933179_0001
2021-07-10 12:47:44,999 INFO mapreduce.JobSubmitter: Executing with tokens: 11
```

```
hadoop@hadoop-VirtualBox: ~/hadoop-3.2.1/sbin
```

```
check: localHostTrusted = false, remoteHostTrusted = false
2021-07-10 12:47:44,998 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1625897933179_0001
2021-07-10 12:47:44,999 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-07-10 12:47:45,383 INFO conf.Configuration: resource-types.xml not found
2021-07-10 12:47:45,383 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-07-10 12:47:45,786 INFO impl.YarnClientImpl: Submitted application application_1625897933179_0001
2021-07-10 12:47:45,989 INFO mapreduce.Job: The url to track the job: http://hadoop-VirtualBox:8088/proxy/application_1625897933179_0001/
2021-07-10 12:47:45,996 INFO mapreduce.Job: Running job: job_1625897933179_0001
2021-07-10 12:47:58,362 INFO mapreduce.Job: Job job_1625897933179_0001 running in uber mode : false
2021-07-10 12:47:58,363 INFO mapreduce.Job: map 0% reduce 0%
2021-07-10 12:48:09,554 INFO mapreduce.Job: map 100% reduce 0%
2021-07-10 12:48:16,638 INFO mapreduce.Job: map 100% reduce 100%
2021-07-10 12:48:17,664 INFO mapreduce.Job: Job job_1625897933179_0001 completed successfully
2021-07-10 12:48:17,891 INFO mapreduce.Job: Counters: 54
      File System Counters
          FILE: Number of bytes read=120
          FILE: Number of bytes written=677699
          FILE: Number of read operations=0
          FILE: Number of large read operations=0
          FILE: Number of write operations=0
          HDFS: Number of bytes read=1427
      
```

```
hadoop@hadoop-VirtualBox: ~/hadoop-3.2.1/sbin
```

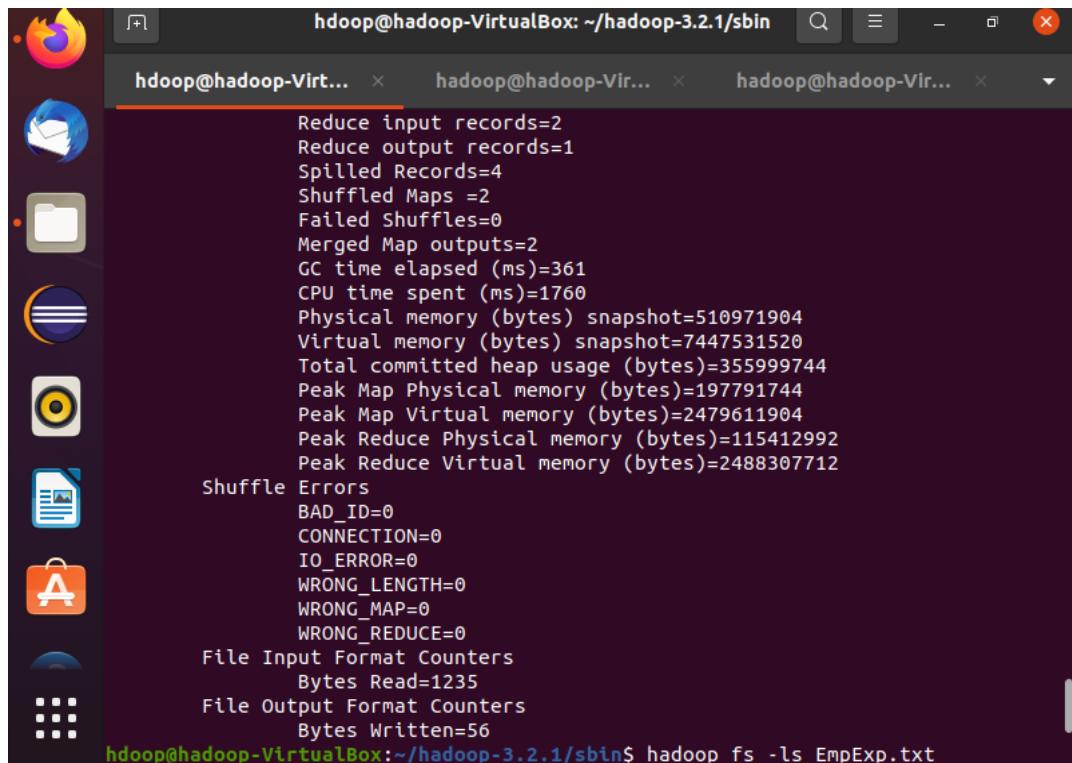
```
      FILE: Number of read operations=0
      FILE: Number of large read operations=0
      FILE: Number of write operations=0
      HDFS: Number of bytes read=1427
      HDFS: Number of bytes written=53
      HDFS: Number of read operations=11
      HDFS: Number of large read operations=0
      HDFS: Number of write operations=2
      HDFS: Number of bytes read erasure-coded=0
      Job Counters
          Launched map tasks=2
          Launched reduce tasks=1
          Data-local map tasks=2
          Total time spent by all maps in occupied slots (ms)=17602
          Total time spent by all reduces in occupied slots (ms)=4771
          Total time spent by all map tasks (ms)=17602
          Total time spent by all reduce tasks (ms)=4771
          Total vcore-milliseconds taken by all map tasks=17602
          Total vcore-milliseconds taken by all reduce tasks=4771
          Total megabyte-milliseconds taken by all map tasks=18024448
          Total megabyte-milliseconds taken by all reduce tasks=4885504
      Map-Reduce Framework
          Map input records=25
          Map output records=8
          Map output bytes=440
          Map output materialized bytes=126
          Input split bytes=192
      
```

```
hadoop@hadoop-VirtualBox: ~/hadoop-3.2.1/sbin$ hadoop fs -cat EmpISE.txt/part-000
00
2021-07-10 12:49:07,491 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
Total no.of employees working in ISE Department :      8
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$
```

```
hadoop@hadoop-VirtualBox: ~/Desktop$ cd /home/hadoop/Desktop
hadoop@hadoop-VirtualBox:~/Desktop$ ls
dataset.csv                                         EmpExperience.jar
'Eclipse IDE for Java Developers - 2021-03.desktop' EmpExper.jar
EmpAddressclass.jar                                EmpExp.jar
EmpAddress.jar                                     EmpISE.jar
EmpExpclass.jar                                    Employee.jar
hadoop@hadoop-VirtualBox:~/Desktop$
```

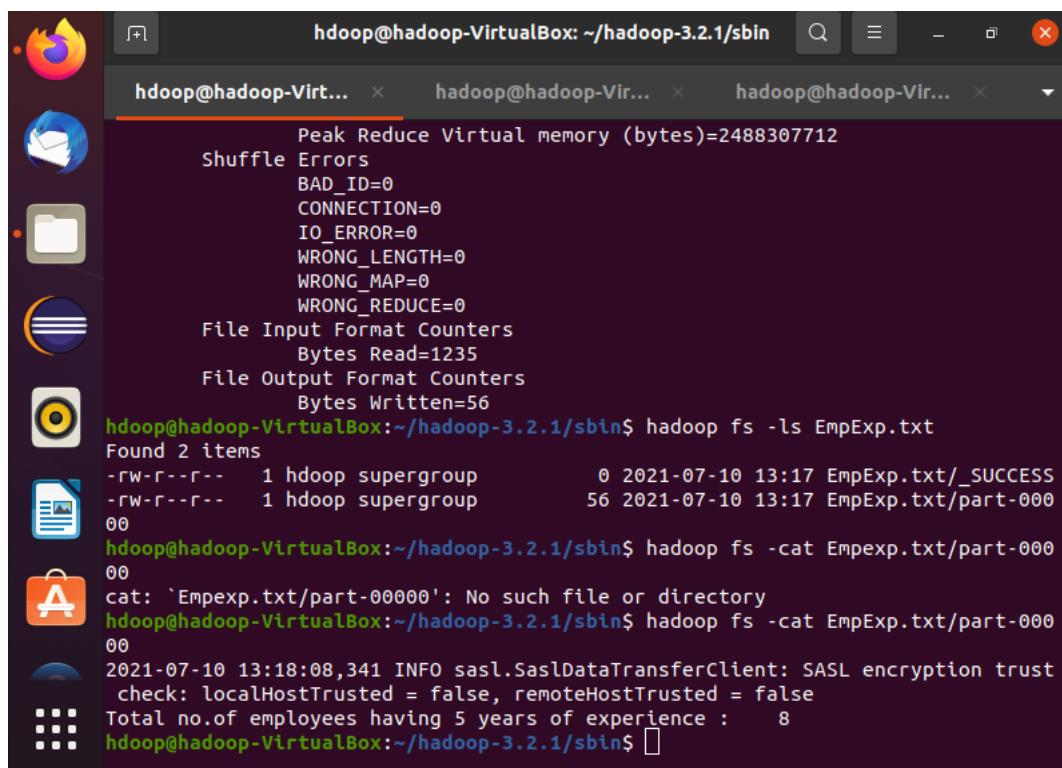
```
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop jar /home/hadoop/Desktop/EmpExperience.jar EmpExper dataset.csv EmpExp.txt
2021-07-10 13:16:38,066 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-10 13:16:38,543 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-10 13:16:38,979 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2021-07-10 13:16:39,055 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1625897933179_0002
2021-07-10 13:16:39,317 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-10 13:16:39,589 INFO mapred.FileInputFormat: Total input files to process : 1
2021-07-10 13:16:39,698 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-10 13:16:39,804 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-10 13:16:39,858 INFO mapreduce.JobSubmitter: number of splits:2
2021-07-10 13:16:40,222 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-10 13:16:40,316 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1625897933179_0002
2021-07-10 13:16:40,317 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-07-10 13:16:40,763 INFO conf.Configuration: resource-types.xml not found
2021-07-10 13:16:40,763 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
```

```
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop jar /home/hadoop/Desktop/EmpExperience.jar EmpExper dataset.csv EmpExp.txt
2021-07-10 13:16:40,763 INFO conf.Configuration: resource-types.xml not found
2021-07-10 13:16:40,763 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-07-10 13:16:40,920 INFO impl.YarnClientImpl: Submitted application application_1625897933179_0002
2021-07-10 13:16:41,044 INFO mapreduce.Job: The url to track the job: http://hadoop-VirtualBox:8088/proxy/application_1625897933179_0002/
2021-07-10 13:16:41,050 INFO mapreduce.Job: Running job: job_1625897933179_0002
2021-07-10 13:16:51,614 INFO mapreduce.Job: Job job_1625897933179_0002 running in uber mode : false
2021-07-10 13:16:51,617 INFO mapreduce.Job: map 0% reduce 0%
2021-07-10 13:17:04,875 INFO mapreduce.Job: map 100% reduce 0%
2021-07-10 13:17:13,048 INFO mapreduce.Job: map 100% reduce 100%
2021-07-10 13:17:14,071 INFO mapreduce.Job: Job job_1625897933179_0002 completed successfully
2021-07-10 13:17:14,368 INFO mapreduce.Job: Counters: 54
      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=1427
                  HDFS: Number of bytes written=56
                  HDFS: Number of read operations=11
                  HDFS: Number of large read operations=0
                  HDFS: Number of write operations=2
```



hadoop@hadoop-VirtualBox: ~/hadoop-3.2.1/sbin

```
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)=361
CPU time spent (ms)=1760
Physical memory (bytes) snapshot=510971904
Virtual memory (bytes) snapshot=7447531520
Total committed heap usage (bytes)=355999744
Peak Map Physical memory (bytes)=197791744
Peak Map Virtual memory (bytes)=2479611904
Peak Reduce Physical memory (bytes)=115412992
Peak Reduce Virtual memory (bytes)=2488307712
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=1235
File Output Format Counters
Bytes Written=56
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop fs -ls EmpExp.txt
```



```
hadoop@hadoop-VirtualBox: ~/hadoop-3.2.1/sbin
```

```
Peak Reduce Virtual memory (bytes)=2488307712
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=1235
File Output Format Counters
Bytes Written=56
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop fs -ls EmpExp.txt
Found 2 items
-rw-r--r--  1 hdoop supergroup      0 2021-07-10 13:17 EmpExp.txt/_SUCCESS
-rw-r--r--  1 hdoop supergroup    56 2021-07-10 13:17 EmpExp.txt/part-00000
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop fs -cat EmpExp.txt/part-00000
cat: `EmpExp.txt/part-00000': No such file or directory
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop fs -cat EmpExp.txt/part-00000
2021-07-10 13:18:08,341 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
Total no.of employees having 5 years of experience :      8
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ 
```

```
hadoop@hadoop-VirtualBox: ~/Desktop
hadoop@hadoop-Virt... x      hadoop@hadoop-Vir... x      hadoop@hadoop-Vir...
hadoop@hadoop-VirtualBox:~$ cd /home/hadoop/Desktop
hadoop@hadoop-VirtualBox:~/Desktop$ ls
dataset.csv
'Eclipse IDE for Java Developers - 2021-03.desktop'
EmpAddressclass.jar
EmpAddress.jar
EmpExpclass.jar
hadoop@hadoop-VirtualBox:~/Desktop$          EmpExperience.jar
                                         EmpExper.jar
                                         EmpExp.jar
                                         EmpISE.jar
                                         Employee.jar
```

```
hadoop@hadoop-VirtualBox: ~/hadoop-3.2.1/sbin
hadoop@hadoop-Virt... x      hadoop@hadoop-Vir... x      hadoop@hadoop-Vir...
total no.of employees having 5 years or experience : 8
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop jar /home/hadoop/Desktop/EmpAddressBangalore.jar EmpAddr dataset.csv EmpAdd.txt
2021-07-10 13:33:04,191 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-10 13:33:04,664 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-10 13:33:04,957 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2021-07-10 13:33:05,023 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1625897933179_0003
2021-07-10 13:33:05,312 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-10 13:33:05,609 INFO mapred.FileInputFormat: Total input files to process : 1
2021-07-10 13:33:05,693 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-10 13:33:05,766 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-10 13:33:05,794 INFO mapreduce.JobSubmitter: number of splits:2
2021-07-10 13:33:06,038 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-10 13:33:06,127 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1625897933179_0003
2021-07-10 13:33:06,127 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-07-10 13:33:06,532 INFO conf.Configuration: resource-types.xml not found
```

```
hadoop@hadoop-VirtualBox: ~/hadoop-3.2.1/sbin
```

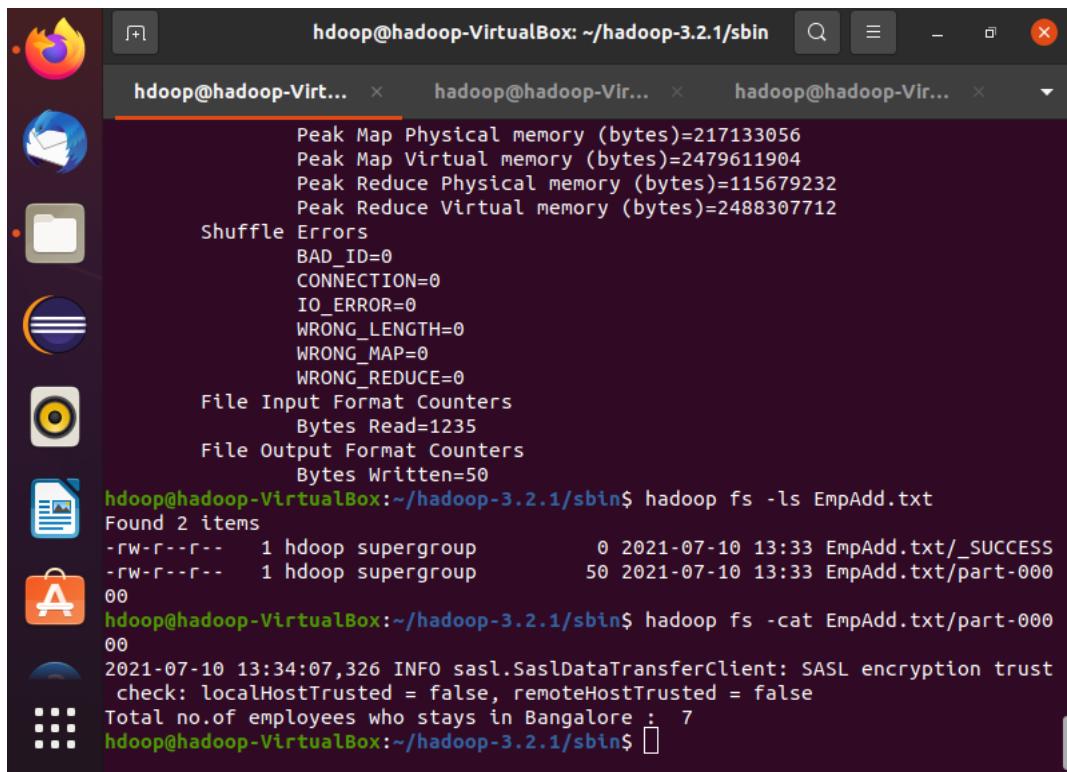
```
hadoop@hadoop-Virt... hadoop@hadoop-Vir... hadoop@hadoop-Vir...
```

```
HDFS: Number of bytes read erasure-coded=0
Job Counters
    Launched map tasks=2
    Launched reduce tasks=1
    Data-local map tasks=2
    Total time spent by all maps in occupied slots (ms)=17366
    Total time spent by all reduces in occupied slots (ms)=4823
    Total time spent by all map tasks (ms)=17366
    Total time spent by all reduce tasks (ms)=4823
    Total vcore-milliseconds taken by all map tasks=17366
    Total vcore-milliseconds taken by all reduce tasks=4823
    Total megabyte-milliseconds taken by all map tasks=17782784
    Total megabyte-milliseconds taken by all reduce tasks=4938752
Map-Reduce Framework
    Map input records=25
    Map output records=7
    Map output bytes=364
    Map output materialized bytes=120
    Input split bytes=192
    Combine input records=7
    Combine output records=2
    Reduce input groups=1
    Reduce shuffle bytes=120
    Reduce input records=2
    Reduce output records=1
    Spilled Records=4
    Shuffled Maps =2
```

```
hadoop@hadoop-VirtualBox: ~/hadoop-3.2.1/sbin
```

```
hadoop@hadoop-Virt... hadoop@hadoop-Vir... hadoop@hadoop-Vir...
```

```
reduce shuffle bytes=120
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)=417
CPU time spent (ms)=1650
Physical memory (bytes) snapshot=549564416
Virtual memory (bytes) snapshot=7447531520
Total committed heap usage (bytes)=355999744
Peak Map Physical memory (bytes)=217133056
Peak Map Virtual memory (bytes)=2479611904
Peak Reduce Physical memory (bytes)=115679232
Peak Reduce Virtual memory (bytes)=2488307712
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=1235
File Output Format Counters
    Bytes Written=50
```



The screenshot shows a terminal window titled "hadoop@hadoop-VirtualBox: ~/hadoop-3.2.1/sbin". The window contains the following text:

```
Peak Map Physical memory (bytes)=217133056
Peak Map Virtual memory (bytes)=2479611904
Peak Reduce Physical memory (bytes)=115679232
Peak Reduce Virtual memory (bytes)=2488307712
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=1235
File Output Format Counters
  Bytes Written=50
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop fs -ls EmpAdd.txt
Found 2 items
-rw-r--r--  1 hdoop supergroup          0  2021-07-10 13:33 EmpAdd.txt/_SUCCESS
-rw-r--r--  1 hdoop supergroup      50  2021-07-10 13:33 EmpAdd.txt/part-000
00
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop fs -cat EmpAdd.txt/part-000
00
2021-07-10 13:34:07,326 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
Total no.of employees who stays in Bangalore :  7
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$
```

Exercise-II

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.

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

1. Insert 5 records using INSERT command.

```
hive> create database if not exists EmployeeDB;
OK
Time taken: 0.081 seconds
hive> show databases;
OK
bank_hive
```

```
default
employeedb
Time taken: 0.035 seconds, Fetched: 3 row(s)
hive> use employeedb;
OK
Time taken: 0.049 seconds
hive> create table employee(name string,ssn int,salary int,address string,dname
string,exp int) row format delimited fields terminated by ",";
OK
Time taken: 0.136 seconds
hive> show tables;
OK
employee
Time taken: 0.06 seconds, Fetched: 1 row(s)
hive> insert into employee
values("Kavya",5000,60000,"Bangalore","ISE",3),("Lishel",5001,25000,"Mumbai","CS
E",6),("Chavi",5002,55000,"Bangalore","ISE",4),("Jothsna",5003,35000,"Hyderabad",
"CSE",7),("Aishwarya",5004,80000,"Pune","ECE",2);

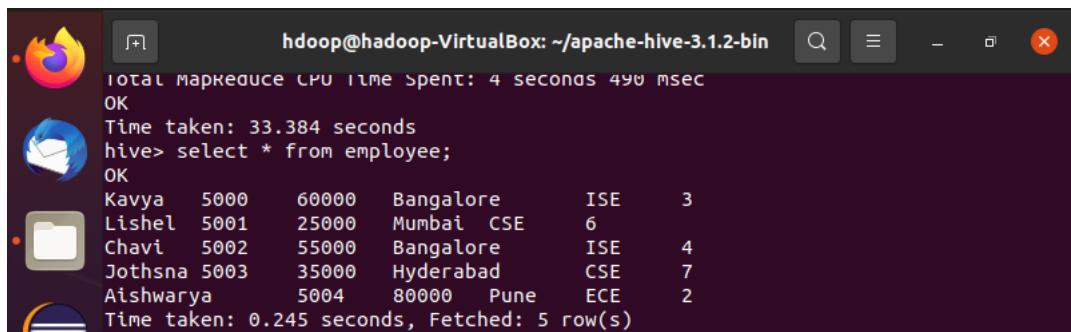
hive> select * from employee;
OK
Kavya 5000 60000 Bangalore ISE 3
Lishel 5001 25000 Mumbai CSE 6
Chavi 5002 55000 Bangalore ISE 4
Jothsna 5003 35000 Hyderabad CSE 7
Aishwarya 5004 80000 Pune ECE 2
Time taken: 0.245 seconds, Fetched: 5 row(s)
```

Output screenshots :

```
hadoop@hadoop-VirtualBox: ~/apache-hive-3.1.2-bin
Time taken: 0.252 seconds, Fetched: 11 row(s)
hive>
>
>
> create database if not exists EmployeeDB;
OK
Time taken: 0.081 seconds
hive> show databases;
OK
bank_hive
default
employeedb
Time taken: 0.035 seconds, Fetched: 3 row(s)
hive> use employeedb;
OK
Time taken: 0.049 seconds
hive> create table employee(name string,ssn int,salary int,address string,dname
string,exp int) row format delimited fields terminated by ",";
OK
Time taken: 0.136 seconds
hive> show tables;
OK
employee
Time taken: 0.06 seconds, Fetched: 1 row(s)
hive> insert into employee values("Kavya",5000,60000,"Bangalore","ISE",3),("Lis
hel",5001,25000,"Mumbai","CSE",6),("Chavi",5002,55000,"Bangalore","ISE",4),("Jo
thsna",5003,35000,"Hyderabad","CSE",7),("Aishwarya",5004,80000,"Pune","ECE",2);
```

```
hadoop@hadoop-VirtualBox: ~/apache-hive-3.1.2-bin
Time taken: 0.06 seconds, Fetched: 1 row(s)
hive> insert into employee values("Kavya",5000,60000,"Bangalore","ISE",3),("Lis
hel",5001,25000,"Mumbai","CSE",6),("Chavi",5002,55000,"Bangalore","ISE",4),("Jo
thsna",5003,35000,"Hyderabad","CSE",7),("Aishwarya",5004,80000,"Pune","ECE",2);

Query ID = hadoop_20210707211609_8d3d102b-5094-425a-b46a-d89557fcfc87
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_1625642025578_0008, Tracking URL = http://hadoop-VirtualBox:
8088/proxy/application_1625642025578_0008/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1625642025578
_0008
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-07 21:16:23,820 Stage-1 map = 0%,  reduce = 0%
2021-07-07 21:16:33,328 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.62 s
ec
2021-07-07 21:16:41,703 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.49
sec
MapReduce Total cumulative CPU time: 4 seconds 490 msec
Ended Job = job_1625642025578_0008
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
```



```
total MapReduce CPU time spent: 4 seconds 490 msec
OK
Time taken: 33.384 seconds
hive> select * from employee;
OK
Kavya 5000 60000 Bangalore ISE 3
Lishel 5001 25000 Mumbai CSE 6
Chavi 5002 55000 Bangalore ISE 4
Jothsna 5003 35000 Hyderabad CSE 7
Aishwarya 5004 80000 Pune ECE 2
Time taken: 0.245 seconds, Fetched: 5 row(s)
```

dataset in .csv file loaded into employee table :-

```
hive> LOAD DATA LOCAL INPATH '/home/hadoop/Desktop/dataset.csv' into table employee;
```

```
Loading data to table employeedb.employee
```

```
OK
```

```
Time taken: 0.864 seconds
```

```
hive> select * from employee;
```

```
OK
```

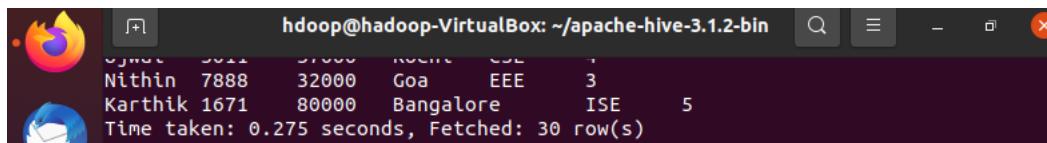
```
Kavya 5000 60000 Bangalore ISE 3
Lishel 5001 25000 Mumbai CSE 6
Chavi 5002 55000 Bangalore ISE 4
Jothsna 5003 35000 Hyderabad CSE 7
Aishwarya 5004 80000 Pune ECE 2
Anusha 6747 70000 Bangalore ISE 5
Avani 4325 30000 Kerala CSE 3
Deepti 7901 58000 Delhi ISE 5
Khushi 2135 25000 Punjab ECE 2
Monika 3984 67000 Kanpur EEE 4
Vanditha 1299 43000 Mumbai ME 6
Gaurav 7733 21000 Rajasthan CSE 7
Anjali 5231 65000 Bangalore ISE 5
Namitha 1177 54000 Pune ECE 4
Ajay 3662 73000 Chennai EEE 6
```

Akash	9815	20000	Bangalore	ISE	5	
Darshan		8000	45000	Kolkata	CSE	8
Anish	3712	35000	Patna	AE	2	
Nandan		8868	15000	Bhopal	TC	3
Purnima		6734	67000	Bangalore	ISE	5
Sharada		4599	34000	Indore	CSE	2
Pratik	2001	75000	Bangalore	ISE	5	
Pooja	6021	28000	Shimla	EEE	4	
Vikas	4378	32000	Agra	CSE	2	
Keerthi	1008	65000	Bangalore	ISE	5	
Sahana	4289	26000	Mangalore	EEE	4	
Deepak		6710	77000	Coimbatore	ECE	6
Ujwal	3011	37000	Kochi	CSE	4	
Nithin	7888	32000	Goa	EEE	3	
Karthik	1671	80000	Bangalore	ISE	5	

Time taken: 0.275 seconds, Fetched: 30 row(s)

```
hive> LOAD DATA LOCAL INPATH '/home/hadoop/Downloads/dataset.csv' into table employee;
      Loading data to table employeedb.employee
      OK
```

```
hive> select * from employee;
OK
+-----+-----+-----+-----+-----+-----+
| Name | ID  | Salary| City  | Branch| Dept  |
+-----+-----+-----+-----+-----+-----+
| Kavya | 5000| 60000| Bangalore| ISE | 3   |
| Lishel | 5001| 25000| Mumbai | CSE | 6   |
| Chavi | 5002| 55000| Bangalore| ISE | 4   |
| Jothsna | 5003| 35000| Hyderabad| CSE | 7   |
| Aishwarya | 5004| 80000| Pune | ECE | 2   |
| Anusha | 6747| 70000| Bangalore| ISE | 5   |
| Avani | 4325| 30000| Kerala | CSE | 3   |
| Deepti | 7901| 58000| Delhi | ISE | 5   |
| Khushi | 2135| 25000| Punjab | ECE | 2   |
| Monika | 3984| 67000| Kanpur | EEE | 4   |
| Vanditha | 1299| 43000| Mumbai | ME | 6   |
| Gaurav | 7733| 21000| Rajasthan| CSE | 7   |
| Anjali | 5231| 65000| Bangalore| ISE | 5   |
| Namitha | 1177| 54000| Pune | ECE | 4   |
| Ajay | 3662| 73000| Chennai | EEE | 6   |
| Akash | 9815| 20000| Bangalore| ISE | 5   |
| Darshan | 8000| 45000| Kolkata | CSE | 8   |
| Anish | 3712| 35000| Patna | AE | 2   |
| Nandan | 8868| 15000| Bhopal | TC | 3   |
| Purnima | 6734| 67000| Bangalore| ISE | 5   |
| Sharada | 4599| 34000| Indore | CSE | 2   |
| Pratik | 2001| 75000| Bangalore| ISE | 5   |
| Pooja | 6021| 28000| Shimla | EEE | 4   |
| Vikas | 4378| 32000| Agra | CSE | 2   |
| Keerthi | 1008| 65000| Bangalore| ISE | 5   |
| Sahana | 4289| 26000| Mangalore| EEE | 4   |
| Deepak | 6710| 77000| Coimbatore| ECE | 6   |
+-----+-----+-----+-----+-----+-----+
```



```
hadoop@hadoop-VirtualBox: ~/apache-hive-3.1.2-bin
+-----+-----+-----+-----+-----+
| Sjani | 5011 | 37000 | Kozhikode | CSE |    1
| Nithin | 7888 | 32000 | Goa | EEE |    3
| Karthik | 1671 | 80000 | Bangalore | ISE |    5
+-----+-----+-----+-----+-----+
Time taken: 0.275 seconds, Fetched: 30 row(s)
```

2. Demonstrate the Alter command for the following cases,

a. Rename the table name to “Emp”.

```
hive> alter table employee rename to emp;
```

OK

Time taken: 0.335 seconds

```
hive> show tables;
```

OK

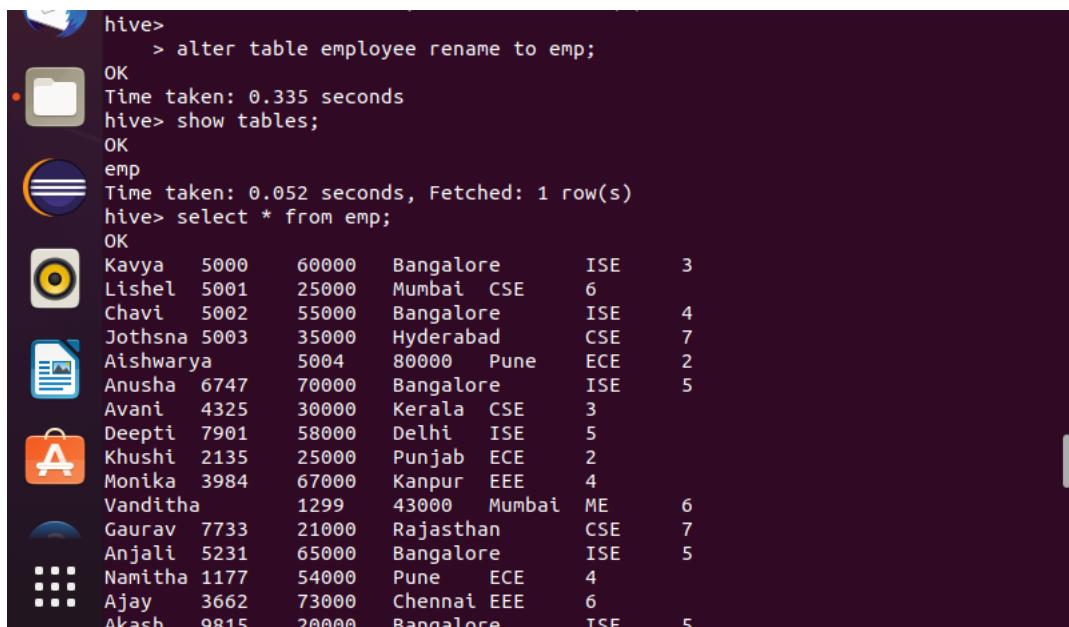
emp

Time taken: 0.052 seconds, Fetched: 1 row(s)

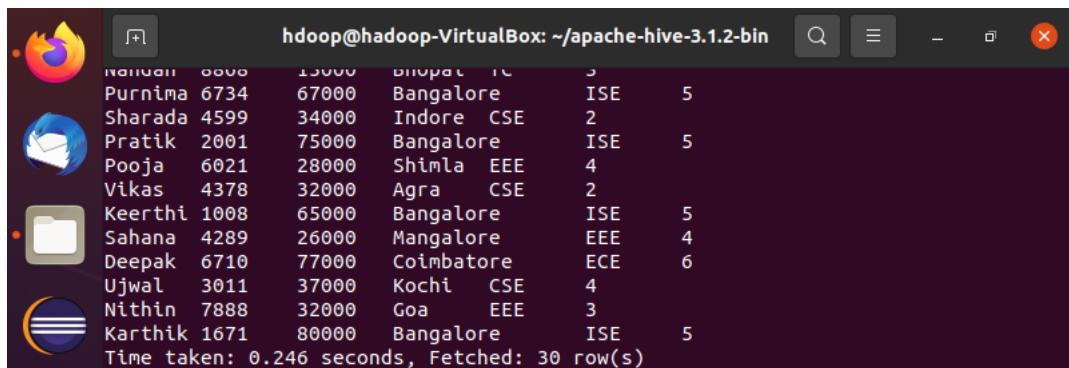
```
hive> select * from emp;
```

OK

Output screenshots :



```
hive> > alter table employee rename to emp;
OK
Time taken: 0.335 seconds
hive> show tables;
OK
emp
Time taken: 0.052 seconds, Fetched: 1 row(s)
hive> select * from emp;
OK
+-----+-----+-----+-----+-----+
| Kavya | 5000 | 60000 | Bangalore | ISE |    3
| Lishel | 5001 | 25000 | Mumbai | CSE |    6
| Chavi | 5002 | 55000 | Bangalore | ISE |    4
| Jothsna | 5003 | 35000 | Hyderabad | CSE |    7
| Aishwarya | 5004 | 80000 | Pune | ECE |    2
| Anusha | 6747 | 70000 | Bangalore | ISE |    5
| Avani | 4325 | 30000 | Kerala | CSE |    3
| Deepti | 7901 | 58000 | Delhi | ISE |    5
| Khushi | 2135 | 25000 | Punjab | ECE |    2
| Monika | 3984 | 67000 | Kanpur | EEE |    4
| Vanditha | 1299 | 43000 | Mumbai | ME |    6
| Gaurav | 7733 | 21000 | Rajasthan | CSE |    7
| Anjali | 5231 | 65000 | Bangalore | ISE |    5
| Namitha | 1177 | 54000 | Pune | ECE |    4
| Ajay | 3662 | 73000 | Chennai | EEE |    6
| Akash | 9815 | 20000 | Bangalore | TSF |    5
+-----+-----+-----+-----+-----+
```



Name	SSN	Salary	Address	Dname	Experience
Narayan	6666	15000	Bhopal	TC	3
Purnima	6734	67000	Bangalore	ISE	5
Sharada	4599	34000	Indore	CSE	2
Pratik	2001	75000	Bangalore	ISE	5
Pooja	6021	28000	Shimla	EEE	4
Vikas	4378	32000	Agra	CSE	2
Keerthi	1008	65000	Bangalore	ISE	5
Sahana	4289	26000	Mangalore	EEE	4
Deepak	6710	77000	Coimbatore	ECE	6
Ujwal	3011	37000	Kochi	CSE	4
Nithin	7888	32000	Goa	EEE	3
Karthik	1671	80000	Bangalore	ISE	5

Time taken: 0.246 seconds, Fetched: 30 row(s)

b. Rename the column name “Dname” to “Dept_name”.

```
hive> alter table emp change dname Dept_name string;
```

```
OK
```

```
Time taken: 0.179 seconds
```

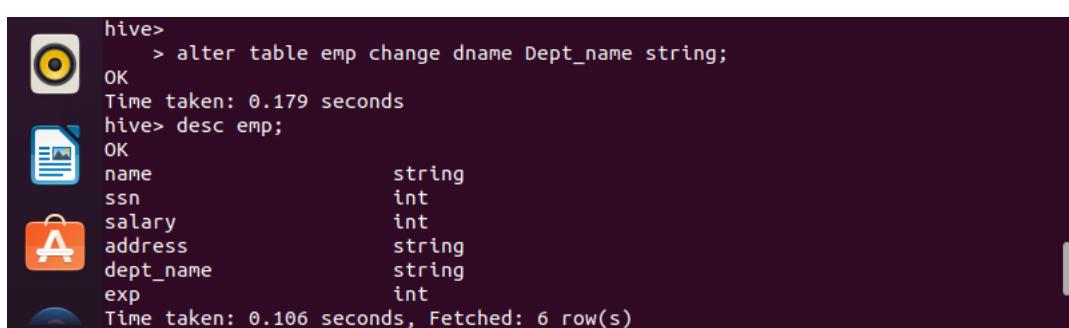
```
hive> desc emp;
```

```
OK
```

name	string
ssn	int
salary	int
address	string
dept_name	string
exp	int

```
Time taken: 0.106 seconds, Fetched: 6 row(s)
```

Output screenshots :

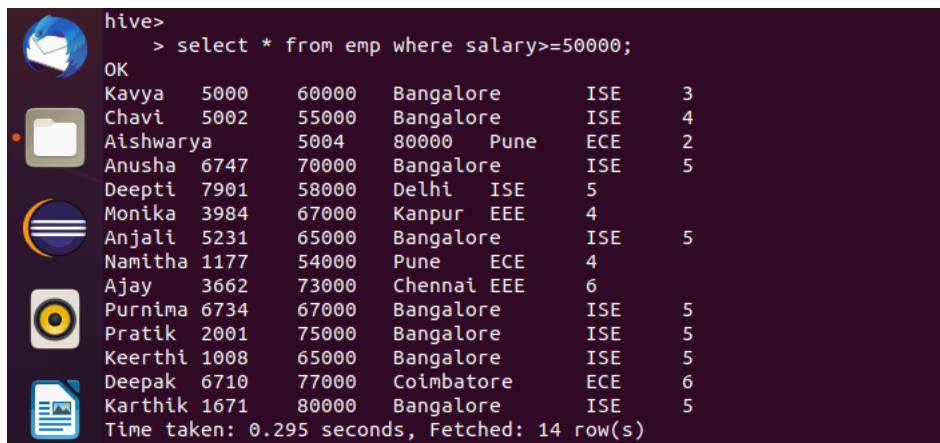


```
hive> alter table emp change dname Dept_name string;
OK
Time taken: 0.179 seconds
hive> desc emp;
OK
name          string
ssn           int
salary         int
address        string
dept_name      string
exp            int
Time taken: 0.106 seconds, Fetched: 6 row(s)
```

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

```
hive> select * from emp where salary>=50000;  
OK
```

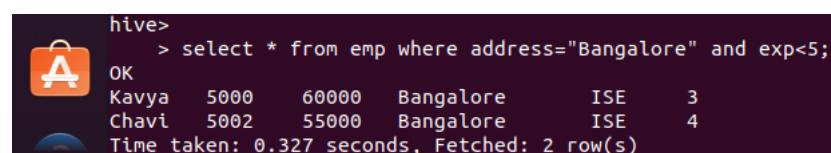
Output screenshots :



```
hive>  
    > select * from emp where salary>=50000;  
OK  
Kavya      5000      60000      Bangalore      ISE      3  
Chavi      5002      55000      Bangalore      ISE      4  
Aishwarya   5004      80000      Pune        ECE      2  
Anusha     6747      70000      Bangalore      ISE      5  
Deepti     7901      58000      Delhi        ISE      5  
Monika     3984      67000      Kanpur       EEE      4  
Anjali     5231      65000      Bangalore      ISE      5  
Namitha    1177      54000      Pune        ECE      4  
Ajay       3662      73000      Chennai      EEE      6  
Purnima    6734      67000      Bangalore      ISE      5  
Pratik     2001      75000      Bangalore      ISE      5  
Keerthi    1008      65000      Bangalore      ISE      5  
Deepak     6710      77000      Coimbatore   ECE      6  
Karthik    1671      80000      Bangalore      ISE      5  
Time taken: 0.295 seconds, Fetched: 14 row(s)
```

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

```
hive> select * from emp where address="Bangalore" and exp<5;  
OK
```



```
hive>  
    > select * from emp where address="Bangalore" and exp<5;  
OK  
Kavya      5000      60000      Bangalore      ISE      3  
Chavi      5002      55000      Bangalore      ISE      4  
Time taken: 0.327 seconds, Fetched: 2 row(s)
```

5. Create separate view containing Name, Dept_name of employees

```
hive> create view emp_dept as select name,dept_name from emp;  
OK  
Time taken: 0.264 seconds  
hive> show tables;  
OK  
emp  
emp_dept  
Time taken: 0.048 seconds, Fetched: 2 row(s)
```



```

hadoop@hadoop-VirtualBox: ~ / apache-hive-3.1.2-bin
2021-07-08 10:26:01,908 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 1.84 sec
2021-07-08 10:26:11,337 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 3.92 sec
MapReduce Total cumulative CPU time: 3 seconds 920 msec
Ended Job = job_1625642025578_0010
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.19 sec HDFS Read: 12805
HDFS Write: 908 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 3.92 sec HDFS Read: 8318 HDFS Write: 809 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 110 msec
OK
Aishwarya      5004
Ajay          3662
Akash          9815
Anish          3712
Anjali          5231
Anusha          6747
Avani          4325
Chavi          5002
Darshan        8000
Deepak          6710
Deepti          7901
Gaurav          7733
Jothsna        5003
Karthik         1671
Kavya          5000
Keerthi        1008
Karthik        1671
Kavya          5000
Keerthi        1008
Khushi          2135
Lishel          5001
Monika          3984
Namitha        1177
Nandan          8868
Nithin          7888
Pooja          6021
Pratik          2001
Purnima         6734
Sahana          4289
Sharada         4599
Ujwal          3011
Vanditha        1299
Vikas          4378
Time taken: 74.877 seconds, Fetched: 30 row(s)

```

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

hive> select max(salary),min(salary),avg(salary) from emp;

Output screenshots :

```

hive>
>
> select max(salary),min(salary),avg(salary) from emp;
Query ID = hdoop_20210708102704_319362e2-3c89-4cf6-90ec-34748ab4faf8
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>

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_1625642025578_0011, Tracking URL = http://hadoop-VirtualBox:
8088/proxy/application_1625642025578_0011/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job -kill job_1625642025578
_0011
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-08 10:27:17,985 Stage-1 map = 0%,  reduce = 0%
2021-07-08 10:27:26,451 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.02 s
ec
2021-07-08 10:27:36,935 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.0
sec
MapReduce Total cumulative CPU time: 5 seconds 0 msec
Ended Job = job_1625642025578_0011
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1  Cumulative CPU: 5.0 sec  HDFS Read: 18239 H
DFS Write: 130 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 0 msec
OK
80000  15000  47633.333333333336
Time taken: 34.083 seconds, Fetched: 1 row(s)
hive> 
```

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

Outer join :-

```
hive>
    > create table department(dno int,dname string) row format delimited fields
terminated by ",";
OK
Time taken: 0.105 seconds
hive>
    > insert into table department values(1,'ISE'),(2,'CSE'),(3,'ISE'),(4,'ECE')
),(5,'EEE'),(6,'ISE');
Query ID = hdoop_20210709130638_6304b267-181b-4de1-aba1-65009c99011d
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_1625642025578_0012, Tracking URL = http://hadoop-VirtualBox:
8088/proxy/application_1625642025578_0012/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1625642025578
_0012
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-09 13:06:53,492 Stage-1 map = 0%,  reduce = 0%
2021-07-09 13:07:04,017 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.59 s
ec
2021-07-09 13:07:13,441 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.54
sec
```

```
Time taken: 36.296 seconds, Fetched: 50 row(s)
hive>
    >
    >
        > SELECT d.dno,e.dept_name FROM emp e FULL OUTER JOIN department d ON(e.dep
t_name=d.dname);
Query ID = hdoop_20210709132028_a7790eed-096f-4e13-afff-f169c2d7a299
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_1625642025578_0017, Tracking URL = http://hadoop-VirtualBox:
8088/proxy/application_1625642025578_0017/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1625642025578
_0017
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2021-07-09 13:20:41,771 Stage-1 map = 0%,  reduce = 0%
2021-07-09 13:20:55,934 Stage-1 map = 50%,  reduce = 0%, Cumulative CPU 1.74 se
c
2021-07-09 13:20:57,035 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.49 s
ec
2021-07-09 13:21:05,435 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.35
sec
MapReduce Total cumulative CPU time: 5 seconds 350 msec
Ended Job = job_1625642025578_0017
```

```
hadoop@hadoop-VirtualBox: ~/apache-hive-3.1.2-bin
Stage-Stage-1: Map: 2  Reduce: 1  Cumulative CPU: 5.35 sec   HDFS Read: 17373
HDFS Write: 987 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 350 msec
OK
NULL      AE
2          CSE
4          ECE
4          ECE
4          ECE
4          ECE
5          EEE
5          EEE
5          EEE
5          EEE
5          EEE
6          ISE
3          ISE
1          ISE
6          ISE
3          ISE
1          ISE
6          ISE
```

```
hadoop@hadoop-VirtualBox: ~/apache-hive-3.1.2-bin
1          ISE
6          ISE
3          ISE
1          ISE
NULL     ME
NULL     TC
Time taken: 38.923 seconds, Fetched: 50 row(s)
```

Left outer join :-

```
hadoop@hadoop-VirtualBox: ~/apache-hive-3.1.2-bin
Time taken: 38.923 seconds, Fetched: 50 row(s)
hive>
> SELECT d.dno,e.dept_name FROM emp e LEFT OUTER JOIN department d ON(e.dept_name=d.dname);
Query ID = hdoop_20210709132119_8c67e36c-c490-4e44-83c0-fd595d516837
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_1625642025578_0018, Tracking URL = http://hadoop-VirtualBox:8088/proxy/application_1625642025578_0018/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1625642025578_0018
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2021-07-09 13:21:44,189 Stage-3 map = 0%,  reduce = 0%
2021-07-09 13:21:52,580 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 2.08 sec
MapReduce Total cumulative CPU time: 2 seconds 80 msec
Ended Job = job_1625642025578_0018
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1  Cumulative CPU: 2.08 sec  HDFS Read: 10141 HDFS Write: 987 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 80 msec
OK
```

```
1      ISE
3      ISE
6      ISE
2      CSE
1      ISE
3      ISE
6      ISE
2      CSE
4      ECE
1      ISE
3      ISE
6      ISE
2      CSE
1      ISE
3      ISE
6      ISE
4      ECE
5      EEE
NULL   ME
2      CSE
1      ISE
3      ISE
6      ISE
4      ECE
5      EEE
1      ISE
3      ISE
```

```
hadoop@hadoop-VirtualBox: ~/apache-hive-3.1.2-bin
```

•	6 ISE
	4 ECE
•	5 EEE
	1 ISE
	3 ISE
•	6 ISE
	2 CSE
•	NULL AE
	NULL TC
	1 ISE
	3 ISE
	6 ISE
	2 CSE
	1 ISE
	3 ISE
	6 ISE
	5 EEE
	2 CSE
	1 ISE
	3 ISE
	6 ISE
	5 EEE
	4 ECE
	2 CSE
	5 EEE
	1 ISE
	3 ISE
	6 ISE
	5 EEE
	4 ECE
	2 CSE
	5 EEE
	1 ISE
	3 ISE
	6 ISE
	..

```
Time taken: 35.582 seconds, Fetched: 50 row(s)
```

Right outer join:-

```
hive>
>
> SELECT distinct d.dno,e.dept_name FROM emp e RIGHT OUTER JOIN department
d ON(e.dept_name=d.dname);
Query ID = hdoop_20210709132359_c6270259-17ed-423d-84b5-2b8e1b293f67
Total jobs = 1
Execution completed successfully
MapredLocal task succeeded
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_1625642025578_0020, Tracking URL = http://hadoop-VirtualBox:
8088/proxy/application_1625642025578_0020/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job -kill job_1625642025578
_0020
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2021-07-09 13:24:22,032 Stage-2 map = 0%,  reduce = 0%
2021-07-09 13:24:32,710 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 2.4 se
c
2021-07-09 13:24:41,140 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 4.59
sec
MapReduce Total cumulative CPU time: 4 seconds 590 msec
Ended Job = job_1625642025578_0020
```

```
hdoop@hadoop-VirtualBox: ~/apache-hive-3.1.2-bin
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_1625642025578_0020, Tracking URL = http://hadoop-VirtualBox:
8088/proxy/application_1625642025578_0020/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job -kill job_1625642025578
_0020
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2021-07-09 13:24:22,032 Stage-2 map = 0%,  reduce = 0%
2021-07-09 13:24:32,710 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 2.4 se
c
2021-07-09 13:24:41,140 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 4.59
sec
MapReduce Total cumulative CPU time: 4 seconds 590 msec
Ended Job = job_1625642025578_0020
MapReduce Jobs Launched:
Stage-Stage-2: Map: 1  Reduce: 1  Cumulative CPU: 4.59 sec  HDFS Read: 17004
HDFS Write: 195 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 590 msec
OK
2      CSE
4      ECE
5      EEE
1      ISE
3      ISE
6      ISE
Time taken: 45.89 seconds, Fetched: 6 row(s)
hive>
```