VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"JnanaSangama", Belgaum -590014, Karnataka.

LAB REPORT on

BIG DATA ANALYTICS
(23CS6PEBDA)

Submitted by

AMRUTHA MURALIDHAR (1BM21CS257)

in partial fulfilment for the award of the degree of
BACHELOR OF ENGINEERING
in
COMPUTER SCIENCE AND ENGINEERING

B.M.S. COLLEGE OF ENGINEERING BENGALURU-560019 May-2022 to July-2022

(Autonomous Institution under VTU)

B. M. S. College of Engineering,

Bull Temple Road, Bangalore 560019

Department of Computer Science and Engineering

<u>CERTIFICATE</u>

This is to certify that the Lab work entitled "**BIG DATA ANALYTICS** " carried out by **Amrutha Muralidhar (1BM21CS257),** who is a bonafide student of **B. M. S. College of Engineering.** It is in partial fulfillment for the award of **Bachelor of Engineering in Computer Science and Engineering** of the Visvesvaraya Technological University, Belgaum during the year 2023. The Lab report has been approved as it satisfies the academic requirements in respect of Big data analytics - (23CS6PEBDA) work prescribed for the said degree.

Lab-In charge
Rekha G S
Assistant Professor
Department of CSE
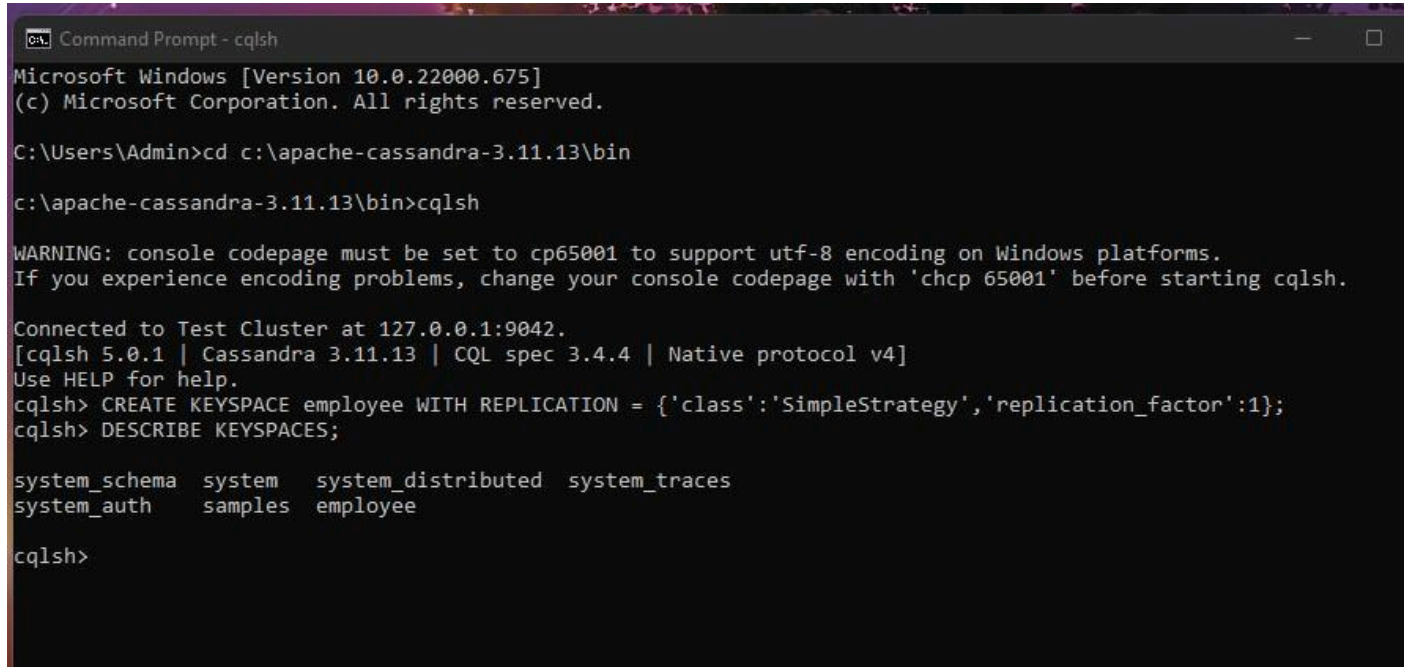BMSCE, Bengaluru

`

Index

Course Outcome

CO1	Apply the concept of NoSQL, Hadoop or Spark for a given task
CO2	Analyze the Big Data and obtain insight using data analytics mechanisms.
CO3	Design and implement Big data applications by applying NoSQL, Hadoop or Spark

Cassandra Lab Program 1: -

Perform the following DB operations using Cassandra.

1. Create a key space by name Employee

```
Command Prompt - cqlsh                                                    —    □
Microsoft Windows [Version 10.0.22000.675]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Admin>cd c:\apache-cassandra-3.11.13\bin

c:\apache-cassandra-3.11.13\bin>cqlsh

WARNING: console codepage must be set to cp65001 to support utf-8 encoding on Windows platforms.
If you experience encoding problems, change your console codepage with 'chcp 65001' before starting cqlsh.

Connected to Test Cluster at 127.0.0.1:9042.
[cqlsh 5.0.1 | Cassandra 3.11.13 | CQL spec 3.4.4 | Native protocol v4]
Use HELP for help.
cqlsh> CREATE KEYSPACE employee WITH REPLICATION = {'class':'SimpleStrategy','replication_factor':1};
cqlsh> DESCRIBE KEYSPACES;

system_schema   system    system_distributed   system_traces
system_auth     samples   employee

cqlsh>
```
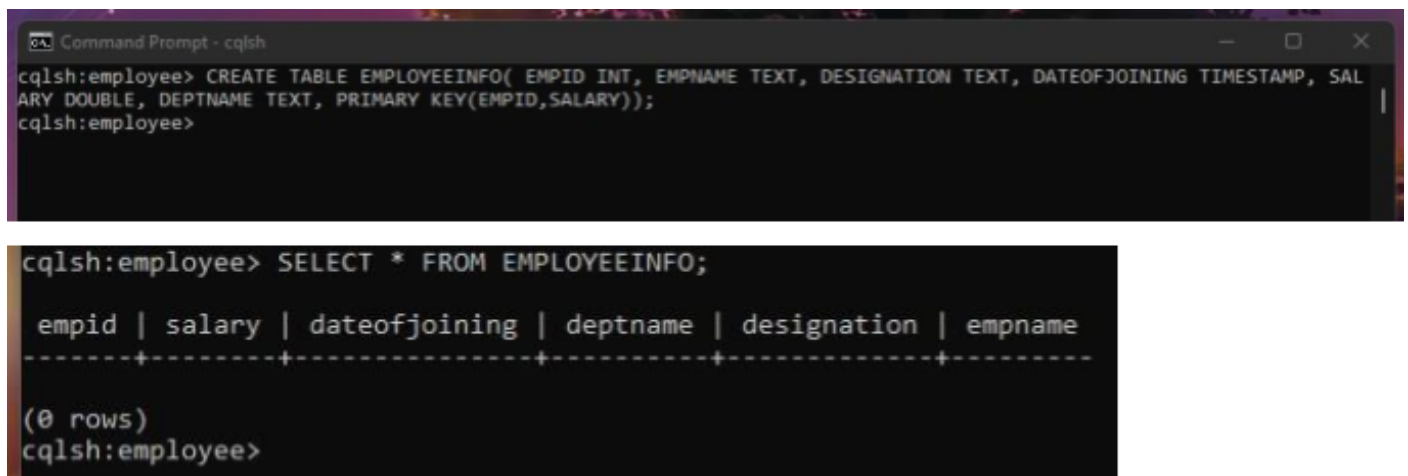
2. Create a column family by name Employee-Info with attributes Emp_Id Primary Key, Emp_Name,

   Designation, Date_of_Joining, Salary, Dept_Name

```
Command Prompt - cqlsh                                                    —    □    X
cqlsh:employee> CREATE TABLE EMPLOYEEINFO( EMPID INT, EMPNAME TEXT, DESIGNATION TEXT, DATEOFJOINING TIMESTAMP, SAL
ARY DOUBLE, DEPTNAME TEXT, PRIMARY KEY(EMPID,SALARY));
cqlsh:employee>
```

```
cqlsh:employee> SELECT * FROM EMPLOYEEINFO;

 empid | salary | dateofjoining | deptname | designation | empname
-------+--------+---------------+----------+-------------+--------

(0 rows)
cqlsh:employee>
```

3. Insert the values into the table in batch

```
Command Prompt - cqlsh                                                                    —  ☐

cqlsh:employee> BEGIN BATCH
          ... INSERT INTO EMPLOYEEINFO (EMPID, EMPNAME, DESIGNATION, DATEOFJOINING, SALARY, DEPTNAME)
          ... VALUES(1,'LOKESH','ASSISTANT MANAGER', '2005-04-6', 50000, 'MARKETING')
          ... INSERT INTO EMPLOYEEINFO (EMPID, EMPNAME, DESIGNATION, DATEOFJOINING, SALARY, DEPTNAME)
          ... VALUES(2,'DHEERAJ','ASSISTANT MANAGER', '2013-11-10', 30000, 'LOGISTICS')
          ... INSERT INTO EMPLOYEEINFO (EMPID, EMPNAME, DESIGNATION, DATEOFJOINING, SALARY, DEPTNAME)
          ... VALUES(3,'CHIRAG','ASSISTANT MANAGER', '2011-07-1', 115000, 'SALES')
          ... INSERT INTO EMPLOYEEINFO (EMPID, EMPNAME, DESIGNATION, DATEOFJOINING, SALARY, DEPTNAME)
          ... VALUES(4,'DHANUSH','ASSISTANT MANAGER', '2010-04-26', 75000, 'MARKETING')
          ...  INSERT INTO EMPLOYEEINFO (EMPID, EMPNAME, DESIGNATION, DATEOFJOINING, SALARY, DEPTNAME)
          ... VALUES(5,'ESHA','ASSISTANT MANAGER', '2010-04-26', 85000, 'TECHNICAL')
          ... INSERT INTO EMPLOYEEINFO (EMPID, EMPNAME, DESIGNATION, DATEOFJOINING, SALARY, DEPTNAME)
          ... VALUES(6,'FARHAN','MANAGER', '2010-04-26', 95000, 'TECHNICAL')
          ... INSERT INTO EMPLOYEEINFO (EMPID, EMPNAME, DESIGNATION, DATEOFJOINING, SALARY, DEPTNAME)
          ... VALUES(7,'JIMMY','MANAGER', '2010-04-26', 95000, 'PR')
          ... INSERT INTO EMPLOYEEINFO (EMPID, EMPNAME, DESIGNATION, DATEOFJOINING, SALARY, DEPTNAME)
          ... VALUES(121,'HARRY','REGIONAL MANAGER', '2010-04-26', 99000, 'MANAGEMENT')
          ... APPLY BATCH;

cqlsh:employee>  SELECT * FROM EMPLOYEEINFO;

 empid | salary   | dateofjoining                    | deptname   | designation       | empname
-------+----------+----------------------------------+------------+-------------------+----------
     5 |    85000 | 2010-04-25 18:30:00.000000+0000  | TECHNICAL  | ASSISTANT MANAGER |    ESHA
     1 |    50000 | 2005-04-05 18:30:00.000000+0000  | MARKETING  | ASSISTANT MANAGER |  LOKESH
     2 |    30000 | 2013-11-09 18:30:00.000000+0000  | LOGISTICS  | ASSISTANT MANAGER | DHEERAJ
     4 |    75000 | 2010-04-25 18:30:00.000000+0000  | MARKETING  | ASSISTANT MANAGER | DHANUSH
   121 |    99000 | 2010-04-25 18:30:00.000000+0000  | MANAGEMENT |  REGIONAL MANAGER |   HARRY
     7 |    95000 | 2010-04-25 18:30:00.000000+0000  |         PR |           MANAGER |   JIMMY
     6 |    95000 | 2010-04-25 18:30:00.000000+0000  | TECHNICAL  |           MANAGER |  FARHAN
     3 | 1.15e+05 | 2011-06-30 18:30:00.000000+0000  |      SALES | ASSISTANT MANAGER |  CHIRAG

(8 rows)
cqlsh:employee>
```

4. Update Employee name and Department of Emp-Id 121

```
cqlsh:employee> UPDATE EMPLOYEEINFO SET EMPNAME='HARRY', DEPTNAME='MANAGEMENT' WHERE EMPID=121 AND SALARY=99000;
cqlsh:employee>  SELECT * FROM EMPLOYEEINFO;

 empid | salary   | dateofjoining                    | deptname   | designation       | empname
-------+----------+----------------------------------+------------+-------------------+----------
     5 |    85000 | 2010-04-25 18:30:00.000000+0000  | TECHNICAL  | ASSISTANT MANAGER |    ESHA
     1 |    50000 | 2005-04-05 18:30:00.000000+0000  | MARKETING  | ASSISTANT MANAGER |  LOKESH
     2 |    30000 | 2013-11-09 18:30:00.000000+0000  | LOGISTICS  | ASSISTANT MANAGER | DHEERAJ
     4 |    75000 | 2010-04-25 18:30:00.000000+0000  | MARKETING  | ASSISTANT MANAGER | DHANUSH
   121 |    99000 | 2010-04-25 18:30:00.000000+0000  | MANAGEMENT |  REGIONAL MANAGER |   HARRY
     7 |    95000 | 2010-04-25 18:30:00.000000+0000  |         PR |           MANAGER |   JIMMY
     6 |    95000 | 2010-04-25 18:30:00.000000+0000  | TECHNICAL  |           MANAGER |  FARHAN
     3 | 1.15e+05 | 2011-06-30 18:30:00.000000+0000  |      SALES | ASSISTANT MANAGER |  CHIRAG

(8 rows)
cqlsh:employee> _
```

5. Sort the details of Employee records based on salary (Note:- cql>PAGING OFF)

```
cqlsh:employee> select * from EMPLOYEEINFO where empid IN(1,2,3,4,5,6,7) ORDER BY salary DESC allow filtering;

 empid | salary   | dateofjoining                   | deptname  | designation       | empname
-------+----------+---------------------------------+-----------+-------------------+---------
     3 | 1.15e+05 | 2011-06-30 18:30:00.000000+0000 |     SALES | ASSISTANT MANAGER |  CHIRAG
     6 |    95000 | 2010-04-25 18:30:00.000000+0000 | TECHNICAL |           MANAGER |  FARHAN
     7 |    95000 | 2010-04-25 18:30:00.000000+0000 |        PR |           MANAGER |   JIMMY
     5 |    85000 | 2010-04-25 18:30:00.000000+0000 | TECHNICAL | ASSISTANT MANAGER |    ESHA
     4 |    75000 | 2010-04-25 18:30:00.000000+0000 | MARKETING | ASSISTANT MANAGER | DHANUSH
     1 |    50000 | 2005-04-05 18:30:00.000000+0000 | MARKETING | ASSISTANT MANAGER |  LOKESH
     2 |    30000 | 2013-11-09 18:30:00.000000+0000 | LOGISTICS | ASSISTANT MANAGER | DHEERAJ

(7 rows)
cqlsh:employee>
```

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

```
(7 rows)
cqlsh:employee> ALTER TABLE EMPLOYEEINFO ADD PROJECTS LIST<TEXT>;
cqlsh:employee> SELECT * FROM EMPLOYEEINFO;

 empid | salary   | dateofjoining                   | deptname   | designation       | empname | projects
-------+----------+---------------------------------+------------+-------------------+---------+----------
     5 |    85000 | 2010-04-25 18:30:00.000000+0000 |  TECHNICAL | ASSISTANT MANAGER |    ESHA |     null
     1 |    50000 | 2005-04-05 18:30:00.000000+0000 |  MARKETING | ASSISTANT MANAGER |  LOKESH |     null
     2 |    30000 | 2013-11-09 18:30:00.000000+0000 |  LOGISTICS | ASSISTANT MANAGER | DHEERAJ |     null
     4 |    75000 | 2010-04-25 18:30:00.000000+0000 |  MARKETING | ASSISTANT MANAGER | DHANUSH |     null
   121 |    99000 | 2010-04-25 18:30:00.000000+0000 | MANAGEMENT |  REGIONAL MANAGER |   HARRY |     null
     7 |    95000 | 2010-04-25 18:30:00.000000+0000 |         PR |           MANAGER |   JIMMY |     null
     6 |    95000 | 2010-04-25 18:30:00.000000+0000 |  TECHNICAL |           MANAGER |  FARHAN |     null
     3 | 1.15e+05 | 2011-06-30 18:30:00.000000+0000 |      SALES | ASSISTANT MANAGER |  CHIRAG |     null

(8 rows)
cqlsh:employee>
```

7. Update the altered table to add project names.

```
Command Prompt - cqlsh                                                                          —    □

cqlsh:employee> UPDATE EMPLOYEEINFO SET PROJECTS=['FACEBOOK','SNAPCHAT'] WHERE EMPID=1 AND SALARY=50000;
cqlsh:employee> UPDATE EMPLOYEEINFO SET PROJECTS=['FACEBOOK','SNAPCHAT'] WHERE EMPID=7 AND SALARY=95000;
cqlsh:employee>  UPDATE EMPLOYEEINFO SET PROJECTS=['PINTEREST','INSTAGRAM'] WHERE EMPID=121 AND SALARY=99000;
cqlsh:employee> UPDATE EMPLOYEEINFO SET PROJECTS=['PINTEREST','INSTAGRAM'] WHERE EMPID=4 AND SALARY=75000;
cqlsh:employee> UPDATE EMPLOYEEINFO SET PROJECTS=['YOUTUBE','SPOTIFY'] WHERE EMPID=2 AND SALARY=30000;
cqlsh:employee> UPDATE EMPLOYEEINFO SET PROJECTS=['YOUTUBE','SPOTIFY'] WHERE EMPID=3 AND SALARY=115000;
cqlsh:employee> UPDATE EMPLOYEEINFO SET PROJECTS=['YOUTUBE','SPOTIFY'] WHERE EMPID=6 AND SALARY=95000;
cqlsh:employee> UPDATE EMPLOYEEINFO SET PROJECTS=['YOUTUBE','SPOTIFY'] WHERE EMPID=5 AND SALARY=85000;
cqlsh:employee> SELECT * FROM EMPLOYEEINFO;

 empid | salary   | dateofjoining                   | deptname   | designation       | empname | projects
-------+----------+---------------------------------+------------+-------------------+---------+----------------------------
     5 |    85000 | 2010-04-25 18:30:00.000000+0000 |  TECHNICAL | ASSISTANT MANAGER |    ESHA |     ['YOUTUBE', 'SPOTIFY']
     1 |    50000 | 2005-04-05 18:30:00.000000+0000 |  MARKETING | ASSISTANT MANAGER |  LOKESH |   ['FACEBOOK', 'SNAPCHAT']
     2 |    30000 | 2013-11-09 18:30:00.000000+0000 |  LOGISTICS | ASSISTANT MANAGER | DHEERAJ |     ['YOUTUBE', 'SPOTIFY']
     4 |    75000 | 2010-04-25 18:30:00.000000+0000 |  MARKETING | ASSISTANT MANAGER | DHANUSH | ['PINTEREST', 'INSTAGRAM']
   121 |    99000 | 2010-04-25 18:30:00.000000+0000 | MANAGEMENT |  REGIONAL MANAGER |   HARRY | ['PINTEREST', 'INSTAGRAM']
     7 |    95000 | 2010-04-25 18:30:00.000000+0000 |         PR |           MANAGER |   JIMMY |   ['FACEBOOK', 'SNAPCHAT']
     6 |    95000 | 2010-04-25 18:30:00.000000+0000 |  TECHNICAL |           MANAGER |  FARHAN |     ['YOUTUBE', 'SPOTIFY']
     3 | 1.15e+05 | 2011-06-30 18:30:00.000000+0000 |      SALES | ASSISTANT MANAGER |  CHIRAG |     ['YOUTUBE', 'SPOTIFY']

(8 rows)
cqlsh:employee>
```

8. Create a TTL of 15 seconds to display the values of Employees.

//BEFORE 15 seconds

Cassandra Lab Program 2: -

Perform the following DB operations using Cassandra.

1.Create a key space by name Library

2. Create a column family by name Library-Info with attributes Stud_Id Primary Key,

   Counter_value of type Counter,
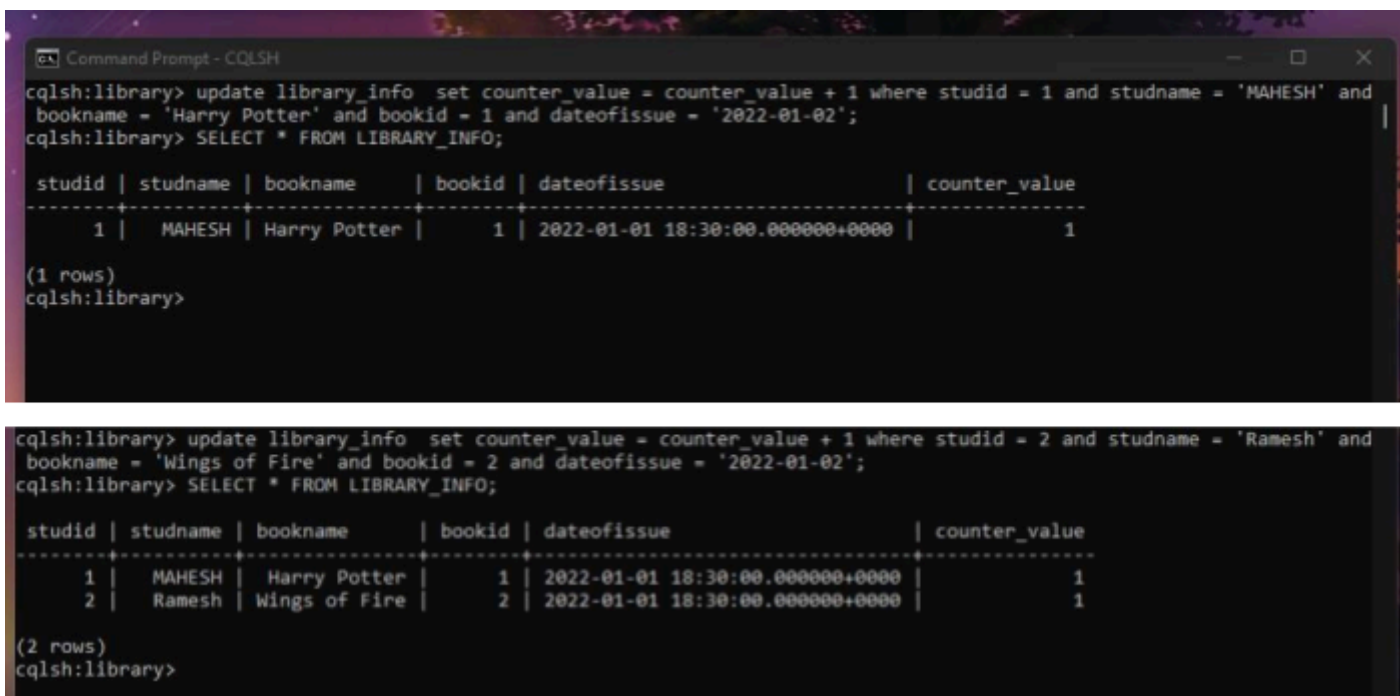
   Stud_Name, Book-Name, Book-Id, Date_of_issue

```
cqlsh> USE library;
cqlsh:library> CREATE TABLE LIBRARY_INFO( STUDID INT PRIMARY KEY, STUDNAME TEXT, BOOKNAME TEXT, DATEOFISSUE TIMESTAMP,
  COUNTER_VALUE COUNTER);
InvalidRequest: Error from server: code=2200 [Invalid query] message="Cannot mix counter and non counter columns in th
e same table"
cqlsh:library> CREATE TABLE LIBRARY_INFO( STUDID INT, STUDNAME TEXT, BOOKNAME TEXT, BOOKID INT, DATEOFISSUE TIMESTAMP,
  COUNTER_VALUE COUNTER, PRIMARY KEY(STUDID, STUDNAME, BOOKNAME, BOOKID, DATEOFISSUE));
cqlsh:library> SELECT * FROM LIBRARYINFO;
InvalidRequest: Error from server: code=2200 [Invalid query] message="unconfigured table libraryinfo"
cqlsh:library> SELECT * FROM LIBRARY_INFO;

 studid | studname | bookname | bookid | dateofissue | counter_value
--------+----------+----------+--------+-------------+---------------

(0 rows)
cqlsh:library>
```

3.Insert the values into the table in batch

```
Command Prompt - CQLSH                                                                              —    □    ×
cqlsh:library> update library_info  set counter_value = counter_value + 1 where studid = 1 and studname = 'MAHESH' and
 bookname = 'Harry Potter' and bookid = 1 and dateofissue = '2022-01-02';
cqlsh:library> SELECT * FROM LIBRARY_INFO;

 studid | studname | bookname     | bookid | dateofissue                     | counter_value
--------+----------+--------------+--------+---------------------------------+---------------
     1  |  MAHESH  | Harry Potter |     1  | 2022-01-01 18:30:00.000000+0000 |             1

(1 rows)
cqlsh:library>
```

```
cqlsh:library> update library_info  set counter_value = counter_value + 1 where studid = 2 and studname = 'Ramesh' and
 bookname = 'Wings of Fire' and bookid = 2 and dateofissue = '2022-01-02';
cqlsh:library> SELECT * FROM LIBRARY_INFO;

 studid | studname | bookname      | bookid | dateofissue                     | counter_value
--------+----------+---------------+--------+---------------------------------+---------------
     1  |  MAHESH  | Harry Potter  |     1  | 2022-01-01 18:30:00.000000+0000 |             1
     2  |  Ramesh  | Wings of Fire |     2  | 2022-01-01 18:30:00.000000+0000 |             1

(2 rows)
cqlsh:library>
```

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

```
cqlsh:library> update library_info  set counter_value = counter_value + 1 where studid = 112 and studname = 'Rajesh' a
nd bookname = 'BDA' and bookid = 3 and dateofissue = '2022-01-02';
cqlsh:library> SELECT * FROM LIBRARY_INFO;

 studid | studname | bookname       | bookid | dateofissue                      | counter_value
--------+----------+----------------+--------+----------------------------------+---------------
      1 |   MAHESH |   Harry Potter |      1 | 2022-01-01 18:30:00.000000+0000  |             1
      2 |   Ramesh | Wings of Fire  |      2 | 2022-01-01 18:30:00.000000+0000  |             1
    112 |   Rajesh |            BDA |      3 | 2022-01-01 18:30:00.000000+0000  |             1

(3 rows)
cqlsh:library>
```

```
(3 rows)
cqlsh:library> update library_info  set counter_value = counter_value + 1 where studid = 112 and studname = 'Rajesh' a
nd bookname = 'BDA' and bookid = 3 and dateofissue = '2022-01-02';
cqlsh:library> SELECT * FROM LIBRARY_INFO;

 studid | studname | bookname       | bookid | dateofissue                      | counter_value
--------+----------+----------------+--------+----------------------------------+---------------
      1 |   MAHESH |   Harry Potter |      1 | 2022-01-01 18:30:00.000000+0000  |             1
      2 |   Ramesh | Wings of Fire  |      2 | 2022-01-01 18:30:00.000000+0000  |             1
    112 |   Rajesh |            BDA |      3 | 2022-01-01 18:30:00.000000+0000  |             2

(3 rows)
cqlsh:library>
```

```
 studid | studname | bookname       | bookid | dateofissue                      | counter_value
--------+----------+----------------+--------+----------------------------------+---------------
    113 |  Ranjith |            rpa |      4 | 2022-01-01 18:30:00.000000+0000  |             1
      1 |   MAHESH |   Harry Potter |      1 | 2022-01-01 18:30:00.000000+0000  |             1
      2 |   Ramesh | Wings of Fire  |      2 | 2022-01-01 18:30:00.000000+0000  |             1
    112 |   Rajesh |            BDA |      3 | 2022-01-01 18:30:00.000000+0000  |             3

(4 rows)
```

5. Write a query to show that a student with id 112 has taken a book "BDA" 3 times.

```
Command Prompt - CQLSH
cqlsh:library> select * from library_info where studid = 112;

 studid | studname | bookname | bookid | dateofissue                      | counter_value
--------+----------+----------+--------+----------------------------------+---------------
    112 |   Rajesh |      BDA |      3 | 2022-01-01 18:30:00.000000+0000  |             3

(1 rows)
cqlsh:library>
```

6. Export the created column to a csv file

```
cqlsh:library> copy library_info (studid, studname, bookname, bookid, dateofissue, counter_value) to 'C:\Users\Admin\O
neDrive\Desktop\BDA Lab\data.csv';
Using 7 child processes

Starting copy of library.library_info with columns [studid, studname, bookname, bookid, dateofissue, counter_value].
Processed: 4 rows; Rate:      2 rows/s; Avg. rate:      1 rows/s
4 rows exported to 1 files in 3.004 seconds.
cqlsh:library>
```


Clipboard				Font				Alignment

POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited

A1 ... 113

	A	B	C	D	E	F	G	H	I
1	113	Ranjith	rpa	4	2022-01-0:	1			
2	2	Ramesh	Wings of F	2	2022-01-0:	1			
3	112	Rajesh	BDA	3	2022-01-0:	3			
4	1	MAHESH	Harry Pott	1	2022-01-0:	1			
5									
6									
7									

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


```
cqlsh:library> copy library_info (studid, studname, bookname, bookid, dateofissue, counter_value) from 'C:\Users\Admin\OneDrive\Desktop\BDA Lab\data.csv';
Using 7 child processes

Starting copy of library.library_info with columns [studid, studname, bookname, bookid, dateofissue, counter_value].
Process ImportProcess-10:      2 rows/s; Avg. rate:      2 rows/s
TPraceback (most recent call last):
rocess ImportProcess-8:
P Process ImportProcess-11:
TTraceback (most recent call last):
rocess ImportProcess-9:
P File "C:\Python27\lib\multiprocessing\process.py", line 267, in _bootstrap
 File "C:\Python27\lib\multiprocessing\process.py", line 267, in _bootstrap
Traceback (most recent call last):
Praceback (most recent call last):
 P File "C:\Python27\lib\multiprocessing\process.py", line 267, in _bootstrap
    self.run()
rocess ImportProcess-12:
  rocess ImportProcess-14:
rocess ImportProcess-13:
  T File "c:\apache-cassandra-3.11.13\bin\..\pylib\cqlshlib\copyutil.py", line 2339, in run
    self.run()
TTraceback (most recent call last):
 File "C:\Python27\lib\multiprocessing\process.py", line 267, in _bootstrap
raceback (most recent call last):
    self.run()
raceback (most recent call last):
      self.run()
 File "C:\Python27\lib\multiprocessing\process.py", line 267, in _bootstrap
 File "c:\apache-cassandra-3.11.13\bin\..\pylib\cqlshlib\copyutil.py", line 2339, in run
 File "C:\Python27\lib\multiprocessing\process.py", line 267, in _bootstrap
    self.close()
    self.run()
 File "c:\apache-cassandra-3.11.13\bin\..\pylib\cqlshlib\copyutil.py", line 2339, in run
    self.run()
 File "c:\apache-cassandra-3.11.13\bin\..\pylib\cqlshlib\copyutil.py", line 2343, in close
  File "c:\apache-cassandra-3.11.13\bin\..\pylib\cqlshlib\copyutil.py", line 2339, in run
    self.close()
 File "C:\Python27\lib\multiprocessing\process.py", line 267, in _bootstrap
 File "c:\apache-cassandra-3.11.13\bin\..\pylib\cqlshlib\copyutil.py", line 2339, in run
  File "c:\apache-cassandra-3.11.13\bin\..\pylib\cqlshlib\copyutil.py", line 2339, in run
    self._session.cluster.shutdown()
    self.run()
    self.close()
 File "c:\apache-cassandra-3.11.13\bin\..\lib\cassandra-driver-internal-only-3.11.0-bb96859b.zip\cassandra-driver-3.11.0-bb96859b\cassandra\cluster.py", line 1259, in shutdown
 File "c:\apache-cassandra-3.11.13\bin\..\pylib\cqlshlib\copyutil.py", line 2343, in close
 File "c:\apache-cassandra-3.11.13\bin\..\pylib\cqlshlib\copyutil.py", line 2339, in run
    self.close()
    self.close()
    self.close()
 File "c:\apache-cassandra-3.11.13\bin\..\pylib\cqlshlib\copyutil.py", line 2343, in close
```

File "c:\apache-cassandra-3.11.13\bin\..\lib\cassandra-driver-internal-only-3.11.0-bb96859b.zip\cassandra-driver-3.11.0-bb96859b\cassandra\io\asyncorereactor.py", line 373, in close
File "c:\apache-cassandra-3.11.13\bin\..\lib\cassandra-driver-internal-only-3.11.0-bb96859b.zip\cassandra-driver-3.11.0-bb96859b\cassandra\io\asyncorereactor.py", line 335, in create_timer
    File "c:\apache-cassandra-3.11.13\bin\..\lib\cassandra-driver-internal-only-3.11.0-bb96859b.zip\cassandra-driver-3.11.0-bb96859b\cassandra\io\asyncorereactor.py", line 373, in close
  self._connection.close()
File "c:\apache-cassandra-3.11.13\bin\..\lib\cassandra-driver-internal-only-3.11.0-bb96859b.zip\cassandra-driver-3.11.0-bb96859b\cassandra\io\asyncorereactor.py", line 335, in create_timer
  self._connection.close()
  AsyncoreConnection.create_timer(0, partial(asyncore.dispatcher.close, self))
  File "c:\apache-cassandra-3.11.13\bin\..\lib\cassandra-driver-internal-only-3.11.0-bb96859b.zip\cassandra-driver-3.11.0-bb96859b\cassandra\io\asyncorereactor.py", line 373, in close
  cls._loop.add_timer(timer)
File "c:\apache-cassandra-3.11.13\bin\..\lib\cassandra-driver-internal-only-3.11.0-bb96859b.zip\cassandra-driver-3.11.0-bb96859b\cassandra\io\asyncorereactor.py", line 373, in close
  cls._loop.add_timer(timer)
  AsyncoreConnection.create_timer(0, partial(asyncore.dispatcher.close, self))
AA   AsyncoreConnection.create_timer(0, partial(asyncore.dispatcher.close, self))
  ttributeError: 'NoneType' object has no attribute 'add_timer'
  File "c:\apache-cassandra-3.11.13\bin\..\lib\cassandra-driver-internal-only-3.11.0-bb96859b.zip\cassandra-driver-3.11.0-bb96859b\cassandra\io\asyncorereactor.py", line 335, in create_timer
ttributeError: 'NoneType' object has no attribute 'add_timer'
  File "c:\apache-cassandra-3.11.13\bin\..\lib\cassandra-driver-internal-only-3.11.0-bb96859b.zip\cassandra-driver-3.11.0-bb96859b\cassandra\io\asyncorereactor.py", line 335, in create_timer
  File "c:\apache-cassandra-3.11.13\bin\..\lib\cassandra-driver-internal-only-3.11.0-bb96859b.zip\cassandra-driver-3.11.0-bb96859b\cassandra\io\asyncorereactor.py", line 335, in create_timer
  AsyncoreConnection.create_timer(0, partial(asyncore.dispatcher.close, self))
      cls._loop.add_timer(timer)
  cls._loop.add_timer(timer)
A File "c:\apache-cassandra-3.11.13\bin\..\lib\cassandra-driver-internal-only-3.11.0-bb96859b.zip\cassandra-driver-3.11.0-bb96859b\cassandra\io\asyncorereactor.py", line 335, in create_timer
ttributeError: 'NoneType' object has no attribute 'add_timer'
A   cls._loop.add_timer(timer)
  cls._loop.add_timer(timer)
ttributeError: 'NoneType' object has no attribute 'add_timer'
AAttributeError: 'NoneType' object has no attribute 'add_timer'
ttributeError: 'NoneType' object has no attribute 'add_timer'
Processed: 4 rows; Rate: 1 rows/s; Avg. rate: 2 rows/s
4 rows imported from 1 files in 2.356 seconds (0 skipped).
cqlsh:library>

MongoDB Lab Program 1 (CRUD Demonstration): -

I. Perform the following DB operations using MongoDB.

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

2. Insert appropriate values

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

4. . Replace the student name from "ABC" to "FEM" of rollno 11

```
Student> db.s.insertMany([{StudName:"JohnDoe", Grade:"VIII", Hobbies:"Painting"},
... {StudName:"SarahSmith", Grade:"VI", Hobbies:"Reading"},
... {StudName:"DavidWilliams", Grade:"IX", Hobbies:"Coding"},
... {StudName:"EmilyJones", Grade:"V", Hobbies:"Dancing"},
... {StudName:"OliverBrown", Grade:"X", Hobbies:"Photography"},
... {StudName:"AvaGarcia", Grade:"IV", Hobbies:"Writing"},
... {StudName:"WilliamParker", Grade:"VII", Hobbies:"Chess"},
... {StudName:"OliviaWilson", Grade:"III", Hobbies:"Cooking"},
... {StudName:"MichaelHall", Grade:"XI", Hobbies:"Gardening"},
... {StudName:"CharlotteRoberts", Grade:"II", Hobbies:"Swimming"}])
{
  acknowledged: true,
  insertedIds: {
    '0': ObjectId('660a7cce2d86dcfefdc00e80'),
    '1': ObjectId('660a7cce2d86dcfefdc00e81'),
    '2': ObjectId('660a7cce2d86dcfefdc00e82'),
    '3': ObjectId('660a7cce2d86dcfefdc00e83'),
    '4': ObjectId('660a7cce2d86dcfefdc00e84'),
    '5': ObjectId('660a7cce2d86dcfefdc00e85'),
    '6': ObjectId('660a7cce2d86dcfefdc00e86'),
    '7': ObjectId('660a7cce2d86dcfefdc00e87'),
    '8': ObjectId('660a7cce2d86dcfefdc00e88'),
    '9': ObjectId('660a7cce2d86dcfefdc00e89')
  }
}
Student> db.s.find()
[
  {
    _id: ObjectId('660a7a7e2d86dcfefdc00e7e'),
    StudName: 'MichelleJacintha',
    Grade: 'VII',
    Hobbies: 'InternsetSurfing'
  },
  {
    _id: ObjectId('660a7ab22d86dcfefdc00e7f'),
    StudName: 'Aryana',
    Grade: 'VIII',
    Hobbies: 'InternsetSurfing'
  },
  {
    _id: ObjectId('660a7cce2d86dcfefdc00e80'),
    StudName: 'JohnDoe',
    Grade: 'VIII',
    Hobbies: 'Painting'
  },
  {
    _id: ObjectId('660a7cce2d86dcfefdc00e81'),
    StudName: 'SarahSmith',
    Grade: 'VI',
    Hobbies: 'Reading'
  },
  {
    _id: ObjectId('660a7cce2d86dcfefdc00e82'),
    StudName: 'DavidWilliams',
    Grade: 'IX',
```

```
    Grade: 'III',
    Hobbies: 'Cooking'
  },
  {
    _id: ObjectId('660a7cce2d86dcfefdc00e88'),
    StudName: 'MichaelHall',
    Grade: 'XI',
    Hobbies: 'Gardening'
  },
  {
    _id: ObjectId('660a7cce2d86dcfefdc00e89'),
    StudName: 'CharlotteRoberts',
    Grade: 'II',
    Hobbies: 'Swimming'
  }
]
Student> db.s.upodateOne({StudName:"AvaGarcia"}, {$set:{StudName: 'Ambika'}})
TypeError: db.s.upodateOne is not a function
Student> db.s.updateOne({StudName:"AvaGarcia"}, {$set:{StudName: 'Ambika'}})
{
  acknowledged: true,
  insertedId: null,
  matchedCount: 1,
  modifiedCount: 1,
  upsertedCount: 0
}
Student> db.s.find({StudName:'Ambika'})
[
  {
    _id: ObjectId('660a7cce2d86dcfefdc00e85'),
    StudName: 'Ambika',
    Grade: 'IV',
    Hobbies: 'Writing'
  }
]
Student> db.s.updateOne({StudName:"Ada"}, {$set:{StudName: 'Ambuja'}}, {upsert: true})
{
  acknowledged: true,
  insertedId: ObjectId('660a7e7c7b7d50080592d516'),
  matchedCount: 0,
  modifiedCount: 0,
  upsertedCount: 1
}
Student> db.s.find({StudName:'Ambuja'})
[ { _id: ObjectId('660a7e7c7b7d50080592d516'), StudName: 'Ambuja' } ]
Student> db.s.updateOne({StudName:"Ambuja"}, {$set:{Grade: 'LKG'}, {Hobbies: 'Staring at Wall'}}, {upsert: true})
Uncaught:
SyntaxError: Unexpected token (1:58)

> 1 | db.s.updateOne({StudName:"Ambuja"}, {$set:{Grade: 'LKG'}, {Hobbies: 'Staring at Wall'}}, {upsert: true})
    |                                                          ^
  2 |

Student> db.s.updateOne({StudName:"Ambuja"}, {$set:{Grade: 'LKG', Hobbies: 'Staring at Wall'}}, {upsert: true})
{
```

```
      }
]
Student> db.s.updateOne({StudName:"Ada"}, {$set:{StudName: 'Ambuja'}}, {upsert: true})
{
  acknowledged: true,
  insertedId: ObjectId('660a7e7c7b7d50080592d516'),
  matchedCount: 0,
  modifiedCount: 0,
  upsertedCount: 1
}
Student> db.s.find({StudName:'Ambuja'})
[ { _id: ObjectId('660a7e7c7b7d50080592d516'), StudName: 'Ambuja' } ]
Student> db.s.updateOne({StudName:"Ambuja"}, {$set:{Grade: 'LKG'}, {Hobbies: 'Staring at Wall'}}, {upsert: true})
Uncaught:
SyntaxError: Unexpected token (1:58)

> 1 | db.s.updateOne({StudName:"Ambuja"}, {$set:{Grade: 'LKG'}, {Hobbies: 'Staring at Wall'}}, {upsert: true})
    |                                                           ^
  2 |

Student> db.s.updateOne({StudName:"Ambuja"}, {$set:{Grade: 'LKG', Hobbies: 'Staring at Wall'}}, {upsert: true})
{
  acknowledged: true,
  insertedId: null,
  matchedCount: 1,
  modifiedCount: 1,
  upsertedCount: 0
}
Student> db.s.find({StudName:'Ambuja'})
[
  {
    _id: ObjectId('660a7e7c7b7d50080592d516'),
    StudName: 'Ambuja',
    Grade: 'LKG',
    Hobbies: 'Staring at Wall'
  }
]
Student> db.s.updateOne({StudName:"Ambuja"}, {$set:{Grade: 'LKG', Hobbies: 'Chewing Gum'}}, {upsert: true})
{
  acknowledged: true,
  insertedId: null,
  matchedCount: 1,
  modifiedCount: 1,
  upsertedCount: 0
}
Student> db.s.find({StudName:'Ambuja'})
[
  {
    _id: ObjectId('660a7e7c7b7d50080592d516'),
    StudName: 'Ambuja',
    Grade: 'LKG',
    Hobbies: 'Chewing Gum'
  }
]
Student> ▯
```

II. Perform the following DB operations using MongoDB.

1. Create a collection by name Customers with the following attributes.

 Cust_id, Acc_Bal, Acc_Type

2. Insert at least 5 values into the table

3. Write a query to display those records whose total account balance is greater than

   1200 of account type 'Z' for each customer_id.

4. Determine Minimum and Maximum account balance for each customer_i

```
Student> db.createCollection("c")
{ ok: 1 }
Student> db.c.insertMany([
...        {Cust_id: "0001", Acc_Bal: 5000, Acc_type: "Savings"},
...        {Cust_id: "0002", Acc_Bal: 12000, Acc_type: "Current"},
...        {Cust_id: "0003", Acc_Bal: 7500, Acc_type: "Savings"},
...        {Cust_id: "0004", Acc_Bal: 25000, Acc_type: "Current"},
...        {Cust_id: "0005", Acc_Bal: 10000, Acc_type: "Savings"},
...        {Cust_id: "0006", Acc_Bal: 17000, Acc_type: "Current"},
...        {Cust_id: "0007", Acc_Bal: 8000, Acc_type: "Savings"},
...        {Cust_id: "0008", Acc_Bal: 32000, Acc_type: "Current"},
...        {Cust_id: "0009", Acc_Bal: 15000, Acc_type: "Savings"},
...        {Cust_id: "0010", Acc_Bal: 20000, Acc_type: "Current"}
... ])
{
  acknowledged: true,
  insertedIds: {
    '0': ObjectId('660a84152d86dcfefdc00e8a'),
    '1': ObjectId('660a84152d86dcfefdc00e8b'),
    '2': ObjectId('660a84152d86dcfefdc00e8c'),
    '3': ObjectId('660a84152d86dcfefdc00e8d'),
    '4': ObjectId('660a84152d86dcfefdc00e8e'),
    '5': ObjectId('660a84152d86dcfefdc00e8f'),
    '6': ObjectId('660a84152d86dcfefdc00e90'),
    '7': ObjectId('660a84152d86dcfefdc00e91'),
    '8': ObjectId('660a84152d86dcfefdc00e92'),
    '9': ObjectId('660a84152d86dcfefdc00e93')
  }
}
Student>
```

```
Student> db.c.find({Acc_Bal: {$gt: 1200}, Acc_type: "Current"})
[
  {
    _id: ObjectId('660a84152d86dcfefdc00e8b'),
    Cust_id: '0002',
    Acc_Bal: 12000,
    Acc_type: 'Current'
  },
  {
    _id: ObjectId('660a84152d86dcfefdc00e8d'),
    Cust_id: '0004',
    Acc_Bal: 25000,
    Acc_type: 'Current'
  },
  {
    _id: ObjectId('660a84152d86dcfefdc00e8f'),
    Cust_id: '0006',
    Acc_Bal: 17000,
    Acc_type: 'Current'
  },
  {
    _id: ObjectId('660a84152d86dcfefdc00e91'),
    Cust_id: '0008',
    Acc_Bal: 32000,
    Acc_type: 'Current'
  },
  {
    _id: ObjectId('660a84152d86dcfefdc00e93'),
    Cust_id: '0010',
    Acc_Bal: 20000,
    Acc_type: 'Current'
  }
]
Student>
```

```
Student> db.c.aggregate([{"$group": {"_id": null ,"max": {"$max": "$Acc_Bal"}, "min": {"$min":"$Acc_Bal" }}}])
[ { _id: null, max: 32000, min: 5000 } ]
Student> db.c.aggregate([{"$group": {"_id": "Cust_id" ,"max": {"$max": "$Acc_Bal"}, "min": {"$min":"$Acc_Bal" }}}])
[ { _id: 'Cust_id', max: 32000, min: 5000 } ]
Student>
```

Hadoop Commands

hdusersbmsce-OptiPlus-3000:-$ sudo su hduser

[sudo] password for hduser:

hdusersbmsce-OptiPlus-3000: $ start-all.sh

This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh

22/06/06 14:43:45 WARN util.NativeCodeLoader: Unable to load native-hadoop Library for your platform... using builtin-java classes where applicable Starting namenodes on [localhost] localhost: nanenade running as process 3396. Stop it first. localhost: datanode running as process 3564, Stop it first.

starting secondary nanenodes [0.0.0.0) 0.0.0.0: secondarynamenode running as process 3773. Stop it first.

O22/06/06 14:43:47 WARN uttt.NativeCodeLoader: Unable to load native-hadoop library for your starting yarn daemons resource process 3932. Stop it first.

Localhost: running as process 4255. stop it first. 6003 Jps

3932 ResourceManager

3773 SecondaryNameNode

4255 NodeManager

hdusersbmsce-OptiPlus-3060:-$ hdfs dfs -mkdir

/Amrutha hdusersbmsce-OptiPlus-3060: $ hdfs dfs -ls /

22/06/06 14:45:30 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable Found 19 itens

drwxr-xr-x   hduser supergroup   02022-   06-06 11:44 /AAA

drwxr-xr-x   -hduser supergroup 2022-06-03 12:17 /Army   drwxr-xr-x   hduser supergroup 02022- 06-06 11:40

/Avnit   drwxr-xr-x   -hduser supergroup 02022-05-31 10:44 /88

drwxr-xr-x -hduser supergroup 02022- 06-01 15:03 /Cath

drwxr-xr-x -hduser supergroup

drwxr-xr-x hduser supergroup

drwxr-xr-x -hduser supergroup

drwxr-xr-x -hduser supergroup

drwxr-xr-x -hduser supergroup

drwxr-xr-x -hduser supergroup

drwxr-xr-x -hduser supergroup

drwxr-xr-x -hduser supergroup

drwxr-xr-x -hduser supergroup

drwxr-xr-x   -   hduser

supergroup drwxr-xr-x -hduser
supergroup 82022-06-04 10:06
/FFF
02022-06-06 14:40 /Kmrv
02022-06-06 14:44 /Amrutha
02022-06-01 15:03 /Neha
02022-06-04 09:54 /WC.txt
0 2022-06-04 09:54 /welcone.txt
02022-06-06 11:36 /abc
62022-06-03 12:13 /akash
0 2022-06-03 15:12 /darshan
0 2022-06-04 09:31 /ghh 8 2022-
06-06 11:45 /hello drwxr-xr-x -
hduser supergroup
62022-06-04 09:35 /rahul
drwxr-xr-x -hduser
supergroup 02022-06-03 12:11
/shre drwxr-xr-x .hduser supergroup 02022-06-03
12:41
/shreshtha
hdusersbmsce-OptiPlus-3060:-$ hdfs dfs put /home/hduser/Desktop/6b.txt
/Amrutha/WC.txt
22/05/06 14:46:40 WARN util.NativeCodeLoader: Unable to load
native-hadoop library for your platform... using butltin-java classes where
applicable hduserabesce-OptiPlex-3060:-$ hdfs dfs cat /Amrutha/WC.txt
22/06/06 14:47:00 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable hello fron
of hdusersbmsce-OptiPlus-3040:-$ hdfs dfs-get /Amrutha/WC.txt
/home/hduser/Downloads/newic.txt
22/05/06 14:51:43 WARN util.NativeCodeLoader: Unable to load nattve-hadoop
library for your platform... using builtin-java classes where applicable
hdusersbmsce-OptiPlus-3066:-$ cd Downloads
hdusersbmsce-OptiPlus-3060:-/Downloads$ cat newwMC.Ext hello from 6E
hdusersbmsce-OptiPlus-3060:-$ hdfs dfs -1s /Amrutha/
22/06/06 14:54:04 WARN util.NativeCodeLoader: Unable to load
native-hadoop Library for your platform... using builtin java classes where
applicable
Found 2 itens
-rw-r--r-- 1 hduser supergroup
23 2822-06-06 14:46 /Amrutha/MC.txt
1 hduser supergroup
23 2022-06-06 14:58 /Amrutha/newwc.txt
hdusersbmsce-OptiPlus-3060:-5 hdfs drs -getmerge /Amrutha/wc.txt
/Amrutha/newwc.txt /bone/hduser/Desktop/newmerge.txt

22/06/06 14:55:18 NARN util.NativeCodeLoader: Unable to load nattve-hadoop library for your platform... using butitin-Java classes where applicable hduserabesce-OptiPlex- 3060:~$ cd Desktop

hduser@besce-OptiPlex-3060:-/Desktops cat newmerge.txt hello from 68

D B

hello

from   68

D B

hdusersbmsce-OptiPlus-3060:-/Desktops hadoop fs getfacl /Amrutha/ 22/06/06 14:56:24 WARN util.NativeCodeLoader: Unable to load native hadoop library for your platform... using builtin java classes where applicable # file: /Amrutha

owner: hduser

group:

supergroup

user::rwx

group::r-x

other::r-x

hdusersbmsce-OptiPlus-3060:-/Desktop5 hdfs dfs copyToLocal /Amrutha/HC.txt /home/hduser/Desktop

22/05/06 14:58:09 WARN util.NativeCodeLoader: Unable to load native-hadoop Library for your platform... using butltin-java classes where applicable hdusersbmsce-OptiPlus-3000:-

/Desktop5 cat MC.txt hello fron 68

hdusersbmsce-OptiPlus-3060:-/Desktops hdfs dfs -cat /Amrutha/MC.txt 22/06/06 14:58:59 WARN util.NativeCodeLoader: Unable to load native-hadoop Library for your platform... ustng bulltin-Java classes where applicable hello from GB B

hdusersbmsce-OptiPlus-3060:-/Desktop5 hadoop fs - /Amrutha /FFF 22/06/06 14:59:46 WARN util.NativeCodeLoader: Unable to load native-hadoop Library for your platform... using builtin-java classes where applicable

hduseransceOptiPlex-3060:-/Desktops hadoop fs-Ls /FFF 22/05/06 15:00:00 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using butltin- java classes where applicable Found 2 itens drwxrxr-x -hduser supergroup TWEE 1 hduser supergroup 02022-05-06 14:50

/FFF/Amrutha 17 2022-05-04 10:06 /FFF/MC.txt

hdusersbmsce-OptiPlus-3060:-/Desktops hadoop fs cp /FFF/ /LLL

22/06/06 15:09:34 WARN util.NativeCodeLoader: Unable to load native hadoop library for your platform... using butltin-java classes where applicable hdusersbmsce-OptiPlus-3060:-

/Desktops hadoop fs -Ls /LLL

22/06/06 15:10:07 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable Found 2 1tens

drwxr-xr-x -hduser supergroup

hdusersbmsce-OptiPlus-3000:-/Desktops 02022-06-06 15:09

Hadoop Programs

1) Word Count WCMapper Java Class file.

```java
// Importing libraries import java.io.IOException; import org.apache.hadoop.io.IntWritable; import
 org.apache.hadoop.io.LongWritable; import org.apache.hadoop.io.Text;
import    org.apache.hadoop.mapred.MapReduceBase;    import
org.apache.hadoop.mapred.Mapper;                      import
org.apache.hadoop.mapred.OutputCollector;             import
org.apache.hadoop.mapred.Reporter;

public class WCMapper extends MapReduceBase implements Mapper<LongWritable,
     Text, Text, IntWritable> {

  // Map function
  public void map(LongWritable key, Text value, OutputCollector<Text, IntWritable> output,
Reporter rep) throws IOException
  {

      String line = value.toString();

      // Splitting  the  line  on  spaces
for (String word : line.split(" "))
      {
         if (word.length() > 0)
         {
             output.collect(new Text(word), new IntWritable(1));
         }  }   } }
```

Reducer Code

```java
// Importing libraries import java.io.IOException; import
java.util.Iterator;                               import
org.apache.hadoop.io.IntWritable; import
```

org.apache.hadoop.io.Text; import org.apache.hadoop.mapred.MapReduceBase; import org.apache.hadoop.mapred.OutputCollector; import org.apache.hadoop.mapred.Reducer; import org.apache.hadoop.mapred.Reporter;

```java
public class WCReducer extends MapReduceBase implements Reducer<Text, IntWritable, Text, IntWritable> {

    // Reduce function    public    void    reduce(Text    key,
Iterator<IntWritable> value,
            OutputCollector<Text,    IntWritable>    output,
                    Reporter rep) throws IOException
{


        int count = 0;

        // Counting  the  frequency  of  each  words
while (value.hasNext())
        {
            IntWritable  i  =  value.next();
count += i.get();
        }

        output.collect(key, new IntWritable(count));
    }
}
```

Driver Code: // Importing libraries import java.io.IOException; import org.apache.hadoop.conf.Configured; import org.apache.hadoop.fs.Path; import org.apache.hadoop.io.IntWritable; import

```java
                                                          import
org.apache.hadoop.io.Text;
org.apache.hadoop.mapred.FileInputFormat;                 import
org.apache.hadoop.mapred.FileOutputFormat;                import
org.apache.hadoop.mapred.JobClient;                       import
org.apache.hadoop.mapred.JobConf;                         import
org.apache.hadoop.util.Tool;                              import
org.apache.hadoop.util.ToolRunner;
public class WCDriver extends Configured implements Tool {


    public int run(String args[]) throws IOException



    {
        if (args.length < 2)
        {
            System.out.println("Please give valid inputs"); return -1;
        }

        JobConf  conf  =  new  JobConf(WCDriver.class);  FileInputFormat.setInputPaths(conf,  new
        Path(args[0]));
FileOutputFormat.setOutputPath(conf, new Path(args[1]));
conf.setMapperClass(WCMapper.class);
conf.setReducerClass(WCReducer.class);
conf.setMapOutputKeyClass(Text.class);
conf.setMapOutputValueClass(IntWritable.class);
conf.setOutputKeyClass(Text.class);
conf.setOutputValueClass(IntWritable.class);
        JobClient.runJob(conf)
        ; return 0;
    }

    // Main Method
    public static void main(String args[]) throws Exception{
```

```
        int     exitCode     =     ToolRunner.run(new     WCDriver(),     args);
        System.out.println(exitCode);
    }
} Output :
```

```
drwxr-xr-x   - hduser supergroup          0 2022-06-01 09:46 /user1
hduser@bmsce-Precision-T1700:~$ hdfs dfs -mkdir /input_khushil
hduser@bmsce-Precision-T1700:~$ hdfs dfs -ls /
Found 19 items
drwxr-xr-x   - hduser supergroup          0 2022-06-06 12:35 /CSE
drwxr-xr-x   - hduser supergroup          0 2022-06-06 12:23 /FFF
drwxr-xr-x   - hduser supergroup          0 2022-06-06 12:36 /LLL
drwxr-xr-x   - hduser supergroup          0 2022-06-20 12:06 /amit_bda
drwxr-xr-x   - hduser supergroup          0 2022-06-03 14:52 /bharath
drwxr-xr-x   - hduser supergroup          0 2022-06-03 14:43 /bharath035
drwxr-xr-x   - hduser supergroup          0 2022-05-31 10:21 /example
drwxr-xr-x   - hduser supergroup          0 2022-06-01 15:13 /foldernew
drwxr-xr-x   - hduser supergroup          0 2022-06-06 15:04 /hemang061
drwxr-xr-x   - hduser supergroup          0 2022-06-20 15:13 /input_khushil
drwxr-xr-x   - hduser supergroup          0 2022-06-03 12:27 /irfan
drwxr-xr-x   - hduser supergroup          0 2022-06-01 15:09 /muskan
drwxr-xr-x   - hduser supergroup          0 2022-06-06 15:04 /new_folder
drwxr-xr-x   - hduser supergroup          0 2022-05-31 10:26 /one
drwxr-xr-x   - hduser supergroup          0 2022-06-20 12:17 /output
drwxr-xr-x   - hduser supergroup          0 2022-06-03 12:08 /saurab
drwxrwxr-x   - hduser supergroup          0 2019-08-01 16:19 /tmp
drwxr-xr-x   - hduser supergroup          0 2019-08-01 16:03 /user
drwxr-xr-x   - hduser supergroup          0 2022-06-01 09:46 /user1
hduser@bmsce-Precision-T1700:~$ hdfs dfs -put /home/hduser/sample.txt /input_khushil
hduser@bmsce-Precision-T1700:~$ hdfs dfs -ls /input_khushil
Found 1 items
-rw-r--r--   1 hduser supergroup         52 2022-06-20 15:15 /input_khushil/sample.txt
hduser@bmsce-Precision-T1700:~$ hadoop jar /home/hduser/khushil/WordCount.jar WCDriver
/input_khushil /input_khushil/output_khushil
22/06/20 15:16:44 INFO Configuration.deprecation: session.id is deprecated. Instead, use
dfs.metrics.session-id
22/06/20 15:16:44 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName=JobTracker,
sessionId=
22/06/20 15:16:44 INFO jvm.JvmMetrics: Cannot initialize JVM Metrics with
processName=JobTracker, sessionId= - already initialized
22/06/20 15:16:44 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not
performed. Implement the Tool interface and execute your application with ToolRunner to remedy
this.
22/06/20 15:16:44 INFO mapred.FileInputFormat: Total input paths to process : 1
22/06/20 15:16:44 INFO mapreduce.JobSubmitter: number of splits:1
22/06/20 15:16:44 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_local230197290_0001
22/06/20 15:16:44 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
22/06/20 15:16:44 INFO mapred.LocalJobRunner: OutputCommitter set in config null
22/06/20 15:16:44 INFO mapreduce.Job: Running job: job_local230197290_0001
22/06/20 15:16:44 INFO mapred.LocalJobRunner: OutputCommitter is
org.apache.hadoop.mapred.FileOutputCommitter
22/06/20 15:16:44 INFO mapred.LocalJobRunner: Waiting for map tasks
22/06/20 15:16:44 INFO mapred.LocalJobRunner: Starting task:
attempt_local230197290_0001_m_000000_0
22/06/20 15:16:44 INFO mapred.Task:  Using ResourceCalculatorProcessTree : [ ]
22/06/20 15:16:44 INFO mapred.MapTask: Processing split:
hdfs://localhost:54310/input_khushil/sample.txt:0+52
22/06/20 15:16:44 INFO mapred.MapTask: numReduceTasks: 1
22/06/20 15:16:44 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
22/06/20 15:16:44 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
22/06/20 15:16:44 INFO mapred.MapTask: soft limit at 83886080
22/06/20 15:16:44 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
22/06/20 15:16:44 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
```

```
                    CPU time spent (ms)=0
                    Physical memory (bytes) snapshot=0
                    Virtual memory (bytes) snapshot=0
                    Total committed heap usage (bytes)=471859200
            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=52
            File Output Format Counters
                    Bytes Written=63
0
hduser@bmsce-Precision-T1700:~$ hdfs dfs -ls /input_khushil
Found 2 items
drwxr-xr-x   - hduser supergroup          0 2022-06-20 15:16 /input_khushil/output_khushil
-rw-r--r--   1 hduser supergroup         52 2022-06-20 15:15 /input_khushil/sample.txt
hduser@bmsce-Precision-T1700:~$ hdfs dfs -ls /input_khushil/output_khushil
Found 2 items
-rw-r--r--   1 hduser supergroup          0 2022-06-20 15:16
/input_khushil/output_khushil/_SUCCESS


-rw-r--r--   1 hduser supergroup         63 2022-06-20 15:16
/input_khushil/output_khushil/part-00000
hduser@bmsce-Precision-T1700:~$ hdfs dfs -cat /input_khushil/output_khushil/part-0000
cat: `/input_khushil/output_khushil/part-0000': No such file or directory
hduser@bmsce-Precision-T1700:~$ hdfs dfs -cat /input_khushil/output_khushil/part-00000
am      1
awesome         1
hadoop 2
hi      1
i       1
im      1
is      1
khushil         1
learing         1
```

2) Top N
Driver-TopN.class

package samples.topn;

```java
import         java.io.IOException;         import         java.util.StringTokenizer;         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.lib.input.FileInputFormat;
import
org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.util.GenericOptionsParser;

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 static final IntWritable one = new IntWritable(1);

    private Text word = new Text();
```

```java
        private String tokens = "[_|$#<>\\^=\\[\\]\\*/\\\\,;,.\\:()?!\"]";
    public void map(Object key, Text value, Mapper<Object, Text, Text, IntWritable>.Context context) throws IOException, InterruptedException {
        String cleanLine = value.toString().toLowerCase().replaceAll(this.tokens, " ");
        StringTokenizer itr = new StringTokenizer(cleanLine);
        while (itr.hasMoreTokens()) {
            this.word.set(itr.nextToken().trim());
            context.write(this.word, one);
        }
    }
}
```

TopNCombiner.class
```java
package samples.topn;

import java.io.IOException;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;

public class TopNCombiner extends Reducer<Text, IntWritable, Text, IntWritable> {
    public void reduce(Text key, Iterable<IntWritable> values, Reducer<Text, IntWritable, Text, IntWritable>.Context context) throws IOException, InterruptedException {
        int sum = 0;
        for (IntWritable val : values)
            sum += val.get();
        context.write(key, new IntWritable(sum));
    }
}
```

TopNMapper.class
```java
package samples.topn;
```

```java
import java.io.IOException; import java.util.StringTokenizer;
import          org.apache.hadoop.io.IntWritable;          import
org.apache.hadoop.io.Text;                                 import
org.apache.hadoop.mapreduce.Mapper;

public class TopNMapper extends Mapper<Object, Text, Text, IntWritable> {  private static
final IntWritable one = new IntWritable(1);
        private Text word = new
Text();
        private                     String                     tokens                     =
"[_|$#<>\\^=\\[\\]\\*/\\\\,;,.\\:()?!\"]";
        public   vo```\\id   map(Object   key,      Text      value,   Mapper<Object,   Text,   Text,
IntWritable>.Context context) throws IOException,
InterruptedException {                          String                    cleanLine                =
value.toString().toLowerCase().replaceAll(this.tokens,   "   ");   StringTokenizer   itr   =   new
StringTokenizer(cleanLine);
while                 (itr.hasMoreTokens())                 {
this.word.set(itr.nextToken().trim());
context.write(this.word, one);
    }
}
}
```

TopNReducer.class
```java
package samples.topn;

import          java.io.IOException;          import
java.util.HashMap;  import java.util.Map;  import
org.apache.hadoop.io.IntWritable;          import
org.apache.hadoop.io.Text;                  import
org.apache.hadoop.mapreduce.Reduc   er;  import
utils.MiscUtils;
```

```java
public class TopNReducer extends Reducer<Text, IntWritable, Text, IntWritable> {
private Map<Text, IntWritable> countMap = new HashMap<>();

    public void reduce(Text key, Iterable<IntWritable> values, Reducer<Text, IntWritable, Text,
IntWritable>.Context context) throws IOException, InterruptedException {                           int
sum = 0; for (IntWritable val : values)                          sum              +=           val.get();
this.countMap.put(new Text(key), new IntWritable(sum)); }
    protected   void   cleanup(Reducer<Text,   IntWritable,   Text,   IntWritable>.Context
context) throws IOException,
InterruptedException {
        Map<Text,        IntWritable>        sortedMap        =
MiscUtils.sortByValues(this.countMap); int counter = 0; for
(Text key : sortedMap.keySet()) {               if (counter++
== 20)                  break;
        context.write(key, sortedMap.get(key));
    }
  }
}
```

Output:

```
hduser@bmsce-Precision-T1700:~/Desktop/temperature$ hdfs dfs -mkdir /khushil_topn
hduser@bmsce-Precision-T1700:~/Desktop/temperature$ hdfs dfs -put ./input.txt /khushil_topn/
hduser@bmsce-Precision-T1700:~/Desktop/temperature$ hdfs dfs -ls /khushil_topn/
Found 1 items
-rw-r--r--   1 hduser supergroup        103 2022-06-27 15:43 /khushil_topn/input.txt
hduser@bmsce-Precision-T1700:~/Desktop/temperature$ hadoop jar topn.jar TopNDriver
/khushil_topn/input.txt /khushil_topn/output
Exception in thread "main" java.lang.ClassNotFoundException: TopNDriver
 at java.net.URLClassLoader.findClass(URLClassLoader.java:382)
 at java.lang.ClassLoader.loadClass(ClassLoader.java:418)
 at java.lang.ClassLoader.loadClass(ClassLoader.java:351)
 at java.lang.Class.forName0(Native Method)
 at java.lang.Class.forName(Class.java:348)
 at org.apache.hadoop.util.RunJar.run(RunJar.java:214)
 at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
hduser@bmsce-Precision-T1700:~/Desktop/temperature$ hadoop jar topn.jar topn.TopNDriver
/khushil_topn/input.txt /khushil_topn/output
22/06/27 15:45:22 INFO Configuration.deprecation: session.id is deprecated. Instead, use
dfs.metrics.session-id
22/06/27 15:45:22 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName=JobTracker,
sessionId=
22/06/27 15:45:22 INFO input.FileInputFormat: Total input paths to process : 1
22/06/27 15:45:22 INFO mapreduce.JobSubmitter: number of splits:1
22/06/27 15:45:22 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local691635730_0001
22/06/27 15:45:22 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
22/06/27 15:45:22 INFO mapreduce.Job: Running job: job_local691635730_0001
22/06/27 15:45:22 INFO mapred.LocalJobRunner: OutputCommitter set in config null
22/06/27 15:45:22 INFO mapred.LocalJobRunner: OutputCommitter is
org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
22/06/27 15:45:22 INFO mapred.LocalJobRunner: Waiting for map tasks
22/06/27 15:45:22 INFO mapred.LocalJobRunner: Starting task: attempt_local691635730_0001_m_000000_0
22/06/27 15:45:22 INFO mapred.Task:  Using ResourceCalculatorProcessTree : [ ]
22/06/27 15:45:22 INFO mapred.MapTask: Processing split:
hdfs://localhost:54310/khushil_topn/input.txt:0+103
22/06/27 15:45:22 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
22/06/27 15:45:22 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
22/06/27 15:45:22 INFO mapred.MapTask: soft limit at 83886080
22/06/27 15:45:22 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
22/06/27 15:45:22 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
22/06/27 15:45:22 INFO mapred.MapTask: Map output collector class =
org.apache.hadoop.mapred.MapTask$MapOutputBuffer
22/06/27 15:45:22 INFO mapred.LocalJobRunner:
22/06/27 15:45:22 INFO mapred.MapTask: Starting flush of map output
22/06/27 15:45:22 INFO mapred.MapTask: Spilling map output
22/06/27 15:45:22 INFO mapred.MapTask: bufstart = 0; bufend = 187; bufvoid = 104857600
22/06/27 15:45:22 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214316(104857264);
length = 81/6553600
22/06/27 15:45:22 INFO mapred.MapTask: Finished spill 0
22/06/27 15:45:22 INFO mapred.Task: Task:attempt_local691635730_0001_m_000000_0 is done. And is in
the process of committing
22/06/27 15:45:22 INFO mapred.LocalJobRunner: map
22/06/27 15:45:22 INFO mapred.Task: Task 'attempt_local691635730_0001_m_000000_0' done.
22/06/27 15:45:22 INFO mapred.LocalJobRunner: Finishing task: attempt_local691635730_0001_m_000000_0
22/06/27 15:45:22 INFO mapred.LocalJobRunner: map task executor complete.
22/06/27 15:45:22 INFO mapred.LocalJobRunner: Waiting for reduce tasks
22/06/27 15:45:22 INFO mapred.LocalJobRunner: Starting task: attempt_local691635730_0001_r_000000_0
22/06/27 15:45:22 INFO mapred.Task:  Using ResourceCalculatorProcessTree : [ ]
```

```
 Map input records=6
 Map output records=21
 Map output bytes=187
 Map output materialized bytes=235
 Input split bytes=110
 Combine input records=0
 Combine output records=0
 Reduce input groups=15
 Reduce shuffle bytes=235
 Reduce input records=21
 Reduce output records=15
 Spilled Records=42
 Shuffled Maps =1
 Failed Shuffles=0
 Merged Map outputs=1
 GC time elapsed (ms)=42
 CPU time spent (ms)=0
 Physical memory (bytes) snapshot=0
 Virtual memory (bytes) snapshot=0
 Total committed heap usage (bytes)=578289664
 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=103
 File Output Format Counters
 Bytes Written=105
hduser@bmsce-Precision-T1700:~/Desktop/temperature$ hdfs dfs -ls /khushil_topn/output/
Found 2 items
-rw-r--r--   1 hduser supergroup          0 2022-06-27 15:45 /khushil_topn/output/_SUCCESS
-rw-r--r--   1 hduser supergroup        105 2022-06-27 15:45 /khushil_topn/output/part-r-00000
hduser@bmsce-Precision-T1700:~/Desktop/temperature$ hdfs dfs -cat /khushil_topn/output/part-r-00000
hadoop  4
i3
am      2
hi      1
im      1
is      1
there   1
bye     1
learing 1
awesome 1
love    1
khushil 1
cool    1
and     1
using   1
hduser@bmsce-Precision-T1700:~/Desktop/temperature$
```

3) Average Temperature

AverageDriver

```
package temp;

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.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
```
package temp;

import java.io.IOException; import org.apache.hadoop.io.IntWritable;
import           org.apache.hadoop.io.LongWritable;            import
org.apache.hadoop.io.Text;                                     import
org.apache.hadoop.mapreduce.Mapper; public class AverageMapper
extends
```

```java
Mapper<LongWritable, Text, Text, IntWritable> {       public
static final int MISSING = 9999;

    public void map(LongWritable key, Text value, Mapper<LongWritable, Text, Text,
IntWritable>.Context context) throws IOException, InterruptedException {                    int
temperature;
        String line = value.toString();              String year = line.substring(15, 19);       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 != 9999 && quality.matches("[01459]")) context.write(new
            Text(year), new
IntWritable(temperature));
    }
}
```

AverageReducer **package**
temp;

```java
import               java.io.IOException;              import
org.apache.hadoop.io.IntWritable;              import
org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;

public class AverageReducer extends Reducer<Text, IntWritable, Text, IntWritable> {
public void reduce(Text key, Iterable<IntWritable> values, Reducer<Text, IntWritable, Text,
IntWritable>.Context context) throws IOException, InterruptedException {   int max_temp = 0;
                                        int count = 0;
```

```java
        for (IntWritable value : values)
{               max_temp  +=  value.get();  count++;
                }
        context.write(key, new IntWritable(max_temp / count));
}
}
```

Output:

```
hduser@bmsce-Precision-T1700:~/Desktop/temperature$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
Starting namenodes on [localhost]
hduser@localhost's password:
localhost: starting namenode, logging to /usr/local/hadoop/logs/hadoop-hduser-namenode-bmsce-
Precision-T1700.out
hduser@localhost's password:
localhost: starting datanode, logging to /usr/local/hadoop/logs/hadoop-hduser-datanode-bmsce-
Precision-T1700.out
Starting secondary namenodes [0.0.0.0]
hduser@0.0.0.0's password:
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-hduser-
secondarynamenode-bmsce-Precision-T1700.out
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-hduser-resourcemanager-bmsce-
Precision-T1700.out
hduser@localhost's password:
localhost: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-hduser-nodemanager-bmsce-
Precision-T1700.out
hduser@bmsce-Precision-T1700:~/Desktop/temperature$ jps
6832 NodeManager
6498 ResourceManager
6339 SecondaryNameNode
4887 org.eclipse.equinox.launcher_1.5.600.v20191014-2022.jar
6954 Jps
6123 DataNode
5951 NameNode
hduser@bmsce-Precision-T1700:~/Desktop/temperature$ hdfs dfs -le /
-le: Unknown command
hduser@bmsce-Precision-T1700:~/Desktop/temperature$ hdfs dfs -ls /
Found 31 items
drwxr-xr-x   - hduser supergroup          0 2022-06-06 12:35 /CSE
drwxr-xr-x   - hduser supergroup          0 2022-06-06 12:23 /FFF
drwxr-xr-x   - hduser supergroup          0 2022-06-06 12:36 /LLL
drwxr-xr-x   - hduser supergroup          0 2022-06-20 12:06 /amit_bda
drwxr-xr-x   - hduser supergroup          0 2022-06-27 11:42 /amit_lab
drwxr-xr-x   - hduser supergroup          0 2022-06-03 14:52 /bharath
drwxr-xr-x   - hduser supergroup          0 2022-06-03 14:43 /bharath035
drwxr-xr-x   - hduser supergroup          0 2022-06-24 14:54 /chi
drwxr-xr-x   - hduser supergroup          0 2022-05-31 10:21 /example
drwxr-xr-x   - hduser supergroup          0 2022-06-01 15:13 /foldernew
drwxr-xr-x   - hduser supergroup          0 2022-06-06 15:04 /hemang061
drwxr-xr-x   - hduser supergroup          0 2022-06-20 15:16 /input_khushil
drwxr-xr-x   - hduser supergroup          0 2022-06-03 12:27 /irfan
drwxr-xr-x   - hduser supergroup          0 2022-06-22 10:44 /lwde
drwxr-xr-x   - hduser supergroup          0 2022-06-27 13:03 /mapreducejoin_amit
drwxr-xr-x   - hduser supergroup          0 2022-06-22 15:32 /muskan
drwxr-xr-x   - hduser supergroup          0 2022-06-22 15:06 /muskan_op
drwxr-xr-x   - hduser supergroup          0 2022-06-22 15:35 /muskan_output
drwxr-xr-x   - hduser supergroup          0 2022-06-06 15:04 /new_folder
drwxr-xr-x   - hduser supergroup          0 2022-05-31 10:26 /one
drwxr-xr-x   - hduser supergroup          0 2022-06-24 15:30 /out55
drwxr-xr-x   - hduser supergroup          0 2022-06-20 12:17 /output
drwxr-xr-x   - hduser supergroup          0 2022-06-27 13:04 /output_TOPn
drwxr-xr-x   - hduser supergroup          0 2022-06-27 12:14 /output_Topn
drwxr-xr-x   - hduser supergroup          0 2022-06-24 12:42 /r1
drwxr-xr-x   - hduser supergroup          0 2022-06-24 12:24 /rgs
```