

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



**LAB REPORT**  
on  
**Big Data Analytics(23CS6PCBDA)**

*Submitted by*

**Madhu Sarika (1BM22CS140)**

*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-2024 to July-2024**

**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** carried out by **Madhu Sarika (1BM22CS1)**, who is 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 **Course Big Data Analytics(23CS6PCBDA)** work prescribed for the said degree.

Leelavaathi. B  
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- CRUD Operations Demonstration (Practice and Self Study)	4
2	Neo4j	22
3	<p>Perform the following DB operations using Cassandra.</p> <p>a) Create a keyspace by name Employee</p> <p>b) Create a column family by name Employee-Info with attributes</p> <p>Emp_Id Primary Key, Emp_Name, Designation, Date_of_Joining, Salary, Dept_Name</p> <p>c) Insert the values into the table in batch</p> <p>d) Update Employee name and Department of Emp-Id 121</p> <p>e) Sort the details of Employee records based on salary</p> <p>f) Alter the schema of the table Employee_Info to add a column Projects which stores a set of Projects done by the corresponding Employee.</p> <p>g) Update the altered table to add project names.</p> <p>h) Create a TTL of 15 seconds to display the values of Employees.</p>	27
4	<p>Perform the following DB operations using Cassandra.</p> <p>a) Create a keyspace by name Library</p> <p>b) Create a column family by name Library-Info with attributes</p> <p>Stud_Id Primary Key, Counter_value of type Counter, Stud_Name, Book-Name, Book-Id, Date_of_issue</p> <p>c) Insert the values into the table in batch</p> <p>d) Display the details of the table created and increase the value of the counter</p> <p>e) Write a query to show that a student with id 112 has taken a book “BDA” 2 times.</p> <p>f) Export the created column to a csv file</p>	30
	g) Import a given csv dataset from local file system into Cassandra column family	

5	Execution of HDFS Commands for interaction with Hadoop Environment. (Minimum 10 commands to be executed)	37
6	Implement Wordcount program on Hadoop framework	41
7	From the following link extract the weather data <a href="https://github.com/tomwhite/hadoop-book/tree/master/input/ncdc/all">https://github.com/tomwhite/hadoop-book/tree/master/input/ncdc/all</a> a) Create a MapReduce program to find average temperature for each year from NCDC data set. b) find the mean max temperature for every month.	48
8	Write a Scala program to print numbers from 1 to 100 using for loop.	55
9	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.	57
10	Write a simple streaming program in Spark to receive text data streams on a particular port, perform basic text cleaning (like white space removal, stop words removal, lemmatization, etc.), and print the cleaned text on the screen. (Open Ended Question).	58

### Course Outcome

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

## Experiment - 1

# MongoDB- CRUD Operations Demonstration

```
Activities Terminal Mar 11 14:30
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
Current Mongosh Log ID: 67cff6b5879eb5f3fc0e7d
Connecting To: mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/?appName=mongosh+2.2.1
Using Mongosh: 0.0.5 (API Version 1)
Using MongoDB: 2.2.1
mongosh 2.4.2 is available for download: https://www.mongodb.com/try/download/shell

For mongosh Info see: https://docs.mongodb.com/mongosh-shell/

Atlas atlas-vvj19l-shard-0 [primary] test>
Atlas atlas-vvj19l-shard-0 [primary] test> use myDB
switched to db myDB
Atlas atlas-vvj19l-shard-0 [primary] myDB> create.collection("Student1");
referencedError: create is not defined
Atlas atlas-vvj19l-shard-0 [primary] myDB> db.createCollection("Student1");
Uncaught:
SyntaxError: Unexpected token, expected ',' (1:29)
> 1 | db.createCollection("Student1");
> 2 |
Atlas atlas-vvj19l-shard-0 [primary] myDB> db.createCollection("Student1");
ok: 1
Atlas atlas-vvj19l-shard-0 [primary] myDB> db.Student1.insert({
...   _id: 1,
...   Studname: "MichelleJacintha",
...   Grade: "VII",
...   Hobbies: "Internetsurfing"
... });
DeprecationWarning: Collection.insert() is deprecated. Use insertOne, insertMany, or bulkWrite.
{
  acknowledged: true,
  insertedId: {
    '$': 1
  }
}
Atlas atlas-vvj19l-shard-0 [primary] myDB> db.Student1.update(
...   { _id: 1 },
...   { Studname: "Aryabhoud", Grade: "VII" },
...   { $set: { Hobbies: "Skating" } },
...   { upsert: true }
... );
DeprecationWarning: Collection.update() is deprecated. Use updateOne, updateMany, or bulkWrite.
{
  acknowledged: true,
  insertedId: 1,
  matchedCount: 0,
  modifiedCount: 0,
  upsertedCount: 0
}
Atlas atlas-vvj19l-shard-0 [primary] myDB> db.Student1.find();
[{
  _id: 1,
  Studname: "MichelleJacintha",
  Grade: "VII",
  Hobbies: "Internetsurfing"
}, {
  _id: 1,
  Grade: "VII",
  Studname: "Aryabhoud",
  Hobbies: "Skating"
} ]
```

```

Activities Terminal Mar 11 14:31
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
[{"_id": 3, "Grade": "VII", "StudName": "AryanDavid", "Hobbies": "Skating"}]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.update(
...   {"_id": 3, "StudName": "AryanDavid"}, {"$set": {"Hobbies": "Chess"}}
... );
{
  acknowledged: 1,
  insertedId: null,
  lastErrorCount: 1,
  modifiedCount: 1,
  upsertedCount: 0
}
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find();
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find();
[
  {
    "_id": 1,
    "StudName": "MichelleJacintha",
    "Grade": "VII",
    "Hobbies": "Internetsurfing"
  },
  {
    "_id": 3, "Grade": "VII", "StudName": "AryanDavid", "Hobbies": "Chess" }
]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find({StudName: "Aryan David"});
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find({StudName: "AryanDavid"});
[{"_id": 3, "Grade": "VII", "StudName": "AryanDavid", "Hobbies": "Chess"}]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find({}, {"StudName": 1, "Grade": 1, "_id": 0});
[
  {"StudName": "MichelleJacintha", "Grade": "VII", "Hobbies": "Aryandavid"}
]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find({Grade: "VII", StudName: "Aryandavid"});
{
  {
    "_id": 3,
    "StudName": "MichelleJacintha",
    "Grade": "VII",
    "Hobbies": "Internetsurfing"
  },
  {
    "_id": 3, "Grade": "VII", "StudName": "Aryandavid", "Hobbies": "Chess" }
}
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find({Grade: {"$eq": "VII"}}).pretty();
[
  {
    "_id": 1,
    "StudName": "MichelleJacintha",
    "Grade": "VII",
    "Hobbies": "Internetsurfing"
  },
  {
    "_id": 3, "Grade": "VII", "StudName": "Aryandavid", "Hobbies": "Chess" }
]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find({Hobbies: {$in: ["Chess", "Skating"]}}).pretty();
[{"_id": 3, "Grade": "VII", "StudName": "Aryandavid", "Hobbies": "Chess"}]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find({StudName: /A/}).pretty();
[
  {
    "_id": 1,
    "StudName": "MichelleJacintha",
    "Grade": "VII",
    "Hobbies": "Internetsurfing"
  }
]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find({StudName: /A/}).pretty();
[
  {
    "_id": 1,
    "StudName": "MichelleJacintha",
    "Grade": "VII",
    "Hobbies": "Internetsurfing"
  },
  {
    "_id": 3, "Grade": "VII", "StudName": "Aryandavid", "Hobbies": "Chess" }
]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.count();
DeprecationWarning: collection.count() is deprecated. Use countDocuments or estimatedDocumentCount.
3
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find().sort({StudName: -1}).pretty();
[
  {
    "_id": 1,
    "StudName": "MichelleJacintha",
    "Grade": "VII",
    "Hobbies": "Internetsurfing"
  },
  {
    "_id": 3, "Grade": "VII", "StudName": "Aryandavid", "Hobbies": "Chess" }
]

```

```

Activities Terminal Mar 11 14:31
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
[{"_id": 3, "Grade": "VII", "StudName": "AryanDavid", "Hobbies": "Skating"}]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.update(
...   {"_id": 3, "StudName": "AryanDavid"}, {"$set": {"Hobbies": "Chess"}}
... );
{
  acknowledged: 1,
  insertedId: null,
  lastErrorCount: 1,
  modifiedCount: 1,
  upsertedCount: 0
}
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find();
ReferenceError: ds is not defined
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find();
[{"_id": 1, "StudName": "MichelleJacintha", "Grade": "VII", "Hobbies": "Internetsurfing"}, {"_id": 3, "Grade": "VII", "StudName": "AryanDavid", "Hobbies": "Chess"}]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find({StudName: "Aryan David"});
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find({StudName: "AryanDavid"});
[{"_id": 3, "Grade": "VII", "StudName": "AryanDavid", "Hobbies": "Chess"}]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find({Grade: "VII", StudName: "AryanDavid"});
[{"_id": 3, "Grade": "VII", "StudName": "AryanDavid"}]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find({Grade: {"$eq": "VII"}});
[{"_id": 1, "StudName": "MichelleJacintha", "Grade": "VII", "Hobbies": "Internetsurfing"}, {"_id": 3, "Grade": "VII", "StudName": "AryanDavid"}]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find({Grade: {"$eq": "VII"}}).pretty();
[{"_id": 1, "StudName": "MichelleJacintha", "Grade": "VII", "Hobbies": "Internetsurfing"}, {"_id": 3, "Grade": "VII", "StudName": "AryanDavid", "Hobbies": "Chess"}]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find({Hobbies: {"$in": ["Chess", "Skating"]}}).pretty();
[{"_id": 3, "Grade": "VII", "StudName": "AryanDavid", "Hobbies": "Chess"}]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find({StudName: "/n"}).pretty();
[{"_id": 1, "StudName": "MichelleJacintha", "Grade": "VII", "Hobbies": "Internetsurfing"}, {"_id": 3, "Grade": "VII", "StudName": "AryanDavid", "Hobbies": "Chess"}]

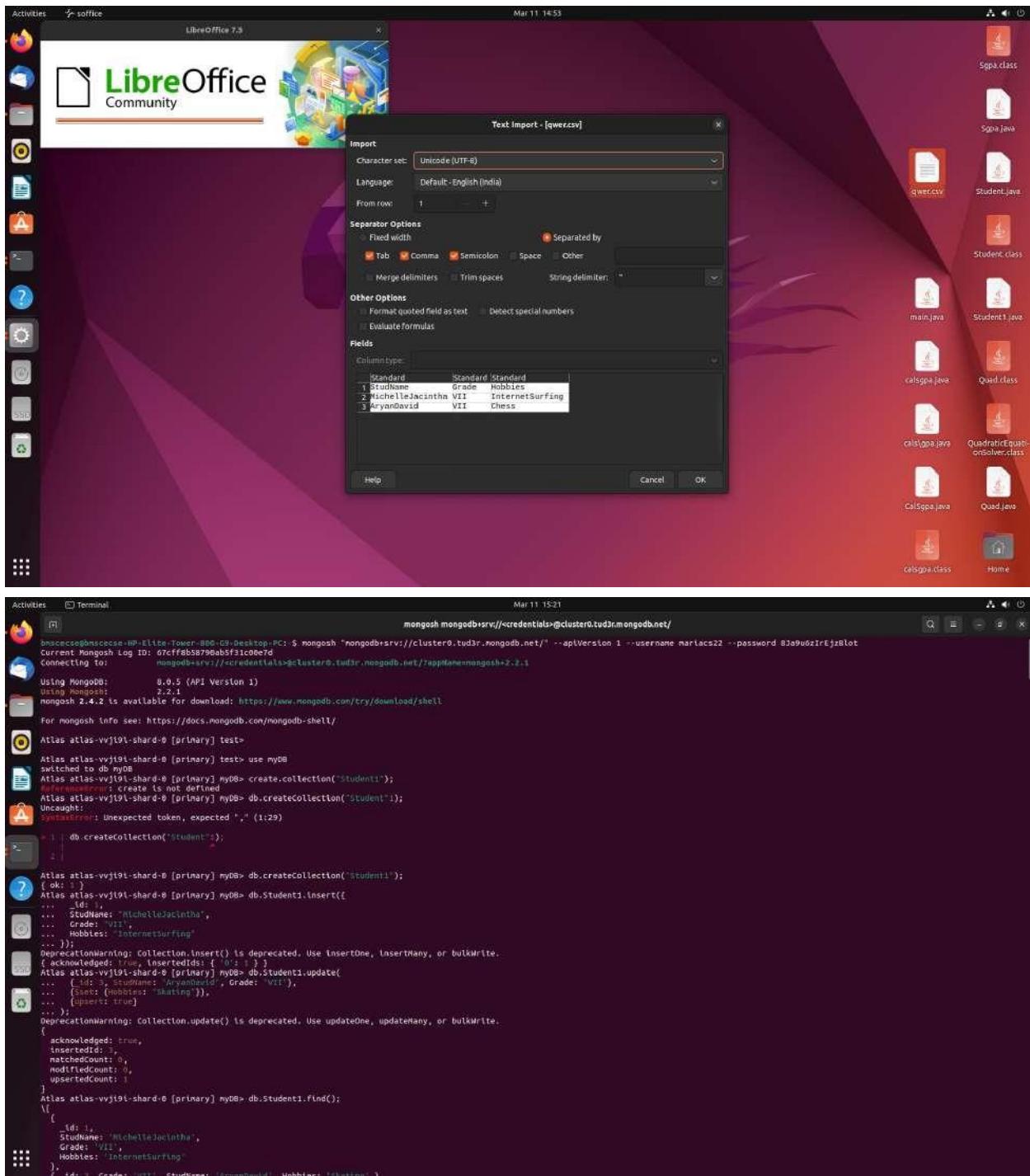
Activities Terminal Mar 11 14:32
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
[{"_id": 3, "Grade": "VII", "StudName": "AryanDavid", "Hobbies": "Chess"}]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find({}, {StudName: 1, Grade: 1, _id: 0});
[{"StudName": "MichelleJacintha", "Grade": "VII"}, {"Grade": "VII", "StudName": "AryanDavid"}]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find({}, {StudName: 1, Grade: 1, _id: 0});
[{"StudName": "MichelleJacintha", "Grade": "VII"}, {"Grade": "VII", "StudName": "AryanDavid"}]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find({Grade: {"$eq": "VII"}});
[{"_id": 1, "StudName": "MichelleJacintha", "Grade": "VII", "Hobbies": "Internetsurfing"}, {"_id": 3, "Grade": "VII", "StudName": "AryanDavid"}]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find({Grade: {"$eq": "VII"}}).pretty();
[{"_id": 1, "StudName": "MichelleJacintha", "Grade": "VII", "Hobbies": "Internetsurfing"}, {"_id": 3, "Grade": "VII", "StudName": "AryanDavid", "Hobbies": "Chess"}]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find({Hobbies: {"$in": ["Chess", "Skating"]}}).pretty();
[{"_id": 3, "Grade": "VII", "StudName": "AryanDavid", "Hobbies": "Chess"}]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find({StudName: "/n"}).pretty();
[{"_id": 1, "StudName": "MichelleJacintha", "Grade": "VII", "Hobbies": "Internetsurfing"}, {"_id": 3, "Grade": "VII", "StudName": "AryanDavid", "Hobbies": "Chess"}]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.count();
DeprecationWarning: collection.count() is deprecated. Use countDocuments or estimatedDocumentCount.
[{"_id": 1, "StudName": "MichelleJacintha", "Grade": "VII", "Hobbies": "Internetsurfing"}, {"_id": 3, "Grade": "VII", "StudName": "AryanDavid", "Hobbies": "Chess"}]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find().sort({StudName: -1}).pretty();
[{"_id": 1, "StudName": "MichelleJacintha", "Grade": "VII", "Hobbies": "Internetsurfing"}, {"_id": 3, "Grade": "VII", "StudName": "AryanDavid", "Hobbies": "Chess"}]

```

```

Activities Terminal Mar 11 14:52
bmseccse@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~
mongodump --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudentName,Grade,Hobbies"
2025-03-11T14:39:39.058+0530 connected to: mongodb+srv://["**REDACTED**"]@cluster0.tud3r.mongodb.net/myDB
2025-03-11T14:39:39.058+0530 exporting 2 records
bmseccse@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~
mongodump --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudentName,Grade,Hobbies"
2025-03-11T14:40:21.068+0530 connected to: mongodb+srv://["**REDACTED**"]@cluster0.tud3r.mongodb.net/myDB
2025-03-11T14:40:21.068+0530 exporting 2 records
bmseccse@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~
mongodump --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudentName,Grade,Hobbies"
2025-03-11T14:40:29.815+0530 connected to: mongodb+srv://["**REDACTED**"]@cluster0.tud3r.mongodb.net/myDB
2025-03-11T14:40:29.815+0530 exporting 2 records
bmseccse@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~
mongodump --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudentName,Grade,Hobbies"
2025-03-11T14:41:21.068+0530 connected to: mongodb+srv://["**REDACTED**"]@cluster0.tud3r.mongodb.net/myDB
2025-03-11T14:41:21.068+0530 exporting 2 records
bmseccse@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~
mongodump --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudentName,Grade,Hobbies"
2025-03-11T14:41:44.222+0530 connected to: mongodb+srv://["**REDACTED**"]@cluster0.tud3r.mongodb.net/myDB
2025-03-11T14:41:44.222+0530 exporting 2 records
bmseccse@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~
whoami
bmseccse@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~
mongorestore --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --file "/home/bmseccse/Desktop/qwer.csv" --headerline
2025-03-11T14:51:32.165+0530 2 document(s) imported successfully. 0 document(s) failed to import.
bmseccse@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~
bmseccse@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~

```



```

Activities Terminal Mar 11 15:22
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
[{"_id": 1, "Grade": "VII", "StudName": "Aryandavid", "Hobbies": "Skating"}]
Atlas atlas-vvji9t-shard-0 [primary] my0B> db.Student1.update(
...   {"_id": 1, "StudName": "Aryandavid", "Grade": "VII"},
...   {"$set": {"Hobbies": "Chess"}}
... );
{
  "acknowledged": true,
  "insertedCount": 0,
  "matchedCount": 1,
  "modifiedCount": 1,
  "upsertedCount": 0
}
Atlas atlas-vvji9t-shard-0 [primary] my0B> ds.Student1.find();
ReferenceError: ds is not defined
Atlas atlas-vvji9t-shard-0 [primary] my0B> db.Student1.find();
[
  {
    "_id": 1,
    "StudName": "MichelleJacintha",
    "Grade": "VII",
    "Hobbies": "Internetsurfing"
  },
  {
    "_id": 2,
    "StudName": "Aryandavid",
    "Grade": "VII",
    "Hobbies": "Chess"
  }
]
Atlas atlas-vvji9t-shard-0 [primary] my0B> db.Student1.find({StudName: "Aryan David"});
Atlas atlas-vvji9t-shard-0 [primary] my0B> db.Student1.find({StudName: "Aryandavid"});
[
  {
    "StudName": "MichelleJacintha", "Grade": "VII" },
  {
    "Grade": "VII", "StudName": "Aryandavid" }
]
Atlas atlas-vvji9t-shard-0 [primary] my0B> db.Student1.find({}, {StudName: 1, GrGrade: 1, _id: 0});
[
  [
    {
      "StudName": "MichelleJacintha", "Grade": "VII" },
      {
        "Grade": "VII", "StudName": "Aryandavid" }
    ]
]
Atlas atlas-vvji9t-shard-0 [primary] my0B> db.Student1.find({Grade: {$eq: "VII"}}).pretty();
[
  {
    "_id": 1,
    "StudName": "MichelleJacintha",
    "Grade": "VII",
    "Hobbies": "Internetsurfing"
  },
  {
    "_id": 2,
    "StudName": "Aryandavid",
    "Grade": "VII",
    "Hobbies": "Chess"
  }
]
Atlas atlas-vvji9t-shard-0 [primary] my0B> db.Student1.find({Hobbies: {$in: ["Skating", "Chess"]}}).pretty();
[{"_id": 1, "Grade": "VII", "StudName": "Aryandavid", "Hobbies": "Chess"}]
Atlas atlas-vvji9t-shard-0 [primary] my0B> db.Student1.find({StudName: /A/}).pretty();
[{"_id": 1, "Grade": "VII", "StudName": "Aryandavid", "Hobbies": "Chess"}]

Activities Terminal Mar 11 15:22
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
[{"_id": 1, "Grade": "VII", "StudName": "MichelleJacintha", "Hobbies": "Internetsurfing"}]
Atlas atlas-vvji9t-shard-0 [primary] my0B> db.Student1.find({studName: /M/}).pretty();
[
  {
    "_id": 1,
    "StudName": "MichelleJacintha",
    "Grade": "VII",
    "Hobbies": "Internetsurfing"
  }
]
Atlas atlas-vvji9t-shard-0 [primary] my0B> db.Student1.find({StudName: /M/}).pretty();
[{"_id": 1, "Grade": "VII", "StudName": "MichelleJacintha", "Hobbies": "Internetsurfing"}]
Atlas atlas-vvji9t-shard-0 [primary] my0B> db.Student1.find({studName: /A/}).pretty();
[{"_id": 1, "Grade": "VII", "StudName": "Aryandavid", "Hobbies": "Chess"}]
Atlas atlas-vvji9t-shard-0 [primary] my0B> db.Student1.count();
DeprecationWarning: Collection.count() is deprecated. Use countDocuments or estimatedDocumentCount.
2
Atlas atlas-vvji9t-shard-0 [primary] my0B> db.Student1.find().sort({$studName: -1}).pretty();
[
  {
    "_id": 1,
    "StudName": "MichelleJacintha",
    "Grade": "VII",
    "Hobbies": "Internetsurfing"
  },
  {
    "_id": 2,
    "Grade": "VII",
    "StudName": "Aryandavid",
    "Hobbies": "Chess"
  }
]
Atlas atlas-vvji9t-shard-0 [primary] my0B> db.Student1.find();
(To exit, press Ctrl+C again or Ctrl+D or type .exit)
Atlas atlas-vvji9t-shard-0 [primary] my0B> db.Student1.find().pretty();
[
  {
    "_id": 1,
    "StudName": "MichelleJacintha",
    "Grade": "VII",
    "Hobbies": "Internetsurfing"
  },
  {
    "_id": 2,
    "Grade": "VII",
    "StudName": "Aryandavid",
    "Hobbies": "Chess"
  }
]
Atlas atlas-vvji9t-shard-0 [primary] my0B> db.Student3.find();
{
  {
    "id": ObjectId("07d0009c20433200f8c4ffcc"),
    "StudName": "Aryandavid",
    "Grade": "VII",
    "Hobbies": "Chess"
  },
  {
    "id": ObjectId("07d0009c20433200f8c4ffcd"),
    "StudName": "MichelleJacintha",
    "Grade": "VII",
    "Hobbies": "Internetsurfing"
  }
]

```

```

Activities Terminal Mar 11 15:21
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
Current Mongosh Log ID: 67cff8b58790ab5f3c00e7d
Connecting to: mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/?appName=mongosh+2.2.1
Using MongoDB: 0.0.5 (API Version 1)
Using Mongosh: 2.2.1
mongosh 2.4.2 is available for download: https://www.mongodb.com/try/download/shell
For mongosh info see: https://docs.mongodb.com/mongodb-shell/
Atlas atlas-vvji9t-shard-0 [primary] test>
Atlas atlas-vvji9t-shard-0 [primary] test> use myDB
switched to db myDB
Atlas atlas-vvji9t-shard-0 [primary] myDB> create.collection("student");
Atlas atlas-vvji9t-shard-0 [primary] myDB> createcollection("student");
Uncaught:
SyntaxError: Unexpected token, expected "," (1:29)
    + 1 | db.createCollection("student");
    + 2 |
    + 3
Atlas atlas-vvji9t-shard-0 [primary] myDB> db.createCollection("student");
{ ok: 1 }
Atlas atlas-vvji9t-shard-0 [primary] myDB> db.Student1.insert([
...   {
...     _id: 1,
...     StudName: "MichelleJacintha",
...     Grade: "VII",
...     Hobbies: "Internetsurfing"
...   }
... ]);
DeprecationWarning: Collection.insert() is deprecated. Use insertOne, insertMany, or bulkWrite.
{
  acknowledged: true,
  insertedId: { $oid: "5e1515a53a2a000000000001" }
}
Atlas atlas-vvji9t-shard-0 [primary] myDB> db.Student1.update(
...   {
...     _id: 1,
...     StudName: "MichelleJacintha",
...     Grade: "VII",
...     Hobbies: "Skating"
...   },
...   {
...     $set: { hobbies: "Skating" },
...     $upsert: true
... });
DeprecationWarning: Collection.update() is deprecated. Use updateOne, updateMany, or bulkWrite.
{
  acknowledged: true,
  insertedId: 1,
  matchedCount: 0,
  modifiedCount: 0,
  upsertedCount: 1
}
Atlas atlas-vvji9t-shard-0 [primary] myDB> db.Student1.find();
\{
  {
    _id: 1,
    StudName: "MichelleJacintha",
    Grade: "VII",
    Hobbies: "Internetsurfing"
  },
  {
    _id: 2,
    StudName: "Aryandavid",
    Grade: "VII",
    Hobbies: "Skating"
  }
\}
Atlas atlas-vvji9t-shard-0 [primary] myDB> db.Student1.update(
...   {
...     _id: 2,
...     StudName: "Aryandavid",
...     Grade: "VII",
...     Hobbies: "Chess"
...   },
...   {
...     acknowledged: true,
...     insertedId: null,
...     matchedCount: 1,
...     modifiedCount: 1,
...     upsertedCount: 0
... });
Atlas atlas-vvji9t-shard-0 [primary] myDB> db.Student1.find();
\{
  {
    _id: 1,
    StudName: "MichelleJacintha",
    Grade: "VII",
    Hobbies: "Internetsurfing"
  },
  {
    _id: 2,
    StudName: "Aryandavid",
    Grade: "VII",
    Hobbies: "Chess"
  }
\}
Atlas atlas-vvji9t-shard-0 [primary] myDB> db.Student.find({studName: "Aryan David"});
Atlas atlas-vvji9t-shard-0 [primary] myDB> db.Student1.find();
\{
  {
    _id: 1,
    StudName: "MichelleJacintha",
    Grade: "VII",
    Hobbies: "Internetsurfing"
  }
\}
Atlas atlas-vvji9t-shard-0 [primary] myDB> db.Student1.find({studName: "Aryan David"});
Atlas atlas-vvji9t-shard-0 [primary] myDB> db.Student1.find({studName: "Aryan David"});
[ { _id: 1, Grade: "VII", StudName: "MichelleJacintha", Hobbies: "Internetsurfing" },
  { _id: 2, Grade: "VII", StudName: "Aryandavid", Hobbies: "Chess" } ]
Atlas atlas-vvji9t-shard-0 [primary] myDB> db.Student1.find({hobbies: { $in: ['Chess', 'Skating'] }}).pretty();
[ {
    _id: 2,
    StudName: "Aryandavid",
    Grade: "VII",
    Hobbies: "Chess"
  },
  {
    _id: 1,
    StudName: "MichelleJacintha",
    Grade: "VII",
    Hobbies: "Internetsurfing"
  }
]
Atlas atlas-vvji9t-shard-0 [primary] myDB> db.Student1.find({studName: /Ary/}).pretty();
[ {
    _id: 2,
    StudName: "Aryandavid",
    Grade: "VII",
    Hobbies: "Chess"
  }
]

```

```

Activities Terminal Mar 11 15:22
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
[1]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find({studName: /VII/}).pretty();
[
  {
    _id: 1,
    StudName: 'MichelleJacintha',
    Grade: 'VII',
    Hobbies: 'Internetsurfing'
  }
]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find({studName: /#/}).pretty();
[
  {
    _id: 1,
    StudName: 'MichelleJacintha',
    Grade: 'VII',
    Hobbies: 'Internetsurfing'
  }
]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.count();
DeprecationWarning: Collection.count() is deprecated. Use countDocuments or estimatedDocumentCount.
2
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find().sort({studName: -1}).pretty();
[
  {
    _id: 1,
    StudName: 'MichelleJacintha',
    Grade: 'VII',
    Hobbies: 'Internetsurfing'
  },
  {
    _id: 1, Grade: 'VIII', StudName: 'Aryandavid', Hobbies: 'Chess' ]
]
Atlas atlas-vvji9t-shard-0 [primary] my08>
(To exit, press Ctrl+C again or Ctrl+B or type .exit)
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find().pretty();
[
  {
    _id: 1,
    StudName: 'MichelleJacintha',
    Grade: 'VII',
    Hobbies: 'Internetsurfing'
  },
  {
    _id: 1, Grade: 'VIII', StudName: 'Aryandavid', Hobbies: 'Chess' ]
]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student3.find();
{
  {
    _id: ObjectId('6700005c20453200f8c4ffcc'),
    StudName: 'Aryandavid',
    Grade: 'VIII',
    Hobbies: 'Chess'
  },
  {
    _id: ObjectId('6700005c20453200f8c4ffcd'),
    StudName: 'MichelleJacintha',
    Grade: 'VII',
    Hobbies: 'Internetsurfing'
  }
]

Activities Terminal Mar 11 15:23
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
[1]
Hobbies: 'Internetsurfing'
]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.save({studName: 'Vansh', Grade: 'VII'});
Error: db.Student1.save is not a function
Atlas atlas-vvji9t-shard-0 [primary] my08> use my08;
already on db my08
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.insertOne({ StudName: "Vansh", Grade: "VII" });
{
  acknowledged: true,
  insertedId: ObjectId('67d0031a8790ab5f31c00e7e')
}
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find();
[
  {
    _id: 1,
    StudName: 'MichelleJacintha',
    Grade: 'VII',
    Hobbies: 'Internetsurfing'
  },
  {
    _id: 1, Grade: 'VII', StudName: 'Aryandavid', Hobbies: 'Chess' ],
  {
    _id: ObjectId('6700031a8790ab5f31c00e7e'),
    StudName: 'Vansh',
    Grade: 'VII'
  }
]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.insertOne({
  ...
  _id: 4,
  ...
  StudName: "JohnDoe",
  ...
  Grade: "VIII",
  ...
  Hobbies: "Reading",
  ...
  Location: "abc"
});
{
  acknowledged: true,
  insertedId: 4
}
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find();
{
  {
    _id: 1,
    StudName: 'MichelleJacintha',
    Grade: 'VII',
    Hobbies: 'Internetsurfing'
  },
  {
    _id: 1, Grade: 'VII', StudName: 'Aryandavid', Hobbies: 'Chess' ],
  {
    _id: ObjectId('6700031a8790ab5f31c00e7e'),
    StudName: 'Vansh',
    Grade: 'VII'
  }
},
{
  _id: 4,
  StudName: 'JohnDoe',
  Grade: 'VIII',
  Hobbies: 'Reading',
  Location: 'abc'
}

```

```

Activities Terminal Mar 11 15:23
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
[{"_id": "5f3e0631a8798ab5f31c00e7e", "StudentName": "MichelleJacintha", "Grade": "VII", "Hobbies": "Internetsurfing"}, {"_id": "5f3e0631a8798ab5f31c00e7e", "StudentName": "AryanDavid", "Grade": "VII", "Hobbies": "Chess"}]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.updateOne(
... { "_id": 3 }, // Find the document with '_id: 3'
... { $set: { Location: "Network" } } // Set the location field to "Network"
)
{
  acknowledged: true,
  insertedId: null,
  lastErrorObject: null,
  modifiedCount: 1,
  upsertedCount: 0
}
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find()
[
  {
    "_id": 3,
    "StudentName": "MichelleJacintha",
    "Grade": "VII",
    "Hobbies": "Internetsurfing"
  },
  {
    "_id": 3,
    "StudentName": "AryanDavid",
    "Grade": "VII",
    "Hobbies": "Chess"
  }
]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find()
[
  {
    "_id": 3,
    "StudentName": "MichelleJacintha",
    "Grade": "VII",
    "Hobbies": "Internetsurfing"
  },
  {
    "_id": 3,
    "Grade": "VII",
    "StudentName": "AryanDavid",
    "Hobbies": "Chess"
  },
  {
    "_id": ObjectId("5f3e0631a8798ab5f31c00e7e"),
    "StudentName": "Vansh",
    "Grade": "VI"
  },
  {
    "_id": 4,
    "StudentName": "JohnDoe",
    "Grade": "VIII",
    "Hobbies": "Reading",
    "Location": "Network"
  }
]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find({ _id: 3, { StudentName: 1, Grade: 1, _id: 0 } })
[ { StudentName: "MichelleJacintha", Grade: "VII" }]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find({ Grade: { $eq: "VII" } })
[
  {
    "_id": ObjectId("5f3e0631a8798ab5f31c00e7e"),
    "StudentName": "Vansh",
    "Grade": "VI"
  }
]
Activities Terminal Mar 11 15:24
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
[{"_id": "5f3e0631a8798ab5f31c00e7e", "StudentName": "Johnson", "Grade": "VIII", "Hobbies": "Reading", "Location": "Network"}]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find({ StudentName: "Johnson" })
[ { _id: "5f3e0631a8798ab5f31c00e7e", "StudentName": "Johnson", "Grade": "VIII", "Hobbies": "Reading", "Location": "Network" }]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.updateOne(
... { "_id": 3 }, // Find the document with '_id: 3'
... { $set: { Location: "Network" } } // Set the location field to "Network"
)
{
  acknowledged: true,
  insertedId: null,
  lastErrorObject: null,
  modifiedCount: 1,
  upsertedCount: 0
}
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.find()
[
  {
    "_id": 3,
    "StudentName": "MichelleJacintha",
    "Grade": "VII",
    "Hobbies": "Internetsurfing"
  },
  {
    "_id": 3,
    "Grade": "VII",
    "StudentName": "AryanDavid",
    "Hobbies": "Chess"
  },
  {
    "_id": ObjectId("5f3e0631a8798ab5f31c00e7e"),
    "StudentName": "Vansh",
    "Grade": "VI"
  },
  {
    "_id": 4,
    "StudentName": "Johnson",
    "Grade": "VIII",
    "Hobbies": "Reading",
    "Location": "Network"
  }
]
Atlas atlas-vvji9t-shard-0 [primary] my08> db.Student1.updateOne({ "_id": 3 }, { $set: { Location: null } })
{
  acknowledged: true,
  insertedId: null,
  lastErrorObject: null,
  modifiedCount: 1,
  upsertedCount: 0
}

```

```
Activities Terminal Mar 11 15:24
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
[{"_id": 1, "StudentName": "MichelleJacintha", "Grade": "VII", "Hobbies": "Internetsurfing"}, {"_id": 2, "StudentName": "AryanDavid", "Grade": "VIII", "Hobbies": "Chess", "Location": null}, {"_id: ObjectId('67d0e31a0790ab5f31c00e7e'), "StudentName": "Vansh", "Grade": "VI"}, {"_id": 4, "StudentName": "JohnDoe", "Grade": "VIII", "Hobbies": "Reading", "Location": "Network"}]
Atlas atlas-vv191-shard-0 [primary] myDB> db.Student1.count();
Atlas atlas-vv191-shard-0 [primary] myDB> db.Student1.countDocuments({Grade: "VII"});
Atlas atlas-vv191-shard-0 [primary] myDB> db.Student1.find({Grade: "VII"}).limit();
[
  {
    "_id": 1,
    "StudentName": "MichelleJacintha",
    "Grade": "VII",
    "Hobbies": "Internetsurfing"
  },
  {
    "_id": 2,
    "Grade": "VIII",
    "StudentName": "AryanDavid",
    "Hobbies": "Chess",
    "Location": null
  }
]
Atlas atlas-vv191-shard-0 [primary] myDB> db.Student1.find().sort({StudentName: 1});
[{"_id": 2,
  "Grade": "VIII",
  "StudentName": "AryanDavid",
  "Hobbies": "Chess",
  "Location": null}
```

```
Activities Terminal Mar 11 15:24
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
[Atlas atlas-vvjt9t-shard-0 {primary} myDB> db.Student1.find().sort({_studentname: 1});
[
  {
    _id: 3,
    Grade: 'VII',
    Studentname: 'AryanDavid',
    Hobbies: 'Chess',
    Location: 'Mell'
  },
  {
    _id: 4,
    Studentname: 'JohnDoe',
    Grade: 'VIII',
    Hobbies: 'Reading',
    Location: 'Network'
  },
  {
    _id: 5,
    Studentname: 'MichelleJackntha',
    Grade: 'VII',
    Hobbies: 'Internetsurfing'
  },
  {
    _id: ObjectId('07de31a0790ab5f31c00e7e'),
    Studentname: 'Vansh',
    Grade: 'VI'
  }
]
Atlas atlas-vvjt9t-shard-0 {primary} myDB> db.Student1.find().skip();
[
  {
    _id: ObjectId('07de31a0790ab5f31c00e7e'),
    Studentname: 'Vansh',
    Grade: 'VI'
  },
  {
    _id: 4,
    Studentname: 'JohnDoe',
    Grade: 'VIII',
    Hobbies: 'Reading',
    Location: 'Network'
  }
]
Atlas atlas-vvjt9t-shard-0 {primary} myDB> db.food.insert({ _id: 1, fruits: ['grapes', 'mango', 'apple'] });
{
  acknowledged: true,
  insertedIds: [ '_id: 1' ]
}
Atlas atlas-vvjt9t-shard-0 {primary} myDB> db.food.insert({ _id: 2, fruits: ['grapes', 'mango', 'cherry'] });
{
  acknowledged: true,
  insertedIds: [ '_id: 2' ]
}
Atlas atlas-vvjt9t-shard-0 {primary} myDB> db.food.insert({ _id: 3, fruits: ['banana', 'mango'] });
{
  acknowledged: true,
  insertedIds: [ '_id: 3' ]
}
Atlas atlas-vvjt9t-shard-0 {primary} myDB>
Atlas atlas-vvjt9t-shard-0 {primary} myDB> db.food.find()
[
  {
    _id: 1,
    fruits: [ 'grapes', 'mango', 'apple' ],
  },
  {
    _id: 2,
    fruits: [ 'grapes', 'mango', 'cherry' ],
  },
  {
    _id: 3,
    fruits: [ 'banana', 'mango' ],
  }
]
```

```

Activities Terminal Mar 11 15:24
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
[{"_id": 1, "fruits": ["grapes", "mango", "cherry"]}, {"_id": 2, "fruits": ["banana", "mango"]}], Atlas atlas-vv19t-shard-0 [primary] my0@> db.food.find({ "fruits": ["grapes", "mango", "apple"] });
[{"_id": 1, "fruits": ["grapes", "mango", "apple"]}], Atlas atlas-vv19t-shard-0 [primary] my0@> db.food.find({ "fruits": ["grapes"] });
[{"_id": 1, "fruits": ["grapes"]}], Atlas atlas-vv19t-shard-0 [primary] my0@> db.food.find({ "fruits": {"$size": 3} });
[{"_id": 1, "fruits": ["grapes", "mango", "apple"]}, {"_id": 2, "fruits": ["grapes", "mango", "cherry"]}], Atlas atlas-vv19t-shard-0 [primary] my0@> db.food.find({ "fruits": {"$all": ["mango", "grapes"]} });
[{"_id": 1, "fruits": ["grapes", "mango", "apple"]}, {"_id": 2, "fruits": ["grapes", "mango", "cherry"]}], Atlas atlas-vv19t-shard-0 [primary] my0@> db.food.find({ "fruits.o": "grapes" });
[{"_id": 1, "fruits": {"$all": ["grapes", "mango", "apple"]}}, {"_id": 2, "fruits": {"$all": ["grapes", "mango", "cherry"]}}], Atlas atlas-vv19t-shard-0 [primary] my0@> db.food.find();
[{"_id": 1, "fruits": ["grapes", "mango", "apple"]}, {"_id": 2, "fruits": ["grapes", "mango", "cherry"]}, {"_id": 3, "fruits": ["banana", "mango"]}], Atlas atlas-vv19t-shard-0 [primary] my0@> db.food.update({ "_id": 3 }, { $set: { "fruits.o": "apple" } });
{
  acknowledged: true,
  insertedId: null,
  matchedCount: 1,
  modifiedCount: 1,
  upsertedCount: 0
}
Atlas atlas-vv19t-shard-0 [primary] my0@> db.food.find();
[{"_id": 1, "fruits": ["grapes", "mango", "apple"]}, {"_id": 2, "fruits": ["grapes", "mango", "cherry"]}, {"_id": 3, "fruits": ["banana", "apple"]}], Atlas atlas-vv19t-shard-0 [primary] my0@> db.food.update({ "_id": 3 }, { $push: { "fruits": "apple" } });
{
  acknowledged: true,
  insertedId: null,
  matchedCount: 1,
  modifiedCount: 1,
  upsertedCount: 0
}
Atlas atlas-vv19t-shard-0 [primary] my0@> db.food.find();
[{"_id": 1, "fruits": ["grapes", "mango", "apple"]}, {"_id": 2, "fruits": ["grapes", "mango", "cherry"]}, {"_id": 3, "fruits": ["banana", "apple", "apple"]}], Activities Terminal Mar 11 15:25
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
[{"price": [{"grapes": 80, "mango": 200, "cherry": 100}], {"_id": 1, "fruits": ["banana", "apple"]}], Atlas atlas-vv19t-shard-0 [primary] my0@> db.food.update({ "_id": 1 }, { $push: { "fruits": {} } });
{
  acknowledged: true,
  insertedId: null,
  matchedCount: 1,
  modifiedCount: 1,
  upsertedCount: 0
}
Atlas atlas-vv19t-shard-0 [primary] my0@> db.food.find();
[{"_id": 1, "fruits": ["grapes", "mango"]}, {"_id": 2, "fruits": {"$all": ["grapes", "mango", "cherry"]}}, {"_id": 3, "fruits": ["banana", "apple"]}], Atlas atlas-vv19t-shard-0 [primary] my0@> db.food.update({ "_id": 2 }, { $push: { "fruits": "orange" } });
{
  acknowledged: true,
  insertedId: null,
  matchedCount: 1,
  modifiedCount: 1,
  upsertedCount: 0
}
Atlas atlas-vv19t-shard-0 [primary] my0@> db.food.find();
[{"_id": 1, "fruits": ["grapes", "mango", "orange"]}, {"_id": 2, "fruits": ["grapes", "mango", "cherry", "orange"]}, {"_id": 3, "fruits": ["banana", "apple"]}], Atlas atlas-vv19t-shard-0 [primary] my0@> db.food.update({ "_id": 2 }, { $pullAll: { "fruits": ["grapes", "mango"] } });
{
  acknowledged: true,
  insertedId: null,
  matchedCount: 1,
  modifiedCount: 1,
  upsertedCount: 0
}
Atlas atlas-vv19t-shard-0 [primary] my0@> db.food.find();
[{"_id": 1, "fruits": ["grapes", "mango", "orange"]}, {"_id": 2, "fruits": ["cherry"]}, {"_id": 3, "fruits": ["banana", "apple"]}],
```

```
Activities Terminal Mar 11 15:25
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
[{"_id": 1, "fruits": ["grapes", "mango", "cherry", "apple"]}, {"_id": 2, "fruits": ["banana", "apple"]}], Atlas atlas-vvjl9t-shard-0 [primary] myDB> db.food.update({ "_id": 3 }, { $pull: { fruits: "banana" } });
{
  acknowledged: true,
  insertedId: null,
  matchedCount: 1,
  modifiedCount: 1,
  upsertedCount: 0
}
Atlas atlas-vvjl9t-shard-0 [primary] myDB> db.food.find()
[{"_id": 1, "fruits": ["grapes", "mango", "orange"]}, {"_id": 2, "fruits": ["cherry"], "price": [{"grapes": 80, "mango": 300, "cherry": 100}], {"_id": 3, "fruits": ["apple"]}], Atlas atlas-vvjl9t-shard-0 [primary] myDB> db.createCollection("customers");
{ "ok": 1 }
Atlas atlas-vvjl9t-shard-0 [primary] myDB>
Atlas atlas-vvjl9t-shard-0 [primary] myDB> db.Customers.insertMany([
  ... { custID: 1, AcctBal: 1000, AcctType: "S" },
  ... { custID: 2, AcctBal: 1500, AcctType: "S" },
  ... { custID: 3, AcctBal: 1200, AcctType: "C" },
  ... { custID: 4, AcctBal: 900, AcctType: "C" },
  ... { custID: 5, AcctBal: 500, AcctType: "C" }
  ... ])
{
  acknowledged: true,
  insertedIds: [
    "5: ObjectId('67d88708a799ab5f31c0e7ff')",
    "6: ObjectId('67d88708a799ab5f31c0e800')",
    "7: ObjectId('67d88708a799ab5f31c0e801')",
    "8: ObjectId('67d88708a799ab5f31c0e802')",
    "9: ObjectId('67d88708a799ab5f31c0e83')"
  ]
}
Atlas atlas-vvjl9t-shard-0 [primary] myDB> db.Customer.find()
Atlas atlas-vvjl9t-shard-0 [primary] myDB> db.Customers.find()
[{"_id": ObjectId("67d88708a799ab5f31c0e7ff"), "custID": 1, "AcctBal": 1000, "AcctType": "S"}, {"_id": ObjectId("67d88708a799ab5f31c0e800"), "custID": 2, "AcctBal": 1500, "AcctType": "S"}, {"_id": ObjectId("67d88708a799ab5f31c0e801"), "custID": 3, "AcctBal": 1200, "AcctType": "C"}, {"_id": ObjectId("67d88708a799ab5f31c0e802"), "custID": 4, "AcctBal": 900, "AcctType": "C"}, {"_id": ObjectId("67d88708a799ab5f31c0e83"), "custID": 5, "AcctBal": 500, "AcctType": "C"}]
```

```
Activities Terminal Mar 11 15:25
mongosh mongodb://credentials@cluster0.tud3r.mongodb.net/
Atlas atlas-vv191-shard-0 [primary] myDB> db.customers.find()
{
  "_id": ObjectId("07d0e7098790ab5f31c08e7f"),
  "custID": 1,
  "AcctBal": 1000,
  "AcctType": "S"
},
{
  "_id": ObjectId("07d0e7098790ab5f31c08e80"),
  "custID": 2,
  "AcctBal": 1000,
  "AcctType": "S"
},
{
  "_id": ObjectId("07d0e7098790ab5f31c08e81"),
  "custID": 3,
  "AcctBal": 1000,
  "AcctType": "C"
},
{
  "_id": ObjectId("07d0e7098790ab5f31c08e82"),
  "custID": 4,
  "AcctBal": 1000,
  "AcctType": "S"
},
{
  "_id": ObjectId("07d0e7098790ab5f31c08e83"),
  "custID": 5,
  "AcctBal": 1000,
  "AcctType": "S"
}
Atlas atlas-vv191-shard-0 [primary] myDB> db.customers.aggregate([
... {
...   $group: {
...     _id: "$custID",
...     TotAccBal: { $sum: "$AcctBal" }
...   }
... }
... ]);
[{"_id": 1, "TotAccBal": 2000}, {"_id": 2, "TotAccBal": 1000}, {"_id": 3, "TotAccBal": 1000}, {"_id": 4, "TotAccBal": 1000}, {"_id": 5, "TotAccBal": 1000}]
Atlas atlas-vv191-shard-0 [primary] myDB> db.customers.aggregate([
... {
...   $group: {
...     _id: "$custID",
...     TotAccBal: { $sum: "$AcctBal" }
...   }
... }
... ]);

```

```
Activities Terminal Mar 11 15:26 bmscscse@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC ~
bmscscse@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ mongoexport
2025-03-11T14:39:05.282+0530 must specify a collection
2025-03-11T14:39:05.282+0530 try 'mongoexport --help' for more information
bmscscse@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ ^[[200~mongosh "mongodb://cluster0.tud3r.mongodb.net/" --apiVersion 1 --username marlacs22 --password B3a9u6zIrEjzBlot
MongoSH: command not found
bmscscse@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ .mongosh "mongodb://cluster0.tud3r.mongodb.net/" --apiVersion 1 --username marlacs22 --password B3a9u6zIrEjzBlot
MongoSH: command not found
bmscscse@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ .mongosh "mongodb://cluster0.tud3r.mongodb.net/" --apiVersion 1 --username marlacs22 --password B3a9u6zIrEjzBlot
Current Mongosh Log ID: 6e7cfdffccba3:86154cb8e07d
Connecting to: mongosh://[credentials@cluster0.tud3r.mongodb.net/mongosh]@2.2.1
Using MongoDB: 8.0.5 (API Version 1)
Using Mongosh: 2.2.1
Using Mongosh: 2.2.1
mongosh 2.2.2 is available for download: https://www.mongodb.com/tryp/download/shell

For mongosh info see: https://docs.mongodb.com/mongosh/shell/

Atlas atlas-vvjl9t-shard-0 [primary] test> use myDB;
switched to db myDB
Atlas atlas-vvjl9t-shard-0 [primary] myDB> mongoexport --uri "mongodb+srv://marlacs22:83a9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "stu
dName,Grade,Hobbies"
Uncought:
SyntaxError: Missing semicolon. (1:14)
> 1 | mongoexport --url "mongodb+srv://marlacs22:83a9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection student1 --type csv --out "D:/output.txt" --fields "studName,Grade,Hobbies"
2 |
Atlas atlas-vvjl9t-shard-0 [primary] myDB> mongoexport --uri "mongodb+srv://marlacs22:83a9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "stu
dName,Grade,Hobbies"
Uncought:
SyntaxError: Missing semicolon. (1:14)
> 1 | mongoexport --url "mongodb+srv://marlacs22:83a9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection student1 --type csv --out "D:/output.txt" --fields "studName,Grade,Hobbies"
2 |
Atlas atlas-vvjl9t-shard-0 [primary] myDB>
(To exit, press Ctrl+C again or Ctrl+D or type .exit)
Atlas atlas-vvjl9t-shard-0 [primary] myDB> mongoexport --uri "mongodb+srv://marlacs22:83a9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "stu
dName,Grade,Hobbies"
Uncought:
SyntaxError: Missing semicolon. (1:14)
> 1 | mongoexport --url "mongodb+srv://marlacs22:83a9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection student1 --type csv --out "D:/output.txt" --fields "studName,Grade,Hobbies"
2 |
3 |

Atlas atlas-vvjl9t-shard-0 [primary] myDB> ext
bmscscse@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ mongoexport --url "mongodb+srv://marlacs22:83a9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "studName,Grade,Hobbies"

2025-03-11T14:42:20.813+0530 connected to: mongodb+srv://[**REDACTED**]@cluster0.tud3r.mongodb.net/myDB
2025-03-11T14:42:23.066+0530 exported 2 records
bmscscse@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ 
bmscscse@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ mongoimport --url "mongodb+srv://marlacs22:83a9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/mynB" --collection Student1 --type csv --out "D:/out.txt" --file
ids "StudentName,Grade,Hobbies"

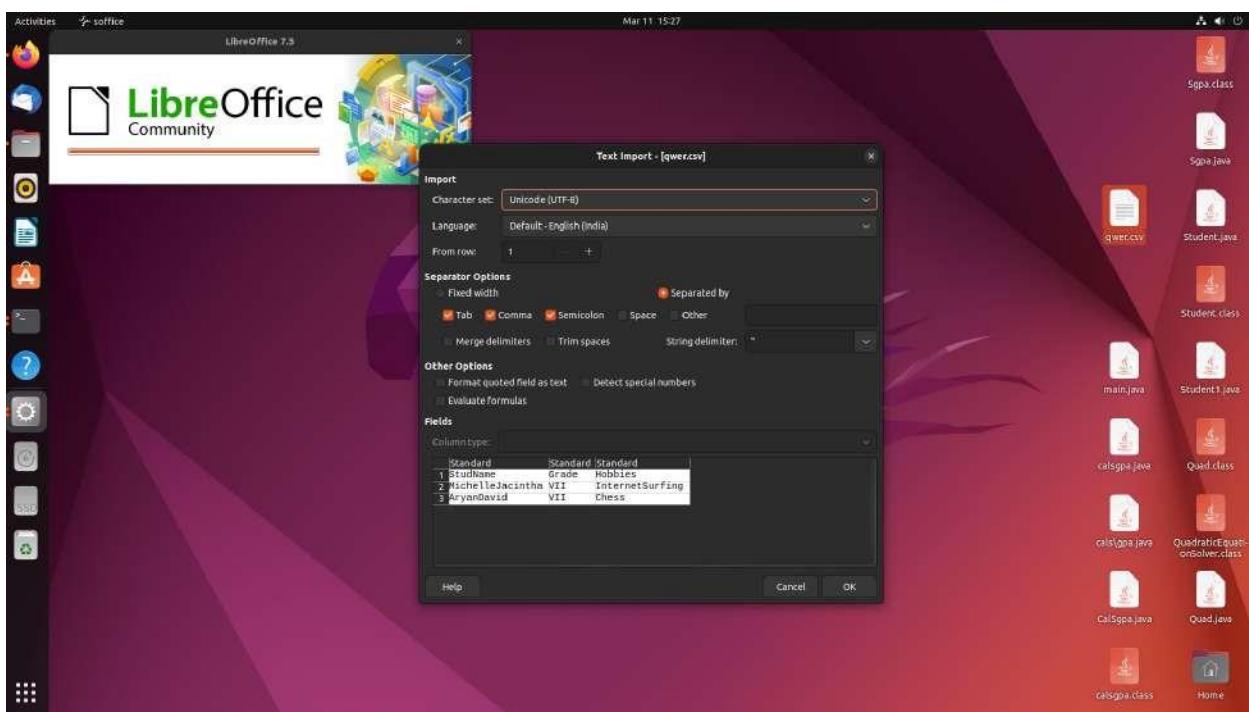
Activities Terminal Mar 11 15:26 bmscscse@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC ~
bmscscse@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ mongoexport
2025-03-11T14:51:35.744+0530 connected to: mongodb+srv://[**REDACTED**]@cluster0.tud3r.mongodb.net/myDB
2025-03-11T14:51:35.744+0530 exported 2 records
bmscscse@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ 
bmscscse@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ mongoexport --url "mongodb+srv://marlacs22:83a9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudentName,Grade,Hobbies"
Uncought:
SyntaxError: Missing semicolon. (1:14)
> 1 | mongoexport --url "mongodb+srv://marlacs22:83a9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection student1 --type csv --out "D:/output.txt" --fields "studName,Grade,Hobbies"
2 |
Atlas atlas-vvjl9t-shard-0 [primary] myDB>
(To exit, press Ctrl+C again or Ctrl+D or type .exit)
Atlas atlas-vvjl9t-shard-0 [primary] myDB> mongoexport --uri "mongodb+srv://marlacs22:83a9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "stu
dName,Grade,Hobbies"
Uncought:
SyntaxError: Missing semicolon. (1:14)
> 1 | mongoexport --url "mongodb+srv://marlacs22:83a9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection student1 --type csv --out "D:/output.txt" --fields "studName,Grade,Hobbies"
2 |
3 |

Atlas atlas-vvjl9t-shard-0 [primary] myDB> ext
bmscscse@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ mongoexport --url "mongodb+srv://marlacs22:83a9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "studName,Grade,Hobbies"

2025-03-11T14:42:20.813+0530 connected to: mongodb+srv://[**REDACTED**]@cluster0.tud3r.mongodb.net/myDB
2025-03-11T14:42:21.066+0530 exported 2 records
bmscscse@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ 
bmscscse@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ mongoexport --url "mongodb+srv://marlacs22:83a9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/mynB" --collection Student1 --type csv --out "D:/out.txt" --file
ids "StudentName,Grade,Hobbies"

2025-03-11T14:43:35.437+0530 connected to: mongodb+srv://[**REDACTED**]@cluster0.tud3r.mongodb.net/mynB
2025-03-11T14:43:35.437+0530 exported 2 records
bmscscse@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ 
bmscscse@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ mongoimport --url "mongodb+srv://marlacs22:83a9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/mynB" --collection Student1 --type csv --out "D:/quer.txt" --fi
elds "StudentName,Grade,Hobbies"
2025-03-11T14:44:44.222+0530 connected to: mongodb+srv://[**REDACTED**]@cluster0.tud3r.mongodb.net/mynB
2025-03-11T14:44:44.529+0530 exported 2 records
bmscscse@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ whoami
bmscscse@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ ^C
bmscscse@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ mongoexport --url "mongodb+srv://marlacs22:83a9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/mynB" --collection Student1 --type csv --out "/home/bmscscse/de
sktop/qwer.csv" --fields "StudentName,Grade,Hobbies"
2025-03-11T14:47:52.993+0530 connected to: mongodb+srv://[**REDACTED**]@cluster0.tud3r.mongodb.net/mynB
2025-03-11T14:47:53.277+0530 exported 2 records
bmscscse@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ mongoimport --url "mongodb+srv://marlacs22:83a9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/mynB" --collection Student3 --type csv --file "/home/bmscscse/de
sktop/qwer.csv" --headerline
2025-03-11T14:51:32.165+0530 connected to: mongodb+srv://[**REDACTED**]@cluster0.tud3r.mongodb.net/mynB
2025-03-11T14:51:32.316+0530 2 document(s) imported successfully. 0 document(s) failed to import.
bmscscse@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ 
bmscscse@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ mongoimport --url "mongodb+srv://marlacs22:83a9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/mynB" --collection Student1 --type csv --headline --file "/de
sktop/airline.csv"
2025-03-11T14:55:09.122+0530 Failed: open /home/bmscscse/Desktop/airline.csv: no such file or directory
2025-03-11T14:55:09.122+0530 0 document(s) imported successfully. 0 document(s) failed to import.
bmscscse@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ mongoimport --url "mongodb+srv://marlacs22:83a9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/mynB" --collection Student1 --type csv --headline --file "/de
sktop/airline.csv"
2025-03-11T14:56:05.413+0530 Failed: open /home/bmscscse/Desktop/airline.csv: no such file or directory
2025-03-11T14:56:05.413+0530 0 document(s) imported successfully. 0 document(s) failed to import.
bmscscse@bmscscse-HP-Elite-Tower-800-G9-Desktop-PC:~$ mongoimport --url "mongodb+srv://marlacs22:83a9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/mynB" --collection Student1 --type csv --headline --file "/de
sktop/airline.csv"
2025-03-11T14:56:05.413+0530 Failed: open /home/bmscscse/Desktop/airline.csv: no such file or directory
2025-03-11T14:56:05.413+0530 0 document(s) imported successfully. 0 document(s) failed to import.
```

```
Activities ┌ Terminal ┌ Mar 11 15:29
bmscsece@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC ~
diane_Grade_Hobbies"
Uncaught:
System.err: Missing semicolon. (1:14)
> i | mongoexport --url "mongodb+srv://mariacs22:8Jaq9u0zIrEjzb0t@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudentName,Grade,Hobbies"
2
Atlas atlas-Vv191 shard-0 [primary] myDB> exit
bmscsece@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC ~ 5 mongoexport --url "mongodb+srv://mariacs22:8Jaq9u0zIrEjzb0t@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" ...
Fields "StudentName,Grade,Hobbies"
2025-03-11T14:42:20.811+0530 connected to: mongodb+srv://[**REDACTED**]@cluster0.tud3r.mongodb.net/myDB
2025-03-11T14:42:21.803+0530 exported 2 records
bmscsece@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC ~ 5 bmscsece@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC ~ 5 bmscsece@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC ~ 5 mongoexport --url "mongodb+srv://mariacs22:8Jaq9u0zIrEjzb0t@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --file "studName,Grade,Hobbies"
ld "studName,Grade,Hobbies"
2025-03-11T14:43:20.811+0530 connected to: mongodb+srv://[**REDACTED**]@cluster0.tud3r.mongodb.net/myDB
2025-03-11T14:43:35.437+0530 exported 2 records
bmscsece@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC ~ 5 bmscsece@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC ~ 5 bmscsece@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC ~ 5 bmscsece@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC ~ 5 mongoexport --url "mongodb+srv://mariacs22:8Jaq9u0zIrEjzb0t@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/quer.txt" --file "StudentName,Grade,Hobbies"
2025-03-11T14:44:44.222+0530 connected to: mongodb+srv://[**REDACTED**]@cluster0.tud3r.mongodb.net/myDB
2025-03-11T14:44:44.529+0530 exported 2 records
bmscsece@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC ~ 5 whoami
bmscsece@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC ~ 5 ^C
bmscsece@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC ~ 5 bmscsece@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC ~ 5 mongoexport --url "mongodb+srv://mariacs22:8Jaq9u0zIrEjzb0t@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "/home/bmscsece/be sktop/quer.csv" --fields "StudentName,Grade,Hobbies"
2025-03-11T14:47:52.993+0530 connected to: mongodb+srv://[**REDACTED**]@cluster0.tud3r.mongodb.net/myDB
2025-03-11T14:47:53.000+0530 exported 2 records
bmscsece@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC ~ 5 bmscsece@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC ~ 5 mongoimport --url "mongodb+srv://mariacs22:8Jaq9u0zIrEjzb0t@cluster0.tud3r.mongodb.net/myDB" --collection Student3 --type csv --file "/home/bmscsece/be sktop/quer.csv" --headerline
2025-03-11T14:51:32.165+0530 connected to: mongodb+srv://[**REDACTED**]@cluster0.tud3r.mongodb.net/myDB
2025-03-11T14:51:32.316+0530 2 document(s) imported successfully. 0 document(s) failed to import.
bmscsece@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC ~ 5 bmscsece@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC ~ 5 mongoimport --url "mongodb+srv://mariacs22:8Jaq9u0zIrEjzb0t@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --headerline --file "/D ownloads/airline.csv"
2025-03-11T14:55:09.112+0530 Failed: open /home/bmscsece/Downloads/airline.csv: no such file or directory
2025-03-11T14:55:09.112+0530 0 document(s) imported successfully. 0 document(s) failed to import.
bmscsece@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC ~ 5 bmscsece@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC ~ 5 mongoimport --url "mongodb+srv://mariacs22:8Jaq9u0zIrEjzb0t@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --headerline --file "/be staircase.csv"
2025-03-11T14:56:05.413+0530 Failed: open /home/bmscsece/Desktop/airline.csv: no such file or directory
2025-03-11T14:56:05.413+0530 0 document(s) imported successfully. 0 document(s) failed to import.
bmscsece@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC ~ 5 bmscsece@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC ~ 5 mongoexport -h localhost -d myDB --collection Student1 --type=csv --out "/D ownloads/airline.csv" --fields "Year","Quarter"
2025-03-11T14:58:48.878+0530 connected to: mongodb+srv://localhost/
2025-03-11T14:58:48.878+0530 exported 0 records
bmscsece@bmscsece-HP-Elite-Tower-800-G9-Desktop-PC ~
```



New... X +

cluster0.tudr.mongodb.net > test > New\_Student

Open MongoDB shell

Documents Aggregations Schema Indexes Validation

Type a query: { field: 'value' } or [Generate query](#)

Add DATA • [EXPORT DATA](#) [UPDATE](#) [DELETE](#)

Explain Reset Find Options

26 1-8 of 6

`_id: ObjectId('67c6cd10cf8aef0fa4218')`  
RollNo: 9  
Age: 21  
Cont: 5576  
email: "anushka.de9@gmail.com"

`_id: ObjectId('67c6cd10cf8aef0fa4215')`  
RollNo: 2  
Age: 22  
Cont: 6976  
email: "anushka.de9@gmail.com"

`_id: ObjectId('67c6cd10cf8aef0fa4214')`  
RollNo: 1  
Age: 22  
Name: "TEH"  
Cont: 2276  
email: "teha.de9@gmail.com"

`_id: ObjectId('67c6cd10cf8aef0fa4213')`  
RollNo: 4  
Age: 28  
Cont: 4476  
email: "santi.design@gmail.com"

`_id: ObjectId('67c6cd10cf8aef0fa4210')`  
RollNo: 18  
Age: 20  
Cont: 2276  
email: "santosh@gmail.com"

`_id: ObjectId('67c6cd10cf8aef0fa4211')`  
RollNo: 3  
Age: 21  
Cont: 5576  
email: "anubhav.de9@gmail.com"

`_id: ObjectId('67c6cd10cf8aef0fa4212')`  
RollNo: 2  
Age: 22  
Cont: 6976  
email: "anushka.de9@gmail.com"

`_id: ObjectId('67c6cd10cf8aef0fa4213')`  
RollNo: 1  
Age: 21  
Cont: 9676  
email: "antara.de9@gmail.com"

`_id: ObjectId('67c6cd10cf8aef0fa4214')`  
RollNo: 11  
Age: 22  
Name: "TEH"  
Cont: 2276  
email: "teha.de9@gmail.com"

`_id: ObjectId('67c6cd10cf8aef0fa4215')`  
RollNo: 4  
Age: 20  
Cont: 4476  
email: "pani.de9@gmail.com"

`_id: ObjectId('67c6cd10cf8aef0fa4216')`  
RollNo: 10  
Age: 23  
Cont: 2276  
email: "Abhinav@gmail.com"

Search

ENG

15:20 4:46

04-03-2022

Compass

cluster0.tud3.mongodb.net > test > student

Documents Aggregations Schema Indexes Validation

Type a query: { field: "value" } or **Generate query**

**ADD DATA** EXPORT DATA UPDATE DELETE

Explain Reset Find Options

20 1-5 of 6

```
_id: ObjectId("5f0c081a10cf096e2bf04214")
RollNo: 1
Age: 21
Cont: 8876
email: "anurika.deo@gmail.com"

_id: ObjectId("5f0c081a10cf096e2bf04215")
RollNo: 2
Age: 21
Cont: 9976
email: "anurika.deo@gmail.com"

_id: ObjectId("5f0c081a10cf096e2bf04216")
RollNo: 3
Age: 21
Cont: 4476
email: "priyanka.deo@gmail.com"

_id: ObjectId("5f0c081a10cf096e2bf04217")
RollNo: 18
Age: 20
Cont: 2276
email: "Akshay.deo@gmail.com"

_id: ObjectId("5f0c081a10cf096e2bf04218")
RollNo: 11
Age: 21
Name: "TEJA"
Cont: 2276
email: "teja.deo@gmail.com"

_id: ObjectId("5f0c081a10cf096e2bf04219")
RollNo: 11
Age: 21
Name: "TEJA"
Cont: 2276
email: "teja.deo@gmail.com"
```

Open MongoDB shell

32% Party sunny

15:21:59 04-09-2025

Compass

cluster0.tud3.mongodb.net > test > student

Documents Aggregations Schema Indexes Validation

Type a query: { field: "value" } or **Generate query**

**ADD DATA** EXPORT DATA UPDATE DELETE

Explain Reset Find Options

20 1-5 of 6

```
_id: ObjectId("5f0c081a10cf096e2bf04215")
RollNo: 2
Age: 21
Cont: 9976
email: "anurika.deo@gmail.com"

_id: ObjectId("5f0c081a10cf096e2bf04216")
RollNo: 3
Age: 21
Cont: 4476
email: "priyanka.deo@gmail.com"

_id: ObjectId("5f0c081a10cf096e2bf04217")
RollNo: 18
Age: 20
Cont: 2276
email: "Akshay.deo@gmail.com"

_id: ObjectId("5f0c081a10cf096e2bf04218")
RollNo: 11
Age: 21
Name: "TEJA"
Cont: 2276
email: "teja.deo@gmail.com"

_id: ObjectId("5f0c081a10cf096e2bf04219")
RollNo: 11
Age: 21
Name: "TEJA"
Cont: 2276
email: "teja.deo@gmail.com"
```

Open MongoDB shell

32% Party sunny

15:21:59 04-09-2025

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
1	"_id": "5\$RollNo1", "Age": 21, "Cont": "+919899910000", "email": "anuradha.ds@gmail.com"}																												
2	"_id": "5\$RollNo2", "Age": 21, "Cont": "+919596666666", "email": "anuradha.ds@gmail.com"}																												
3	"_id": "5\$RollNo3", "Age": 21, "Cont": "+915576666666", "email": "anuradha.ds@gmail.com"}																												
4	"_id": "5\$RollNo4", "Age": 20, "Cont": "+918427666666", "email": "pankaj.ds@gmail.com"}																												
5	"_id": "5\$RollNo10", "Age": 23, "Cont": "+919276666666", "email": "Abhinav.ds@gmail.com"}																												
6	"_id": "5\$RollNo11", "Age": 22, "Name": "FE", "Cont": "+919276666666", "email": "rea.ds@gmail.com"}																												

```

C:\Users\3 STUDENT> python.json -
[{"_id": "$oid": "57a5c681bcf8a62f84210", "RollNo": "1", "Age": "21", "Cont": "+919596666666", "email": "anuradha.ds@gmail.com"}, {"_id": "$oid": "57a5c681bcf8a62f84210", "RollNo": "2", "Age": "21", "Cont": "+919596666666", "email": "anuradha.ds@gmail.com"}, {"_id": "$oid": "57a5c681bcf8a62f84210", "RollNo": "3", "Age": "21", "Cont": "+915576666666", "email": "anuradha.ds@gmail.com"}, {"_id": "$oid": "57a5c681bcf8a62f84210", "RollNo": "4", "Age": "20", "Cont": "+918427666666", "email": "pankaj.ds@gmail.com"}, {"_id": "$oid": "57cnc8351bcf8a62f84210", "RollNo": "10", "Age": "23", "Cont": "+919276666666", "email": "Abhinav.ds@gmail.com"}, {"_id": "$oid": "57cnc8351bcf8a62f84210", "RollNo": "11", "Age": "22", "Name": "FE", "Cont": "+919276666666", "email": "rea.ds@gmail.com"}]

```

```
Command Prompt      X + v - □ ×
Microsoft Windows [Version 10.0.22631.4890]
(c) Microsoft Corporation. All rights reserved.

C:\Users\STUDENT>mongoexport --uri="mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/test" --collection=Student --out=abc.csv
2025-03-04T15:09:05.137+0530      connected to: mongodb+srv://[**REDACTED**]@cluster0.tud3r.mongodb.net/test
2025-03-04T15:09:05.359+0530      exported 6 records

C:\Users\STUDENT>mongoexport --uri="mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/test" --collection=Student --out=efg.json
2025-03-04T15:14:50.067+0530      connected to: mongodb+srv://[**REDACTED**]@cluster0.tud3r.mongodb.net/test
2025-03-04T15:14:50.288+0530      exported 6 records

C:\Users\STUDENT>mongoimport --uri="mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/test" --collection>New_Student --type=json --file="C:\Users\STUDENT\efg.json"
2025-03-04T15:15:35.501+0530      connected to: mongodb+srv://[**REDACTED**]@cluster0.tud3r.mongodb.net/test
2025-03-04T15:15:35.590+0530      6 document(s) imported successfully. 0 document(s) failed to import.

C:\Users\STUDENT>
```

## EXPERIMENT-2

Neo4j

Code with Screenshots:

The screenshot shows the Neo4j browser interface. On the left, there is a sidebar with three tabs: 'Table' (selected), 'Code', and 'Graph'. The main area contains a command line window and a graph visualization.

**Command Line Window:**

```
neo4j$ CREATE (a:STUDENT {name:"Ram" , id: 12})
```

Output:

Added 1 label, created 1 node, set 2 properties, completed after 27 ms.

**Graph Visualization:**

```
neo4j$ MATCH (n) RETURN n
```

The graph view displays a single pink circular node labeled "Ram". To the right of the graph, the "Overview" panel shows the following information:

- Node labels:** \* (1) STUDENT (1)
- Displaying 1 nodes, 0 relationships.

Below the graph are three small icons: a magnifying glass, a search icon, and a refresh icon.

```
neo4j$ CREATE (a:STUDENT {name:"Abhi" , id: 14})
```

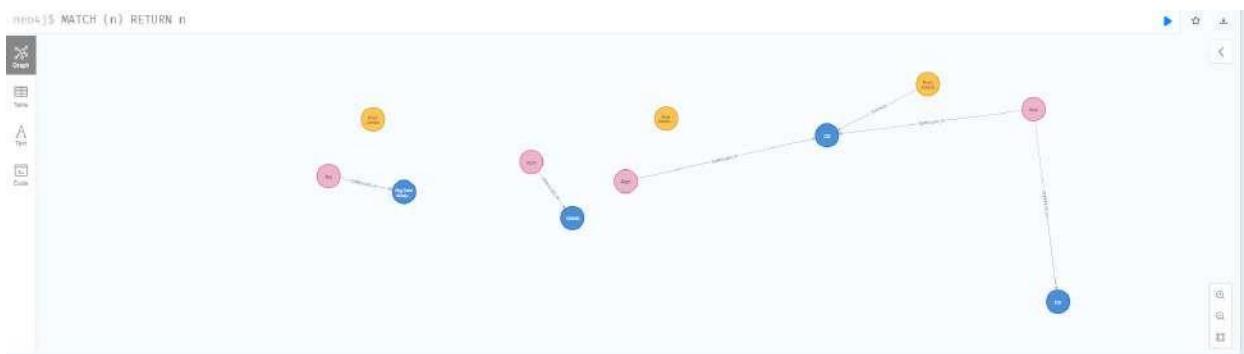
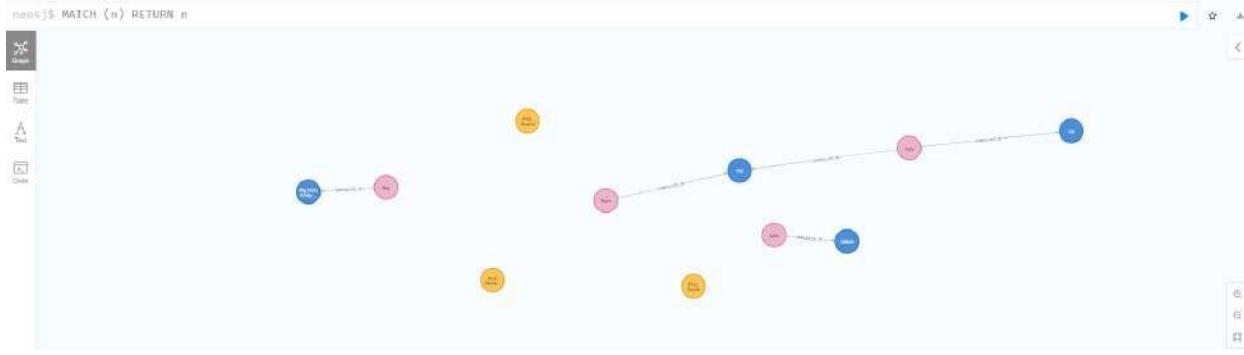
Added 1 label, created 1 node, set 2 properties, completed after 6 ms.

```
neo4j$ CREATE (a:STUDENT {name:"Raj" , id: 13})
```

Added 1 label, created 1 node, set 2 properties, completed after 6 ms.

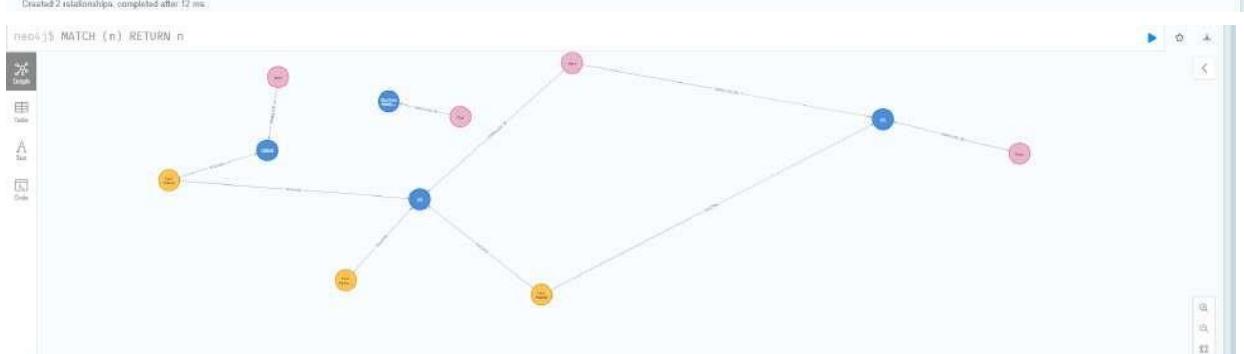
```
neo4j$ CREATE (a:STUDENT {name:"Ajay" , id: 15})
```

Added 1 label, created 1 node, set 2 properties, completed after 6 ms.



neo4j:~> MATCH (a:TEACHER) , (b:COURSE) where a.name = "Prof. Anand" AND b.name = "OS" CREATE (a)-[r:TEACHES]→(b)

Created 2 relationships, completed after 12 ms.



neo4j:~> MATCH (a:TEACHER) , (b:COURSE) where a.name = "Prof. Aman" AND b.name = "DBMS" OR b.name = "DS" CREATE (a)-[r:TEACHES]→(b)

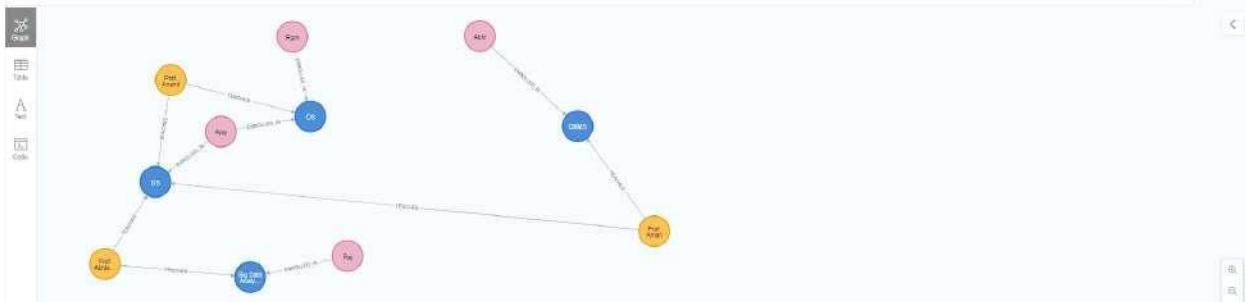
Created 4 relationships, completed after 12 ms.

Created 4 relationships, completed after 12 ms.

```
neo4j$ MATCH (a:STUDENT {name:"Ajay"}) SET a.id = 21 RETURN a
```



```
neo4j$ MATCH (n) RETURN n
```

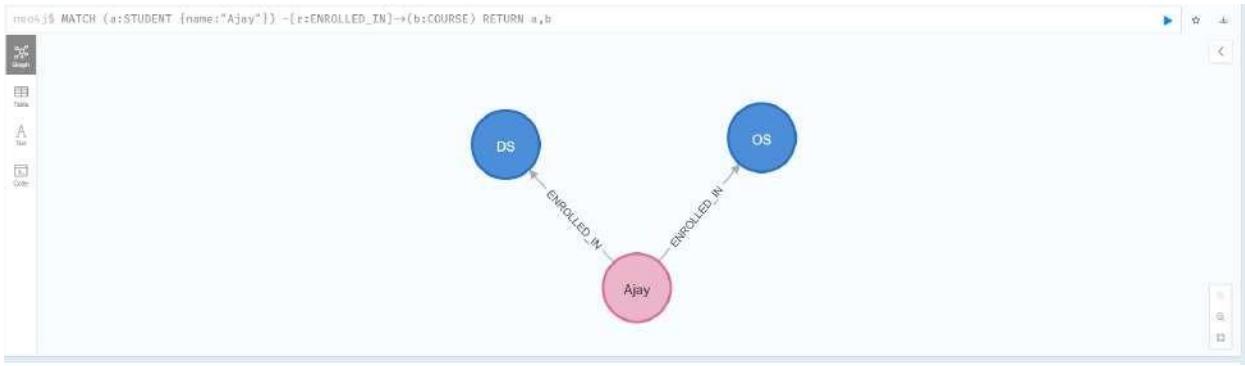


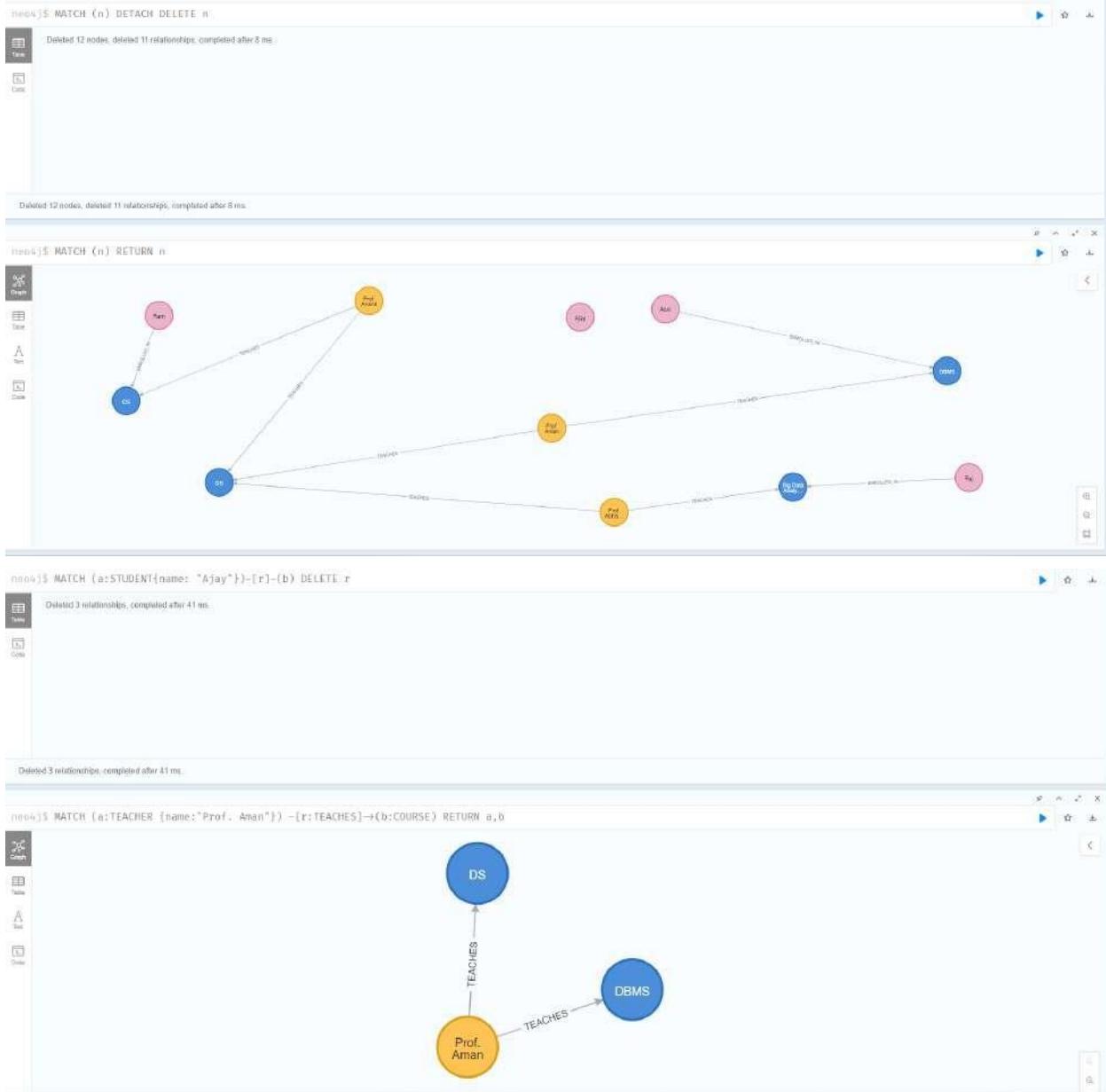
```
neo4j$ MATCH (a:TEACHER) , (b:COURSE) where a.name = "Prof. Abhishek" AND b.name = "Big Data Analysis" CREATE (a)-[r:TEACHES]-(b)
```

Created 1 relationship, completed after 10 ms.

Created 1 relationship, completed after 10 ms.

```
neo4j$ MATCH (a:STUDENT {name:"Ajay"}) -[r:ENROLLED_IN]-(b:COURSE) RETURN a,b
```





## EXPERIMENT-3

Perform the following DB operations using Cassandra.

- Create a keyspace by name Employee
- Create a column family by name Employee-Info with attributes  
Emp\_Id Primary Key, Emp\_Name,  
Designation, Date\_of\_Joining,  
Salary, Dept\_Name
- Insert the values into the table in batch
- Update Employee name and Department of Emp-Id 121
- Sort the details of Employee records based on salary
- Alter the schema of the table Employee\_Info to add a column Projects which stores a set of Projects done by the corresponding Employee.
- Update the altered table to add project names.
- Create a TTL of 15 seconds to display the values of Employees.

CODE WITH SCREENSHOTS:

```
Activities Terminal Apr 8 14:12
bmscse@bmscse-HP-Elite-Tower-800-C9-Desktop-PC:~$ cqlsh
Connected to test cluster at 127.0.0.1:9042
Cassandra 4.1.0 | CQL spec 3.4.6 | Native protocol v5
Use HELP for help.

cqlsh> CREATE KEYSPACE Employee
...   WITH replication = {'class': 'SimpleStrategy', 'replication_factor': 1};

cqlsh>
cqlsh> USE Employee;
cqlsh> employee> CREATE TABLE Employee_Info (
...     Emp_Id int PRIMARY KEY,
...     Emp_Name text,
...     Designation text,
...     Date_of_Joining date,
...     Salary double,
...     Dept_Name text
... );

cqlsh> employee> BEGIN BATCH
...     INSERT INTO Employee_Info (Emp_Id, Emp_Name, Designation, Date_of_Joining, Salary, Dept_Name)
...     VALUES (121, 'Alice', 'Manager', '2020-05-01', 75000, 'HR');
...     INSERT INTO Employee_Info (Emp_Id, Emp_Name, Designation, Date_of_Joining, Salary, Dept_Name)
...     VALUES (122, 'Bob', 'Developer', '2021-01-15', 60000, 'IT');
...     INSERT INTO Employee_Info (Emp_Id, Emp_Name, Designation, Date_of_Joining, Salary, Dept_Name)
...     VALUES (123, 'Charlie', 'Analyst', '2019-09-10', 50000, 'Finance');
...     APPLY BATCH;
cqlsh> employee> SELECT * FROM Employee_Info;
emp_id | date_of_joining | dept_name | designation | emp_name | salary
--- --- --- --- --- ---
123 | 2019-09-10 | Finance | Analyst | Charlie | 50000
122 | 2021-01-15 | IT | Developer | Bob | 60000
121 | 2020-05-01 | HR | Manager | Alice | 75000
(3 rows)
cqlsh> employee> UPDATE Employee_Info SET Emp_Name = 'Alicia', Dept_Name = 'Admin' WHERE Emp_Id = 121;
cqlsh> employee> SELECT * FROM Employee_Info;
emp_id | date_of_joining | dept_name | designation | emp_name | salary
--- --- --- --- --- ---
123 | 2019-09-10 | Finance | Analyst | Charlie | 50000
122 | 2021-01-15 | IT | Developer | Bob | 60000
121 | 2020-05-01 | Admin | Manager | Alicia | 75000
(3 rows)
cqlsh> employee> SELECT * FROM Employee_Info WHERE Salary >= 6 ALLOW FILTERING;
emp_id | date_of_joining | dept_name | designation | emp_name | salary
--- --- --- --- --- ---
123 | 2019-09-10 | Finance | Analyst | Charlie | 50000
122 | 2021-01-15 | IT | Developer | Bob | 60000
121 | 2020-05-01 | Admin | Manager | Alicia | 75000
***
```

```

Activities Terminal Apr 8 14:12
[1] bmscse@bmscse-HP-Elite-Tower-800-G9-Desktop-PC: ~
121 | 2020-05-01 | HR | Manager | Alice | 75000
(3 rows)
cqsh:employees> UPDATE Employee_Info SET Emp_Name = 'Alicia', Dept_Name = 'Admin' WHERE Emp_Id = 121;
cqsh:employees> SELECT*FROM Employee_Info;
+-----+-----+-----+-----+-----+
| emp_id | date_of_joining | dept_name | designation | emp_name | salary |
+-----+-----+-----+-----+-----+
| 123 | 2019-09-10 | Finance | Analyst | Charlie | 50000 |
| 122 | 2021-01-15 | IT | Developer | Bob | 60000 |
| 121 | 2020-05-01 | Admin | Manager | Alicia | 75000 |
+-----+-----+-----+-----+-----+
(3 rows)
cqsh:employees> SELECT * FROM Employee_Info WHERE Salary >= 0 ALLOW FILTERING;
+-----+-----+-----+-----+-----+
| emp_id | date_of_joining | dept_name | designation | emp_name | salary |
+-----+-----+-----+-----+-----+
| 123 | 2019-09-10 | Finance | Analyst | Charlie | 50000 |
| 122 | 2021-01-15 | IT | Developer | Bob | 60000 |
| 121 | 2020-05-01 | Admin | Manager | Alicia | 75000 |
+-----+-----+-----+-----+-----+
(3 rows)
cqsh:employees> ALTER TABLE Employee_Info ADD Projects set<text>;
cqsh:employees> SELECT*FROM Employee_Info;
+-----+-----+-----+-----+-----+-----+
| emp_id | date_of_joining | dept_name | designation | emp_name | projects | salary |
+-----+-----+-----+-----+-----+-----+
| 123 | 2019-09-10 | Finance | Analyst | Charlie | null | 50000 |
| 122 | 2021-01-15 | IT | Developer | Bob | null | 60000 |
| 121 | 2020-05-01 | Admin | Manager | Alicia | null | 75000 |
+-----+-----+-----+-----+-----+-----+
(3 rows)
cqsh:employees> UPDATE Employee_Info SET Projects = ('Payroll System', 'Attendance Tracker') WHERE Emp_Id = 121;
cqsh:employees> SELECT*FROM Employee_Info;
+-----+-----+-----+-----+-----+-----+-----+
| emp_id | date_of_joining | dept_name | designation | emp_name | projects | salary |
+-----+-----+-----+-----+-----+-----+-----+
| 123 | 2019-09-10 | Finance | Analyst | Charlie | null | 50000 |
| 122 | 2021-01-15 | IT | Developer | Bob | null | 60000 |
| 121 | 2020-05-01 | Admin | Manager | Alicia | ('Attendance Tracker', 'Payroll System') | 75000 |
+-----+-----+-----+-----+-----+-----+
(3 rows)
cqsh:employees> 

```

```

Activities Terminal Apr 8 14:24
[1] bmscse@bmscse-HP-Elite-Tower-800-G9-Desktop-PC: ~
bmscse@bmscse-HP-Elite-Tower-800-G9-Desktop-PC: $ cqsh
Connected to Test Cluster at 127.0.0.1:9042
[cqlsh 6.1.0 - Cassandra 4.1.8 | CQL spec 3.4.6 | Native protocol v5]
Use HELP for help.
cqlsh> CREATE KEYSPACE Library
... WITH replication = {'class': 'SimpleStrategy', 'replication_factor': 1};
cqlsh> USE Library;
cqlsh:library> CREATE TABLE Library_Info (
... Stud_Id int PRIMARY KEY,
... Counter_Value counter,
... Stud_Name text,
... Book_Name text,
... Book_Id int,
... Date_of_Issue date
... );
InvalidRequest: Error from server: code=2000 [Invalid query] Message="Cannot mix counter and non counter columns in the same table"
cqlsh:library> CREATE TABLE Library_Info (
... Stud_Id int PRIMARY KEY,
... Stud_Name text,
... Book_Name text,
... Book_Id int,
... Date_of_Issue date
... );
cqlsh:library> CREATE TABLE Library_Counter (
... Stud_Id int PRIMARY KEY,
... Counter_Value counter
... );
cqlsh:library> SELECT*FROM Library_Info;
+-----+-----+-----+-----+
| stud_id | book_id | book_name | date_of_issue | stud_name |
+-----+-----+-----+-----+
(0 rows)
cqlsh:library> CREATE TABLE Library_Info1(
... Stud_Id int,
... Book_Name text,
... Counter_Value counter,
... PRIMARY KEY (Stud_Id, Book_Name)
... );
cqlsh:library> DROP TABLE LIBRARY_INFO;
cqlsh:library> 
cqlsh:library> 
cqlsh:library> 
```

```

Activities Terminal Apr 8 14:25
bmscecse@bmscecse-HP-Slate-Tower-800-G9-Desktop-PC ~
...
    ... Counter_Value counter,
    ... PRIMARY KEY (Stud_Id, Book_Name)
    ... };

cqlsh:library> BEGIN BATCH
... UPDATE Library_Info SET Counter_Value = Counter_Value + 1 WHERE Stud_Id = 112 AND Book_Name = 'BDA';
... UPDATE Library_Info SET Counter_Value = Counter_Value + 1 WHERE Stud_Id = 113 AND Book_Name = 'DSA';
... UPDATE Library_Info SET Counter_Value = Counter_Value + 1 WHERE Stud_Id = 112 AND Book_Name = 'DSA';
... APPLY BATCH;
cqlsh:library> SELECT * FROM Library_Info;
stud_id | book_name | counter_value
-----+
113 | DSA | 1
112 | BDA | 2
(2 rows)
cqlsh:library> SELECT * FROM Library_Info WHERE Stud_Id = 112 AND Book_Name = 'BDA';
stud_id | book_name | counter_value
-----+
112 | BDA | 2
(1 rows)
cqlsh:library> COPY Library_Info TO '/home/your_username/library_info.csv' WITH HEADER = TRUE;
Using 16 child processes
Can't open '/home/your_username/library_info.csv' for writing: [Errno 2] No such file or directory: '/home/your_username/library_info.csv'
cqlsh:library> COPY library_info (stud_id, book_name, counter_value) TO '/home/bmscecse/library_info.csv' WITH HEADER = TRUE;
Using 16 child processes

Starting copy of library.library_info with columns (stud_id, book_name, counter_value).
Processed: 2 rows; Rate: 22 rows/s; Avg. rate: 22 rows/s
2 rows exported to 1 files in 0.105 seconds.
cqlsh:library> select*from library.library_info;
stud_id | book_name | counter_value
-----+
113 | DSA | 1
112 | BDA | 4
(2 rows)
cqlsh:library>

```

## EXPERIMENT:4

Perform the following DB operations using Cassandra.

- a) Create a keyspace by name Library
  - b) 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
  - c) Insert the values into the table in batch
  - d) Display the details of the table created and increase the value of the counter
  - e) Write a query to show that a student with id 112 has taken a book “BDA” 2 times.
  - f) Export the created column to a csv file
  - g) Import a given csv dataset from local file system into Cassandra column family

```

[1]+  Stopped                  qclsh
bmscece@bmscece-HP-Elite-Tower-800-C9-Desktop-PC:~$ BEGIN BATCH
INSERT INTO Students_Info (Roll_No, Studname, DateOfJoining, last_exam_Percent)
VALUES (1, 'Asha', '2012-03-12', 79.9);
INSERT INTO Students_Info (Roll_No, Studname, DateOfJoining, last_exam_Percent)
VALUES (2, 'Krian', '2012-03-12', 89.9);
INSERT INTO Students_Info (Roll_No, Studname, DateOfJoining, last_exam_Percent)
VALUES (3, 'Tarun', '2012-03-12', 78.9);
INSERT INTO Students_Info (Roll_No, Studname, DateOfJoining, last_exam_Percent)
VALUES (4, 'Sonrth', '2012-03-12', 90.9);
INSERT INTO Students_Info (Roll_No, Studname, DateOfJoining, last_exam_Percent)
VALUES (5, 'Silma', '2012-03-12', 67.9);
INSERT INTO Students_Info (Roll_No, Studname, DateOfJoining, last_exam_Percent)
VALUES (6, 'Rohan', '2012-03-12', 56.9);
APPLY BATCH;
[1]+  Stopped                  qclsh
bmscece@bmscece-HP-Elite-Tower-800-C9-Desktop-PC:~$ DESCRIBE KEYSPACES;
company    students    system_distributed    system_views
eng       system      system_schema        system_virtual_schemas
eng_data   system_auth  system_traces

qclsh:students> DESCRIBE TABLES;
qclsh:students> DESCRIBE TABLES;
qclsh:students> CREATE TABLE Students_Info (
...     Roll_No int PRIMARY KEY,
...     Studname text,
...     DateOfJoining timestamp,
...     last_exam_Percent double
... );
qclsh:students> BEGIN BATCH
...     INSERT INTO Students_Info (Roll_No, Studname, DateOfJoining, last_exam_Percent)
...     VALUES (1, 'Asha', '2012-03-12', 79.9);
...     INSERT INTO Students_Info (Roll_No, Studname, DateOfJoining, last_exam_Percent)
...     VALUES (2, 'Krian', '2012-03-12', 89.9);
...     INSERT INTO Students_Info (Roll_No, Studname, DateOfJoining, last_exam_Percent)
...     VALUES (3, 'Tarun', '2012-03-12', 78.9);
...     INSERT INTO Students_Info (Roll_No, Studname, DateOfJoining, last_exam_Percent)
...     VALUES (4, 'Sonrth', '2012-03-12', 90.9);
...     INSERT INTO Students_Info (Roll_No, Studname, DateOfJoining, last_exam_Percent)
...     VALUES (5, 'Silma', '2012-03-12', 67.9);
...     INSERT INTO Students_Info (Roll_No, Studname, DateOfJoining, last_exam_Percent)
...     VALUES (6, 'Rohan', '2012-03-12', 56.9);
APPLY BATCH;
[1]+  Stopped                  qclsh
bmscece@bmscece-HP-Elite-Tower-800-C9-Desktop-PC:~$ DESCRIBE KEYSERVERS;
Datacenter: datacenter1
=====
Status=Up/Down
-- State:Normal/Leaving/Joining/Moving
-- Address      Load      Tokens  Owns    Host ID  Rack
UN 127.0.0.1  360.28  K16  16      ?          rack1

Note: Non-system keyspaces don't have the same replication settings, effective ownership information is meaningless
bmscece@bmscece-HP-Elite-Tower-800-C9-Desktop-PC:~$ nodetool status
Connected to Test Cluster at 127.0.0.1:9642
[cqlsh 6.1.0 | Cassandra 4.1.8 | CQL spec 3.4.0 | Native protocol v5]
Use HELP for help.
cqlsh> CREATE KEYSPACE Students WITH REPLICATION =
...  {'class':'SimpleStrategy','replication_factor':1};

cqlsh>
cqlsh>
cqlsh>
```

```

[1] es Terminal Apr 1 14:26 bmscece@bmscece-HP-Elite-Tower-800-C9-Desktop-PC:-
cqsh:students> BEGIN BATCH
... INSERT INTO Students_Info (Roll_No, StudName, DateOfJoining, last_exam_Percent)
... VALUES (1, 'Asha', '2012-03-12', 79.9);
...
... INSERT INTO Students_Info (Roll_No, StudName, DateOfJoining, last_exam_Percent)
... VALUES (2, 'Kiran', '2012-03-12', 89.9);
...
... INSERT INTO Students_Info (Roll_No, StudName, DateOfJoining, last_exam_Percent)
... VALUES (3, 'Tarun', '2012-03-12', 78.9);
...
... INSERT INTO Students_Info (Roll_No, StudName, DateOfJoining, last_exam_Percent)
... VALUES (4, 'Santh', '2012-03-12', 89.9);
...
... INSERT INTO Students_Info (Roll_No, StudName, DateOfJoining, last_exam_Percent)
... VALUES (5, 'Sinha', '2012-03-12', 67.9);
...
... INSERT INTO Students_Info (Roll_No, StudName, DateOfJoining, last_exam_Percent)
... VALUES (6, 'Rohan', '2012-03-12', 56.9);
... APPLY BATCH;
cqsh:students> SELECT*FROM students_Info;
roll_no | dateofjoining | last_exam_percent | studname
-----+-----+-----+-----
  5 | 2012-03-11 18:30:00.000000+0000 |    67.9 | Sinha
  1 | 2012-03-11 18:30:00.000000+0000 |    79.9 | Asha
  2 | 2012-03-11 18:30:00.000000+0000 |    89.9 | Kiran
  4 | 2012-03-11 18:30:00.000000+0000 |    89.9 | Santh
  6 | 2012-03-11 18:30:00.000000+0000 |    56.9 | Rohan
  3 | 2012-03-11 18:30:00.000000+0000 |    78.9 | Tarun

(6 rows)
cqsh:students> SELECT*FROM student_info WHERE roll_no IN(1,2,3);
InvalidRequest: Error from server: code=2200 [Invalid query] message="table student_info does not exist"
cqsh:students> SELECT*FROM student_info WHERE roll_no IN(1,2,3);
InvalidRequest: Error from server: code=2200 [Invalid query] message="table student_info does not exist"
cqsh:students> SELECT*FROM students_Info WHERE roll_no IN(1,2,3);
roll_no | dateofjoining | last_exam_percent | studname
-----+-----+-----+-----
  1 | 2012-03-11 18:30:00.000000+0000 |    79.9 | Asha
  2 | 2012-03-11 18:30:00.000000+0000 |    89.9 | Kiran
  3 | 2012-03-11 18:30:00.000000+0000 |    78.9 | Tarun

(3 rows)
cqsh:students> SELECT*FROM students_info WHERE studname='Asha';
SyntaxException: line 1:47 no viable alternative at input '1' ((SELECT*FROM students_Info WHERE studname='Asha)a);
cqsh:students> SELECT * FROM students_Info WHERE studname = 'Asha';
InvalidRequest: Error from server: code=2200 [Invalid query] message="Cannot execute this query as it might involve data filtering and thus may have unpredictable performance. If you want to execute this query despite the performance unpredictability, use ALLOW FILTERING"
cqsh:students> SELECT * FROM students_Info WHERE studname = 'Asha' ALLOW FILTERING;
roll_no | dateofjoining | last_exam_percent | studname
-----+-----+-----+-----
  1 | 2012-03-11 18:30:00.000000+0000 |    79.9 | Asha
(1 rows)

[2] es Terminal Apr 1 14:27 bmscece@bmscece-HP-Elite-Tower-800-C9-Desktop-PC:-
cqsh:students> select roll_no as "USN" from students_Info;
USN
-----
1
4
2
4
6
3

(6 rows)
cqsh:students> UPDATE roll_no as "USN" From students_Info;
SyntaxException: line 1:15 mismatched input 'as' expecting NAME_SET (UPDATE roll_no [as] "US...")
cqsh:students> ALTER TABLE students_Info RENAME roll_no TO USN;
cqsh:students> select*from students_Info;
usn | dateofjoining | last_exam_percent | studname
-----+-----+-----+-----
  5 | 2012-03-11 18:30:00.000000+0000 |    67.9 | Sinha
  1 | 2012-03-11 18:30:00.000000+0000 |    79.9 | Asha
  2 | 2012-03-11 18:30:00.000000+0000 |    89.9 | Kiran
  4 | 2012-03-11 18:30:00.000000+0000 |    89.9 | Santh
  6 | 2012-03-11 18:30:00.000000+0000 |    56.9 | Rohan
  3 | 2012-03-11 18:30:00.000000+0000 |    78.9 | Tarun

(6 rows)
cqsh:students> UPDATE students_Info SET studname='David Sheen' WHERE usn=2;
cqsh:students> select*from students_Info;
usn | dateofjoining | last_exam_percent | studname
-----+-----+-----+-----
  4 | 2012-03-11 18:30:00.000000+0000 |    67.9 | Sinha
  1 | 2012-03-11 18:30:00.000000+0000 |    79.9 | Asha
  2 | 2012-03-11 18:30:00.000000+0000 |    89.9 | David Sheen
  4 | 2012-03-11 18:30:00.000000+0000 |    90.9 | Santh
  6 | 2012-03-11 18:30:00.000000+0000 |    56.9 | Rohan
  3 | 2012-03-11 18:30:00.000000+0000 |    78.9 | Tarun

(6 rows)
cqsh:students> UPDATE students_Info SET rollno=6 WHERE rollno=3;
InvalidRequest: Error from server: code=2200 [Invalid query] message="Undefined column name rollno in table students.students_Info"
cqsh:students> cqsh:students> UPDATE students_Info SET usn=0 WHERE usn=3;
[2]+ Stopped                  cqsh
bmscece@bmscece-HP-Elite-Tower-800-C9-Desktop-PC:~$ cqsh
Connected to Cassandra Cluster at 127.0.0.1:9042
CQLsh 6.1.0 | Cassandra 4.1.10 | CQL spec 3.4.6 | Native protocol v5
Use HELP for help.
cqsh> use students
...
...
...
fail > stopped cqsh

```

```

les Terminal Apr 1 14:28
bmscse@bmscse-HP-Elite-Tower-800-C9-Desktop-PC:~ [3]+ Stopped cqlsh
bmscse@bmscse-HP-Elite-Tower-800-C9-Desktop-PC:~ $ cqlsh
Connected to Test Cluster at 127.0.0.1:9642
[cqlsh 6.1.0 | Cassandra 4.1.0 | CQL spec 3.4.6 | Native protocol v5]
Use HELP for help.
[cqlsh] USE students;
[cqlsh:students]> SELECT * FROM students_info;
+-----+-----+-----+
| usn | dateofjoining | last_exam_percent | studname |
+-----+-----+-----+
| 5 | 2012-03-11 18:30:00.000000+0000 | 67.9 | Srinha |
| 1 | 2012-03-11 18:30:00.000000+0000 | 79.9 | Asha |
| 2 | 2012-03-11 18:30:00.000000+0000 | null | David Sheen |
| 3 | 2012-03-11 18:30:00.000000+0000 | 90.9 | Sanarth |
| 4 | 2012-03-11 18:30:00.000000+0000 | 56.9 | Rohan |
| 6 | 2012-03-11 18:30:00.000000+0000 | 78.9 | Tarun |
+-----+
(6 rows)
[cqlsh:students]> UPDATE students_info SET usn=6 WHERE usn=3;
InvalidRequest: Error from server: code=200 [Invalid query] message="PRIMARY KEY part usn found in SET part"
[cqlsh:students]> UPDATE students_info SET usn = 6 WHERE usn = 3;
InvalidRequest: Error from server: code=200 [Invalid query] message="PRIMARY KEY part usn found in SET part"
[cqlsh:students]> UPDATE students_info SET usn = 6 WHERE usn = 3;
InvalidRequest: Error from server: code=200 [Invalid query] message="PRIMARY KEY part usn found in SET part"
[cqlsh:students]> DELETE FROM students_info WHERE usn = 3;
[cqlsh:students]> INSERT INTO students_info (usn, dateofjoining, last_exam_percent, studname)
    ... VALUES (6, '2012-03-11 18:30:00', 78.9, 'Tarun');
[cqlsh:students]> select*from STUDENTS_INFO;
+-----+-----+-----+
| usn | dateofjoining | last_exam_percent | studname |
+-----+-----+-----+
| 5 | 2012-03-11 18:30:00.000000+0000 | 67.9 | Srinha |
| 1 | 2012-03-11 18:30:00.000000+0000 | 79.9 | Asha |
| 2 | 2012-03-11 18:30:00.000000+0000 | null | David Sheen |
| 3 | 2012-03-11 18:30:00.000000+0000 | 90.9 | Sanarth |
| 6 | 2012-03-11 18:30:00.000000+0000 | 78.9 | Tarun |
+-----+
(5 rows)
[cqlsh:students]> DELETE last_exam_percent FROM students_info WHERE usn=2;
[cqlsh:students]> select*from students_info;
+-----+-----+-----+
| usn | dateofjoining | last_exam_percent | studname |
+-----+-----+-----+
| 5 | 2012-03-11 18:30:00.000000+0000 | 67.9 | Srinha |
| 1 | 2012-03-11 18:30:00.000000+0000 | 79.9 | Asha |
| 2 | 2012-03-11 18:30:00.000000+0000 | null | David Sheen |
| 3 | 2012-03-11 18:30:00.000000+0000 | 90.9 | Sanarth |
| 6 | 2012-03-11 18:30:00.000000+0000 | 78.9 | Tarun |
+-----+
(5 rows)
[cqlsh:students]> DELETE FROM students_info WHERE usn=2;
[cqlsh:students]> SELECT*FROM STUDENTS_INFO;
+-----+-----+-----+
| usn | dateofjoining | last_exam_percent | studname |
+-----+-----+-----+
| 5 | 2012-03-11 18:30:00.000000+0000 | 67.9 | Srinha |
| 1 | 2012-03-11 18:30:00.000000+0000 | 79.9 | Asha |
| 2 | 2012-03-11 18:30:00.000000+0000 | null | David Sheen |
| 3 | 2012-03-11 18:30:00.000000+0000 | 90.9 | Sanarth |
| 6 | 2012-03-11 18:30:00.000000+0000 | 78.9 | Tarun |
+-----+
(5 rows)
les Terminal Apr 1 14:29
bmscse@bmscse-HP-Elite-Tower-800-C9-Desktop-PC:~ [3]+ Stopped cqlsh
bmscse@bmscse-HP-Elite-Tower-800-C9-Desktop-PC:~ $ cqlsh
Connected to Test Cluster at 127.0.0.1:9642
[cqlsh 6.1.0 | Cassandra 4.1.0 | CQL spec 3.4.6 | Native protocol v5]
Use HELP for help.
[cqlsh] USE students;
[cqlsh:students]> select*from students_info;
+-----+-----+-----+
| usn | dateofjoining | last_exam_percent | studname |
+-----+-----+-----+
| 5 | 2012-03-11 18:30:00.000000+0000 | 67.9 | Srinha |
| 1 | 2012-03-11 18:30:00.000000+0000 | 79.9 | Asha |
| 2 | 2012-03-11 18:30:00.000000+0000 | null | David Sheen |
| 3 | 2012-03-11 18:30:00.000000+0000 | 90.9 | Sanarth |
| 6 | 2012-03-11 18:30:00.000000+0000 | 78.9 | Tarun |
+-----+
(5 rows)
[cqlsh:students]> DELETE FROM students_info WHERE usn=2;
[cqlsh:students]> SELECT*FROM STUDENTS_INFO;
+-----+-----+-----+
| usn | dateofjoining | last_exam_percent | studname |
+-----+-----+-----+
| 5 | 2012-03-11 18:30:00.000000+0000 | 67.9 | Srinha |
| 1 | 2012-03-11 18:30:00.000000+0000 | 79.9 | Asha |
| 3 | 2012-03-11 18:30:00.000000+0000 | null | David Sheen |
| 4 | 2012-03-11 18:30:00.000000+0000 | 90.9 | Sanarth |
| 6 | 2012-03-11 18:30:00.000000+0000 | 78.9 | Tarun |
+-----+
(4 rows)
[cqlsh:students]> ALTER TABLE STUDENTS_INFO ADD hobbies set<text>;
...
...
Statements are terminated with a ';' . You can press CTRL-C to cancel an incomplete statement.
[cqlsh:students]> ALTER TABLE STUDENTS_INFO ADD hobbies set<text>;
...
[cqlsh:students]> ALTER TABLE STUDENTS_INFO ADD hobbies set<text>;
[cqlsh:students]> select*from students_info;
+-----+-----+-----+
| usn | dateofjoining | hobbies | last_exam_percent | studname |
+-----+-----+-----+
| 5 | 2012-03-11 18:30:00.000000+0000 | null | 67.9 | Srinha |
| 1 | 2012-03-11 18:30:00.000000+0000 | null | 79.9 | Asha |
| 3 | 2012-03-11 18:30:00.000000+0000 | null | 90.9 | Sanarth |
| 6 | 2012-03-11 18:30:00.000000+0000 | null | 78.9 | Tarun |
+-----+
(4 rows)
[cqlsh:students]> ALTER TABLE students_info ADD language list<text>;
[cqlsh:students]> select*from students_info;
+-----+-----+-----+-----+
| usn | dateofjoining | hobbies | language | last_exam_percent | studname |
+-----+-----+-----+-----+
| 5 | 2012-03-11 18:30:00.000000+0000 | null | null | 67.9 | Srinha |
| 1 | 2012-03-11 18:30:00.000000+0000 | null | null | 79.9 | Asha |
| 3 | 2012-03-11 18:30:00.000000+0000 | null | null | 90.9 | Sanarth |
| 6 | 2012-03-11 18:30:00.000000+0000 | null | null | 78.9 | Tarun |
+-----+
(4 rows)

```

```

les Terminal April 1 14:29 bmscse@bmscse-HP-Elite-Tower-800-G9-Desktop-PC:-
cqsh:students>
cqsh:students> select*from students_info;
+-----+-----+-----+-----+
| usn | dateofjoining | hobbies | last_exam_percent | studname |
+-----+-----+-----+-----+
| 5 | 2012-03-11 16:30:00.000000+0000 | null | 67.9 | Simha |
| 2 | 2012-03-11 16:30:00.000000+0000 | null | 79.9 | Asha |
| 4 | 2012-03-11 16:30:00.000000+0000 | null | 90.9 | Sanarth |
| 6 | 2012-03-11 13:00:00.000000+0000 | null | 78.9 | Tarun |
+-----+-----+-----+-----+
(4 rows)
cqsh:students> ALTER TABLE students_info ADD language list<text>;
cqsh:students> select*from students_info;
+-----+-----+-----+-----+-----+
| usn | dateofjoining | hobbies | language | last_exam_percent | studname |
+-----+-----+-----+-----+-----+
| 5 | 2012-03-11 16:30:00.000000+0000 | null | null | 67.9 | Simha |
| 2 | 2012-03-11 16:30:00.000000+0000 | null | null | 79.9 | Asha |
| 4 | 2012-03-11 16:30:00.000000+0000 | null | null | 90.9 | Sanarth |
| 6 | 2012-03-11 13:00:00.000000+0000 | null | null | 78.9 | Tarun |
+-----+-----+-----+-----+
(4 rows)
cqsh:students> UPDATE students_info SET hobbies=hobbies||'Chess,Table Tennis' where usn=1;
cqsh:students> SELECT*FROM STUDENTS_INFO;
+-----+-----+-----+-----+-----+
| usn | dateofjoining | hobbies | language | last_exam_percent | studname |
+-----+-----+-----+-----+-----+
| 5 | 2012-03-11 16:30:00.000000+0000 | null | null | 67.9 | Simha |
| 2 | 2012-03-11 16:30:00.000000+0000 | ('Chess,Table Tennis') | null | 79.9 | Asha |
| 4 | 2012-03-11 16:30:00.000000+0000 | null | null | 90.9 | Sanarth |
| 6 | 2012-03-11 13:00:00.000000+0000 | null | null | 78.9 | Tarun |
+-----+-----+-----+-----+
(4 rows)
cqsh:students> SELECT*FROM student_info where usn=1;
InvalidRequest: Error from server: code=2200 [Invalid query] message="Table student_info does not exist"
cqsh:students> select*from students_info where usn=1;
+-----+-----+-----+-----+
| usn | dateofjoining | hobbies | language | last_exam_percent | studname |
+-----+-----+-----+-----+
| 1 | 2012-03-11 16:30:00.000000+0000 | ('Chess,Table Tennis') | null | 79.9 | Asha |
+-----+-----+-----+-----+
(1 rows)
cqsh:students> update students_info set language=language||'Hindi,English' where usn=1;
cqsh:students> select*from students_info;
+-----+-----+-----+-----+-----+
| usn | dateofjoining | hobbies | language | last_exam_percent | studname |
+-----+-----+-----+-----+-----+
| 5 | 2012-03-11 16:30:00.000000+0000 | null | ('Hindi,English') | 67.9 | Simha |
| 2 | 2012-03-11 16:30:00.000000+0000 | ('Chess,Table Tennis') | ('Hindi,English') | 79.9 | Asha |
| 4 | 2012-03-11 16:30:00.000000+0000 | null | null | 90.9 | Sanarth |
| 6 | 2012-03-11 13:00:00.000000+0000 | null | null | 78.9 | Tarun |
+-----+-----+-----+-----+
(4 rows)
cqsh:students> 

```

```

Activities Terminal April 1 14:40 bmscse@bmscse-HP-Elite-Tower-800-G9-Desktop-PC:-
cqsh:students>
cqsh:students> CREATE TABLE library_book (
...     counter_value integer,
...     book_name varchar,
...     stud_name varchar,
...     PRIMARY KEY (book_name, stud_name)
... );
cqsh:students> SELECT*FROM LIBRARY_BOOK;
+-----+-----+-----+
| book_name | stud_name | counter_value |
+-----+-----+-----+
(0 rows)
cqsh:students> BEGIN BATCH
...     Inserting 4 records into library_book table
...     INSERT INTO library_book (counter_value, book_name, stud_name)
...     VALUES (0, 'Big Data Analytics', 'Jeet');
...
...     INSERT INTO library_book (counter_value, book_name, stud_name)
...     VALUES (0, 'Machine Learning', 'Amit');
...
...     INSERT INTO library_book (counter_value, book_name, stud_name)
...     VALUES (0, 'Data Science Essentials', 'Ravi');
...
...     INSERT INTO library_book (counter_value, book_name, stud_name)
...     VALUES (0, 'Cloud Computing', 'Prity');
...
...     APPLY BATCH;
InvalidRequest: Error from server: code=2200 [Invalid query] message="INSERT statements are not allowed on counter tables, use UPDATE instead"
cqsh:students> BEGIN BATCH
...     Inserting 4 records into library_book table with counter initialized to 0
...     INSERT INTO library_book (counter_value, book_name, stud_name)
...     VALUES (0, 'Big Data Analytics', 'Jeet');
...
...     INSERT INTO library_book (counter_value, book_name, stud_name)
...     VALUES (0, 'Machine Learning', 'Amit');
...
...     INSERT INTO library_book (counter_value, book_name, stud_name)
...     VALUES (0, 'Data Science Essentials', 'Ravi');
...
...     INSERT INTO library_book (counter_value, book_name, stud_name)
...     VALUES (0, 'Cloud Computing', 'Prity');
...
...     APPLY BATCH;
InvalidRequest: Error from server: code=2200 [Invalid query] message="INSERT statements are not allowed on counter tables, use UPDATE instead"
cqsh:students> UPDATE library_book
...     SET counter_value = counter_value + 1
...     WHERE book_name = 'Big Data Analytics'
...     AND stud_name = 'jeet';
cqsh:students> SELECT*FROM library_book;
+-----+-----+-----+
| book_name | stud_name | counter_value |
+-----+-----+-----+
| Big Data Analytics | jeet | 1 |
+-----+-----+-----+
(1 rows)
cqsh:students> 

```

```

Activities Terminal Apr 1 14:41 bmscsece@bmscsece-HP-Elite-Tower-800-G9-Desktop:~ 
... INSERT INTO library_book (counter_value, book_name, stud_name)
... VALUES (0, 'Data Science Essentials', 'Ravi');
...
... INSERT INTO library_book (counter_value, book_name, stud_name)
... VALUES (0, 'Cloud Computing', 'Priya');
...
APPLY BATCH;
InvalidRequest: Error from server: code=2200 [Invalid query] message="INSERT statements are not allowed on counter tables; use UPDATE instead"
cqish:students> -- Load data into the counter column
cqish:students> UPDATE library_book
... SET counter_value = counter_value + 1
... WHERE book_name = 'Big data Analytics'
... AND stud_name = 'Jeth';
cqish:students> SELECT*FROM library_book;
book_name | stud_name | counter_value
-----+-----+-----
Big data Analytics | Jeth |      1
(1 rows)
cqish:students> -- TIME TO LIVE (TTL) Example
cqish:students> CREATE TABLE userlogin (
...   userId int PRIMARY KEY,
...   password text
... );
cqish:students>
cqish:students>
cqish:students> SELECT*FROM userlogin;
userId | password
-----+-----
1 |    luffy
(1 rows)
cqish:students> INSERT INTO userlogin(userId, password)
... VALUES (1, 'luffy')
... USING TTL 30;
cqish:students> SELECT*FROM userlogin;
userId | password
-----+-----
1 |    luffy
(1 rows)
cqish:students> SELECT TTL(password)
... FROM userlogin
... WHERE userId = 1;
ttl(password)
-----+-----
(0 rows)
cqish:students>

```

```
Activities Terminal Apr 1 14:49 bmscse@bmscse-HP-Elite-Tower-800-G9-Desktop-PC ~
CREATE TABLE students.elearninglists (
    id uuid PRIMARY KEY,
    course_id uuid,
    course_order int,
    courseowner text,
    title text
) WITH additional.write_policy = '999'
AND bloom_filter_fp_chance = 0.01
AND caching = ['ALL', 'rows_per_partition': 'NONE']
AND cdc = false
AND comment = ''
AND compaction = {'class': 'org.apache.cassandra.db.compaction.SizeTieredCompactionStrategy', 'max_threshold': '32', 'min_threshold': '4'}
AND durable_writes = ['chunk_length_in_kb': '16', 'class': 'org.apache.cassandra.io.compress.LZ4Compressor']
AND memtable = 'default'
AND crc_check_chance = 1.0
AND default_time_to_live = 0
AND extensions = []
AND gc_grace_seconds = 864000
AND max_index_interval = 1000000
AND maxstable.flush_period_in_ms = 0
AND min_index_interval = 128
AND read_repair = 'BLOCKING'
AND speculative_retry = '999';
cqsh:students> COPY elearninglists(id, course_order, course_id, courseowner, title)
          TO '/home/bmscse/Desktop/elearninglists.csv' WITH HEADER = TRUE;
Using 16 child processes
Can't open 'd:/elearninglists.csv' for writing: [Errno 2] No such file or directory: 'd:/elearninglists.csv'
cqsh:students> COPY elearninglists(id, course_order, course_id, courseowner, title)
          TO '/home/bmscse/Desktop/elearninglists.csv' WITH HEADER = TRUE;
Using 16 child processes
Starting copy of students.elearninglists with columns [id, course_order, course_id, courseowner, title].
Processed: 0 rows; Rate: 0 rows/s; Avg. rate: 0 rows/s
0 rows exported to 1 files in 0.133 seconds.
cqsh:students> BEGIN BATCH
...     INSERT INTO elearninglists (id, course_order, course_id, courseowner, title)
...       VALUES (uuid(), 1, uuid(), 'John Doe', 'Introduction to Python');
...
...     INSERT INTO elearninglists (id, course_order, course_id, courseowner, title)
...       VALUES (uuid(), 2, uuid(), 'Jane Smith', 'Machine Learning Basics');
...
...     INSERT INTO elearninglists (id, course_order, course_id, courseowner, title)
...       VALUES (uuid(), 3, uuid(), 'Alice Johnson', 'Data Structures and Algorithms');
...   APPLY BATCH;
cqsh:students> select*from elearninglists;
+-----+-----+-----+-----+
| id   | course_id | course_order | courseowner | title           |
+-----+-----+-----+-----+
| fccb9530-c3c6-4619-b26a-7f1f14675290 | 10cded27-2d81-4481-ac73-d156a2b01a0f | 3 | Alice Johnson | Data Structures and Algorithms
| f9892ac5-a6e2-4156-a996-2f62e0f77962 | 5b588bf5-5bb5-4726-9460-786a98838883 | 2 | Jane Smith   | Machine Learning Basics
| 5ca576d-0253-440b-b137-003751ff31f1 | 5d5c2b42-97ef-483d-97ba-3e201ef7b36 | 1 | John Doe    | Introduction to Python
+-----+-----+-----+-----+
(3 rows)
cqsh:students> COPY elearninglists(id, course_order, course_id, courseowner, title)
          TO '/home/bmscse/Desktop/elearninglists.csv' WITH HEADER = TRUE;
Activities Terminal Apr 1 14:50 bmscse@bmscse-HP-Elite-Tower-800-G9-Desktop-PC ~
cqsh:students> COPY elearninglists(id, course_order, course_id, courseowner, title)
          ... TO 'd:/elearninglists.csv' WITH HEADER = TRUE;
Using 16 child processes
Can't open 'd:/elearninglists.csv' for writing: [Errno 2] No such file or directory: 'd:/elearninglists.csv'
cqsh:students> COPY elearninglists(id, course_order, course_id, courseowner, title)
          ... TO '/home/bmscse/Desktop/elearninglists.csv' WITH HEADER = TRUE;
Using 16 child processes
Starting copy of students.elearninglists with columns [id, course_order, course_id, courseowner, title].
Processed: 0 rows; Rate: 0 rows/s; Avg. rate: 0 rows/s
0 rows exported to 1 files in 0.133 seconds.
cqsh:students> BEGIN BATCH
...     INSERT INTO elearninglists (id, course_order, course_id, courseowner, title)
...       VALUES (uuid(), 1, uuid(), 'John Doe', 'Introduction to Python');
...
...     INSERT INTO elearninglists (id, course_order, course_id, courseowner, title)
...       VALUES (uuid(), 2, uuid(), 'Jane Smith', 'Machine Learning Basics');
...
...     INSERT INTO elearninglists (id, course_order, course_id, courseowner, title)
...       VALUES (uuid(), 3, uuid(), 'Alice Johnson', 'Data Structures and Algorithms');
...   APPLY BATCH;
cqsh:students> select*from elearninglists;
+-----+-----+-----+-----+
| id   | course_id | course_order | courseowner | title           |
+-----+-----+-----+-----+
| fccb9530-c3c6-4619-b26a-7f1f14675290 | 10cded27-2d81-4481-ac73-d156a2b01a0f | 3 | Alice Johnson | Data Structures and Algorithms
| f9892ac5-a6e2-4156-a996-2f62e0f77962 | 5b588bf5-5bb5-4726-9460-786a98838883 | 2 | Jane Smith   | Machine Learning Basics
| 5ca576d-0253-440b-b137-003751ff31f1 | 5d5c2b42-97ef-483d-97ba-3e201ef7b36 | 1 | John Doe    | Introduction to Python
+-----+-----+-----+-----+
(3 rows)
cqsh:students> COPY elearninglists(id, course_order, course_id, courseowner, title)
          ... TO '/home/bmscse/Desktop/elearninglists.csv' WITH HEADER = TRUE;
Using 16 child processes
Starting copy of students.elearninglists with columns [id, course_order, course_id, courseowner, title].
Processed: 3 rows; Rate: 54 rows/s; Avg. rate: 54 rows/s
3 rows imported from 1 files in 0.870 seconds.
cqsh:students> COPY elearninglists(id, course_order, course_id, courseowner, title)
          ... FROM '/home/bmscse/Desktop/elearninglists.csv' WITH HEADER = TRUE;
Using 16 child processes
Starting copy of students.elearninglists with columns [id, course_order, course_id, courseowner, title].
Processed: 3 rows; Rate: 54 rows/s; Avg. rate: 54 rows/s
3 rows imported from 1 files in 0.452 seconds (0 skipped).
cqsh:students> cat persons.csv | cqsh -e "COPY persons(id, fname, lname) FROM STDIN WITH HEADER = TRUE;">
Invalid syntax at char 17
  cat persons.csv | cqsh -e "COPY persons(id, fname, lname) FROM STDIN WITH HEADER = TRUE;">
cqsh:students> COPY elearninglists(id, course_order, course_id, courseowner, title)
          ... TO STDIN WITH HEADER = TRUE;
id, course_order, course_id, courseowner, title
f9892ac5-a6e2-4156-a996-2f62e0f77962, 2,b588bf5-5bb5-4726-9460-786a98838883,Jane Smith,Machine Learning Basics
5ca576d-0253-440b-b137-003751ff31f1, 1,5d5c2b42-97ef-483d-33e201ef7b36,John Doe,Introduction to Python
f0c09530-c3c6-4619-b26a-7f1f14675290, 3,10cded27-2d81-4481-ac73-d156a2b01a0f,Alice Johnson,Data Structures and Algorithms
cqsh:students>
```

## **EXPERIMENT-5**

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

```

hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -cat /abc/file2.txt
ht how are you
how is your job
how is your family
how is your brother
how is your sister

hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -get /abc/WC.txt /home/hduser/Downloads/WC.txt
get: '/home/hduser/Downloads/WC.txt': No such file or directory
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ ls -l /home/hduser/Downloads/WC.txt
ls: cannot access '/home/hduser/Downloads/WC.txt': No such file or directory
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ mkdir -p /home/hduser/Downloads/
mkdir: cannot create directory '/home/hduser': Permission denied
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ sudo mkdir -p /home/hduser/Downloads/
[sudo] password for hadoop:
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ ls /home/hduser/
[1] 11886 pts/0 0:00 bash

A hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -get /abc/WC.txt /home/hduser/Downloads/WC.txt
get: '/abc/WC.txt': No such file or directory
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -get /abc/WC.txt /home/hduser/Downloads/WC.txt
get: '/abc/WC.txt': No such file or directory
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -ls /abc/
Found 2 items
-rw-r--r-- 1 hadoop supergroup 99 2025-04-15 14:34 /abc/file1.txt
-rw-r--r-- 1 hadoop supergroup 99 2025-04-15 14:35 /abc/file2.txt
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -get /abc/file1.txt /home/hduser/Downloads/file1.txt
get: '/home/hduser/Downloads/file1.txt': COPYNG (Permission denied)
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ ls -l /home/hduser/Downloads/
drwxr-xr-x 2 hadoop hadoop 4096 Apr 15 14:37 /home/hduser/Downloads/
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ sudo hadoop fs -rm /home/hduser/Downloads/
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ ls -l /home/hduser/Downloads/
drwxr-xr-x 2 hadoop hadoop 4096 Apr 15 14:37 /home/hduser/Downloads/
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -get /abc/file1.txt /home/hduser/Downloads/file1.txt
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -getmerge /abc/WC.txt /abc/WC2.txt /home/hduser/Desktop/Merge.txt
getmerge: '/abc/WC.txt': No such file or directory
getmerge: '/abc/WC2.txt': No such file or directory
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -getmerge /abc/file1.txt /abc/file2.txt /home/hduser/Desktop/Merge.txt
getmerge: HDFS failed to create file /home/hduser/Desktop/Merge (exists=false, cwd=file:///home/hduser/Desktop/sbin)
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ sudo mkdir -p /home/hduser/Desktop
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ sudo chmod -R hadoop:hadoop /home/hduser/Desktop
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -getmerge /abc/file1.txt /abc/file2.txt /home/hduser/Desktop/Merge.txt
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ cat /home/hduser/Desktop/Merge.txt
M. how are you
ht how are you
how is your job
how is your family
how is your brother
how is your sister

ht how are you
ht how are you
how is your job
how is your family
how is your brother
how is your sister

hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -cat /abc/file2.txt
ht how are you
how is your job
how is your family
how is your brother
how is your sister

hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -copyToLocal /abc/WC.txt /home/hduser/Desktop
copyToLocal: '/abc/WC.txt': No such file or directory
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -copyToLocal /abc/file1.txt /home/hduser/Desktop
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ cat /home/hduser/Desktop/file1.txt
ht how are you
ht how are you
how is your job
how is your family
how is your brother
how is your sister

hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -cat /abc/file1.txt
ht how are you
ht how are you
how is your job
how is your family
how is your brother
how is your sister

hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -cat /abc/File2.txt
ht how are you
how is your job
how is your family
how is your brother
how is your sister

hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -mv /abc /FFF
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -ls /FFF
FileStatus[0]
-rw-r--r-- 1 hadoop supergroup 98 2025-04-15 14:34 /FFF/file1.txt
-rw-r--r-- 1 hadoop supergroup 98 2025-04-15 14:35 /FFF/file2.txt
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ 9. Cp Hadoop HDFS
cp Command Usage cp Hadoop HDFS
cp Command Example
hadoop fs -cp /CSE/ /LLL
hadoop fs -ls /LLL

The cp command copies a file from one directory to another directory within the HDFS.
9.: command not found
cp: target '/CSE/' is not a directory
cp: target '/LLL': No such file or directory
cp: '/CSE/': No such file or directory
ls: '/LLL': No such file or directory
Command 'The' not found, did you mean:
  command 'the' from deb node-he (1:2.0-3)
  command 'the' from deb node-he (1:2.0-3)
  command 'the' from deb node-he (1:2.0-3)
  command 'the' from deb node-he (1:2.0-3)
Try: sudo apt install deb-node-he
HDFS: command not found
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hadoop fs -cp /abc/file1.txt /newDir/
cp: '/newDir/': No such file or directory: 'hdfs://localhost:9000/newDir'
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -mkdir /destinationDir
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hadoop fs -cp /abc/file1.txt /destinationDir/

```

```

hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin
how is your brother
how is your sister
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin$ hdfs dfs -mv /abc /FFF
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin$ hdfs dfs -ls /FFF
Found 2 items
-rw-r--r-- 1 hadoop supergroup 90 2025-04-15 14:34 /FFF/file1.txt
-rw-r--r-- 1 hadoop supergroup 90 2025-04-15 14:35 /FFF/file2.txt
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin$ 9. Cp Hadoop HDFS
cp Command Usage cp Hadoop HDFS
cp Command Example
hadoop fs -cp /CSE/ /LLL
hadoop fs -ls /LLL

A The cp command copies a file from one directory to another directory within the HDFS.
9.: command not found
cp: target '/LLL' is not a directory
cp: command 'cp' or 'command': No such file or directory
ls: '/CSE/': No such file or directory
ls: '/LLL': No such file or directory
Command 'ls' not found, did you mean:
  command 'ls' from deb node-he (1.2.0-3)
  command 'ls' from deb the (3.3-c1-3huldi)
  To search for a package name:
    apt search ls
HDFS: command not found
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin$ hdfs fs -cp /abc/file1.txt /newDir/
cp: '/newDir/': No such file or directory: 'hdfs://localhost:9000/newDir'
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin$ hdfs dfs -mkdir /destinationDir
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin$ hadoop fs -cp /abc/file1.txt /destinationDir/
cp: '/abc/file1.txt': No such file or directory
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin$ hdfs dfs -ls /abc/
ls: '/abc/': No such file or directory
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin$ hdfs dfs -ls /
Found 5 items
drwxr-xr-x  - hadoop supergroup 0 2025-04-15 14:35 /FFF
drwxr-xr-x  - hadoop supergroup 0 2025-04-15 15:12 /newdataFlair
drwxr-xr-x  - hadoop supergroup 0 2024-05-13 14:36 /olddata
drwxr-xr-x  - hadoop supergroup 0 2025-04-15 14:19 /rgs
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin$ hdfs dfs -ls /FFF
[Found 2 items]
-rw-r--r-- 1 hadoop supergroup 90 2025-04-15 14:34 /FFF/file1.txt
-rw-r--r-- 1 hadoop supergroup 90 2025-04-15 14:35 /FFF/file2.txt
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin$ hadoop fs -cp /FFF/file1.txt /destinationDir/
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin$ hadoop fs -cp /FFF/file2.txt /destinationDir/
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin$ hdfs dfs -ls /destinationDir/
[Found 2 items]
-rw-r--r-- 1 hadoop supergroup 90 2025-04-15 14:49 /destinationDir/file1.txt
-rw-r--r-- 1 hadoop supergroup 90 2025-04-15 14:49 /destinationDir/file2.txt
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin$ 5-5

hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin
SUBCOMMAND may print help when invoked w/o parameters or with -h.
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: $ cd ~/hadoop/hadoop/sbin
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin$ ./start-dfs.sh
Starting namenodes on [localhost]
localhost: namenode is running as process 9900. Stop it first and ensure /tmp/hadoop-hadoop-namenode.pid file is empty before retry.
Starting datanodes
localhost: datanode is running as process 10056. Stop it first and ensure /tmp/hadoop-hadoop-datanode.pid file is empty before retry.
Starting secondary namenodes on [bmseccse-HP-Elite-Tower-800-G9-Desktop-PC]
bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: secondarynamenode is running as process 10334. Stop it first and ensure /tmp/hadoop-hadoop-secondarynamenode.pid file is empty before retry.
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin$ ./start-yarn.sh
Starting resourcemanager
resourcemanager is running as process 10644. Stop it first and ensure /tmp/hadoop-hadoop-resourcemanager.pid file is empty before retry.
Starting nodemanagers
localhost: nodemanager is running as process 10807. Stop it first and ensure /tmp/hadoop-hadoop-nodemanager.pid file is empty before retry.
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin$ jps
12226 Jps
10644 ResourceManager
10807 NodeManager
10956 DataNode
10956 SecondaryNameNode
10956 DataNode
10334 SecondaryNameNode
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin$ ./start-dfs.sh
Starting namenodes on [localhost]
localhost: namenode is running as process 9900. Stop it first and ensure /tmp/hadoop-hadoop-namenode.pid file is empty before retry.
Starting datanodes
localhost: datanode is running as process 10056. Stop it first and ensure /tmp/hadoop-hadoop-datanode.pid file is empty before retry.
Starting secondary namenodes on [bmseccse-HP-Elite-Tower-800-G9-Desktop-PC]
bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: secondarynamenode is running as process 10334. Stop it first and ensure /tmp/hadoop-hadoop-secondarynamenode.pid file is empty before retry.
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin$ ./start-yarn.sh
Starting resourcemanager
resourcemanager is running as process 10644. Stop it first and ensure /tmp/hadoop-hadoop-resourcemanager.pid file is empty before retry.
Starting nodemanagers
localhost: nodemanager is running as process 10807. Stop it first and ensure /tmp/hadoop-hadoop-nodemanager.pid file is empty before retry.
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin$ jps
13012 Jps
10642 ResourceManager
10807 NodeManager
10956 DataNode
9900 NameNode
10334 SecondaryNameNode
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin$ hdfs fs -ls /
[Found 2 items]
drwxr-xr-x  - hadoop supergroup 0 2024-05-13 15:12 /newdataFlair
drwxr-xr-x  - hadoop supergroup 0 2024-05-13 14:36 /olddata
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin$ hadoop fs -copyFromLocal /home/rekhags/Desktop/file1.txt /rgs/test.txt
copyFromLocal: /rgs/test.txt: No such file or directory: 'hdfs://localhost:9000/rgs/test.txt'
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin$ hdfs fs -mkdir /rgs
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin$ hdfs fs -copyFromLocal /home/rekhags/Desktop/file1.txt /rgs/test.txt
copyFromLocal: /home/rekhags/Desktop/file1.txt: No such file or directory
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin$ ls /home/hadoop/Desktop/
file1.txt  first.txt  test.txt
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin$ nano /home/hadoop/Desktop/file1.txt
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin$ hdfs fs -copyFromLocal /home/hadoop/Desktop/file1.txt /rgs/test.txt
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin$ hdfs fs -mkdir /rgs

```

```

hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~[200]sudo apt update
[sudo] password for hadoop
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~$ sudo apt update
[sudo] password for hadoop
Command 'sudo' not found, did you mean:
  command 'su' from deb sudo [1.9.9-1ubuntu2.4]
  command 'sudo' from deb sudo-ldap [1.9.9-1ubuntu2.4]
Try: sudo apt install <deb name>
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~$ tar -xvf hadoop-3.x.x.tar.gz
tar: (child): hadoop-3.x.x.tar.gz: cannot open: No such file or directory
tar: Child returned status 2
tar: Error is not recoverable: exiting now
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~$ sudo apt update
[sudo] password for hadoop
[sudo] password for hadoop
http://ppa.launchpad.net/ericpfir/con-epp/code stable InRelease
HIt:2 http://security.ubuntu.com/ubuntu jammy-security InRelease
Get:3 https://dl.google.com/linux/chrome/deb stable InRelease [1,825 B]
HIt:4 http://in.archive.ubuntu.com/ubuntu jammy InRelease
Get:5 http://in.archive.ubuntu.com/ubuntu jammy-updates InRelease [128 kB]
HIt:6 http://repo.mongodb.org/apt/ubuntu/jammy/mongodb-org/6.0 InRelease
Get:7 https://repo.mongodb.org/apt/ubuntu/jammy/mongodb-org/6.0/bionic Packages [1,214 B]
HIt:9 https://repo.mongodb.org/apt/ubuntu/jammy/mongodb-org/7.0 InRelease
HIt:10 http://in.archive.ubuntu.com/ubuntu jammy-backports InRelease
HIt:11 https://apache.jfrog.io/artifactory/cassandra-deb 41x InRelease
Fetched 131 KB in 3s (42.9 kB/s)
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
344 packages can be upgraded, run 'apt list --upgradable' to see them.
W: https://repo.mongodb.org/apt/ubuntu/dists/jammy/mongodb-org/6.0/InRelease: Key is stored in legacy trusted.gpg keyring (/etc/apt/trusted.gpg), see the DEPRECATION section in apt-key(8) for details.
W: https://debian.cassandra.apache.org/dists/41x/InRelease: Key is stored in legacy trusted.gpg keyring (/etc/apt/trusted.gpg), see the DEPRECATION section in apt-key(8) for details.
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~$ sudo apt install openjdk-11-jdk
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
openjdk-11-jdk is already the newest version (11.0.2+7-Ubuntu-12.04).
The following packages were automatically installed and are no longer required:
  linux-headers-6.2.0-26-generic linux-hwe-6.2-headers-6.2.0-26
  linux-image-6.2.0-26-generic linux-modules-6.2.0-26-generic
  linux-modules-extra-6.2.0-26-generic
Use 'sudo apt autoremove' to remove them.
0 upgraded, 0 newly installed, 0 to remove and 344 not upgraded.
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~$ java -version
openjdk version "11.0.20" 2025-01-21
OpenJDK Runtime Environment (build 11.0.20+post-Ubuntu-1ubuntu122.04)
OpenJDK 64-Bit Server VM (build 11.0.20+post-Ubuntu-1ubuntu122.04, mixed mode, sharing)
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~$ mkdir -p ~/hadoop_wordcount
cd ~/hadoop_wordcount
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop_wordcount$ nano WordCount.java
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop_wordcount$ javac -classpath $(hadoop classpath) -d /home/hadoop/wordcount_classes WordCount.java
WordCount.java:8: error: cannot find symbol
  public static class TokenizerMapper extends MapReduceBase implements Mapper<Object, Text, Text, IntWritable> {
                                         ^
  symbol:  class MapReduceBase
  location: class WordCount
WordCount.java:17: error: cannot find symbol
  }
                                         ^
  symbol:  class IntWritable
location: class WordCount
hadoop@bmseccse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop_wordcount$ ls
WordCount.java  WordCountMapper.class  WordCountReducer.class

```

## EXPERIMENT-6

### Implement Wordcount program on Hadoop framework

Code with Screenshots:

```
hadoop@hmsccse-HP-Elite-Tower-800-G9-Desktop-PC: ~$ hadoop fs -mkdir /input
hadoop fs -copyFromLocal /home/hadoop/Desktop/sample.txt /input/
mkdir: /input: File exists
copyFromLocal: /home/hadoop/Desktop/sample.txt: No such file or directory
hadoop@hmsccse-HP-Elite-Tower-800-G9-Desktop-PC: ~$ hadoop fs -rm -r /input
hadoop fs -mkdir /input
Deleted /input
hadoop@hmsccse-HP-Elite-Tower-800-G9-Desktop-PC: ~$ /home/hadoop/Desktop/HC.jar
hadoop@hmsccse-HP-Elite-Tower-800-G9-Desktop-PC: ~$ nano /home/hadoop/Desktop/sample.txt
hadoop@hmsccse-HP-Elite-Tower-800-G9-Desktop-PC: ~$ hadoop fs -copyFromLocal /home/hadoop/Desktop/sample.txt /input/
hadoop@hmsccse-HP-Elite-Tower-800-G9-Desktop-PC: ~$ hadoop fs -mkdir /input
mkdir: /input: File exists
hadoop@hmsccse-HP-Elite-Tower-800-G9-Desktop-PC: ~$ hadoop fs -copyFromLocal /home/hadoop/Desktop/sample.txt /input/
copyFromLocal: /input/sample.txt: File exists
hadoop@hmsccse-HP-Elite-Tower-800-G9-Desktop-PC: ~$ hadoop fs -ls /input
Found 1 items
-rw-r--r-- 1 hadoop supergroup 89 2025-04-29 15:21 /input/sample.txt
hadoop@hmsccse-HP-Elite-Tower-800-G9-Desktop-PC: ~$ hadoop jar /home/hadoop/HC.jar MyDriver /input/sample.txt /output
2025-04-29 15:22:52,183 INFO Client:DefaultNamenodeFallbackProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2025-04-29 15:22:52,243 INFO Client:DefaultNamenodeFallbackProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2025-04-29 15:22:52,245 INFO Mapreduce:JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/job_1745916871358_0002
2025-04-29 15:22:52,508 INFO Mapreduce:FileInputFormat: Total input files to process: 1
2025-04-29 15:22:52,578 INFO Mapreduce:Jobsubmitter: Number of splits:2
2025-04-29 15:22:52,678 INFO Mapreduce:Jobsubmitter: Submitting tokens for job: job_1745916871358_0002
2025-04-29 15:22:52,679 INFO Mapreduce:Jobsubmitter: Executing with tokens: []
2025-04-29 15:22:52,740 INFO Mapreduce:JobResourceUploader: Unable to find resource-types.xml'
2025-04-29 15:22:52,740 INFO Mapreduce:JobResourceUploader: Unable to find resource-types.xml'
2025-04-29 15:22:52,755 INFO Mapreduce:Job: The url to track the Job: http://hmsccse-HP-Elite-Tower-800-G9-Desktop-PC:8088/proxy/application_1745916871358_0002/
2025-04-29 15:22:52,781 INFO Mapreduce:Job: Running job: job_1745916871358_0002
2025-04-29 15:22:55,795 INFO Mapreduce:Job: Job job_1745916871358_0002 running in uber mode : false
2025-04-29 15:22:55,803 INFO Mapreduce:Job: map 0% reduce 0%
2025-04-29 15:22:55,803 INFO Mapreduce:Job: Job job_1745916871358_0002 failed with state FAILED due to: Application application_1745916871358_0002 failed 2 times due to AM Container for appattempt_1745916871358_0002 exited with exitcode 1
Failing this attempt:[diagnostics: [2025-04-29 15:22:55.864]Exception from container-launch,
Container id: container_1745916871358_0002_02_00001
Exit code: 1
[2025-04-29 15:22:55.866]Container exited with a non-zero exit code 1. Error file: prelaunch.err,
Last 4096 bytes of prelaunch.err :
Last 4096 bytes of stderr :
Error: Could not find or load main class org.apache.hadoop.mapreduce.v2.app.MRAppMaster
Caused by: java.lang.ClassNotFoundException: org.apache.hadoop.mapreduce.v2.app.MRAppMaster
Please check whether your <HADOOP_HOME>/etc/hadoop/mapred-site.xml contains the below configuration:
<property>
  <name>yarn.app.mapreduce.an-envs</name>
  <value>HADOOP_MAPRED_HOME=${full path of your hadoop distribution directory}</value>
</property>
<property>
  <name>yarn.mapreduce.map.envs</name>
  <value>HADOOP_MAPRED_HOME=${full path of your hadoop distribution directory}</value>
</property>
<property>
  <name>yarn.mapreduce.reduce.envs</name>
  <value>HADOOP_MAPRED_HOME=${full path of your hadoop distribution directory}</value>
</property>
[2025-04-29 15:22:55.866]Container exited with a non-zero exit code 1. Error file: prelaunch.err,
Last 4096 bytes of prelaunch.err :
Last 4096 bytes of stderr :
Error: Could not find or load main class org.apache.hadoop.mapreduce.v2.app.MRAppMaster
Caused by: java.lang.ClassNotFoundException: org.apache.hadoop.mapreduce.v2.app.MRAppMaster
Please check whether your <HADOOP_HOME>/etc/hadoop/mapred-site.xml contains the below configuration:
<property>
  <name>yarn.app.mapreduce.an-envs</name>
  <value>HADOOP_MAPRED_HOME=${full path of your hadoop distribution directory}</value>
</property>
<property>
  <name>yarn.mapreduce.map.envs</name>
  <value>HADOOP_MAPRED_HOME=${full path of your hadoop distribution directory}</value>
</property>
<property>
  <name>yarn.mapreduce.reduce.envs</name>
  <value>HADOOP_MAPRED_HOME=${full path of your hadoop distribution directory}</value>
</property>
For more detailed output, check the application tracking page: http://hmsccse-HP-Elite-Tower-800-G9-Desktop-PC:8088/cluster/app/application_1745916871358_0002 Then click on links to logs of each attempt.
Failing the application.
2025-04-29 15:22:55,814 INFO Mapreduce:Job: Counters: 0
Exception in thread "main" java.io.IOException: Job failed!
at org.apache.hadoop.mapred.JobClient.runJob(JobClient.java:876)
at MyDriver.run(MyDriver.java:28)
at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:82)
at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:97)
at MyDriver.main(MyDriver.java:33)
at java.base/jdk.internal.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at java.base/jdk.internal.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at java.base/jdk.internal.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.base/java.lang.reflect.Method.invoke(Method.java:569)
at org.apache.hadoop.util.RunJar.main(RunJar.java:132)
at org.apache.hadoop.util.RunJar.main(RunJar.java:241)
hadoop@hmsccse-HP-Elite-Tower-800-G9-Desktop-PC: ~$ hadoop fs -ls /output
ls: /output: No such file or directory
hadoop@hmsccse-HP-Elite-Tower-800-G9-Desktop-PC: ~$ nano /home/hadoop/hadoop/etc/hadoop/mapred-site.xml
hadoop@hmsccse-HP-Elite-Tower-800-G9-Desktop-PC: ~$ hadoop namenode -format
Starting namenodes on [localhost]
localhost: namenode is running as process 7941. Stop it first and ensure /tmp/hadoop-hadoop-namenode.pid file is empty before retry..
Starting secondary namenodes [hadoop@hmsccse-HP-Elite-Tower-800-G9-Desktop-PC]
localhost: datanode is running as process 8118. Stop it first and ensure /tmp/hadoop-hadoop-datanode.pid file is empty before retry.
Starting secondary namenodes [hadoop@hmsccse-HP-Elite-Tower-800-G9-Desktop-PC]
secondarynamenode is running as process 8400. Stop it first and ensure /tmp/hadoop-hadoop-secondarynamenode.pid file is empty before retry.
Starting resourcemanager
resourceManager is running as process 8687. Stop it first and ensure /tmp/hadoop-hadoop-resourcemanager.pid file is empty before retry.
```

```

Activities Terminal Apr 29 15:31
hadoop@bmseccse-HP-Elite-Tower-800-09-Desktop-PC: ~
resourceManager is running as process 8887. Stop it first and ensure /tmp/hadoop-hadoop-resourceManager.pid file is empty before retry.
Starting nodemanagers
localhost: nodemanager is running as process 8852. Stop it first and ensure /tmp/hadoop-hadoop-nodemanager.pid file is empty before retry.
hadoop@bmseccse-HP-Elite-Tower-800-09-Desktop-PC: ~
8408 SecondaryNameNode
15538 Jps
8852 NodeManager
7941 NameNode
8118 DataNode
1099 org.eclipse.equinox.launcher_1.6.1000.v20250227-1734.jar
8687 ResourceManager
hadoop@bmseccse-HP-Elite-Tower-800-09-Desktop-PC: $ cd ~/Desktop # or navigate to any other directory where you want to create the file
hadoop@bmseccse-HP-Elite-Tower-800-09-Desktop-PC: ~/Desktop$ nano sample.txt
hadoop@bmseccse-HP-Elite-Tower-800-09-Desktop-PC: ~/Desktop$ cd -
hadoop@bmseccse-HP-Elite-Tower-800-09-Desktop-PC: ~
8852 NodeManager
7941 NameNode
8118 DataNode
15593 Jps
18992 org.eclipse.equinox.launcher_1.6.1000.v20250227-1734.jar
8689 ResourceManager
hadoop@bmseccse-HP-Elite-Tower-800-09-Desktop-PC: ~ hadoop fs -mkdir /rgs
hadoop@bmseccse-HP-Elite-Tower-800-09-Desktop-PC: ~ hadoop fs -copyFromLocal ~/Desktop/sample.txt /rgs/test.txt
copyFromLocal: /rgs/test.txt: File exists
hadoop@bmseccse-HP-Elite-Tower-800-09-Desktop-PC: ~ hadoop fs -ls /rgs
File: /rgs/test.txt
-rw----r-- 1 hadoop supergroup 98 2025-04-15 14:19 /rgs/test.txt
hadoop@bmseccse-HP-Elite-Tower-800-09-Desktop-PC: ~ hadoop fs -cat /rgs/test.txt
hi how are you
how is your job
how is your family
how is your brother
how is your sister

hadoop@bmseccse-HP-Elite-Tower-800-09-Desktop-PC: ~ hadoop jar /home/hadoop/Desktop/WC.jar MyDriver /rgs /output
2025-04-29 15:29:14,897 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-04-29 15:29:14,935 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-04-29 15:29:14,935 INFO impl.MetricsSystemImpl: JobMetrics initialized
2025-04-29 15:29:14,944 WARN impl.MetricsSystemImpl: Jobtracker metrics system already initialized!
2025-04-29 15:29:14,996 WARN mapred.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-04-29 15:29:15,059 INFO mapred.FileInputFormat: Total input files to process : 1
2025-04-29 15:29:15,069 INFO mapred.JobSubmitter: number of splits:1
2025-04-29 15:29:15,126 INFO mapred.JobSubmitter: Submitting tokens for job: job_local537176479_0001
2025-04-29 15:29:15,126 INFO mapred.JobSubmitter: Executing with tokens: []
2025-04-29 15:29:15,182 INFO mapred.JobSubmitter: Submitting job: http://localhost:8080/
2025-04-29 15:29:15,183 INFO mapred.Job: Running job: job_local537176479_0001
2025-04-29 15:29:15,183 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-04-29 15:29:15,188 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapred.FileOutputCommitter
2025-04-29 15:29:15,188 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-04-29 15:29:15,188 INFO output.FileOutputCommitter: File Output Committer skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-04-29 15:29:15,221 INFO mapred.LocalJobRunner: Waiting for map tasks
2025-04-29 15:29:15,232 INFO mapred.LocalJobRunner: Starting task attempt_local537176479_0001_m_000000_0
2025-04-29 15:29:15,233 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-04-29 15:29:15,233 INFO output.FileOutputCommitter: File Output Committer skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false

Activities Terminal Apr 29 15:31
hadoop@bmseccse-HP-Elite-Tower-800-09-Desktop-PC: ~
hadoop@bmseccse-HP-Elite-Tower-800-09-Desktop-PC: ~ hadoop jar /home/hadoop/Desktop/WC.jar MyDriver /rgs /output
2025-04-29 15:29:14,897 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-04-29 15:29:14,935 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-04-29 15:29:14,935 INFO impl.MetricsSystemImpl: JobMetrics initialized
2025-04-29 15:29:14,944 WARN impl.MetricsSystemImpl: Jobtracker metrics system already initialized!
2025-04-29 15:29:15,059 INFO mapred.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-04-29 15:29:15,059 INFO mapred.FileInputFormat: Total input files to process : 1
2025-04-29 15:29:15,069 INFO mapred.JobSubmitter: number of splits:1
2025-04-29 15:29:15,126 INFO mapred.JobSubmitter: Submitting tokens for job: job_local537176479_0001
2025-04-29 15:29:15,126 INFO mapred.JobSubmitter: Executing with tokens: []
2025-04-29 15:29:15,182 INFO mapred.JobSubmitter: Submitting job: http://localhost:8080/
2025-04-29 15:29:15,183 INFO mapred.Job: Running job: job_local537176479_0001
2025-04-29 15:29:15,183 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-04-29 15:29:15,188 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapred.FileOutputCommitter
2025-04-29 15:29:15,188 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-04-29 15:29:15,188 INFO output.FileOutputCommitter: File Output Committer skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-04-29 15:29:15,221 INFO mapred.LocalJobRunner: Waiting for map tasks
2025-04-29 15:29:15,222 INFO mapred.LocalJobRunner: Starting task: attempt_local537176479_0001_m_000000_0
2025-04-29 15:29:15,232 INFO output.FileOutputCommitter: File Output Committer skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-04-29 15:29:15,232 INFO output.FileOutputCommitter: File Output Committer skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-04-29 15:29:15,232 INFO mapred.Task: Processing split: hdfs://localhost:9600/rgs/test.txt@0+98
2025-04-29 15:29:15,242 INFO mapred.MapTask: Processing split: hdfs://localhost:9600/rgs/test.txt@0+98
2025-04-29 15:29:15,252 INFO mapred.MapTask: numReduceTasks: 1
2025-04-29 15:29:15,281 INFO mapred.MapTask: [EQUATOR] 0 kvl 26214396(10485784)
2025-04-29 15:29:15,281 INFO mapred.MapTask: aggrreduce.task.lo.sort.mb: 100
2025-04-29 15:29:15,281 INFO mapred.MapTask: sortInMemorySize: 100000000
2025-04-29 15:29:15,281 INFO mapred.MapTask: sortInMemorySizeMB = 104857600
2025-04-29 15:29:15,281 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
2025-04-29 15:29:15,283 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2025-04-29 15:29:15,316 INFO mapred.LocalJobRunner: 
2025-04-29 15:29:15,316 INFO mapred.MapTask: Starting flush of map output
2025-04-29 15:29:15,316 INFO mapred.MapTask: splitting map output
2025-04-29 15:29:15,316 INFO mapred.MapTask: Buffer length = 100; bufend = 100; bufvoid = 104857600
2025-04-29 15:29:15,316 INFO mapred.MapTask: finished spill: 0
2025-04-29 15:29:15,325 INFO mapred.Task: Task@attempt_local537176479_0001_m_000000_0 is done. And is in the process of committing
2025-04-29 15:29:15,327 INFO mapred.LocalJobRunner: hdfs://localhost:9600/rgs/test.txt@0+98
2025-04-29 15:29:15,327 INFO mapred.Task: Task@attempt_local537176479_0001_m_000000_0 is done.
2025-04-29 15:29:15,329 INFO mapred.Task: Final Counters for attempt_local537176479_0001_m_000000_0: Counters: 23
File System Counters
FILE: Number of bytes read=4150
FILE: Number of bytes written=642471
FILE: Number of read operations=0
FILE: Number of write operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=0
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=26
Map output records=26
Map output bytes=169
Map output Materialized bytes=215

```

```

Activities Terminal Apr 29 15:32
hadoop@hmscece-HP-Elite-Tower-800-D-Desktop-PC: ~
2025-04-29 15:29:15,316 INFO mapred.MapTask: burststart = 0; butend = 169; butvold = 104657600
2025-04-29 15:29:15,316 INFO mapred.MapTask: mapreduce.map.ReduceTaskId=20214320(104657584); kvnd = 20214320(104657280); length = 77/6553600
2025-04-29 15:29:15,319 INFO mapred.Task: mapreduce.map.ReduceTaskId=20214320(104657584); finished=true
2025-04-29 15:29:15,325 INFO mapred.Task: Task attempt_local537170479_0001_m_000000_0 is done. And is in the process of committing
2025-04-29 15:29:15,327 INFO mapred.Task: Task attempt_local537170479_0001_m_000000_0 done.
2025-04-29 15:29:15,329 INFO mapred.Task: Final Counters for attempt_local537170479_0001_m_000000_0: Counters: 23
  File System Counters
    FILE: Number of bytes read=416
    FILE: Number of bytes written=642471
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=98
    HDFS: Number of bytes written=8
    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
    Combine input records=0
    Map output records=20
    Map output bytes=169
    Map output materialized bytes=215
    Input split bytes=80
    Combine output records=0
    Spilled Records=20
    Failed Shuffles=0
    Merged Map outputs=0
    GC time elapsed (ms)=8
    Total committed heap usage (bytes)=633339904
  File Input Format Counters
    FILE: Number of bytes read=416
2025-04-29 15:29:15,329 INFO mapred.LocalJobRunner: Finishing task: attempt_local537170479_0001_m_000000_0
2025-04-29 15:29:15,330 INFO mapred.LocalJobRunner: map task executor complete.
2025-04-29 15:29:15,331 INFO mapred.LocalJobRunner: Waiting for reduce tasks
2025-04-29 15:29:15,331 INFO mapred.LocalJobRunner: Starting task: attempt_local537170479_0001_r_000000_0
2025-04-29 15:29:15,335 INFO mapred.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory=false, ignore cleanup failures: false
2025-04-29 15:29:15,335 INFO mapred.Task: Using ResourceCalculatorProcessTree : []
2025-04-29 15:29:15,336 INFO mapred.Task: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@2749269
2025-04-29 15:29:15,337 WARN impl.MetricsSystemImpl: JobTracker metric system already initialized!
2025-04-29 15:29:15,340 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=5827985408, maxSingleShuffleLimit=1456990352, mergeThreshold=3846470400, toSortFactor=10, nonToMemMergeOutputsThreshold=10
2025-04-29 15:29:15,341 INFO reduce.EventFetcher: attempt_local537170479_0001_r_000000_0 Thread started: EventFetcher for Fetching Map Completion Events
2025-04-29 15:29:15,342 INFO reduce.EventFetcher: attempt_local537170479_0001_r_000000_0 Thread started: EventFetcher for Fetching Reduce Completion Events
2025-04-29 15:29:15,356 INFO reduce.InMemoryMapOutput: Read 211 bytes from map-output For attempt_local537170479_0001_m_000000_0 decom: 211 len: 215 to MEMORY
2025-04-29 15:29:15,356 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of size: 211, inMemoryMapOutputs.size() -> 1, commitMemory -> 0, usedMemory ->211
2025-04-29 15:29:15,357 INFO reduce.EventFetcher: EventFetcher is interrupted. Returning
2025-04-29 15:29:15,357 INFO mapred.LocalJobRunner: / i copied.
2025-04-29 15:29:15,357 INFO reduce.MergeManagerImpl: merger called with 1 in-memory map-outputs and 0 on-disk map-outputs
2025-04-29 15:29:15,358 INFO reduce.MergeManagerImpl: sorted segments
2025-04-29 15:29:15,362 INFO mapred.Herger: Down to the last merge-pass, with 1 segments left of total size: 265 bytes
2025-04-29 15:29:15,362 INFO reduce.MergeManagerImpl: Merged 1 segments, 211 bytes to disk to satisfy reduce memory limit
2025-04-29 15:29:15,362 INFO reduce.MergeManagerImpl: Merging 1 files, 215 bytes from disk
2025-04-29 15:29:15,362 INFO mapred.Herger: Merging 0 segments, 0 bytes from memory into reduce
2025-04-29 15:29:15,363 INFO mapred.Herger: Merging 1 sorted segments
2025-04-29 15:29:15,363 INFO mapred.Herger: Down to the last merge-pass, with 1 segments left of total size: 265 bytes
2025-04-29 15:29:15,363 INFO mapred.Task: Task attempt_local537170479_0001_r_000000_0 is allowed to commit now
2025-04-29 15:29:15,416 INFO mapred.Task: Task attempt_local537170479_0001_r_000000_0 done. And is in the process of committing
2025-04-29 15:29:15,416 INFO mapred.Task: Task attempt_local537170479_0001_r_000000_0 allowed to commit now
2025-04-29 15:29:15,418 INFO mapred.Task: Task attempt_local537170479_0001_r_000000_0 to hdfs://localhost:9000/output
2025-04-29 15:29:15,420 INFO mapred.LocalJobRunner: reduce > reduce
2025-04-29 15:29:15,424 INFO mapred.Task: Task attempt_local537170479_0001_r_000000_0 done.
2025-04-29 15:29:15,424 INFO mapred.Task: Final counters for attempt_local537170479_0001_r_000000_0: Counters: 30
  File System Counters
    FILE: Number of bytes read=418
    FILE: Number of bytes written=642686
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes written=6
    HDFS: Number of read operations=18
    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
    Spilled Records=10
    spilled Records=20
    shuffled Maps=1
    Failed Shuffles=0
    Merged Map outputs=1

```

```
Activities ◻ Terminal Apr 29 15:33
hadoop@bmsece-HP-Elite-Tower-300-C9-Desktop-PC: ~

File System Counters:
  FILE: Number of bytes read=4618
  FILE: Number of bytes written=42086
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=39
  HDFS: Number of bytes written=49
  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=0
  Reduce shuffle bytes=215
  Reduce input records=20
  Redundant map records=16
  Spilled Records=20
  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
  ID_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0

File Output Format Counters:
  Bytes Written=69

2023-04-29 15:29:15,424 INFO mapred.LocalJobRunner: Flushing task: attempt_local537370479_0001_r_000000_0
2023-04-29 15:29:15,424 INFO mapred.LocalJobRunner: map task executor complete,
2023-04-29 15:29:16,185 INFO mapreduce.Job: Job job_local537370479_0001 running in uber mode : false
2023-04-29 15:29:16,186 INFO mapreduce.Job: map 100% reduce 100%
2023-04-29 15:29:16,186 INFO mapreduce.Job: Job job_local537370479_0001 completed successfully
2023-04-29 15:29:16,198 INFO mapreduce.Job: Counters: 36
  File System Counters:
    FILE: Number of bytes read=8774
    FILE: Number of bytes written=285157
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=0
    HDFS: Number of bytes written=49
    HDFS: Number of read operations=1
    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 output records=20
  Map output bytes=169
```

```
Activities Terminal Apr 29 15:34
hadoop@bmcsse-HP-Elite-Tower-800-C9-Desktop-PC: ~
2025-04-29 15:29:15,424 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-04-29 15:29:16,185 INFO mapreduce.Job: Job job_local537170479_0001 running in uber mode : false
2025-04-29 15:29:16,186 INFO mapreduce.Job: map 10% reduce 100%
2025-04-29 15:29:16,186 INFO mapreduce.Job: Job job_local537170479_0001 completed successfully
2025-04-29 15:29:16,186 org.apache.hadoop.mapreduce.Job: Counters: 50
  File System Counters:
    FILE: Number of bytes read=8774
    FILE: Number of bytes written=1285157
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    FILE: Number of large write operations=0
    HDFS: Number of bytes read=180
    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=6
    Map output records=20
    Map output bytes=20
    Map output total initial bytes=215
    Input split bytes=80
    Combine input records=0
    Combine output records=0
    Reduce Input groups=10
    Reduce Input bytes=215
    Reduce Input福录=20
    Reduce output records=10
    Spilled Records=48
    Shuffled Maps=1
    Failed Shuffles=0
    Merged Map outputs=1
    Coarsegrained (m)=0
    Total committed heap usage (bytes)=1266079800
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    DUPLICATE=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=0
  File Output Format Counters
    Bytes Written=69
0
hadoop@bmcsse-HP-Elite-Tower-800-C9-Desktop-PC: $ hadoop fs -ls /output
Found 2 items
-rw-r--r-- 1 hadoop supergroup 8 2025-04-29 15:29 /output/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 69 2025-04-29 15:29 /output/part-00000
hadoop@bmcsse-HP-Elite-Tower-800-C9-Desktop-PC: $ hadoop fs -cat /output/part-00000
are 1
brother 1
family 1
```

```

Activities Terminal Apr 29 15:34
hadoop@bmseccse-HP-Elite-Tower-800-D9-Desktop-PC: ~
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=180
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=0
  Map output records=0
  Map output bytes=169
  Map output materialized bytes=215
  Input split bytes=80
  Combiner output records=0
  Combine output records=0
  Reduce input groups=18
  Reduce shuffle bytes=215
  Reduce input records=20
  Reduce output records=10
  Spilled records=10
  Shuffled Maps=1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=0
  Total committed heap usage (bytes)=1266679898
Shuffle
  Bytes=0
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=99
  File Output Format Counters
    Bytes Written=0
hadoop@bmseccse-HP-Elite-Tower-800-D9-Desktop-PC: $ hadoop fs -ls /output
Found 2 items
-rw-r--r-- 1 hadoop supergroup          0 2025-04-29 15:29 /output/_SUCCESS
-rw-r--r-- 1 hadoop supergroup       69 2025-04-29 15:29 /output/part-00000
hadoop@bmseccse-HP-Elite-Tower-800-D9-Desktop-PC: $ hadoop fs -cat /output/part-00000
are
  brother
  finally
  h1
  how
  is
  4
  job
  sister
  you
  your
  4
hadoop@bmseccse-HP-Elite-Tower-800-D9-Desktop-PC: $

```

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

- Project Explorer:** Shows the WordCount project structure, including packages (src, settings), Java files (MyDriver.java, MyMapper.java, MyReducer.java), and referenced libraries.
- Code Editor:** Displays the MyDriver.java code, which is a Hadoop Driver class. It imports various Hadoop classes and defines a main method that sets up a JobConf, configures input and output paths, and runs the job.
- Task List:** Shows a task related to Java 24 Support.
- Outline:** Shows the outline of the MyDriver class.
- Problems:** Shows a message about Java 24 being released and asks if the user wants to install support.

```

eclipse-workspace · WordCount/src/MyDriver.java - Eclipse IDE
File Edit Source Refactor Navigate Search Project Run Window Help
src
  WordCount
    JRE System Library[JavaSE-1.8]
    src
      (default package)
        MyDriver.java
        MyMapper.java
        MyReducer.java
    settings
    classpath
    project

import java.io.IOException;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapred.FileInputFormat;
import org.apache.hadoop.mapred.FileOutputFormat;
import org.apache.hadoop.mapred.JobClient;
import org.apache.hadoop.mapred.JobConf;
import org.apache.hadoop.util.Tool;
import org.apache.hadoop.util.ToolRunner;
public class MyDriver extends Configuration implements Tool {
    public int run(String[] args) throws IOException {
        if (args.length < 2) {
            System.out.println("Please provide input and output paths");
            return -1;
        }
        JobConf conf = new JobConf(MyDriver.class);
        FileInputFormat.setInputPaths(conf, new Path(args[0]));
        FileOutputFormat.setOutputPath(conf, new Path(args[1]));
        conf.setMapperClass(MyMapper.class);
        conf.setCombinerClass(MyReducer.class);
        conf.setMapOutputKeyClass(Text.class);
        conf.setMapOutputValueClass(IntWritable.class);
        conf.setOutputKeyClass(Text.class);
        conf.setOutputValueClass(IntWritable.class);
        JobClient.runJob(conf);
        return 0;
    }
    public static void main(String[] args) throws Exception {
        int exitCode = ToolRunner.run(new MyDriver(), args);
        System.out.println(exitCode);
    }
}

```

Activities Eclipse Apr 29 15:35

eclipse-workspace - WordCount/src/MyMapper.java - Eclipse IDE

```
File Edit Source Refactor Navigate Search Project Run Window Help
```

Package Explorer X MyMapper.java MyDriver.java

JRE System Library [JavaSE-1.8]

src (default package)

MyDriver.java

MyMapper.java

MyReducer.java

Referenced Libraries

settings

classpath

project

```
1 import java.io.IOException;
2 import org.apache.hadoop.io.IntWritable;
3 import org.apache.hadoop.io.LongWritable;
4 import org.apache.hadoop.io.Text;
5 import org.apache.hadoop.mapreduce.Mapper;
6 import org.apache.hadoop.mapred.MapReduceBase;
7 import org.apache.hadoop.mapred.OutputCollector;
8 import org.apache.hadoop.mapred.Reporter;
9
10 public class MyMapper extends MapReduceBase implements Mapper<LongWritable, Text, Text, IntWritable> {
11     public void map(LongWritable key, Text value, OutputCollector<Text, IntWritable> output, Reporter rep) throws IOException {
12         String line = value.toString();
13         for (String word : line.split(" ")) {
14             if (word.length() > 0) {
15                 output.collect(new Text(word), new IntWritable(1));
16             }
17         }
18     }
19 }
20 }
```

Task List X

Find All Activate...

Outline X

MyMapper

Problems Javadoc Declaration Install Java 24 Support

Java 24 has been released.

Would you like to install Eclipse Java Development Tools support for Java 24 into your IDE?

Install Java 24 Support Remind Me Later Don't Ask Again

Would you like to download a Java 24 JDK?

Download Temurin JDK 24 Download OpenJDK 24

Writable Smartinsert 21:1:779

Activities Eclipse Apr 29 15:35

eclipse-workspace - WordCount/src/MyReducer.java - Eclipse IDE

```
File Edit Source Refactor Navigate Search Project Run Window Help
```

Package Explorer X MyMapper.java MyReducer.java MyDriver.java

JRE System Library [JavaSE-1.8]

src (default package)

MyDriver.java

MyMapper.java

MyReducer.java

Referenced Libraries

settings

classpath

project

```
1 import java.io.IOException;
2 import java.util.Iterator;
3 import org.apache.hadoop.io.IntWritable;
4 import org.apache.hadoop.io.Text;
5 import org.apache.hadoop.mapred.MapReduceBase;
6 import org.apache.hadoop.mapred.OutputCollector;
7 import org.apache.hadoop.mapred.Reducer;
8 import org.apache.hadoop.mapred.Reporter;
9
10 public class MyReducer extends MapReduceBase implements Reducer<Text, IntWritable, Text, IntWritable> {
11     public void reduce(Text key, Iterator<IntWritable> values, OutputCollector<Text, IntWritable> output, Reporter rep) throws IOException {
12         int count = 0;
13         while (values.hasNext()) {
14             count += values.next().get();
15         }
16         output.collect(key, new IntWritable(count));
17     }
18 }
```

Task List X

Find All Activate...

Outline X

MyReducer

Problems Javadoc Declaration Install Java 24 Support

Java 24 has been released.

Would you like to install Eclipse Java Development Tools support for Java 24 into your IDE?

Install Java 24 Support Remind Me Later Don't Ask Again

Would you like to download a Java 24 JDK?

Download Temurin JDK 24 Download OpenJDK 24

Writable Smartinsert 19:1:726

Activities Eclipse Apr 29 15:35

eclipse-workspace - WordCount/src/MyDriver.java - Eclipse IDE

File Edit Source Refactor Navigate Search Project Run Window Help

Package Explorer X

JRE System Library [JavaSE-1.8]

src (default package)

WordCount

MyMapper.java MyReducer.java MyDriver.java

Referenced Libraries

settings classpath project

```
1 import java.io.IOException;
2 import org.apache.hadoop.conf.Configuration;
3 import org.apache.hadoop.fs.Path;
4 import org.apache.hadoop.mapreduce.Job;
5 import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
6 import org.apache.hadoop.mapred.FileOutputFormat;
7 import org.apache.hadoop.mapred.JobClient;
8 import org.apache.hadoop.mapred.lib.Counters;
9 import org.apache.hadoop.util.Tool;
10 import org.apache.hadoop.util.ToolRunner;
11
12 public class MyDriver extends Configuration implements Tool {
13     public int run(String[] args) throws IOException {
14         if (args.length < 2) {
15             System.out.println("Please provide input and output paths");
16             return -1;
17         }
18         JobConf conf = new JobConf(MyDriver.class);
19         FileInputFormat.setInputPaths(conf, new Path(args[0]));
20         FileOutputFormat.setOutputPath(conf, new Path(args[1]));
21         conf.setMapperClass(MyMapper.class);
22         conf.setReducerClass(MyReducer.class);
23         conf.setMapOutputKeyClass(IntWritable.class);
24         conf.setMapOutputValueClass(IntWritable.class);
25         conf.setOutputKeyClass(Text.class);
26         conf.setOutputValueClass(IntWritable.class);
27         JobClient.runJob(conf);
28     }
29     return 0;
30 }
31
32 public static void main(String[] args) throws Exception {
33     int exitCode = ToolRunner.run(new MyDriver(), args);
34     System.out.println(exitCode);
35 }
```

Problems Javadoc Declaration Install Java 24 Support X

Java 24 has been released.

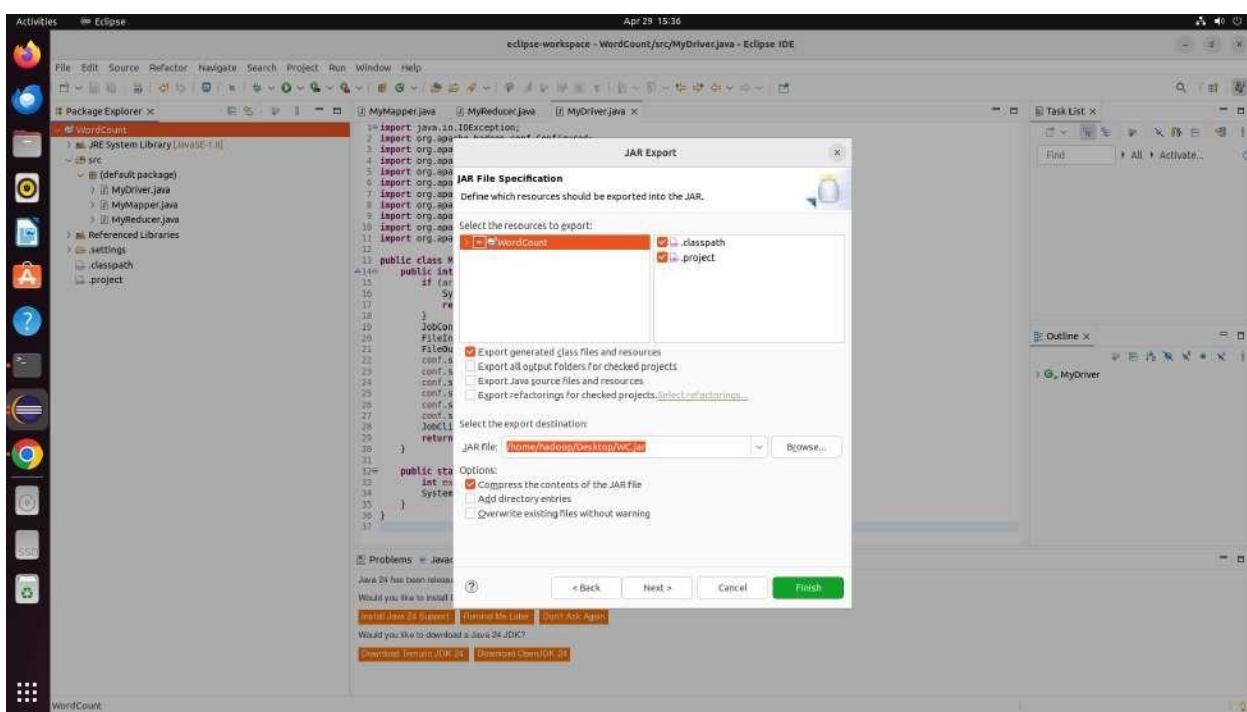
Would you like to install Eclipse Java Development Tools support for Java 24 into your IDE?

Install Java 24 Support Remind Me Later Don't Ask Again

Would you like to download a Java 24 JDK?

Download Temurin JDK 24 Download Oracle JDK 24

Writable Smart insert 37/11 1391



## EXPERIMENT-7

From the following link extract the weather data

<https://github.com/tomwhite/hadoop-book/tree/master/input/ncdc/all>

- a) Create a MapReduce program to  
find average temperature for each year from NCDC  
data set.
- b) find the mean max temperature for every month.

Code with screenshots:

eclipse-workspace the - NCDCWeatherAnalysis/src/temp/AverageDriver.java - Eclipse IDE

```
File Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer Task List
src AverageDriver AverageMapper AverageReducer MeanMaxMapper MeanMaxReducer
src MeanMaxDriver.java
src MeanMaxReducer.java
temp AverageDriver.java AverageMapper.java AverageReducer.java
Referenced Libraries

1 package temp;
2 import org.apache.hadoop.fs.Path;
3 import org.apache.hadoop.io.IntWritable;
4 import org.apache.hadoop.mapreduce.Job;
5 import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
6 import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
7 public class AverageDriver {
8     public static void main(String[] args) throws Exception {
9         if (args.length != 1) {
10             System.out.println("Please enter input and output parameters");
11             System.exit(-1);
12         }
13         Job job = Job.getJobConf();
14         job.setJarByClass(AverageDriver.class); // Set the main class for the job
15         job.setJobName("Average Temperature by Year"); // Set a name for the job
16         FileInputFormat.addInputPath(job, new Path(args[0]));
17         FileOutputFormat.setOutputPath(job, new Path(args[1]));
18
19         // Set the Mapper and Reducer classes
20         job.setMapperClass(AverageMapper.class);
21         job.setReducerClass(AverageReducer.class);
22
23         // Set the output key and value types
24         job.setOutputKeyClass(Text.class);
25         job.setOutputValueClass(IntWritable.class);
26
27         // Submit the job and wait for completion
28         System.exit((job.waitForCompletion(true)) ? 0 : 1);
29     }
30 }
31
32 }
```

Problems Declaration Console Install Java 24 Support
terminated: AverageDriver [Java Application] /usr/lib/jvm/java-8-openjdk-amd64/bin/java (06-May-2025, 2:49:49 pm - 2:49:51 pm elapsed: 0:00:01.788) [pid: 7013]
Please enter input and output parameters?

eclipse-workspace the - NCDCWeatherAnalysis/src/temp/AverageDriver.java - Eclipse IDE

```
File Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer Task List
src AverageDriver AverageMapper AverageReducer MeanMaxMapper MeanMaxReducer
src MeanMaxDriver.java
src MeanMaxReducer.java
temp AverageDriver.java AverageMapper.java AverageReducer.java
Referenced Libraries

1 package temp;
2 import org.apache.hadoop.fs.Path;
3 import org.apache.hadoop.io.IntWritable;
4 import org.apache.hadoop.mapreduce.Job;
5 import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
6 import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
7 public class AverageDriver {
8     public static void main(String[] args) throws Exception {
9         if (args.length != 1) {
10             System.out.println("Please enter input and output parameters");
11             System.exit(-1);
12         }
13         Job job = Job.getJobConf();
14         job.setJarByClass(AverageDriver.class); // Set the main class for the job
15         job.setJobName("Average Temperature by Year"); // Set a name for the job
16         FileInputFormat.addInputPath(job, new Path(args[0]));
17         FileOutputFormat.setOutputPath(job, new Path(args[1]));
18
19         // Set the Mapper and Reducer classes
20         job.setMapperClass(AverageMapper.class);
21         job.setReducerClass(AverageReducer.class);
22
23         // Set the output key and value types
24         job.setOutputKeyClass(Text.class);
25         job.setOutputValueClass(IntWritable.class);
26
27         // Submit the job and wait for completion
28         System.exit((job.waitForCompletion(true)) ? 0 : 1);
29     }
30 }
31
32 }
```

Problems Declaration Console Install Java 24 Support
terminated: AverageDriver [Java Application] /usr/lib/jvm/java-8-openjdk-amd64/bin/java (06-May-2025, 2:49:49 pm - 2:49:51 pm elapsed: 0:00:01.788) [pid: 7013]
Please enter input and output parameters?

eclipse-workspace the - NCDCWeatherAnalysis/src/temp/AverageDriver.java - Eclipse IDE

```

File Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer X
  NCDCWeatherAnalysis
    JRE System Library [JavaSE-1.8]
      src
        meanmax
          MeanMaxDriver.java
          MeanMaxMapper.java
          MeanMaxReducer.java
        temp
          AverageDriver.java
          AverageMapper.java
          AverageReducer.java
      Referenced Libraries
Task List X
  Find All Activate...
  Connect Mylyn
  Connect to your task and ALM tools or create a local task.
Outline X
  temp
    AverageDriver.java
      main(String[])

```

```

package temp;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
public class AverageDriver {
  public static void main(String[] args) throws Exception {
    if (args.length != 1) {
      System.out.println("Please enter input and output parameters");
      System.exit(-1);
    }
    Job job = Job.getJobConf();
    job.setJarByClass(AverageDriver.class); // Set the main class for the job
    job.setJobName("Average Temperature by Year"); // Set a name for the job
    FileInputFormat.addInputPath(job, new Path(args[0]));
    FileOutputFormat.setOutputPath(job, new Path(args[1]));
    // Set the Mapper and Reducer classes
    job.setMapperClass(AverageMapper.class);
    job.setReducerClass(AverageReducer.class);
    // Set the output key and value types
    job.setOutputKeyClass(Text.class);
    job.setOutputValueClass(IntWritable.class);
    // Submit the job and wait for completion
    System.exit((job.waitForCompletion(true)) ? 0 : 1);
  }
}

```

Problems Declaration Console Install Java 24 Support  
 terminated>AverageDriver [Java Application] /usr/lib/jvm/java-8-openjdk-amd64/bin/java (06-May-2025, 2:49:49 pm - 2:49:51 pm elapsed: 0:00:01.788) [pid: 7013]  
 Please enter input and output parameters?

meanmax.MeanMaxReducer.java - NCDCWeatherAnalysis/src

eclipse-workspace the - NCDCWeatherAnalysis/src/temp/AverageDriver.java - Eclipse IDE

```

File Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer X
  NCDCWeatherAnalysis
    JRE System Library [JavaSE-1.8]
      src
        meanmax
          MeanMaxDriver.java
          MeanMaxMapper.java
          MeanMaxReducer.java
        temp
          AverageDriver.java
          AverageMapper.java
          AverageReducer.java
      Referenced Libraries
Task List X
  Find All Activate...
  Connect Mylyn
  Connect to your task and ALM tools or create a local task.
Outline X
  temp
    AverageDriver.java
      main(String[])

```

```

package temp;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
public class AverageDriver {
  public static void main(String[] args) throws Exception {
    if (args.length != 1) {
      System.out.println("Please enter input and output parameters");
      System.exit(-1);
    }
    Job job = Job.getJobConf();
    job.setJarByClass(AverageDriver.class); // Set the main class for the job
    job.setJobName("Average Temperature by Year"); // Set a name for the job
    FileInputFormat.addInputPath(job, new Path(args[0]));
    FileOutputFormat.setOutputPath(job, new Path(args[1]));
    // Set the Mapper and Reducer classes
    job.setMapperClass(AverageMapper.class);
    job.setReducerClass(AverageReducer.class);
    // Set the output key and value types
    job.setOutputKeyClass(Text.class);
    job.setOutputValueClass(IntWritable.class);
    // Submit the job and wait for completion
    System.exit((job.waitForCompletion(true)) ? 0 : 1);
  }
}

```

Problems Declaration Console Install Java 24 Support  
 terminated>AverageDriver [Java Application] /usr/lib/jvm/java-8-openjdk-amd64/bin/java (06-May-2025, 2:49:49 pm - 2:49:51 pm elapsed: 0:00:01.788) [pid: 7013]  
 Please enter input and output parameters?

temp.AverageDriver.java - NCDCWeatherAnalysis/src

May 6 15:22

eclipse-workspace the - NCDCWeatherAnalysis/src/temp/AverageDriver.java - Eclipse IDE

```

1 package temp;
2 import org.apache.hadoop.fs.Path;
3 import org.apache.hadoop.io.IntWritable;
4 import org.apache.hadoop.mapreduce.Job;
5 import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
6 import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
7 import org.apache.hadoop.mapreduce.Mapper;
8 import org.apache.hadoop.mapreduce.Reducer;
9 public class AverageDriver {
10     public static void main(String[] args) throws Exception {
11         if (args.length != 2) {
12             System.err.println("Please enter input and output parameters");
13             System.exit(1);
14         }
15         Job job = Job.getInstance();
16         job.setJarByClass(AverageDriver.class); // Set the main class for the job
17         job.setJobName("Average Temperature by Year"); // Set a name for the job
18         FileInputFormat.addInputPath(job, new Path(args[0]));
19         FileOutputFormat.setOutputPath(job, new Path(args[1]));
20
21         // Set the Mapper and Reducer classes
22         job.setMapperClass(AverageMapper.class);
23         job.setReducerClass(AverageReducer.class);
24
25         // Set the output key and value types
26         job.setOutputKeyClass(Text.class);
27         job.setOutputValueClass(IntWritable.class);
28
29         // Submit the job and wait for completion
30         System.exit(job.waitForCompletion(true) ? 0 : 1);
31     }
32 }
33

```

May 6 15:23

eclipse-workspace the - NCDCWeatherAnalysis/src/temp/AverageDriver.java - Eclipse IDE

```

1 package temp;
2 import org.apache.hadoop.fs.Path;
3 import org.apache.hadoop.io.IntWritable;
4 import org.apache.hadoop.mapreduce.Job;
5 import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
6 import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
7 import org.apache.hadoop.mapreduce.Mapper;
8 import org.apache.hadoop.mapreduce.Reducer;
9 public class AverageDriver {
10     public static void main(String[] args) throws Exception {
11         if (args.length != 2) {
12             System.err.println("Please enter input and output parameters");
13             System.exit(1);
14         }
15         Job job = Job.getInstance();
16         job.setJarByClass(AverageDriver.class); // Set the main class for the job
17         job.setJobName("Average Temperature by Year"); // Set a name for the job
18         FileInputFormat.addInputPath(job, new Path(args[0]));
19         FileOutputFormat.setOutputPath(job, new Path(args[1]));
20
21         // Set the Mapper and Reducer classes
22         job.setMapperClass(AverageMapper.class);
23         job.setReducerClass(AverageReducer.class);
24
25         // Set the output key and value types
26         job.setOutputKeyClass(Text.class);
27         job.setOutputValueClass(IntWritable.class);
28
29         // Submit the job and wait for completion
30         System.exit(job.waitForCompletion(true) ? 0 : 1);
31     }
32 }
33

```

```

Activities Terminal May 6 15:25
hadoop@bmsccse-HP-Elite-Tower-600-G9-Desktop-PC: ~/Desktop
hadoop@bmsccse-HP-Elite-Tower-600-G9-Desktop-PC: $ hdfs dfs -mkdirr /ncdc_input
mkdir: Call From bmsccse-HP-Elite-Tower-600-G9-Desktop-PC/127.0.1.1 to localhost:9000 failed on connection exception: java.net.ConnectException: Connection refused; For more details see: http://wiki.apache.org/hadoop/ConnectionRefused
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes (bmsccse-HP-Elite-Tower-600-G9-Desktop-PC)
Starting resourcemanager
Starting nodemanagers
16032 SecondaryNameNode
15745 DataNode
15558 NameNode
7093 org.eclipse.equinox.launcher_1.6.1000.v20250227-1734.jar
16504 NodeManager
16292 ResourceManager
16310 ResourceManager
hadoop@bmsccse-HP-Elite-Tower-600-G9-Desktop-PC: $ jps
16032 SecondaryNameNode
15745 DataNode
15558 NameNode
7093 org.eclipse.equinox.launcher_1.6.1000.v20250227-1734.jar
16504 NodeManager
16292 ResourceManager
16310 ResourceManager
hadoop@bmsccse-HP-Elite-Tower-600-G9-Desktop-PC: $ hdfs dfs -mkdirr /ncdc_input
hadoop@bmsccse-HP-Elite-Tower-600-G9-Desktop-PC: $ git clone https://github.com/tomwhite/hadoop-book.git
fatal: destination path 'hadoop-book' already exists and is not an empty directory.
hadoop@bmsccse-HP-Elite-Tower-600-G9-Desktop-PC: $ cd hadoop-book/input/ncdc/all/
hdfs dfs -put /ncdc_input
hadoop@bmsccse-HP-Elite-Tower-600-G9-Desktop-PC: ~/hadoop-book/input/ncdc/all$ hdfs dfs -ls /ncdc_input
Found 2 items
-rw-r--r-- 1 hadoop supergroup 73867 2025-05-06 15:11 /ncdc_input/1901.gz
-rw-r--r-- 1 hadoop supergroup 74105 2025-05-06 15:11 /ncdc_input/1982.gz
hadoop@bmsccse-HP-Elite-Tower-600-G9-Desktop-PC: ~/hadoop/book/input/ncdc/all$ hadoop jar AverageTemp.jar temp.AverageDriver /ncdc_input /output_avg
JAR does not exist or is not a normal file: /home/hadoop/hadoop-book/input/ncdc/all/AverageTemp.jar
hadoop@bmsccse-HP-Elite-Tower-600-G9-Desktop-PC: ~/hadoop/book/input/ncdc/all$ cd ~/Desktop
hadoop@bmsccse-HP-Elite-Tower-600-G9-Desktop-PC: ~/Desktop$ hadoop jar temp.AverageDriver /ncdc_input /output_avg
JAR does not exist or is not a normal file: /home/hadoop/Desktop/AverageTemp.jar
hadoop@bmsccse-HP-Elite-Tower-600-G9-Desktop-PC: ~/Desktop$ ls
ACFrqGz1080DXZ9jzb1lxQ8_3_MyTt4Qzqp_teWZ0d59ehdtaf_urnhi-d7nxM55849x4MgjzFx5zAK8tKMyxYNBKEw7y0GFL117d183C8kldjx9GAsRZj/JpZGQKyTrpgSUCU.pdf
file.txt
hadoop@bmsccse-HP-Elite-Tower-600-G9-Desktop-PC: ~$ hadoop jar NDCDWeather.jar temp.AverageDriver /ncdc_input /output_avg
2025-05-06 15:14:57,391 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-05-06 15:14:57,428 INFO impl.MetricSystemImpl: Scheduled Metrics snapshot period at 10 second(s).
2025-05-06 15:14:57,429 INFO impl.MetricSystemImpl: JobTracker metrics system started
2025-05-06 15:14:57,485 WARN mapred.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool Interface and execute your application with ToolRunner to remedy this.
2025-05-06 15:14:57,533 INFO mapred.JobSubmitter: Number of splits:2
2025-05-06 15:14:57,607 INFO mapred.JobSubmitter: Submitting tokens for job: job_local1944593155_0001
2025-05-06 15:14:57,607 INFO mapred.JobSubmitter: Executing with tokens: []
2025-05-06 15:14:57,656 INFO mapred.Job: The url to track the job: http://localhost:8088/
2025-05-06 15:14:57,657 INFO mapred.Job: Running job: job_local1944593155_0001
2025-05-06 15:14:57,657 INFO mapred.LocalJobRunner: OutputCommitter: FileOutputCommitter configured null
2025-05-06 15:14:57,661 INFO mapred.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-05-06 15:14:57,661 INFO output.FileOutputCommitter: File Output Committer skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-05-06 15:14:57,661 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.mrunit.FileOutputCommitter
Activities Terminal May 6 15:26
hadoop@bmsccse-HP-Elite-Tower-600-G9-Desktop-PC: ~/Desktop
hadoop@bmsccse-HP-Elite-Tower-600-G9-Desktop-PC: ~/Desktop$ hadoop jar NDCDWeather.jar temp.AverageDriver /ncdc_input /output_avg
2025-05-06 15:14:57,391 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-05-06 15:14:57,428 INFO impl.MetricSystemImpl: Scheduled Metrics snapshot period at 10 second(s).
2025-05-06 15:14:57,429 INFO impl.MetricSystemImpl: JobTracker metrics system started
2025-05-06 15:14:57,537 INFO input.FileInputFormat: Total input files to process : 2
2025-05-06 15:14:57,537 INFO input.FileInputFormat: InputFormat tokens for job: job_local1944593155_0001
2025-05-06 15:14:57,537 INFO input.FileInputFormat: Execution tokens: []
2025-05-06 15:14:57,537 INFO input.FileInputFormat: Job: The url to track the job: http://localhost:8088/
2025-05-06 15:14:57,537 INFO input.FileInputFormat: Job: Running job: job_local1944593155_0001
2025-05-06 15:14:57,537 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-05-06 15:14:57,601 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-05-06 15:14:57,601 INFO mapred.LocalJobRunner: OutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-05-06 15:14:57,602 INFO mapred.LocalJobRunner: Waiting for map tasks
2025-05-06 15:14:57,699 INFO mapred.LocalJobRunner: Starting task: attempt_local1944593155_m_000000_0
2025-05-06 15:14:57,700 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-05-06 15:14:57,700 INFO output.FileOutputCommitter: File Output Committer skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-05-06 15:14:57,707 INFO mapred.Task: Using ResourceCalculatorProcessTree []
2025-05-06 15:14:57,707 INFO mapred.Task: Using ResourceCalculatorProcessTree []
2025-05-06 15:14:57,708 INFO mapred.MapTask: Processing split: hdfs://localhost:9000/ncdc_input/1902.gz:0+74105
2025-05-06 15:14:57,739 INFO mapred.MapTask: (EQUATOR) 0 kvt 28214396(104857584)
2025-05-06 15:14:57,739 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2025-05-06 15:14:57,739 INFO mapred.MapTask: soft limit at 83886080
2025-05-06 15:14:57,739 INFO mapred.MapTask: hard limit at 164887680
2025-05-06 15:14:57,739 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
2025-05-06 15:14:57,741 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2025-05-06 15:14:57,754 INFO zillio.ZillibFactory: Successfully loaded & initialized native zillib library
2025-05-06 15:14:57,754 INFO compress.CodecPool: Codec name=decompressor (.gz)
2025-05-06 15:14:57,816 INFO mapred.LocalJobRunner: LocalJobRunner: 
2025-05-06 15:14:57,816 INFO mapred.Task: Starting Flush of map output
2025-05-06 15:14:57,816 INFO mapred.Task: Spilling map output
2025-05-06 15:14:57,816 INFO mapred.Task: bufstart = 0; bufend = 59805; bufvoid = 104857600
2025-05-06 15:14:57,816 INFO mapred.Task: kvstart = 26214396(104857584); length = 26257/6553600
2025-05-06 15:14:57,822 INFO mapred.Task: Finished spill 0
2025-05-06 15:14:57,828 INFO mapred.LocalJobRunner: mapreduce.task.attempt_local1944593155_m_000000_0 is done. And is in the process of committing
2025-05-06 15:14:57,828 INFO mapred.LocalJobRunner: mapreduce.task.attempt_local1944593155_m_000000_0 is done.
2025-05-06 15:14:57,829 INFO mapred.Task: Task 'attempt_local1944593155_m_000000_0' done.
2025-05-06 15:14:57,831 INFO mapred.Task: Final Counters for attempt_local1944593155_m_000000_0: Counters: 23
FILE System Counters
FILE: Number of bytes read=75923
FILE: Number of read operations=0
FILE: Number of write operations=0
FILE: Number of large read operations=0
HDFS: Number of bytes read=74105
HDFS: Number of bytes written=0
HDFS: Number of read operations=0
HDFS: Number of large read operations=0
HDFS: Number of write operations=1
HDFS: Number of bytes read erasure-coded=0
Map-Reduce Framework
MAPREDUCE_FRAMEWORK_LOCAL_MAPS=0
map output records=6565
map output bytes=39885

```

```
Activities Terminal May 15 15:26 hadoop@bmseccse-HP-Elite-Tower-600-G9-Desktop-PC: ~/Desktop

Map-Reduce Framework
  Map input records=6365
  Map output records=6365
  Map output bytes=59085
  Map output materialized bytes=72221
  Input split bytes=105
  Combine input records=0
  Spilled Records=6595
  Failed records=0
  Merged Map outputs=0
  GC Time elapsed (ms)=0
    Total committed heap usage (bytes)=526385152
File Input Format Counters
  Bytes Read=74105
2025-05-06 15:14:57.831 INFO mapred.LocalJobRunner: Finishing task: attempt_local1944593155_0001_n_000001_0
2025-05-06 15:14:57.831 INFO mapred.LocalJobRunner: Starting task: attempt_local1944593155_0001_n_000001_0
2025-05-06 15:14:57.832 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-05-06 15:14:57.832 INFO mapred.Task: Using ResourceCalculatorProcessTree : []
2025-05-06 15:14:57.833 INFO mapred.Task: Processing splits=1
2025-05-06 15:14:57.833 INFO mapred.Task: (attempt_local1944593155_0001_n_000001) (attempt_local1944593155_0001_n_000001)
2025-05-06 15:14:57.859 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2025-05-06 15:14:57.859 INFO mapred.MapTask: soft limit at 8388608
2025-05-06 15:14:57.859 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
2025-05-06 15:14:57.859 INFO mapred.MapTask: kvstart = 26234396; length = 6533000
2025-05-06 15:14:57.887 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2025-05-06 15:14:57.887 INFO mapred.MapTask: Starting flush of map output
2025-05-06 15:14:57.887 INFO mapred.MapTask: Spilling map output
2025-05-06 15:14:57.887 INFO mapred.MapTask: bufstart = 0; bufvoid = 59076; bufvoid = 104857600
2025-05-06 15:14:57.887 INFO mapred.MapTask: kvstart = 26234396(104857584); kvend = 20188144(104752576); length = 26233/6553000
2025-05-06 15:14:57.892 INFO mapred.Task: Task attempt_local1944593155_0001_n_000001_0 killed!
2025-05-06 15:14:57.892 INFO mapred.Task: Task attempt_local1944593155_0001_n_000001_0 is done. And is in the process of committing
2025-05-06 15:14:57.894 INFO mapred.LocalJobRunner: map
2025-05-06 15:14:57.894 INFO mapred.Task: Task attempt_local1944593155_0001_n_000001_0 done.
2025-05-06 15:14:57.895 INFO mapred.Task: Final Counters for attempt_local1944593155_0001_n_000001_0: Counters: 23
  File System Counters
    FILE: Number of bytes read=9102
    FILE: Number of bytes written=792600
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=14972
    HDFS: Number of read operations=2
    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=6365
  Map output records=6364
  Map output bytes=59076
  Map output materialized bytes=72210
  Input split bytes=105
  Combine input records=0
  Call limit Report=459d
```

```
Activities Terminal May 6 15:26 hadoop@bmsece-HP-Elite-Tower-600-G9-Desktop-PC: ~/Desktop
Input split bytes=165
Combine input records=0
Splitted Records=0564
Fallen bytes=0
Merged Map outputs=0
GC time elapsed (ms)=0
Total committed heap usage (bytes)=526395152
FILE Input Format Counters
  FILE Input Format Counters
  2025-05-06 15:14:57.895 INFO mapred.LocalJobRunner: Finishing task: attempt_local1944593155_0001_r_000001_0
  2025-05-06 15:14:57.897 INFO mapred.LocalJobRunner: map task executor complete.
  2025-05-06 15:14:57.897 INFO mapred.LocalJobRunner: Waiting for reduce tasks
  2025-05-06 15:14:57.897 INFO mapred.LocalJobRunner: Starting task: attempt_local1944593155_0001_r_000000_0
  2025-05-06 15:14:57.904 INFO output.FileOutputCommitter: FileOutputCommitter algorithm version is 1
  2025-05-06 15:14:57.904 INFO output.FileOutputCommitter: Using configuration settings temporary folders under output directory:false, ignore cleanup failures: false
  2025-05-06 15:14:57.905 INFO mapred.Task: Using ResourcecalculatorForMapred: []
  2025-05-06 15:14:57.905 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@777e85
  2025-05-06 15:14:57.915 INFO reduce.MergeManagerImpl: JobTracker metrics system already initialized
  2025-05-06 15:14:57.915 INFO reduce.MergeManagerImpl: MergerManager memoryLimit=5827985488, maxSingleShuffleFileInitialSize=50960320, mergeThreshold=3840479460, noSortFactor=10, memToMemMergeOutputsThreshold=10
  2025-05-06 15:14:57.915 INFO reduce.MergeManagerImpl: MergerManager has been started; for fetching Map Completion Events
  2025-05-06 15:14:57.920 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map attempt_local1944593155_0001_m_000001_0 decompt: 7220 len: 72210 to MEMORY
  2025-05-06 15:14:57.928 INFO reduce.InMemoryMapOutput: Read 72206 bytes from map-output for attempt_local1944593155_0001_m_000001_0
  2025-05-06 15:14:57.929 INFO reduce.MergeManagerImpl: closeInMemoryFile --> map-output of size: 72206, inMemoryMapOutputs.size() -> 1, commitMemory --> 0, usedMemory --> 72206
  2025-05-06 15:14:57.929 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map attempt_local1944593155_0001_m_000000_0 decompt: 72217 len: 72221 to MEMORY
  2025-05-06 15:14:57.930 INFO reduce.MergeManagerImpl: closeInMemoryFile --> map-output of size: 72217, inMemoryMapOutputs.size() -> 2, commitMemory --> 72206, usedMemory --> 144423
  2025-05-06 15:14:57.930 INFO EventFetcher: EventFetcher is interrupted. Returning
  2025-05-06 15:14:57.930 WARN org.mapred.JobClient: Failed readAhead on file
EBADF: Bad file descriptor
  at org.apache.hadoop.io.nativeio.NativeIO$Posix$posix_fadviseNativeMethod()
  at org.apache.hadoop.io.nativeio.NativeIO$Posix$posix_fadviseNative$0(NativeIO.java:419)
  at org.apache.hadoop.io.nativeio.NativeIO$Posix$posixCacheManipulator$posix_fadviseNative$0(NativeIO.java:296)
  at org.apache.hadoop.io.FadvisePool$ReadaheadPoolRequestImpl.run(ReadaheadPool.java:226)
  at java.base/java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1128)
  at java.base/java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:628)
  at java.base/java.lang.Thread.run(Thread.java:834)
  2025-05-06 15:14:57.936 INFO mapred.LocalJobRunner: / 2 copied.
  2025-05-06 15:14:57.931 INFO reduce.MergeManagerImpl: finalMerge called with 2 in-memory map-outputs and 0 on-disk map-outputs
  2025-05-06 15:14:57.933 INFO mapred.Merger: Merging 2 sorted segments
  2025-05-06 15:14:57.933 INFO mapred.Merger: Down to the last merge-pass, with 2 segments left of total size: 144409 bytes
  2025-05-06 15:14:57.933 INFO reduce.MergeManagerImpl: reduce.MergerManagerImpl: 2 segments, 144409 bytes to disk to satisfy reduce memory limit
  2025-05-06 15:14:57.938 INFO reduce.MergeManagerImpl: Reducing 2 segments, 144409 bytes to disk
  2025-05-06 15:14:57.938 INFO reduce.MergeManagerImpl: Merging 2 segments, 144409 bytes to disk
  2025-05-06 15:14:57.938 INFO mapred.Merger: Merging 1 sorted segments
  2025-05-06 15:14:57.939 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 144414 bytes
  2025-05-06 15:14:57.939 INFO mapred.LocalJobRunner: / 2 copied.
  2025-05-06 15:14:57.940 INFO mapred.Merger: Reducing 1 sorted segments, 144414 bytes to disk
  2025-05-06 15:14:58.022 INFO mapred.Task: Task attempt_local1944593155_0001_r_000000_0 is done. And it is in the process of committing
  2025-05-06 15:14:58.023 INFO mapred.LocalJobRunner: / 2 copied.
  2025-05-06 15:14:58.023 INFO mapred.Task: Task attempt_local1944593155_0001_r_000000_0 is allowed to commit now
  2025-05-06 15:14:58.038 INFO mapred.Task: Task attempt_local1944593155_0001_r_000000_0 is allowed to commit now
  2025-05-06 15:14:58.038 INFO mapred.Task: Task attempt_local1944593155_0001_r_000000_0 is done.
  2025-05-06 15:14:58.038 INFO mapred.Task: Final counters for attempt_local1944593155_0001_r_000000_0: Counters: 30
  FILE System Counters
    FILE System Counters
      Number of bytes read=297022
```

```

Activities Terminal May 6 15:27
hadoop@bmseccse-HP-Elite-Tower-600-G9-Desktop-PC: ~/Desktop

2025-05-06 15:14:58,038 INFO mapred.Task: Final Counters for attempt_local1944593155_0001_r_000000_0: Counters: 30
  FILE: Number of bytes read=297022
  FILE: Number of bytes written=936493
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=147972
  HDFS: Number of bytes written=16
  HDFS: Number of read operations=12
  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=2
  Total compressed input bytes=244431
  Reduce Input records=13129
  Reduce output records=2
  Spilled Records=13129
  Shuffled Maps =2
  Failed Shuffles=0
  Merged Map outputs=2
  GC time elapsed (ms)=2
  Total committed heap usage (bytes)=633339904
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=72
2025-05-06 15:14:58,039 INFO mapred.LocalJobRunner: Finishing task; attempt_local1944593155_0001_r_000000_0
2025-05-06 15:14:58,059 INFO mapred.LocalJobRunner: reduce task executor complete
2025-05-06 15:14:58,059 INFO mapred.Job: Job job_local1944593155_0001 running in uber mode : false
2025-05-06 15:14:58,060 INFO mapred.Job: map 100% reduce 100%
2025-05-06 15:14:58,060 INFO mapred.Job: Job job_local1944593155_0001 completed successfully
2025-05-06 15:14:58,060 INFO mapred.Job: Counters: 36
  File System Counters
    FILE: Number of bytes read=312987
    FILE: Number of bytes written=2448387
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=297022
    HDFS: Number of bytes written=16
    HDFS: Number of read operations=24
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=5
    HDFS: Number of bytes read erasure-coded=0
  Map-Reduce framework
    Map input records=13129
    Map output records=13129
    Map output bytes=13129

Activities Terminal May 6 15:27
hadoop@bmseccse-HP-Elite-Tower-600-G9-Desktop-PC: ~/Desktop

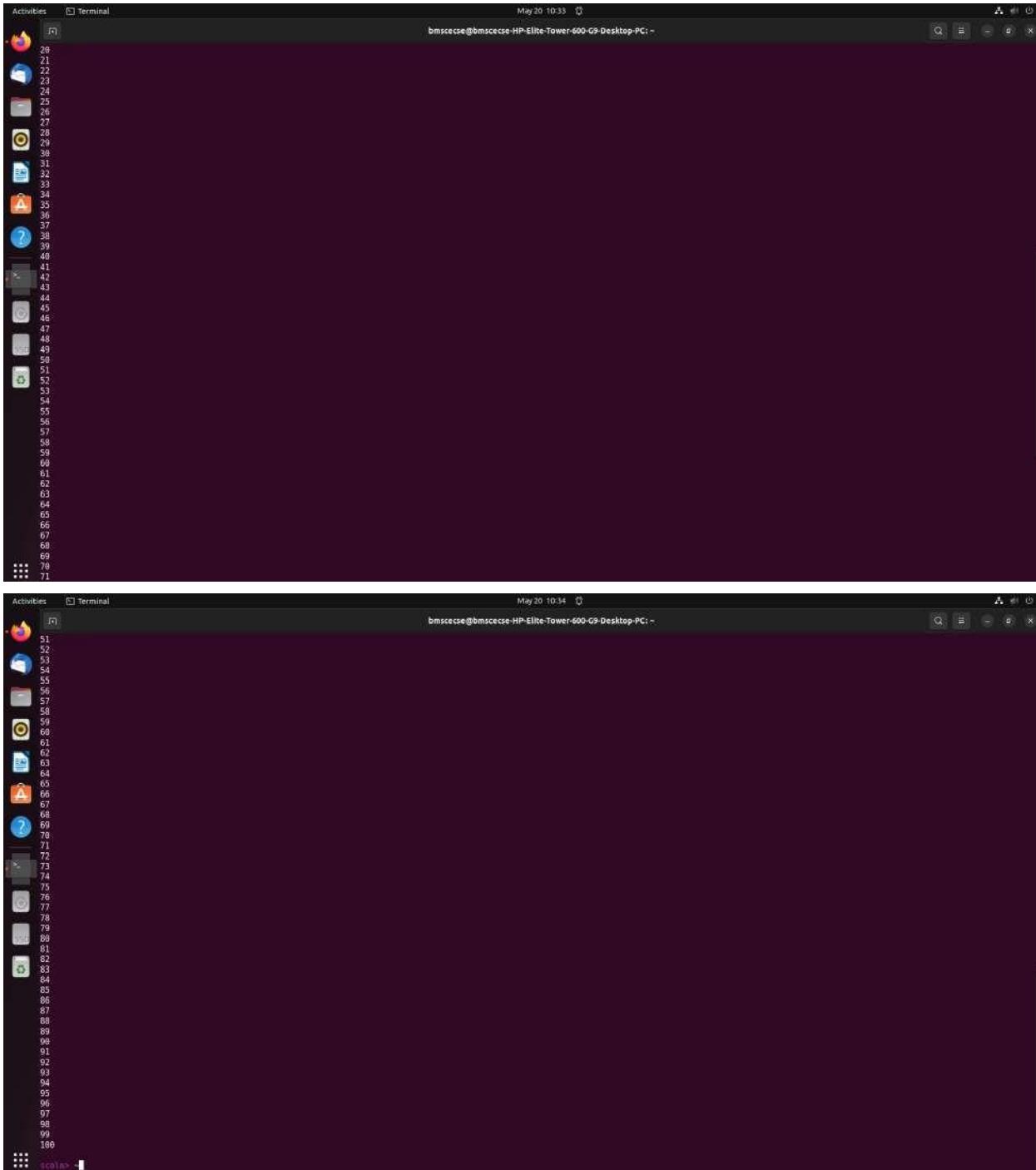
Reduce output records=12
Spilled Records=26288
Shuffled Maps =2
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=3
Total committed heap usage (bytes)=1686110288
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
File Input Format Counters
  Bytes Written=72
hadoop@bmseccse-HP-Elite-Tower-600-G9-Desktop-PC: ~/Desktop$ hdfs dfs -ls /output_avg
Found 2 items
-rw-r--r-- 1 hadoop supergroup 0 2025-05-06 15:14 /output_avg/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 16 2025-05-06 15:14 /output_avg/part-r-00000
hadoop@bmseccse-HP-Elite-Tower-600-G9-Desktop-PC: ~/Desktop$ hdfs dfs -cat /output_avg/part-r-00000
1901 46
1902 21
hadoop@bmseccse-HP-Elite-Tower-600-G9-Desktop-PC: ~/Desktop$ hdfs dfs -ls /output_avg
Found 2 items
-rw-r--r-- 1 hadoop supergroup 0 2025-05-06 15:14 /output_avg/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 16 2025-05-06 15:14 /output_avg/part-r-00000
hadoop@bmseccse-HP-Elite-Tower-600-G9-Desktop-PC: ~/Desktop$ hdfs dfs -cat /output_avg/part-r-00000
1901 46
1902 21
hadoop@bmseccse-HP-Elite-Tower-600-G9-Desktop-PC: ~/Desktop$ hdfs dfs -cat /output_mean/part-r-00000
01 4
02 1
03 0
04 34
05 89
06 143
07 182
08 172
09 123
10 73
11 21
12 3
hadoop@bmseccse-HP-Elite-Tower-600-G9-Desktop-PC: ~/Desktop$ hadoop@bmseccse-HP-Elite-Tower-600-G9-Desktop-PC: ~/Desktop$ hdfs dfs -cat /output_avg/part-r-00000
1901 46
1902 21
hadoop@bmseccse-HP-Elite-Tower-600-G9-Desktop-PC: ~/Desktop$ hdfs dfs -cat /output_avg/part-r-00000
1901 46
1902 21

```

## EXPERIMENT-8

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

## Code with screenshots:



## EXPERIMENT-9

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 Screenshots:

```
bnscecse@bnscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ spark-shell
25/05/20 15:32:38 WARN Utils: Your hostname, bnscecse-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


$$\begin{array}{c} \diagup \diagdown \\ \diagup \diagdown \\ \diagup \diagdown \\ \diagup \diagdown \\ \diagup \diagdown \end{array}$$
 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/bnscecse/Desktop/sparkdata.txt")
textFile: org.apache.spark.rdd.RDD[String] = /home/bnscecse/Desktop/sparkdata.txt MapPartitionsRDD[1] at textFile at <console>:2
4

scala>

scala> val counts = textFile
counts: org.apache.spark.rdd.RDD[String] = /home/bnscecse/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)
```

## EXPERIMENT-10

Write a simple streaming program in Spark to receive text data streams on a particular port, perform basic text cleaning (like white space removal, stop words removal, lemmatization, etc.), and print the cleaned text on the screen. (Open Ended Question).

Code with Screenshots:

```
import org.apache.spark.SparkConf  
  
import org.apache.spark.streaming.{Seconds, StreamingContext}  
  
import org.apache.spark.streaming.dstream.DStream  
  
  
object StreamingTextCleaner {  
  
    val stopWords = Set(  
        "a", "an", "the", "is", "are", "was", "were", "this", "that",  
        "on", "in", "and", "or", "of", "to", "for", "with", "as", "by"  
    )  
  
  
    def simpleLemmatize(word: String): String = {  
        if (word.endsWith("ing")) word.dropRight(3)  
        else if (word.endsWith("ed")) word.dropRight(2)  
        else word  
    }  
}
```

```

def main(args: Array[String]): Unit = {
    if (args.length < 2) {
        System.err.println("Usage: StreamingTextCleaner <hostname> <port>")
        System.exit(1)
    }

    val hostname = args(0)
    val port = args(1).toInt

    val conf = new SparkConf().setAppName("StreamingTextCleaner").setMaster("local[*]")
    val ssc = new StreamingContext(conf, Seconds(5))

    val lines = ssc.socketTextStream(hostname, port)

    val cleanedLines: DStream[String] = lines.map(_.toLowerCase)
        .map(_.trim)
        .map(line => line.replaceAll("[^a-zA-Z\\s]", ""))
        .flatMap(_.split("\\s+"))
        .filter(word => word.nonEmpty && !stopWords.contains(word))
        .map(simpleLemmatize)
        .filter(_.nonEmpty)
        .reduceByWindow(

```

```

(a: String, b: String) => a + " " + b,
Seconds(5),
Seconds(5)

)

cleanedLines.print()

ssc.start()
ssc.awaitTermination()
}
}

```

```

Requirement already satisfied: nltk in /usr/local/lib/python3.11/dist-packages (3.9.1)
Requirement already satisfied: click in /usr/local/lib/python3.11/dist-packages (from nltk) (8.2.0)
Requirement already satisfied: joblib in /usr/local/lib/python3.11/dist-packages (from nltk) (1.5.0)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.11/dist-packages (from nltk) (2024.11.6)
Requirement already satisfied: tqdm in /usr/local/lib/python3.11/dist-packages (from nltk) (4.67.1)
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
[nltk_data] Downloading package wordnet to /root/nltk_data...
+---+
|word |
+---+
|hello|
|hate |
|hate |
|love |
|dont |
|want |
|cant |
|put |
|nobody|
|else |
+---+

```