B.M.S. COLLEGE OF ENGINEERING BENGALURU
Autonomous Institute, Affiliated to VTU

Lab Record

Big-Data Analytics

Submitted in partial fulfillment for the 6th Semester Laboratory

Bachelor of Technology
in
Computer Science and Engineering

Submitted by:

P Sai Deekshith
1BM18S148

Department of Computer Science and Engineering
B.M.S. College of Engineering
Bull Temple Road, Basavanagudi, Bangalore 560 019
Mar-June 2021

B.M.S. COLLEGE OF ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the Big-Data Analytics (20CS6PEBDA) laboratory has been carried out by P Sai Deekshith (1BM18CS148) during the 6th Semester Mar-June-2021.

Signature of the Faculty Incharge:

Bhoomika A P
Associate Professor
Department of Computer Science and Engineering
B.M.S. College of Engineering, Bangalore

TABLE OF CONTENTS

PROGRAM - 1

PERFORM THE FOLLOWING DB OPERATIONS USING CASSANDRA LIBRARY.

Create a keyspace by name Employee

```
cqlsh> create keyspace employee with replication = {'class': 'SimpleStrategy', 'replication_factor': 1};
cqlsh> use employee;
```

Create a column family by name Employee-Info with attributes Emp_Id Primary Key, Emp_Name, Designation, Date_of_Joining, Salary, Dept_Name

```
cqlsh:employee> create table employeeinfo(emp_id int primary key, emp_name text, designation text, doj timestamp, salary double, dept_name text);
```

Insert the values into the table in batch

```
cqlsh:employee> begin batch
        ... insert into employeeinfo(emp_id, emp_name, designation, doj, salary, dept_name) values (1, 'Ajay', 'Data analyst', '2018-04-16', 20000, 'Corporate');
        ... insert into employeeinfo(emp_id, emp_name, designation, doj, salary, dept_name) values (121, 'Chaitra', 'web design', '2019-08-06', 15000, 'web_designer');
        ... apply batch;
cqlsh:employee> select * from employeeinfo;
```

Update Employee name and Department of Emp-Id 121

```
cqlsh:employee> update employeeinfo set emp_name = 'Joy', dept_name = 'Management' where emp_id = 121;
cqlsh:employee> select * from employeeinfo;
```

Alter the schema of the table Employee_Info to add a column Projects which stores a set of Projects done by the corresponding Employee.

```
cqlsh:employee> alter table employeeinfo add projects set<text>;
```

Update the altered table to add project names.

cqlsh:employee> update employeeinfo set projects = {'project1', 'project2'} where emp_id in(1,121);
cqlsh:employee> select * from employeeinfo;

<u>Create a TTL of 15 seconds to display the values of Employees.</u>

cqlsh:employee> begin batch
... insert into employeeinfo(emp_id, emp_name, designation, doj, salary, dept_name) values (121, 'Boris', 'MTO', '2001-08-05', 12212, 'Corporate') using ttl 15;
... apply batch;
cqlsh:employee> select ttl(designation) from employeeinfo where emp_id = 121;

Output :

```
Terminal
Your Interactive Bash Terminal.
$ cqlsh
Connected to Test Cluster at 127.0.0.1:9042.
[cqlsh 5.0.1 | Cassandra 4.0-beta2 | CQL spec 3.4.5 | Native protocol v4]
Use HELP for help.
cqlsh>
cqlsh>
cqlsh> create keyspace employee with replication = {'class': 'SimpleStrategy', 'replication_factor': 1};
cqlsh> use employee;
cqlsh:employee> create table employeeinfo(emp_id int primary key, emp_name text, designation text, doj timestamp, salary double, dept_name text);
cqlsh:employee> begin batch
           ... insert into employeeinfo(emp_id, emp_name, designation, doj, salary, dept_name) values (1, 'Ajay', 'Data analyst', '2018-04-16', 2000
0, 'Corporate');
           ... insert into employeeinfo(emp_id, emp_name, designation, doj, salary, dept_name) values (121, 'Chaitra', 'web design', '2019-08-06', 1
5000, 'web_designer');
           ... apply batch;
cqlsh:employee> select * from employeeinfo;

 emp_id | dept_name    | designation  | doj                             | emp_name | salary
--------+--------------+--------------+---------------------------------+----------+--------
      1 |    Corporate | Data analyst | 2018-04-16 00:00:00.000000+0000 |     Ajay |  20000
    121 | web_designer |   web design | 2019-08-06 00:00:00.000000+0000 |  Chaitra |  15000

(2 rows)
cqlsh:employee> update employeeinfo set emp_name = 'Joy', dept_name = 'Management' where emp_id = 121;
cqlsh:employee> select * from employeeinfo;

 emp_id | dept_name  | designation  | doj                             | emp_name | salary
--------+------------+--------------+---------------------------------+----------+--------
      1 |  Corporate | Data analyst | 2018-04-16 00:00:00.000000+0000 |     Ajay |  20000
    121 | Management |   web design | 2019-08-06 00:00:00.000000+0000 |      Joy |  15000

(2 rows)
cqlsh:employee> alter table employeeinfo add projects set<text>;
cqlsh:employee> update employeeinfo set projects = {'project1', 'project2'} where emp_id in(1,121);
cqlsh:employee> select * from employeeinfo;

 emp_id | dept_name  | designation  | doj                             | emp_name | projects                  | salary
--------+------------+--------------+---------------------------------+----------+---------------------------+--------
      1 |  Corporate | Data analyst | 2018-04-16 00:00:00.000000+0000 |     Ajay | {'project1', 'project2'} |  20000
    121 | Management |   web design | 2019-08-06 00:00:00.000000+0000 |      Joy | {'project1', 'project2'} |  15000

(2 rows)
```

```
Terminal
--------+------------+--------------+---------------------------------+----------+--------
      1 |  Corporate | Data analyst | 2018-04-16 00:00:00.000000+0000 |     Ajay |  20000
    121 | Management |   web design | 2019-08-06 00:00:00.000000+0000 |      Joy |  15000

(2 rows)
cqlsh:employee> alter table employeeinfo add projects set<text>;
cqlsh:employee> update employeeinfo set projects = {'project1', 'project2'} where emp_id in(1,121);
cqlsh:employee> select * from employeeinfo;

 emp_id | dept_name  | designation  | doj                             | emp_name | projects                  | salary
--------+------------+--------------+---------------------------------+----------+---------------------------+--------
      1 |  Corporate | Data analyst | 2018-04-16 00:00:00.000000+0000 |     Ajay | {'project1', 'project2'} |  20000
    121 | Management |   web design | 2019-08-06 00:00:00.000000+0000 |      Joy | {'project1', 'project2'} |  15000

(2 rows)
cqlsh:employee> begin batch
           ... insert into employeeinfo(emp_id, emp_name, designation, doj, salary, dept_name) values (121, 'Boris', 'MTO', '2001-08-05', 12212, 'Co
rporate') using ttl 15;
           ... apply batch;
cqlsh:employee> select ttl(designation) from employeeinfo where emp_id = 121;

 ttl(designation)
------------------
             null

(1 rows)
cqlsh:employee> select * from employeeinfo;

 emp_id | dept_name | designation  | doj                             | emp_name | projects                  | salary
--------+-----------+--------------+---------------------------------+----------+---------------------------+--------
      1 | Corporate | Data analyst | 2018-04-16 00:00:00.000000+0000 |     Ajay | {'project1', 'project2'} |  20000
    121 |      null |         null |                            null |     null | {'project1', 'project2'} |   null

(2 rows)
cqlsh:employee> begin batch insert into employeeinfo(emp_id, emp_name, designation, doj, salary, dept_name) values (121, 'Boris', 'MTO', '2001-08-05'
, 12212, 'Corporate') using ttl 120; apply batch;
cqlsh:employee> select ttl(designation) from employeeinfo where emp_id = 121;
                                                ttl(designation)
------------------
               109

(1 rows)
cqlsh:employee>
```

PROGRAM – 2

PERFORM THE FOLLOWING DB OPERATIONS USING CASSANDRA:

Create a keyspace by name Library

```
cqlsh> create keyspace library  with replication = { 'class' : 'SimpleStrategy','replication_factor':1};
cqlsh> use library;
```

Create a column family by name Library-Info with attributes Stud_Id Primary Key, Counter_value of type Counter, Stud_Name, Book-Name, Book-Id, Date_of_issue

```
cqlsh:library> create table library_info( id int, counter_val counter, stud_name text, book_name text, book_id int, issue_date timestamp,primary key(id,stud_name,book_name,book_id,issue_date));
```

Insert the values into the table in batch

```
cqlsh:library> update library_info SET counter_val = counter_val +1 where id = 1 and stud_name = 'Anand' and book_name = 'CNS' and book_id = 121 and issue_date='2020-12-31';
cqlsh:library> update library_info SET counter_val = counter_val +1 where id = 3 and stud_name = 'Arjun' and book_name = 'ML' and book_id = 112 and issue_date='2021-02-01';

cqlsh:library> update library_info SET counter_val = counter_val +1 where id = 5 and stud_name = 'Chaitra' and book_name = 'Python' and book_id = 114 and issue_date='2009-08-27';
cqlsh:library> select * from library_info;
```

Display the details of the table created and increase the value of the counter

```
cqlsh:library> update library_info SET counter_val = counter_val +1 where id = 3 and stud_name = 'Arjun' and book_name = 'ML' and book_id = 112 and issue_date='2021-02-01';
```

Write a query to show that a student with id 112 has taken a book “BDA” 2 times.

```
cqlsh:library> select * from library_info where counter_val = 2 allow filtering;
```

Export the created column to a csv file

```
cqlsh:library> copy library_info(id,counter_val,stud_name,book_name,book_id,issue_date) to 'Desktop/library_data.csv';
```

Import a given csv dataset from local file system into Cassandra column family

cqlsh:library> copy library_info(id,counter_val,stud_name,book_name,book_id,issue_date) from 'Desktop/library_data.csv';

Output :

```
Terminal
Your Interactive Bash Terminal.
$ cqlsh
Connected to Test Cluster at 127.0.0.1:9042.
[cqlsh 5.0.1 | Cassandra 4.0-beta2 | CQL spec 3.4.5 | Native protocol v4]
Use HELP for help.
cqlsh>
cqlsh>
cqlsh> create keyspace library  with replication = { 'class' : 'SimpleStrategy','replication_factor':1};
cqlsh> use library;
cqlsh:library> create table library_info( id int, counter_val counter, stud_name text, book_name text, book_id int, issue_date timestamp,primary key(
id,stud_name,book_name,book_id,issue_date));
cqlsh:library> update library_info SET counter_val = counter_val +1 where id = 1 and stud_name = 'Anand' and book_name = 'CNS' and book_id = 121 and
issue_date='2020-12-31';
cqlsh:library> update library_info SET counter_val = counter_val +1 where id = 3 and stud_name = 'Arjun' and book_name = 'ML' and book_id = 112 and i
ssue_date='2021-02-01';
cqlsh:library> update library_info SET counter_val = counter_val +1 where id = 5 and stud_name = 'Chaitra' and book_name = 'Python' and book_id = 114
 and issue_date='2009-08-27';
cqlsh:library> select * from library_info;

 id | stud_name | book_name | book_id | issue_date                      | counter_val
----+-----------+-----------+---------+---------------------------------+-------------
  5 |   Chaitra |    Python |     114 | 2009-08-27 00:00:00.000000+0000 |           1
  1 |     Anand |       CNS |     121 | 2020-12-31 00:00:00.000000+0000 |           1
  3 |     Arjun |        ML |     112 | 2021-02-01 00:00:00.000000+0000 |           1

(3 rows)
cqlsh:library> update library_info SET counter_val = counter_val +1 where id = 3 and stud_name = 'Arjun' and book_name = 'BDA' and book_id = 112 and
issue_date='2011-12-20';
cqlsh:library> select * from library_info where counter_val = 2 allow filtering;

 id | stud_name | book_name | book_id | issue_date | counter_val
----+-----------+-----------+---------+------------+-------------

(0 rows)
cqlsh:library> update library_info SET counter_val = counter_val +1 where id = 3 and stud_name = 'Arjun' and book_name = 'ML' and book_id = 112 and i
ssue_date='2011-12-20';
cqlsh:library> select * from library_info where counter_val = 2 allow filtering;
 id | stud_name | book_name | book_id | issue_date | counter_val
----+-----------+-----------+---------+------------+-------------

(0 rows)
cqlsh:library> update library_info SET counter_val = counter_val +1 where id = 3 and stud_name = 'Arjun' and book_name = 'ML' and book_id = 112 and i
ssue_date='2021-02-01';
```

```
Terminal

(3 rows)
cqlsh:library> update library_info SET counter_val = counter_val +1 where id = 3 and stud_name = 'Arjun' and book_name = 'BDA' and book_id = 112 and
issue_date='2011-12-20';
cqlsh:library> select * from library_info where counter_val = 2 allow filtering;

 id | stud_name | book_name | book_id | issue_date | counter_val
----+-----------+-----------+---------+------------+-------------

(0 rows)
cqlsh:library> update library_info SET counter_val = counter_val +1 where id = 3 and stud_name = 'Arjun' and book_name = 'ML' and book_id = 112 and i
ssue_date='2011-12-20';
cqlsh:library> select * from library_info where counter_val = 2 allow filtering;
 id | stud_name | book_name | book_id | issue_date | counter_val
----+-----------+-----------+---------+------------+-------------

(0 rows)
cqlsh:library> update library_info SET counter_val = counter_val +1 where id = 3 and stud_name = 'Arjun' and book_name = 'ML' and book_id = 112 and i
ssue_date='2021-02-01';
cqlsh:library> select * from library_info where counter_val = 2 allow filtering;
 id | stud_name | book_name | book_id | issue_date                      | counter_val
----+-----------+-----------+---------+---------------------------------+-------------
  3 |     Arjun |        ML |     112 | 2021-02-01 00:00:00.000000+0000 |           2

(1 rows)
cqlsh:library> copy library_info(id,counter_val,stud_name,book_name,book_id,issue_date) to 'Desktop/library_data.csv';
Using 1 child processes

Starting copy of library.library_info with columns [id, counter_val, stud_name, book_name, book_id, issue_date].
cqlshlib.copyutil.ExportProcess.write_rows_to_csv(): writing row
cqlshlib.copyutil.ExportProcess.write_rows_to_csv(): writing row
cqlshlib.copyutil.ExportProcess.write_rows_to_csv(): writing row
cqlshlib.copyutil.ExportProcess.write_rows_to_csv(): writing row
cqlshlib.copyutil.ExportProcess.write_rows_to_csv(): writing row
Processed: 5 rows; Rate:       8 rows/s; Avg. rate:       9 rows/s
5 rows exported to 1 files in 0.555 seconds.
cqlsh:library> copy library_info(id,counter_val,stud_name,book_name,book_id,issue_date) from 'Desktop/library_data.csv';
Using 1 child processes

Starting copy of library.library_info with columns [id, counter_val, stud_name, book_name, book_id, issue_date].
Processed: 5 rows; Rate:       9 rows/s; Avg. rate:      14 rows/s
5 rows imported from 1 files in 0.365 seconds (0 skipped).
cqlsh:library>
```

PROGRAM – 3

PERFORM THE FOLLOWING DB OPERATIONS USING MONGODB:

Create a database “Student” with the following attributes Rollno, Age, ContactNo, Email-Id.

use student

Insert appropriate values

db.student.insert({Roll: 10, Name: "suma", age: 21, contact: "7723112389", email: "suma@gmail.com"})
db.student.insert({Roll: 11, Name: "ABC", age: 20, contact: "9263532389", email: "abc@gmail.com"})
db.student.insert({Roll: 12, Name: "shek", age: 21, contact: "7788996655", email: "shek@gmail.com"})
db.student.insert({Roll: 13, Name: "raj", age: 20, contact: "1234123412", email: "raj@gmail.com"})

Write a query to update Email-Id of a student with rollno 10.

db.student.update({Roll:10}, {$set: {email: "suma123@gmail.com"}})

Replace the student name from “ABC” to “FEM” of rollno 11.

db.student.update({Roll:11}, {$set: {Name: "FEM"}})

Export the created table into local file system

mongoexport --db student --collection student --type csv --out D:\export.csv --fields “Roll,Name,age,contact,email”

Drop the table

db.student.drop()

Import a given csv dataset from the local file system into mongodb collection.

mongoimport --db student --collection student --type csv --file D:\export.csv --headerline

Output :

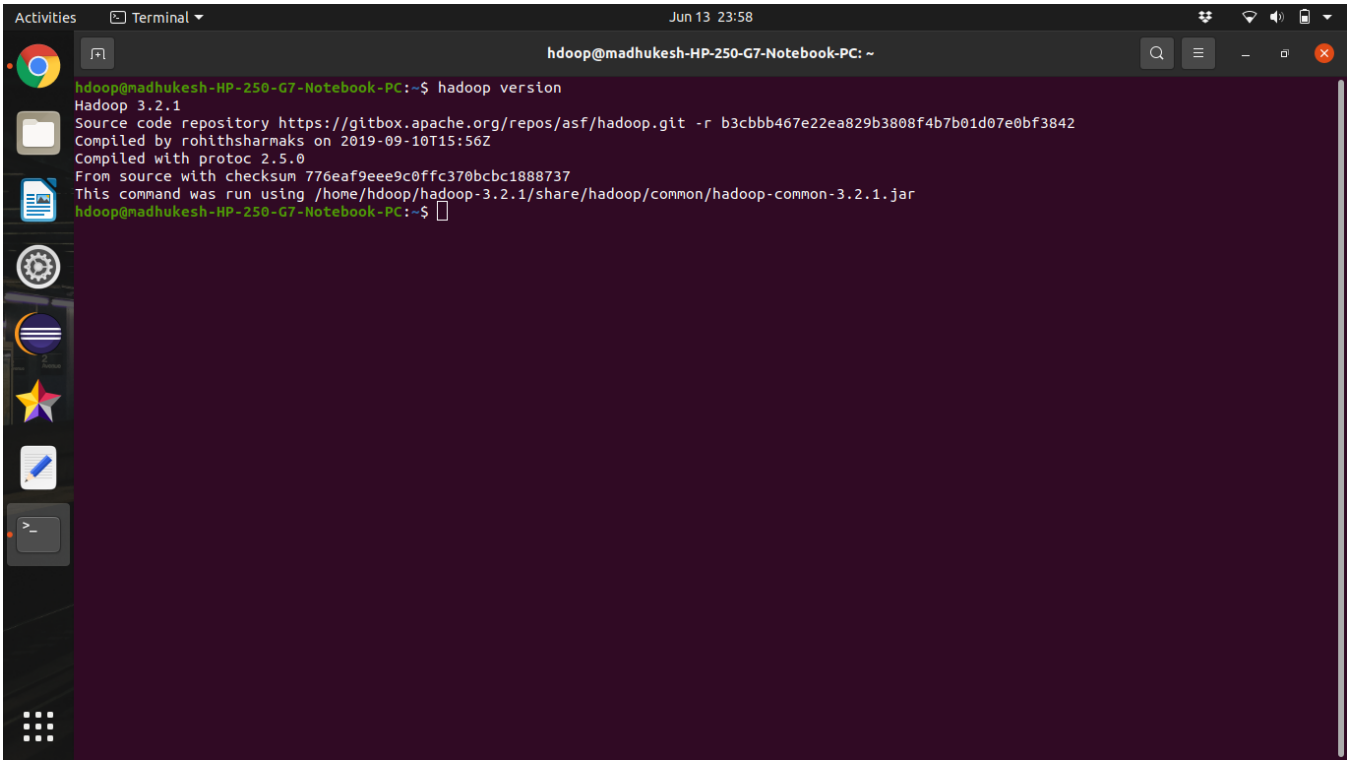
```
use student
db.student.insert({Roll: 10, Name: "suma", age: 21, contact: "7723112389", email: "suma@gmail.com"})
db.student.insert({Roll: 11, Name: "ABC", age: 20, contact: "9263532389", email: "abc@gmail.com"})
db.student.insert({Roll: 12, Name: "shek", age: 21, contact: "7788996655", email: "shek@gmail.com"})
db.student.insert({Roll: 13, Name: "raj", age: 20, contact: "1234123412", email: "raj@gmail.com"})

db.student.update({Roll:10}, {$set: {email: "suma123@gmail.com"}})
db.student.update({Roll:11}, {$set: {Name: "FEM"}})

show collections
db.student.find()

mongoexport --db testdb --collection student --out C:\Users\SUMALATA\OneDrive\Desktop\output.json

Drop student

mongoimport --db testdb --collection Student C:\Users\SUMALATA\OneDrive\Desktop\output.json
```

student 0.059 sec. 0 50

Key	Value	Type
(1) ObjectId("6067b01b80764c83bffc7871")	{ 6 fields }	Object
_id	ObjectId("6067b01b80764c83bffc7871")	ObjectId
Roll	10.0	Double
Name	suma	String
age	21.0	Double
contact	7723112389	String
email	suma123@gmail.com	String
(2) ObjectId("6067b0b980764c83bffc7872")	{ 6 fields }	Object
_id	ObjectId("6067b0b980764c83bffc7872")	ObjectId
Roll	11.0	Double
Name	FEM	String

PROGRAM – 4

SCREENSHOT OF HADOOP INSTALLED:



PROGRAM – 5

EXECUTION OF HDFS COMMANDS FOR INTERACTION WITH HADOOP ENVIRONMENT. (MINIMUM 10 COMMANDS TO BE EXECUTED:

version

hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop version
Hadoop 3.2.1
Source code repository https://gitbox.apache.org/repos/asf/hadoop.git -r b3cbbb467e22ea829b3808f4b7b01d07e0bf3842
Compiled by rohithsharmaks on 2019-09-10T15:56Z
Compiled with protoc 2.5.0
From source with checksum 776eaf9eee9c0ffc370bcbc1888737
This command was run using /home/hdoop/hadoop-3.2.1/share/hadoop/common/hadoop-common-3.2.1.jar

mkdir

hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -mkdir /samplefile1
2021-04-20 13:37:25,376 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -mkdir /samplefile2
2021-04-20 13:37:43,271 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -mkdir /samplefile3
2021-04-20 13:38:18,887 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

ls

hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -ls /
2021-04-20 13:38:41,762 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

```
Found 3 items
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:37 /samplefile1
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:37 /samplefile2
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:38 /samplefile3


hdoop@madhu-HP-250-G7-Notebook-PC:~$ ls
dfsdata  hadoop-3.2.1  hadoop-3.2.1.tar.gz  tmpdata

hdoop@madhu-HP-250-G7-Notebook-PC:~$ mkdir lab

hdoop@madhu-HP-250-G7-Notebook-PC:~$ mkdir lab/samples/

hdoop@madhu-HP-250-G7-Notebook-PC:~$ mkdir lab/samples/file1

hdoop@madhu-HP-250-G7-Notebook-PC:~$ mkdir lab/samples/file2

hdoop@madhu-HP-250-G7-Notebook-PC:~$ mkdir lab/samples/file3

hdoop@madhu-HP-250-G7-Notebook-PC:~$ mkdir lab/samples/file4
```

<u>put / copyFromLocal</u>

```
hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -put ~/lab/samples/file1 /
2021-04-20 13:48:24,640 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable

hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -put ~/lab/samples/file2 /samplefile1
2021-04-20 13:49:04,048 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable

hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -ls /
2021-04-20 13:50:32,226 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Found 4 items
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:48 /file1
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:49 /samplefile1
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:37 /samplefile2
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:38 /samplefile3
```

```
hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -ls -R /
2021-04-20 13:52:21,533 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:48 /file1
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:49 /samplefile1
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:49 /samplefile1/file2
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:37 /samplefile2
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:38 /samplefile3

hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -copyFromLocal ~/lab/samples/file3
/samplefile2
2021-04-20 13:58:22,912 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable

hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -copyFromLocal ~/lab/sam'ples/file4
/samplefile3
2021-04-20 13:58:38,623 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable

hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -ls -R /
2021-04-20 13:58:49,088 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:48 /file1
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:49 /samplefile1
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:49 /samplefile1/file2
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:58 /samplefile2
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:58 /samplefile2/file3
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:58 /samplefile3
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:58 /samplefile3/file4
```

<u>get / copyToLocal</u>

```
hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -get /file1 ~/lab/copies
2021-04-20 19:16:54,079 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable

hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -get /samplefile1/file2 ~/lab/copies
2021-04-20 19:17:59,535 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
```

```
hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -copyToLocal /samplefile2/file3
~/lab/copies
2021-04-20 19:19:09,548 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable

hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -copyToLocal /samplefile3/file4
~/lab/copies
2021-04-20 19:19:30,733 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable

hdoop@madhu-HP-250-G7-Notebook-PC:~$ ls -l lab/copies
total 12
drwxr-xr-x 2 hdoop hdoop 4096 Apr 20 19:18 file2
drwxr-xr-x 2 hdoop hdoop 4096 Apr 20 19:19 file3
drwxr-xr-x 2 hdoop hdoop 4096 Apr 20 19:19 file4

hdoop@madhu-HP-250-G7-Notebook-PC:~$ ls -l lab
total 8
drwxr-xr-x 5 hdoop hdoop 4096 Apr 20 19:19 copies
drwxrwxr-x 6 hdoop hdoop 4096 Apr 20 13:47 samples

hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -get /file1 ~/lab/copies
2021-04-20 19:22:17,555 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable

hdoop@madhu-HP-250-G7-Notebook-PC:~$ ls -l lab/copies
total 16
drwxr-xr-x 2 hdoop hdoop 4096 Apr 20 19:22 file1
drwxr-xr-x 2 hdoop hdoop 4096 Apr 20 19:18 file2
drwxr-xr-x 2 hdoop hdoop 4096 Apr 20 19:19 file3
drwxr-xr-x 2 hdoop hdoop 4096 Apr 20 19:19 file4


hdoop@madhu-HP-250-G7-Notebook-PC:~$ mkdir lab/text

hdoop@madhu-HP-250-G7-Notebook-PC:~$ cd lab/text/

hdoop@madhu-HP-250-G7-Notebook-PC:~/lab/text$ cat > text1
Hi
I'm executing hadoop commands
```

```
hdoop@madhu-HP-250-G7-Notebook-PC:~/lab/text$ cd

hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -copyFromLocal ~/lab/text/text1 /
2021-04-20 19:26:31,016 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
2021-04-20 19:26:33,108 INFO sasl.SaslDataTransferClient: SASL encryption trust check:
localHostTrusted = false, remoteHostTrusted = false

hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -ls /
2021-04-20 19:27:17,423 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Found 5 items
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:48 /file1
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:49 /samplefile1
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:58 /samplefile2
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:58 /samplefile3
-rw-r--r--   1 hdoop supergroup         33 2021-04-20 19:26 /text1
```

<u>cat</u>

```
hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -cat /text1
2021-04-20 19:28:24,990 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
2021-04-20 19:28:26,530 INFO sasl.SaslDataTransferClient: SASL encryption trust check:
localHostTrusted = false, remoteHostTrusted = false
Hi
I'm executing hadoop commands
```

<u>mv</u>

```
hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -mv /file1 /samplefile1
2021-04-20 19:31:09,926 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable

hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -ls /samplefile1
2021-04-20 19:31:49,316 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Found 2 items
```

```
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:48 /samplefile1/file1
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:49 /samplefile1/file2

hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -ls /
2021-04-20 19:32:12,458 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Found 4 items
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 19:31 /samplefile1
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:58 /samplefile2
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:58 /samplefile3
-rw-r--r--   1 hdoop supergroup         33 2021-04-20 19:26 /text1
```

cp

```
hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -cp /text1 /samplefile3
2021-04-20 19:33:32,689 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
2021-04-20 19:33:34,093 INFO sasl.SaslDataTransferClient: SASL encryption trust check:
localHostTrusted = false, remoteHostTrusted = false
2021-04-20 19:33:34,332 INFO sasl.SaslDataTransferClient: SASL encryption trust check:
localHostTrusted = false, remoteHostTrusted = false

hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -ls -R /
2021-04-20 19:33:52,862 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 19:31 /samplefile1
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:48 /samplefile1/file1
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:49 /samplefile1/file2
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:58 /samplefile2
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:58 /samplefile2/file3
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 19:33 /samplefile3
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:58 /samplefile3/file4
-rw-r--r--   1 hdoop supergroup         33 2021-04-20 19:33 /samplefile3/text1
-rw-r--r--   1 hdoop supergroup         33 2021-04-20 19:26 /text1
```

rm

```
hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -rm /text1
```

```
2021-04-20 19:49:33,071 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Deleted /text1

hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -ls /
2021-04-20 19:49:44,650 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Found 3 items
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 19:31 /samplefile1
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 13:58 /samplefile2
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 19:33 /samplefile3

hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -rm -r /samplefile2
2021-04-20 19:51:13,448 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Deleted /samplefile2

hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -ls /
2021-04-20 19:51:21,573 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Found 2 items
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 19:31 /samplefile1
drwxr-xr-x   - hdoop supergroup          0 2021-04-20 19:33 /samplefile3
```

du

```
hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -du -s /samplefile3/text1
2021-04-20 19:54:16,666 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
33  33  /samplefile3/text1
```

df

```
hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -df
2021-04-20 19:55:56,239 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Filesystem                  Size   Used     Available  Use%
hdfs://127.0.0.1:9000  267221413888  45056  155517390848    0%
```

```
hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -df -h
2021-04-20 19:56:16,756 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Filesystem              Size  Used  Available  Use%
hdfs://127.0.0.1:9000  248.9 G  44 K    144.8 G    0%
```

count

```
hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -count -v -h /
2021-04-20 20:01:44,154 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
   DIR_COUNT   FILE_COUNT       CONTENT_SIZE PATHNAME
           6            1                 33 /

hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -count -v -h -q /
2021-04-20 20:02:07,036 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
      QUOTA       REM_QUOTA     SPACE_QUOTA REM_SPACE_QUOTA    DIR_COUNT
FILE_COUNT       CONTENT_SIZE PATHNAME
      8.0 E          8.0 E            none             inf            6            1                 33 /

hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -count -v -q /
2021-04-20 20:02:22,339 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
      QUOTA       REM_QUOTA     SPACE_QUOTA REM_SPACE_QUOTA    DIR_COUNT
FILE_COUNT       CONTENT_SIZE PATHNAME
9223372036854775807 9223372036854775800            none             inf            6            1
33 /

hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -count -v -u /
2021-04-20 20:02:33,945 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicabl/e
      QUOTA       REM_QUOTA     SPACE_QUOTA REM_SPACE_QUOTA PATHNAME
9223372036854775807 9223372036854775800            none             inf /
```

PROGRAM – 6

CREATE A MAP REDUCE PROGRAM TO

FIND AVERAGE TEMPERATURE FOR EACH YEAR FROM NCDC DATA SET.

Java Files:

AverageReducer.java

```
package temperature;

import org.apache.hadoop.io.*;

import org.apache.hadoop.mapreduce.*;

import java.io.IOException;

public class AverageReducer extends Reducer <Text, IntWritable,Text, IntWritable >

{

public void reduce(Text key,  Iterable<IntWritable> values, Context context) throws
IOException, InterruptedException

        {

        int max_temp = 0;

        int count = 0;

        for (IntWritable value : values)

                {

                        max_temp += value.get();

                        count+=1;

                }

        context.write(key, new IntWritable(max_temp/count));

        }

}
```

AverageeDriver.java

```
package temperature;
```

```
import org.apache.hadoop.io.*;
import org.apache.hadoop.fs.*;
import org.apache.hadoop.mapreduce.*;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
public class AverageDriver
{
        public static void main (String[] args) throws Exception
        {
                if (args.length != 2)
                {
                        System.err.println("Please Enter the input and output parameters");
                        System.exit(-1);
                }
                Job job = new Job();
                job.setJarByClass(AverageDriver.class);
                job.setJobName("Max temperature");
                FileInputFormat.addInputPath(job,new Path(args[0]));
                FileOutputFormat.setOutputPath(job,new Path (args[1]));
                job.setMapperClass(AverageMapper.class);
                job.setReducerClass(AverageReducer.class);
                job.setOutputKeyClass(Text.class);
                job.setOutputValueClass(IntWritable.class);
                System.exit(job.waitForCompletion(true)?0:1);
        }
}
```

AverageMapper.java:

```
package temperature;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapreduce.*;
import java.io.IOException;
public class AverageMapper extends Mapper <LongWritable, Text, Text, IntWritable>
{
public static final int MISSING = 9999;
public void map(LongWritable key, Text value, Context context) throws
IOException,  InterruptedException
        {
                String line = value.toString();
                String year = line.substring(15,19);
                int temperature;
                if (line.charAt(87)=='+')
                        temperature = Integer.parseInt(line.substring(88, 92));
                else
                        temperature = Integer.parseInt(line.substring(87, 92));

                String quality = line.substring(92, 93);
                if(temperature != MISSING && quality.matches("[01459]"))
                context.write(new Text(year),new IntWritable(temperature));
        }
}
```

Output:

```
Activities  Terminal  May 10 15:54
hdoop@madhu-HP-250-G7-Notebook-PC: ~
        at org.apache.hadoop.util.RunJar.run(RunJar.java:323)
        at org.apache.hadoop.util.RunJar.main(RunJar.java:236)
hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop jar /home/madhu/LABS/Big\ Data\ Analysis/temperature.jar temperature.AverageDriver /temp_input/tem
p.txt /temp_out
2021-05-10 15:35:13,580 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
 applicable
2021-05-10 15:35:14,400 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-05-10 15:35:15,377 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and
 execute your application with ToolRunner to remedy this.
2021-05-10 15:35:15,836 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hdoop/.staging/job_162
0639438976_0002
2021-05-10 15:35:17,011 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-05-10 15:35:18,077 INFO input.FileInputFormat: Total input files to process : 1
2021-05-10 15:35:18,424 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-05-10 15:35:18,915 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-05-10 15:35:19,125 INFO mapreduce.JobSubmitter: number of splits:1
2021-05-10 15:35:20,417 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-05-10 15:35:20,863 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1620639438976_0002
2021-05-10 15:35:20,863 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-05-10 15:35:22,111 INFO conf.Configuration: resource-types.xml not found
2021-05-10 15:35:22,113 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-05-10 15:35:22,334 INFO impl.YarnClientImpl: Submitted application application_1620639438976_0002
2021-05-10 15:35:22,484 INFO mapreduce.Job: The url to track the job: http://madhu-HP-250-G7-Notebook-PC:8088/proxy/application_1620639438976_
0002/
2021-05-10 15:35:22,486 INFO mapreduce.Job: Running job: job_1620639438976_0002
2021-05-10 15:35:40,775 INFO mapreduce.Job: Job job_1620639438976_0002 running in uber mode : false
2021-05-10 15:35:40,780 INFO mapreduce.Job:  map 0% reduce 0%
2021-05-10 15:36:14,723 INFO mapreduce.Job:  map 100% reduce 0%
2021-05-10 15:36:47,309 INFO mapreduce.Job:  map 100% reduce 100%
2021-05-10 15:36:49,352 INFO mapreduce.Job: Job job_1620639438976_0002 completed successfully
2021-05-10 15:36:49,573 INFO mapreduce.Job: Counters: 54
        File System Counters
                FILE: Number of bytes read=72210
                FILE: Number of bytes written=594915
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=894861
```

```
Activities  Terminal  May 10 15:55
hdoop@madhu-HP-250-G7-Notebook-PC: ~
                Input split bytes=113
                Combine input records=0
                Combine output records=0
                Reduce input groups=2
                Reduce shuffle bytes=61
                Reduce input records=5
                Reduce output records=2
                Spilled Records=10
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=121
                CPU time spent (ms)=2680
                Physical memory (bytes) snapshot=441757696
                Virtual memory (bytes) snapshot=5079216128
                Total committed heap usage (bytes)=300941312
                Peak Map Physical memory (bytes)=267333632
                Peak Map Virtual memory (bytes)=2536779776
                Peak Reduce Physical memory (bytes)=174424064
                Peak Reduce Virtual memory (bytes)=2542436352
        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=533
        File Output Format Counters
                Bytes Written=15
hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -cat /temp_out2/*
2021-05-10 15:50:55,745 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
 applicable
2021-05-10 15:50:56,986 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
1949    94
1950    3
hdoop@madhu-HP-250-G7-Notebook-PC:~$ a
```

FIND THE MEAN MAX TEMPERATURE FOR EVERY MONTH:

Java Files:

MaxDriver:

```
package tempMax;
import org.apache.hadoop.io.*;
import org.apache.hadoop.fs.*;
import org.apache.hadoop.mapreduce.*;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
public class MaxDriver
{
    public static void main (String[] args) throws Exception
    {
        if (args.length != 2)
        {
            System.err.println("Please Enter the input and output parameters");
            System.exit(-1);
        }
        Job job = new Job();
        job.setJarByClass(MaxDriver.class);
        job.setJobName("Max temperature");
        FileInputFormat.addInputPath(job,new Path(args[0]));
        FileOutputFormat.setOutputPath(job,new Path (args[1]));
        job.setMapperClass(MaxMapper.class);
        job.setReducerClass(MaxReducer.class);
        job.setOutputKeyClass(Text.class);
        job.setOutputValueClass(IntWritable.class);
        System.exit(job.waitForCompletion(true)?0:1);
```

```
    }
}
```

MaxMapper:

```
package tempMax;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapreduce.*;
import java.io.IOException;
public class MaxMapper extends Mapper <LongWritable, Text, Text, IntWritable>
{
public static final int MISSING = 9999;
public void map(LongWritable key, Text value, Context context) throws IOException,
InterruptedException
{
    String line = value.toString();
    String month = line.substring(19,21);
    int temperature;
    if (line.charAt(87)=='+')
            temperature = Integer.parseInt(line.substring(88, 92));
    else
      temperature = Integer.parseInt(line.substring(87, 92));
    String quality = line.substring(92, 93);
    if(temperature != MISSING && quality.matches("[01459]"))
      context.write(new Text(month),new IntWritable(temperature));
    }
}
```

MaxReducer:

```
package tempMax;
```

```
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapreduce.*;
import java.io.IOException;
public class MaxReducer extends Reducer <Text, IntWritable,Text, IntWritable>
{
    public void reduce(Text key, Iterable<IntWritable> values, Context context) throws
IOException,InterruptedException
    {
        int max_temp = 0;
        for (IntWritable value : values)
        {
            if(max_temp<value.get()) {
                max_temp = value.get();
            }
        }
        context.write(key, new IntWritable(max_temp));
    }
}
```

Output:

```
hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop jar /home/madhu/LABS/Big\ Data\ Analysis/temperature.jar tempMax.MaxDriver /temp_input/sample_temp
.txt /temp_out3
2021-05-10 16:53:21,877 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
 applicable
2021-05-10 16:53:23,289 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-05-10 16:53:24,201 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and
 execute your application with ToolRunner to remedy this.
2021-05-10 16:53:24,322 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hdoop/.staging/job_162
0639438976_0004
2021-05-10 16:53:24,683 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-05-10 16:53:25,362 INFO input.FileInputFormat: Total input files to process : 1
2021-05-10 16:53:25,574 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-05-10 16:53:25,753 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-05-10 16:53:25,830 INFO mapreduce.JobSubmitter: number of splits:1
2021-05-10 16:53:26,208 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-05-10 16:53:26,331 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1620639438976_0004
2021-05-10 16:53:26,331 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-05-10 16:53:26,639 INFO conf.Configuration: resource-types.xml not found
2021-05-10 16:53:26,639 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-05-10 16:53:26,709 INFO impl.YarnClientImpl: Submitted application application_1620639438976_0004
2021-05-10 16:53:26,754 INFO mapreduce.Job: The url to track the job: http://madhu-HP-250-G7-Notebook-PC:8088/proxy/application_1620639438976_
0004/
2021-05-10 16:53:26,754 INFO mapreduce.Job: Running job: job_1620639438976_0004
2021-05-10 16:53:38,177 INFO mapreduce.Job: Job job_1620639438976_0004 running in uber mode : false
2021-05-10 16:53:38,198 INFO mapreduce.Job:  map 0% reduce 0%
2021-05-10 16:53:49,426 INFO mapreduce.Job:  map 100% reduce 0%
2021-05-10 16:54:01,554 INFO mapreduce.Job:  map 100% reduce 100%
2021-05-10 16:54:04,613 INFO mapreduce.Job: Job job_1620639438976_0004 completed successfully
2021-05-10 16:54:04,798 INFO mapreduce.Job: Counters: 54
        File System Counters
                FILE: Number of bytes read=51
                FILE: Number of bytes written=450581
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=646
                HDFS: Number of bytes written=13
                HDFS: Number of read operations=8
```

```
                Merged Map outputs=1
                GC time elapsed (ms)=113
                CPU time spent (ms)=2890
                Physical memory (bytes) snapshot=456769536
                Virtual memory (bytes) snapshot=5081296896
                Total committed heap usage (bytes)=348127232
                Peak Map Physical memory (bytes)=279945216
                Peak Map Virtual memory (bytes)=2533858176
                Peak Reduce Physical memory (bytes)=176824320
                Peak Reduce Virtual memory (bytes)=2545438720
        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=894755
        File Output Format Counters
                Bytes Written=80
hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -cat /temp_out4/*
2021-05-10 16:56:14,221 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
 applicable
2021-05-10 16:56:15,144 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
01      44
02      17
03      50
04      194
05      256
06      278
07      317
08      283
09      211
10      156
11      89
12      117
hdoop@madhu-HP-250-G7-Notebook-PC:~$
```

PROGRAM – 7

FOR A GIVEN TEXT FILE, CREATE A MAP REDUCE PROGRAM TO SORT THE CONTENT IN AN ALPHABETIC ORDER LISTING ONLY TOP 10 MAXIMUM OCCURRENCES OF WORDS:

Java Files:

TopN.java:

```
package sortwords;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.util.GenericOptionsParser;
import utils.MiscUtils;
import java.io.IOException;
import java.util.*;
public class TopN {
    public static void main(String[] args) throws Exception {
        Configuration conf = new Configuration();
        String[] otherArgs = new GenericOptionsParser(conf, args).getRemainingArgs();
        if (otherArgs.length != 2) {
            System.err.println("Usage: TopN <in> <out>");
            System.exit(2);
        }
        Job job = Job.getInstance(conf);
```

```
        job.setJobName("Top N");
        job.setJarByClass(TopN.class);
        job.setMapperClass(TopNMapper.class);
        job.setReducerClass(TopNReducer.class);
        job.setOutputKeyClass(Text.class);
        job.setOutputValueClass(IntWritable.class);
        FileInputFormat.addInputPath(job, new Path(otherArgs[0]));
        FileOutputFormat.setOutputPath(job, new Path(otherArgs[1]));
        System.exit(job.waitForCompletion(true) ? 0 : 1);
    }
    public static class TopNMapper extends Mapper<Object, Text, Text, IntWritable> {
        private final static IntWritable one = new IntWritable(1);
        private Text word = new Text();
        private String tokens = "[_|$#<>\\^=\\[\\]\\*/\\\\,;,.\\-:()?!\"']";
        @Override
        public void map(Object key, Text value, Context context) throws IOException,
InterruptedException {
            String cleanLine = value.toString().toLowerCase().replaceAll(tokens, " ");
            StringTokenizer itr = new StringTokenizer(cleanLine);
            while (itr.hasMoreTokens()) {
                word.set(itr.nextToken().trim());
                context.write(word, one);
            }
        }
    }
    public static class TopNReducer extends Reducer<Text, IntWritable, Text, IntWritable> {
        private Map<Text, IntWritable> countMap = new HashMap<>();
        @Override
```

```
        public void reduce(Text key, Iterable<IntWritable> values, Context context) throws
IOException, InterruptedException {
            int sum = 0;
            for (IntWritable val : values) {
                sum += val.get();
            }
            countMap.put(new Text(key), new IntWritable(sum));
        }
        @Override
        protected void cleanup(Context context) throws IOException, InterruptedException {
            Map<Text, IntWritable> sortedMap = MiscUtils.sortByValues(countMap);
            int counter = 0;
            for (Text key : sortedMap.keySet()) {
                if (counter++ == 3) {
                    break;
                }
                context.write(key, sortedMap.get(key));
            }
        }
    }
    public static class TopNCombiner extends Reducer<Text, IntWritable, Text, IntWritable> {
        @Override
        public void reduce(Text key, Iterable<IntWritable> values, Context context) throws
IOException, InterruptedException {
            int sum = 0;
            for (IntWritable val : values) {
                sum += val.get();
            }
            context.write(key, new IntWritable(sum));
```

```
        }
    }
}
```

MiscUtils.java

```
package utils;
import java.util.*;
public class MiscUtils {
    public static <K extends Comparable, V extends Comparable> Map<K, V>
sortByValues(Map<K, V> map) {
        List<Map.Entry<K, V>> entries = new LinkedList<Map.Entry<K, V>>(map.entrySet());
        Collections.sort(entries, new Comparator<Map.Entry<K, V>>() {
            @Override
            public int compare(Map.Entry<K, V> o1, Map.Entry<K, V> o2) {
                return o2.getValue().compareTo(o1.getValue());
            }
        });
        Map<K, V> sortedMap = new LinkedHashMap<K, V>();
        for (Map.Entry<K, V> entry : entries) {
            sortedMap.put(entry.getKey(), entry.getValue());
        }
        return sortedMap;
    }
}
```

Output:

```
hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -cat /input/inputFile.txt
2021-05-03 15:48:03,486 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
 applicable
2021-05-03 15:48:05,243 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
Mother  Father  Car     Office  Father
Sister  Mother  Bike    Car     Travel
Father  Travel  Car     Travel  Home
Sister  College Brother Bike    Father
Brother Car     Bike    College Home
hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop jar /home/madhu/LABS/Big\ Data\ Analysis/sortwords.jar sortwords.TopN /input /output/
2021-05-03 15:49:18,042 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
 applicable
2021-05-03 15:49:19,305 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-05-03 15:49:20,735 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hdoop/.staging/job_162
0036386996_0001
2021-05-03 15:49:21,352 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-05-03 15:49:23,658 INFO input.FileInputFormat: Total input files to process : 1
2021-05-03 15:49:23,869 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-05-03 15:49:23,994 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-05-03 15:49:24,093 INFO mapreduce.JobSubmitter: number of splits:1
2021-05-03 15:49:25,567 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-05-03 15:49:25,651 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1620036386996_0001
2021-05-03 15:49:25,652 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-05-03 15:49:26,373 INFO conf.Configuration: resource-types.xml not found
2021-05-03 15:49:26,373 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-05-03 15:49:27,358 INFO impl.YarnClientImpl: Submitted application application_1620036386996_0001
2021-05-03 15:49:27,644 INFO mapreduce.Job: The url to track the job: http://madhu-HP-250-G7-Notebook-PC:8088/proxy/application_1620036386996_
0001/
2021-05-03 15:49:27,645 INFO mapreduce.Job: Running job: job_1620036386996_0001
2021-05-03 15:50:15,043 INFO mapreduce.Job: Job job_1620036386996_0001 running in uber mode : false
2021-05-03 15:50:15,045 INFO mapreduce.Job:  map 0% reduce 0%
2021-05-03 15:51:25,361 INFO mapreduce.Job:  map 100% reduce 0%
        2021-05-03 15:52:07,054 INFO mapreduce.Job:  map 100% reduce 100%
2021-05-03 15:52:11,118 INFO mapreduce.Job: Job job_1620036386996_0001 completed successfully
2021-05-03 15:52:11,939 INFO mapreduce.Job: Counters: 54
        File System Counters
                FILE: Number of bytes read=313
                FILE: Number of bytes written=451415
```

```
                Reduce output records=3
                Spilled Records=50
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=298
                CPU time spent (ms)=3040
                Physical memory (bytes) snapshot=401829888
                Virtual memory (bytes) snapshot=5074657280
                Total committed heap usage (bytes)=273678336
                Peak Map Physical memory (bytes)=228569088
                Peak Map Virtual memory (bytes)=2533502976
                Peak Reduce Physical memory (bytes)=173260800
                Peak Reduce Virtual memory (bytes)=2541154304
        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=160
        File Output Format Counters
                Bytes Written=24
hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -ls /output
]2021-05-03 15:57:29,781 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes wher
e applicable
Found 2 items
-rw-r--r--   1 hdoop supergroup          0 2021-05-03 15:52 /output/_SUCCESS
-rw-r--r--   1 hdoop supergroup         24 2021-05-03 15:52 /output/part-r-00000
hdoop@madhu-HP-250-G7-Notebook-PC:~$ hadoop fs -cat /output/part-r-00000
2021-05-03 15:58:35,442 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
 applicable
2021-05-03 15:58:36,532 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
car     4
father  4
travel  3
```

PROGRAM – 8

CREATE A MAP REDUCE PROGRAM TO DEMONSTRATING JOIN OPERATION:

Java Files:

JoinDriver.java

```
package DatasetJoin;
import org.apache.hadoop.conf.Configured;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.*;
import org.apache.hadoop.mapred.lib.MultipleInputs;
import org.apache.hadoop.util.*;
public class JoinDriver extends Configured implements Tool {
    public static class KeyPartitioner implements Partitioner<TextPair, Text> {
        @Override
        public void configure(JobConf job) {}
        @Override
        public int getPartition(TextPair key, Text value, int numPartitions) {
                return (key.getFirst().hashCode() & Integer.MAX_VALUE) % numPartitions;
        }
    }
    @Override
    public int run(String[] args) throws Exception {
        if (args.length != 3) {
                System.out.println("Usage: <Department Emp Strength input> <Department
Name input> <output>");
                return -1;
        }
```

```
        JobConf conf = new JobConf(getConf(), getClass());
        conf.setJobName("Join 'Department Emp Strength input' with 'Department Name
input'");
        Path AInputPath = new Path(args[0]);
        Path BInputPath = new Path(args[1]);
        Path outputPath = new Path(args[2]);
        MultipleInputs.addInputPath(conf, AInputPath, TextInputFormat.class, Posts.class);
        MultipleInputs.addInputPath(conf, BInputPath, TextInputFormat.class, Users.class);
        FileOutputFormat.setOutputPath(conf, outputPath);
        conf.setPartitionerClass(KeyPartitioner.class);
        conf.setOutputValueGroupingComparator(TextPair.FirstComparator.class);
        conf.setMapOutputKeyClass(TextPair.class);
        conf.setReducerClass(JoinReducer.class);
        conf.setOutputKeyClass(Text.class);
        JobClient.runJob(conf);
        return 0;
    }
    public static void main(String[] args) throws Exception {
        int exitCode = ToolRunner.run(new JoinDriver(), args);
        System.exit(exitCode);
    }
}
```

JoinReducer.java

```
package DatasetJoin;
import java.io.IOException;
import java.util.Iterator;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.*;
```

```
public class JoinReducer extends MapReduceBase implements Reducer<TextPair, Text, Text, Text> {
    @Override
    public void reduce (TextPair key, Iterator<Text> values, OutputCollector<Text, Text> output, Reporter reporter)
        throws IOException
    {
        Text nodeId = new Text(values.next());
        while (values.hasNext()) {
            Text node = values.next();
            Text outValue = new Text(nodeId.toString() + "\t\t" + node.toString());
            output.collect(key.getFirst(), outValue);
        }
    }
}
```

Posts.java

```
package DatasetJoin;
import java.io.IOException;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapred.*;
public class Posts extends MapReduceBase implements Mapper<LongWritable, Text, TextPair, Text>  {
    @Override
    public void map(LongWritable key, Text value, OutputCollector<TextPair, Text> output, Reporter reporter)
            throws IOException
    {
        String valueString = value.toString();
        String[] SingleNodeData = valueString.split("\t");
```

```
            output.collect(new TextPair(SingleNodeData[3], "0"), new Text(SingleNodeData[9]));
        }
}
```

Users.java

```
package DatasetJoin;
import java.io.IOException;
import java.util.Iterator;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.FSDataInputStream;
import org.apache.hadoop.fs.FSDataOutputStream;
import org.apache.hadoop.fs.FileSystem;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.*;
import org.apache.hadoop.io.IntWritable;
public class Users extends MapReduceBase implements Mapper<LongWritable, Text, TextPair,
Text> {
        @Override
        public void map(LongWritable key, Text value, OutputCollector<TextPair, Text> output,
Reporter reporter)
            throws IOException
        {
            String valueString = value.toString();
            String[] SingleNodeData = valueString.split("\t");
            output.collect(new TextPair(SingleNodeData[0], "1"), new Text(SingleNodeData[1]));
        }
}
```

TextPair.java

```
package DatasetJoin;
import java.io.*;
import org.apache.hadoop.io.*;
public class TextPair implements WritableComparable<TextPair> {
  private Text first;
  private Text second;
  public TextPair() {
    set(new Text(), new Text());
  }
  public TextPair(String first, String second) {
    set(new Text(first), new Text(second));
  }
  public TextPair(Text first, Text second) {
    set(first, second);
  }
  public void set(Text first, Text second) {
     this.first = first;
     this.second = second;
  }
  public Text getFirst() {
    return first;
  }
  public Text getSecond() {
    return second;
  }
  @Override
  public void write(DataOutput out) throws IOException {
```

```
  first.write(out);
  second.write(out);
}
@Override
public void readFields(DataInput in) throws IOException {
  first.readFields(in);
  second.readFields(in);
}
@Override
public int hashCode() {
  return first.hashCode() * 163 + second.hashCode();
}
@Override
public boolean equals(Object o) {
  if (o instanceof TextPair) {
    TextPair tp = (TextPair) o;
    return first.equals(tp.first) && second.equals(tp.second);
  }
  return false;
}
@Override
public String toString() {
  return first + "\t" + second;
}
@Override
public int compareTo(TextPair tp) {
  int cmp = first.compareTo(tp.first);
  if (cmp != 0) {
```

```
    return cmp;
  }
  return second.compareTo(tp.second);
}
public static class Comparator extends WritableComparator {
  private static final Text.Comparator TEXT_COMPARATOR = new Text.Comparator();
  public Comparator() {
    super(TextPair.class);
  }
  @Override
  public int compare(byte[] b1, int s1, int l1, byte[] b2, int s2, int l2) {
    try {
      int firstL1 = WritableUtils.decodeVIntSize(b1[s1]) + readVInt(b1, s1);
      int firstL2 = WritableUtils.decodeVIntSize(b2[s2]) + readVInt(b2, s2);
      int cmp = TEXT_COMPARATOR.compare(b1, s1, firstL1, b2, s2, firstL2);
      if (cmp != 0) {
        return cmp;
      }
      return TEXT_COMPARATOR.compare(b1, s1 + firstL1, l1 - firstL1,
                                     b2, s2 + firstL2, l2 - firstL2);
    } catch (IOException e) {
      throw new IllegalArgumentException(e);
    }
  }
}
static {
  WritableComparator.define(TextPair.class, new Comparator());
}
```

```
  public static class FirstComparator extends WritableComparator {
    private static final Text.Comparator TEXT_COMPARATOR = new Text.Comparator();
    public FirstComparator() {
      super(TextPair.class);
    }
    @Override
    public int compare(byte[] b1, int s1, int l1,  byte[] b2, int s2, int l2) {
       try {
        int firstL1 = WritableUtils.decodeVIntSize(b1[s1]) + readVInt(b1, s1);
        int firstL2 = WritableUtils.decodeVIntSize(b2[s2]) + readVInt(b2, s2);
        return TEXT_COMPARATOR.compare(b1, s1, firstL1, b2, s2, firstL2);
      } catch (IOException e) {
        throw new IllegalArgumentException(e);
      }
    }
    @Override
    public int compare(WritableComparable a, WritableComparable b) {
      if (a instanceof TextPair && b instanceof TextPair) {
        return ((TextPair) a).first.compareTo(((TextPair) b).first);
      }
      return super.compare(a, b);
    }
  }
}
```

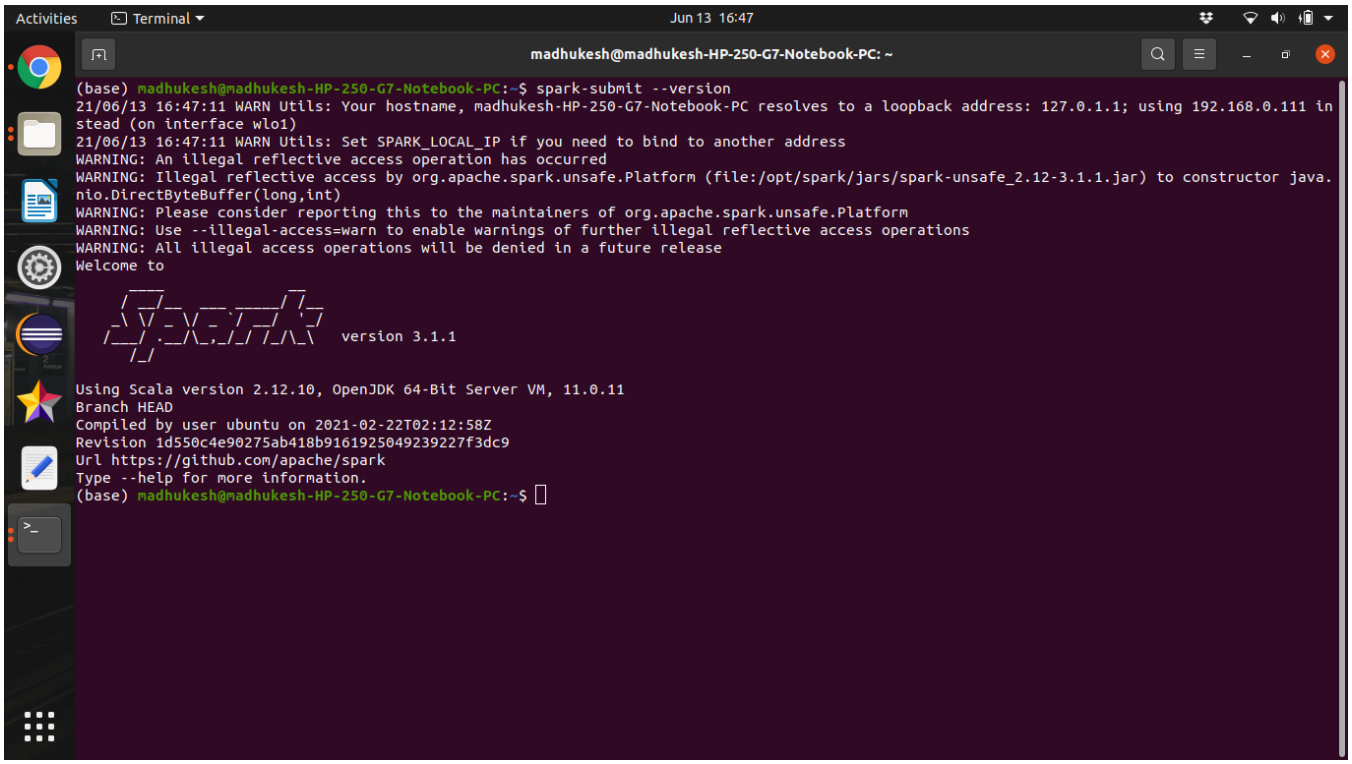
Output :

```
hdoop@madhukesh-HP-250-G7-Notebook-PC:~$ hadoop jar /home/madhukesh/SIXTH\ SEMESTER/LABS/Big\ Data\ Analysis/DatasetJoins.jar DatasetJoin.Join
Driver /post_input /user_input /joins_output/
2021-06-13 16:15:24,437 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
 applicable
2021-06-13 16:15:25,144 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-06-13 16:15:25,418 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-06-13 16:15:25,708 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hdoop/.staging/job_162
3580806936_0001
2021-06-13 16:15:25,906 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-06-13 16:15:26,670 INFO mapred.FileInputFormat: Total input files to process : 1
2021-06-13 16:15:26,702 INFO mapred.FileInputFormat: Total input files to process : 1
2021-06-13 16:15:26,777 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-06-13 16:15:26,846 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-06-13 16:15:27,278 INFO mapreduce.JobSubmitter: number of splits:4
2021-06-13 16:15:27,811 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-06-13 16:15:28,235 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1623580806936_0001
2021-06-13 16:15:28,235 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-06-13 16:15:28,514 INFO conf.Configuration: resource-types.xml not found
2021-06-13 16:15:28,514 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-06-13 16:15:28,786 INFO impl.YarnClientImpl: Submitted application application_1623580806936_0001
2021-06-13 16:15:28,843 INFO mapreduce.Job: The url to track the job: http://madhukesh-HP-250-G7-Notebook-PC:8088/proxy/application_1623580806
936_0001/
2021-06-13 16:15:28,846 INFO mapreduce.Job: Running job: job_1623580806936_0001
2021-06-13 16:15:38,008 INFO mapreduce.Job: Job job_1623580806936_0001 running in uber mode : false
2021-06-13 16:15:38,011 INFO mapreduce.Job:  map 0% reduce 0%
2021-06-13 16:15:52,338 INFO mapreduce.Job:  map 100% reduce 0%
2021-06-13 16:16:14,145 INFO mapreduce.Job:  map 100% reduce 100%
2021-06-13 16:16:16,173 INFO mapreduce.Job: Job job_1623580806936_0001 completed successfully
2021-06-13 16:16:16,426 INFO mapreduce.Job: Counters: 54
        File System Counters
                FILE: Number of bytes read=155
                FILE: Number of bytes written=1131403
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=2665
                HDFS: Number of bytes written=71
                HDFS: Number of read operations=17
```

```
                Spilled Records=14
                Shuffled Maps =4
                Failed Shuffles=0
                Merged Map outputs=4
                GC time elapsed (ms)=1016
                CPU time spent (ms)=6410
                Physical memory (bytes) snapshot=1224622080
                Virtual memory (bytes) snapshot=12675289088
                Total committed heap usage (bytes)=1072168960
                Peak Map Physical memory (bytes)=276529152
                Peak Map Virtual memory (bytes)=2535211008
                Peak Reduce Physical memory (bytes)=153554944
                Peak Reduce Virtual memory (bytes)=2541469696
        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=0
        File Output Format Counters
                Bytes Written=71
hdoop@madhukesh-HP-250-G7-Notebook-PC:~$ hadoop fs -ls /joins_output
2021-06-13 16:21:44,490 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
 applicable
Found 2 items
-rw-r--r--   1 hdoop supergroup          0 2021-06-13 16:16 /joins_output/_SUCCESS
-rw-r--r--   1 hdoop supergroup         71 2021-06-13 16:16 /joins_output/part-00000
hdoop@madhukesh-HP-250-G7-Notebook-PC:~$ hadoop fs -cat /joins_output/part-00000
2021-06-13 16:22:19,641 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
 applicable
2021-06-13 16:22:20,543 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
"100005361"     "2"             "36134"
"100018705"     "2"             "76"
"100022094"     "0"             "6354"
hdoop@madhukesh-HP-250-G7-Notebook-PC:~$
```

PROGRAM – 9

SCREENSHOT OF SPARK INSTALLED:



PROGRAM – 10

USING RDD AND FLAMAP COUNT HOW MANY TIMES EACH WORD APPEARS IN A FILE AND WRITE OUT A LIST OF WORDS WHOSE COUNT IS STRICTLY GREATER THAN 4 USING SPARK:

Input:

Mother Father Car Office Father

Sister Mother Bike Car Travel

Father Travel Car Travel Home

Sister College Brother Bike Father

Brother Car Bike College Home

Car Father Travel Bike Bike

Output: