

NAME : FAIZAH ALI

USN : 1NT18IS057

BIG DATA LAB LA2

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.

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 :-

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);

    }

}

```

eclipse-workspace - Employee/src/EmpISE.java - Eclipse IDE

```

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

```

eclipse-workspace - Employee/src/EmpISE.java - Eclipse IDE

```

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

```

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]));
}

```

```

        FileOutputStream.setOutputPath(conf, new Path(args[1]));

    JobClient.runJob(conf);

}

}

```

eclipse-workspace - EmpExperience/src/EmpExper.java - Eclipse IDE

```

import java.io.IOException;
public class EmpExper {
    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);
            }
        }
    }
}

```

eclipse-workspace - EmpExperience/src/EmpExper.java - Eclipse IDE

```

    }
}
//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 {
        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);
    TextInputFormat.setInputPaths(conf, new Path(args[0]));
    TextOutputFormat.setOutputPath(conf, new Path(args[1]));
    JobClient.runJob(conf);
}
}

```

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]));

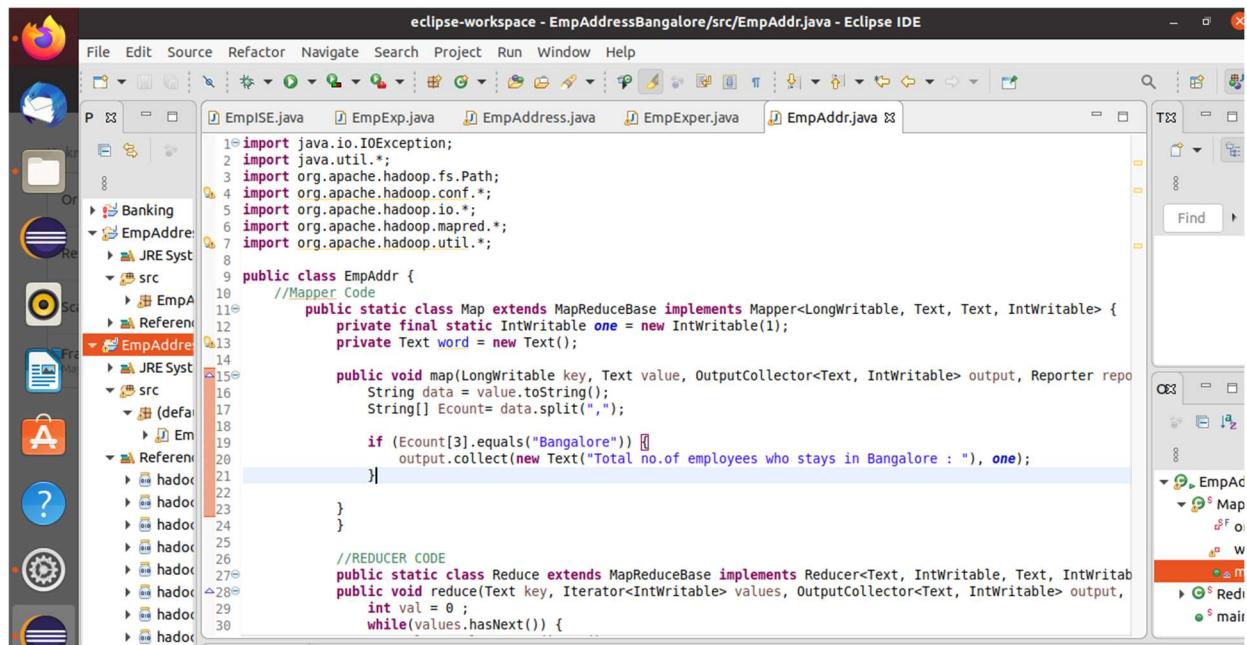
        JobClient.runJob(conf);

    }

}

```

}



```
25
26
27
28 //REDUCER CODE
29 public static class Reducer extends MapReduceBase implements Reducer<Text, IntWritable, Text, IntWritable>
30     public void reduce(Text key, Iterator<IntWritable> values, OutputCollector<Text, IntWritable> output,
31         int val = 0;
32         while(values.hasNext()) {
33             val += values.next().get();
34         }
35         output.collect(key, new IntWritable(val));
36     }
37
38 //DRIVER CODE
39 public static void main(String[] args) throws Exception {
40     JobConf conf = new JobConf(EmpAddr.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(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 }
```

Hadoop Map-reduce commands and outputs :

```
hadoop@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:~$ cd /home/hadoop/Desktop
```

```
hdoop@hadoop-VirtualBox:/home/hadoop/Desktop$ ls
```

```
dataset.csv          EmpExpclass.jar

'Eclipse IDE for Java Developers - 2021-03.desktop'  EmpExp.jar

EmpAddressclass.jar      EmplSE.jar

EmpAddress.jar          Employee.jar

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/hadoop
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/hadoop
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/hadoop/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 fs -ls
Found 2 items

-rw-r--r-- 1 hdoop supergroup    823 2021-07-10 12:11 dataset.csv
drwxr-xr-x - hdoop supergroup    0 2021-07-10 12:11 faizah
```

```
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop jar Employee.jar Employee.EmpISE  
dataset.csv EmpISE.txt
```

JAR does not exist or is not a normal file: /home/hadoop/hadoop-3.2.1/sbin/Employee.jar

```
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop jar Employee.jar EmpISE dataset.csv  
EmpISE.txt
```

JAR does not exist or is not a normal file: /home/hadoop/hadoop-3.2.1/sbin/Employee.jar

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

```
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop jar Employee.jar EmpISE dataset.csv  
EmpISE.txt
```

JAR does not exist or is not a normal file: /home/hadoop/hadoop-3.2.1/sbin/Employee.jar

```
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop jar  
/home/hadoop/Desktop/Employee.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 application with  
ToolRunner to remedy this.
```

```
2021-07-10 12:47:43,992 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding 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 process : 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: []

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

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

Combine input records=8

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)=379

CPU time spent (ms)=1620

Physical memory (bytes) snapshot=537788416

Virtual memory (bytes) snapshot=7447535616

Total committed heap usage (bytes)=355999744

Peak Map Physical memory (bytes)=211656704

Peak Map Virtual memory (bytes)=2479611904

Peak Reduce Physical memory (bytes)=115912704

Peak Reduce Virtual memory (bytes)=2488311808

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=53

```
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop fs -ls EmpISE.txt
```

Found 2 items

```
-rw-r--r-- 1 hdoop supergroup      0 2021-07-10 12:48 EmpISE.txt/_SUCCESS
```

```
-rw-r--r-- 1 hdoop supergroup    53 2021-07-10 12:48 EmpISE.txt/part-00000
```

```
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop fs -cat EmpISE.txt/part-00000
```

```
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:~/hadoop-3.2.1/sbin$  
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'.  
  
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

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)=22438
Total time spent by all reduces in occupied slots (ms)=5547
Total time spent by all map tasks (ms)=22438
Total time spent by all reduce tasks (ms)=5547
Total vcore-milliseconds taken by all map tasks=22438
Total vcore-milliseconds taken by all reduce tasks=5547
Total megabyte-milliseconds taken by all map tasks=22976512
Total megabyte-milliseconds taken by all reduce tasks=5680128

Map-Reduce Framework

Map input records=25
Map output records=8
Map output bytes=464
Map output materialized bytes=132
Input split bytes=192
Combine input records=8
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)=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

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

hdoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin\$ hadoop fs -cat Empexp.txt/part-00000

cat: `Empexp.txt/part-00000': No such file or directory

hdoop@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

hdoop@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/hdoop/.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

2021-07-10 13:33:06,533 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.

2021-07-10 13:33:06,658 INFO impl.YarnClientImpl: Submitted application
application_1625897933179_0003

2021-07-10 13:33:06,744 INFO mapreduce.Job: The url to track the job: http://hadoop-VirtualBox:8088/proxy/application_1625897933179_0003/

2021-07-10 13:33:06,749 INFO mapreduce.Job: Running job: job_1625897933179_0003

2021-07-10 13:33:17,074 INFO mapreduce.Job: Job job_1625897933179_0003 running in uber mode : false

2021-07-10 13:33:17,078 INFO mapreduce.Job: map 0% reduce 0%

2021-07-10 13:33:27,336 INFO mapreduce.Job: map 50% reduce 0%

2021-07-10 13:33:28,433 INFO mapreduce.Job: map 100% reduce 0%

2021-07-10 13:33:35,550 INFO mapreduce.Job: map 100% reduce 100%

2021-07-10 13:33:36,575 INFO mapreduce.Job: Job job_1625897933179_0003 completed successfully

2021-07-10 13:33:36,795 INFO mapreduce.Job: Counters: 54

File System Counters

FILE: Number of bytes read=114

FILE: Number of bytes written=677705

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=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)=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
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

```
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop fs -ls EmpAdd.txt
```

Found 2 items

```
-rw-r--r-- 1 hadoop supergroup 0 2021-07-10 13:33 EmpAdd.txt/_SUCCESS
```

```
-rw-r--r-- 1 hadoop supergroup 50 2021-07-10 13:33 EmpAdd.txt/part-00000
```

```
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop fs -cat EmpAdd.txt/part-00000
```

```
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$
```

HADOOP MAP REDUCE SCREENSHOTS :

```
hadoop@hadoop-VirtualBox: ~$ su - hdoop
Password:
hadoop@hadoop-VirtualBox: ~$ cd hadoop-3.2.1/sbin
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: ~$ su - hdoop
Password:
hadoop@hadoop-VirtualBox: ~$ cd hadoop-3.2.1/sbin
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: /home/hadoop/Desktop$ hadoop fs -copyFromLocal dataset
```

```
hadoop@hadoop-VirtualBox: ~/hadoop... x      hadoop@hadoop-VirtualBox: ~ x
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/hadoop
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/hadoop
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/hadoop/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 fs -ls
Found 2 items
-rw-r--r--  1 hdoop supergroup     823 2021-07-10 12:11 dataset.csv
drwxr-xr-x  - hdoop supergroup          0 2021-07-10 12:11 faizah
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop jar Employee.jar Employee.EmpISE dataset.csv EmpISE.txt
JAR does not exist or is not a normal file: /home/hadoop/hadoop-3.2.1/sbin/Employee.jar
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop jar Employee.jar EmpISE dat
```

```
hadoop@hadoop-VirtualBox: ~/hadoop... x      hadoop@hadoop-VirtualBox: ~ x
drwxr-xr-x  - hdoop supergroup          0 2021-07-10 12:11 faizah
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop jar Employee.jar Employee.EmpISE dataset.csv EmpISE.txt
JAR does not exist or is not a normal file: /home/hadoop/hadoop-3.2.1/sbin/Employee.jar
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop jar Employee.jar EmpISE dataset.csv EmpISE.txt
JAR does not exist or is not a normal file: /home/hadoop/hadoop-3.2.1/sbin/Employee.jar
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ 
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop jar Employee.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 application with ToolRunner to remedy this.
2021-07-10 12:47:43,992 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding 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
```

```
option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2021-07-10 12:47:43,992 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding 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 process : 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: []
```

```
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$ hadoop fs -report
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 -report
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=379
CPU time spent (ms)=1620
Physical memory (bytes) snapshot=537788416
Virtual memory (bytes) snapshot=7447535616
Total committed heap usage (bytes)=355999744
Peak Map Physical memory (bytes)=211656704
Peak Map Virtual memory (bytes)=2479611904
Peak Reduce Physical memory (bytes)=115912704
Peak Reduce Virtual memory (bytes)=2488311808
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=53
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop fs -ls EmpISE.txt
Found 2 items
-rw-r--r--  1 hdoop supergroup          0 2021-07-10 12:48 EmpISE.txt/_SUCCESS
-rw-r--r--  1 hdoop supergroup      53 2021-07-10 12:48 EmpISE.txt/part-000000
```

```
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  
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 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@hadoop-Virt... x hadoop@hadoop-Vir... x hadoop@hadoop-Vir...  
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-t
```

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

```
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$ 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-000
00
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop fs -cat Empexp.txt/part-000
00
cat: `Empexp.txt/part-00000': No such file or directory
hadoop@hadoop-VirtualBox:~/hadoop-3.2.1/sbin$ hadoop fs -cat EmpExp.txt/part-000
00
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$ 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$
```

```
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 ap  
plication with ToolRunner to remedy this.  
2021-07-10 13:33:05,023 INFO mapreduce.JobResourceUploader: Disabling Erasure C  
oding 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 proce  
ss : 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
```

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

```
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","CSE",
6),("Chavi",5002,55000,"Bangalore","ISE",4),("Jothsna",5003,35000,"Hyderabad","CSE",
7),("Aishwarya",5004,80000,"Pune","ECE",2);

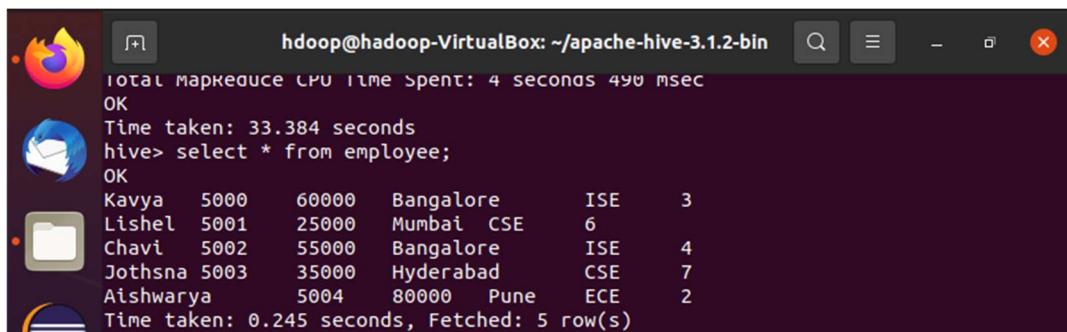
hive> select * from employee;
OK
Kavya 5000 60000 Bangalore ISE 3
Lishel 5001 25000 Mumbai CSE 6
Chavi 5002 55000 Bangalore ISE 4
Jothsna 5003 35000 Hyderabad CSE 7
Aishwarya 5004 80000 Pune ECE 2
Time taken: 0.245 seconds, Fetched: 5 row(s)
```

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/Downloads/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| Grade |
+-----+-----+-----+-----+-----+-----+
| 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
+-----+-----+-----+-----+-----+
| Name | ID  | Salary| City  | Branch| Grade |
+-----+-----+-----+-----+-----+
| 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
```

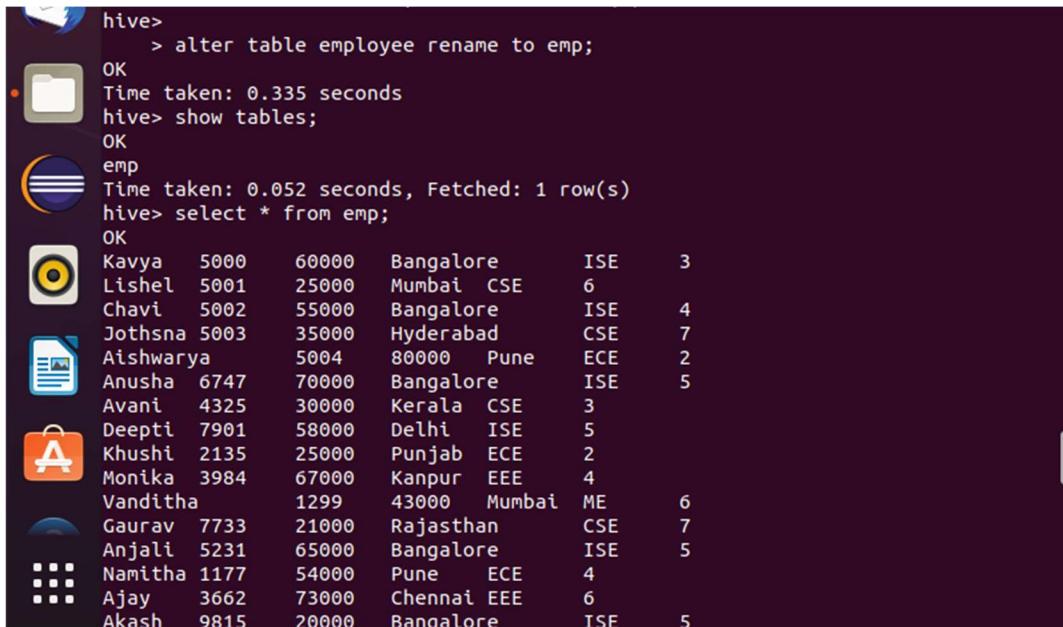
```
emp
```

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

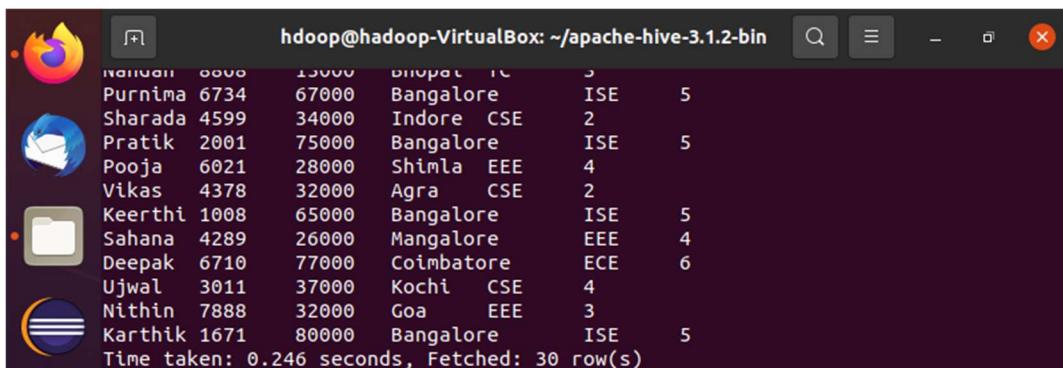
```
hive> select * from emp;
```

```
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 ISF 5
```



```
hadoop@hadoop-VirtualBox: ~/apache-hive-3.1.2-bin
hive> alter table emp change dname Dept_name string;
OK
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
```

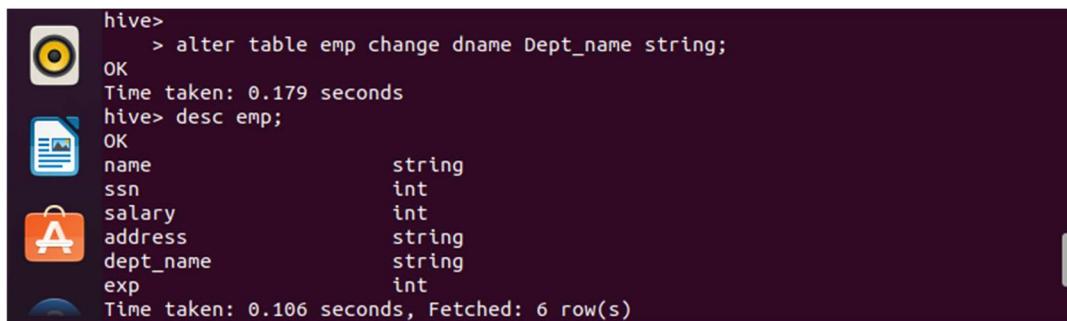
```
name      string
```

```
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)

```
hive> select * from emp_dept;
```

OK

Output screenshots :

```
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)
hive> select * from emp_dept;
OK
Kavya    ISE
Lishel   CSE
Chavi    ISE
Jothsna  CSE
Aishwarya      ECE
Anusha   ISE
Avani    CSE
Deepti   ISE
Khushi   ECE
Monika   EEE
Vanditha      ME
Gaurav   CSE
Anjali   ISE
Namitha  ECE
.....  EEE
Time taken: 0.233 seconds, Fetched: 30 row(s)
hive>
```

```
hadoop@hadoop-VirtualBox: ~/apache-hive-3.1.2-bin
Jothsna  CSE
Aishwarya      ECE
Anusha   ISE
Avani    CSE
Deepti   ISE
Khushi   ECE
Monika   EEE
Vanditha      ME
Gaurav   CSE
Anjali   ISE
Namitha  ECE
Ajay     EEE
Akash    ISE
Darshan  CSE
Anish    AE
Nandan   TC
Purnima  ISE
Sharada  CSE
Pratik   ISE
Pooja    EEE
Vikas    CSE
Keerthi  ISE
Sahana   EEE
Deepak   ECE
Ujwal    CSE
Nithin   EEE
Karthik  ISE
Time taken: 0.233 seconds, Fetched: 30 row(s)
hive>
```

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

```
hive> select name,ssn from emp group by ssn,name order by name;
```

Output screenshots :

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 H
    DFS 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
  ...

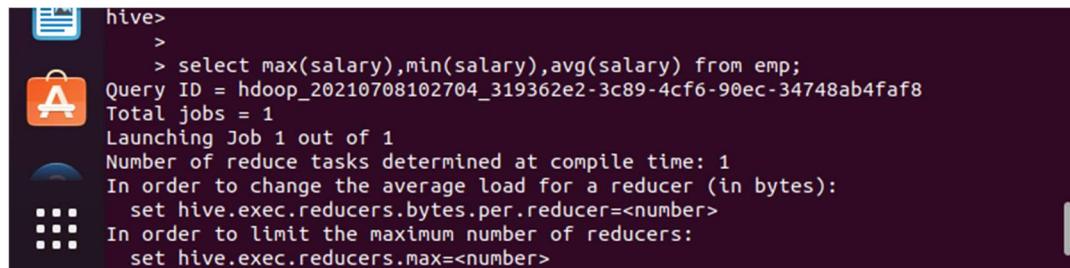
```

```
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)
```

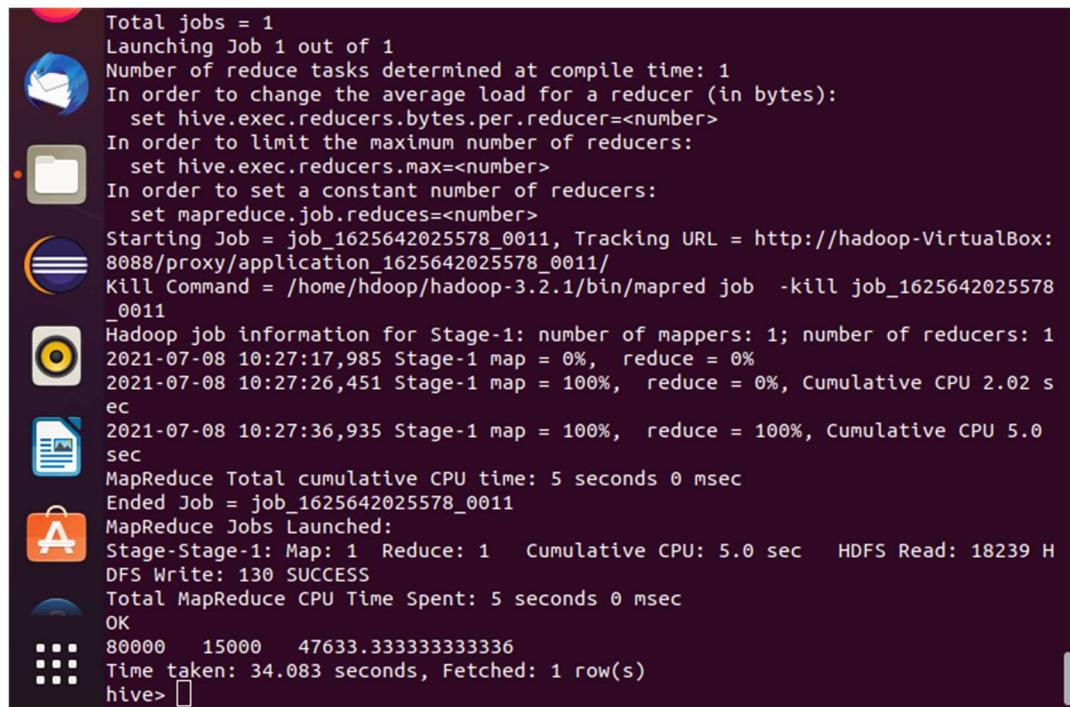
7. Retrieve Maximum salary, minimum salary and Average salary of the employees

```
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/hdoop/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
```

```
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
5       EEE
5       EEE
5       EEE
5       EEE
6       ISE
3       ISE
1       ISE
6       ISE
3       ISE
1       ISE
6       ISE
```

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
```

```
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
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_20210709132357_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
```

```
hadoop@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/hadoop/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>