

Nitte Meenakshi Institute of Technology
Yelahanka, Bangalore

BIGDATA LAB
Learning Activity – 2

Submitted by-

Bhyravi C

1NT18IS041

6'A' (A2 Batch)

A Brief about 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, faulttolerant 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 filesystem. 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.

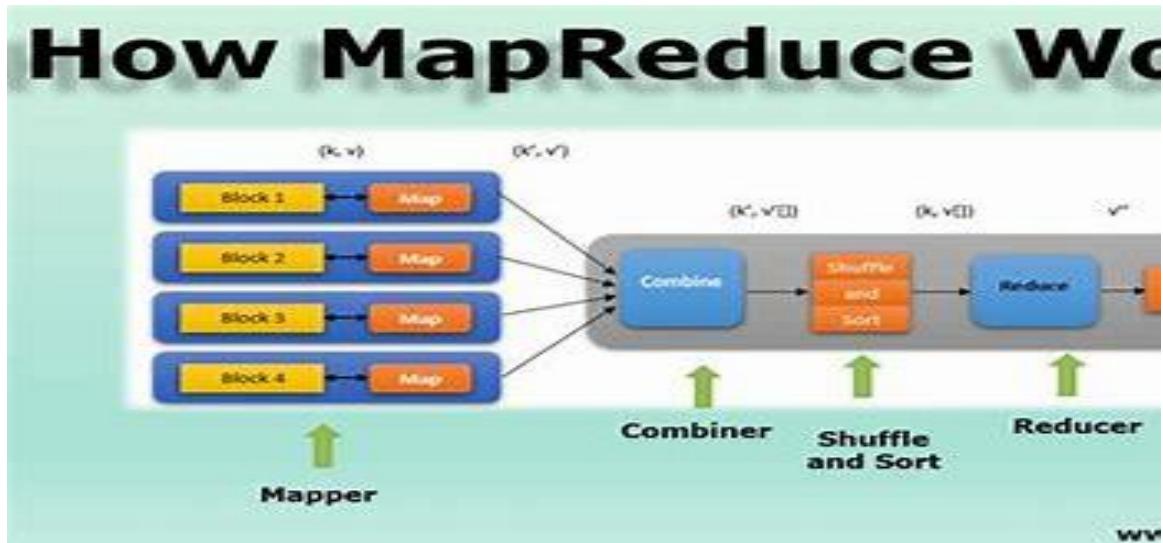
Inputs and Outputs

The MapReduce framework operates exclusively on pairs, that is, the framework views the input to the job as a set of pairs and produces a set of pairs as the output of the job, conceivably of different types.

The key and value classes have to be serializable by the framework and hence need to implement the Writable interface. Additionally, the key classes have to implement the WritableComparable interface to facilitate sorting by the framework.

Input and Output types of a MapReduce job:

(input) -> map -> -> combine -> -> reduce -> (output)



GitHub link:

https://github.com/1nt18is041-bhyravi/Int18is041_Bhyravi_BDlab_LA2

Exercise-I

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

Name SSN Salary Address Dname Experience

Harsha 5000 30000 Bangalore ISE 5

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

1. Total number of employees who 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(.csv)

Activities Firefox Web Browser Fri 07:13

Untitled spreadsheet - G x + https://docs.google.com/spreadsheets/d/1IzgsEycBho94sco-C-z1YHBQXhHj7zl5tEcDjv-7HQ/edit#gid=0 110% Share B

Untitled spreadsheet File Edit View Insert Format Data Tools Add-ons Help Last edit was 11 minutes ago

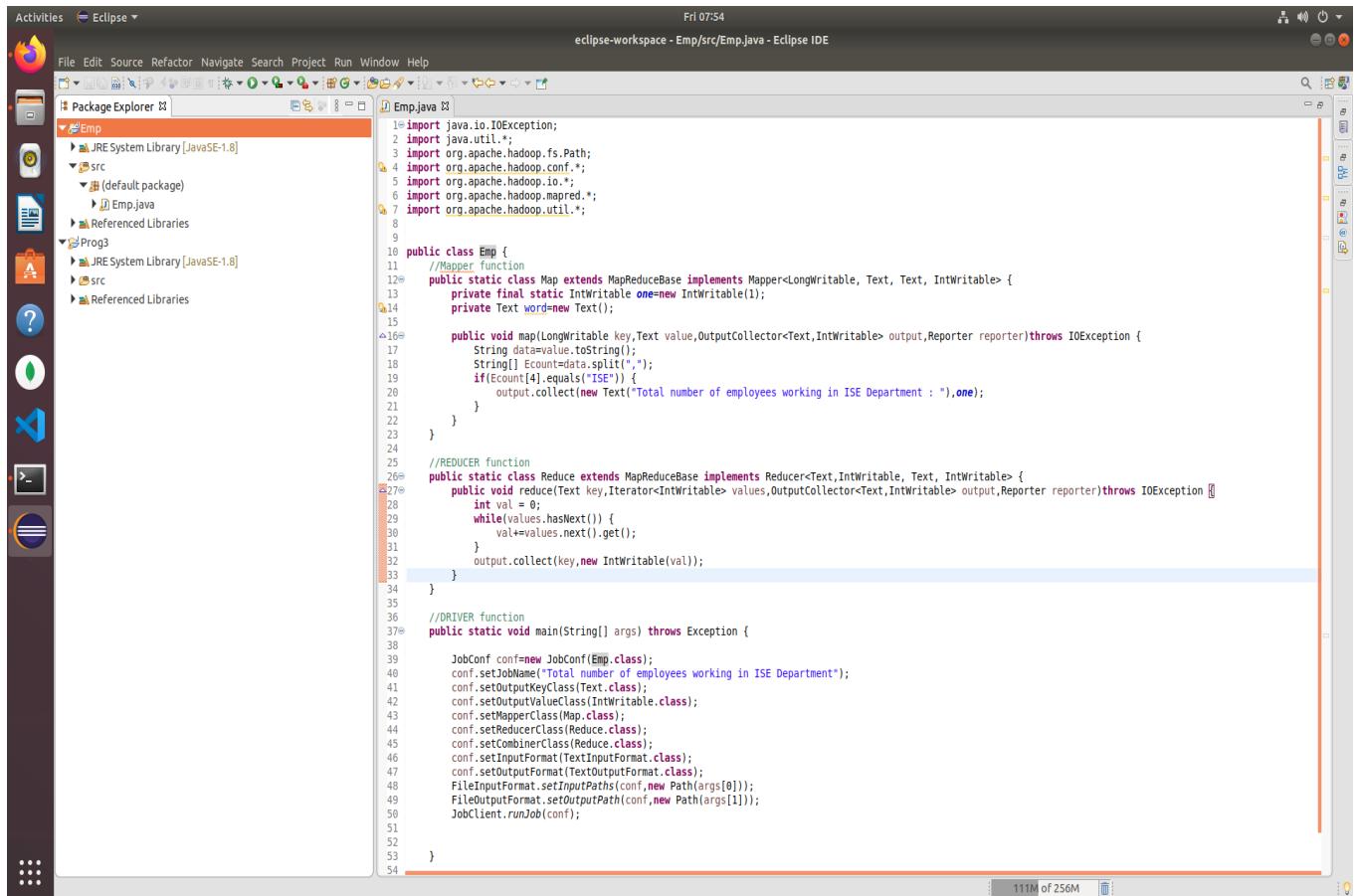
L10 A B C D E F G H I J K L M N O P

1	Harsha	5000	30000	Bangalore	ISE	5
2	Bhyravi	5001	35000	Mysore	CSE	6
3	Sadhana	5002	40000	Dharwad	EEE	5
4	Vishnupriya	5003	45000	Kerla	CVE	3
5	Kavya	5004	50000	Mangalore	ISE	2
6	Bhargavi	5005	55000	Bangalore	ISE	1
7	Anusha	5006	60000	Mumbai	ECE	6
8	Ajay	5007	65000	Yellapur	ISE	5
9	Bharath	5008	70000	Bangalore	CSE	7
10	Priya	5009	75000	Mandy	EEE	3
11	Nikitha	5010	35000	Bangalore	ISE	4
12	Keerthi	5011	60000	Kerala	CVE	5
13	Jeevan	5012	75000	Mumbai	EEE	1
14	Monika	5013	60000	Mandy	ISE	5
15	Prathap	5014	70000	Hassan	CSE	5
16	Dashmini	5015	50000	Jharkand	ECE	7
17	Pavan	5016	65000	Bangalore	CVE	3
18	Ajay	5017	70000	Kerala	CSE	2
19	Asha	5018	80000	Bangalore	ISE	4
20	Anjali	5019	60000	Jharkand	CSE	5
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						

Result and Snapshot of hadoop program

1. Total number of employees who work in ISE department

Java Program Snapshot:



The screenshot shows the Eclipse IDE interface with the following details:

- Title Bar:** Activities - Eclipse - Fri 07:54 - eclipse-workspace - Emp/src/Emp.java - Eclipse IDE
- Left Sidebar (Activities):** Shows various application icons.
- Top Menu Bar:** File, Edit, Source, Refactor, Navigate, Search, Project, Run, Window, Help.
- Package Explorer:** Shows the project structure:
 - Emp (selected)
 - JRE System Library [JavaSE-1.8]
 - src
 - (default package)
 - Emp.java
 - Referenced Libraries
- Right Sidebar:** Shows various tool icons.
- Central Editor Area:** Displays the Java code for `Emp.java`. The code implements a MapReduce program to count employees working in the ISE department.

```
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
10 public class Emp {
11     //Mapper function
12     public static class Map extends MapReduceBase implements Mapper<LongWritable, Text, Text, IntWritable> {
13         private final static IntWritable one=new IntWritable(1);
14         private Text word=new Text();
15
16         public void map(LongWritable key,Text value,OutputCollector<Text,IntWritable> output,Reporter reporter) throws IOException {
17             String data=value.toString();
18             String[] Ecount=data.split(",");
19             if(Ecount[4].equals("ISE")){
20                 output.collect(new Text("Total number of employees working in ISE Department : "),one);
21             }
22         }
23     }
24
25     //REDUCER function
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,Reporter reporter) throws IOException {
28             int val = 0;
29             while(values.hasNext()){
30                 val+=values.next().get();
31             }
32             output.collect(key,new IntWritable(val));
33         }
34     }
35
36     //DRIVER function
37     public static void main(String[] args) throws Exception {
38
39         JobConf conf=new JobConf(Emp.class);
40         conf.setJobName("Total number of employees working in ISE Department");
41         conf.setOutputKeyClass(Text.class);
42         conf.setOutputValueClass(IntWritable.class);
43         conf.setMapperClass(Map.class);
44         conf.setReducerClass(Reduce.class);
45         conf.setCombinerClass(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 }
```

MapReducing through jar file:

```

Activities Terminal Fri 09:06
hadoop@ubuntu:~/eclipse-workspace

hdop@ubuntu:~$ 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 [ubuntu]
Starting resourcemanager
Starting nodemanagers
hdop@ubuntu:~$ jps
2021 3338 Jps
2021 3292 SecondaryNameNode
2021 3168 NodeManager
2021 2312 DataNode
2021 2473 TaskTracker
2021 2473 DataNode
hdop@ubuntu:~$ cd eclipse-workspace/
hdop@ubuntu:~/eclipse-workspace$ ls
derby.log en.jar Emp hs_err_pid4063.log input1.txt input2.txt input.txt metastore_db Prog3 prog3.jar Program3 program3.jar program.jar Programs3
hdop@ubuntu:~/eclipse-workspace$ hadoop fs -ls
hdop@ubuntu:~/eclipse-workspace$ hadoop fs -ls
drwxr-xr-x 1 hdoop supergroup 0 2021-07-09 07:59 emp.csv
drwxr-xr-x 1 hdoop supergroup 0 2021-07-09 08:02 emp1.txt
drwxr-xr-x 1 hdoop supergroup 0 2021-07-09 18:33 empty.txt
drwxr-xr-x 1 hdoop supergroup 37 2021-07-09 18:36 input.txt
drwxr-xr-x 1 hdoop supergroup 46 2021-07-09 08:06 input1.txt
drwxr-xr-x 1 hdoop supergroup 33 2021-07-09 23:05 input2.txt
drwxr-xr-x 1 hdoop supergroup 0 2021-07-09 23:37 out
drwxr-xr-x 1 hdoop supergroup 0 2021-07-09 23:41 out1.txt
drwxr-xr-x 1 hdoop supergroup 0 2021-07-09 01:13 out2
drwxr-xr-x 1 hdoop supergroup 0 2021-07-09 00:20 output1
hdop@ubuntu:~/eclipse-workspaces$ hadoop jar en.jar Emp emp.csv ou.txt
2021-07-09 09:03:07,181 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-09 09:03:07,722 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-09 09:03:08,345 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool Interface and execute your application with ToolRunner to remedy this.
2021-07-09 09:03:08,427 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path /tmp/hadoop-yarn/staging/job_1625846505558_0001
2021-07-09 09:03:09,896 INFO mapred.FileInputFormat: Total input files to process : 1
2021-07-09 09:03:10,089 INFO mapred.FileInputFormat: Input split 0 has 1 partitions
2021-07-09 09:03:10,089 INFO mapred.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-09 09:03:10,125 INFO mapreduce.JobSubmitter: number of splits:2
2021-07-09 09:03:10,711 INFO mapred.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-09 09:03:10,802 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1625846505558_0001
2021-07-09 09:03:10,802 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-07-09 09:03:11,543 INFO conf.Configuration: resource-types.xml not found
2021-07-09 09:03:11,544 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-07-09 09:03:12,771 INFO mapred.YarnClientImpl: Submitted application application_1625846505558_0001
2021-07-09 09:03:13,071 INFO mapreduce.Job: The url to track the job: http://ubuntu:8088/proxy/application_1625846505558_0001/
2021-07-09 09:03:13,072 INFO mapreduce.Job: map 0% reduce 0%
2021-07-09 09:03:32,881 INFO mapreduce.Job: Job job_1625846505558_0001 running in uber mode : false
2021-07-09 09:03:32,804 INFO mapreduce.Job: map 0% reduce 0%
2021-07-09 09:03:59,172 INFO mapreduce.Job: map 100% reduce 100%
2021-07-09 09:03:58,263 INFO mapreduce.Job: map 100% reduce 100%
2021-07-09 09:03:59,302 INFO mapreduce.Job: Job job_1625846505558_0001 completed successfully
2021-07-09 09:03:59,512 INFO mapreduce.Job: Counters: 54
          Filesystem Counters
```

OUTPUT

```

Activities Terminal Fri 09:07
hadoop@ubuntu:~/eclipse-workspace

hdop@ubuntu:~$ hadoop fs -?
[-createSnapshot <snapshotDir> [<snapshotName>]
[-deleteSnapshot <snapshotDir> [<snapshotName>]
[-df [-h] [<path> ...]
[-du [-s] [<s>] [<v>] [<x>] <path> ...]
[-exunge [-t]immediate]
[-find <path> [<expression> ...]
[-get [-f] [<p>] [-ignorecrc] [<crc>] <src> ... <localdst>]
[-getfacl [-R] <path>]
[-getfattr [-R] [<name> | -d] [-e en] <path>]
[-getmerge [-n] [<n>] [-skipEmptyFile] <src> <localdst>
[-head <file>]
[-help [cmd ...]]
[-ls [-c] [<d>] [-h] [-q] [-R] [-t] [-S] [-r] [-u] [-e] [<path> ...]
[-mkdir [-p] <path>]
[-moveFromLocal <localsrc> ... <dst>]
[-moveToLocal <src> <localdst>]
[-mv <src> ... <dst>]
[-put [-f] [-l] [<d>] <localsrc> ... <dst>]
[-renamesnapshot <snapshotDir> <oldName> <newName>]
[-rm [-f] [<r>R] [-skipTrash] [<safely>] <src> ...
[-rmdir [--ignore-fail-on-non-empty] <dir> ...]
[-setfacl [-R] [<b>K] [<m>] <acl_spec> <path>][[-set <acl_spec> <path>]]
[-setfattr [<n> [<v> value] | -x name] <path>]
[-setrep [-R] [<w>] <rep> <path> ...]
[-stat [format] <path> ...]
[-tail [-f] [<s> sleep interval] <file>]
[-test [-defwzr] <path>]
[-text [-ignorecrc] <src> ...]
[-touch [-a] [-m] [-t TIMESTAMP] [<c>] <path> ...]
[-touch <path> ...]
[-truncate [-w] <length> <path> ...]
[-usage [cmd ...]]]

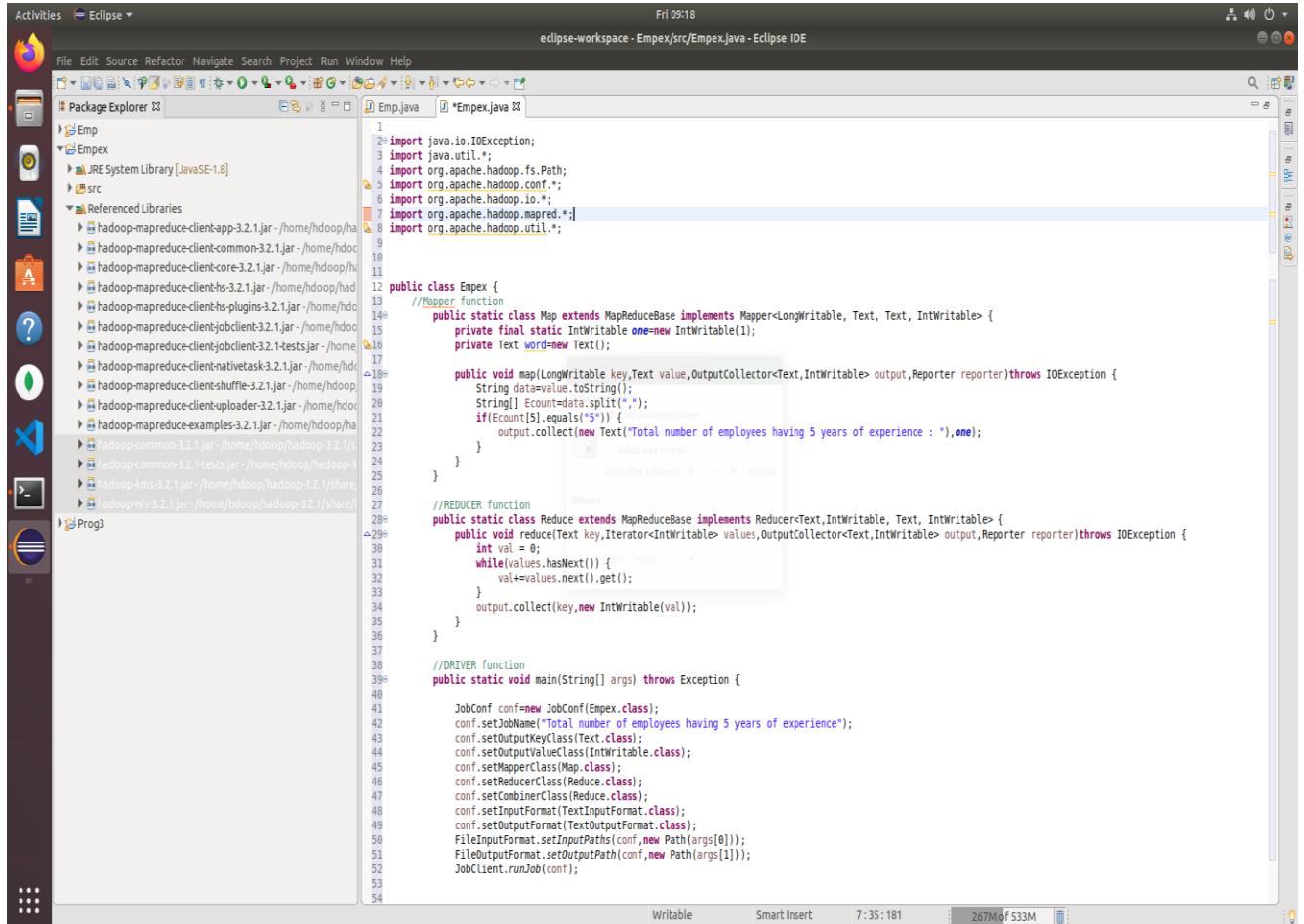
Generic options supported are:
-conf <configuration file> specify an application configuration file
-D <property>=<value> define a value for a given property
-fs <file:///|hdfs://>namenode:>port> specify default filesystem URL to use, overrides 'fs.defaultFS' property from configurations.
-jt <local>resourcemanager:>port> specify a ResourceManager
-files <file1,>...> specify a comma-separated list of files to be copied to the map reduce cluster
-libjars <jar1,>...> specify a comma-separated list of jar files to be included in the classpath
-archives <archive1,>...> specify a comma-separated list of archives to be unarchived on the compute machines

The general command line syntax is:
command [genericOptions] [commandOptions]

hdop@ubuntu:~/eclipse-workspace$ hadoop fs -ls ou.txt
Found 2 items
-rw-r--r-- 1 hdoop supergroup 0 2021-07-09 09:03 ou.txt/_SUCCESS
-rw-r--r-- 1 hdoop supergroup 57 2021-07-09 09:03 ou.txt/part-00000
hdop@ubuntu:~/eclipse-workspace$ hadoop fs -ls ou.txt/part-00000
-rw-r--r-- 1 hdoop supergroup 57 2021-07-09 09:03 ou.txt/part-00000
hdop@ubuntu:~/eclipse-workspace$ hadoop fs -cat ou.txt/part-00000
2021-07-09 09:03:27,984 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
Total number of employees working in ISE Department : 7
hdop@ubuntu:~/eclipse-workspace$
```

2. Total number of employees with experience=5 years

Java Program Snapshot:



The screenshot shows the Eclipse IDE interface with the following details:

- Title Bar:** Activities - Eclipse - Fri 09:18 - eclipse-workspace - Empex/src/Empex.java - Eclipse IDE
- Left Sidebar (Package Explorer):** Shows the project structure with packages like Emp and Empex, and referenced libraries including JRE System Library [JavaSE-1.8] and various Hadoop jars.
- Central Area (Empex.java):** Displays the Java code for the Empex class. The code implements a MapReduce job to count employees with 5 years of experience. It includes imports for java.util, org.apache.hadoop.conf, org.apache.hadoop.io, org.apache.hadoop.mapred, and org.apache.hadoop.util.
- Code Snippet:** The code defines the Mapper function to map each line of input to a key-value pair where the value is 1 if it contains "5" in the word field. The Reducer function then sums up these values to get the total count.

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

public class Empex {
    //Mapper function
    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 number of employees having 5 years of experience : "),one);
            }
        }
    }
    //REDUCER function
    public static class Reduce extends MapReduceBase implements Reducer<Text,Iterator<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 function
    public static void main(String[] args) throws Exception {
        JobConf conf=new JobConf(Empex.class);
        conf.setJobName("Total number of employees having 5 years of experience");
        conf.setOutputKeyClass(Text.class);
        conf.setOutputValueClass(IntWritable.class);
        conf.setMapperClass(Map.class);
        conf.setReducerClass(Reduce.class);
        conf.setCombinerClass(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);
    }
}
```

- Bottom Status Bar:** Writable, Smart Insert, 7:35:181, 267M of 533M

MapReducing through jar file:

```
Fri 09:19
hadoop@ubuntu:~$ 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.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [ubuntu]
Starting resourcemanager
Starting nodemanagers
hadoop@ubuntu:~$ jps
5585 NameNode
6801 SecondaryNameNode
6801 Jps
6435 NodeManager
6265 ResourceManager
5757 DataNode
hadoop@ubuntu:~$ eclipse
org.eclipse.m2e.logback.configuration: The org.eclipse.m2e.logback.configuration bundle was activated before the state location was initialized. Will retry after the state location is initialized.
org.eclipse.m2e.logback.configuration: Logback config file: /home/hadoop/eclipse-workspace/.metadata/.plugins/org.eclipse.m2e.logback.configuration/logback.1.14.0.20191209-1925.xml
org.eclipse.m2e.logback.configuration: Initializing logback
(Eclipse:6847): Gtk-WARNING **: 09:13:59.392: Negative content width -4 (allocation 2, extents 3x3) while allocating gadget (node progressbar, owner GtkProgressBar)
(Eclipse:6847): Gtk-WARNING **: 09:13:59.393: Negative content width -2 (allocation 0, extents 1x1) while allocating gadget (node trough, owner GtkProgressBar)
(Eclipse:6847): Gtk-WARNING **: 09:13:59.396: Negative content width -4 (allocation 2, extents 3x3) while allocating gadget (node progressbar, owner GtkProgressBar)
(Eclipse:6847): Gtk-WARNING **: 09:13:59.397: Negative content width -2 (allocation 0, extents 1x1) while allocating gadget (node trough, owner GtkProgressBar)
(Eclipse:6847): Gtk-WARNING **: 09:13:59.397: Negative content width -4 (allocation 2, extents 3x3) while allocating gadget (node progressbar, owner GtkProgressBar)
(Eclipse:6847): Gtk-WARNING **: 09:13:59.400: Negative content width -2 (allocation 0, extents 1x1) while allocating gadget (node trough, owner GtkProgressBar)
(Eclipse:6847): Gtk-WARNING **: 09:13:59.403: Negative content width -4 (allocation 2, extents 3x3) while allocating gadget (node progressbar, owner GtkProgressBar)
(Eclipse:6847): Gtk-WARNING **: 09:13:59.403: Negative content width -2 (allocation 0, extents 1x1) while allocating gadget (node trough, owner GtkProgressBar)
hadoop@ubuntu:~$
```

```
Fri 09:26
hadoop@ubuntu:~/eclipse-workspace
File Edit View Search Terminal Help
hadoop@ubuntu:~$ 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.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [ubuntu]
Starting resourcemanager
Starting nodemanagers
hadoop@ubuntu:~$ jps
5585 NameNode
6801 SecondaryNameNode
6801 Jps
6435 NodeManager
6265 ResourceManager
5757 DataNode
hadoop@ubuntu:~$ eclipse
org.eclipse.m2e.logback.configuration: The org.eclipse.m2e.logback.configuration bundle was activated before the state location was initialized. Will retry after the state location is initialized.
org.eclipse.m2e.logback.configuration: Logback config file: /home/hadoop/eclipse-workspace/.metadata/.plugins/org.eclipse.m2e.logback.configuration/logback.1.14.0.20191209-1925.xml
org.eclipse.m2e.logback.configuration: Initializing logback
(Eclipse:6847): Gtk-WARNING **: 09:13:59.392: Negative content width -4 (allocation 2, extents 3x3) while allocating gadget (node progressbar, owner GtkProgressBar)
(Eclipse:6847): Gtk-WARNING **: 09:13:59.393: Negative content width -2 (allocation 0, extents 1x1) while allocating gadget (node trough, owner GtkProgressBar)
(Eclipse:6847): Gtk-WARNING **: 09:13:59.396: Negative content width -4 (allocation 2, extents 3x3) while allocating gadget (node progressbar, owner GtkProgressBar)
(Eclipse:6847): Gtk-WARNING **: 09:13:59.397: Negative content width -2 (allocation 0, extents 1x1) while allocating gadget (node trough, owner GtkProgressBar)
(Eclipse:6847): Gtk-WARNING **: 09:13:59.397: Negative content width -4 (allocation 2, extents 3x3) while allocating gadget (node progressbar, owner GtkProgressBar)
(Eclipse:6847): Gtk-WARNING **: 09:13:59.400: Negative content width -2 (allocation 0, extents 1x1) while allocating gadget (node trough, owner GtkProgressBar)
(Eclipse:6847): Gtk-WARNING **: 09:13:59.403: Negative content width -4 (allocation 2, extents 3x3) while allocating gadget (node progressbar, owner GtkProgressBar)
(Eclipse:6847): Gtk-WARNING **: 09:13:59.403: Negative content width -2 (allocation 0, extents 1x1) while allocating gadget (node trough, owner GtkProgressBar)
hadoop@ubuntu:~$ clear
hadoop@ubuntu:~/eclipse-workspace$ ls
derby.log em2.jar enj.jar Emp Empe hs_err_pid4063.log input1.txt input2.txt input.txt metastore_db Prog3 prog3.jar Program3 program3.jar program.jar Programs3
hadoop@ubuntu:~/eclipse-workspace$ hadoop fs -ls
Found 11 items
-rw-r--r-- 1 hdoop supergroup 671 2021-07-09 07:59 emp.csv
drwxr-xr-x - hdoop supergroup 0 2021-07-09 08:02 empl.txt
-rw-r--r-- 1 hdoop supergroup 0 2021-06-08 18:33 empty.txt
-rw-r--r-- 1 hdoop supergroup 37 2021-06-08 18:36 input.txt
-rw-r--r-- 1 hdoop supergroup 40 2021-06-16 00:06 input1.txt
-rw-r--r-- 1 hdoop supergroup 33 2021-06-15 23:05 input2.txt
drwxr-xr-x - hdoop supergroup 0 2021-07-09 09:03 ou.txt
drwxr-xr-x - hdoop supergroup 0 2021-06-15 23:37 out
drwxr-xr-x - hdoop supergroup 0 2021-07-09 08:30 out1.txt
drwxr-xr-x - hdoop supergroup 0 2021-06-16 01:13 out2
drwxr-xr-x - hdoop supergroup 0 2021-06-16 00:28 output1
hadoop@ubuntu:~/eclipse-workspace$ hadoop fs -copyFromLocal em2.csv
copyFromLocal: src=em2.csv, dst=/user/hadoop/emp.csv failed: java.io.IOException
```

```
Activities Terminal Fri 09:27
hadoop@ubuntu: ~/eclipse-workspace
File Edit View Search Terminal Help
hadoop@ubuntu:~/eclipse-workspace$ hadoop fs -copyFromLocal em2.csv
copyFromLocal: 'em2.csv': No such file or directory
hadoop@ubuntu:~/eclipse-workspace$ hadoop jar em2.jar Emplex emp.csv em.txt
hadoop@ubuntu:~/eclipse-workspace$ hadoop fs -copyFromLocal /home/hadoop/Downloads/emp.csv
copyFromLocal: 'emp.csv': File exists
hadoop@ubuntu:~/eclipse-workspace$ hadoop jar em2.jar Emplex emp.csv em.txt
2021-07-09 09:24:13,120 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-09 09:24:13,689 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-09 09:24:14,265 INFO mapreduce.JobResourceUploader: Disabling feature Coding for path: /tmp/hadoop-yarn/staging/hadoop/_staging/job_1625847192845_0001
2021-07-09 09:24:14,999 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-09 09:24:14,999 INFO mapred.FileInputFormat: Total input files to process : 1
2021-07-09 09:24:15,069 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-09 09:24:15,133 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-09 09:24:15,474 INFO mapred.JobSubmitter: number of splits:2
2021-07-09 09:24:15,581 INFO mapred.JobSubmitter: Submitting tokens for job: job_1625847192845_0001
2021-07-09 09:24:15,581 INFO mapred.JobSubmitter: Executing with tokens: []
2021-07-09 09:24:16,031 INFO conf.Configuration: resource-types.xml not found
2021-07-09 09:24:16,515 INFO resourcemanager.ResourceManager: Submitted application_1625847192845_0001
2021-07-09 09:24:16,809 INFO mapreduce.Job: The url to track the job: http://ubuntu:8088/proxy/application_1625847192845_0001/
2021-07-09 09:24:16,813 INFO mapreduce.Job: Running job: job_1625847192845_0001
2021-07-09 09:24:31,474 INFO mapreduce.Job: Job job_1625847192845_0001 running in uber mode : false
2021-07-09 09:24:31,476 INFO mapreduce.Job: map 0% reduce 0%
2021-07-09 09:24:31,476 INFO mapreduce.Job: map 100% reduce 0%
2021-07-09 09:25:13,888 INFO mapreduce.Job: Job job_1625847192845_0001 completed successfully
2021-07-09 09:25:14,939 INFO mapreduce.Job: Job job_1625847192845_0001 completed successfully
2021-07-09 09:25:15,122 INFO mapreduce.Job: Counters: 54
File System Counter
    FILE: Number of bytes read=134
    FILE: Number of bytes written=877692
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=1191
    HDFS: Number of bytes written=60
    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 Counter
    Launched map tasks=2
    Launched reduce tasks=1
    Data-local map tasks=2
    Total time spent by all maps in occupied slots (ms)=61966
    Total time spent by all reduces in occupied slots (ms)=6289
    Total time spent by all map tasks (ms)=61966
    Total time spent by all reduce tasks (ms)=6289
    Total vcore-milliseconds taken by all map tasks=61966
    Total vcore-milliseconds taken by all reduce tasks=6289
    Total negabyte-milliseconds taken by all map tasks=63453184
    Total negabyte-milliseconds taken by all reduce tasks=6439936
Map-Reduce Framework
    Map input records=20
    Map output records=7
```

OUTPUT

```

Activities Terminal Fri 09:27
hdooop@ubuntu: ~/eclipse-workspace

File Edit View Search Terminal Help
Launched map tasks=2
Launched reduce tasks=1
Data-local map tasks=2
Total time spent by all maps in occupied slots (ms)=61966
Total time spent by all reduces in occupied slots (ms)=6289
Total time spent by all map tasks (ms)=61966
Total time spent by all reduce tasks (ms)=6289
Total vcore-milliseconds taken by all map tasks=61966
Total vcore-milliseconds taken by all reduce tasks=6289
Total megabyte-milliseconds taken by all map tasks=63453184
Total megabyte-milliseconds taken by all reduce tasks=6439936

Map-Reduce Framework
  Map Input records=20
  Map output records=7
  Map output bytes=434
  Map output materialized bytes=140
  Input split bytes=184
  Combine input records=7
  Combine output records=2
  Reduce input groups=1
  Reduce shuffle bytes=140
  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)=2201
  CPU time spent (ms)=35830
  Physical memory (bytes) snapshot=709795840
  Virtual memory (bytes) snapshot=7738068800
  Total committed heap usage (bytes)=543686656
  Peak Map Physical memory (bytes)=287043584
  Peak Map Virtual memory (bytes)=2577231872
  Peak Reduce Physical memory (bytes)=188062976
  Peak Reduce Virtual memory (bytes)=2585169920

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=1007
File Output Format Counters
  Bytes Written=0

hdooop@ubuntu:~/eclipse-workspace$ hadoop fs -ls em.txt
Found 2 items
-rw-r--r-- 1 hdooop supergroup 0 2021-07-09 09:25 em.txt/_SUCCESS
-rw-r--r-- 1 hdooop supergroup 60 2021-07-09 09:25 em.txt/part-00000
hdooop@ubuntu:~/eclipse-workspace$ hadoop fs -cat em.txt/part-00000
2021-07-09 09:26:16,148 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
...
Total number of employees having 5 years of experience :      7
hdooop@ubuntu:~/eclipse-workspace$ 

```

3. Count the number of employees who lives in Bangalore

Java Program Snapshot:

```

Activities Eclipse Fri 09:41
eclipse-workspace - Eclipse IDE

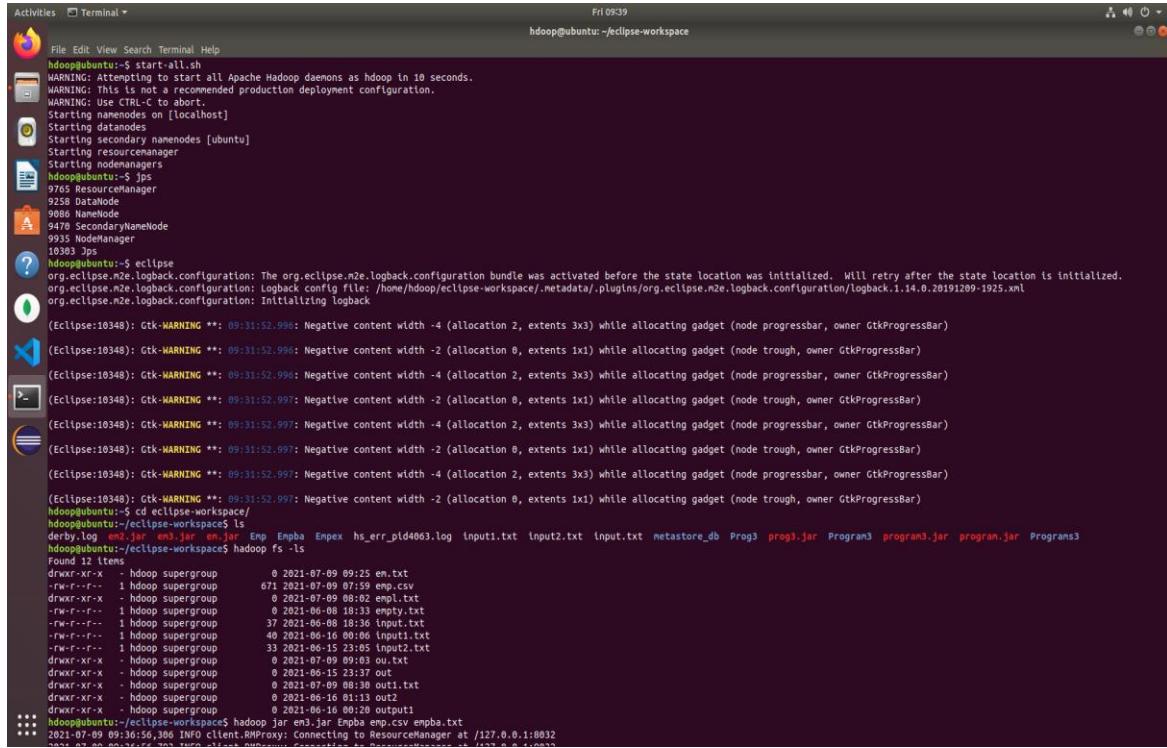
File Edit Source Refactor Search Project Run Window Help

Package Explorer [2]
  + Emp
  + Empba
  + Empex
  + Prog3

  + Import java.io.IOException;[]
  0
  1  public class Empba{
  2    //Mapper Function
  3    public static class Map extends MapReduceBase implements Mapper<LongWritable, Text, Text, IntWritable> {
  4      private final static IntWritable one=new IntWritable(1);
  5      private Text word=new Text();
  6
  7      public void map(LongWritable key,Text value,OutputCollector<Text,IntWritable> output,Reporter reporter) throws IOException {
  8        StringTokenizer str=new StringTokenizer(value.toString());
  9        String[] Ecount=str.nextToken();
 10        if(Ecount[3].equals("Bangalore")){
 11          output.collect(new Text("Total number of employees who lives in 'Bangalore' : "),one);
 12        }
 13      }
 14    }
 15
 16    //REDUCE Function
 17    public static class Reduce extends MapReduceBase implements Reducer<Text,IntWritable,Text,IntWritable>{
 18      public void reduce(Text key,Iterator<IntWritable> values,OutputCollector<Text,IntWritable> output,Reporter reporter) throws IOException {
 19        int val = 0;
 20        while(values.hasNext()){
 21          val+=values.next().get();
 22        }
 23        output.collect(key,new IntWritable(val));
 24      }
 25    }
 26
 27    //DRIVER Function
 28    public static void main(String[] args) throws Exception {
 29
 30      JobConf conf=new JobConf(Empba.class);
 31      conf.setJobName("Total number of employees who lives in Bangalore");
 32      conf.setOutputKeyClass(Text.class);
 33      conf.setOutputValueClass(IntWritable.class);
 34      conf.setMapperClass(Map.class);
 35      conf.setCombinerClass(Reduce.class);
 36      conf.setInputFormat(TextInputFormat.class);
 37      conf.setOutputFormat(TextOutputFormat.class);
 38      FileInputFormat.setInputPaths(conf,new Path(args[0]));
 39      FileOutputFormat.setOutputPath(conf,new Path(args[1]));
 40      JobClient.runJob(conf);
 41
 42    }
 43
 44  }
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57

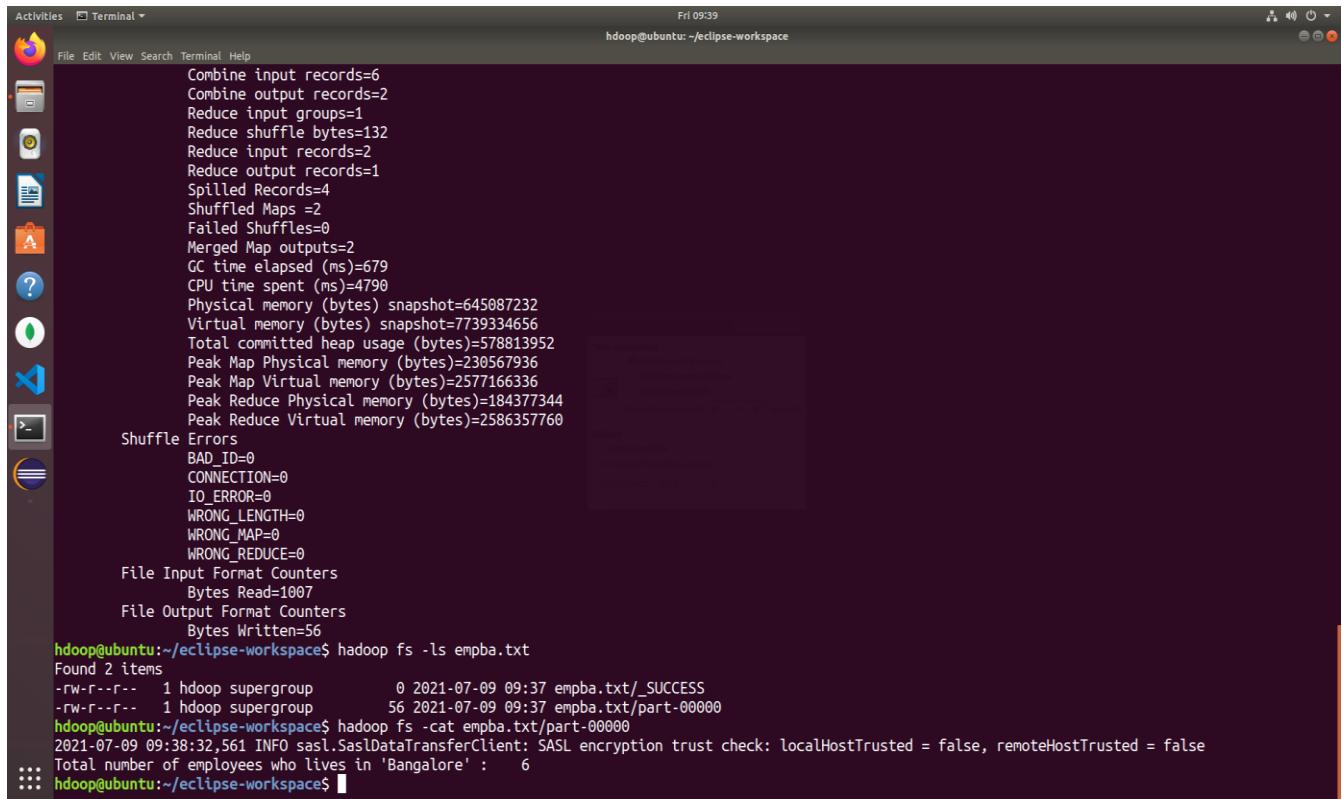
```

MapReducing through jar file:



```
Activities Terminal Fri 09:39
hadoop@ubuntu:~$ 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.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [ubuntu]
Starting resourcemanager
Starting nodemanagers
hadoop@ubuntu:~$ jps
9768 ResourceManager
9769 NodeManager
9800 NameNode
9470 SecondaryNameNode
9935 NodeManager
10383 Jps
?
hadoop@ubuntu:~$ org.eclipse.m2e.logback.configuration: The org.eclipse.m2e.logback.configuration bundle was activated before the state location was initialized. Will retry after the state location is initialized.
org.eclipse.m2e.logback.configuration: Logback config file: /home/hadoop/eclipse-workspace/.metadata/.plugins/org.eclipse.m2e.logback.configuration/logback.i.14.0.20191209-1925.xml
org.eclipse.m2e.logback.configuration: Initializing logback
(Eclipse:10348): Gtk-WARNING **: 09:31:52.998: Negative content width -4 (allocation 2, extents 3x3) while allocating gadget (node progressbar, owner GtkProgressBar)
(Eclipse:10348): Gtk-WARNING **: 09:31:52.998: Negative content width -2 (allocation 0, extents 1x1) while allocating gadget (node trough, owner GtkProgressBar)
(Eclipse:10348): Gtk-WARNING **: 09:31:52.998: Negative content width -4 (allocation 2, extents 3x3) while allocating gadget (node progressbar, owner GtkProgressBar)
(Eclipse:10348): Gtk-WARNING **: 09:31:52.997: Negative content width -2 (allocation 0, extents 1x1) while allocating gadget (node trough, owner GtkProgressBar)
(Eclipse:10348): Gtk-WARNING **: 09:31:52.997: Negative content width -4 (allocation 2, extents 3x3) while allocating gadget (node progressbar, owner GtkProgressBar)
(Eclipse:10348): Gtk-WARNING **: 09:31:52.997: Negative content width -2 (allocation 0, extents 1x1) while allocating gadget (node trough, owner GtkProgressBar)
(Eclipse:10348): Gtk-WARNING **: 09:31:52.997: Negative content width -4 (allocation 2, extents 3x3) while allocating gadget (node progressbar, owner GtkProgressBar)
(Eclipse:10348): Gtk-WARNING **: 09:31:52.997: Negative content width -2 (allocation 0, extents 1x1) while allocating gadget (node trough, owner GtkProgressBar)
hadoop@ubuntu:~$ cd eclipse-workspace/
hadoop@ubuntu:~/eclipse-workspace$ ls
Derby-10.14.2-bin.tar.gz  Emp  Emph  Emplex  hs_err_pid4063.log  input1.txt  input2.txt  input.txt  metastore_db  Prog3  prog3.jar  Program3  program3.jar  program.jar  Programs3
hadoop@ubuntu:~/eclipse-workspace$ hadoop fs -ls
Found 12 items
drwxr-xr-x  - hadoop supergroup          0 2021-07-09 09:25 em.txt
-rw-r--r--  1 hadoop supergroup       671 2021-07-09 07:59 emp.csv
drwxr-xr-x  - hadoop supergroup          0 2021-07-09 08:02 emp1.txt
-rw-r--r--  1 hadoop supergroup      29216 2021-06-15 23:33 empty.txt
drwxr-xr-x  - hadoop supergroup          0 2021-06-15 18:33 emp1
drwxr-xr-x  - hadoop supergroup        46 2021-06-16 00:06 input1.txt
drwxr-xr-x  - hadoop supergroup      33 2021-06-15 23:05 input2.txt
drwxr-xr-x  - hadoop supergroup      0 2021-07-09 09:03 ou.txt
drwxr-xr-x  - hadoop supergroup      0 2021-06-15 23:37 out
drwxr-xr-x  - hadoop supergroup      0 2021-07-09 08:30 out1.txt
drwxr-xr-x  - hadoop supergroup      0 2021-06-16 01:13 out2
drwxr-xr-x  - hadoop supergroup      0 2021-06-16 00:20 output1
hadoop@ubuntu:~/eclipse-workspace$ hadoop jar em3.jar Empb emp.csv empba.txt
2021-07-09 09:36:56,366 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8083
```

Output:



The screenshot shows a terminal window on a Linux desktop environment. The title bar indicates it's a Terminal window, and the date and time are Fri 09:39. The command hdoop@ubuntu: ~/eclipse-workspace\$ is at the prompt. The output of the hadoop fs -ls command is displayed, showing two files: empba.txt and empba.txt/_SUCCESS. The file empba.txt has a size of 56 bytes and was modified on 2021-07-09 at 09:37. The file empba.txt/_SUCCESS has a size of 0 bytes and was modified on 2021-07-09 at 09:37. Below this, the command hadoop fs -cat empba.txt/part-00000 is run, followed by a summary line: "Total number of employees who lives in 'Bangalore' : 6". The terminal window also shows various system statistics and error counts related to Hadoop's memory usage and shuffle errors.

```
Combine input records=6
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)=679
CPU time spent (ms)=4790
Physical memory (bytes) snapshot=645087232
Virtual memory (bytes) snapshot=7739334656
Total committed heap usage (bytes)=578813952
Peak Map Physical memory (bytes)=230567936
Peak Map Virtual memory (bytes)=2577166336
Peak Reduce Physical memory (bytes)=184377344
Peak Reduce Virtual memory (bytes)=2586357760
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=1007
File Output Format Counters
Bytes Written=56
hdoop@ubuntu:~/eclipse-workspace$ hadoop fs -ls empba.txt
Found 2 items
-rw-r--r-- 1 hdoop supergroup 0 2021-07-09 09:37 empba.txt/_SUCCESS
-rw-r--r-- 1 hdoop supergroup 56 2021-07-09 09:37 empba.txt/part-00000
hdoop@ubuntu:~/eclipse-workspace$ hadoop fs -cat empba.txt/part-00000
2021-07-09 09:38:32,561 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
Total number of employees who lives in 'Bangalore' : 6
hdoop@ubuntu:~/eclipse-workspace$
```

A Brief about 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-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 Mapreduce operations,

1. Insert 5 records using INSERT command.
2. Demonstrate the Alter command for the following cases,
 - a. Rename the table name to “Emp”.
 - b. Rename the column name “Dname” to “Dept_name”.
3. Retrieve all the employees who’s salary is not less than 50000.
4. Extract all employees who live in Bangalore but having less than 5 years of experience
5. Create separate view containing Name, Dept_name of employees
6. Display Name and SSN and use group by SSN and order by Name
7. Retrieve Maximum salary, minimum salary and Average salary of the employees
8. Create Another table called Department with the following fields (Dname = Dept_name and perform the following joins(outer, left outer, right outer) over Dname

Dno Dname

6 ISE

Result and Snapshot of Hive Querie

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Harsha	5000	30000	Bangalore	ISE	5										
2	Bhyravi	5001	35000	Mysore	CSE	6										
3	Sadhana	5002	40000	Dhanwad	EEE	5										
4	Vishnupriya	5003	45000	Kerala	CVE	3										
5	Kavya	5004	50000	Mangalore	ISE	2										
6	Bhargavi	5005	55000	Bangalore	ISE	1										
7	Anusha	5006	60000	Mumbai	ECE	6										
8	Ajay	5007	65000	Yellapur	ISE	5										
9	Bharath	5008	70000	Bangalore	CSE	7										
10	Priya	5009	75000	Mandy	EEE	3										
11	Nikitha	5010	35000	Bangalore	ISE	4										
12	Keerthi	5011	60000	Kerala	CVE	5										
13	Jeevan	5012	75000	Mumbai	EEE	1										
14	Monika	5013	60000	Mandy	ISE	5										
15	Prathap	5014	70000	Hassan	CSE	5										
16	Darsmini	5015	50000	Jharkand	ECE	7										
17	Pavan	5016	65000	Bangalore	CVE	3										
18	Ajay	5017	70000	Kerala	CSE	2										
19	Asha	5018	80000	Bangalore	ISE	4										
20	Anjali	5019	60000	Jharkand	CSE	5										
21																
22																
23																
24																
25																
26																
27																
28																
29																
30																
31																

Creating database EmployeeDB and creating table Employee:

```
File Edit View Search Terminal Help Fri 09:50 hdoop@ubuntu: ~/apache-hive-3.1.2-bin/bin
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:498)
at org.apache.hadoop.util.RunJar.run(RunJar.java:323)
at org.apache.hadoop.util.RunJar.main(RunJar.java:236)
FAILED: ParseException line 1:5 cannot recognize input near 'show' 'database' '<EOF>' in ddl statement
hive> show databases;
OK
bank
customer
default
Time taken: 0.432 seconds, Fetched: 3 row(s)
hive> create database EmployeeDB;
OK
Time taken: 0.476 seconds
hive> use EmployeeDB;
OK
Time taken: 0.065 seconds
hive> create table employee(name string,ssn int,salary float,address string,dname string,experience int)row format delimited fields terminated by ",";
OK
Time taken: 1.713 seconds
hive> desc employee;
OK
name          string
ssn           int
salary         float
address        string
dname          string
experience     int
Time taken: 0.854 seconds, Fetched: 6 row(s)
hive>
```

Importing csv file:

```
Activities Terminal Fri 09:52
File Edit View Search Terminal Help hdoop@ubuntu: ~/apache-hive-3.1.2-bin/bin
address          string
dname           string
experience      int
Time taken: 0.854 seconds, Fetched: 6 row(s)
hive> load data local inpath '/home/hdoop/Downloads/emp.csv' into table employee;
Loading data to table employeedb.employee
OK
hive> select* from employee;
OK
Harsha 5000 30000.0 Bangalore ISE 5
Bhyravi 5001 35000.0 Mysore CSE 6
Sadhana 5002 40000.0 Dharwad EEE 5
Vishnupriya 5003 45000.0 Kerla CVE 3
Kavya 5004 50000.0 Mangalore ISE 2
Bhargavi 5005 55000.0 Bangalore ISE 1
Anusha 5006 60000.0 Mumbai ECE 6
Ajay 5007 65000.0 Yellapur ISE 5
Bharath 5008 70000.0 Bangalore CSE 7
Priya 5009 75000.0 Mandya EEE 3
Nikitha 5010 35000.0 Bangalore ISE 4
Keerthi 5011 60000.0 Kerala CVE 5
Jeevan 5012 75000.0 Mumbai EEE 1
Monika 5013 60000.0 Mandya ISE 5
Prathap 5014 70000.0 Hassan CSE 5
Darshini 5015 50000.0 Jharkand ECE 7
Pavan 5016 65000.0 Bangalore CVE 3
Ajay 5017 70000.0 Kerala CSE 2
Asha 5018 80000.0 Bangalore ISE 4
Anjali 5019 60000.0 Jharkand CSE 5
Time taken: 4.711 seconds, Fetched: 20 row(s)
hive>
```

1. Insert 5 records using INSERT command.

```
Activities Terminal Fri 10:00
File Edit View Search Terminal Help hdoop@ubuntu: ~/apache-hive-3.1.2-bin/bin
Prathap 5014 70000.0 Hassan CSE 5
Darshini 5015 50000.0 Jharkand ECE 7
Pavan 5016 65000.0 Bangalore CVE 3
Ajay 5017 70000.0 Kerala CSE 2
Asha 5018 80000.0 Bangalore ISE 4
Anjali 5019 60000.0 Jharkand CSE 5
Time taken: 4.711 seconds, Fetched: 20 row(s)
hive> insert into employee values("Akash",5020,35000,"goa","ISE",5),("Abhay",5021,45000,"bangalore","CSE",6),("Arjun",5022,55000,"Bidar","CVE",3),("Ananya",5023,85000,"Blajapur","EEE",6),("Ani",5024,60000,"Kerala","ECE",5);
Query ID = hdoop_20210709095835_1480d0db-24ef-4b09-a57c-78ace4f422d4
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_1625849013163_0001, Tracking URL = http://ubuntu:8088/proxy/application_1625849013163_0001/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job -kill job_1625849013163_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-09 09:59:00,462 Stage-1 map = 0%, reduce = 0%
2021-07-09 09:59:14,787 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 8.27 sec
2021-07-09 09:59:28,010 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 12.98 sec
MapReduce Total cumulative CPU time: 12 seconds 980 msec
Ended Job = job_1625849013163_0001
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://127.0.0.1:9000/user/hive/warehouse/employeedb.db/employee/.hive-staging_hive_2021-07-09_09-58-35_775_3798681951976185312-1/-ext-10000
Loading data to table employeedb.employee
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 12.98 sec HDFS Read: 21886 HDFS Write: 676 SUCCESS
Total MapReduce CPU Time Spent: 12 seconds 980 msec
OK
Time taken: 54.488 seconds
```

```

Activities Terminal Fri 10:02
hadoop@ubuntu: ~/apache-hive-3.1.2-bin/bin

Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://127.0.0.1:9000/user/hive/warehouse/employeedb.db/employee/.hive-staging_hive_2021-07-09_09-58-35_775_3798681951976185312-1-ext-10000
Loading data to table employeedb.employee
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 12.98 sec HDFS Read: 21886 HDFS Write: 676 SUCCESS
Total MapReduce CPU Time Spent: 12 seconds 980 msec
OK
Time taken: 54.488 seconds
hive> select* from employee;
OK
Akash 5020 35000.0 goa ISE 5
Abhay 5021 45000.0 bangalore CSE 6
Arjun 5022 55000.0 Bidar CVE 3
Ananya 5023 85000.0 Bijapur EEE 6
Ani 5024 60000.0 Kerala ECE 5
Harsha 5000 30000.0 Bangalore ISE 5
Bhyravi 5001 35000.0 Mysore CSE 6
Sadhana 5002 40000.0 Dharwad EEE 5
Vishnupriya 5003 45000.0 Kerla CVE 3
Kavya 5004 50000.0 Mangalore ISE 2
Bhargavi 5005 55000.0 Bangalore ISE 1
Anusha 5006 60000.0 Mumbai ECE 6
Ajay 5007 65000.0 Yellapur ISE 5
Bharath 5008 70000.0 Bangalore CSE 7
Priya 5009 75000.0 Mandya EEE 3
Nikitha 5010 35000.0 Bangalore ISE 4
Keerthi 5011 60000.0 Kerala CVE 5
Jeevan 5012 75000.0 Mumbai EEE 1
Monika 5013 60000.0 Mandya ISE 5
Prathap 5014 70000.0 Hassan CSE 5
Darshini 5015 50000.0 Jharkand ECE 7
Pavan 5016 65000.0 Bangalore CVE 3
Ajay 5017 70000.0 Kerala CSE 2
Asha 5018 80000.0 Bangalore ISE 4
Anjali 5019 60000.0 Jharkand CSE 5
Time taken: 0.287 seconds, Fetched: 25 row(s)
hive>

```

2. Demonstrate the Alter command for the following cases,

a. Rename the table name to “Emp”.

```

Activities Terminal Fri 10:04
hadoop@ubuntu: ~/apache-hive-3.1.2-bin/bin

Arjun 5022 55000.0 Bidar CVE 3
Ananya 5023 85000.0 Bijapur EEE 6
Ani 5024 60000.0 Kerala ECE 5
Harsha 5000 30000.0 Bangalore ISE 5
Bhyravi 5001 35000.0 Mysore CSE 6
Sadhana 5002 40000.0 Dharwad EEE 5
Vishnupriya 5003 45000.0 Kerla CVE 3
Kavya 5004 50000.0 Mangalore ISE 2
Bhargavi 5005 55000.0 Bangalore ISE 1
Anusha 5006 60000.0 Mumbai ECE 6
Ajay 5007 65000.0 Yellapur ISE 5
Bharath 5008 70000.0 Bangalore CSE 7
Priya 5009 75000.0 Mandya EEE 3
Nikitha 5010 35000.0 Bangalore ISE 4
Keerthi 5011 60000.0 Kerala CVE 5
Jeevan 5012 75000.0 Mumbai EEE 1
Monika 5013 60000.0 Mandya ISE 5
Prathap 5014 70000.0 Hassan CSE 5
Darshini 5015 50000.0 Jharkand ECE 7
Pavan 5016 65000.0 Bangalore CVE 3
Ajay 5017 70000.0 Kerala CSE 2
Asha 5018 80000.0 Bangalore ISE 4
Anjali 5019 60000.0 Jharkand CSE 5
Time taken: 0.287 seconds, Fetched: 25 row(s)
hive> alter table employee rename to emp;
OK
Time taken: 0.156 seconds
hive> show tables;
OK
emp
Time taken: 0.044 seconds, Fetched: 1 row(s)
hive>

```

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

```
Activities Terminal Fri 10:05
File Edit View Search Terminal Help
hadoop@ubuntu: ~/apache-hive-3.1.2-bin/bin
Priya 5009 75000.0 Mandya EEE 3
Nikitha 5010 35000.0 Bangalore ISE 4
Keerthi 5011 60000.0 Kerala CVE 5
Jeevan 5012 75000.0 Mumbai EEE 1
Monika 5013 60000.0 Mandya ISE 5
Prathap 5014 70000.0 Hassan CSE 5
Darshini 5015 50000.0 Jharkand ECE 7
Pavan 5016 65000.0 Bangalore CVE 3
Ajay 5017 70000.0 Kerala CSE 2
Asha 5018 80000.0 Bangalore ISE 4
Anjali 5019 60000.0 Jharkand CSE 5
Time taken: 0.287 seconds, Fetched: 25 row(s)
hive> alter table employee rename to emp;
OK
Time taken: 0.156 seconds
hive> show tables;
OK
emp
Time taken: 0.044 seconds, Fetched: 1 row(s)
hive> alter table emp change dname dept_name string;
OK
Time taken: 0.225 seconds
hive> desc emp;
OK
name          string
ssn           int
salary         float
address        string
dept_name     string
experience    int
Time taken: 0.052 seconds, Fetched: 6 row(s)
hive> 
```

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

```
Activities Terminal Fri 10:07
File Edit View Search Terminal Help
hive> Time taken: 0.225 seconds
hive> desc emp;
OK
+-----+-----+
| name      | string |
| ssn       | int    |
| salary    | float  |
| address   | string |
| dept_name | string |
| experience| int   |
+-----+-----+
Time taken: 0.052 seconds, Fetched: 6 row(s)
hive> select name,ssn,salary from emp where salary>=50000;
OK
+-----+-----+-----+
| Arjun    | 5022  | 55000.0 |
| Ananya   | 5023  | 85000.0 |
| Ani      | 5024  | 60000.0 |
| Kavya    | 5004  | 50000.0 |
| Bhargavi | 5005  | 55000.0 |
| Anusha   | 5006  | 60000.0 |
| Ajay     | 5007  | 65000.0 |
| Bharath  | 5008  | 70000.0 |
| Priya    | 5009  | 75000.0 |
| Keerthi  | 5011  | 60000.0 |
| Jeevan   | 5012  | 75000.0 |
| Monika   | 5013  | 60000.0 |
| Prathap  | 5014  | 70000.0 |
| Darshini | 5015  | 50000.0 |
| Pavan    | 5016  | 65000.0 |
| Ajay     | 5017  | 70000.0 |
| Asha    | 5018  | 80000.0 |
| Anjali   | 5019  | 60000.0 |
+-----+-----+-----+
Time taken: 0.55 seconds, Fetched: 18 row(s)
hive>
```

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

```
Activities Terminal Fri 10:09
File Edit View Search Terminal Help
hive> dept_name          string
experience        int
Time taken: 0.052 seconds, Fetched: 6 row(s)
hive> select name,ssn,salary from emp where salary>=50000;
OK
+-----+-----+-----+
| Arjun    | 5022  | 55000.0 |
| Ananya   | 5023  | 85000.0 |
| Ani      | 5024  | 60000.0 |
| Kavya    | 5004  | 50000.0 |
| Bhargavi | 5005  | 55000.0 |
| Anusha   | 5006  | 60000.0 |
| Ajay     | 5007  | 65000.0 |
| Bharath  | 5008  | 70000.0 |
| Priya    | 5009  | 75000.0 |
| Keerthi  | 5011  | 60000.0 |
| Jeevan   | 5012  | 75000.0 |
| Monika   | 5013  | 60000.0 |
| Prathap  | 5014  | 70000.0 |
| Darshini | 5015  | 50000.0 |
| Pavan    | 5016  | 65000.0 |
| Ajay     | 5017  | 70000.0 |
| Asha    | 5018  | 80000.0 |
| Anjali   | 5019  | 60000.0 |
+-----+-----+-----+
Time taken: 0.55 seconds, Fetched: 18 row(s)
hive> select name,address,experience from emp where address="Bangalore" and experience<5;
OK
+-----+-----+-----+
| Bhargavi | Bangalore | 1 |
| Nikitha  | Bangalore | 4 |
| Pavan    | Bangalore | 3 |
| Asha    | Bangalore | 4 |
+-----+-----+-----+
Time taken: 0.313 seconds, Fetched: 4 row(s)
hive>
```

5. Create separate view containing Name, Dept_name of employees

```
Activities Terminal Fri 10:11
File Edit View Search Terminal Help
hive> create view emp_dept as select name,dept_name from emp;
OK
Time taken: 0.168 seconds
hive> select* from emp_dept;
OK
Akash ISE
Abhay CSE
Arjun EEE
Ananya EEE
Anil ECE
Harsha ISE
Bhyravt CSE
Sadhana EEE
Vlshnupriya CVE
Kavya ISE
Bhargavi ISE
Anusha ECE
Ajay ISE
Bharath CSE
Priya EEE
Nikitha ISE
Keerthi CVE
Jeevan EEE
Monika ISE
Prathap CSE
Darshini ECE
Pavan CVE
Ajay CSE
Asha ISE
Anjali CSE
Time taken: 0.233 seconds, Fetched: 25 row(s)
::: hive>
```

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

```
Activities Terminal Fri 10:34
File Edit View Search Terminal Help
Time taken: 0.233 seconds, Fetched: 25 row(s)
hive> select name,ssn from emp group by ssn,name order by name;
Query ID = hdoop_20210709101222_6f8523fe-cfe5-4423-a165-b91ee03fa561
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1625849013163_0002, Tracking URL = http://ubuntu:8088/proxy/application_1625849013163_0002/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job -kill job_1625849013163_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-09 10:12:31,592 Stage-1 map = 0%, reduce = 0%
2021-07-09 10:12:53,796 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 17.85 sec
2021-07-09 10:12:58,927 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 19.13 sec
MapReduce Total cumulative CPU time: 19 seconds 130 msec
Ended Job = job_1625849013163_0002
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1625849013163_0003, Tracking URL = http://ubuntu:8088/proxy/application_1625849013163_0003/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job -kill job_1625849013163_0003
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2021-07-09 10:13:09,918 Stage-2 map = 0%, reduce = 0%
2021-07-09 10:13:14,042 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 0.77 sec
```

```
Activities Terminal Fri 10:16
hadoop@ubuntu:~/apache-hive-3.1.2-bin/bin

MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 19.13 sec HDFS Read: 12773 HDFS Write: 771 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 1.87 sec HDFS Read: 8171 HDFS Write: 687 SUCCESS
Total MapReduce CPU Time Spent: 21 seconds 0 msec

OK
Abhay      5021
Ajay       5007
Ajay       5017
Akash      5020
Ananya     5023
Ani        5024
Anjali     5019
Anusha     5006
Arjun      5022
Asha       5018
Bharath    5008
Bhargavi   5005
Bhyravil  5001
Darshini   5015
Harsha     5000
Jeevan     5012
Kavya      5004
Keerthi    5011
Monika     5013
Nikitha    5010
Pavan      5016
Prathap    5014
Priya      5009
Sadhana    5002
Vishnupriya 5003
Time taken: 59.221 seconds, Fetched: 25 row(s)
hive> 
```

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

```
Activities Terminal Fri 10:16
hadoop@ubuntu:~/apache-hive-3.1.2-bin/bin

Pavan      5016
Prathap    5014
Priya      5009
Sadhana    5002
Vishnupriya 5003
Time taken: 59.221 seconds, Fetched: 25 row(s)
hive> select max(salary),min(salary),avg(salary) from emp;
Query ID = hadoop_20210709101545_c071f701-2ae4-4cf4-af3a-0ae5d692a7c4
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_1625849013163_0004, Tracking URL = http://ubuntu:8088/proxy/application_1625849013163_0004/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1625849013163_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-09 10:15:51,967 Stage-1 map = 0%, reduce = 0%
2021-07-09 10:15:56,090 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 0.99 sec
2021-07-09 10:16:01,231 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.75 sec
MapReduce Total cumulative CPU time: 2 seconds 750 msec
Ended Job = job_1625849013163_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.75 sec HDFS Read: 18306 HDFS Write: 123 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 750 msec
OK
85000.0 30000.0 57200.0
Time taken: 16.434 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

```
File Edit View Search Terminal Help Fri 10:25
hive> create table department(dno int,dname string)row format delimited fields terminated by ",";
OK
Time taken: 0.166 seconds
hive> desc department;
OK
dno          int
dname        string
Time taken: 0.045 seconds, Fetched: 2 row(s)
hive> insert into department values(1,'ISE'),(2,'ECE'),(3,'CSE'),(4,'EEE'),(5,'CVE')
Query ID = hdoop_20210709101955_8a203c4a-7078-4626-8dab-fead65ea8b43
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_1625849013163_0005, Tracking URL = http://ubuntu:8088/proxy/application_1625849013163_0005
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job -kill job_1625849013163_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-09 10:20:02,442 Stage-1 map = 0%,  reduce = 0%
2021-07-09 10:20:08,678 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.72 sec
2021-07-09 10:20:13,835 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.62 sec
MapReduce Total cumulative CPU time: 3 seconds 620 msec
Ended Job = job_1625849013163_0005
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://127.0.0.1:9000/user/hive/warehouse/employeedb.db/department/.hive-staging_hive_2021-07-09_10-19-55_396_18614444424343508
7-1/-ext-10000
Loading data to table employeedb.department
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1  Cumulative CPU: 3.62 sec  HDFS Read: 15804 HDFS Write: 314 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 620 msec
OK
Time taken: 19.776 seconds
```

Performing outer Join over Dname:

```
File Edit View Search Terminal Help Fri 10:26
hive> select* from department;
OK
1    ISE
2    ECE
3    CSE
4    EEE
5    CVE
Time taken: 0.108 seconds, Fetched: 5 row(s)
hive> alter table department change dname dept_name string;
OK
Time taken: 0.135 seconds
hive> desc department;
OK
dno          int
dept_name    string
Time taken: 0.024 seconds, Fetched: 2 row(s)
hive> select name,ssn,d.dept_name,dno from emp e full outer join department d on e.dept_name=d.dept_name;
Query ID = hdoop_20210709102216_459b6ae6-5659-4df4-b61a-89306839c3fb
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_1625849013163_0006, Tracking URL = http://ubuntu:8088/proxy/application_1625849013163_0006/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job -kill job_1625849013163_0006
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2021-07-09 10:22:22,581 Stage-1 map = 0%,  reduce = 0%
2021-07-09 10:23:11,037 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 67.59 sec
2021-07-09 10:23:16,161 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 68.82 sec
MapReduce Total cumulative CPU time: 1 minutes 8 seconds 820 msec
Ended Job = job_1625849013163_0006
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2  Reduce: 1  Cumulative CPU: 68.82 sec  HDFS Read: 17842 HDFS Write: 837 SUCCESS
Total MapReduce CPU Time Spent: 1 minutes 8 seconds 820 msec
```

```

Activities Terminal Fri 10:27
File Edit View Search Terminal Help
Starting Job = job_1625849013163_0006, Tracking URL = http://ubuntu:8088/proxy/application_1625849013163_0006/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1625849013163_0006
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2021-07-09 10:22:22,581 Stage-1 map = 0%, reduce = 0%
2021-07-09 10:23:11,037 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 67.59 sec
2021-07-09 10:23:16,161 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 68.82 sec
MapReduce Total cumulative CPU time: 1 minutes 8 seconds 820 msec
Ended Job = job_1625849013163_0006
MapReduce Jobs Launched:
A Stage-Stage-1: Map: 2 Reduce: 1 Cumulative CPU: 68.82 sec HDFS Read: 17842 HDFS Write: 837 SUCCESS
Total MapReduce CPU Time Spent: 1 minutes 8 seconds 820 msec
OK
Abhay 5021 CSE 3
Bharath 5008 CSE 3
Ajay 5017 CSE 3
Prathap 5014 CSE 3
Anjali 5019 CSE 3
Bhyravi 5001 CSE 3
Vishnupriya 5003 CVE 5
Arjun 5022 CVE 5
Keerthi 5011 CVE 5
Pavan 5016 CVE 5
Anusha 5006 ECE 2
Darshini 5015 ECE 2
Ani 5024 ECE 2
Priya 5009 EEE 4
Jeevan 5012 EEE 4
Ananya 5023 EEE 4
Sadhana 5002 EEE 4
Ajay 5007 ISE 1
Asha 5018 ISE 1
Monika 5013 ISE 1
Nikitha 5010 ISE 1
Bhargavi 5005 ISE 1
Kavya 5004 ISE 1
Harsha 5000 ISE 1
Akash 5020 ISE 1
Time taken: 61.626 seconds, Fetched: 25 row(s)
hive> select name,ssn,d.dept_name,dno from emp e left outer join department d on e.dept_name=d.dept_name;

```

Performing Left outer join over Dname:

```

Activities Terminal Fri 10:40
File Edit View Search Terminal Help
Time taken: 61.626 seconds, Fetched: 25 row(s)
hive> select name,ssn,d.dept_name,dno from emp e left outer join department d on e.dept_name=d.dept_name;
Query ID = hdoop_20210709102837_e2df5347-f9f7-4d2a-9f99-cc5319ac99b0
Total jobs = 1
SLF4J: Found binding in [jar:file:/home/hadoop/apache-hive-3.1.2-bin/lib/log4j-slf4j-impl-2.10.0.jar!/_org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/hadoop/hadoop-3.2.1/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/_org/slf4j/impl/StaticLoggerBinder.class]
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_1625849013163_0007, Tracking URL = http://ubuntu:8088/proxy/application_1625849013163_0007/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1625849013163_0007
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2021-07-09 10:28:52,936 Stage-3 map = 0%, reduce = 0%
2021-07-09 10:28:58,056 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 1.24 sec
MapReduce Total cumulative CPU time: 1 seconds 240 msec
Ended Job = job_1625849013163_0007
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 1.24 sec HDFS Read: 10168 HDFS Write: 837 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 240 msec
OK
Akash 5020 ISE 1
Abhay 5021 CSE 3
Arjun 5022 CVE 5
Ananya 5023 EEE 4
Ani 5024 ECE 2
Harsha 5000 ISE 1
Bhyravi 5001 CSE 3
Sadhana 5002 EEE 4
Vishnupriya 5003 CVE 5
Kavya 5004 ISE 1
Bhargavi 5005 ISE 1
Anusha 5006 ECE 2
Ajay 5007 ISE 1
Bharath 5008 CSE 3
Priya 5009 EEE 4
Nikitha 5010 ISE 1
Keerthi 5011 CVE 5

```

Performing Right outer Join over Dname:

```
Activities Terminal Fri 10:41
hadoop@ubuntu: ~/apache-hive-3.1.2-bin/bin

File Edit View Search Terminal Help
Akash 5020 ISE 1
Abhay 5021 CSE 3
Arjun 5022 CVE 5
Ananya 5023 EEE 4
Anu 5024 ECE 2
Harsha 5000 ISE 1
Bhyravt. 5001 CSE 3
Sadhana 5002 EEE 4
Vishnupriya 5003 CVE 5
Kavy 5004 ISE 1
Bhargavi 5005 ISE 1
Anusha 5006 ECE 2
Ajay 5007 ISE 1
Bharath 5008 CSE 3
Priya 5009 EEE 4
Nikitha 5010 ISE 1
Keerthi 5011 CVE 5
Jeevan 5012 EEE 4
Monika 5013 ISE 1
Prathap 5014 CSE 3
Darshini 5015 ECE 2
Pavan 5016 CVE 5
Ajay 5017 CSE 3
Asha 5018 ISE 1
Anjali 5019 CSE 3
Time taken: 21.319 seconds, Fetched: 25 row(s)
hive> select name,ssn,d.dept_name,dno from emp e right outer join department d on e.dept_name=d.dept_name;
Query ID = hadoop_20210709103129_13813147-7ebe-455b-ae2c-f4aff5126ad9
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_1625849013163_0008, Tracking URL = http://ubuntu:8088/proxy/application_1625849013163_0008/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1625849013163_0008
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2021-07-09 10:31:42,038 Stage-3 map = 0%, reduce = 0%
2021-07-09 10:31:48,312 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 1.82 sec
2021-07-09 10:31:48,312 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 1.82 sec
MapReduce Total cumulative CPU time: 1 seconds 820 msec
```

```
Activities Terminal Fri 10:42
hadoop@ubuntu: ~/apache-hive-3.1.2-bin/bin

File Edit View Search Terminal Help
Starting Job = job_1625849013163_0008, Tracking URL = http://ubuntu:8088/proxy/application_1625849013163_0008/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1625849013163_0008
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2021-07-09 10:31:42,038 Stage-3 map = 0%, reduce = 0%
2021-07-09 10:31:48,312 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 1.82 sec
MapReduce Total cumulative CPU time: 1 seconds 820 msec
Ended Job = job_1625849013163_0008
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 1.82 sec HDFS Read: 9141 HDFS Write: 837 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 820 msec
OK
Akash 5028 ISE 1
Harsha 5000 ISE 1
Kavy 5004 ISE 1
Bhargavi 5005 ISE 1
Ajay 5007 ISE 1
Nikitha 5010 ISE 1
Monika 5013 ISE 1
Asha 5018 ISE 1
Anu 5024 ECE 2
Anusha 5006 ECE 2
Darshini 5015 ECE 2
Abhay 5021 CSE 3
Bhyravt. 5001 CSE 3
Bharath 5008 CSE 3
Prathap 5014 CSE 3
Ajay 5017 CSE 3
Anjali 5019 CSE 3
Ananya 5023 EEE 4
Sadhana 5002 EEE 4
Priya 5009 EEE 4
Jeevan 5012 EEE 4
Arjun 5022 CVE 5
Vishnupriya 5003 CVE 5
Keerthi 5011 CVE 5
Pavan 5016 CVE 5
Time taken: 19.764 seconds, Fetched: 25 row(s)
::: hive>
```