

VISVESVARAYA TECHNOLOGICAL UNIVERSITY
“JnanaSangama”, Belgaum -590014, Karnataka.



LAB REPORT
on
Big Data Analytics (23CS6PCBDA)

Submitted by

Arbaj Wadagera (1BM22CS051)

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



B.M.S. COLLEGE OF ENGINEERING
(Autonomous Institution under VTU)
BENGALURU-560019
Feb-2025 to June-2025

**B. M. S. College of Engineering,
Bull Temple Road, Bangalore 560019**
(Affiliated To Visvesvaraya Technological University, Belgaum)
Department of Computer Science and Engineering



CERTIFICATE

This is to certify that the Lab work entitled "**Big Data Analytics(23CS6PCBDA)**" carried out by **Arbaj Wadagera(1BM22CS051)**, 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 2024. The Lab report has been approved as it satisfies the academic requirements in respect of a **Big Data Analytics(23CS6PCBDA)** work prescribed for the said degree.

Vikranth B M
Assistant Professor
Department of CSE
BMSCE, Bengaluru

Dr. Kavitha Sooda
Professor and Head
Department of CSE
BMSCE, Bengaluru

Index Sheet

Sl. No.	Experiment Title	Page No.
1	MongoDB	1-13
2	MongoDB	14-16
3	Neo4j	17-20
4	Cassandra	21-25
5	Cassandra	26-37
6	Hadoop	38-44
7	Hadoop	45-56
8	Scala and Spark	57-61

Course Outcome

CO1	Apply the concept of NoSQL, Hadoop or Spark for a given task
CO2	Analyse big data analytics mechanisms that can be applied to obtain solution for a given problem.
CO3	Design and implement solutions using data analytics mechanisms for a given problem.

Lab 1: MongoDB- CRUD Demonstration

Question: Perform basic CRUD (Create, Read, Update, Delete) operations in MongoDB.

Code with Output:

```
Atlas atlas-ws5rct-shard-0 [primary] mydb> db.createCollection('Student')
{ ok: 1 }

Atlas atlas-ws5rct-shard-0 [primary] mydb> db.createCollection('Student')
{ ok: 1 }
...
Atlas atlas-ws5rct-shard-0 [primary] mydb> db.createCollection('Customers')
{ ok: 1 }

Atlas atlas-ws5rct-shard-0 [primary] test> use mydb
switched to db mydb

Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Student.deleteMany({Grade:'VII'})
{ acknowledged: true, deletedCount: 3 }

Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Student.deleteOne({StudName:'JacobAdam'})
{ acknowledged: true, deletedCount: 0 }

[Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Student.drop()
true
]

Atlas atlas-ws5rct-shard-0 [primary] mydb> db.dropDatabase()
{ ok: 1, dropped: 'mydb' }

Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Student.remove({StudName:'JacobAdam'})
)
DeprecationWarning: Collection.remove() is deprecated. Use deleteOne, deleteMany, findOneAndDelete, or bulkWrite.
{ acknowledged: true, deletedCount: 0 }
```

```
[Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Customers.find()
[
  {
    _id: ObjectId('67c6c71f812483cc27dd4a64'),
    cust_id: 1,
    balance: 200,
    type: 'S'
  },
  {
    _id: ObjectId('67c6c739812483cc27dd4a65'),
    cust_id: 1,
    balance: 1000,
    type: 'Z'
  },
  {
    _id: ObjectId('67c6c74d812483cc27dd4a66'),
    cust_id: 2,
    balance: 100,
    type: 'Z'
  },
  {
    _id: ObjectId('67c6c75e812483cc27dd4a67'),
    cust_id: 2,
    balance: 1000,
    type: 'C'
  },
  {
    _id: ObjectId('67c6c76e812483cc27dd4a68'),
    cust_id: 2,
    balance: 500,
    type: 'C'
  },
  {
    _id: ObjectId('67c6c781812483cc27dd4a69'),
    cust_id: 2,
    balance: 50,
    type: 'S'
  },
  {
    _id: ObjectId('67c6c795812483cc27dd4a6a'),
    cust_id: 3,
    balance: 500,
    type: 'Z'
  }
]
```

Google Classroom

```
Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Student.find()
[
  {
    _id: ObjectId('67c6c3c3812483cc27dd4a5d'),
    RollNo: 1,
    Age: 21,
    Cont: 9876,
    email: 'antara.de9@gmail.com'
  },
  {
    _id: ObjectId('67c6c3d8812483cc27dd4a5e'),
    RollNo: 1,
    Age: 21,
    Cont: 9876,
    email: 'antara.de9@gmail.com'
  },
  {
    _id: ObjectId('67c6c458812483cc27dd4a5f'),
    RollNo: 2,
    Age: 22,
    Cont: 9976,
    email: 'anushka.de@gmail.com'
  },
  {
    _id: ObjectId('67c6c47f812483cc27dd4a60'),
    RollNo: 3,
    Age: 21,
    Cont: 5576,
    email: 'anubhav.de@gmail.com'
  },
  {
    _id: ObjectId('67c6c4a2812483cc27dd4a61'),
    RollNo: 4,
    Age: 20,
    Cont: 4476,
    email: 'pani.de9@gmail.com'
  },
  {
    _id: ObjectId('67c6c4db812483cc27dd4a62'),
    RollNo: 10,
    Age: 23,
    Cont: 2276,
    email: 'rekha.de9@gmail.com'
  }
]
```

```
Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Student.find()
[{"_id": ObjectId('67c6c3c3812483cc27dd4a5d'), "RollNo": 1, "Age": 21, "Cont": 9876, "email": "antara.de9@gmail.com"}, {"_id": ObjectId('67c6c3d8812483cc27dd4a5e'), "RollNo": 1, "Age": 21, "Cont": 9876, "email": "antara.de9@gmail.com"}, {"_id": ObjectId('67c6c458812483cc27dd4a5f'), "RollNo": 2, "Age": 22, "Cont": 9976, "email": "anushka.de@gmail.com"}, {"_id": ObjectId('67c6c47f812483cc27dd4a60'), "RollNo": 3, "Age": 21, "Cont": 5576, "email": "anubhav.de@gmail.com"}, {"_id": ObjectId('67c6c4a2812483cc27dd4a61'), "RollNo": 4, "Age": 20, "Cont": 4476, "email": "pani.de9@gmail.com"}, {"_id": ObjectId('67c6c4db812483cc27dd4a62'), "RollNo": 10, "Age": 23, "Cont": 2276, "email": "abhinav@gmail.com"}]
```

```
{  
  _id: ObjectId('67c6c616812483cc27dd4a63'),  
  RollNo: 11,  
  Age: 22,  
  Name: 'FEM',  
  cont: 2276,  
  email: 'rea.de9@gmail.com'  
,  
{  
  _id: 1,  
  StudName: 'Michelle Jacintha',  
  Grade: 'VII',  
  Hobbies: 'InternetSurfing'  
,  
  { _id: 2, StudName: 'Jannie', Grade: 'VIII', Hobbies: 'Music' },  
  { _id: 3, StudName: 'Jacob Adam', Grade: 'VII', Hobbies: 'Swimming' },  
  {  
    _id: 4,  
    StudName: 'Amy Jacks',  
    Grade: 'X',  
    Hobbies: 'Dancing',  
    Location: 'Network'  
,  
  { _id: 6, StudName: 'Aryan David', Grade: 'VII', Hobbies: 'Skating' }  
]
```

```

Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Customers.insert({cust_id:1,balance:200,type:'S'})
{
  acknowledged: true,
  insertedIds: { '0': ObjectId('67c6c71f812483cc27dd4a64') }
}
Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Customers.insert({cust_id:1,balance:1000,type:'Z'})
{
  acknowledged: true,
  insertedIds: { '0': ObjectId('67c6c739812483cc27dd4a65') }
}
[Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Customers.insert({cust_id:2,balance:100,type:'Z'})
{
  acknowledged: true,
  insertedIds: { '0': ObjectId('67c6c74d812483cc27dd4a66') }
}
[Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Customers.insert({cust_id:2,balance:1000,type:'C'})
{
  acknowledged: true,
  insertedIds: { '0': ObjectId('67c6c75e812483cc27dd4a67') }
}
[Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Customers.insert({cust_id:2,balance:500,type:'C'})
{
  acknowledged: true,
  insertedIds: { '0': ObjectId('67c6c76e812483cc27dd4a68') }
}
Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Customers.insert({cust_id:2,balance:50,type:'S'})
{
  acknowledged: true,
  insertedIds: { '0': ObjectId('67c6c781812483cc27dd4a69') }
}
[Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Customers.insert({cust_id:3,balance:500,type:'Z'})
{
  acknowledged: true,
  insertedIds: { '0': ObjectId('67c6c795812483cc27dd4a6a') }
}

```

```

Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Student.insert({RollNo:1,Age:21,Cont:t:9876,email:'antara.de9@gmail.com'});
DeprecationWarning: Collection.insert() is deprecated. Use insertOne, insertMany, or bulkWrite.
{
  acknowledged: true,
  insertedIds: { '0': ObjectId('67c6c3c3812483cc27dd4a5d') }
}
Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Student.insertOne({RollNo:1,Age:21,Cont:t:9876,email:'antara.de9@gmail.com'});
{
  acknowledged: true,
  insertedId: ObjectId('67c6c3d8812483cc27dd4a5e')
}
Atlas atlas-ws5rct-shard-0 [primary] mydb> show mydb
MongoshInvalidInputError: [COMMON-10001] 'mydb' is not a valid argument for "show".
Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Student.insert({RollNo:2,Age:22,Cont:t:9976,email:'anushka.de@gmail.com'});
{
  acknowledged: true,
  insertedIds: { '0': ObjectId('67c6c458812483cc27dd4a5f') }
}
Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Student.insert({RollNo:3,Age:21,Cont:t:5576,email:'anubhav.de@gmail.com'});
{
  acknowledged: true,
  insertedIds: { '0': ObjectId('67c6c47f812483cc27dd4a60') }
}
Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Student.insert({RollNo:4,Age:20,Cont:t:4476,email:'pani.de9@gmail.com'});
{
  acknowledged: true,
  insertedIds: { '0': ObjectId('67c6c4a2812483cc27dd4a61') }
}
Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Student.insert({RollNo:10,Age:23,Cont:t:2276,email:'rekha.de9@gmail.com'});
{
  acknowledged: true,
  insertedIds: { '0': ObjectId('67c6c4db812483cc27dd4a62') }
}

Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Student.insertMany([{_id:3,StudName:'Jacob Adam',Grade:'VII',Hobbies:'Swimming'}, {_id:4,StudName:'Amy Jacks',Grade:'X',Hobbies:'Dancing'}])
{ acknowledged: true, insertedIds: { '0': 3, '1': 4 } }

Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Student.insert({_id:1,StudName:'Michelle Jacintha',Grade:'VII',Hobbies:'InternetSurfing'})
{ acknowledged: true, insertedIds: { '0': 1 } }
Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Student.insertOne({_id:2,StudName:'Janie',Grade:'VIII',Hobbies:'Music'})
{ acknowledged: true, insertedId: 2 }

```

```
Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Student.find().pretty()
[
  {
    _id: ObjectId('67c6c3c3812483cc27dd4a5d'),
    RollNo: 1,
    Age: 21,
    Cont: 9876,
    email: 'antara.de9@gmail.com'
  },
  {
    _id: ObjectId('67c6c3d8812483cc27dd4a5e'),
    RollNo: 1,
    Age: 21,
    Cont: 9876,
    email: 'antara.de9@gmail.com'
  },
  {
    _id: ObjectId('67c6c458812483cc27dd4a5f'),
    RollNo: 2,
    Age: 22,
    Cont: 9976,
    email: 'anushka.de@gmail.com'
  },
  {
    _id: ObjectId('67c6c47f812483cc27dd4a60'),
    RollNo: 3,
    Age: 21,
    Cont: 5576,
    email: 'anubhav.de@gmail.com'
  },
  {
    _id: ObjectId('67c6c4a2812483cc27dd4a61'),
    RollNo: 4,
    Age: 20,
    Cont: 4476,
    email: 'pani.de9@gmail.com'
  },
  {
    _id: ObjectId('67c6c4db812483cc27dd4a62'),
    RollNo: 10,
    Age: 23,
    Cont: 2276,
    email: 'rekha.de9@gmail.com'
  }
]
```

```

Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Student.save({StudName:'Vamsi',Grade:'VI'})
Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Student.updateOne({_id:6,StudName:'Aryan David',Grade:'VII'},{$set:{Hobbies:'Skating'}},{upsert:true})
{
  acknowledged: true,
  insertedId: 6,
  matchedCount: 0,
  modifiedCount: 0,
  upsertedCount: 1
}

Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Student.insert({RollNo:11,Age:22,Name :"ABC",cont:2276,email:"rea.de9@gmail.com"})
{
  acknowledged: true,
  insertedIds: { '0': ObjectId('67c6c616812483cc27dd4a63') }
}
Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Student.update({RollNo:11,Name:"ABC"},{$set:{Name:"FEM"}})
{
  acknowledged: true,
  insertedId: null,
  matchedCount: 1,
  modifiedCount: 1,
  upsertedCount: 0
}

Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Student.updateMany({Grade:'VII'},{$set:{status:'Active'}})
{
  acknowledged: true,
  insertedId: null,
  matchedCount: 3,
  modifiedCount: 2,
  upsertedCount: 0
}

Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Student.updateOne({Grade:'VII'},{$set:{status:'Active'}})
{
  acknowledged: true,
  insertedId: null,
  matchedCount: 1,
  modifiedCount: 1,
  upsertedCount: 0
}

```

```
Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Student.update({_id:4},{$set:{Location:'Network'}})
{
  acknowledged: true,
  insertedId: null,
  matchedCount: 1,
  modifiedCount: 1,
  upsertedCount: 0
}
[{"$unset": {"Location": "Network"}}

Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Student.update({_id:4},{$unset:{Location:'Network'}})
{
  acknowledged: true,
  insertedId: null,
  matchedCount: 1,
  modifiedCount: 1,
  upsertedCount: 0
}

Atlas atlas-ws5rct-shard-0 [primary] mydb> db.Student.update({RollNo:10},{$set:{email:'abhinav@gmail.com'}})
DeprecationWarning: Collection.update() is deprecated. Use updateOne, updateMany, or bulkWrite.
{
  acknowledged: true,
  insertedId: null,
  matchedCount: 1,
  modifiedCount: 1,
  upsertedCount: 0
}
```

Lab - I

```
→ mongosh  
→ use myDB  
→ db.createCollection("student");  
→ db.student.insert({ RollNo: 1, Age: 21,  
    cout: 9876, email: "abcd"  
    @gmail.com })
```

```
→ db.student.find()  
{  
  {  
    _id: ObjectId("63bFef..."),  
    RollNo: 1,  
    Age: 21,  
    cout: 9876,  
    email: "abcd@gmail.com"  
  }  
}
```

```
→ db.student.update({ RollNo: 10 }, { $set: { email:  
    abc@gmail.com } })
```

```
{  
  acknowledged: true,  
  insertedId: null,
```

matchedCount: 1,

modifiedCount: 1,

upsertedCount: 0,

lastErrorObject: null,

writeConcernError: null,

```

→ db.student.insert( { RollNo: 11, Age: 22,
                      Name: "ABC", Cout: 2276,
                      email: "xyz@gmail.com" }

{
    -id: ObjectId("63bfed..."),
    RollNo: 11,
    Age: 22,
    Name: "ABC"
    Cout: 2276
    email: "xyz@gmail.com"
}

```

```

→ db.student.update({RollNo: 11, Name: "ABC"},
```

```

    { $set: {Name: "FEM"} }

```

```

{
    -id: ObjectId("63bfed4..."),
    RollNo: 11,
    Age: 22,
    Name: "FEM",
    Cout: 2276
    email: "xyz@gmail.com"
}

```

```

→ mongoexport <copied connection url | Database  

        Name> --collection=Tablename  

--out <output file name on local system>  

Connected to : <connection url | Database name>  

Exported 3 records

```

```
→ mongoimport <copied connection url|Database name>
  --collection = Table name --type json --file
  <file name>
  connected to : <connection url|Database name>
  3 documents imported successfully. 0 documents
  failed to import.
```

```
→ mongoimport --connection-string <connection url>
  --db <DB name>
  --collection <Collection name>
  --type <JSON|CSV>
  --file <file name>
```

Lab 2: Cassandra

Question: Perform the following DB operations using Cassandra.

1. Create a keyspace by name Employee
2. Create a column family by name Employee-Info with attributes Emp_Id Primary Key, Emp_Name, Designation, Date_of_Joining, Salary, Dept_Name
3. Insert the values into the table in batch
4. Update Employee name and Department of Emp-Id 121
5. Sort the details of Employee records based on salary
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. Update the altered table to add project names.
8. Create a TTL of 15 seconds to display the values of Employees.

Code with Output:

```
cqlsh> CREATE KEYSPACE Employee WITH REPLICATION = { 'class' : 'SimpleStrategy', 'replication_factor' : 1 };
cqlsh> CREATE TABLE Employee.Employee_Info (
...     Emp_Id int,
...     Salary DECIMAL,
...     Emp_Name TEXT,
...     Designation TEXT,
...     Date_of_Joining DATE,
...     Dept_Name TEXT,
...     PRIMARY KEY (Emp_Id, Salary)
... ) WITH CLUSTERING ORDER BY (Salary ASC);
cqlsh> BEGIN BATCH
...     INSERT INTO Employee.Employee_Info (Emp_Id, Salary, Emp_Name, Designation, Date_of_Joining, Dept_Name) VALUES (121, 60000, 'John Doe', 'Developer', '2023-01-15',
..., 'IT');
...     INSERT INTO Employee.Employee_Info (Emp_Id, Salary, Emp_Name, Designation, Date_of_Joining, Dept_Name) VALUES (122, 80000, 'Jane Smith', 'Manager', '2022-05-20',
..., 'HR');
...     INSERT INTO Employee.Employee_Info (Emp_Id, Salary, Emp_Name, Designation, Date_of_Joining, Dept_Name) VALUES (123, 55000, 'Alice Johnson', 'Analyst', '2021-11-10',
..., 'Finance');
...     APPLY BATCH;
cqlsh> UPDATE Employee.Employee_Info SET Emp_Name = 'Johnathan Doe', Dept_Name = 'Engineering' WHERE Emp_Id = 121 AND Salary = 60000;
cqlsh> SELECT * FROM Employee.Employee_Info WHERE Emp_Id = 121 ORDER BY Salary;
emp_id | salary | date_of_joining | dept_name | designation | emp_name
-----+-----+-----+-----+-----+-----+
  121 |   60000 | 2023-01-15 | Engineering | Developer | Johnathan Doe
(1 rows)

cqlsh> ALTER TABLE Employee.Employee_Info ADD Projects SET<TEXT>;
cqlsh> UPDATE Employee.Employee_Info SET Projects = {'Project A', 'Project B'} WHERE Emp_Id = 121 AND Salary = 60000;
cqlsh> INSERT INTO Employee.Employee_Info (Emp_Id, Salary, Emp_Name, Designation, Date_of_Joining, Dept_Name) VALUES (124, 30000, 'Temp Employee', 'Intern', '2023-10-01', 'Temp Dept') USING TTL 15;
cqlsh> SELECT * FROM Employee.Employee_Info;
emp_id | salary | date_of_joining | dept_name | designation | emp_name | projects
-----+-----+-----+-----+-----+-----+-----+
  123 | 55000 | 2021-11-10 | Finance | Analyst | Alice Johnson | null
  122 | 80000 | 2022-05-20 | HR | Manager | Jane Smith | null
  121 | 60000 | 2023-01-15 | Engineering | Developer | Johnathan Doe | {'Project A', 'Project B'}
(3 rows)
cqlsh> []
```

LAB - II

→ Mongosh

→ use bankDB;

→ db.customers.insertOne({
 cust-id: 1,
 Acc-Bal: 1500,
 Acc-Type: 'Z'
});

→ db.customers.insertMany([
 { cust-id: 1, Acc-Bal: 1500, Acc-Type: 'Z' },
 { cust-id: 2, Acc-Bal: 1100, Acc-Type: 'Z' },
 { cust-id: 3, Acc-Bal: 1300, Acc-Type: 'X' },
 { cust-id: 4, Acc-Bal: 1600, Acc-Type: 'Z' },
 { cust-id: 5, Acc-Bal: 900, Acc-Type: 'Y' }
]);

→ db.customers.find({
 Acc-Bal: { \$gt: 1200 },
 Acc-Type: 'Z'
});

→ db.customers.aggregate([
 {
 \$group: {

```
-id : "$cust_id",
minBalance : { $min : "$Acc_Bal" },
maxBalance : { $max : "$Acc_Bal" }
}

]);
→ db.customers.find()
new cmd
→ mongoexport mongodb+srv://arbajcs22:2
password>@cluster0.fsgaw.mongodb.net/
bankDB --collection=Customers --out
c:\Users\student\Downloads\output2.csv
Exported 6 records.
```

Back to ~~earlier~~ ~~other~~ cmd.

```
→ db.customers.drop();
```

for import

~~mongoimport~~ ~~Customerabc --file~~
~~10/13/23~~

6 document(s) imported successfully.

Lab 3: Cassandra

Question: Perform the following DB operations using Cassandra.

1. Create a keyspace 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
3. Insert the values into the table in batch
4. Display the details of the table created and increase the value of the counter
5. Write a query to show that a student with id 112 has taken a book “BDA” 2 times.
6. Export the created column to a csv file
7. Import a given csv dataset from local file system into Cassandra column family

Code with Output:

```
cqlsh> CREATE KEYSPACE Library WITH REPLICATION = { 'class' : 'SimpleStrategy', 'replication_factor' : 1 };
cqlsh> CREATE TABLE Library.Library_Info (
...     Stud_Id int,
...     Book_Name TEXT,
...     Book_Id int,
...     Date_of_issue DATE,
...     PRIMARY KEY (Stud_Id, Book_Name, Date_of_issue)
... );
cqlsh> BEGIN BATCH
...     INSERT INTO Library.Library_Info (Stud_Id, Book_Name, Book_Id, Date_of_issue) VALUES (112, 'BDA', 1, '2023-09-01');
...     INSERT INTO Library.Library_Info (Stud_Id, Book_Name, Book_Id, Date_of_issue) VALUES (112, 'BDA', 1, '2023-09-05');
...     INSERT INTO Library.Library_Info (Stud_Id, Book_Name, Book_Id, Date_of_issue) VALUES (113, 'ML', 2, '2023-09-02');
...     INSERT INTO Library.Library_Info (Stud_Id, Book_Name, Book_Id, Date_of_issue) VALUES (114, 'AI', 3, '2023-09-03');
...     INSERT INTO Library.Library_Info (Stud_Id, Book_Name, Book_Id, Date_of_issue) VALUES (115, 'DBMS', 4, '2023-09-04');
...     APPLY BATCH;
cqlsh> SELECT * FROM Library.Library_Info;

stud_id | book_name | date_of_issue | book_id
-----+-----+-----+-----+
  114 |      AI | 2023-09-03 |      3
  113 |      ML | 2023-09-02 |      2
  112 |      BDA | 2023-09-01 |      1
  112 |      BDA | 2023-09-05 |      1
  115 |    DBMS | 2023-09-04 |      4

(5 rows)
cqlsh> SELECT COUNT(*) FROM Library.Library_Info WHERE Stud_Id = 112 AND Book_Name = 'BDA';

count
-----
  2

(1 rows)
```

```
cqlsh> COPY Library.Library_Info TO 'library_info.csv' WITH HEADER = TRUE;
Using 16 child processes

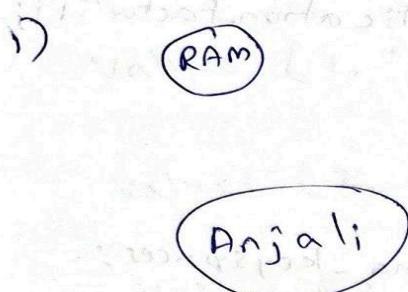
Starting copy of library.library_info with columns [stud_id, book_name, date_of_issue, book_id].
Processed: 5 rows; Rate:      96 rows/s; Avg. rate:      96 rows/s
5 rows exported to 1 files in 0.089 seconds.
cqlsh> COPY Library.Library_Info FROM 'library_info.csv' WITH HEADER = TRUE;
Using 16 child processes

Starting copy of library.library_info with columns [stud_id, book_name, date_of_issue, book_id].
Processed: 5 rows; Rate:      9 rows/s; Avg. rate:     13 rows/s
5 rows imported from 1 files in 0.375 seconds (0 skipped).
cqlsh> █
```

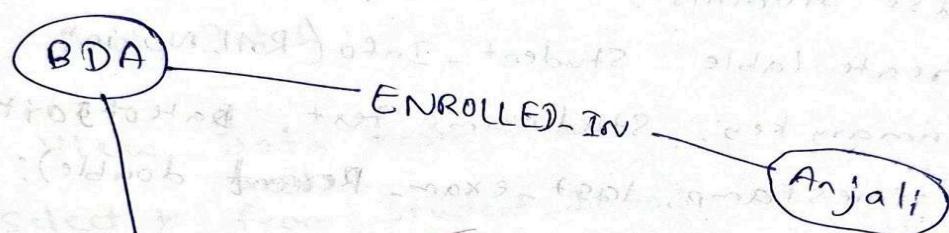
Lab-3

```
→ neo4j $ CREATE (a: STUDENT {name: "ram",  
id: 12})  
→ MATCH (n) RETURN n  
→ CREATE (a: STUDENT {name: "raj", id: 13})  
→ CREATE (a: STUDENT {name: "abhi", id: 14})  
→ CREATE (a: STUDENT {name: "anjali", id: 15})  
→ MATCH (n) RETURN n  
→ CREATE (a: COURSE {name: "DBMS", credits: 3})  
RETURN a  
→ CREATE (a: COURSE {name: "BDA", credits: 4})  
RETURN a  
→ CREATE (a: TEACHERS {name: "VBM", 3})  
→ MATCH (a: STUDENT), (b: COURSE) WHERE a.name =  
"Anjali" AND b.name = "BDA"  
CREATE (a) - [r: ENROLLED-IN] → (b)  
→ MATCH (n) RETURN n  
→ MATCH (a: TEACHER), (b: COURSE) WHERE a.name =  
"VBM" AND b.name = "BOA"  
CREATE (a) - [r: ENROLLED-IN] → (b)
```

① p -



2)



~~✓~~ 17/3/25

Lab 4: Cassandra

Question: Execution of HDFS Commands for interaction with Hadoop Environment. (Minimum 10 commands to be executed).

Code with Output:

```
hadoop@bmscsecse-HP-Elite-Tower-800-G9-Desktop-PC: $ cd ./Desktop/
hadoop@bmscsecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [bmscsecse-HP-Elite-Tower-800-G9-Desktop-PC]
Starting resourcemanager
Starting nodemanagers
hadoop@bmscsecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hdfs dfs -mkdir /Lab05
```

```
hadoop@bmscsecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hadoop fs -ls /Hadoop
ls: '/Hadoop': No such file or directory
hadoop@bmscsecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hadoop fs -ls /Lab05
```

```
hadoop@bmscsecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ touch test.txt
hadoop@bmscsecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ nano text.txt
hadoop@bmscsecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hdfs dfs -put ./text.txt /Lab05/text.txt
hadoop@bmscsecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hadoop fs -ls /Lab05
Found 1 items
-rw-r--r-- 1 hadoop supergroup 19 2024-05-13 14:33 /Lab05/text.txt
hadoop@bmscsecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hdfs dfs -cat /Lab05/text.txt
Hello
How are you?
```

```
hadoop@bmscsecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hadoop fs -ls /Lab05
Found 2 items
-rw-r--r-- 1 hadoop supergroup 15 2024-05-13 14:40 /Lab05/test.txt
-rw-r--r-- 1 hadoop supergroup 19 2024-05-13 14:33 /Lab05/text.txt
hadoop@bmscsecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hdfs dfs -getmerge /Lab05 /text.txt /Lab05 /test.txt ../Downloads/Merged.txt
getmerge: '/text.txt': No such file or directory
getmerge: '/test.txt': No such file or directory
hadoop@bmscsecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hdfs dfs -getmerge /Lab05/text.txt /Lab05/test.txt ../Downloads/Merged.txt
hadoop@bmscsecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hadoop fs -getfacl /Lab05
# file: /Lab05
# owner: hadoop
# group: supergroup
user::rwx
group::r-x
other::r-x
```

```
hadoop@bmscsecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hdfs dfs -copyToLocal /Lab05/text.txt ..//Documents
hadoop@bmscsecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hdfs dfs -copyToLocal /Lab05/test.txt ..//Documents
```

```
hadoop@bmscsecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hdfs dfs -cat /Lab05/text.txt
Hello
How are you?
hadoop@bmscsecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hdfs dfs -mv /Lab05 /test_Lab05
```

```
hadoop@bmscsecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hdfs dfs -ls /test_Lab05
Found 2 items
-rw-r--r-- 1 hadoop supergroup 15 2024-05-13 14:40 /test_Lab05/test.txt
-rw-r--r-- 1 hadoop supergroup 19 2024-05-13 14:33 /test_Lab05/text.txt
hadoop@bmscsecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hdfs dfs -cp /test_Lab05/ /Lab05
hadoop@bmscsecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hdfs dfs -ls /Lab05
Found 2 items
-rw-r--r-- 1 hadoop supergroup 15 2024-05-13 14:51 /Lab05/test.txt
-rw-r--r-- 1 hadoop supergroup 19 2024-05-13 14:51 /Lab05/text.txt
hadoop@bmscsecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hdfs dfs -ls /test_Lab05
Found 2 items
-rw-r--r-- 1 hadoop supergroup 15 2024-05-13 14:40 /test_Lab05/test.txt
-rw-r--r-- 1 hadoop supergroup 19 2024-05-13 14:33 /test_Lab05/text.txt
```

Lab-4

1) Create keyspace

Create keyspace Students with REPLICATION=

{ 'class': 'strategy', replication-factor': 1};

2) Check existing keyspace

③ Describe keyspaces:

④ Select * from system.schema_keyspaces;

use Students;

⑤ Create Table student_Info (Roll-No int
Primary key, StudName text, DateofJoin timestamp, last_exam_Percent double);

⑥ Describe Tables:

⑦ Insert into student_Info (Roll-No, StudName,
Date of Join, last_exam_percent)
Value (1, 'Asha', '2012-03-12');

(Select * from select_Info where
Roll-No IN (1, 2, 3);)

insert multiple records using Batch Begin
Batch

Insert Into Students_info (rollno, studname, date of joining, last exampercent)

Values (1, "Asha", "2012-05-04", 78);

Values (2, "Kiran", "2012-06-13", 85);

Values (3, "Tarun", "2012-03-12", 74);

Apply Batch;

8) View Data from Table

Select * from Student_Info;

O/P	rollno	studname	(date of joining)	lastexampercent
1	1	Asha	2012-05-04	78.0
2	2	Kiran	2012-06-13	85.0
3	3	Tarun	2012-03-12	74.0

9) Filter using where clause.

Select * from Student_info where roll-no
in (1, 2, 3);

Op	roll-no	student name	date of joining	last exam percent
1	1	Asha	2012-05-04	78.0
2	1	Kiran	2012-06-13	85.0
3	1	Tarun	2012-08-12	74.0

10. Errors when querying non-primary

key column

• Select * from student_info where
 studentname = 'Asha';

i) Create an index on student name

Create Index on student_info (studentname);

ii) Query using the index

Select * from student_info where
 studentname = 'Asha';

Op	roll-no	student name	date of joining	last exam percent
1	1	Asha	2012-05-04	78.0

13) Update a record

update student-info set studentname='David'
where roll-no=2;

14) Delete a record

Delete from student-info where roll-no=3;

22015 = notes

15) Alter table to add a column

Alter table student-info add hobbies text;

16) Export Data to CSV

copy student-info TO 'student-data.csv';

17) import data from CSV

copy student-info from 'student-data.csv';

✓
7/11/2015

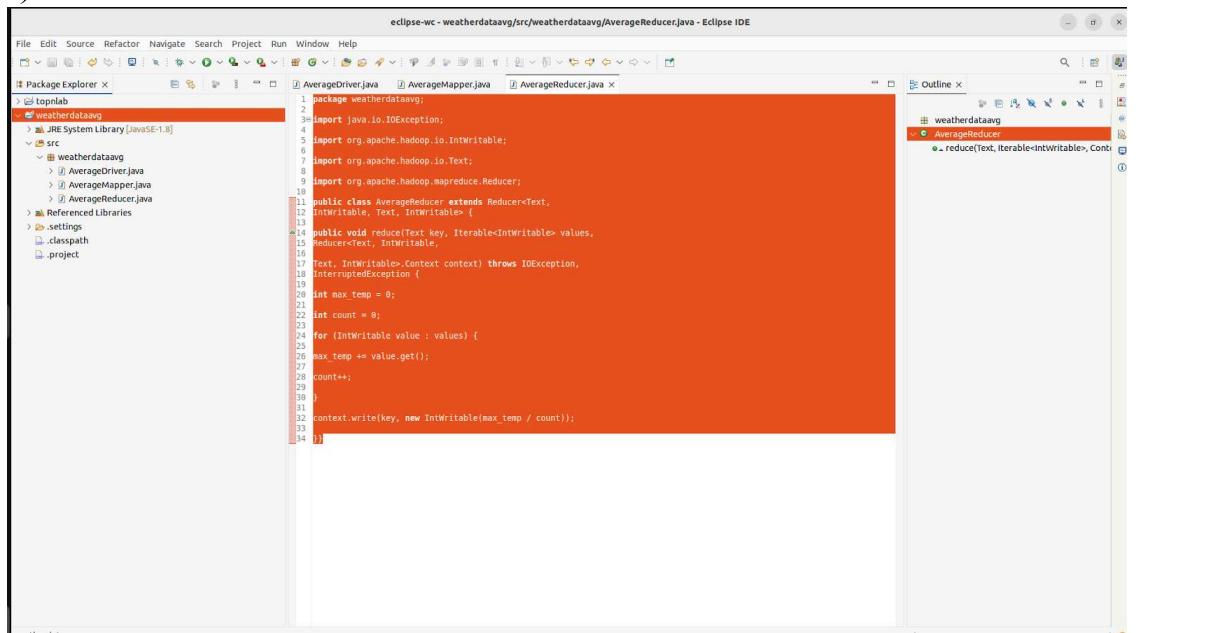
Lab 5: Hadoop

Question: From the following link extract the weather data
<https://github.com/tomwhite/hadoop-book/tree/master/input/ncdc/all> Create a Map Reduce program to

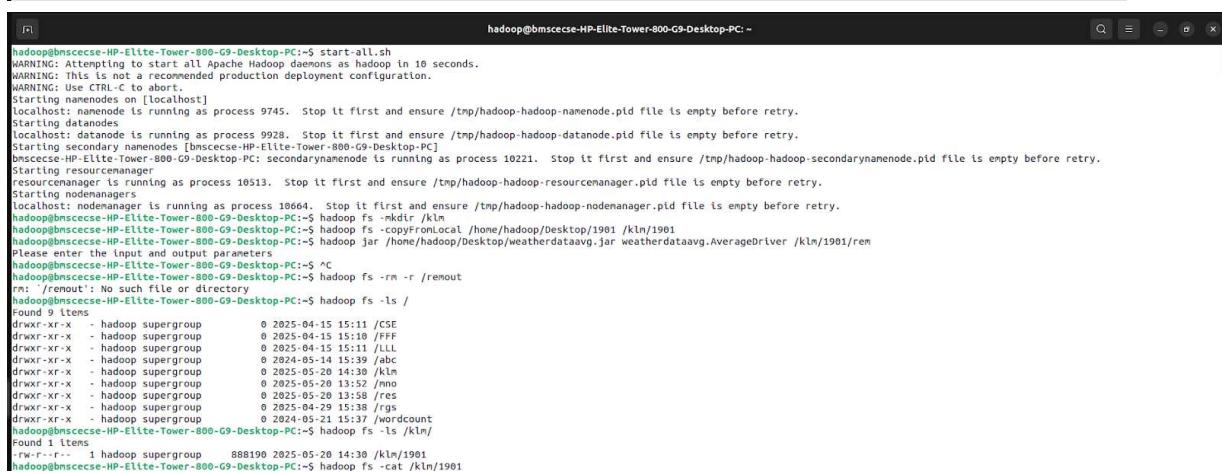
- find average temperature for each year from NCDC data set.
- find the mean max temperature for every month

Code with Output:

a)



```
eclipse-wc - weatherdataavg/src/weatherdataavg/AverageReducer.java - Eclipse IDE
File Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer X AverageDriver.java AverageMapper.java AverageReducer.java
JRE System Library [JavaSE-1.8]
src
weatherdataavg
AverageDriver.java
AverageMapper.java
AverageReducer.java
settings
classpath
project
Outline X
weatherdataavg
AverageReducer
reduce(Text, Iterable<IntWritable>, Context)
1 package weatherdataavg;
2 import java.io.IOException;
3 import org.apache.hadoop.io.IntWritable;
4 import org.apache.hadoop.io.Text;
5 import org.apache.hadoop.mapreduce.Reducer;
6
7 public class AverageReducer extends Reducer<Text, IntWritable, Text, IntWritable> {
8
9     public void reduce(Text key, Iterable<IntWritable> values,
10                        Reducer<Text, IntWritable, Text, IntWritable>.Context context) throws IOException,
11                        InterruptedException {
12
13         int max_temp = 0;
14         int count = 0;
15
16         for (IntWritable value : values) {
17             max_temp += value.get();
18             count++;
19         }
20
21         context.write(key, new IntWritable(max_temp / count));
22     }
23 }
```

```
hadoop@bmsccse-HP-Elite-Tower-800-G9-Desktop-PC:~ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
localhost: namenode is running as process 9745. Stop it first and ensure /tmp/hadoop-hadoop-namenode.pid file is empty before retry.
Starting datanodes
localhost: datanode is running as process 9928. Stop it first and ensure /tmp/hadoop-hadoop-datanode.pid file is empty before retry.
Starting secondarynamenodes [bmsccse-HP-Elite-Tower-800-G9-Desktop-PC]
bmsccse-HP-Elite-Tower-800-G9-Desktop-PC: secondarynamenode is running as process 10221. Stop it first and ensure /tmp/hadoop-hadoop-secondarynamenode.pid file is empty before retry.
Starting resourcemanager
resourcemanager is running as process 10515. Stop it first and ensure /tmp/hadoop-hadoop-resourcemanager.pid file is empty before retry.
Starting nodemanagers
localhost: nodemanager is running as process 18664. Stop it first and ensure /tmp/hadoop-hadoop-nodemanager.pid file is empty before retry.
hadoop@bmsccse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -mkdir /kln
hadoop@bmsccse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -copyFromLocal /home/hadoop/Desktop/1901 /kln/1901
hadoop@bmsccse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop jar /home/hadoop/Desktop/weatherdataavg.jar weatherdataavg.AverageDriver /kln/1901/rem
Please enter the input and output parameters
hadoop@bmsccse-HP-Elite-Tower-800-G9-Desktop-PC:~$ ^C
hadoop@bmsccse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -rm -r /rem
rem: /rem: No such file or directory
hadoop@bmsccse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -ls /
Found 9 items
drwxr-xr-x - hadoop supergroup 0 2025-04-15 15:11 /CSE
drwxr-xr-x - hadoop supergroup 0 2025-04-15 15:10 /FFF
drwxr-xr-x - hadoop supergroup 0 2025-04-15 15:11 /LLL
drwxr-xr-x - hadoop supergroup 0 2024-05-14 15:39 /abc
drwxr-xr-x - hadoop supergroup 0 2025-05-20 13:52 /mno
drwxr-xr-x - hadoop supergroup 0 2025-05-20 13:58 /res
drwxr-xr-x - hadoop supergroup 0 2025-04-29 15:35 /rgs
drwxr-xr-x - hadoop supergroup 0 2024-05-21 15:37 /wordcount
hadoop@bmsccse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -ls /kln/
Found 1 item
-rw-r--r-- 1 hadoop supergroup 888190 2025-05-20 14:38 /kln/1901
hadoop@bmsccse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -cat /kln/1901
```

```

hadoop@bmsccse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop jar /home/hadoop/Desktop/weatherdataavg.jar weatherdataavg.AverageDriver /klm/1901 /rem
2025-05-20 14:34:21,074 INFO impl.MetricsConfig: loaded properties from hadoop-metrics2.properties
2025-05-20 14:34:21,116 INFO lml.MetricsSystemImpl: Scheduled Metric snapshot period at 10 seconds(s).
2025-05-20 14:34:21,116 INFO lml.MetricsSystemImpl: JobTracker metrics system started
2025-05-20 14:34:21,230 INFO input.FileInputFormat: Total input files to process : 1
2025-05-20 14:34:21,262 INFO mapreduce.JobSubmitter: number of splits:1
2025-05-20 14:34:21,334 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local261815800_0001
2025-05-20 14:34:21,334 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-05-20 14:34:21,400 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-05-20 14:34:21,400 INFO mapreduce.Job: Running local in directory: /local261815800
2025-05-20 14:34:21,407 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-05-20 14:34:21,405 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2025-05-20 14:34:21,405 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-05-20 14:34:21,406 INFO output.FileOutputCommitter: skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-05-20 14:34:21,406 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2025-05-20 14:34:21,406 INFO mapred.LocalJobRunner: Waiting for map tasks
2025-05-20 14:34:21,406 INFO mapred.LocalJobRunner: Starting task attempt_local261815800_0001_m_000000_0
2025-05-20 14:34:21,459 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2025-05-20 14:34:21,460 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-05-20 14:34:21,460 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-05-20 14:34:21,469 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
2025-05-20 14:34:21,472 INFO mapred.MapTask: Processing split: hdfs://localhost:9000/klm/1901/0+888198
2025-05-20 14:34:21,472 INFO mapred.MapTask: Input split bytes=104857584(1048575784)
2025-05-20 14:34:21,534 INFO mapred.MapTask: mapred.task.io.sort.mb: 100
2025-05-20 14:34:21,512 INFO mapred.MapTask: soft limit at 83886080
2025-05-20 14:34:21,512 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
2025-05-20 14:34:21,512 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
2025-05-20 14:34:21,586 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2025-05-20 14:34:21,586 INFO mapred.MapTask: Starting flush of map output
2025-05-20 14:34:21,587 INFO mapred.MapTask: Spilling map output
2025-05-20 14:34:21,587 INFO mapred.MapTask: bufstart = 0; bufend = 59076; bufvoid = 104857600
2025-05-20 14:34:21,587 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26188144(104752576); length = 26253/6553600
2025-05-20 14:34:21,595 INFO mapred.MapTask: Flushed spill 0
2025-05-20 14:34:21,600 INFO mapred.Task: Task attempt_local261815800_0001_m_000000_0 is done. And is in the process of committing
2025-05-20 14:34:21,600 INFO mapred.LocalJobRunner: map
2025-05-20 14:34:21,607 INFO mapred.Task: Task attempt_local261815800_0001_m_000000_0 done.
2025-05-20 14:34:21,605 INFO mapred.Task: Final Counters for attempt_local261815800_0001_m_000000_0: Counters: 23
File System Counters
FILE: Number of bytes read=4430
FILE: Number of bytes written=713998
FILE: Number of read operations=5
FILE: Number of large read operations=0
FILE: Number of write operations=4
HDFS: Number of bytes read=888198
HDFS: Number of bytes written=0
HDFS: Number of read operations=5
HDFS: Number of large read operations=0
HDFS: Number of write operations=1
HDFS: Number of bytes read erasure-coded=0
Map-Reduce Framework
  Map Input records=6565
  Map output records=6564
  HDFS: Number of bytes written=8
  HDFS: Number of read operations=10
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=3
  HDFS: Number of bytes read erasure-coded=0
Map-Reduce Framework
  Combine input records=0
  Combine output records=0
  Reduce input groups=1
  Reduce shuffle bytes=72210
  Reduce input records=6564
  Reduce output records=1
  Spilled Records=6564
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=0
  Total committed heap usage (bytes)=633339904
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAGIC=0
  WRONG_REDUCE=0
File Output Format Counters
  Bytes Written=8
2025-05-20 14:34:21,803 INFO mapred.LocalJobRunner: Finishing task: attempt_local261815800_0001_r_000000_0
2025-05-20 14:34:21,803 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-05-20 14:34:22,404 INFO mapreduce.Job: Job job_local261815800_0001 running in uber mode : false
2025-05-20 14:34:22,404 INFO mapreduce.Job: map 100% reduce 100%
2025-05-20 14:34:22,407 INFO mapreduce.Job: Job job_local261815800_0001 completed successfully
2025-05-20 14:34:22,414 INFO mapreduce.Job: Counters: 30
File System Counters
FILE: Number of bytes read=153312
FILE: Number of bytes written=1502000
FILE: Number of read operations=1
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=1776388
HDFS: Number of bytes written=8
HDFS: Number of read operations=15
HDFS: Number of large read operations=0
HDFS: Number of write operations=4
HDFS: Number of bytes read erasure-coded=0
Map-Reduce Framework
  Map Input records=6565
  Map output records=6564
  Map output bytes=59076
  Map output serialized bytes=72210
  Input split bytes=85
  Combine input records=0
  Combine output records=0
  Reduce input groups=1
  Reduce shuffle bytes=72210
  Reduce input records=6564

```

```

Map Input records=6565
Map output records=6564
Map output bytes=59076
Map output initialized bytes=72210
Input split bytes=0
Combine input records=0
Spilled Records=6564
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=0
Total committed map usage (bytes)=633339904
File Input Format Counters
Bytes Read=888198
2025-05-20 14:34:21,605 INFO mapred.LocalJobRunner: Finishing task: attempt_local261815800_0001_m_000000
2025-05-20 14:34:21,606 INFO mapred.LocalJobRunner: map task executor complete.
2025-05-20 14:34:21,607 INFO mapred.LocalJobRunner: Waiting for reduce tasks.
2025-05-20 14:34:21,607 INFO mapred.LocalJobRunner: 59 reduce tasks in progress.
2025-05-20 14:34:21,612 INFO output.PathOutputCommitterFactory: No output Committer factory defined, defaulting to FileOutputCommitterFactory
2025-05-20 14:34:21,612 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-05-20 14:34:21,612 INFO mapred.Task: Using ResourceCalculatorProcessFree: []
2025-05-20 14:34:21,613 INFO mapred.Task: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@cd4d167
2025-05-20 14:34:21,613 INFO mapred.Task: Using MetricsSystem: metrics
2025-05-20 14:34:21,613 INFO reduce.MergeManagerImpl: Merging 1 segments, 8829453212, maxSingleShuffleInit=1457363328, mergeThreshold=3847439368, ioSortFactor=10, memToMemMergeOutputsThreshold=10
2025-05-20 14:34:21,623 INFO reduce.EventFetcher: attempt_local261815800_0001_r_000000_0 Thread started: EventFetcher for fetching Map Completion Events
2025-05-20 14:34:21,635 INFO reduce.LocalFetcher: localFetcher#1 about to shuffle output of map attempt_local261815800_0001_m_000000_0 decompt: 72206 len: 72210 to MEMORY
2025-05-20 14:34:21,637 INFO reduce.InMemoryMapOutput: Read 72206 bytes from map-output for attempt_local261815800_0001_m_000000_0
2025-05-20 14:34:21,637 INFO reduce.MergeManagerImpl: closeInMemoryFiles -> map-output of size: 72206, inMemoryMapOutputs.size() -> 1, commitMemory -> 0, usedMemory ->72206
2025-05-20 14:34:21,638 INFO reduce.MergeManagerImpl: merge is interrupted.. Returning
2025-05-20 14:34:21,638 INFO mapred.LocalJobRunner: 1 / 1 copied.
2025-05-20 14:34:21,638 INFO reduce.MergeManagerImpl: finalMerge called with 1 in-memory map-outputs and 0 on-disk map-outputs
2025-05-20 14:34:21,641 INFO mapred.Merger: Merging 1 sorted segments
2025-05-20 14:34:21,641 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 72199 bytes
2025-05-20 14:34:21,641 INFO reduce.MergeManagerImpl: Merged 1 segments, 72206 bytes to disk to satisfy reduce memory limit
2025-05-20 14:34:21,641 INFO reduce.MergeManagerImpl: Merged 1 files, 72210 bytes from disk
2025-05-20 14:34:21,645 INFO reduce.MergeManagerImpl: merging 0 segments, 0 bytes from memory into reduce
2025-05-20 14:34:21,645 INFO mapred.Merger: Merging 1 sorted segments
2025-05-20 14:34:21,645 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 72199 bytes
2025-05-20 14:34:21,646 INFO mapred.Merger: 1 / 1 copied.
2025-05-20 14:34:21,679 INFO configuration.deprecations: mapred.skip.on is deprecated. Instead, use mapreduce.job.skiprecords
2025-05-20 14:34:21,767 INFO mapred.Task: Task attemp_local261815800_0001_r_000000_0 is done. And is in the process of committing
2025-05-20 14:34:21,771 INFO mapred.Task: Task attemp_local261815800_0001_r_000000_0 committed
2025-05-20 14:34:21,800 INFO output.FileOutputCommitter: Task attempt_local261815800_0001_r_000000_0 is allowed to commit now
2025-05-20 14:34:21,800 INFO output.FileOutputCommitter: Saved output of task 'attempt_local261815800_0001_r_000000_0' to hdfs://localhost:9000/ren
2025-05-20 14:34:21,802 INFO mapred.LocalJobRunner: reduce > reduce
2025-05-20 14:34:21,803 INFO mapred.Task: Final Counters for attempt_local261815800_0001_r_000000_0: Counters: 30
File System Counters
FILE: Number of bytes read=14882
FILE: Number of bytes written=78208
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=888198
HDFS: Number of bytes written=8
HDFS: Number of read operations=10
Reduce shuffle bytes=72210
Reduce Input records=6564
Reduce Output records=1
Spilled Records=13128
Shuffled Maps=1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=0
Total committed heap usage (bytes)=1266679808
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAGIC=0
WRONG_REPLACE=0
File Input Format Counters
Bytes Read=888198
File Output Format Counters
Bytes Written=8
hadoop@bmsccese-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -cat /rem/part-00000
cat: '/rem/part-00000': No such file or directory
hadoop@bmsccese-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -cat /rem/part-r-00000
1981 -46
hadoop@bmsccese-HP-Elite-Tower-800-G9-Desktop-PC:~$ package weatherdataavg;
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> context) throws IOException, InterruptedException {
int max_temp = 0;
int count = 0;
for (IntWritable value : values) {
max_temp += value.get();
count++;
}
context.write(key, new IntWritable(max_temp / count));
}
}

```

b)

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

- Eclipse IDE Title Bar:** eclipse-wc - weatherdatameanmax/src/weatherdatameanmax/MeanMaxReducer.java - Eclipse IDE
- Package Explorer:** Shows the project structure with packages like topnlab, weatherdataawg, and weatherdatameanmax.
- MeanMaxDriver.java:** A Java file containing the main class for the driver.
- MeanMaxReducer.java:** The current file being edited, showing the implementation of the Reducer class.
- MeanMaxMapper.java:** Another Java file in the project.
- Outline View:** Shows the class hierarchy and methods defined in MeanMaxReducer.

MeanMaxReducer.java Content:

```
1 package weatherdatameanmax;
2
3 import java.io.IOException;
4 import org.apache.hadoop.io.IntWritable;
5 import org.apache.hadoop.io.Text;
6 import org.apache.hadoop.mapreduce.Reducer;
7
8 public class MeanMaxReducer extends Reducer<Text, IntWritable, Text, IntWritable> {
9
10    public void reduce(Text key, Iterable<IntWritable> values, Context context) throws IOException, InterruptedException {
11        int maxTemp = Integer.MIN_VALUE;
12        int totalTemp = 0;
13        int count = 0;
14        int days = 0;
15
16        for (IntWritable value : values) {
17            int temp = value.get();
18            if (temp > maxTemp) {
19                maxTemp = temp;
20            }
21
22            count++;
23
24            // After every 3 temperatures, assume a day-group
25            if (count == 3) {
26                totalTemp += maxTemp;
27                maxTemp = Integer.MIN_VALUE;
28                count = 0;
29                days++;
30            }
31        }
32
33        // Avoid division by 0
34        if (days > 0) {
35            context.write(key, new IntWritable(totalTemp / days));
36        } else {
37            context.write(key, new IntWritable(0)); // or handle differently
38        }
39    }
40
41 }
42
```

Terminal Window:

```
hadoop@bmscse-HP-Elite-Tower-800-G9-Desktop-PC: ~
hadoop@bmscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: If you see a CTR-ABORT:
Starting namenodes on [localhost]
localhost: namenode is running as process 9745. Stop it first and ensure /tmp/hadoop-hadoop-namenode.pid file is empty before retry.
Starting datanodes
localhost: datanode is running as process 9928. Stop it first and ensure /tmp/hadoop-hadoop-datanode.pid file is empty before retry.
Starting secondary namenodes [bmscse-HP-Elite-Tower-800-G9-Desktop-PC]
bmscse-HP-Elite-Tower-800-G9-Desktop-PC: secondarynamenode is running as process 10221. Stop it first and ensure /tmp/hadoop-hadoop-secondarynamenode.pid file is empty before retry.
starting resourcemanager
resourcemanager is running as process 18513. Stop it first and ensure /tmp/hadoop-hadoop-resourcemanager.pid file is empty before retry.
starting nodemanagers
localhost: nodemanager is running process 10664. Stop it first and ensure /tmp/hadoop-hadoop-nodemanager.pid file is empty before retry.
hadoop@bmscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop ls -mkdir /tmp
ERROR: ls is not COMMAND, fully qualified CLASSNAME.
Usage: hadoop [OPTIONS] SUBCOMMAND [SUBCOMMAND OPTIONS]
      or hadoop [OPTIONS] CLASSNAME [CLASSNAME OPTIONS]
      where CLASSNAME is a user-provided Java class

OPTIONS is none or any of:
buildpaths          attempt to add class files from build tree
--config dir        Hadoop config directory
--debug             turn on shell script debug mode
--help              usage information
hostnames list[,of,host,names] hosts to use in worker mode
hosts filename      list of hosts to use in worker mode
loglevel level     set the Log4j level for this command
workers            turn on worker mode

SUBCOMMAND is one of:
  Admin Commands:
daemonlog          get/set the log level for each daemon
  Client Commands:
archive            create a Hadoop archive
checknative        check native Hadoop and compression libraries availability
classpath          prints the class path needed to get the Hadoop jar and the required libraries
conftest           validate configuration XML files
credential         interact with credential providers
distch             distributed metadata changes
distcp             copy files or dirs recursively
duftool           operation related to delegation tokens
envvars            display computed Hadoop environment variables
fs                run a generic filesystem user client
gridmix            submit a mix of synthetic job, modeling a profiled from production load
jar <jar>           run a jar file. NOTE: please use "yarn jar" to launch YARN applications, not this command.
minirth           prints the Java library path
```

```

hadoop@bmseccse-HP-Elite-Tower-800-C9-Desktop-PC: ~

jnpipath      applications, not this command.
kdc        prints the kerberos principal path
kdc        Diagnose Kerberos problems
kerbname    show auth_to_local principal conversion
key        manage keys via the KeyProvider
rumenfolder scale a rumen input trace
rumentrace  convert logs into a rumen trace
s3guard     S3 Commands
trace      view and modify Hadoop tracing settings
version    print the version

  Daemon Commands:

kms        run KMS, the Key Management Server
regstrydns run the registry DNS server

SUBCOMMAND may print help when invoked w/o parameters or with -h.
hadoop@bmseccse-HP-Elite-Tower-800-C9-Desktop-PC:~$ hadoop fs -mkdir /omn
hadoop@bmseccse-HP-Elite-Tower-800-C9-Desktop-PC:~$ hadoop fs -copyFromLocal /home/hadoop/Desktop/1901 /omn/1901
hadoop@bmseccse-HP-Elite-Tower-800-C9-Desktop-PC:~$ hadoop jar /home/hadoop/Desktop/weatherdatameanmax.jar weatherdatameanmax.MeanMaxDriver /omn/1901 /ren
2025-05-20 14:53:15.976 INFO LocalMetrics: Local metrics are loaded from conf/hadoop-metrics2.properties
2025-05-20 14:53:15.979 INFO LocalMetricsSystemImpl: Scheduling metrics snapshot every 10 second(s).
2025-05-20 14:53:15.915 INFO Impl.MetricsSystemImpl: JobTracker metrics system started
2025-05-20 14:53:15.679 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-05-20 14:53:15.737 INFO Input.FileInputFormat: Total input files to process : 1
2025-05-20 14:53:15.764 INFO mapreduce.JobSubmitter: number of splits:1
2025-05-20 14:53:15.836 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local2143084439_0001
2025-05-20 14:53:15.841 INFO mapreduce.Job: Job tracking page available: http://localhost:8080/jobs/job_local2143084439_0001
2025-05-20 14:53:15.896 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-05-20 14:53:15.891 INFO mapreduce.Job: Running Job: job_local2143084439_0001
2025-05-20 14:53:15.891 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-05-20 14:53:15.895 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2025-05-20 14:53:15.895 INFO output.FileOutputCommitter: File Output Committer algorithm version is 2
2025-05-20 14:53:15.895 INFO output.PathOutputCommitter: Output Committer will skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-05-20 14:53:15.896 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapred.lib.output.FileOutputCommitter
2025-05-20 14:53:15.981 INFO mapred.LocalJobRunner: Waiting for map tasks
2025-05-20 14:53:15.982 INFO mapred.LocalJobRunner: Starting task: attempt_local2143084439_0001_m_000000_0
2025-05-20 14:53:15.996 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2025-05-20 14:53:15.996 INFO output.FileOutputCommitter: File Output Committer algorithm version is 2
2025-05-20 14:53:15.996 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-05-20 14:53:16.001 INFO mapred.Task: ResourceCalculatorBackend free: [ ]
2025-05-20 14:53:16.007 INFO mapred.Task: Processing split: hdfs://localhost:9008/omn/1901:0+888190
2025-05-20 14:53:16.044 INFO mapred.MapTask: Processing split: [kv1 26214396/104857584]
2025-05-20 14:53:16.044 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2025-05-20 14:53:16.044 INFO mapred.MapTask: soft limit at 83886088
2025-05-20 14:53:16.044 INFO mapred.MapTask: bufstart = 0; bufend = 104857600
2025-05-20 14:53:16.044 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 2618814(10475276); length = 6553600
2025-05-20 14:53:16.044 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2025-05-20 14:53:16.118 INFO mapred.LocalJobRunner:
2025-05-20 14:53:16.119 INFO mapred.MapTask: Starting flush of map output
2025-05-20 14:53:16.119 INFO mapred.MapTask: Spilling map output
2025-05-20 14:53:16.119 INFO mapred.MapTask: bufstart = 0; bufend = 45948; bufvoid = 104857600
2025-05-20 14:53:16.119 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 2618814(10475276); length = 6553600
2025-05-20 14:53:16.120 INFO mapred.Task: Task attempt_local2143084439_0001_M_000000_0 is done. And is in the process of committing
2025-05-20 14:53:16.120 INFO mapred.LocalJobRunner: main

hadoop@bmseccse-HP-Elite-Tower-800-C9-Desktop-PC: ~

2025-05-20 14:53:16.145 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-05-20 14:53:16.146 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin org.apache.hadoop.mapreduce.task.reduce.Shuffle@6d4ffe0
2025-05-20 14:53:16.146 INFO reduce.EventFetcher: EventFetcher initialized
2025-05-20 14:53:16.155 INFO reduce.MergeManagerImpl: Merging 1 segments into 1 sorted segments
2025-05-20 14:53:16.170 INFO reduce.MergeManagerImpl: attempt_local2143084439_0001_r_000000_0 Thread started: EventFetcher for fetching Map Completion Events
2025-05-20 14:53:16.172 INFO reduce.InMemoryMapOutput: Read 59078 bytes from map-output for attempt_local2143084439_0001_m_000000_0
2025-05-20 14:53:16.172 INFO reduce.InMemoryMapOutput: closeInMemoryFile() > map-output of size: 59078, InMemoryMapOutputs.size() -> 1, commitMemory -> 0, usedMemory ->59078
2025-05-20 14:53:16.173 INFO reduce.MergeManagerImpl: EventFetcher is interrupted.. Returning
2025-05-20 14:53:16.173 INFO reduce.MergeManagerImpl: 1 copy completed
2025-05-20 14:53:16.174 INFO reduce.MergeManagerImpl: finalMerge called with 1 in-memory map-outputs and 0 on-disk map-outputs
2025-05-20 14:53:16.176 INFO mapred.Merger: Merging 1 sorted segments
2025-05-20 14:53:16.176 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 59073 bytes
2025-05-20 14:53:16.180 INFO reduce.MergeManagerImpl: Merged 1 segments, 59078 bytes to disk to satisfy reduce memory limit
2025-05-20 14:53:16.181 INFO reduce.MergeManagerImpl: Merging 1 files, 59082 bytes from disk
2025-05-20 14:53:16.181 INFO reduce.MergeManagerImpl: Merged 0 segments, 0 bytes from memory into reduce
2025-05-20 14:53:16.182 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 59073 bytes
2025-05-20 14:53:16.182 INFO mapred.LocalJobRunner: 1 / 1 copied
2025-05-20 14:53:16.182 INFO configuration.deprecations: mapred.skip.on is deprecated. Instead, use mapreduce.job.skipprecords
2025-05-20 14:53:16.273 INFO mapred.Task: Task attempt_local2143084439_0001_r_000000_0 is done. And is in the process of committing
2025-05-20 14:53:16.273 INFO mapred.Task: Task attempt_local2143084439_0001_r_000000_0 is committed
2025-05-20 14:53:16.289 INFO mapred.Task: Task attempt_local2143084439_0001_r_000000_0 is allowed to commit now
2025-05-20 14:53:16.289 INFO fileoutputcommitter: Saved output of task 'attempt_local2143084439_0001_r_000000_0' to hdfs://localhost:9000/ren
2025-05-20 14:53:16.290 INFO mapred.Task: Task 'attempt_local2143084439_0001_r_000000_0' done.
2025-05-20 14:53:16.290 INFO mapred.Task: Task 'attempt_local2143084439_0001_r_000000_0' done.

File System Counters
  Bytes Read: 59073
  Bytes Written: 76193
  File: Number of bytes read:122769
  File: Number of bytes written:76193
  File: Number of read operations=0
  File: Number of large read operations=0
  File: Number of write operations=0
  HDFS: Number of bytes read:888190
  HDFS: Number of bytes written:81
  HDFS: Number of read operations=10
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=3
  HDFS: Number of bytes read erasure-coded=0

Map-Reduce Framework
  Combine input records=0
  Combine output records=0
  Combiner input records=0
  Reduce input groups=12
  Reduce shuffle bytes=59082
  Reduce input records=6564
  Reduce output records=12
  Splitted Records=6564
  Shuffled Maps=0
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=0
  Total committed heap usage (bytes)=526385152

Shuffle Errors
  File ID mismatch: 0
  File corruption: 0

```

Screenshot captured
You can paste the image from the clipboard.

```

2025-05-20 14:53:15,982 INFO mapred.LocalJobRunner: Starting task: attempt_local2143084439_0001_m_000000_0
2025-05-20 14:53:15,986 INFO output.FileOutputCommitter: File Output Committer algorithm version is 2
2025-05-20 14:53:15,986 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-05-20 14:53:16,004 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
2025-05-20 14:53:16,007 INFO mapred.MapTask: Processing split: hdfs://localhost:9000/omn/1901:0+888190
2025-05-20 14:53:16,044 INFO mapred.MapTask: (EQUATOR) 0 kv1 26214396(104857584)
2025-05-20 14:53:16,044 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2025-05-20 14:53:16,044 INFO mapred.MapTask: mapreduce.map.memory.mb: 1000
2025-05-20 14:53:16,044 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
2025-05-20 14:53:16,044 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
2025-05-20 14:53:16,046 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2025-05-20 14:53:16,118 INFO mapred.LocalJobRunner:
2025-05-20 14:53:16,119 INFO mapred.MapTask: Starting flush of map output
2025-05-20 14:53:16,119 INFO mapred.MapTask: Spilling map output
2025-05-20 14:53:16,119 INFO mapred.MapTask: bufstart = 0; bufend = 45948; bufvoid = 104857600
2025-05-20 14:53:16,128 INFO mapred.MapTask: Flushed spill 0
2025-05-20 14:53:16,133 INFO mapred.Task: Task 'attempt_local2143084439_0001_m_000000_0' is done. And is in the process of committing
2025-05-20 14:53:16,135 INFO mapred.LocalJobRunner: map
2025-05-20 14:53:16,135 INFO mapred.Task: Final Counters for attempt_local2143084439_0001_m_000000_0: Counters: 23
  File System Counters
    FILE: Number of bytes read=4573
    FILE: Number of bytes written=704111
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of writes operations=0
    HDFS: Number of bytes read=888190
    HDFS: Number of bytes written=0
    HDFS: Number of read operations=5
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=1
    HDFS: Number of bytes read erasure-coded=0
  Map-Reduce Framework
    Map Input records=565
    Map output records=6564
    Map output bytes=45948
    Map output materialized bytes=59082
    Input split bytes=95
    Combine input records=0
    Spilled Records=6564
    Failed Shuffles=0
    Merged Map outputs=0
    GC time elapsed (ms)=0
    Total committed heap usage (bytes)=526385152
  File Input Format Counters
    Bytes Read=888190
2025-05-20 14:53:16,138 INFO mapred.LocalJobRunner: Finishing task: attempt_local2143084439_0001_m_000000_0
2025-05-20 14:53:16,139 INFO mapred.LocalJobRunner: map task executor complete.
2025-05-20 14:53:16,140 INFO mapred.LocalJobRunner: Waiting for reduce tasks
2025-05-20 14:53:16,140 INFO mapred.LocalJobRunner: Starting task: attempt_local2143084439_0001_r_000000_0
2025-05-20 14:53:16,145 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2025-05-20 14:53:16,145 INFO output.FileOutputCommitter: File Output Committer algorithm version is 2
2025-05-20 14:53:16,145 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-05-20 14:53:16,146 INFO mapred.Task: Reopen ResourceCalculatorProcessTree : [ ]

```

Screenshot captured
You can paste the image from the clipboard.

```

FILE: Number of bytes written=763193
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=888198
HDFS: Number of bytes written=81
HDFS: Number of read operations=10
HDFS: Number of large read operations=0
HDFS: Number of write operations=3
HDFS: Number of bytes read erasure-coded=0

Map-Reduce Framework
  Combine input records=0
  Combine output records=0
  Reduce input groups=1
  Reduce shuffle bytes=59882
  Reduce input records=6564
  Spilled Records=6564
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=0
  Total committed heap usage (bytes)=526385152

Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0

File Output Format Counters
  Bytes Written=81

2025-05-20 14:53:16,298 INFO mapred.LocalJobRunner: Finishing task: attempt_local21430884439_0001_r_000000_0
2025-05-20 14:53:16,298 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-05-20 14:53:16,897 INFO mapreduce.Job: Job local21430884439_0001 running in uber mode : false
2025-05-20 14:53:16,897 INFO mapreduce.Job: map 100% reduced 100%
2025-05-20 14:53:16,898 INFO mapreduce.Job: Job local21430884439_0001 completed successfully
2025-05-20 14:53:16,908 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=127342
    FILE: Number of bytes written=1467384
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=1776380
    HDFS: Number of bytes written=81
    HDFS: Number of read operations=15
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=4
    HDFS: Number of bytes read erasure-coded=0

Map-Reduce Framework
  Map Input records=565
  Map output records=6564
  Map output bytes=45948
  Map output materialized bytes=59882
  Input split bytes=95
  Combine output records=0
  Combine output records=0
  Reduce input groups=12
  Reduce shuffle bytes=59882
  Reduce input records=6564
  Reduce output records=1
  Spilled Records=13128
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=0
  Total committed heap usage (bytes)=526385152

Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0

File Input Format Counters
  Bytes Written=81
  File Output Format Counters
  Bytes Written=81

hadoop@bmsccse-HP-Elite-Tower-800-G9-Desktop-PC:~
```

Lab-05

① Perform the foll. DB operations using

Cassandra

② Create keyspace by name Employee

Ans → Create keyspace employee with
replication = { 'class': 'SimpleStrategy',
replication factor: 1 }

③ Create a column family by name employee-

info, with attributes Emp-ID, Primary key,
Emp-name, Designation, Date-of-Joining,

Salary, Dept-name

Ans → Create Table employee-Info {

Emp-ID int,
Emp-name text,
Designation text,
Date-of-Joining date,

Salary double,
Dept-name text,

PRIMARY KEY (Dept-name, Salary
Emp-ID)

With Clustering Order By
(salary DESC);

③ Insert values at start batch and store it

Begin Batch

Insert INTO employee-Info (Dept-name,
Salary, Emp-ID, Emp-name, Designation,
Date-of-Joining)

Values (R&D, 75000, 121, Alicia,
Software Eng., '2021-01-01');

Select * from employee-Info;

Output

Output

Dept-name	Salary	Emp-ID	Designation	Emp-name
HR	90000	123	Manager	Carol
IT	70000	121	Developer	Alicia
R&D	75000	121	Soft. Eng.	Alicia
Finance	60000	122	Analyst	Bob
Analytics	65000	122	Data Analyst	Bob

④ Sort the details of employee records based
on Salary

⑤ Alter schema of table employee-Info
to add a column Project which stores
a set of Project done by corr. employee.

ALTER table employee-Info ADD
Projects set (text);

6. Update the altered table to add project names

→ UPDATE employee-Info
SET Projects = {'Project Alpha', 'Project Beta'}

where Dept-Name = 'R&D' AND

Salary = 75000 AND Emp-ID = 121,

7. Create a TTL of 15 seconds to display the values of employees

→ Update employee-Info using TTL 15 SET

Emp-Name = 'Temporary_name'

where Dept-name = 'HR' AND

Salary > 90000 AND Emp-ID > 123,

⑧ Perform the foll. DB operation using CASSANDRA

1. Create a keyspace by name Library

→ create keyspace library with

replication = {'class': 'SimpleStrategy',
replication_factor: 3};

2. Create a column family by name Library-with

attributes Stud-ID, primary key, counter-value,
of type counter, Stud-Name, book-Name,
Book-ID, Date-of-Issue,

(3. Insert values into what two tables?)

BEGIN Batch

Insert into Employee-Info (Dept-name,
Salary-Emp-ID),

Create table Library-Info

stud-ID int,

stud-name text,

Book-Name text,

Book-ID text,

Date-of-issue date

PRIMARY KEY (stud-ID, Book-Name);

Create table Book-Counter(

stud-ID int,

Book-name text,

Counter-value counter,

PRIMARY KEY (stud-ID, Book-Name);

(Q) Display Details & increase the counter,

→ Select * from Library-Info;

Select * from Book-Counter where

Update Book-Counter SET Counter-value

Counter-value + where stud-ID = 112 &

Book-Name = 'BDA';

8) Show student with ID 112 has taken
Book "BDA" 2 times

Ans → select counter-value from Book-Counter
where stud-ID = 112 & Book-Name=BDA.

Output

counter-value

2

9) Cqlsh - e "Copy Library.Library-Info.
INTO 'Library-info.csv'
with HEADER='TRUE';"

10) Cqlsh - e "Copy Library.Library-Info
(Book-name, Id, issue) FROM 'library
info.csv' with HEADER=TRUE;"

Lab 6: Hadoop

Question: Implement Wordcount program on Hadoop framework

Code with Output:

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

- File Explorer:** Shows a project named "wordcount" containing three files: WCDriver.java, WCMapper.java, and WCReducer.java.
- Code Editor:** Displays the WCDriver.java code, which is a Hadoop WordCount implementation.
- Output Console:** Shows the command-line output of the Hadoop job execution.

```
WCDriver.java
1 import java.io.IOException;
2 import org.apache.hadoop.conf.Configuration;
3 import org.apache.hadoop.fs.Path;
4 import org.apache.hadoop.io.IntWritable;
5 import org.apache.hadoop.io.Text;
6 import org.apache.hadoop.mapred.FileInputFormat;
7 import org.apache.hadoop.mapred.FileOutputFormat;
8 import org.apache.hadoop.mapred.JobClient;
9 import org.apache.hadoop.mapred.JobConf;
10 import org.apache.hadoop.util.Tool;
11 import org.apache.hadoop.util.ToolRunner;
12
13 public class WCDriver extends Configuration implements Tool {
14
15     public int run(String[] args) throws IOException {
16         if (args.length < 2) {
17             System.out.println("Please provide input and output paths");
18             return -1;
19         }
20
21         JobConf conf = new JobConf(WCDriver.class);
22         conf.setJobName("WordCount");
23         conf.setJarByClass(WCDriver.class); // Ensures job runs from correct JAR
24
25         FileInputFormat.setInputPaths(conf, new Path(args[0]));
26         FileOutputFormat.setOutputPath(conf, new Path(args[1]));
27
28         conf.setMapperClass(WCMapper.class);
29         conf.setReducerClass(WCReducer.class);
30
31         conf.setMapOutputKeyClass(Text.class);
32     }
33 }
```

```
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ hadoop fs -mkdir /rsg
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ hadoop fs -copyFromLocal /home/hadoop/Desktop/sample.txt /rsg/sample.txt
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ hadoop jar /home/hadoop/Desktop/wordcount.jar WCDriver /rsg/sample.txt /result
2025-05-06 15:05:01,260 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-05-06 15:05:01,299 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-05-06 15:05:01,299 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-05-06 15:05:01,305 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2025-05-06 15:05:01,365 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-05-06 15:05:01,414 INFO mapred.FileInputFormat: Total input files to process : 1
2025-05-06 15:05:01,445 INFO mapreduce.JobSubmitter: number of splits:1
2025-05-06 15:05:01,511 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local90897529_0001
2025-05-06 15:05:01,511 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-05-06 15:05:01,565 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-05-06 15:05:01,566 INFO mapreduce.Job: Running job: job_local90897529_0001
2025-05-06 15:05:01,566 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-05-06 15:05:01,567 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapred.FileOutputCommitter
2025-05-06 15:05:01,569 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
```

```

hadoop@bmscse-HP-Elite-Tower-800-G9-Desktop-PC: ~
2025-05-06 15:05:01,299 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-05-06 15:05:01,299 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-05-06 15:05:01,305 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2025-05-06 15:05:01,365 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-05-06 15:05:01,414 INFO mapred.FileInputFormat: Total input files to process : 1
2025-05-06 15:05:01,445 INFO mapreduce.JobSubmitter: number of splits:1
2025-05-06 15:05:01,511 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local90897529_0001
2025-05-06 15:05:01,511 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-05-06 15:05:01,565 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-05-06 15:05:01,566 INFO mapreduce.Job: Running job: job_local90897529_0001
2025-05-06 15:05:01,566 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-05-06 15:05:01,567 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapred.FileOutputCommitter
2025-05-06 15:05:01,569 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-05-06 15:05:01,569 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false , ignore cleanup failures: false
2025-05-06 15:05:01,606 INFO mapred.LocalJobRunner: Waiting for map tasks
2025-05-06 15:05:01,607 INFO mapred.LocalJobRunner: Starting task: attempt_local90897529_0001_m_000000_0
2025-05-06 15:05:01,618 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-05-06 15:05:01,618 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false , ignore cleanup failures: false
2025-05-06 15:05:01,624 INFO mapred.Task: Using ResourceCalculatorProcessTree : []
2025-05-06 15:05:01,631 INFO mapred.MapTask: Processing split: hdfs://localhost:9000/rsg/sample.txt:0+89
2025-05-06 15:05:01,640 INFO mapred.MapTask: numReduceTasks: 1
2025-05-06 15:05:01,671 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
2025-05-06 15:05:01,671 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2025-05-06 15:05:01,671 INFO mapred.MapTask: soft limit at 83886080
2025-05-06 15:05:01,671 INFO mapred.MapTask: bufstart = 0; bufend = 104857600
2025-05-06 15:05:01,671 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
2025-05-06 15:05:01,673 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2025-05-06 15:05:01,742 INFO mapred.LocalJobRunner: 
2025-05-06 15:05:01,742 INFO mapred.MapTask: Starting flush of map output
2025-05-06 15:05:01,742 INFO mapred.MapTask: Spilling map output
2025-05-06 15:05:01,742 INFO mapred.MapTask: bufstart = 0; bufend = 169; bufvoid = 104857600
2025-05-06 15:05:01,742 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214320(104857280); length = 77/6553600
2025-05-06 15:05:01,745 INFO mapred.MapTask: Finished spill 0
2025-05-06 15:05:01,751 INFO mapred.Task: Task:attempt_local90897529_0001_m_000000_0 is done. And is in the process of committing
2025-05-06 15:05:01,753 INFO mapred.LocalJobRunner: hdfs://localhost:9000/rsg/sample.txt:0+89
2025-05-06 15:05:01,753 INFO mapred.Task: Task 'attempt_local90897529_0001_m_000000_0' done.
2025-05-06 15:05:01,756 INFO mapred.Task: Final Counters for attempt_local90897529_0001_m_000000_0: Counters: 23
    File System Counters
        FILE: Number of bytes read=4273
        FILE: Number of bytes written=639534
        FILE: Number of read operations=0

hadoop@bmscse-HP-Elite-Tower-800-G9-Desktop-PC: ~
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Output Format Counters
Bytes Written=69
2025-05-06 15:05:01,897 INFO mapred.LocalJobRunner: Finishing task: attempt_local90897529_0001_r_000000_0
2025-05-06 15:05:01,897 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-05-06 15:05:02,569 INFO mapreduce.Job: Job job_local90897529_0001 running in uber mode : false
2025-05-06 15:05:02,572 INFO mapreduce.Job: map 100% reduce 100%
2025-05-06 15:05:02,574 INFO mapreduce.Job: Job job_local90897529_0001 completed successfully
2025-05-06 15:05:02,584 INFO mapreduce.Job: Counters: 36
    File System Counters
        FILE: Number of bytes read=9008
        FILE: Number of bytes written=1279283
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=178
        HDFS: Number of bytes written=69
        HDFS: Number of read operations=15
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=4
        HDFS: Number of bytes read erasure-coded=0
    Map-Reduce Framework
        Map input records=5
        Map output records=20
        Map output bytes=169
        Map output materialized bytes=215
        Input split bytes=88
        Combine input records=0
        Combine output records=0
        Reduce input groups=10
        Reduce shuffle bytes=215
        Reduce input records=20
        Reduce output records=10
        Spilled Records=40
        Shuffled Maps =1
        Failed Shuffles=0
        Merged Map outputs=1
        GC time elapsed (ms)=0
        Total committed heap usage (bytes)=1052770304

```

```

hadoop@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC: ~
FILE: Number of write operations=0
HDFS: Number of bytes read=89
HDFS: Number of bytes written=0
HDFS: Number of read operations=5
HDFS: Number of large read operations=0
HDFS: Number of write operations=1
HDFS: Number of bytes read erasure-coded=0
Map-Reduce Framework
  Map input records=5
  Map output records=20
  Map output bytes=169
  Map output materialized bytes=215
  Input split bytes=88
  Combine input records=0
  Spilled Records=20
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=0
  Total committed heap usage (bytes)=526385152
  File Input Format Counters
    Bytes Read=89
2025-05-06 15:05:01,756 INFO mapred.LocalJobRunner: Finishing task: attempt_local90897529_0001_m_000000_0
2025-05-06 15:05:01,757 INFO mapred.LocalJobRunner: map task executor complete.
2025-05-06 15:05:01,758 INFO mapred.LocalJobRunner: Waiting for reduce tasks
2025-05-06 15:05:01,758 INFO mapred.LocalJobRunner: Starting task: attempt_local90897529_0001_r_000000_0
2025-05-06 15:05:01,762 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-05-06 15:05:01,762 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false , ignore cleanup failures: false
2025-05-06 15:05:01,762 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
2025-05-06 15:05:01,763 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@636a90e9
2025-05-06 15:05:01,764 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2025-05-06 15:05:01,771 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=5827985408, maxSingleShuffleLimit=1456996352, mergeThreshold=3846470400, ioSortFactor=10, memToMemMergeOutputsThreshold=10
2025-05-06 15:05:01,772 INFO reduce.EventFetcher: attempt_local90897529_0001_r_000000_0 Thread started: EventFetcher for fetching Map Completion Events
2025-05-06 15:05:01,785 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map attempt_local90897529_0001_m_000000_0 decomp: 211 len: 215 to MEMORY
2025-05-06 15:05:01,787 INFO reduce.InMemoryMapOutput: Read 211 bytes from map-output for attempt_local90897529_0001_m_000000_0
2025-05-06 15:05:01,788 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of size: 211, inMemoryMapOutputs.size() -> 1, committedMemory -> 0, usedMemory ->211
2025-05-06 15:05:01,788 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning
2025-05-06 15:05:01,789 INFO mapred.LocalJobRunner: 1 / 1 copied.
2025-05-06 15:05:01,789 INFO reduce.MergeManagerImpl: finalMerge called with 1 in-memory map-outputs and 0 on-disk map-outputs
2025-05-06 15:05:01,792 INFO mapred.Merger: Merging 1 sorted segments

```



```

hadoop@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC: ~
2025-05-06 15:05:01,792 INFO reduce.MergeManagerImpl: Merged 1 segments, 211 bytes to disk to satisfy reduce memory limit
2025-05-06 15:05:01,793 INFO reduce.MergeManagerImpl: Merging 1 files, 215 bytes from disk
2025-05-06 15:05:01,793 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce
2025-05-06 15:05:01,793 INFO mapred.Merger: Merging 1 sorted segments
2025-05-06 15:05:01,793 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 205 bytes
2025-05-06 15:05:01,793 INFO mapred.LocalJobRunner: 1 / 1 copied.
2025-05-06 15:05:01,867 INFO mapred.Task: Task:attempt_local90897529_0001_r_000000_0 is done. And is in the process of committing
2025-05-06 15:05:01,869 INFO mapred.LocalJobRunner: 1 / 1 copied.
2025-05-06 15:05:01,869 INFO mapred.Task: Task attempt_local90897529_0001_r_000000_0 is allowed to commit now
2025-05-06 15:05:01,894 INFO output.FileOutputCommitter: Saved output of task 'attempt_local90897529_0001_r_000000_0' to hdfs://localhost:9000/result
2025-05-06 15:05:01,896 INFO mapred.LocalJobRunner: reduce > reduce
2025-05-06 15:05:01,896 INFO mapred.Task: Task 'attempt_local90897529_0001_r_000000_0' done.
2025-05-06 15:05:01,897 INFO mapred.Task: Final Counters for attempt_local90897529_0001_r_000000_0: Counters: 30
  File System Counters
    FILE: Number of bytes read=4735
    FILE: Number of bytes written=639749
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=89
    HDFS: Number of bytes written=69
    HDFS: Number of read operations=10
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=3
    HDFS: Number of bytes read erasure-coded=0
Map-Reduce Framework
  Combine input records=0
  Combine output records=0
  Reduce input groups=10
  Reduce shuffle bytes=215
  Reduce input records=20
  Reduce output records=10
  Spilled Records=20
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=0
  Total committed heap usage (bytes)=526385152
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0

```

```
hadoop@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC: ~
HDFS: Number of write operations=4
HDFS: Number of bytes read erasure-coded=0
Map-Reduce Framework
  Map input records=5
  Map output records=20
  Map output bytes=169
  Map output materialized bytes=215
  Input split bytes=88
  Combine input records=0
  Combine output records=0
  Reduce input groups=10
  Reduce shuffle bytes=215
  Reduce input records=20
  Reduce output records=10
  Spilled Records=40
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=0
  Total committed heap usage (bytes)=1052770304
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=89
File Output Format Counters
  Bytes Written=69
Exit Code: 0
hadoop@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC: $ hadoop fs -cat /result/part-00000
are      1
brother  1
family   1
hi       1
how      5
is       4
job      1
sister   1
you      1
your     4
hadoop@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC: $
```

Lab-06

Word Count Program Map Reduce

1) mapper.java

```
public class mapperCode extends mapper
<long, Text, Text, IntWritable>
public void map (longWritable key,
Text value, context context) throws IOException {
    String[] words = value.toString().split ("\\.");
    for (String word : words) {
        context.write (new Text (word) new
IntWritable (1));
    }
}
```

2) Reducer Code.java

```
public class ReducerCode extends Reducer
<Text, IntWritable, Text, IntWritable>
throws IOException {
    int sum = 0;
```

```
for (IntWritable val : values) {
    sum += val.get();
}
```

```
context.write (key, new IntWritable
(sum));
```

3

3) Driver Code

```
public class DriverCode {
    public static void main (String [] args)
        throws Exception {
        Configuration conf = new Configuration();
        Job job = Job get Instance (conf,
            "word");
        job.setMapperClass (Mapper.class);
        job.setReducerClass (Reducer.class);
        job.setOutputKeyClass (Text.class);
        job.setOutputValueClass (Invertable.class);
        FileInputFormat.add Input (Job, new Path
            ("src/test/resources/test1"));
        FileOutputFormat.setOutputPath (Job, new Path
            ("src/test/resources/test1"));
        System.exit (job.waitForCompletion
            (true));
    }
}
```

Hadoop codes:

```

> hdfs dfs -mkdir -p /input
> hdfs dfs -put /home/hadoop/sample.txt /input
> hadoop jar wordcount.jar DriverCode /input/
sample.txt /output
> hdfs dfs -cat /output/part-r-00000

```

3 <file:///user/hadoop/wordcount/output/part-r-00000>

Output

Are 3 same local file storing

Hadoop 1 same local file storing

Hello 1 different local word count

How 2

I 1 different local = how local

am 1

an 1

are 2

assistant 1 same local

can 2

Coding 1

days 1

during 1

engineer 1

for 2

help 1

interview 1

looking 2

questions 1

strengths 1

there 1

what 2

you 6

Your 1

Lab 7: Hadoop

Question: 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.

Code with Output:

```
hadoop@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC:~$ start -all.sh
Command 'start' not found, did you mean:
  command 'stars' from snap stars (2.7jrc3)
  command 'rstart' from deb x11-session-utils (7.7+4build2)
  command 'kstart' from deb kde-cli-tools (4:5.24.4-0ubuntu1)
  command 'startx' from deb xinit (1.4.1-0ubuntu4)
  command 'stat' from deb coreutils (8.32-4.1ubuntu1.2)
  command 'tart' from deb tart (3.10-1build1)
See 'snap info <snapname>' for additional versions.
hadoop@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC:~$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [bmscsece-HP-Elite-Tower-800-G9-Desktop-PC]
Starting resourcemanager
Starting nodemanagers
hadoop@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -mkdir /mno
mkdir: Cannot create directory /mno. Name node is in safe mode.
hadoop@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC:~$ ^C
hadoop@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC:~$ ^C
hadoop@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC:~$ hdfs dfsadmin -safemode get
Safe mode is OFF
hadoop@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -mkdir /mno
hadoop@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -copyFromLocal /home/hadoop/Desktop/sample.txt /mno/sample.txt
hadoop@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop jar /home/hadoop/Desktop/topn.jar TopN /mno/sample.txt /res
Exception in thread "main" java.lang.ClassNotFoundException: TopN
  at java.base/java.net.URLClassLoader.findClass(URLClassLoader.java:476)
  at java.base/java.lang.ClassLoader.loadClass(ClassLoader.java:594)
  at java.base/java.lang.ClassLoader.loadClass(ClassLoader.java:527)
  at java.base/java.lang.Class.forName0(Native Method)
  at java.base/java.lang.Class.forName(Class.java:398)
  at org.apache.hadoop.util.RunJar.run(RunJar.java:321)
  at org.apache.hadoop.util.RunJar.main(RunJar.java:241)
hadoop@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop jar /home/hadoop/Desktop/topn.jar topnlab /mno/sample.txt /res
Exception in thread "main" java.lang.ClassNotFoundException: topnlab
  at java.base/java.net.URLClassLoader.findClass(URLClassLoader.java:476)
  at java.base/java.lang.ClassLoader.loadClass(ClassLoader.java:594)
  at java.base/java.lang.ClassLoader.loadClass(ClassLoader.java:527)
  at java.base/java.lang.Class.forName0(Native Method)
  at java.base/java.lang.Class.forName(Class.java:398)
  at org.apache.hadoop.util.RunJar.run(RunJar.java:321)
  at org.apache.hadoop.util.RunJar.main(RunJar.java:241)
hadoop@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop jar /home/hadoop/Desktop/topn.jar TopN /mno/sample.txt /res
Exception in thread "main" java.lang.NoClassDefFoundError: topnlab/TopN (wrong name: TopN)
  at java.base/java.lang.ClassLoader.defineClass1(Native Method)
```

```
hadoop@bmscse-HP-Elite-Tower-800-G9-Desktop-PC: ~
at java.base/java.lang.ClassLoader.loadClass(ClassLoader.java:594)
at java.base/java.lang.ClassLoader.loadClass(ClassLoader.java:527)
at java.base/java.lang.Class.forName0(Native Method)
at java.base/java.lang.Class.forName(Class.java:398)
at org.apache.hadoop.util.RunJar.run(RunJar.java:321)
at org.apache.hadoop.util.RunJar.main(RunJar.java:241)
hadoop@bmscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop jar /home/hadoop/Desktop/topn.jar topnlab /mnosample.txt /res
Exception in thread "main" java.lang.ClassNotFoundException: topnlab
    at java.base/java.net.URLClassLoader.findClass(URLClassLoader.java:476)
    at java.base/java.lang.ClassLoader.loadClass(ClassLoader.java:594)
    at java.base/java.lang.ClassLoader.loadClass(ClassLoader.java:527)
    at java.base/java.lang.Class.forName0(Native Method)
    at java.base/java.lang.Class.forName(Class.java:398)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:321)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:241)
hadoop@bmscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop jar /home/hadoop/Desktop/topn.jar TopN /mnosample.txt /res
Exception in thread "main" java.lang.NoClassDefFoundError: topnlab/TopN (wrong name: TopN)
    at java.base/java.lang.ClassLoader.defineClass1(Native Method)
    at java.base/java.lang.ClassLoader.defineClass(ClassLoader.java:1022)
    at java.base/java.security.SecureClassLoader.defineClass(SecureClassLoader.java:174)
    at java.base/java.net.URLClassLoader.defineClass(URLClassLoader.java:555)
    at java.base/java.net.URLClassLoader$1.run(URLClassLoader.java:458)
    at java.base/java.net.URLClassLoader$1.run(URLClassLoader.java:452)
    at java.base/java.security.AccessController.doPrivileged(Native Method)
    at java.base/java.net.URLClassLoader.findClass(URLClassLoader.java:451)
    at java.base/java.lang.ClassLoader.loadClass(ClassLoader.java:594)
    at java.base/java.lang.ClassLoader.loadClass(ClassLoader.java:527)
    at java.base/java.lang.Class.forName0(Native Method)
    at java.base/java.lang.Class.forName(Class.java:398)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:321)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:241)
hadoop@bmscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop jar /home/hadoop/Desktop/topn.jar topnlab.TopN /mnosample.txt /res
2025-05-20 13:58:09,506 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-05-20 13:58:09,545 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-05-20 13:58:09,545 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-05-20 13:58:09,658 INFO input.FileInputFormat: Total input files to process : 1
2025-05-20 13:58:09,709 INFO mapreduce.JobSubmitter: number of splits:1
2025-05-20 13:58:09,777 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local408680812_0001
2025-05-20 13:58:09,778 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-05-20 13:58:09,836 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-05-20 13:58:09,837 INFO mapreduce.Job: Running job: job_local408680812_0001
2025-05-20 13:58:09,838 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-05-20 13:58:09,841 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2025-05-20 13:58:09,842 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-05-20 13:58:09,842 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
```

```
hadoop@bmscse-HP-Elite-Tower-800-G9-Desktop-PC: ~
2025-05-20 13:58:09,842 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-05-20 13:58:09,842 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-05-20 13:58:09,842 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2025-05-20 13:58:09,884 INFO mapred.LocalJobRunner: Waiting for map tasks
2025-05-20 13:58:09,885 INFO mapred.LocalJobRunner: Starting task: attempt_local408680812_0001_m_000000_0
2025-05-20 13:58:09,895 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2025-05-20 13:58:09,895 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-05-20 13:58:09,895 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-05-20 13:58:09,903 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
2025-05-20 13:58:09,906 INFO mapred.MapTask: Processing split: hdfs://localhost:9000/mno/sample.txt:0+75
2025-05-20 13:58:09,945 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
2025-05-20 13:58:09,945 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2025-05-20 13:58:09,945 INFO mapred.MapTask: soft limit at 83886080
2025-05-20 13:58:09,945 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
2025-05-20 13:58:09,945 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
2025-05-20 13:58:09,947 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2025-05-20 13:58:10,006 INFO mapred.LocalJobRunner:
2025-05-20 13:58:10,007 INFO mapred.MapTask: Starting flush of map output
2025-05-20 13:58:10,007 INFO mapred.MapTask: Spilling map output
2025-05-20 13:58:10,007 INFO mapred.MapTask: bufstart = 0; bufend = 135; bufvoid = 104857600
2025-05-20 13:58:10,007 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214340(104857360); length = 57/6553600
2025-05-20 13:58:10,010 INFO mapred.MapTask: Finished spill 0
2025-05-20 13:58:10,014 INFO mapred.Task: Task 'attempt_local408680812_0001_m_000000_0' is done. And is in the process of committing
2025-05-20 13:58:10,016 INFO mapred.LocalJobRunner: map
2025-05-20 13:58:10,017 INFO mapred.Task: Task 'attempt_local408680812_0001_m_000000_0' done.
2025-05-20 13:58:10,020 INFO mapred.Task: Final Counters for attempt_local408680812_0001_m_000000_0: Counters: 23
    File System Counters
        FILE: Number of bytes read=7513
        FILE: Number of bytes written=645435
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=75
        HDFS: Number of bytes written=0
        HDFS: Number of read operations=5
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=1
        HDFS: Number of bytes read erasure-coded=0
    Map-Reduce Framework
        Map input records=2
        Map output records=15
        Map output bytes=135
```

```
hadoop@bmscse-HP-Elite-Tower-800-G9-Desktop-PC: ~
2025-05-20 13:58:10,168 INFO mapred.Task: Task:attempt_local408680812_0001_r_000000_0 is done. And is in the process of committing.
2025-05-20 13:58:10,169 INFO mapred.LocalJobRunner: 1 / 1 copied.
2025-05-20 13:58:10,169 INFO mapred.Task: Task attempt_local408680812_0001_r_000000_0 is allowed to commit now
2025-05-20 13:58:10,194 INFO output.FileOutputCommitter: Saved output of task 'attempt_local408680812_0001_r_000000_0' to hdfs://localhost:9000/res
2025-05-20 13:58:10,195 INFO mapred.LocalJobRunner: reduce > reduce
2025-05-20 13:58:10,196 INFO mapred.Task: Task 'attempt_local408680812_0001_r_000000_0' done.
2025-05-20 13:58:10,197 INFO mapred.Task: Final Counters for attempt_local408680812_0001_r_000000_0: Counters: 30
    File System Counters
        FILE: Number of bytes read=7887
        FILE: Number of bytes written=645606
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=75
        HDFS: Number of bytes written=105
        HDFS: Number of read operations=10
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=3
        HDFS: Number of bytes read erasure-coded=0
    Map-Reduce Framework
        Combine input records=0
        Combine output records=0
        Reduce input groups=15
        Reduce shuffle bytes=171
        Reduce input records=15
        Reduce output records=15
        Spilled Records=15
        Shuffled Maps =1
        Failed Shuffles=0
        Merged Map outputs=1
        GC time elapsed (ms)=0
        Total committed heap usage (bytes)=526385152
    Shuffle Errors
        BAD_ID=0
        CONNECTION=0
        IO_ERROR=0
        WRONG_LENGTH=0
        WRONG_MAP=0
        WRONG_REDUCE=0
    File Output Format Counters
        Bytes Written=105
2025-05-20 13:58:10,197 INFO mapred.LocalJobRunner: Finishing task: attempt_local408680812_0001_r_000000_0
2025-05-20 13:58:10,197 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-05-20 13:58:10,840 INFO mapreduce.Job: Job job_local408680812_0001 running in uber mode : false
2025-05-20 13:58:10,842 INFO mapreduce.Job: map 100% reduce 100%
2025-05-20 13:58:10,843 INFO mapreduce.Job: Job job_local408680812_0001 completed successfully
```

```
hadoop@bmscse-HP-Elite-Tower-800-G9-Desktop-PC: ~
2025-05-20 13:58:10,168 INFO mapred.Task: Task:attempt_local408680812_0001_r_000000_0 is done. And is in the process of committing.
2025-05-20 13:58:10,169 INFO mapred.LocalJobRunner: 1 / 1 copied.
2025-05-20 13:58:10,169 INFO mapred.Task: Task attempt_local408680812_0001_r_000000_0 is allowed to commit now
2025-05-20 13:58:10,194 INFO output.FileOutputCommitter: Saved output of task 'attempt_local408680812_0001_r_000000_0' to hdfs://localhost:9000/res
2025-05-20 13:58:10,195 INFO mapred.LocalJobRunner: reduce > reduce
2025-05-20 13:58:10,196 INFO mapred.Task: Task 'attempt_local408680812_0001_r_000000_0' done.
2025-05-20 13:58:10,197 INFO mapred.Task: Final Counters for attempt_local408680812_0001_r_000000_0: Counters: 30
    File System Counters
        FILE: Number of bytes read=7887
        FILE: Number of bytes written=645606
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=75
        HDFS: Number of bytes written=105
        HDFS: Number of read operations=10
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=3
        HDFS: Number of bytes read erasure-coded=0
    Map-Reduce Framework
        Combine input records=0
        Combine output records=0
        Reduce input groups=15
        Reduce shuffle bytes=171
        Reduce input records=15
        Reduce output records=15
        Spilled Records=15
        Shuffled Maps =1
        Failed Shuffles=0
        Merged Map outputs=1
        GC time elapsed (ms)=0
        Total committed heap usage (bytes)=526385152
    Shuffle Errors
        BAD_ID=0
        CONNECTION=0
        IO_ERROR=0
        WRONG_LENGTH=0
        WRONG_MAP=0
        WRONG_REDUCE=0
    File Output Format Counters
        Bytes Written=105
2025-05-20 13:58:10,197 INFO mapred.LocalJobRunner: Finishing task: attempt_local408680812_0001_r_000000_0
2025-05-20 13:58:10,197 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-05-20 13:58:10,840 INFO mapreduce.Job: Job job_local408680812_0001 running in uber mode : false
2025-05-20 13:58:10,842 INFO mapreduce.Job: map 100% reduce 100%
2025-05-20 13:58:10,843 INFO mapreduce.Job: Job job_local408680812_0001 completed successfully
```

```
hadoop@bmscse-HP-Elite-Tower-800-G9-Desktop-PC: ~
Map input records=2
Map output records=15
Map output bytes=135
Map output materialized bytes=171
Input split bytes=101
Combine input records=0
Spilled Records=15
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=0
Total committed heap usage (bytes)=526385152
File Input Format Counters
Bytes Read=75
2025-05-20 13:58:10,020 INFO mapred.LocalJobRunner: Finishing task: attempt_local408680812_0001_m_000000_0
2025-05-20 13:58:10,020 INFO mapred.LocalJobRunner: map task executor complete.
2025-05-20 13:58:10,021 INFO mapred.LocalJobRunner: Waiting for reduce tasks
2025-05-20 13:58:10,022 INFO mapred.LocalJobRunner: Starting task: attempt_local408680812_0001_r_000000_0
2025-05-20 13:58:10,027 INFO output.PathOutputCommitterFactory: No output committer factory defined, defaulting to FileOutputCommitterFactory
2025-05-20 13:58:10,027 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-05-20 13:58:10,027 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-05-20 13:58:10,027 INFO mapred.Task: Using ResourceCalculatorProcessTree : []
2025-05-20 13:58:10,028 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@6622090a
2025-05-20 13:58:10,029 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2025-05-20 13:58:10,037 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=5829453312, maxSingleShuffleLimit=1457363328, mergeThreshold=3847439360, ioSortFactor=10, memToMemMergeOutputsThreshold=10
2025-05-20 13:58:10,038 INFO reduce.EventFetcher: attempt_local408680812_0001_r_000000_0 Thread started: EventFetcher for fetching Map Completion Events
2025-05-20 13:58:10,053 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map attempt_local408680812_0001_m_000000_0 decomp: 167 len: 171 to MEMORY
2025-05-20 13:58:10,054 INFO reduce.InMemoryMapOutput: Read 167 bytes from map-output for attempt_local408680812_0001_m_000000_0
2025-05-20 13:58:10,055 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of size: 167, inMemoryMapOutputs.size() -> 1, commitMemory -> 0, usedMemory ->167
2025-05-20 13:58:10,056 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning
2025-05-20 13:58:10,056 INFO mapred.LocalJobRunner: 1 / 1 copied.
2025-05-20 13:58:10,056 INFO reduce.MergeManagerImpl: finalMerge called with 1 in-memory map-outputs and 0 on-disk map-outputs
2025-05-20 13:58:10,059 INFO mapred.Merger: Merging 1 sorted segments
2025-05-20 13:58:10,059 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 162 bytes
2025-05-20 13:58:10,060 INFO reduce.MergeManagerImpl: Merged 1 segments, 167 bytes to disk to satisfy reduce memory limit
2025-05-20 13:58:10,060 INFO reduce.MergeManagerImpl: Merging 1 files, 171 bytes from disk
2025-05-20 13:58:10,060 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce
2025-05-20 13:58:10,060 INFO mapred.Merger: Merging 1 sorted segments
2025-05-20 13:58:10,061 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 162 bytes
2025-05-20 13:58:10,061 INFO mapred.LocalJobRunner: 1 / 1 copied.
```

```
hadoop@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC: ~
```

```
2025-05-20 13:58:10,840 INFO mapreduce.Job: Job job_local408680812_0001 running in uber mode : false
2025-05-20 13:58:10,842 INFO mapreduce.Job: map 100% reduce 100%
2025-05-20 13:58:10,843 INFO mapreduce.Job: Job job_local408680812_0001 completed successfully
2025-05-20 13:58:10,854 INFO mapreduce.Job: Counters: 36
    File System Counters
        FILE: Number of bytes read=15400
        FILE: Number of bytes written=1291041
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=150
        HDFS: Number of bytes written=105
        HDFS: Number of read operations=15
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=4
        HDFS: Number of bytes read erasure-coded=0
    Map-Reduce Framework
        Map input records=2
        Map output records=15
        Map output bytes=135
        Map output materialized bytes=171
        Input split bytes=101
        Combine input records=0
        Combine output records=0
        Reduce input groups=15
        Reduce shuffle bytes=171
        Reduce input records=15
        Reduce output records=15
        Spilled Records=30
        Shuffled Maps =1
        Failed Shuffles=0
        Merged Map outputs=1
        GC time elapsed (ms)=0
        Total committed heap usage (bytes)=1052770304
    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=75
    File Output Format Counters
        Bytes Written=105
hadoop@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -cat /res/part-00000
cat: `/res/part-00000': No such file or directory
hadoop@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -ls /res
```

```
hadoop@bmscse-HP-Elite-Tower-800-G9-Desktop-PC: ~
Input split bytes=101
Combine input records=0
Combine output records=0
Reduce input groups=15
Reduce shuffle bytes=171
Reduce input records=15
Reduce output records=15
Spilled Records=30
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=0
Total committed heap usage (bytes)=1052770304
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=75
File Output Format Counters
Bytes Written=105
hadoop@bmscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -cat /res/part-00000
cat: '/res/part-00000': No such file or directory
hadoop@bmscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -ls /res
Found 2 items
-rw-r--r-- 1 hadoop supergroup 0 2025-05-20 13:58 /res/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 105 2025-05-20 13:58 /res/part-r-00000
hadoop@bmscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ ^C
hadoop@bmscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -cat /res/part-r-00000
college 1
in 1
bms 1
hi 1
i 1
inna 1
am 1
m 1
bhuvana 1
how 1
are 1
avyukth 1
of 1
you 1
engineeering 1
hadoop@bmscse-HP-Elite-Tower-800-G9-Desktop-PC:~$
```

Lab 2

~~MinMax program map reduce~~

MinMax Mapper

```
public class MinMaxMapper extends Mapper<Object, Text, Text, DoubleWritable> {
    private DoubleWritable value02d = new DoubleWritable();
    private Double localMin = Double.MAX_VALUE;
    private double localMax = Double.MIN_VALUE;
    contact() throws IOException, InterruptedException {
        double val = Double.parseDouble(value,
                                         toString());
        localMin = Math.min(localMin, val);
        localMax = Math.max(localMax, val);
    }
    protected void cleanup(Context context) throws
    IOException, InterruptedException {
        context.write(new Text("min"), new DoubleWritable(
            localMin));
        context.write(new Text("max"), new DoubleWritable(
            localMax));
    }
}
```

3

}

Min Max Reducer

```
public class MinMaxReducer extends Reducer<Text,  
    DoubleWritable> {  
    protected void reduce(Text key,  
        Iterable<DoubleWritable> values,  
        Context context) throws IOException,  
        InterruptedException {  
        if (key.toString().equals("min")) {  
            double min = Double.MAX_VALUE;  
            for (DoubleWritable val : values) {  
                min = Math.min(min, val.get());  
            }  
            context.write(new Text("global_min"),  
                new DoubleWritable(min));  
        } else if (key.toString().equals("max")) {  
            double max = Double.MIN_VALUE;  
            for (DoubleWritable val : values) {  
                max = Math.max(max, val.get());  
            }  
            context.write(new Text("global_max"),  
                new DoubleWritable(max));  
        }  
    }  
}
```

bash

hadoop com.sun.tools.javac.Main MainDriver.java
minmaxMapper.java
minmaxReducer.java
minmaxDriver.java
jar of minmaxjar minmax.class
hadoop jar minmax.jar /input /outputminmax
minmax priva /input /data.txt /outputminmax
minmax driver class

Sample Input File: min

10
3
55
36
23
8
91
17
global min = 3
global max = 91

Output

Mapper 1 exits!
min 3
max 55
Mapper 2 exits!
min 8
max 91;

Reducer receives!

min 3
min 8
max 55
max 91

$$\text{global min} = \min(3, 8) = 3$$
$$\text{global max} = \max(55, 91) = 91$$

Lab 8: Scala

Question: Write a Scala program to print numbers from 1 to 100 using for loop.

Code with Output:

```
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ nano pi.scala
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ scalac pi.scala
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ scala pi
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 5
7 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83
84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
```

The screenshot shows a terminal window titled "bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~". The window contains the following Scala code:

```
GNU nano 6.2                                     pi.scala
object pi {
  def main(args: Array[String]): Unit = {
    for(counter <- 1 to 100)
      print(counter + " ")
    println()
  }
}
```

At the bottom of the terminal window, there is a menu bar with the following options: [Read 7 lines], ^G Help, ^O Write Out, ^W Where Is, ^K Cut, ^T Execute, ^C Location, ^X Exit, ^R Read File, ^\ Replace, ^U Paste, ^J Justify, ^/ Go To Line.

Lab 9: Spark

Question: Using RDD and FlatMap 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.

Code with Output:

```
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ spark-shell
25/05/20 15:32:38 WARN Utils: Your hostname, bmscecse-HP-Elite-Tower-800-G9-Desktop-PC resolves to a loopback address: 127.0.1.1
: using 10.124.2.8 instead (on interface eno1)
25/05/20 15:32:38 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.spark.unsafe.Platform (file:/opt/spark/jars/spark-unsafe_2.12-3.0.3.jar) to constructor java.nio.DirectByteBuffer(long,int)
WARNING: Please consider reporting this to the maintainers of org.apache.spark.unsafe.Platform
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
25/05/20 15:32:38 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Spark context Web UI available at http://10.124.2.8:4040
Spark context available as 'sc' (master = local[*], app id = local-1747735361481).
Spark session available as 'spark'.
Welcome to

    / \   / \   / \   / \
   / \ / - \ / - \ / \ / \
  / \ / . \ / , / \ / \ / \
 / \ / . \ / , / \ / \ / \
version 3.0.3

Using Scala version 2.12.10 (OpenJDK 64-Bit Server VM, Java 11.0.26)
Type in expressions to have them evaluated.
Type :help for more information.

scala> val textFile = sc.textFile("/home/bmscecse/Desktop/sparkdata.txt")
textFile: org.apache.spark.rdd.RDD[String] = /home/bmscecse/Desktop/sparkdata.txt MapPartitionsRDD[1] at textFile at <console>:2
4

scala>

scala> val counts = textFile
counts: org.apache.spark.rdd.RDD[String] = /home/bmscecse/Desktop/sparkdata.txt MapPartitionsRDD[1] at textFile at <console>:24

scala> .flatMap(line => line.split(" "))
res0: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[2] at flatMap at <console>:26

scala> .map(word => (word, 1))

scala> val data = sc.textFile("sparkdata.txt")
data: org.apache.spark.rdd.RDD[String] = sparkdata.txt MapPartitionsRDD[1] at textFile at <console>:25

scala> val splitdata = data.flatMap(line => line.split(" "))
splitdata: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[2] at flatMap at <console>:26

scala> val mapdata = splitdata.map(word => (word, 1))
mapdata: org.apache.spark.rdd.RDD[(String, Int)] = MapPartitionsRDD[3] at map at <console>:26

scala> val reducedata = mapdata.reduceByKey(_ + _)
reducedata: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[4] at reduceByKey at <console>:26

scala> reducedata.collect.foreach(println)
(,1)
(hello,2)
(world,1)
(spark,1)
```

```
scala> val textFile = sc.textFile("/home/bmscecse/Desktop/WC.txt")
textFile: org.apache.spark.rdd.RDD[String] = /home/bmscecse/Desktop/WC.txt MapPartitionsRDD[31] at textFile at <console>:31

scala> val words = textFile.flatMap(line => line.split(" "))
words: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[32] at flatMap at <console>:32

scala>

scala> val pairs = words.map(word => (word, 1))
pairs: org.apache.spark.rdd.RDD[(String, Int)] = MapPartitionsRDD[33] at map at <console>:32

scala>

scala> val counts = pairs.reduceByKey(_ + _)
counts: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[34] at reduceByKey at <console>:32

scala> val countsArray = counts.collect() // This is Array[(String, Int)]
countsArray: Array[(String, Int)] = Array(( "", 1), (hello, 6), (world, 1), (spark, 1))

scala> val sorted = ListMap(countsArray.sortWith(_.value > _.value): _*)
sorted: scala.collection.immutable.ListMap[String,Int] = ListMap(hello -> 6, "" -> 1, world -> 1, spark -> 1)

scala> for ((k, v) <- sorted) {
    |   if (v > 4) println(s"$k, $v")
    | }
hello, 6
scala> ■
```

Lab VII

1) write a scala print no's 1 to 100

```
for (i <- 1 to 100) {  
    print(i)}
```

y

out

1

2

3

.

:

98

99

100

[closed]

"...236" on 3

"...236" on 3

two stdans

2) >cd Desktop

> nano file.txt

> realpath ~/Desktop/file.txt

/home/bmscsecs/Desktop/file.txt

> val rdd = sc.textFile("/home/bmscsecs/Desktop/file.txt")

val words = rdd.flatMap(line =>

split(" \s+"))

val pairs = words.map(word => (word, 1))

val counts = pairs.reduceByKey(_ + _)

filter { case (word, count) => count > 1}

counts. collect (1. for each & case (word, count))
prints ("word & count")

3

(3) Add slot repository keys & repos.

bash

→ curl -sL "https://keyserver.ubuntu.com/"

echo "deb..."

echo "deb..."

→ update apt

→ sudo apt install sbt

→ slot sbt version

→ make directory

→ mkdir streaming Textcleaner

cd streaming Textcleaner

Sbt new scalars/scalar-seed.g8

→ Add spark (luence) dependencies

→ know build.sbt

→ dependencies added to streaming

→ dependencies added to Textcleaner