VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"JnanaSangama", Belgaum -590014, Karnataka.

LAB REPORT
on

Big Data Analytics(23CS6PCBDA)

Submitted by

Maria Sayeema(1BM22CS151)

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 **Maria Sayeema(1BM22CS151),** 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.

Prof.Amrutha 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	Perform the following DB operations using Cassandra. a) Create a keyspace by name Employee b) Create a column family by name Employee-Info with attributes Emp_Id Primary Key, Emp_Name, Designation, Date_of_Joining, Salary, Dept_Name c) Insert the values into the table in batch d) Update Employee name and Department of Emp-Id 121 e) Sort the details of Employee records based on salary 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. g) Update the altered table to add project names. h) Create a TTL of 15 seconds to display the values of Employees.	27
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	30

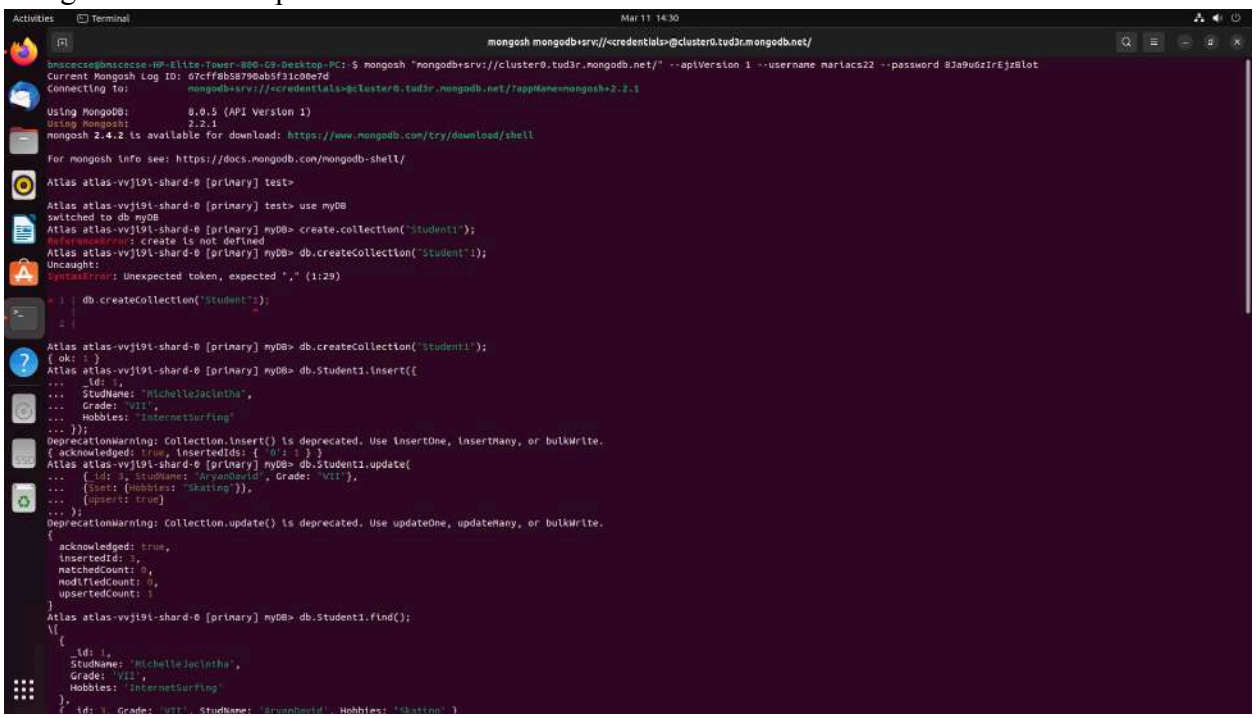
	g) Import a given csv dataset from local file system into Cassandra column family	
5	Execution of HDFS Commands for interaction with Hadoop Environment. (Minimum 10 commands to be executed)	37
6	Implement Wordcount program on Hadoop framework	41
7	From the following link extract the weather data https://github.com/tomwhite/hadoop-book/tree/master/input/ncdc/all 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.	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

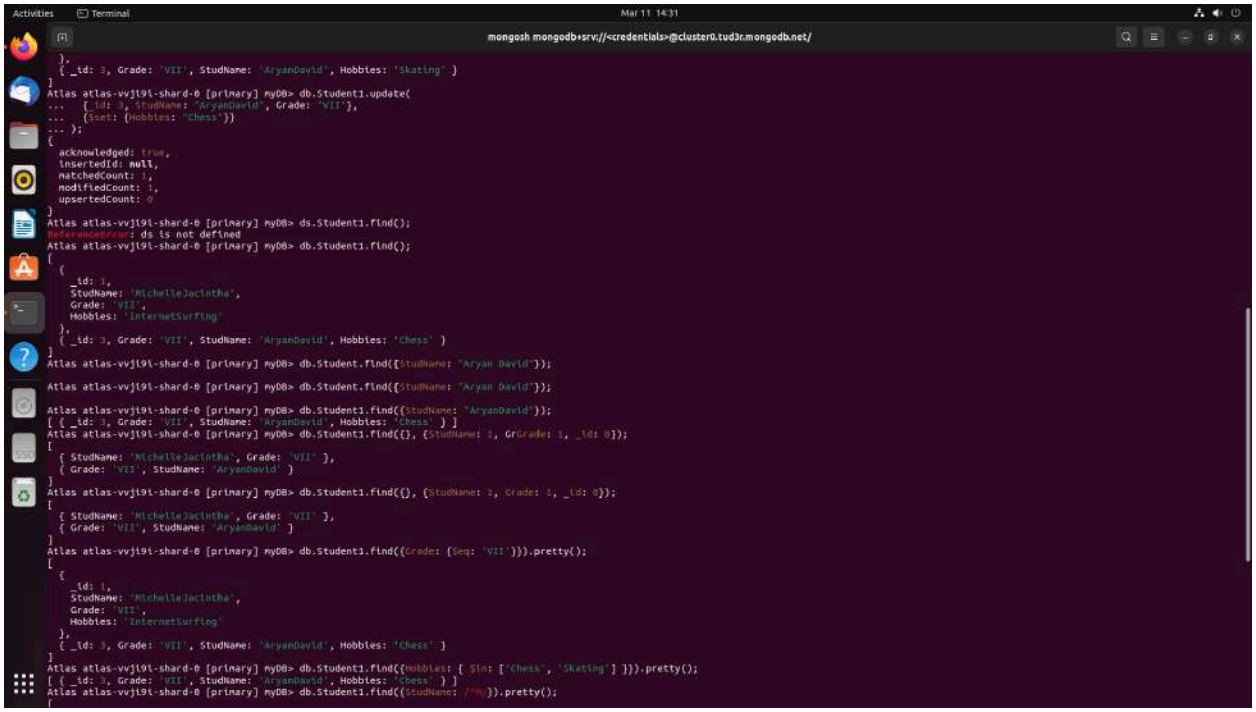
Course Outcome

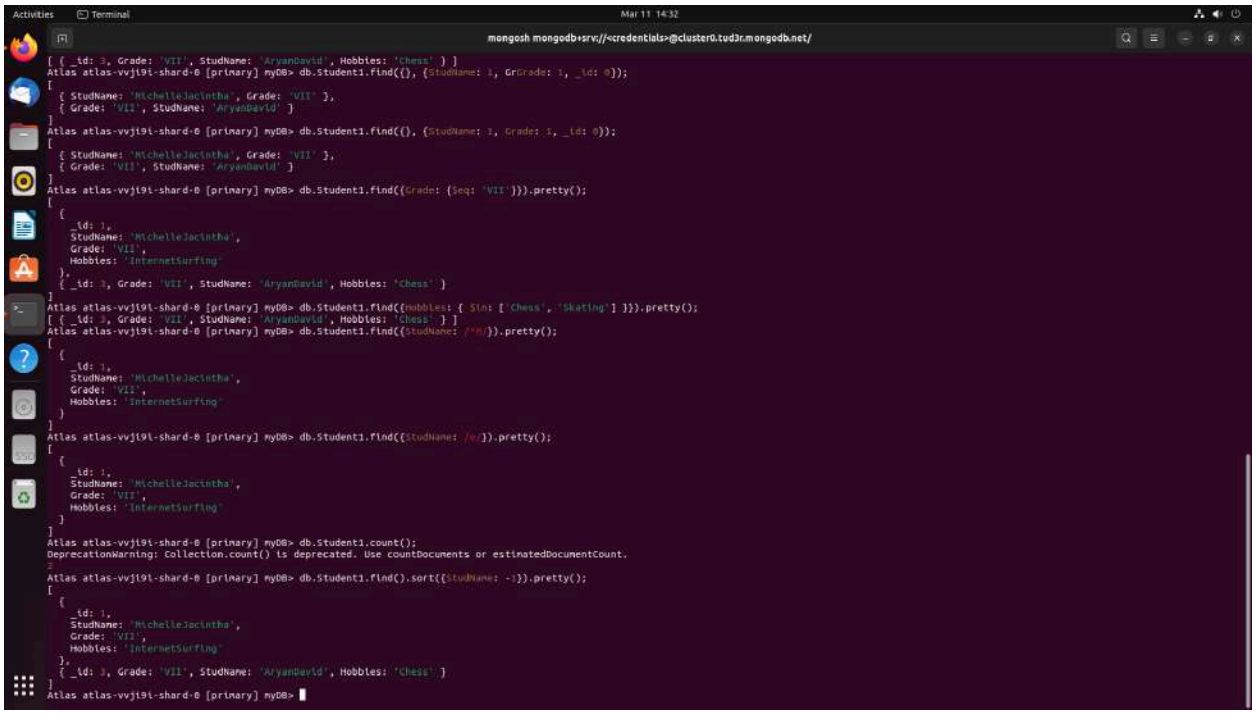
CO1	Apply the concept of NoSQL, Hadoop or Spark for a given task
CO2	Analyse big data analytics mechanisms that can be applied to obtain solution for a given problem.
CO3	Design and implement solutions using data analytics mechanisms for a given problem.

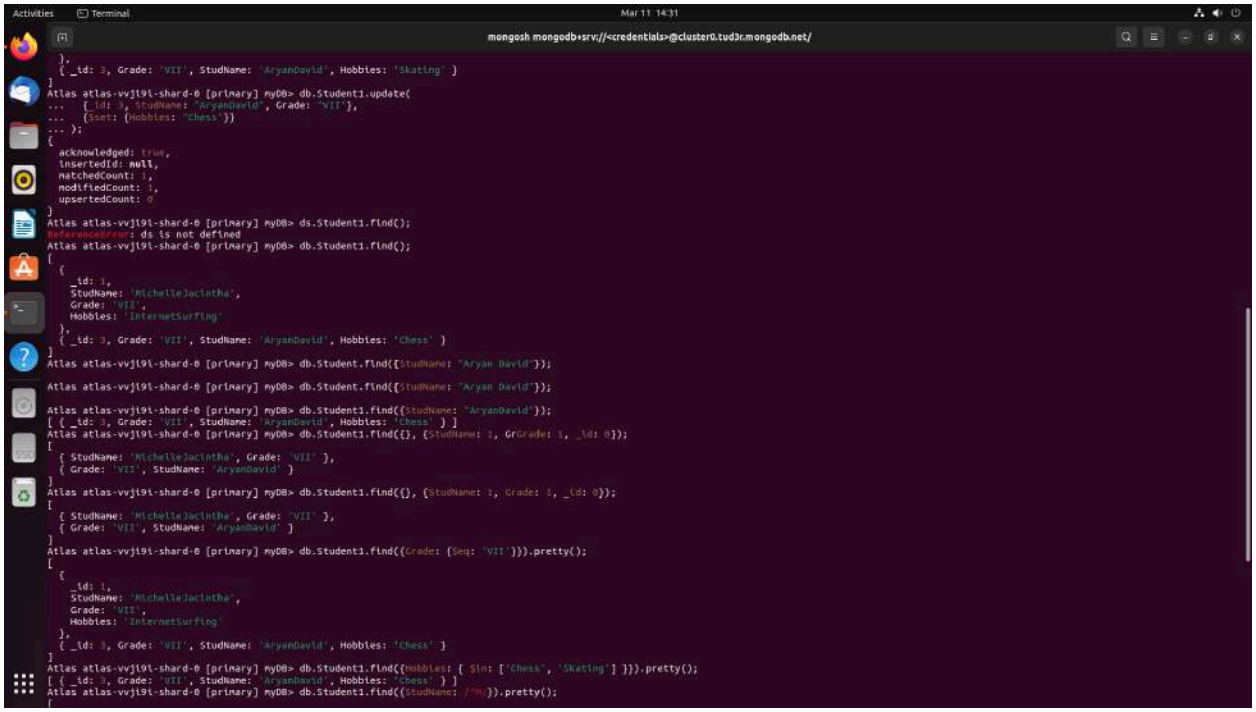
Experiment - 1

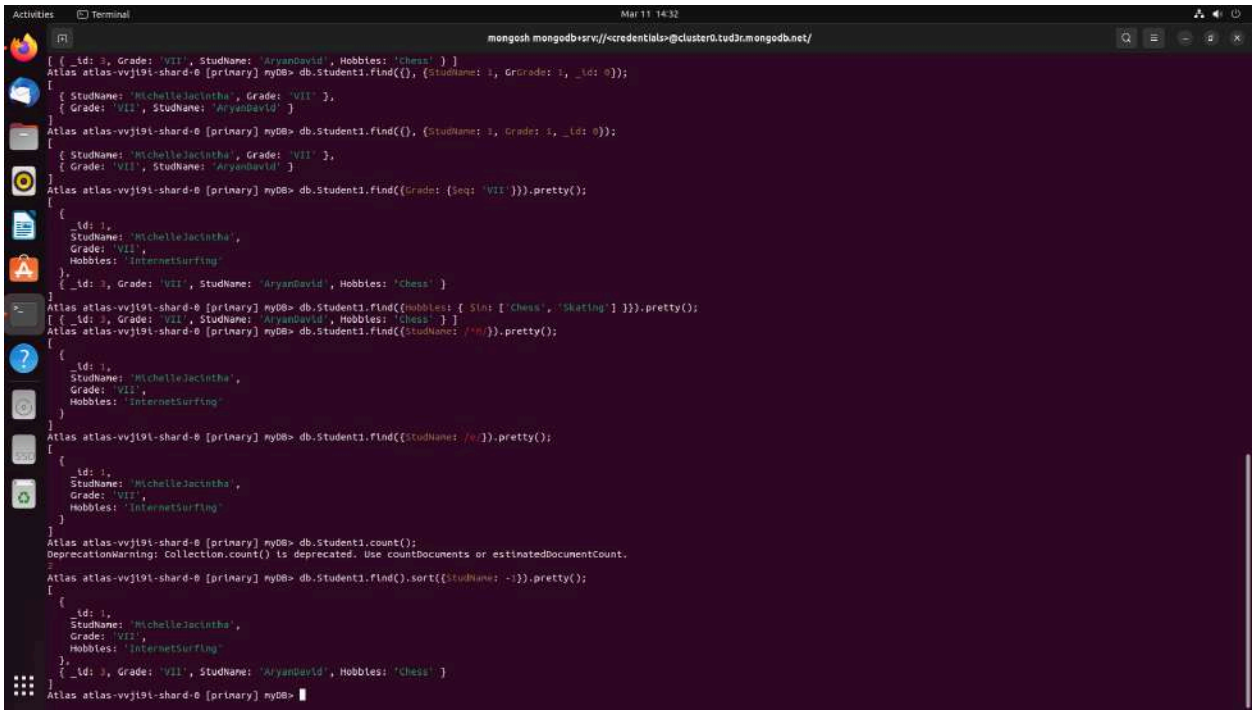
MongoDB- CRUD Operations Demonstration

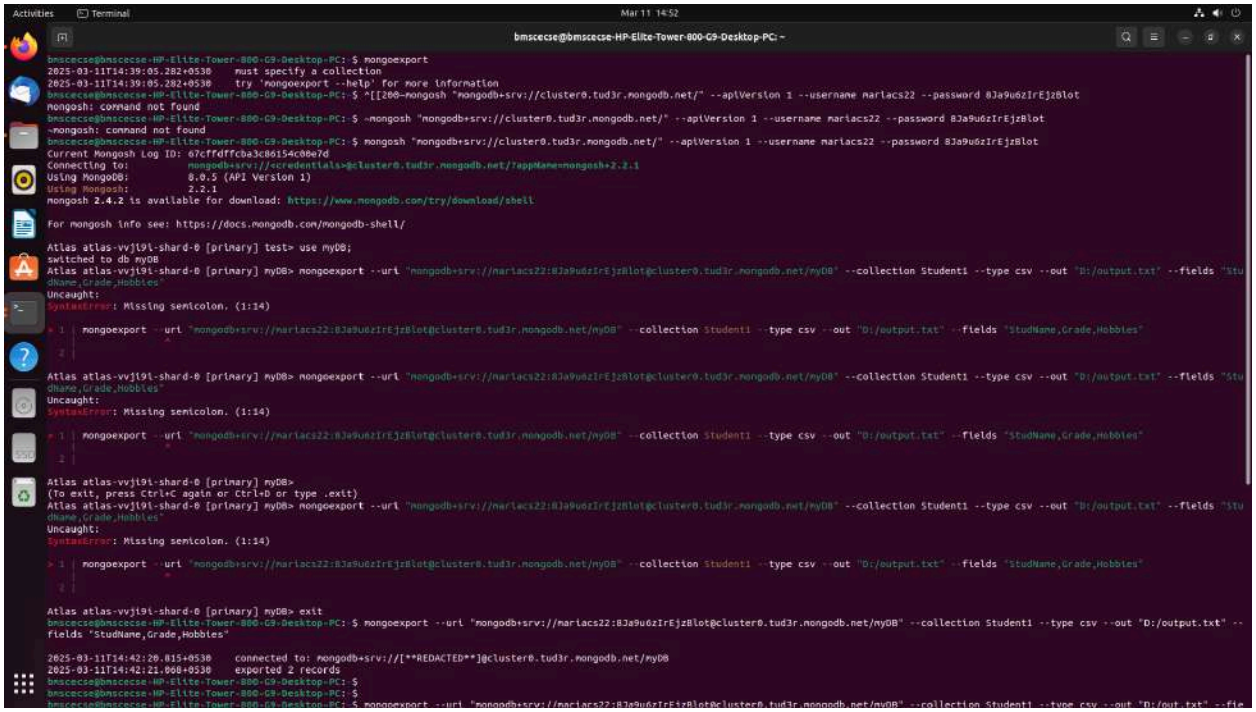


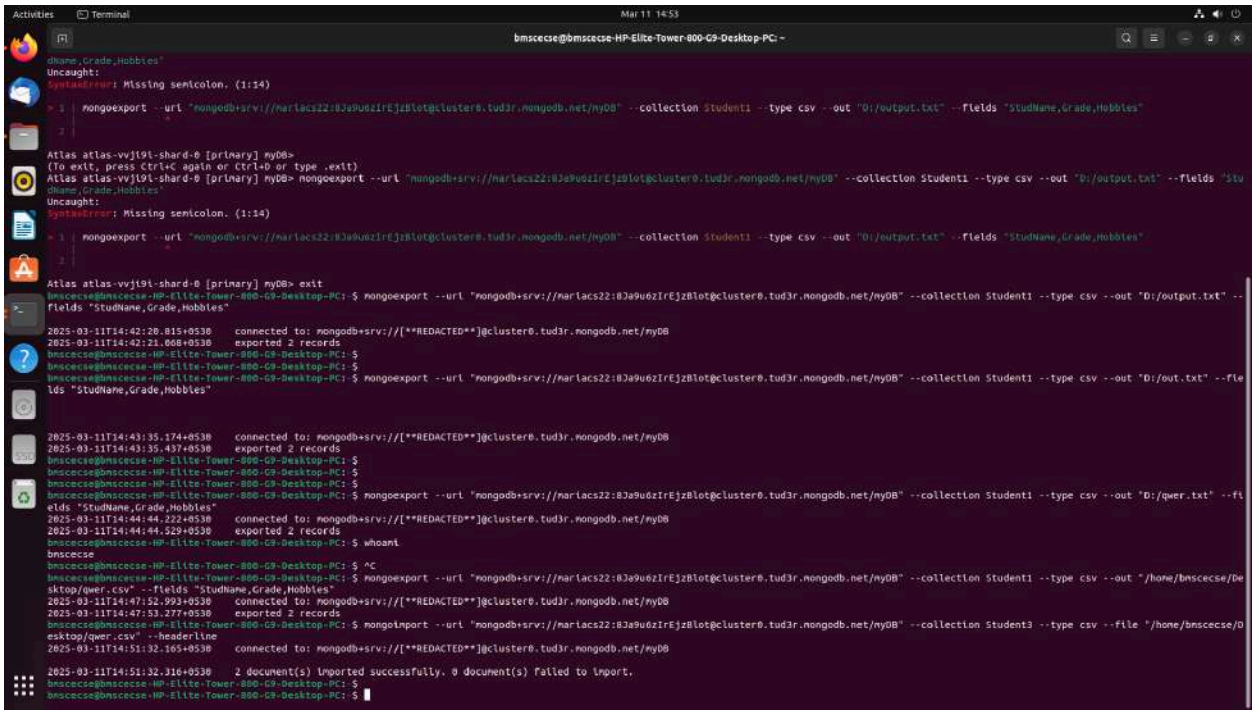

Activities
Terminal
Mar 11 14:31
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
},
{ _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Skating' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.update(
... {_id: 3, StudName: "AryanDavid", Grade: "VII"},
... {$set: {Hobbies: "Chess"}}
...);
{
acknowledged: true,
insertedId: null,
matchedCount: 1,
modifiedCount: 1,
upsertedCount: 0
}
Atlas atlas-vvji9i-shard-0 [primary] myDB> ds.Student1.find();
ReferenceError: ds is not defined
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find();
[
{
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
},
{ _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student.find({StudName: "Aryan David"});
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student.find({StudName: "Aryan David"});
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({StudName: "AryanDavid"});
[{ _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' }]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({}, {StudName: 1, GrGrade: 1, _id: 0});
[
{ StudName: 'MichelleJacintha', Grade: 'VII' },
{ Grade: 'VII', StudName: 'AryanDavid' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({}, {StudName: 1, Grade: 1, _id: 0});
[
{ StudName: 'MichelleJacintha', Grade: 'VII' },
{ Grade: 'VII', StudName: 'AryanDavid' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({Grade: {$eq: 'VII'}}).pretty();
[
{
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
},
{ _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({Hobbies: { $in: ['Chess', 'Skating'] }}).pretty();
[{ _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' }]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({StudName: /^M/}).pretty();
[

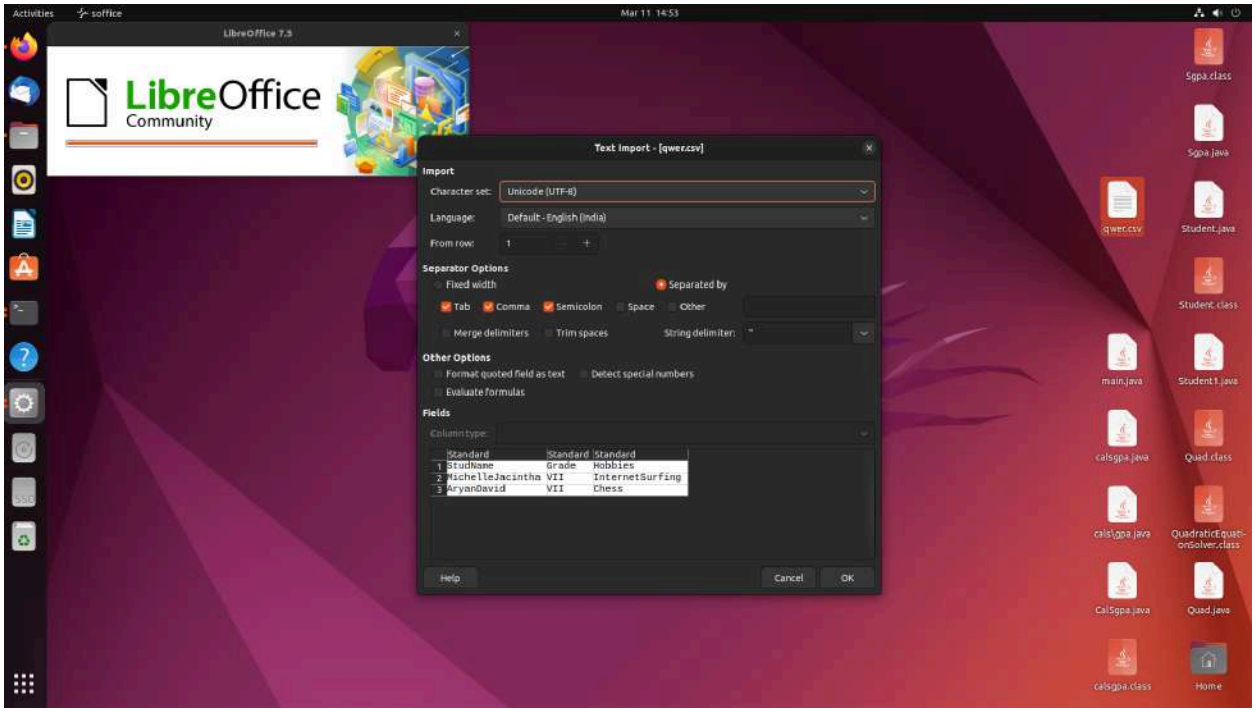

Activities
Terminal
Mar 11 14:32
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
[{ _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' }]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({}, {StudName: 1, GrGrade: 1, _id: 0});
[
{ StudName: 'MichelleJacintha', Grade: 'VII' },
{ Grade: 'VII', StudName: 'AryanDavid' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({}, {StudName: 1, Grade: 1, _id: 0});
[
{ StudName: 'MichelleJacintha', Grade: 'VII' },
{ Grade: 'VII', StudName: 'AryanDavid' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> 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-vvji9i-shard-0 [primary] myDB> db.Student1.find({Hobbies: { $in: ['Chess', 'Skating'] }}).pretty();
[{ _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' }]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({StudName: /^M/}).pretty();
[
{
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
}
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({StudName: /e/}).pretty();
[
{
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
}
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.count();
DeprecationWarning: Collection.count() is deprecated. Use countDocuments or estimatedDocumentCount.
3
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find().sort({StudName: -1}).pretty();
[
{
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
},
{ _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB>

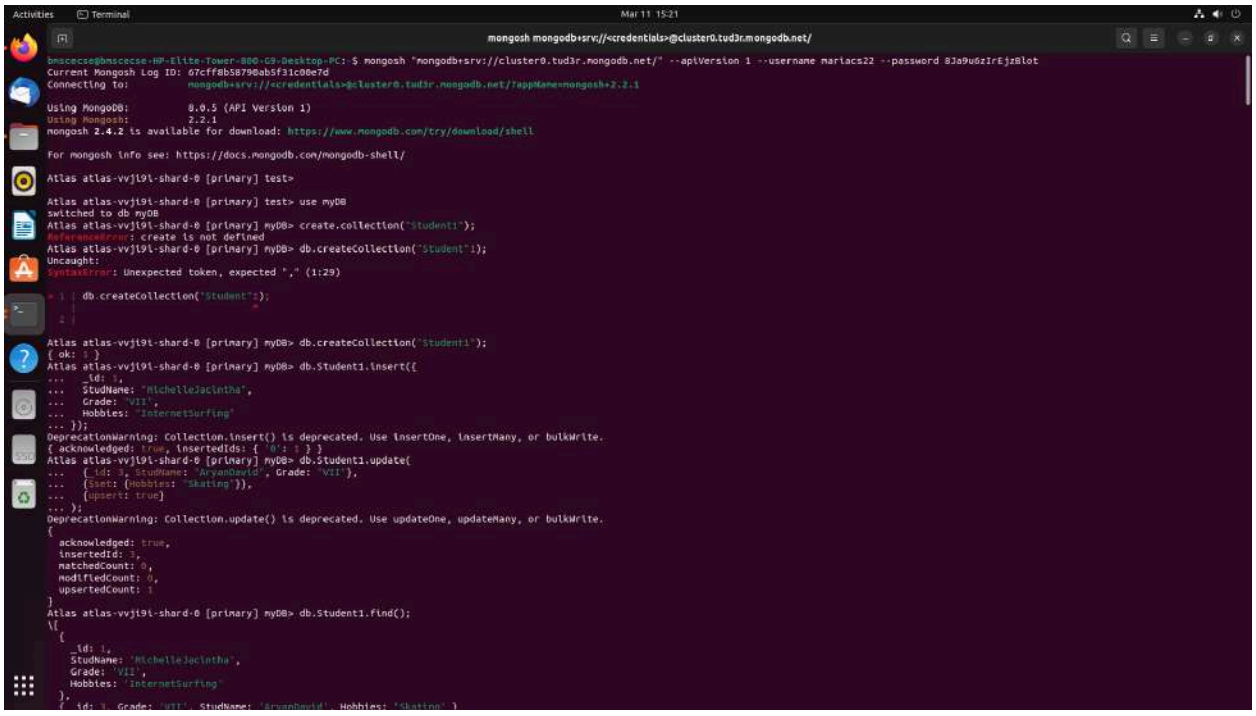

Activities
Terminal
Mar 11 14:31
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
},
{ _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Skating' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.update(
... {_id: 3, StudName: "AryanDavid", Grade: "VII"},
... {$set: {Hobbies: "Chess"}}
...);
{
acknowledged: true,
insertedId: null,
matchedCount: 1,
modifiedCount: 1,
upsertedCount: 0
}
Atlas atlas-vvji9i-shard-0 [primary] myDB> ds.Student1.find();
ReferenceError: ds is not defined
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find();
[
{
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
},
{ _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student.find({StudName: "Aryan David"});
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student.find({StudName: "Aryan David"});
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({StudName: "AryanDavid"});
[{ _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' }]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({}, {StudName: 1, GrGrade: 1, _id: 0});
[
{ StudName: 'MichelleJacintha', Grade: 'VII' },
{ Grade: 'VII', StudName: 'AryanDavid' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({}, {StudName: 1, Grade: 1, _id: 0});
[
{ StudName: 'MichelleJacintha', Grade: 'VII' },
{ Grade: 'VII', StudName: 'AryanDavid' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({Grade: {$eq: 'VII'}}).pretty();
[
{
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
},
{ _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({Hobbies: { $in: ['Chess', 'Skating'] }}).pretty();
[{ _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' }]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({StudName: /^M/}).pretty();

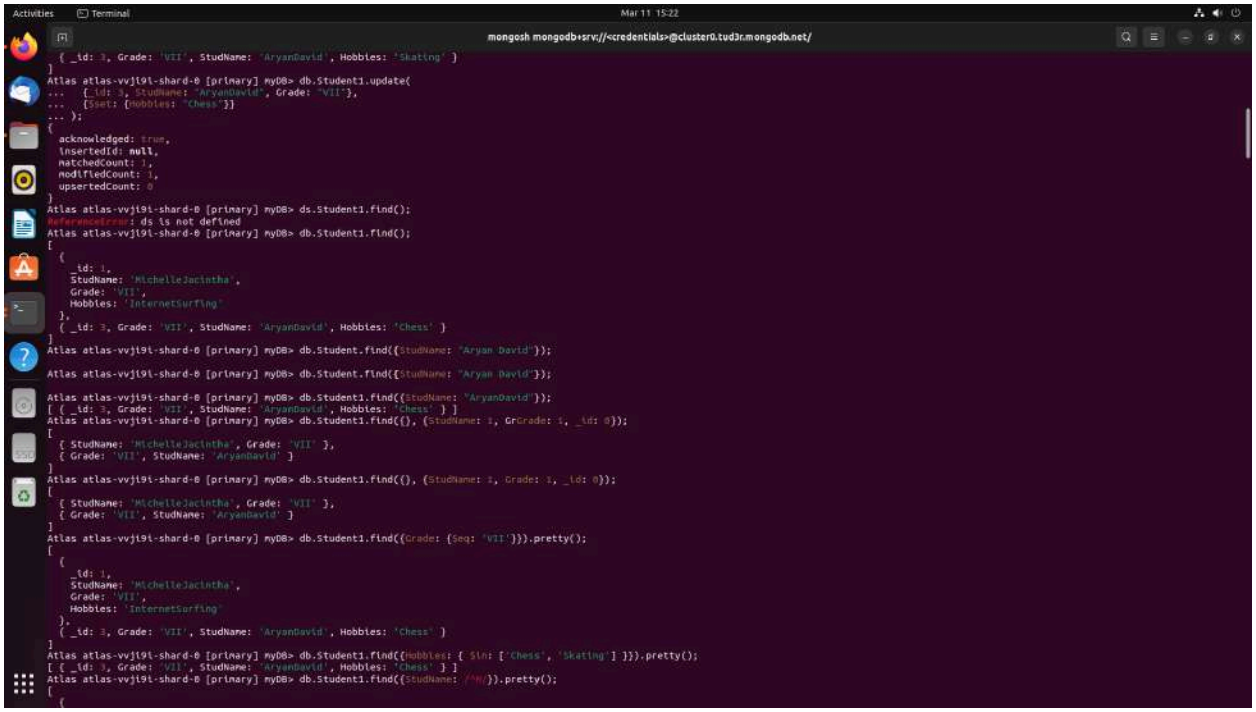

Activities
Terminal
Mar 11 14:32
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
[{ _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' }]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({}, {StudName: 1, GrGrade: 1, _id: 0});
[
{ StudName: 'MichelleJacintha', Grade: 'VII' },
{ Grade: 'VII', StudName: 'AryanDavid' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({}, {StudName: 1, Grade: 1, _id: 0});
[
{ StudName: 'MichelleJacintha', Grade: 'VII' },
{ Grade: 'VII', StudName: 'AryanDavid' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> 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-vvji9i-shard-0 [primary] myDB> db.Student1.find({Hobbies: { $in: ['Chess', 'Skating'] }}).pretty();
[{ _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' }]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({StudName: /^M/}).pretty();
[
{
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
}
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({StudName: /e/}).pretty();
[
{
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
}
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.count();
DeprecationWarning: Collection.count() is deprecated. Use countDocuments or estimatedDocumentCount.
2
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find().sort({StudName: -1}).pretty();
[
{
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
},
{ _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB>

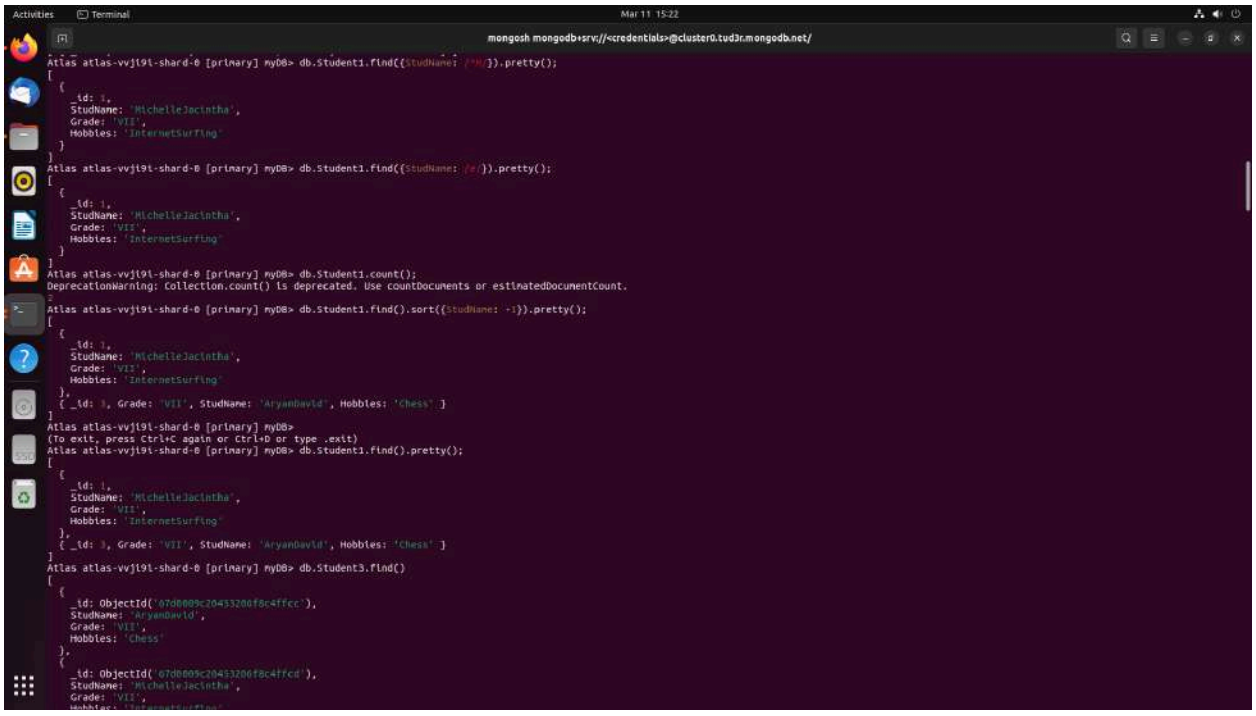

Activities
Terminal
Mar 11 14:52
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ mongoexport
2025-03-11T14:39:05.282+0530 must specify a collection
2025-03-11T14:39:05.282+0530 try 'mongoexport --help' for more information
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ ^[[200~mongosh "mongodb+srv://cluster0.tud3r.mongodb.net/" --apiVersion 1 --username mariacs22 --password 8Ja9u6zIrEjzBlot
mongosh: command not found
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ ~mongosh "mongodb+srv://cluster0.tud3r.mongodb.net/" --apiVersion 1 --username mariacs22 --password 8Ja9u6zIrEjzBlot
~mongosh: command not found
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ mongosh "mongodb+srv://cluster0.tud3r.mongodb.net/" --apiVersion 1 --username mariacs22 --password 8Ja9u6zIrEjzBlot
Current Mongosh Log ID: 67cffdffcba3c86154c08e7d
Connecting to: mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/?appName=mongosh+2.2.1
Using MongoDB: 8.0.5 (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-vvji9i-shard-0 [primary] test> use myDB;
switched to db myDB
Atlas atlas-vvji9i-shard-0 [primary] myDB> mongoexport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudName,Grade,Hobbies"
Uncaught:
SyntaxError: Missing semicolon. (1:14)
> 1 | mongoexport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudName,Grade,Hobbies"
| ^
2 |
Atlas atlas-vvji9i-shard-0 [primary] myDB> mongoexport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudName,Grade,Hobbies"
Uncaught:
SyntaxError: Missing semicolon. (1:14)
> 1 | mongoexport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudName,Grade,Hobbies"
| ^
2 |
Atlas atlas-vvji9i-shard-0 [primary] myDB>
(To exit, press Ctrl+C again or Ctrl+D or type .exit)
Atlas atlas-vvji9i-shard-0 [primary] myDB> mongoexport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudName,Grade,Hobbies"
Uncaught:
SyntaxError: Missing semicolon. (1:14)
> 1 | mongoexport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudName,Grade,Hobbies"
| ^
2 |
Atlas atlas-vvji9i-shard-0 [primary] myDB> exit
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ mongoexport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudName,Grade,Hobbies"
2025-03-11T14:42:20.815+0530 connected to: mongodb+srv://[**REDACTED**]@cluster0.tud3r.mongodb.net/myDB
2025-03-11T14:42:21.068+0530 exported 2 records
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $

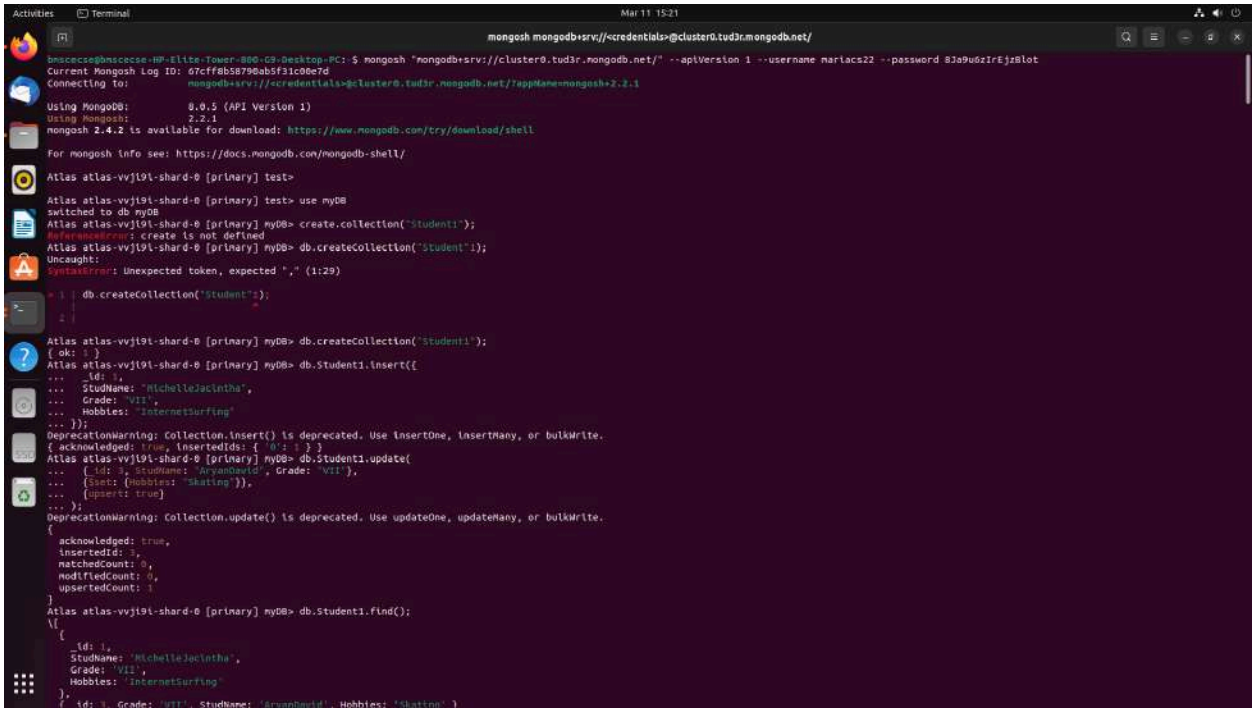

Activities
Terminal
Mar 11 14:53
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~
dName,Grade,Hobbies"
Uncaught:
SyntaxError: Missing semicolon. (1:14)
> 1 | mongoexport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudName,Grade,Hobbies"
| ^
2 |
Atlas atlas-vvji9i-shard-0 [primary] myDB>
(To exit, press Ctrl+C again or Ctrl+D or type .exit)
Atlas atlas-vvji9i-shard-0 [primary] myDB> mongoexport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudName,Grade,Hobbies"
Uncaught:
SyntaxError: Missing semicolon. (1:14)
> 1 | mongoexport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudName,Grade,Hobbies"
| ^
2 |
Atlas atlas-vvji9i-shard-0 [primary] myDB> exit
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ mongoexport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudName,Grade,Hobbies"
2025-03-11T14:42:20.815+0530 connected to: mongodb+srv://[**REDACTED**]@cluster0.tud3r.mongodb.net/myDB
2025-03-11T14:42:21.068+0530 exported 2 records
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ mongoexport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/out.txt" --fields "StudName,Grade,Hobbies"
2025-03-11T14:43:35.174+0530 connected to: mongodb+srv://[**REDACTED**]@cluster0.tud3r.mongodb.net/myDB
2025-03-11T14:43:35.437+0530 exported 2 records
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ mongoexport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/qwer.txt" --fields "StudName,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
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ whoami
bmscecse
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ ^C
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ mongoexport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "/home/bmscecse/Desktop/qwer.csv" --fields "StudName,Grade,Hobbies"
2025-03-11T14:47:52.993+0530 connected to: mongodb+srv://[**REDACTED**]@cluster0.tud3r.mongodb.net/myDB
2025-03-11T14:47:53.277+0530 exported 2 records
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ mongoimport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student3 --type csv --file "/home/bmscecse/Desktop/qwer.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.
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $


Activities
soffice
Mar 11 14:53
LibreOffice 7.5
LibreOffice
Community
Text Import - [qwer.csv]
Import
Character set:
Unicode (UTF-8)
Language:
Default - English (India)
From row:
Separator Options
Fixed width
Separated by
Tab
Comma
Semicolon
Space
Other
Merge delimiters
Trim spaces
String delimiter:
Other Options
Format quoted field as text
Detect special numbers
Evaluate formulas
Fields
Column type:
Standard
Standard
Standard
1 StudName Grade Hobbies
2 MichelleJacintha VII InternetSurfing
3 AryanDavid VII Chess
Help
Cancel
OK
Sgpa.class
Sgpa.java
qwer.csv
Student.java
Student.class
main.java
Student1.java
calsgpa.java
Quad.class
cals\gpa.java
QuadraticEquationSolver.class
CalSgpa.java
Quad.java
calsgpa.class
Home

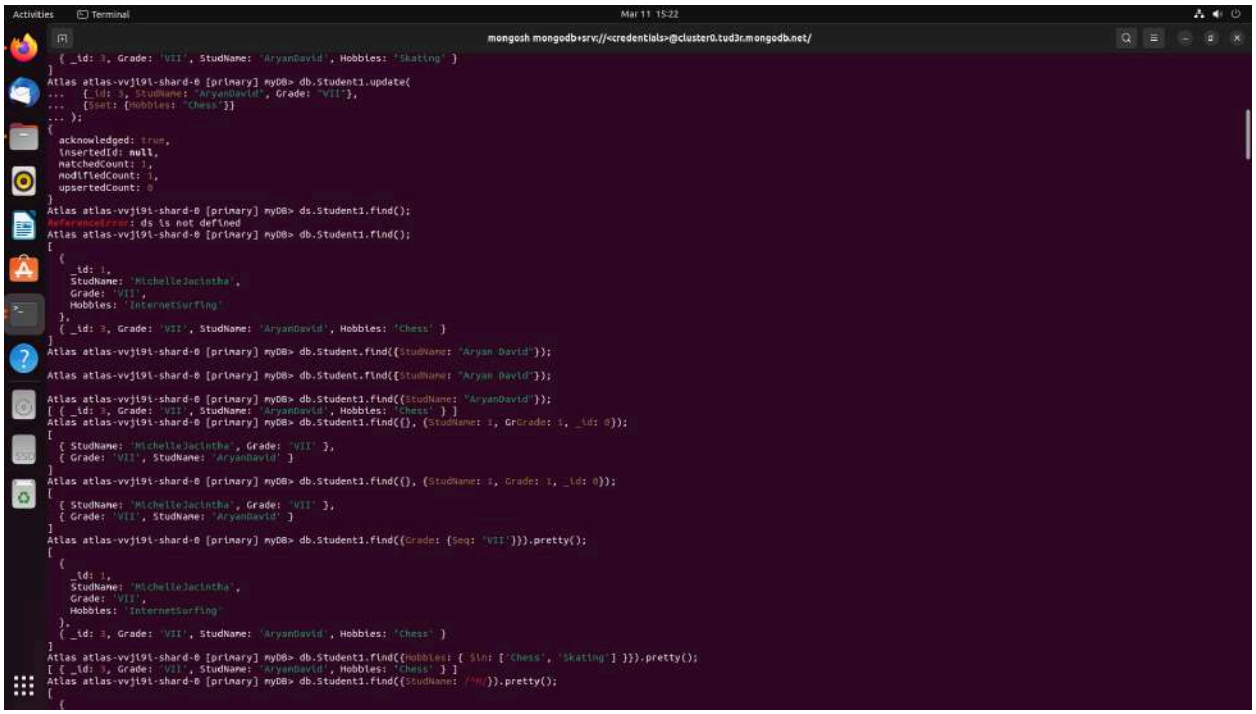

Activities
Terminal
Mar 11 15:21
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
bnscecse@bnscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ mongosh "mongodb+srv://cluster0.tud3r.mongodb.net/" --apiVersion 1 --username marlacs22 --password 8Ja9u6zIrEjzBlot
Current Mongosh Log ID: 67cff8b58790ab5f31c00e7d
Connecting to: mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/?appName=mongosh+2.2.1
Using MongoDB: 8.0.5 (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-vvji9i-shard-0 [primary] test>
Atlas atlas-vvji9i-shard-0 [primary] test> use myDB
switched to db myDB
Atlas atlas-vvji9i-shard-0 [primary] myDB> create.collection("Student1");
ReferenceError: create is not defined
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.createCollection("Student"1);
Uncaught:
SyntaxError: Unexpected token, expected "," (1:29)
1 | db.createCollection("Student"1);
2 |
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.createCollection('Student1');
{ ok: 1 }
Atlas atlas-vvji9i-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, insertedIds: { '0': 1 } }
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.update(
... {_id: 3, StudName: 'AryanDavid', Grade: 'VII'},
... {$set: {Hobbies: 'Skating'}},
... {upsert: true}
...);
DeprecationWarning: Collection.update() is deprecated. Use updateOne, updateMany, or bulkWrite.
{
acknowledged: true,
insertedId: 3,
matchedCount: 0,
modifiedCount: 0,
upsertedCount: 1
}
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find();
[
{
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
},


Activities
Terminal
Mar 11 15:22
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
{ _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Skating' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.update(
... {_id: 3, StudName: "AryanDavid", Grade: "VII"},
... {$set: {Hobbies: "Chess"}}
...);
{
acknowledged: true,
insertedId: null,
matchedCount: 1,
modifiedCount: 1,
upsertedCount: 0
}
Atlas atlas-vvji9i-shard-0 [primary] myDB> ds.Student1.find();
ReferenceError: ds is not defined
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find();
[
{
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
},
{ _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student.find({StudName: "Aryan David"});
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student.find({StudName: "Aryan David"});
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({StudName: "AryanDavid"});
[{ _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' }]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({}, {StudName: 1, Grade: 1, _id: 0});
[
{ StudName: 'MichelleJacintha', Grade: 'VII' },
{ Grade: 'VII', StudName: 'AryanDavid' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({}, {StudName: 1, Grade: 1, _id: 0});
[
{ StudName: 'MichelleJacintha', Grade: 'VII' },
{ Grade: 'VII', StudName: 'AryanDavid' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({Grade: {$eq: 'VII'}}).pretty();
[
{
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
},
{ _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({Hobbies: { $in: ['Chess', 'Skating'] }}).pretty();
[{ _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' }]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({StudName: /^M/}).pretty();
[

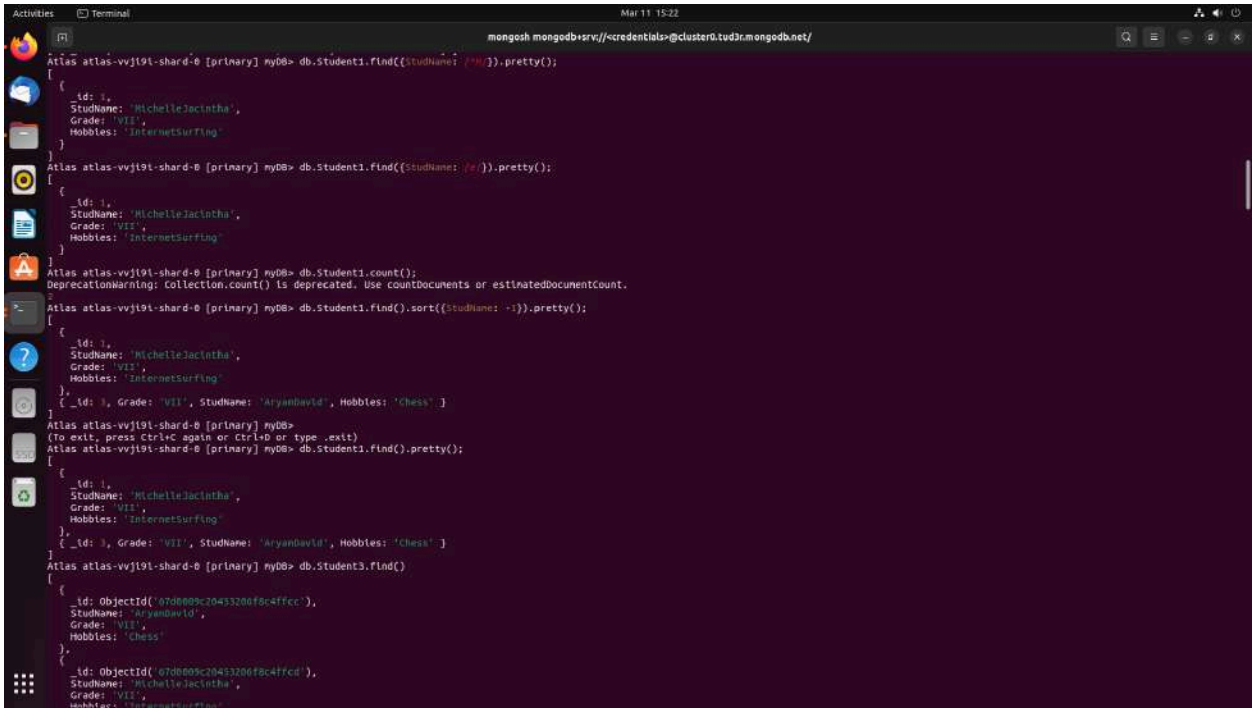

Activities
Terminal
Mar 11 15:22
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({StudName: /^M/}).pretty();
[
{
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
}
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({StudName: /e/}).pretty();
[
{
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
}
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.count();
DeprecationWarning: Collection.count() is deprecated. Use countDocuments or estimatedDocumentCount.
3
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find().sort({StudName: -1}).pretty();
[
{
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
},
{ _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB>
(To exit, press Ctrl+C again or Ctrl+D or type .exit)
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find().pretty();
[
{
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
},
{ _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student3.find()
[
{
_id: ObjectId('67d0809c20433206f8c4ffcc'),
StudName: 'AryanDavid',
Grade: 'VII',
Hobbies: 'Chess'
},
{
_id: ObjectId('67d0809c20433206f8c4ffcd'),
StudName: 'MichelleJacintha',
Grade: 'VII',

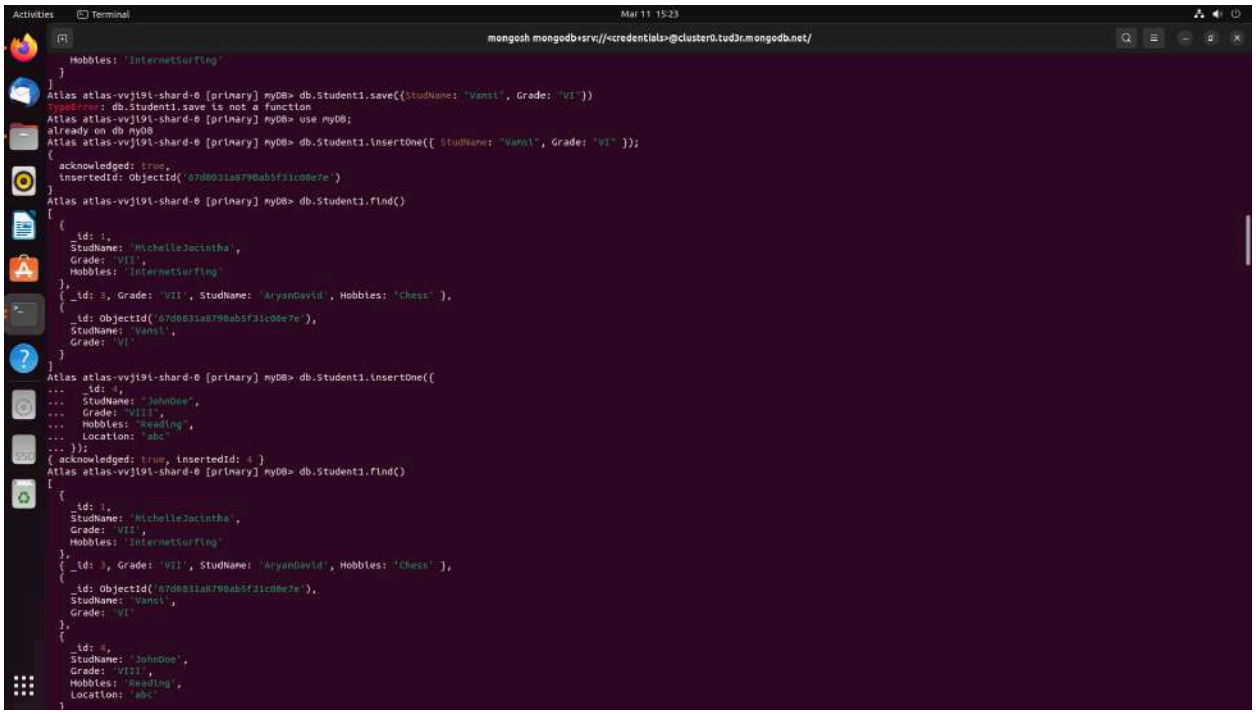


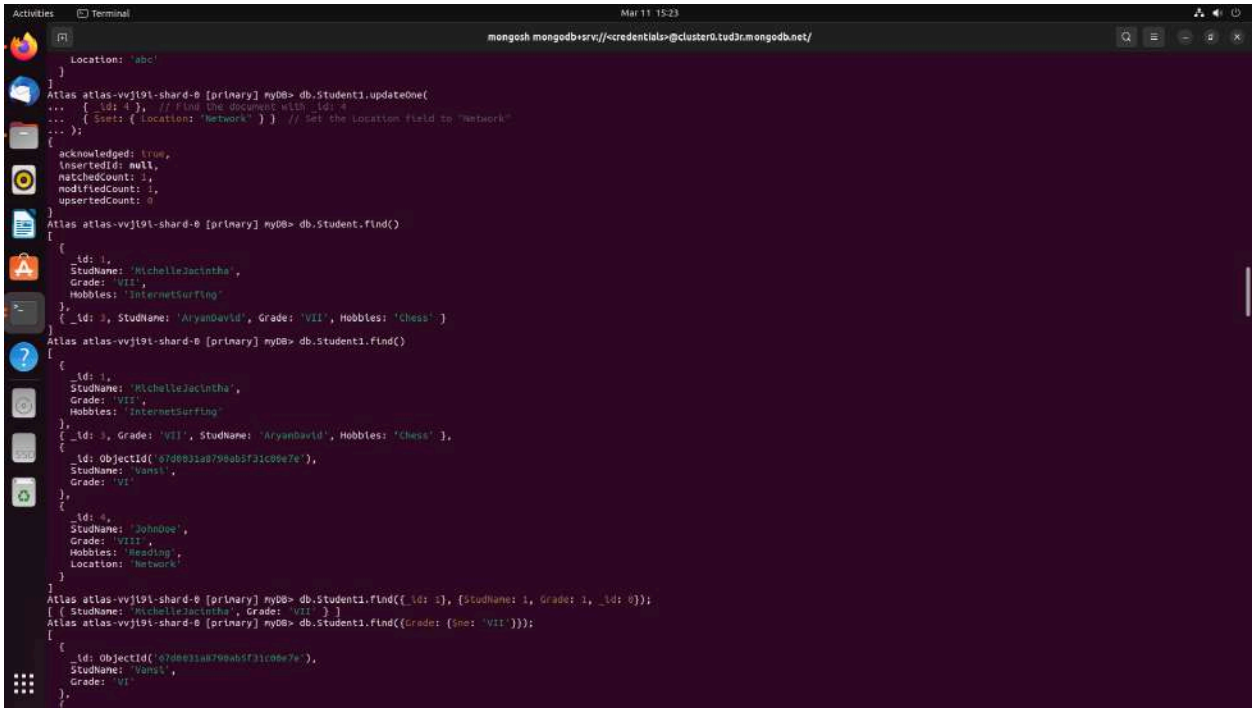

```
bnsccsce@bnsccsce-HP-Elite-Tower-800-G9-Desktop-PC: $ mongosh "mongodb+srv://cluster0.tud3r.mongodb.net/" --apiVersion 1 --username marlacs22 --password 8Ja9u6zIrEjz8lot
Current Mongosh Log ID: 67cff8b58790ab5f31c00e7d
Connecting to:          mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/?appName=mongosh+2.2.1

Using MongoDB:          8.0.5 (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-vvji9i-shard-0 [primary] test>

Atlas atlas-vvji9i-shard-0 [primary] test> use myDB
switched to db myDB
Atlas atlas-vvji9i-shard-0 [primary] myDB> create.collection("Student1");
ReferenceError: create is not defined
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.createCollection("Student1");
Uncaught:
SyntaxError: Unexpected token, expected "," (1:29)

> 1 | db.createCollection("Student1");
    |                             ^
  2 |

Atlas atlas-vvji9i-shard-0 [primary] myDB> db.createCollection("Student1");
{ ok: 1 }
Atlas atlas-vvji9i-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, insertedIds: { '0': 1 } }
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.update(
...   {_id: 3, StudName: "AryanDavid", Grade: "VII"},
...   {$set: {Hobbies: "Skating"}},
...   {upsert: true}
... );
DeprecationWarning: Collection.update() is deprecated. Use updateOne, updateMany, or bulkWrite.
{
  acknowledged: true,
  insertedId: 3,
  matchedCount: 0,
  modifiedCount: 0,
  upsertedCount: 1
}
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find();
[
  {
    _id: 1,
    StudName: 'MichelleJacintha',
    Grade: 'VII',
    Hobbies: 'InternetSurfing'
  },
  { _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Skating' }
]
```

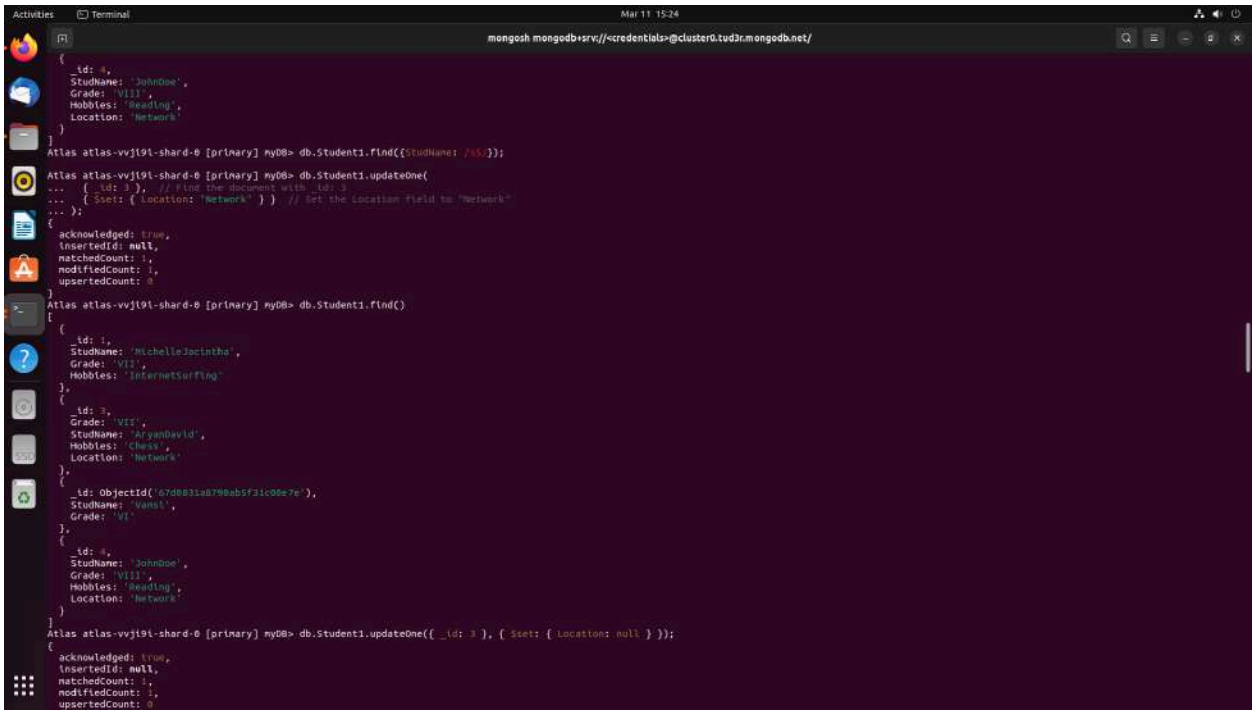


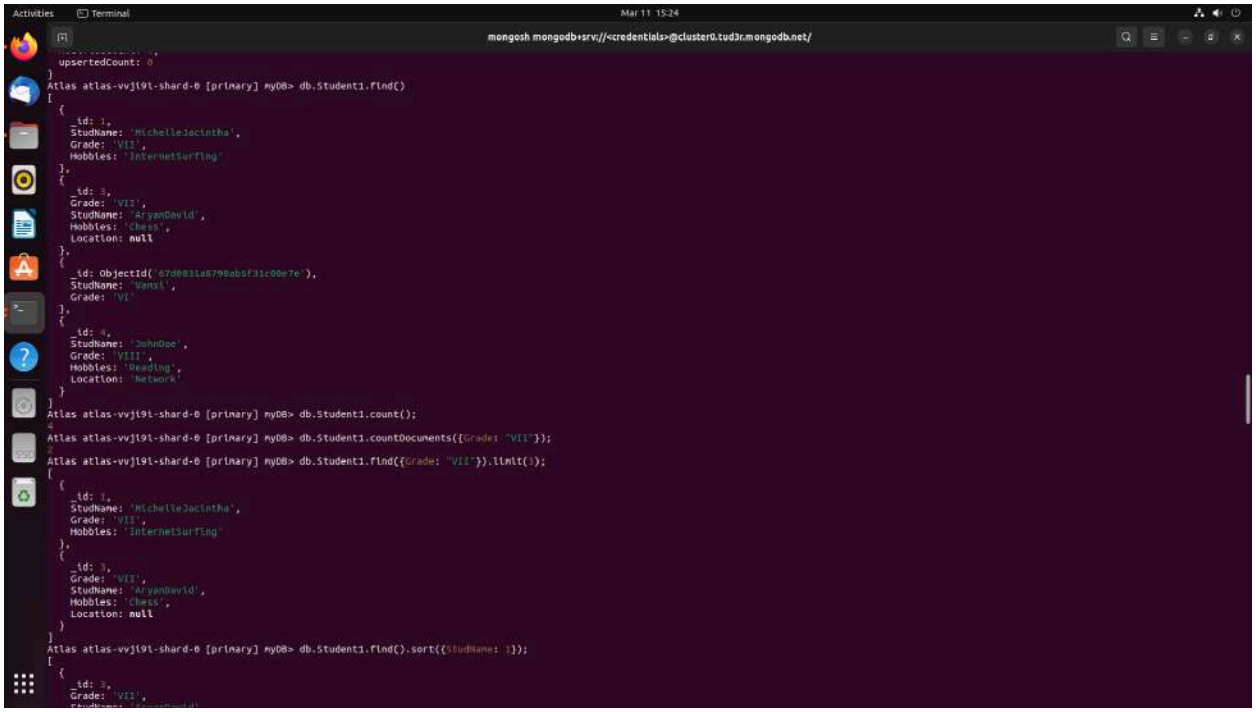

```
  { _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Skating' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.update(
...   {_id: 3, StudName: "AryanDavid", Grade: "VII"},
...   {$set: {Hobbies: "Chess"}}
... );
{
  acknowledged: true,
  insertedId: null,
  matchedCount: 1,
  modifiedCount: 1,
  upsertedCount: 0
}
Atlas atlas-vvji9i-shard-0 [primary] myDB> ds.Student1.find();
ReferenceError: ds is not defined
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find();
[
  {
    _id: 1,
    StudName: 'MichelleJacintha',
    Grade: 'VII',
    Hobbies: 'InternetSurfing'
  },
  { _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student.find({StudName: "Aryan David"});

Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student.find({StudName: "Aryan David"});

Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({StudName: "AryanDavid"});
[ { _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' } ]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({}, {StudName: 1, GrGrade: 1, _id: 0});
[
  { StudName: 'MichelleJacintha', Grade: 'VII' },
  { Grade: 'VII', StudName: 'AryanDavid' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({}, {StudName: 1, Grade: 1, _id: 0});
[
  { StudName: 'MichelleJacintha', Grade: 'VII' },
  { Grade: 'VII', StudName: 'AryanDavid' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({Grade: {Seq: 'VII'}}).pretty();
[
  {
    _id: 1,
    StudName: 'MichelleJacintha',
    Grade: 'VII',
    Hobbies: 'InternetSurfing'
  },
  { _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({Hobbies: { $in: ['Chess', 'Skating'] }}).pretty();
[ { _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' } ]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({StudName: /^M/}).pretty();
[
  {
```

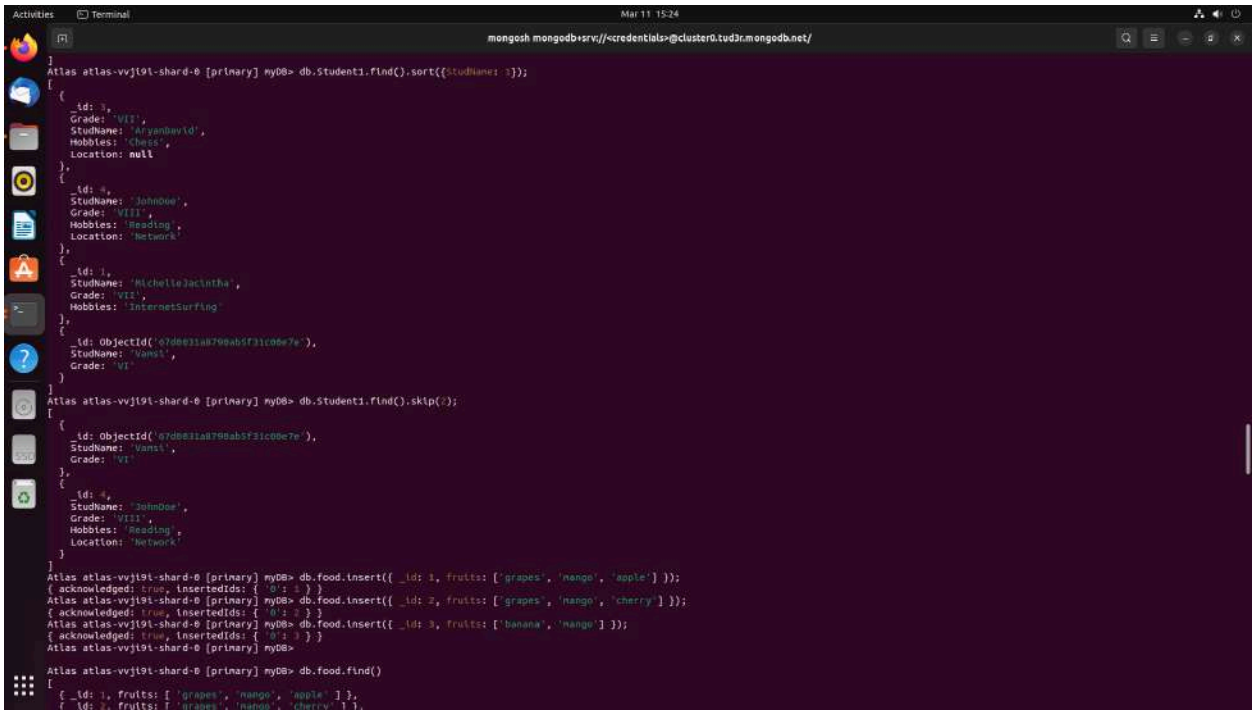

Activities
Terminal
Mar 11 15:22
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({StudName: /^M/}).pretty();
[
{
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
}
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({StudName: /e/}).pretty();
[
{
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
}
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.count();
DeprecationWarning: Collection.count() is deprecated. Use countDocuments or estimatedDocumentCount.
2
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find().sort({StudName: -1}).pretty();
[
{
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
},
{ _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB>
(To exit, press Ctrl+C again or Ctrl+D or type .exit)
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find().pretty();
[
{
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
},
{ _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student3.find()
[
{
_id: ObjectId('67d0809c20453206f8c4ffcc'),
StudName: 'AryanDavid',
Grade: 'VII',
Hobbies: 'Chess'
},
{
_id: ObjectId('67d0809c20453206f8c4ffcd'),
StudName: 'MichelleJacintha',
Grade: 'VII',

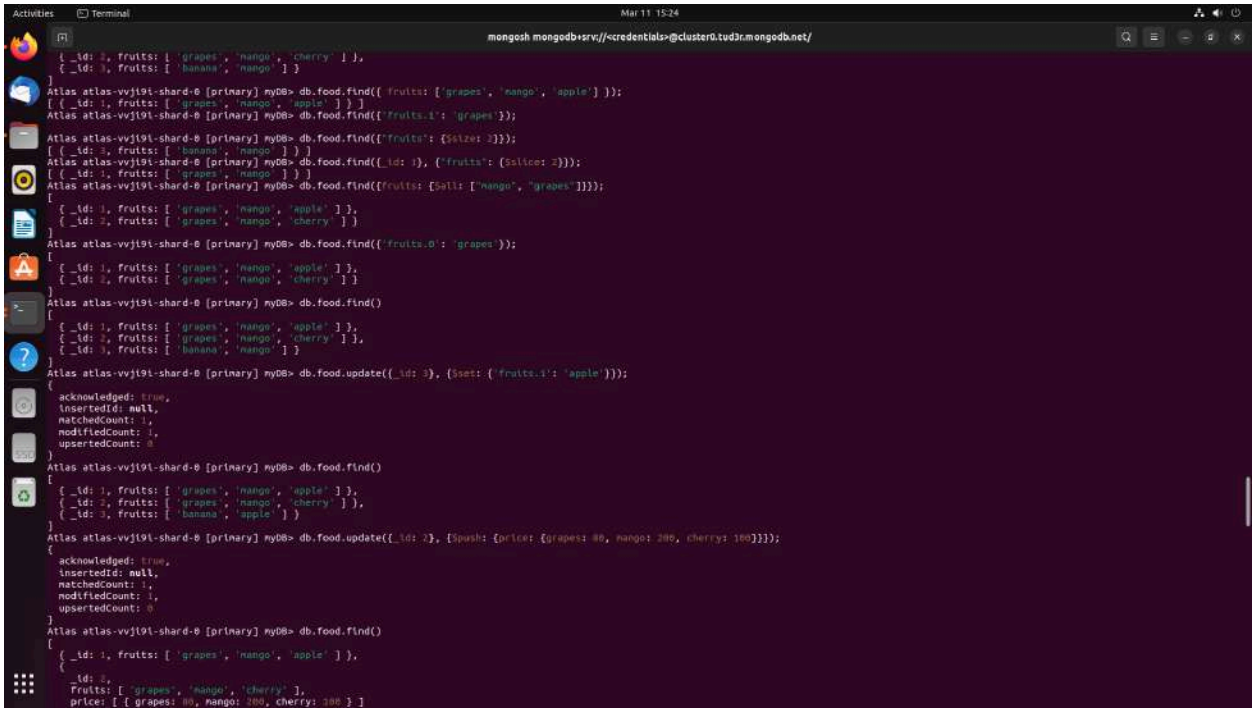

Activities
Terminal
Mar 11 15:23
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
Hobbies: 'InternetSurfing'
}
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.save({StudName: "Vansi", Grade: "VI"})
TypeError: db.Student1.save is not a function
Atlas atlas-vvji9i-shard-0 [primary] myDB> use myDB;
already on db myDB
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.insertOne({ StudName: "Vansi", Grade: "VI" });
{
acknowledged: true,
insertedId: ObjectId('67d0831a8790ab5f31c00e7e')
}
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find()
[
{
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
},
{ _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' },
{
_id: ObjectId('67d0831a8790ab5f31c00e7e'),
StudName: 'Vansi',
Grade: 'VI'
}
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.insertOne({
... _id: 4,
... StudName: "JohnDoe",
... Grade: "VIII",
... Hobbies: "Reading",
... Location: "abc"
... });
{ acknowledged: true, insertedId: 4 }
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find()
[
{
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
},
{ _id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess' },
{
_id: ObjectId('67d0831a8790ab5f31c00e7e'),
StudName: 'Vansi',
Grade: 'VI'
},
{
_id: 4,
StudName: 'JohnDoe',
Grade: 'VIII',
Hobbies: 'Reading',
Location: 'abc'
}

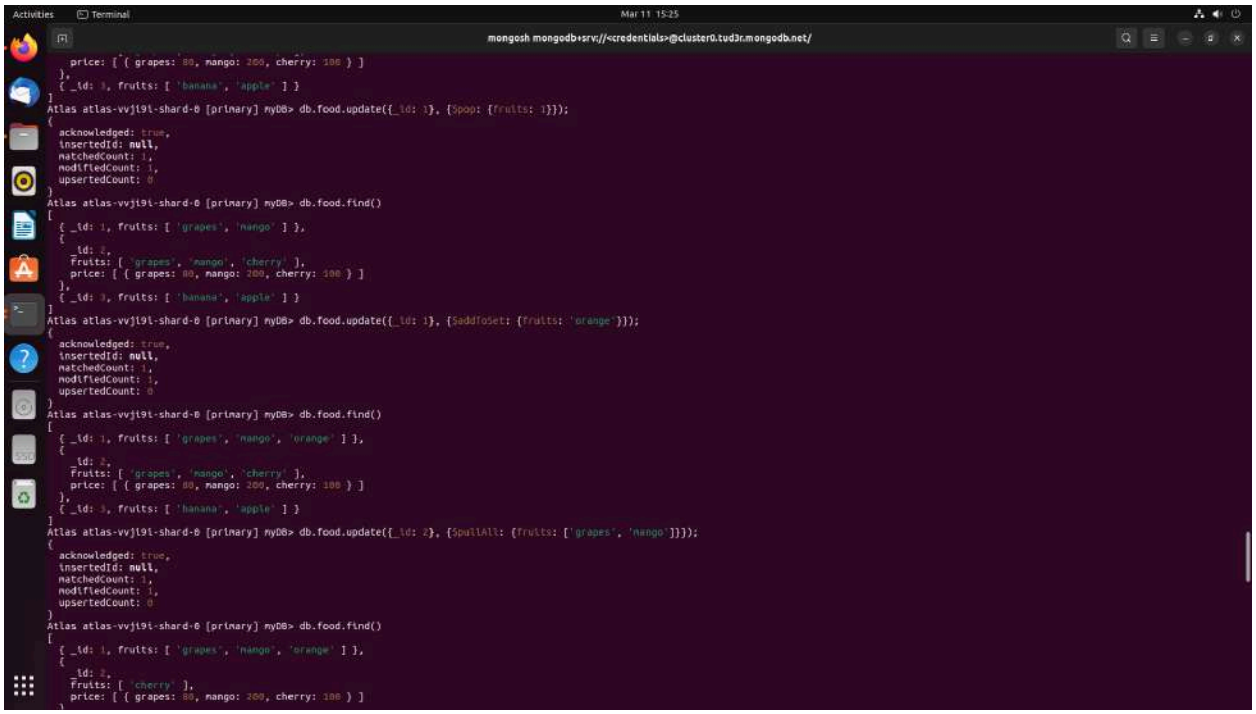

Activities
Terminal
Mar 11 15:23
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
Location: 'abc'
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.updateOne(
... { _id: 4 }, // Find the document with _id: 4
... { $set: { Location: 'Network' } } // Set the Location field to "Network"
...);
acknowledged: true,
insertedId: null,
matchedCount: 1,
modifiedCount: 1,
upsertedCount: 0
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student.find()
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
_id: 3, StudName: 'AryanDavid', Grade: 'VII', Hobbies: 'Chess'
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find()
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
_id: 3, Grade: 'VII', StudName: 'AryanDavid', Hobbies: 'Chess'
_id: ObjectId('67d0831a8790ab5f31c08e7e'),
StudName: 'Vansi',
Grade: 'VI'
_id: 4,
StudName: 'JohnDoe',
Grade: 'VIII',
Hobbies: 'Reading',
Location: 'Network'
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({ _id: 1}, {StudName: 1, Grade: 1, _id: 0});
[{ StudName: 'MichelleJacintha', Grade: 'VII' }]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({Grade: {$ne: 'VII'}});
_id: ObjectId('67d0831a8790ab5f31c08e7e'),
StudName: 'Vansi',
Grade: 'VI'

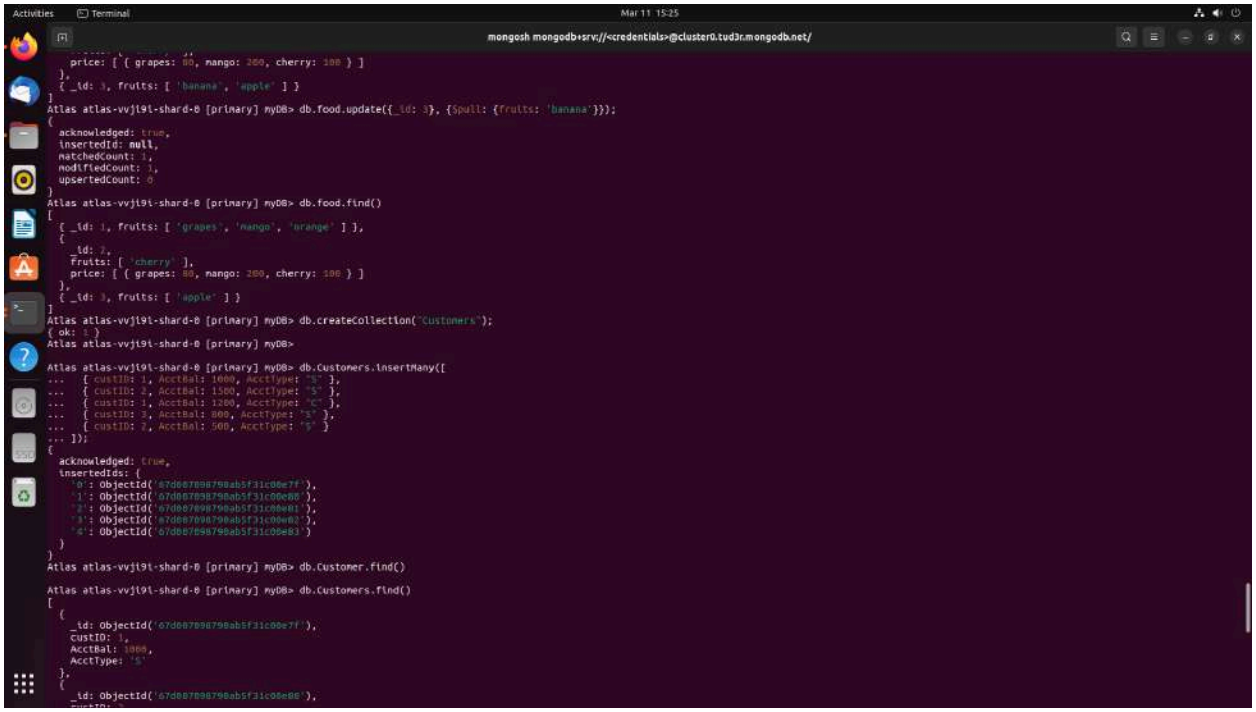

Activities
Terminal
Mar 11 15:24
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
_id: 4,
StudName: 'JohnDoe',
Grade: 'VIII',
Hobbies: 'Reading',
Location: 'Network'
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({StudName: /^S/});
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.updateOne(
... { _id: 3 }, // Find the document with _id: 3
... { $set: { Location: 'Network' } } // Set the Location field to "Network"
...);
acknowledged: true,
insertedId: null,
matchedCount: 1,
modifiedCount: 1,
upsertedCount: 0
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find()
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
_id: 3,
Grade: 'VII',
StudName: 'AryanDavid',
Hobbies: 'Chess',
Location: 'Network'
_id: ObjectId('67d0831a8790ab5f31c08e7e'),
StudName: 'Vansi',
Grade: 'VI'
_id: 4,
StudName: 'JohnDoe',
Grade: 'VIII',
Hobbies: 'Reading',
Location: 'Network'
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.updateOne({ _id: 3 }, { $set: { Location: null } });
acknowledged: true,
insertedId: null,
matchedCount: 1,
modifiedCount: 1,
upsertedCount: 0

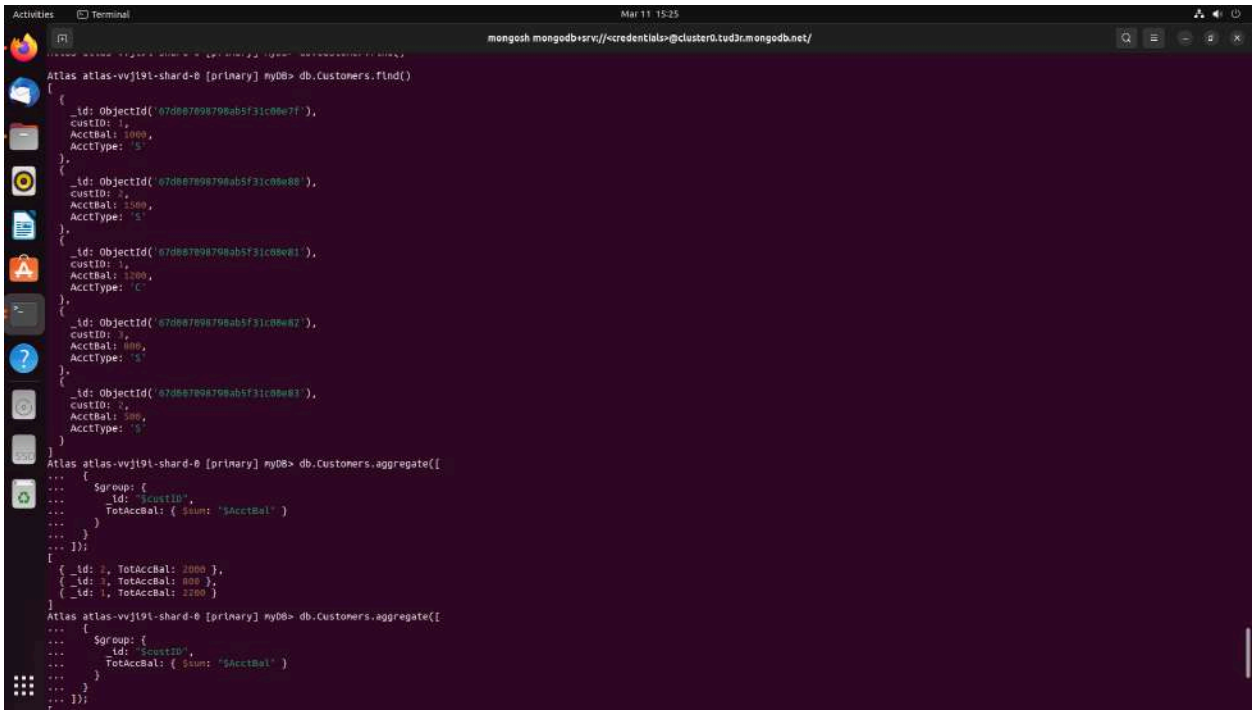

Activities
Terminal
Mar 11 15:24
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
upsertedCount: 0
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find()
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
_id: 3,
Grade: 'VII',
StudName: 'AryanDavid',
Hobbies: 'Chess',
Location: null
_id: ObjectId('67d0831a8790ab5f31c00e7e'),
StudName: 'Vansi',
Grade: 'VI'
_id: 4,
StudName: 'JohnDoe',
Grade: 'VIII',
Hobbies: 'Reading',
Location: 'Network'
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.count();
4
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.countDocuments({Grade: "VII"});
2
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find({Grade: "VII"}).limit(1);
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
_id: 3,
Grade: 'VII',
StudName: 'AryanDavid',
Hobbies: 'Chess',
Location: null
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find().sort({StudName: 1});
_id: 3,
Grade: 'VII',

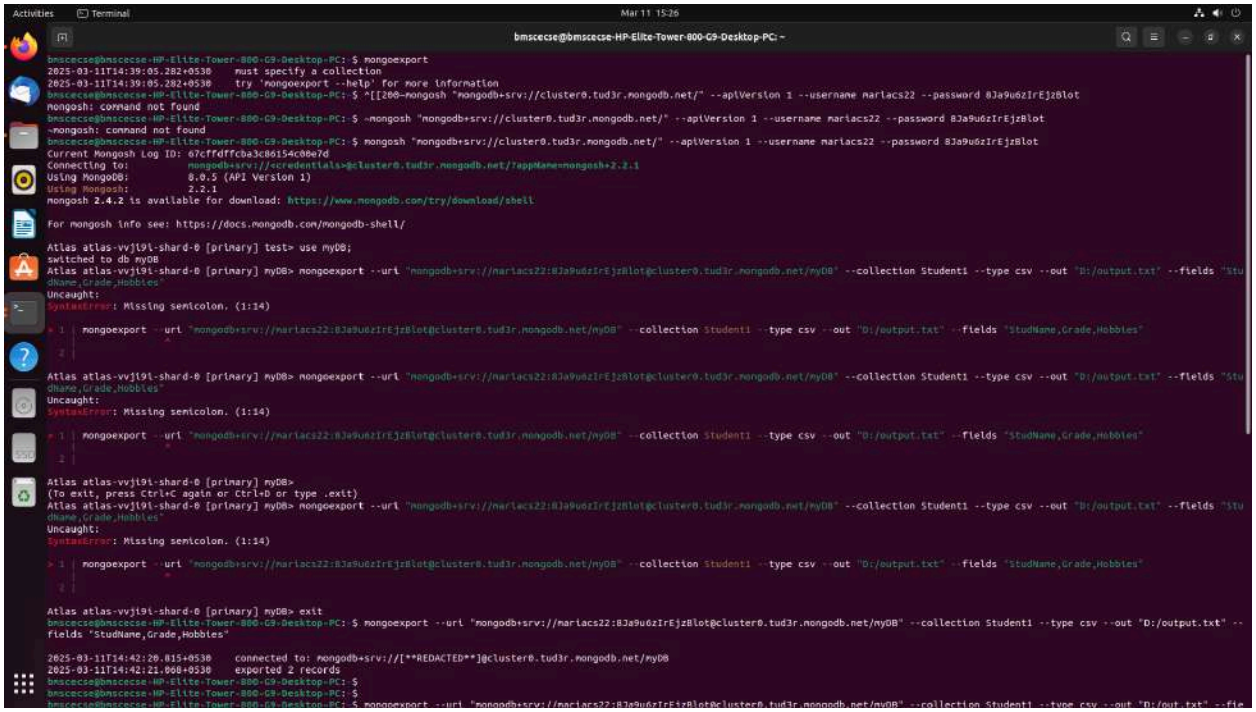

Activities
Terminal
Mar 11 15:24
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find().sort({StudName: 1});
_id: 3,
Grade: 'VII',
StudName: 'AryanDavid',
Hobbies: 'Chess',
Location: null
_id: 4,
StudName: 'JohnDoe',
Grade: 'VIII',
Hobbies: 'Reading',
Location: 'Network'
_id: 1,
StudName: 'MichelleJacintha',
Grade: 'VII',
Hobbies: 'InternetSurfing'
_id: ObjectId('67d0831a8790ab5f31c00e7e'),
StudName: 'Vansi',
Grade: 'VI'
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Student1.find().skip(2);
_id: ObjectId('67d0831a8790ab5f31c00e7e'),
StudName: 'Vansi',
Grade: 'VI'
_id: 4,
StudName: 'JohnDoe',
Grade: 'VIII',
Hobbies: 'Reading',
Location: 'Network'
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.food.insert({ _id: 1, fruits: ['grapes', 'mango', 'apple'] });
{ acknowledged: true, insertedIds: { '0': 1 } }
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.food.insert({ _id: 2, fruits: ['grapes', 'mango', 'cherry'] });
{ acknowledged: true, insertedIds: { '0': 2 } }
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.food.insert({ _id: 3, fruits: ['banana', 'mango'] });
{ acknowledged: true, insertedIds: { '0': 3 } }
Atlas atlas-vvji9i-shard-0 [primary] myDB>
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.food.find()
{ _id: 1, fruits: ['grapes', 'mango', 'apple'] },

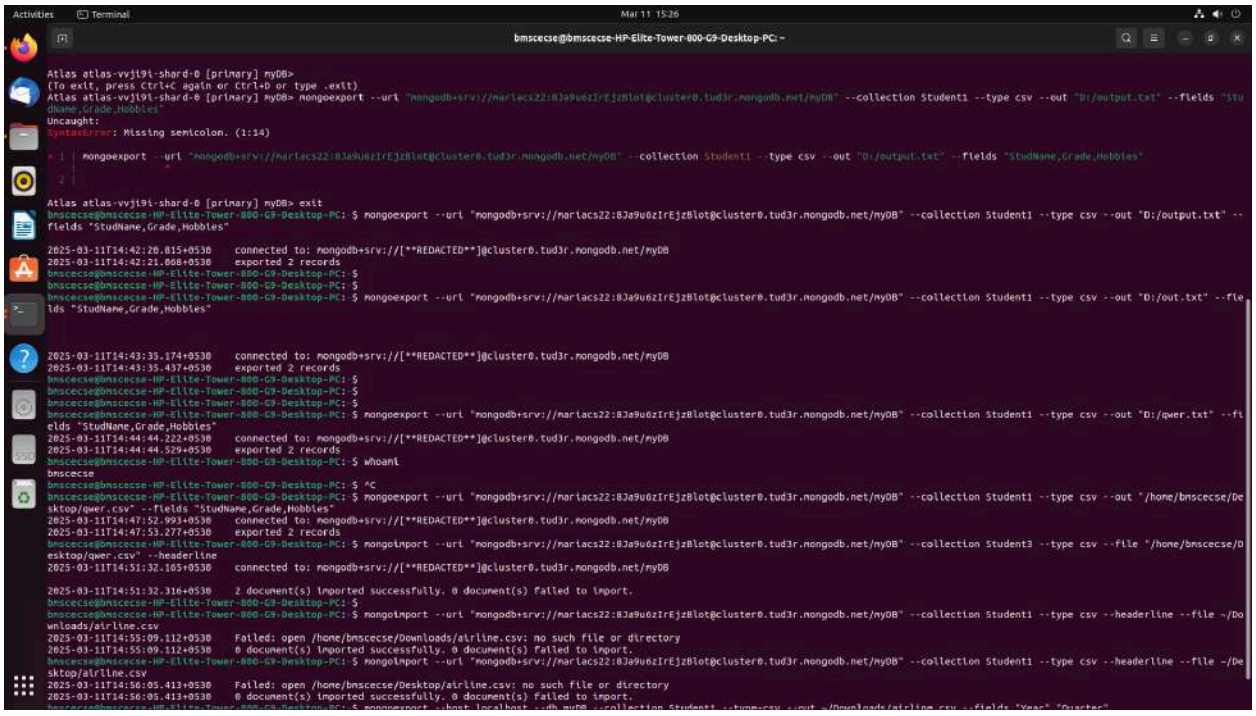

Activities
Terminal
Mar 11 15:24
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
{ _id: 2, fruits: ['grapes', 'mango', 'cherry'] },
{ _id: 3, fruits: ['banana', 'mango'] }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.food.find({ fruits: ['grapes', 'mango', 'apple'] });
[{ _id: 1, fruits: ['grapes', 'mango', 'apple'] }]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.food.find({'fruits.1': 'grapes'});
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.food.find({"fruits": {$size: 2}});
[{ _id: 3, fruits: ['banana', 'mango'] }]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.food.find({_id: 1}, {"fruits": {$slice: 2}});
[{ _id: 1, fruits: ['grapes', 'mango'] }]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.food.find({fruits: {$all: ["mango", "grapes"]}});
[
{ _id: 1, fruits: ['grapes', 'mango', 'apple'] },
{ _id: 2, fruits: ['grapes', 'mango', 'cherry'] }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.food.find({'fruits.0': 'grapes'});
[
{ _id: 1, fruits: ['grapes', 'mango', 'apple'] },
{ _id: 2, fruits: ['grapes', 'mango', 'cherry'] }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.food.find()
[
{ _id: 1, fruits: ['grapes', 'mango', 'apple'] },
{ _id: 2, fruits: ['grapes', 'mango', 'cherry'] },
{ _id: 3, fruits: ['banana', 'mango'] }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.food.update({_id: 3}, {$set: {'fruits.1': 'apple'}});
{
acknowledged: true,
insertedId: null,
matchedCount: 1,
modifiedCount: 1,
upsertedCount: 0
}
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.food.find()
[
{ _id: 1, fruits: ['grapes', 'mango', 'apple'] },
{ _id: 2, fruits: ['grapes', 'mango', 'cherry'] },
{ _id: 3, fruits: ['banana', 'apple'] }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.food.update({_id: 2}, {$push: {price: {grapes: 80, mango: 200, cherry: 100}}});
{
acknowledged: true,
insertedId: null,
matchedCount: 1,
modifiedCount: 1,
upsertedCount: 0
}
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.food.find()
[
{ _id: 1, fruits: ['grapes', 'mango', 'apple'] },
{
_id: 2,
fruits: ['grapes', 'mango', 'cherry'],
price: [{ grapes: 80, mango: 200, cherry: 100 }]

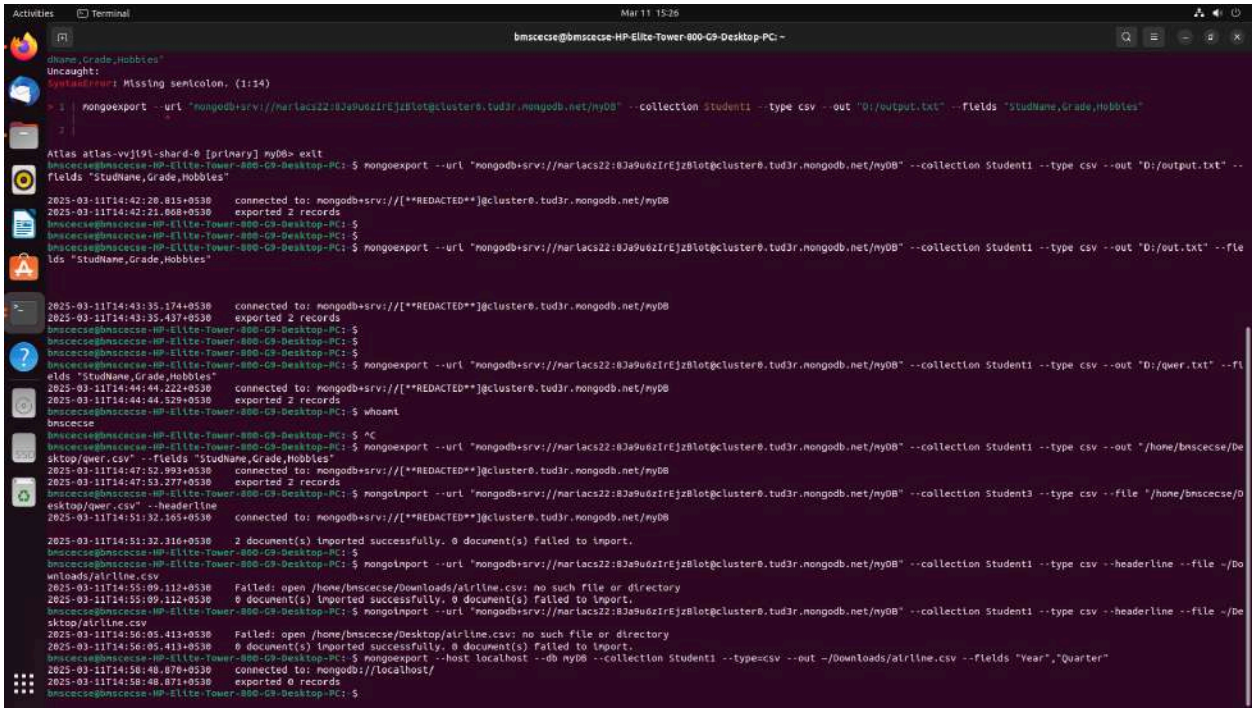

Activities
Terminal
Mar 11 15:25
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
price: [{ grapes: 80, mango: 200, cherry: 100 }]
},
{ _id: 3, fruits: ['banana', 'apple'] }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.food.update({_id: 1}, {$pop: {fruits: 1}});
{
acknowledged: true,
insertedId: null,
matchedCount: 1,
modifiedCount: 1,
upsertedCount: 0
}
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.food.find()
[
{ _id: 1, fruits: ['grapes', 'mango'] },
{
_id: 2,
fruits: ['grapes', 'mango', 'cherry'],
price: [{ grapes: 80, mango: 200, cherry: 100 }]
},
{ _id: 3, fruits: ['banana', 'apple'] }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.food.update({_id: 1}, {$addToSet: {fruits: 'orange'}});
{
acknowledged: true,
insertedId: null,
matchedCount: 1,
modifiedCount: 1,
upsertedCount: 0
}
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.food.find()
[
{ _id: 1, fruits: ['grapes', 'mango', 'orange'] },
{
_id: 2,
fruits: ['grapes', 'mango', 'cherry'],
price: [{ grapes: 80, mango: 200, cherry: 100 }]
},
{ _id: 3, fruits: ['banana', 'apple'] }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.food.update({_id: 2}, {$pullAll: {fruits: ['grapes', 'mango']}});
{
acknowledged: true,
insertedId: null,
matchedCount: 1,
modifiedCount: 1,
upsertedCount: 0
}
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.food.find()
[
{ _id: 1, fruits: ['grapes', 'mango', 'orange'] },
{
_id: 2,
fruits: ['cherry'],
price: [{ grapes: 80, mango: 200, cherry: 100 }]

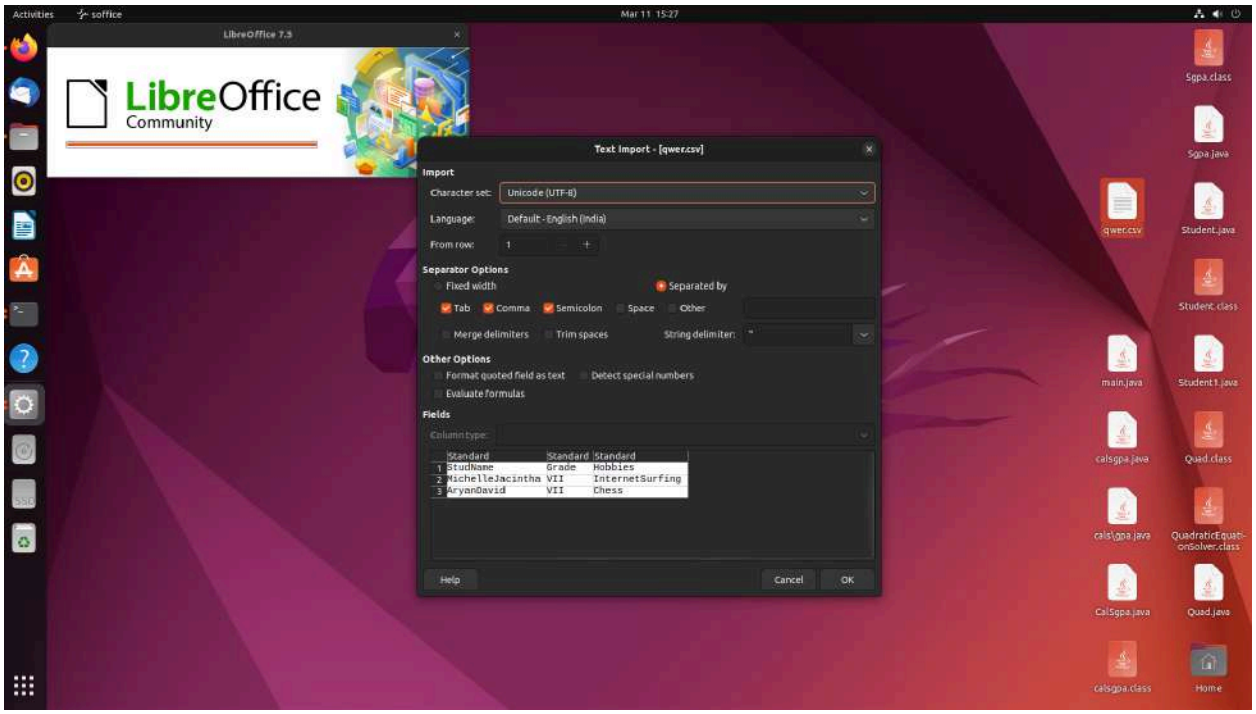

Activities
Terminal
Mar 11 15:25
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
price: [{ grapes: 80, mango: 200, cherry: 100 }]
},
{ _id: 3, fruits: ['banana', 'apple'] }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.food.update({_id: 3}, {$pull: {fruits: 'banana'}});
{
acknowledged: true,
insertedId: null,
matchedCount: 1,
modifiedCount: 1,
upsertedCount: 0
}
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.food.find()
[
{ _id: 1, fruits: ['grapes', 'mango', 'orange'] },
{
_id: 2,
fruits: ['cherry'],
price: [{ grapes: 80, mango: 200, cherry: 100 }]
},
{ _id: 3, fruits: ['apple'] }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.createCollection("Customers");
{ ok: 1 }
Atlas atlas-vvji9i-shard-0 [primary] myDB>
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Customers.insertMany([
... { custID: 1, AcctBal: 1000, AcctType: "S" },
... { custID: 2, AcctBal: 1500, AcctType: "S" },
... { custID: 1, AcctBal: 1200, AcctType: "C" },
... { custID: 3, AcctBal: 800, AcctType: "S" },
... { custID: 2, AcctBal: 500, AcctType: "S" }
...]);
{
acknowledged: true,
insertedIds: {
'0': ObjectId('67d087098798ab5f31c06e7f'),
'1': ObjectId('67d087098798ab5f31c06e80'),
'2': ObjectId('67d087098798ab5f31c06e81'),
'3': ObjectId('67d087098798ab5f31c06e82'),
'4': ObjectId('67d087098798ab5f31c06e83')
}
}
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Customer.find()
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Customers.find()
[
{
_id: ObjectId('67d087098798ab5f31c06e7f'),
custID: 1,
AcctBal: 1000,
AcctType: 'S'
},
{
_id: ObjectId('67d087098798ab5f31c06e80'),

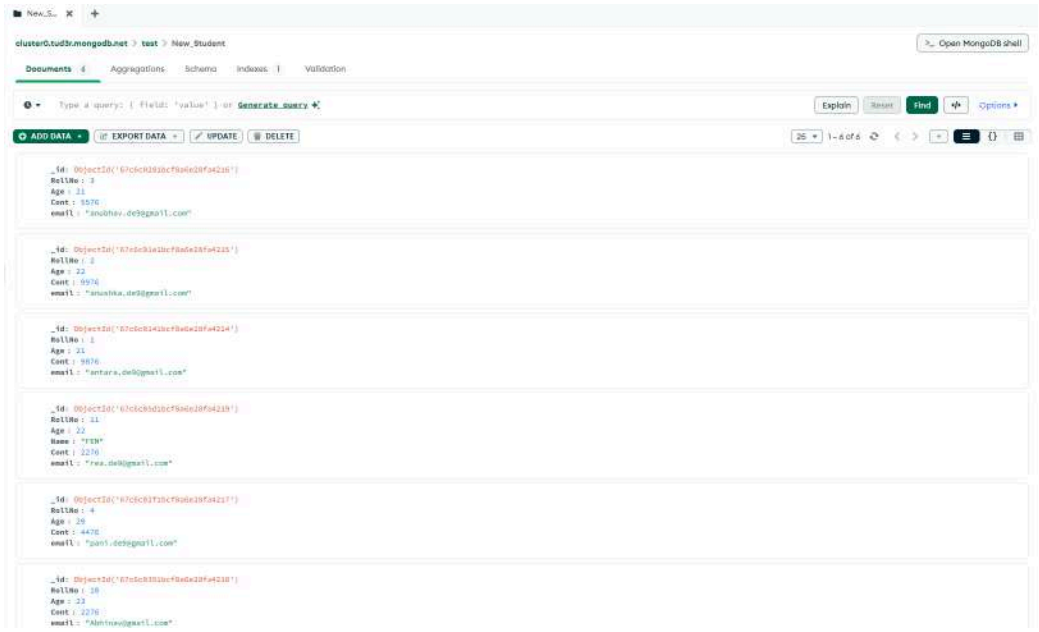

Activities
Terminal
Mar 11 15:25
mongosh mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Customers.find()
[
{
_id: ObjectId('67d087098798ab5f31c06e7f'),
custID: 1,
AcctBal: 1000,
AcctType: 'S'
},
{
_id: ObjectId('67d087098798ab5f31c06e80'),
custID: 2,
AcctBal: 1500,
AcctType: 'S'
},
{
_id: ObjectId('67d087098798ab5f31c06e81'),
custID: 1,
AcctBal: 1200,
AcctType: 'C'
},
{
_id: ObjectId('67d087098798ab5f31c06e82'),
custID: 3,
AcctBal: 800,
AcctType: 'S'
},
{
_id: ObjectId('67d087098798ab5f31c06e83'),
custID: 2,
AcctBal: 500,
AcctType: 'S'
}
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Customers.aggregate([
... {
... $group: {
... _id: "$custID",
... TotAccBal: { $sum: '$AcctBal' }
... }
... }
...]);
[
{ _id: 2, TotAccBal: 2000 },
{ _id: 3, TotAccBal: 800 },
{ _id: 1, TotAccBal: 2200 }
]
Atlas atlas-vvji9i-shard-0 [primary] myDB> db.Customers.aggregate([
... {
... $group: {
... _id: "$custID",
... TotAccBal: { $sum: "$AcctBal" }
... }
... }
...]);


Activities
Terminal
Mar 11 15:26
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~
bmscecse@bmscecse-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
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ ^[[200~mongosh "mongodb+srv://cluster0.tud3r.mongodb.net/" --apiVersion 1 --username mariacs22 --password 8Ja9u6zIrEjzBlot
mongosh: command not found
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ ~mongosh "mongodb+srv://cluster0.tud3r.mongodb.net/" --apiVersion 1 --username mariacs22 --password 8Ja9u6zIrEjzBlot
~mongosh: command not found
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ mongosh "mongodb+srv://cluster0.tud3r.mongodb.net/" --apiVersion 1 --username mariacs22 --password 8Ja9u6zIrEjzBlot
Current Mongosh Log ID: 67cffdffcba3c86154c08e7d
Connecting to: mongodb+srv://<credentials>@cluster0.tud3r.mongodb.net/?appName=mongosh+2.2.1
Using MongoDB: 8.0.5 (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-vvji9i-shard-0 [primary] test> use myDB;
switched to db myDB
Atlas atlas-vvji9i-shard-0 [primary] myDB> mongoexport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudName,Grade,Hobbies"
Uncaught:
SyntaxError: Missing semicolon. (1:14)
> 1 | mongoexport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudName,Grade,Hobbies"
2 |
Atlas atlas-vvji9i-shard-0 [primary] myDB> mongoexport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudName,Grade,Hobbies"
Uncaught:
SyntaxError: Missing semicolon. (1:14)
> 1 | mongoexport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudName,Grade,Hobbies"
2 |
Atlas atlas-vvji9i-shard-0 [primary] myDB>
(To exit, press Ctrl+C again or Ctrl+D or type .exit)
Atlas atlas-vvji9i-shard-0 [primary] myDB> mongoexport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudName,Grade,Hobbies"
Uncaught:
SyntaxError: Missing semicolon. (1:14)
> 1 | mongoexport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudName,Grade,Hobbies"
2 |
Atlas atlas-vvji9i-shard-0 [primary] myDB> exit
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ mongoexport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudName,Grade,Hobbies"
2025-03-11T14:42:20.815+0530 connected to: mongodb+srv://[**REDACTED**]@cluster0.tud3r.mongodb.net/myDB
2025-03-11T14:42:21.068+0530 exported 2 records
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $


Activities
Terminal
Mar 11 15:26
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~
Atlas atlas-vvji9i-shard-0 [primary] myDB>
(To exit, press Ctrl+C again or Ctrl+D or type .exit)
Atlas atlas-vvji9i-shard-0 [primary] myDB> mongoexport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudName,Grade,Hobbies"
Uncaught:
SyntaxError: Missing semicolon. (1:14)
> 1 | mongoexport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudName,Grade,Hobbies"
2 |
Atlas atlas-vvji9i-shard-0 [primary] myDB> exit
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ mongoexport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudName,Grade,Hobbies"
2025-03-11T14:42:20.815+0530 connected to: mongodb+srv://[**REDACTED**]@cluster0.tud3r.mongodb.net/myDB
2025-03-11T14:42:21.068+0530 exported 2 records
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ mongoexport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/out.txt" --fields "StudName,Grade,Hobbies"
2025-03-11T14:43:35.174+0530 connected to: mongodb+srv://[**REDACTED**]@cluster0.tud3r.mongodb.net/myDB
2025-03-11T14:43:35.437+0530 exported 2 records
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ mongoexport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/qwer.txt" --fields "StudName,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
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ whoaml
bmscecse
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ ^C
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ mongoexport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "/home/bmscecse/Desktop/qwer.csv" --fields "StudName,Grade,Hobbies"
2025-03-11T14:47:52.993+0530 connected to: mongodb+srv://[**REDACTED**]@cluster0.tud3r.mongodb.net/myDB
2025-03-11T14:47:53.277+0530 exported 2 records
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ mongoimport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student3 --type csv --file "/home/bmscecse/Desktop/qwer.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.
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ mongoimport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --headerline --file ~/Downloads/airline.csv
2025-03-11T14:55:09.112+0530 Failed: open /home/bmscecse/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.
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ mongoimport --uri "mongodb+srv://mariacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --headerline --file ~/Desktop/airline.csv
2025-03-11T14:56:05.413+0530 Failed: open /home/bmscecse/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 15:26
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~
dName,Grade,Hobbies"
Uncaught:
SyntaxError: Missing semicolon. (1:14)
> 1 | mongoexport --uri "mongodb+srv://marlacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudName,Grade,Hobbies"
| ^
2 |
Atlas atlas-vvj19i-shard-0 [primary] myDB> exit
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ mongoexport --uri "mongodb+srv://marlacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/output.txt" --fields "StudName,Grade,Hobbies"
2025-03-11T14:42:20.815+0530 connected to: mongodb+srv://[**REDACTED**]@cluster0.tud3r.mongodb.net/myDB
2025-03-11T14:42:21.868+0530 exported 2 records
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ mongoexport --uri "mongodb+srv://marlacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/out.txt" --fields "StudName,Grade,Hobbies"
2025-03-11T14:43:35.174+0530 connected to: mongodb+srv://[**REDACTED**]@cluster0.tud3r.mongodb.net/myDB
2025-03-11T14:43:35.437+0530 exported 2 records
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ mongoexport --uri "mongodb+srv://marlacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "D:/qwer.txt" --fields "StudName,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
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ whoami
bmscecse
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ ^C
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ mongoexport --uri "mongodb+srv://marlacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --out "/home/bmscecse/Desktop/qwer.csv" --fields "StudName,Grade,Hobbies"
2025-03-11T14:47:52.993+0530 connected to: mongodb+srv://[**REDACTED**]@cluster0.tud3r.mongodb.net/myDB
2025-03-11T14:47:53.277+0530 exported 2 records
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ mongoinport --uri "mongodb+srv://marlacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student3 --type csv --file "/home/bmscecse/Desktop/qwer.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.
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ mongoinport --uri "mongodb+srv://marlacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --headerline --file ~/Downloads/airline.csv
2025-03-11T14:55:09.112+0530 Failed: open /home/bmscecse/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.
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ mongoinport --uri "mongodb+srv://marlacs22:8Ja9u6zIrEjzBlot@cluster0.tud3r.mongodb.net/myDB" --collection Student1 --type csv --headerline --file ~/Desktop/airline.csv
2025-03-11T14:56:05.413+0530 Failed: open /home/bmscecse/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.
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ mongoexport --host localhost --db myDB --collection Student1 --type=csv --out ~/Downloads/airline.csv --fields "Year","Quarter"
2025-03-11T14:58:48.870+0530 connected to: mongodb://localhost/
2025-03-11T14:58:48.871+0530 exported 0 records
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $


Activities
soffice
Mar 11 15:27
LibreOffice 7.5
LibreOffice
Community
Text Import - [qwer.csv]
Import
Character set:
Unicode (UTF-8)
Language:
Default - English (India)
From row:
Separator Options
Fixed width
Separated by
Tab
Comma
Semicolon
Space
Other
Merge delimiters
Trim spaces
String delimiter:
Other Options
Format quoted field as text
Detect special numbers
Evaluate formulas
Fields
Column type:
Standard
Standard
Standard
1 StudName Grade Hobbies
2 MichelleJacintha VII InternetSurfing
3 AryanDavid VII Chess
Help
Cancel
OK
Sgpa.class
Sgpa.java
qwer.csv
Student.java
Student.class
main.java
Student1.java
calsgpa.java
Quad.class
calsSgpa.java
QuadraticEquationSolver.class
CalSgpa.java
Quad.java
calsgpa.class
Home



RollNo : 3
Age : 21
Cont : 5576
email : "anubhav.de9@gmail.com"

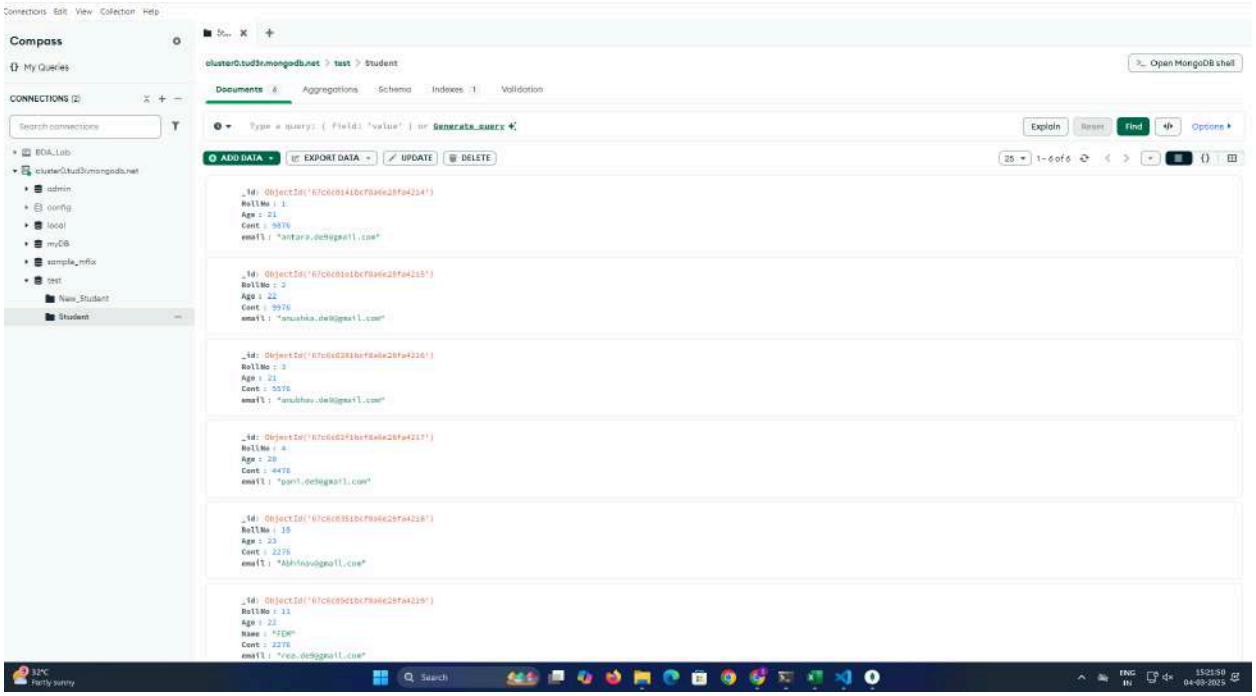
_id: ObjectId('67c6c91a1bcf8a6e28fa4215')
RollNo : 2
Age : 22
Cont : 9976
email : "anushka.de9@gmail.com"

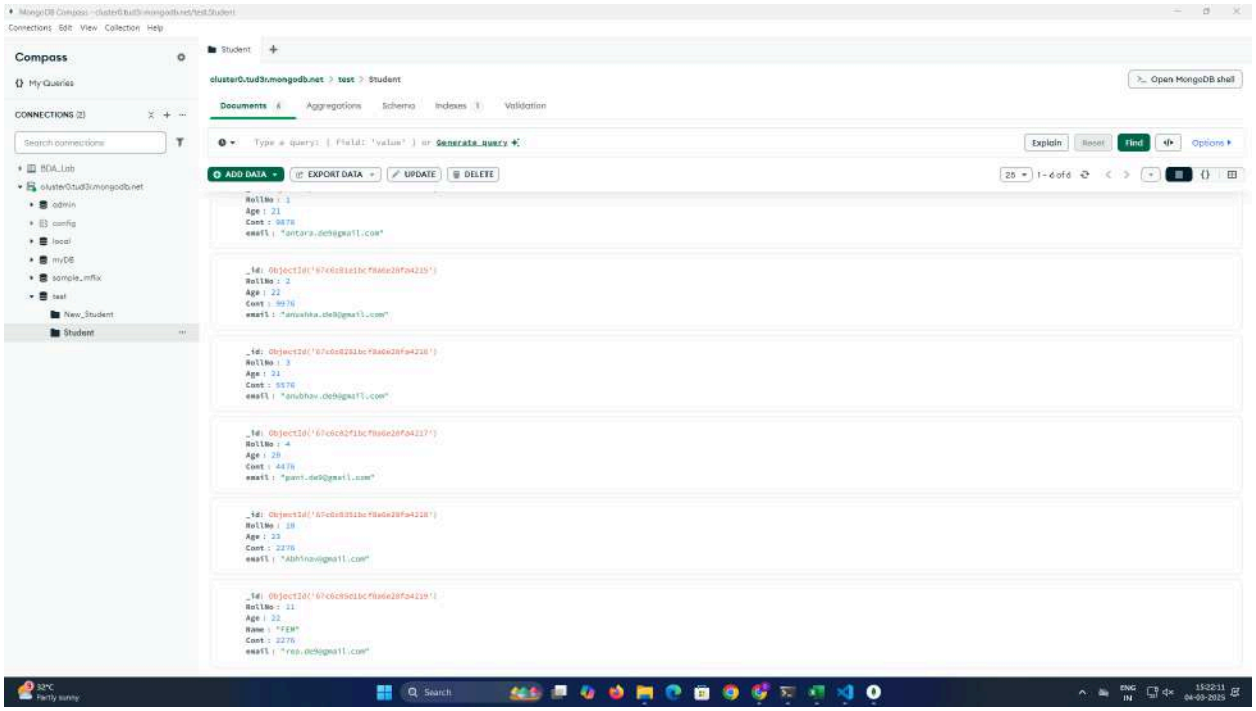
_id: ObjectId('67c6c9141bcf8a6e28fa4214')
RollNo : 1
Age : 21
Cont : 9876
email : "antara.de9@gmail.com"

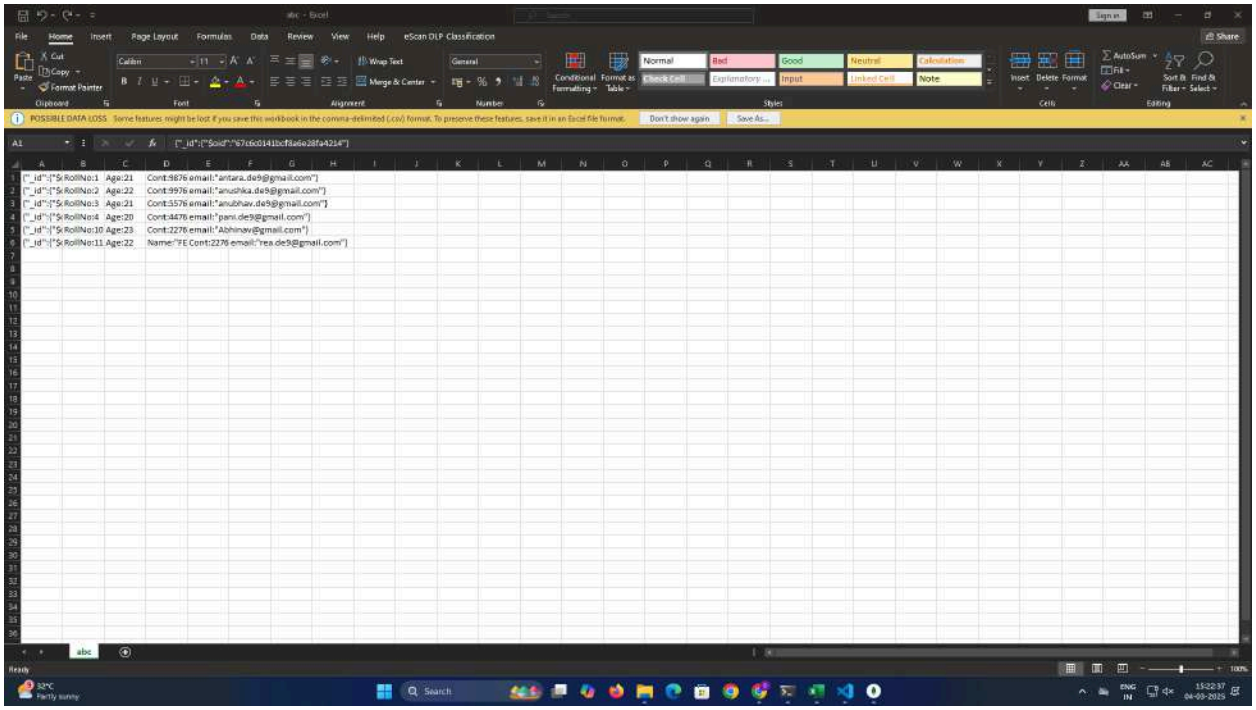
_id: ObjectId('67c6c95d1bcf8a6e28fa4219')
RollNo : 11
Age : 22
Name : "TEN"
Cont : 2276
email : "roa.de9@gmail.com"

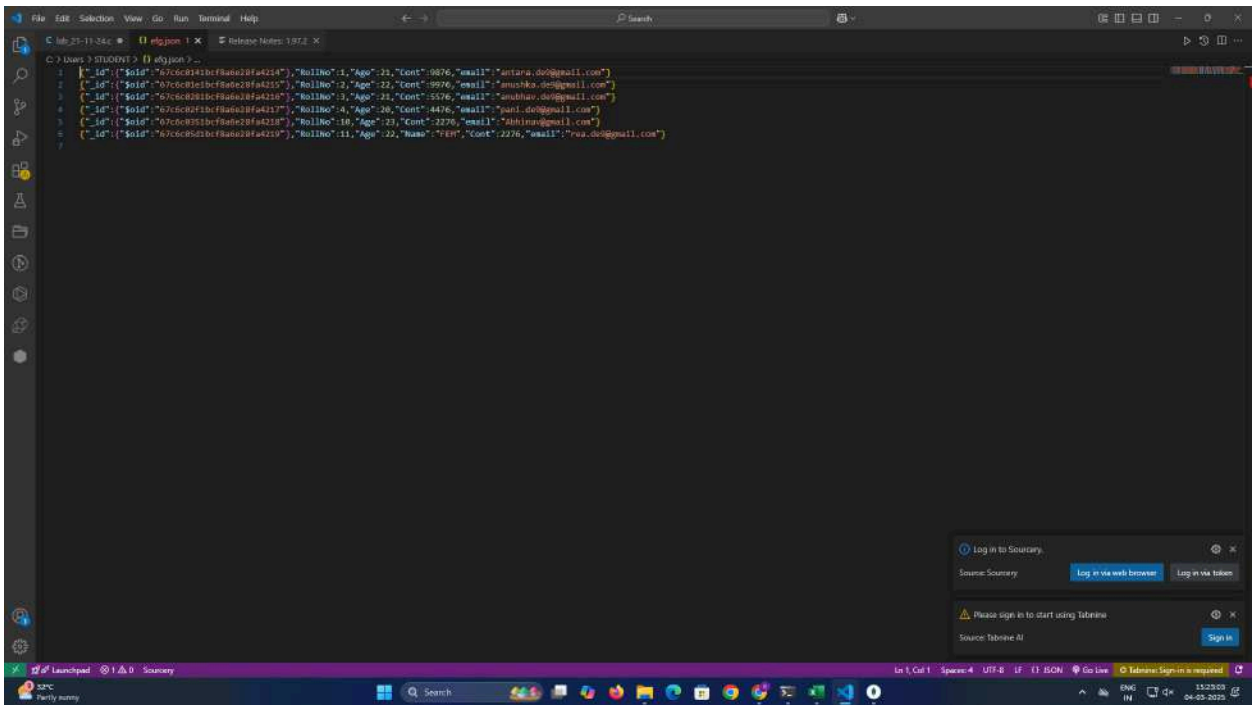
_id: ObjectId('67c6c91f1bcf8a6e28fa4217')
RollNo : 4
Age : 20
Cont : 4476
email : "pan1.de9@gmail.com"

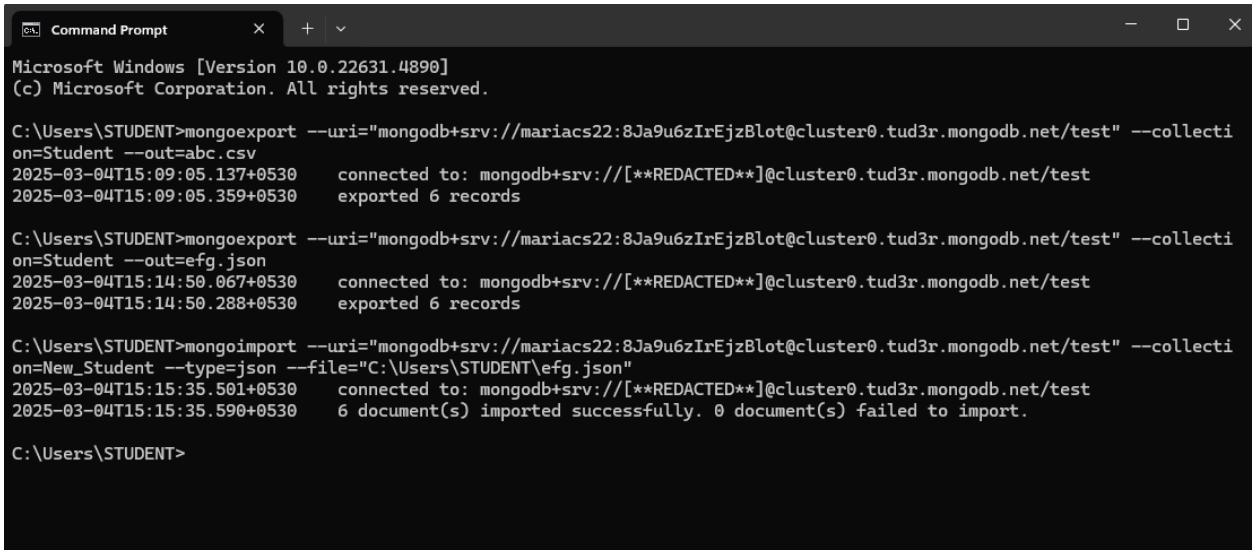
_id: ObjectId('67c6c9351bcf8a6e28fa4218')
RollNo : 10
Age : 23
Cont : 2276
email : "Abhinav@gmail.com"


Compass
My Queries
CONNECTIONS (2)
cluster0.tud3r.mongodb.net > test > Student
Open MongoDB shell
Documents
Aggregations
Schema
Indexes
Validation
Explain
Reset
Find
Options
ADD DATA
EXPORT DATA
UPDATE
DELETE
New_Student
Student


Compass
My Queries
CONNECTIONS (2)
cluster0.tud3r.mongodb.net > test > Student
Open MongoDB shell
Documents
Aggregations
Schema
Indexes
Validation
Explain
Reset
Find
Options
ADD DATA
EXPORT DATA
UPDATE
DELETE
New_Student
Student






Command Prompt
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" --collecti
on=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" --collecti
on=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" --collecti
on=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

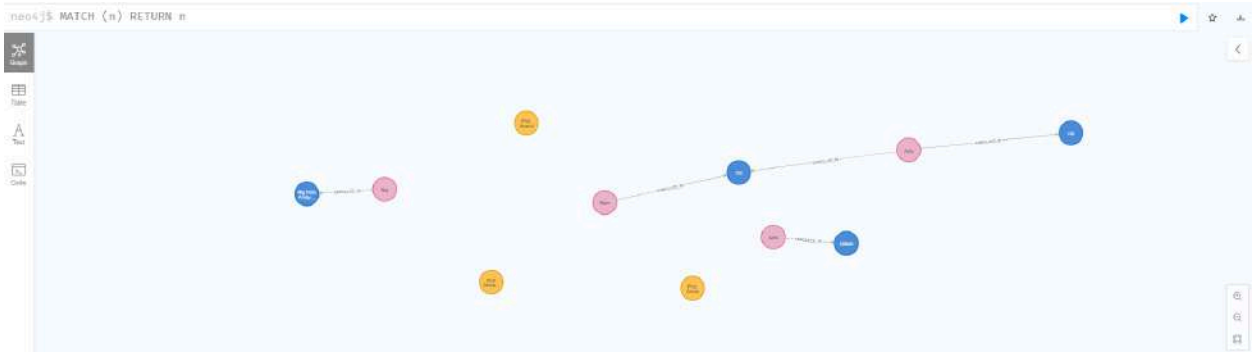
Neo4j

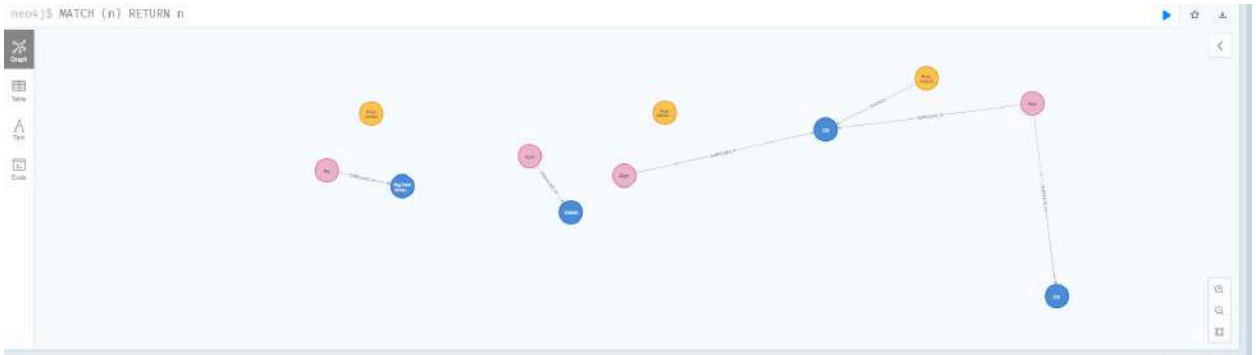
Code with Screenshots:

neo4j$ CREATE (a:STUDENT {name:"Abhi" , id: 14})
Table
Code
Added 1 label, created 1 node, set 2 properties, completed after 6 ms.
Added 1 label, created 1 node, set 2 properties, completed after 6 ms.

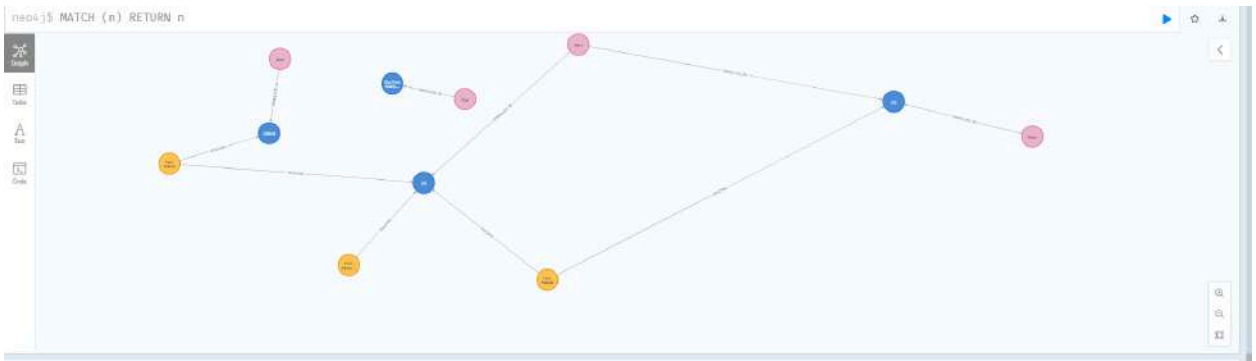
neo4j$ CREATE (a:STUDENT {name:"Raj" , id: 13})
Table
Code
Added 1 label, created 1 node, set 2 properties, completed after 6 ms.
Added 1 label, created 1 node, set 2 properties, completed after 6 ms.

neo4j$ CREATE (a:STUDENT {name:"Ajay" , id: 15})
Table
Code
Added 1 label, created 1 node, set 2 properties, completed after 6 ms.
Added 1 label, created 1 node, set 2 properties, completed after 6 ms.

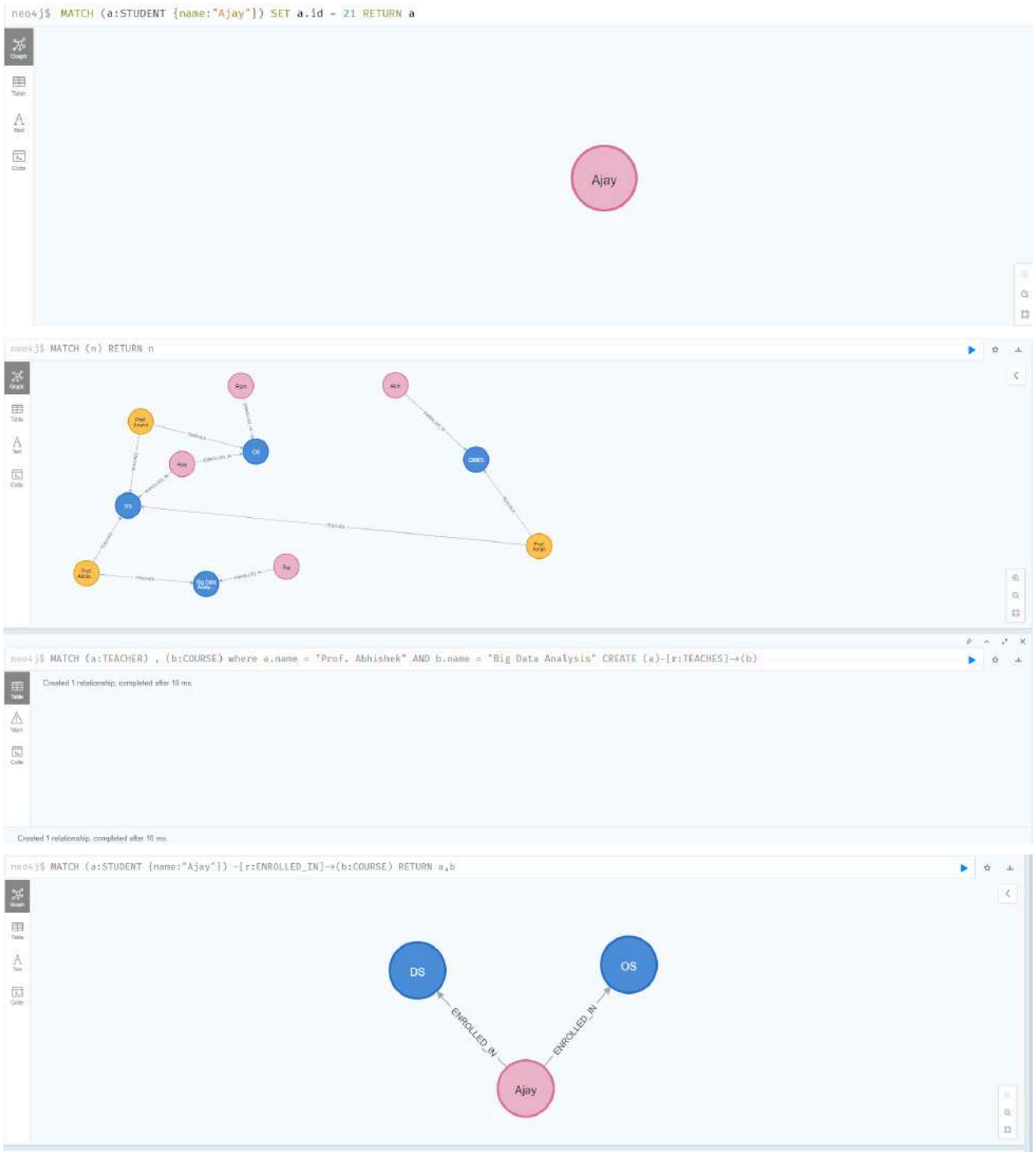

neo4j$ MATCH (n) RETURN n

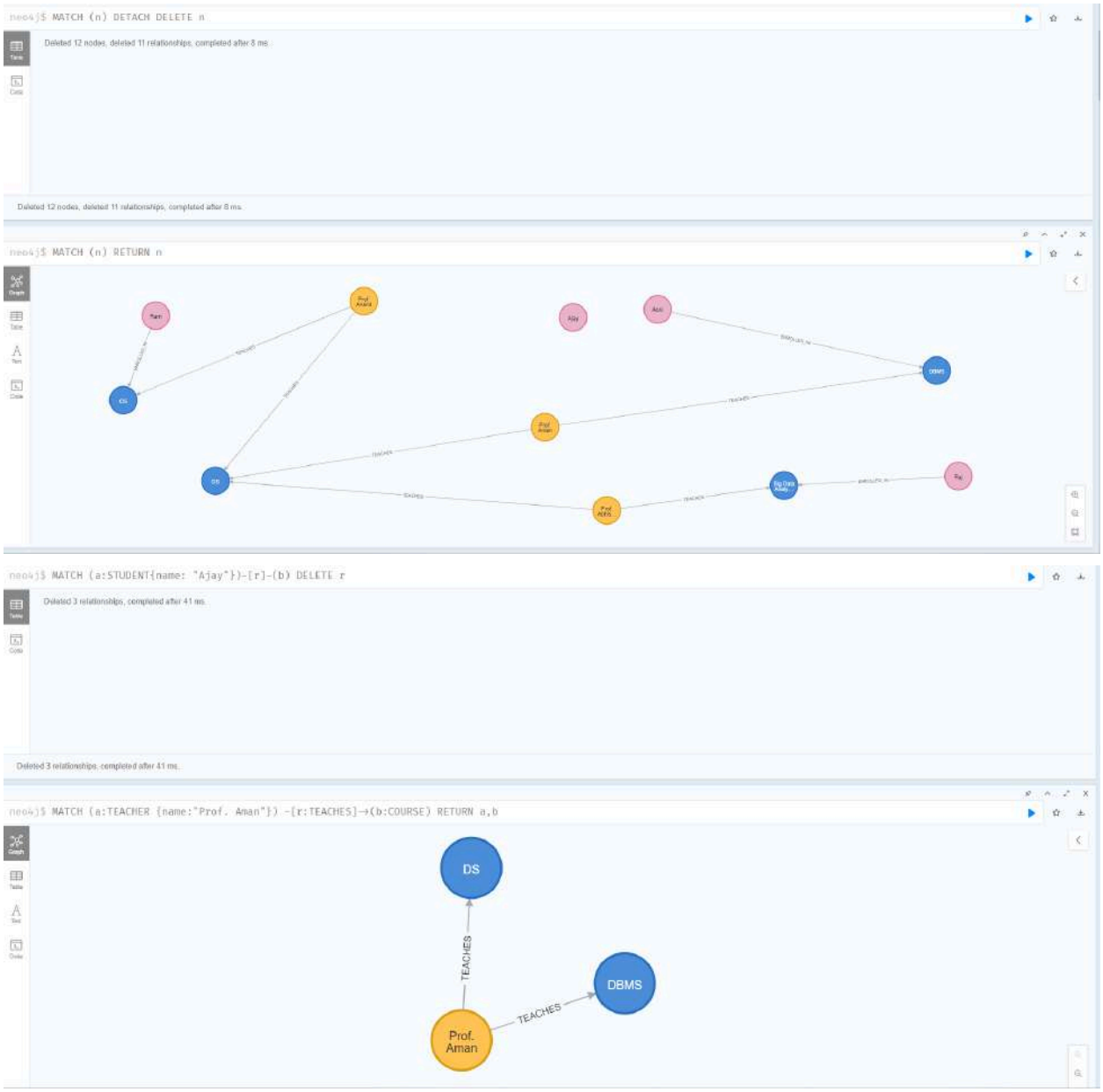

neo4j$ MATCH (n) RETURN n

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.
Created 2 relationships, completed after 12 ms.


neo4j$ MATCH (n) RETURN n

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.


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


neo4j$ MATCH (n) DETACH DELETE n
Deleted 12 nodes, deleted 11 relationships, completed after 8 ms.
neo4j$ MATCH (n) RETURN n
neo4j$ MATCH (a:STUDENT{name: "Ajay"})-[r]-(b) DELETE r
Deleted 3 relationships, completed after 41 ms.
neo4j$ MATCH (a:TEACHER {name:"Prof. Aman"}) -[r:TEACHES]→(b:COURSE) RETURN a,b
DS
TEACHES
DBMS
TEACHES
Prof. Aman

EXPERIMENT-3

Perform the following DB operations using Cassandra.

a) Create a keyspace by name Employee

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

c) Insert the values into the table in batch

d) Update Employee name and Department of Emp-Id 121

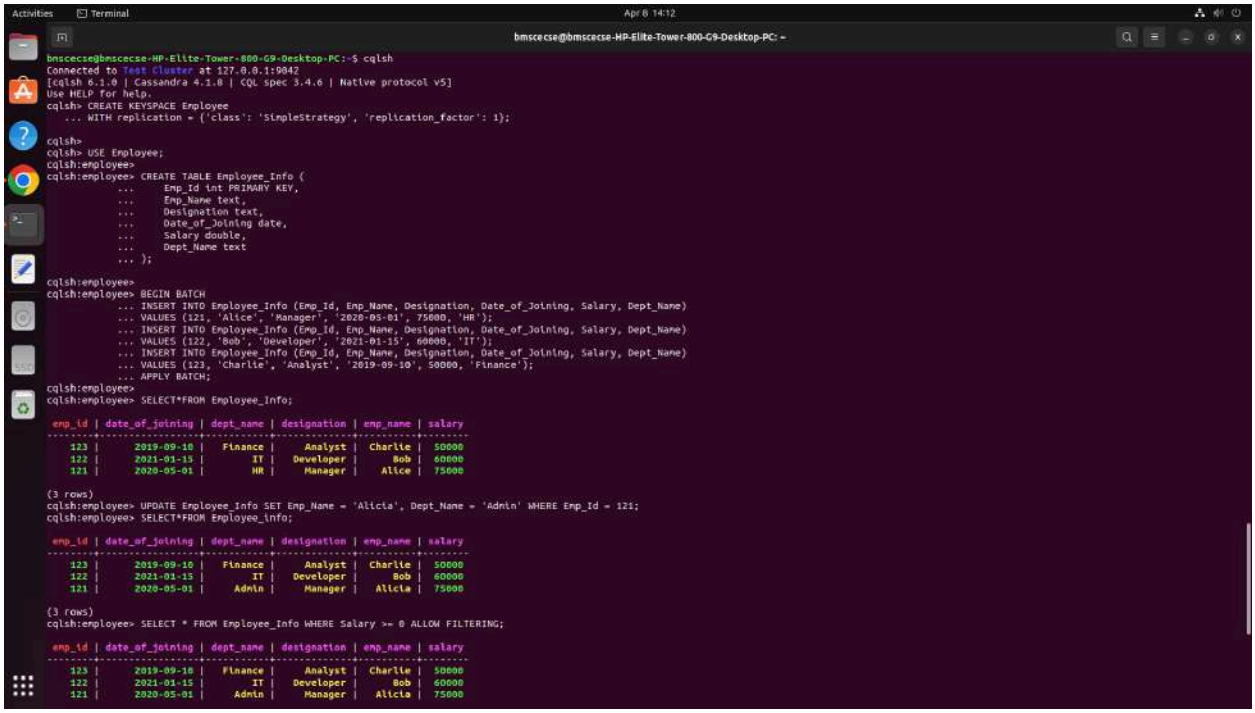
e) Sort the details of Employee records based on salary

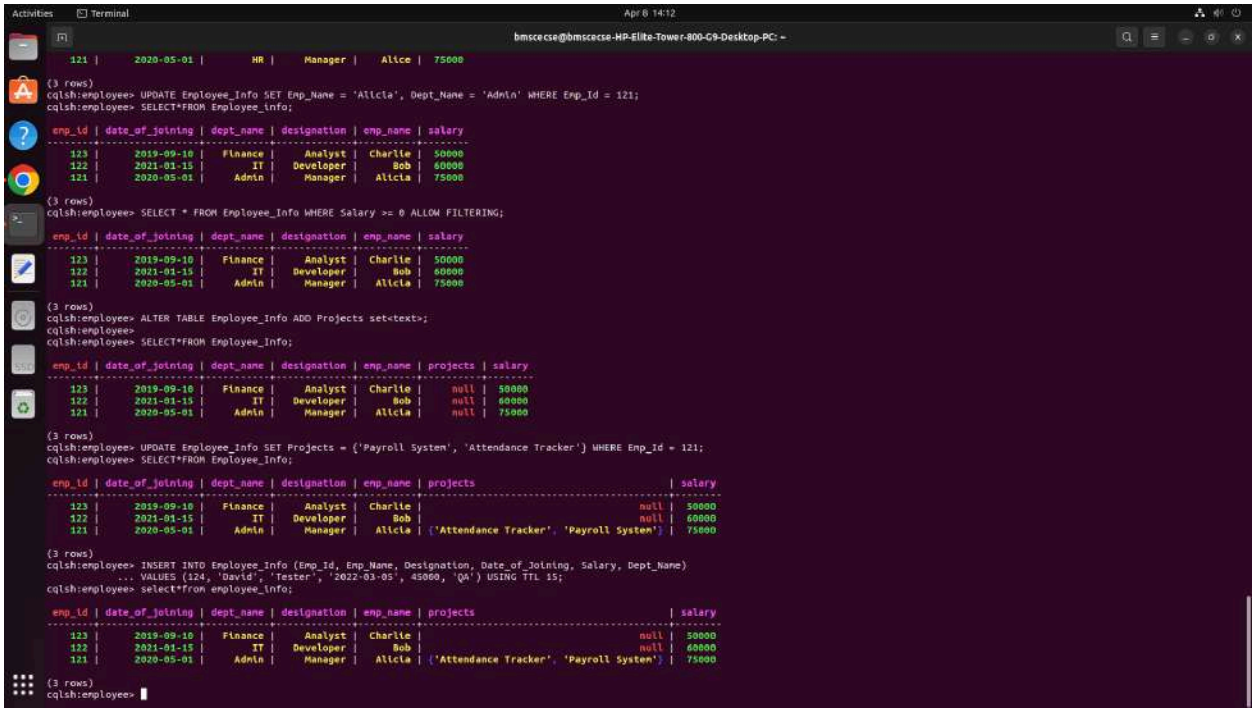
f) Alter the schema of the table Employee_Info to add a column Projects which stores a set of Projects done by the corresponding Employee.

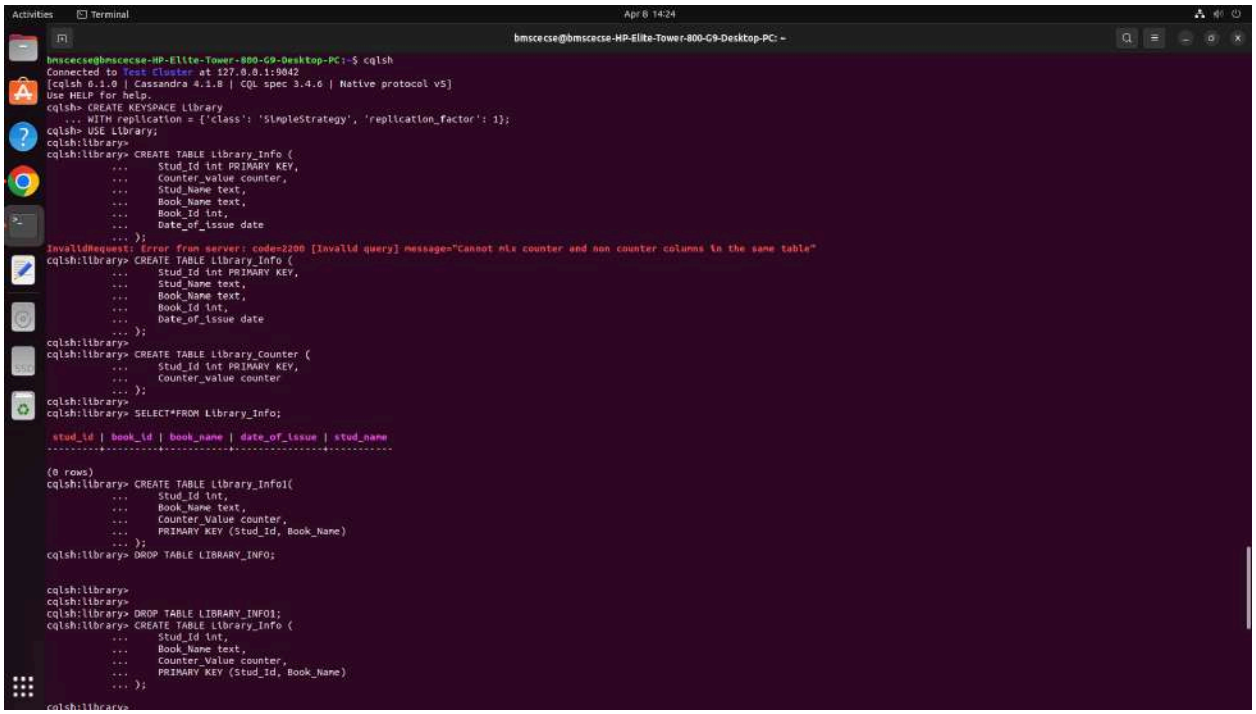
g) Update the altered table to add project names.

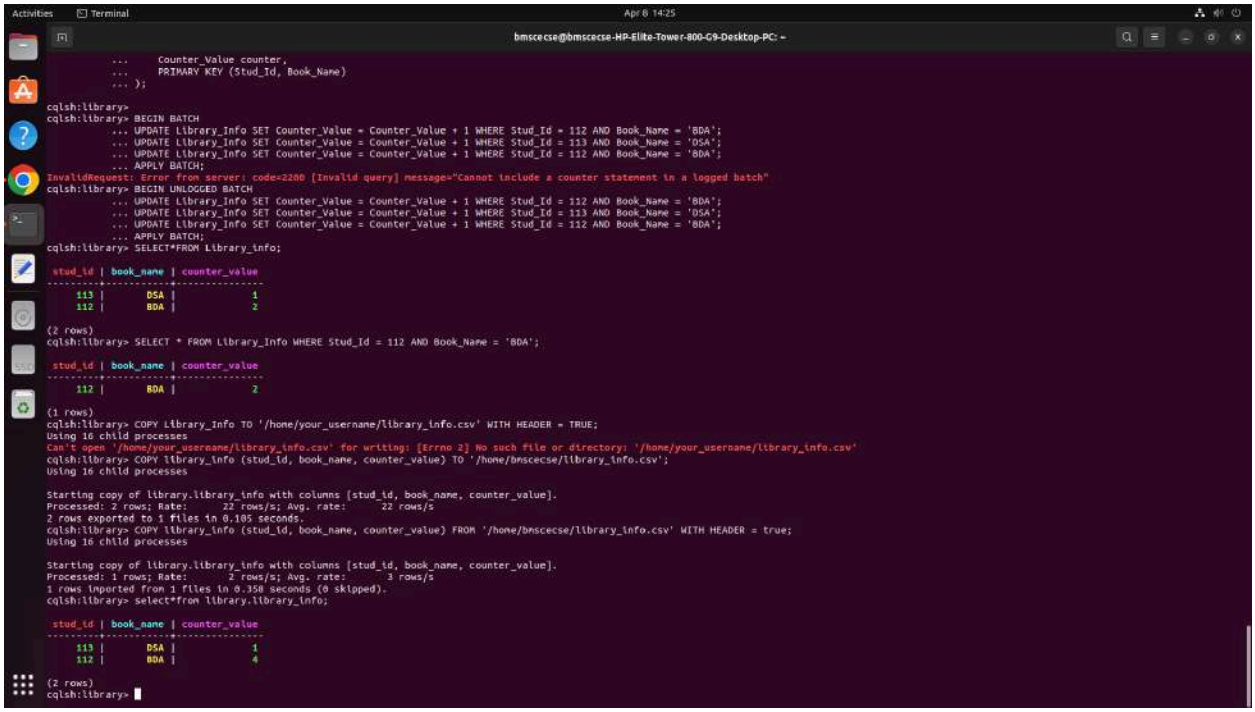
h) Create a TTL of 15 seconds to display the values of Employees.

CODE WITH SCREENSHOTS:



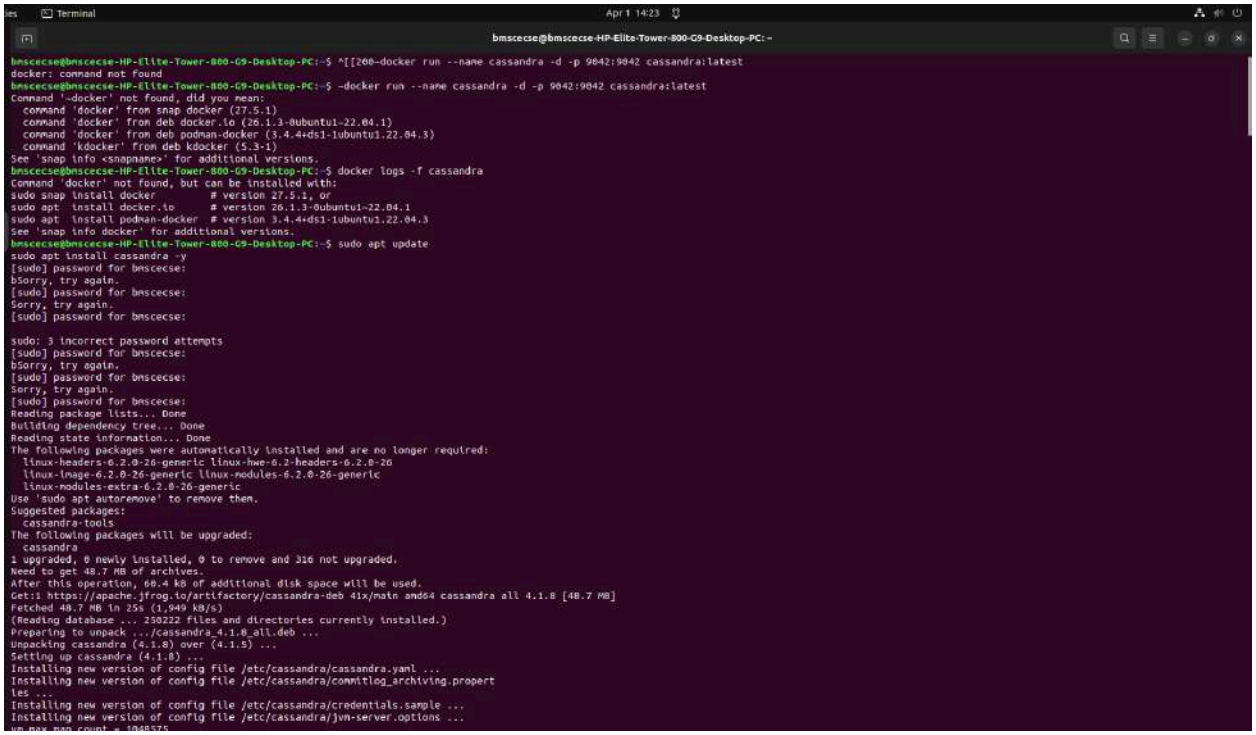

Activities
Terminal
Apr 8 14:12
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~
121 | 2020-05-01 | HR | Manager | Alice | 75000
(3 rows)
cqlsh:employee> UPDATE Employee_Info SET Emp_Name = 'Alicia', Dept_Name = 'Admin' WHERE Emp_Id = 121;
cqlsh:employee> SELECT*FROM Employee_info;
emp_id | date_of_joining | 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 >= 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)
cqlsh:employee> ALTER TABLE Employee_Info ADD Projects set<text>;
cqlsh:employee>
cqlsh:employee> 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)
cqlsh:employee> UPDATE Employee_Info SET Projects = {'Payroll System', 'Attendance Tracker'} WHERE Emp_Id = 121;
cqlsh:employee> 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)
cqlsh:employee> INSERT INTO Employee_Info (Emp_Id, Emp_Name, Designation, Date_of_Joining, Salary, Dept_Name)
... VALUES (124, 'David', 'Tester', '2022-03-05', 45000, 'QA') USING TTL 15;
cqlsh:employee> 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)
cqlsh:employee>

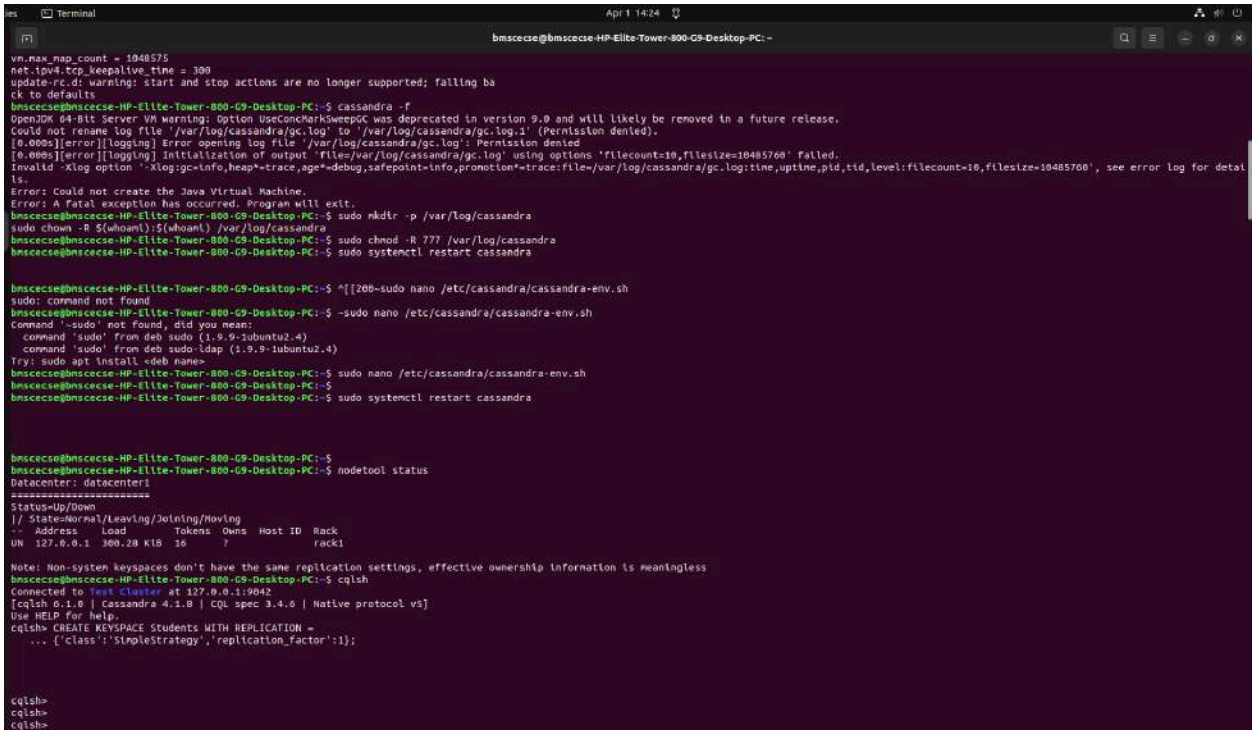

Activities
Terminal
Apr 8 14:24
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ cqlsh
Connected to Test Cluster at 127.0.0.1:9042
[cqlsh 6.1.0 | Cassandra 4.1.8 | CQL spec 3.4.6 | Native protocol v5]
Use HELP for help.
cqlsh> CREATE KEYSPACE Library
... WITH replication = {'class': 'SimpleStrategy', 'replication_factor': 1};
cqlsh> USE Library;
cqlsh:library>
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=2200 [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>
cqlsh:library> CREATE TABLE Library_Counter (
... Stud_Id int PRIMARY KEY,
... Counter_value counter
...);
cqlsh:library>
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> DROP TABLE LIBRARY_INFO1;
cqlsh:library> CREATE TABLE Library_Info (
... Stud_Id int,
... Book_Name text,
... Counter_Value counter,
... PRIMARY KEY (Stud_Id, Book_Name)
...);
cqlsh:library>

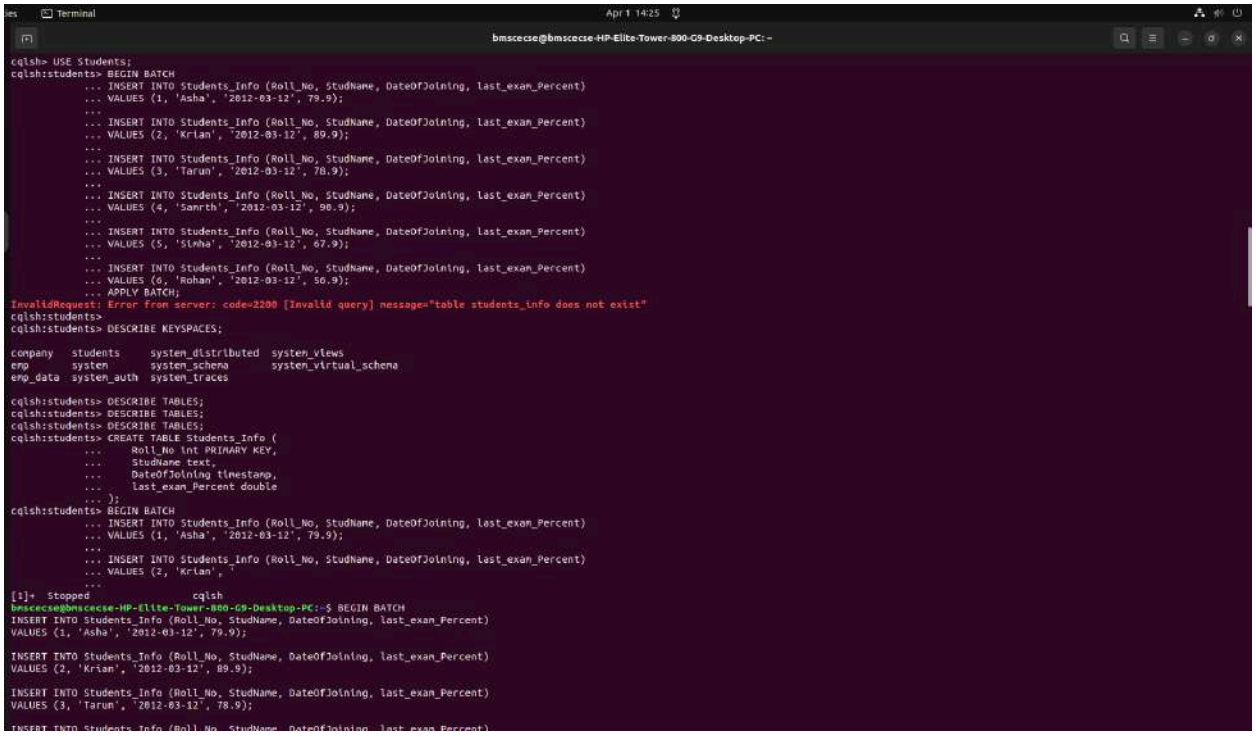

Activities
Terminal
Apr 8 14:25
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~
... Counter_Value counter,
... PRIMARY KEY (Stud_Id, Book_Name)
...);
cqlsh:library>
cqlsh:library> BEGIN BATCH
... UPDATE Library_Info SET Counter_Value = Counter_Value + 1 WHERE Stud_Id = 112 AND Book_Name = 'BDA';
... 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 = 'BDA';
... APPLY BATCH;
InvalidRequest: Error from server: code=2200 [Invalid query] message="Cannot include a counter statement in a logged batch"
cqlsh:library> BEGIN UNLOGGED 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 = 'BDA';
... 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';
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> COPY library_info (stud_id, book_name, counter_value) FROM '/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: 1 rows; Rate: 2 rows/s; Avg. rate: 3 rows/s
1 rows imported from 1 files in 0.350 seconds (0 skipped).
cqlsh:library> select*from library.library_info;
stud_id | book_name | counter_value
113 | DSA | 1
112 | BDA | 4
(2 rows)
cqlsh:library>

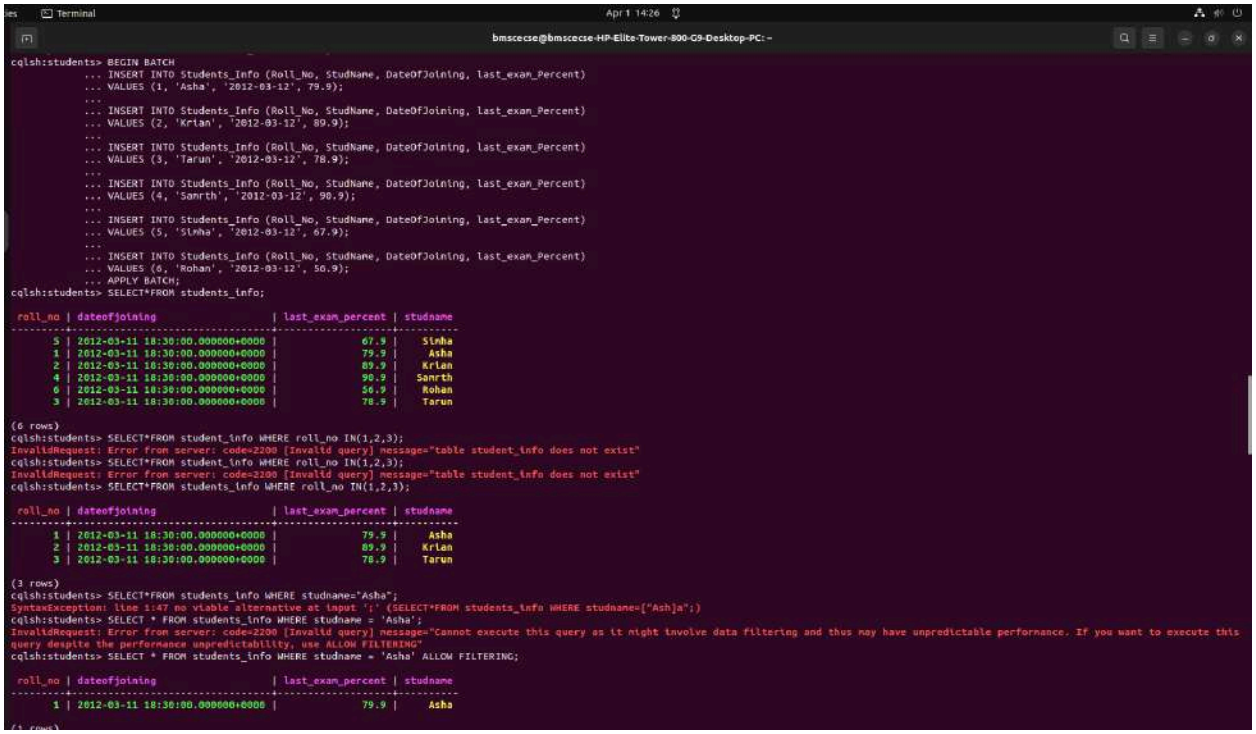
EXPERIMENT:4

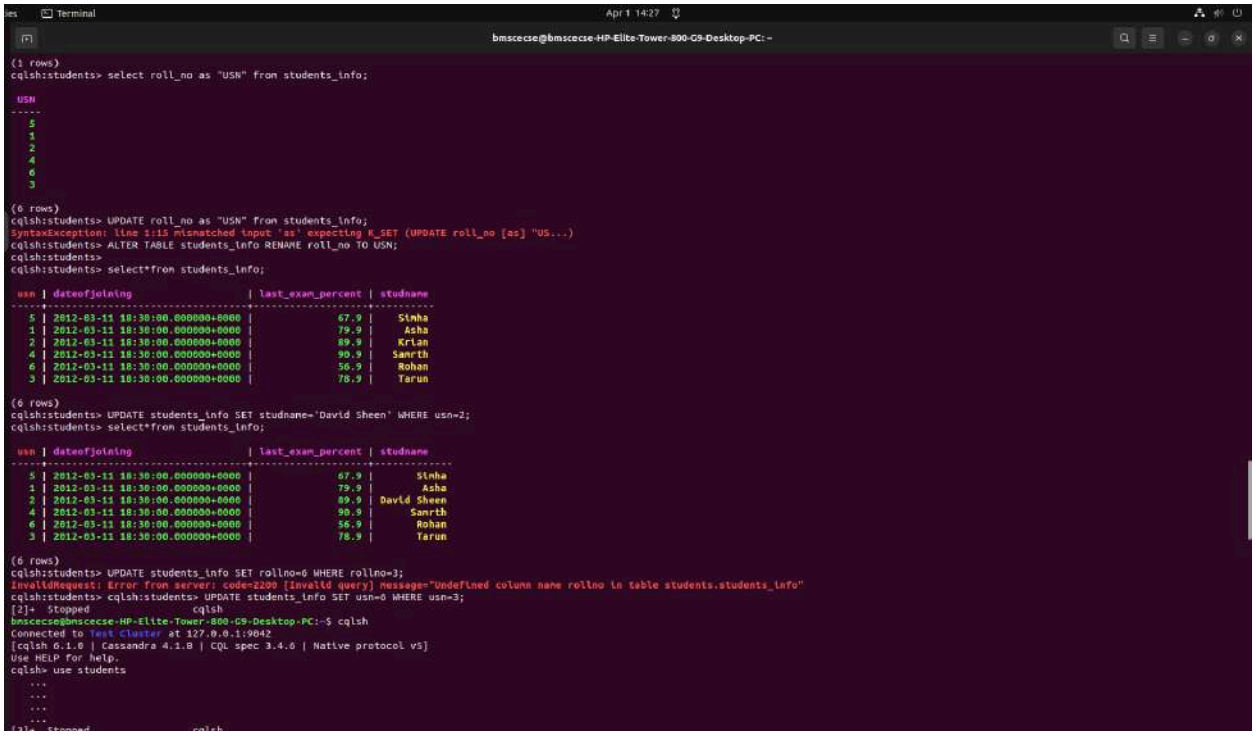
Perform the following DB operations using Cassandra.
a) Create a keyspace by name Library
b) Create a column family by name Library-Info with attributes
Stud_Id Primary Key,
Counter_value of type Counter,
Stud_Name, Book-Name, Book-Id,
Date_of_issue
c) Insert the values into the table in batch
d) Display 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

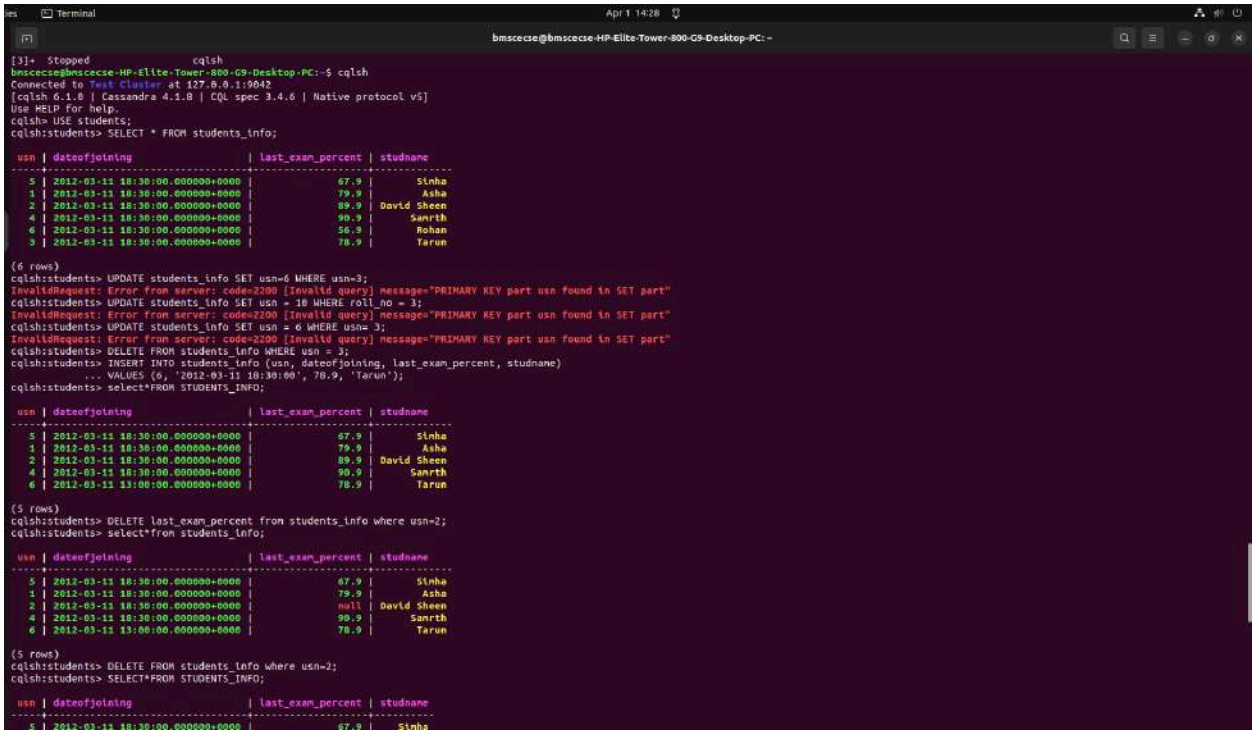


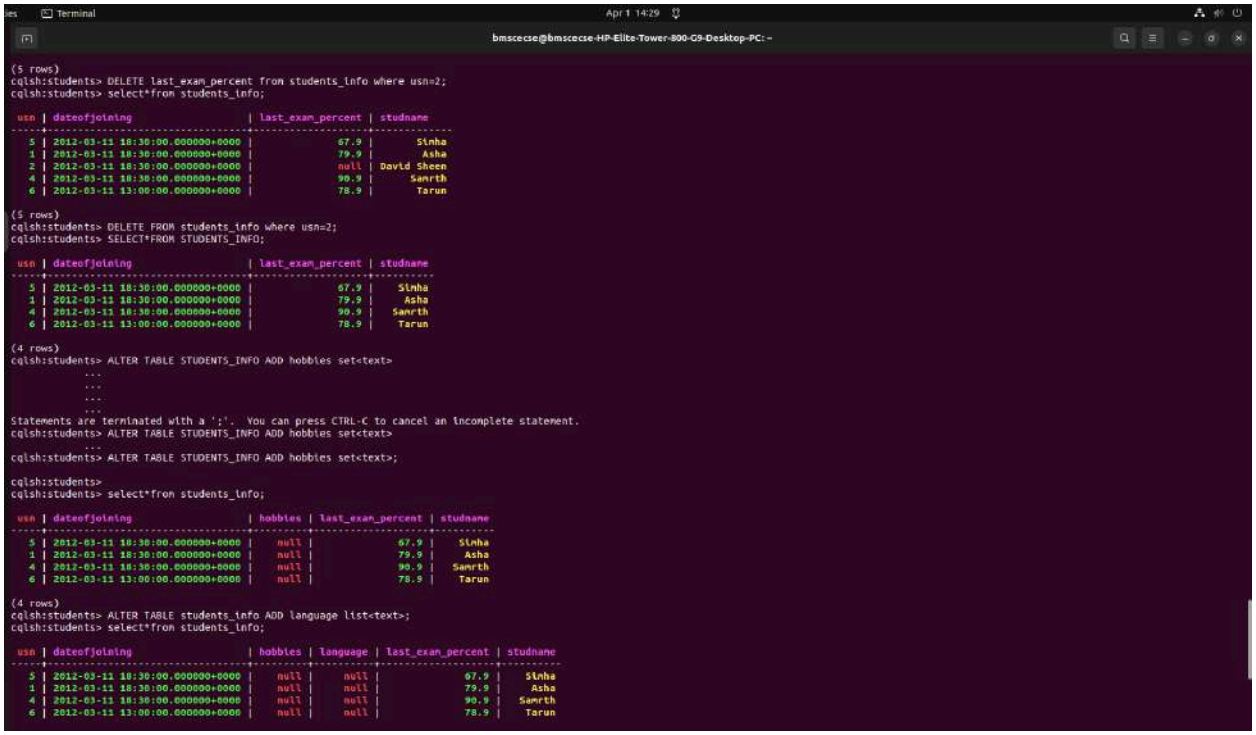

Terminal
Apr 1 14:24
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~
vm.max_map_count = 1048575
net.ipv4.tcp_keepalive_time = 300
update-rc.d: warning: start and stop actions are no longer supported; falling ba
ck to defaults
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ cassandra -f
OpenJDK 64-Bit Server VM warning: Option UseConcMarkSweepGC was deprecated in version 9.0 and will likely be removed in a future release.
Could not rename log file '/var/log/cassandra/gc.log' to '/var/log/cassandra/gc.log.1' (Permission denied).
[0.008s][error][logging] Error opening log file '/var/log/cassandra/gc.log': Permission denied
[0.008s][error][logging] Initialization of output 'file=/var/log/cassandra/gc.log' using options 'filecount=10,filesize=10485760' failed.
Invalid -Xlog option '-Xlog:gc=info,heap*=trace,age*=debug,safepoint=info,promotion*=trace:file=/var/log/cassandra/gc.log:time,uptime,pid,tid,level:filecount=10,filesize=10485760', see error log for detai
ls.
Error: Could not create the Java Virtual Machine.
Error: A fatal exception has occurred. Program will exit.
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ sudo mkdir -p /var/log/cassandra
sudo chown -R $(whoami):$(whoami) /var/log/cassandra
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ sudo chmod -R 777 /var/log/cassandra
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ sudo systemctl restart cassandra
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ ^[[200~sudo nano /etc/cassandra/cassandra-env.sh
sudo: command not found
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ ~sudo nano /etc/cassandra/cassandra-env.sh
Command '~sudo' not found, did you mean:
command 'sudo' 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>
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ sudo nano /etc/cassandra/cassandra-env.sh
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ sudo systemctl restart cassandra
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ nodetool status
Datacenter: datacenter1
=======================
Status=Up/Down
|/ State=Normal/Leaving/Joining/Moving
-- Address   Load   Tokens Owns Host ID Rack
UN 127.0.0.1 300.28 KiB 16   ?   rack1
Note: Non-system keyspaces don't have the same replication settings, effective ownership information is meaningless
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ cqlsh
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 Students WITH REPLICATION =
... {'class':'SimpleStrategy','replication_factor':1};
cqlsh>
cqlsh>
cqlsh>

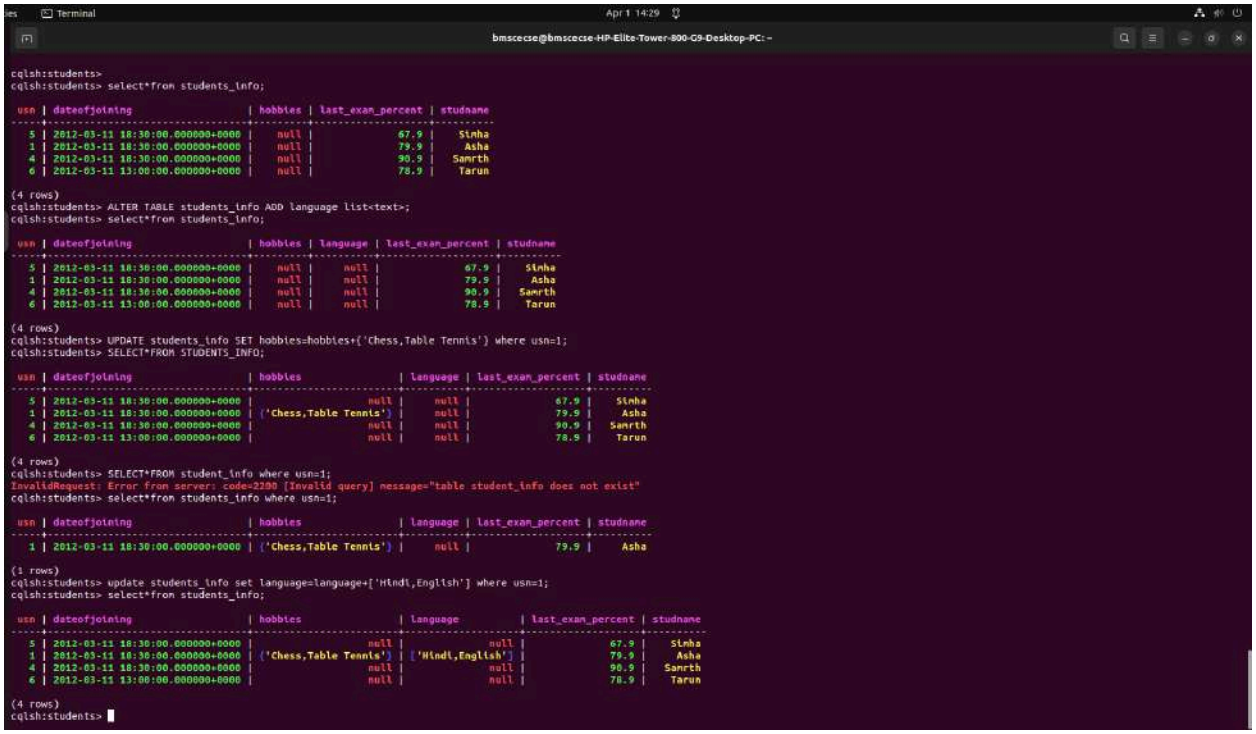

Terminal
Apr 1 14:25
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~
cqlsh> USE Students;
cqlsh:students> BEGIN BATCH
... INSERT INTO Students_Info (Roll_No, StudName, DateOfJoining, last_exan_Percent)
... VALUES (1, 'Asha', '2012-03-12', 79.9);
...
... INSERT INTO Students_Info (Roll_No, StudName, DateOfJoining, last_exan_Percent)
... VALUES (2, 'Krian', '2012-03-12', 89.9);
...
... INSERT INTO Students_Info (Roll_No, StudName, DateOfJoining, last_exan_Percent)
... VALUES (3, 'Tarun', '2012-03-12', 78.9);
...
... INSERT INTO Students_Info (Roll_No, StudName, DateOfJoining, last_exan_Percent)
... VALUES (4, 'Sanrth', '2012-03-12', 90.9);
...
... INSERT INTO Students_Info (Roll_No, StudName, DateOfJoining, last_exan_Percent)
... VALUES (5, 'Sinha', '2012-03-12', 67.9);
...
... INSERT INTO Students_Info (Roll_No, StudName, DateOfJoining, last_exan_Percent)
... VALUES (6, 'Rohan', '2012-03-12', 50.9);
... APPLY BATCH;
InvalidRequest: Error from server: code=2200 [Invalid query] message="table students_info does not exist"
cqlsh:students>
cqlsh:students> DESCRIBE KEYSPACES;
company students   system_distributed system_views
emp   system   system_schema   system_virtual_schema
emp_data system_auth system_traces
cqlsh:students> DESCRIBE TABLES;
cqlsh:students> DESCRIBE TABLES;
cqlsh:students> DESCRIBE TABLES;
cqlsh:students> CREATE TABLE Students_Info (
...   Roll_No int PRIMARY KEY,
...   StudName text,
...   DateOfJoining timestamp,
...   last_exan_Percent double
...);
cqlsh:students> BEGIN BATCH
... INSERT INTO Students_Info (Roll_No, StudName, DateOfJoining, last_exan_Percent)
... VALUES (1, 'Asha', '2012-03-12', 79.9);
...
... INSERT INTO Students_Info (Roll_No, StudName, DateOfJoining, last_exan_Percent)
... VALUES (2, 'Krian', '
...
[1]+ Stopped   cqlsh
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ BEGIN BATCH
INSERT INTO Students_Info (Roll_No, StudName, DateOfJoining, last_exan_Percent)
VALUES (1, 'Asha', '2012-03-12', 79.9);
INSERT INTO Students_Info (Roll_No, StudName, DateOfJoining, last_exan_Percent)
VALUES (2, 'Krian', '2012-03-12', 89.9);
INSERT INTO Students_Info (Roll_No, StudName, DateOfJoining, last_exan_Percent)
VALUES (3, 'Tarun', '2012-03-12', 78.9);

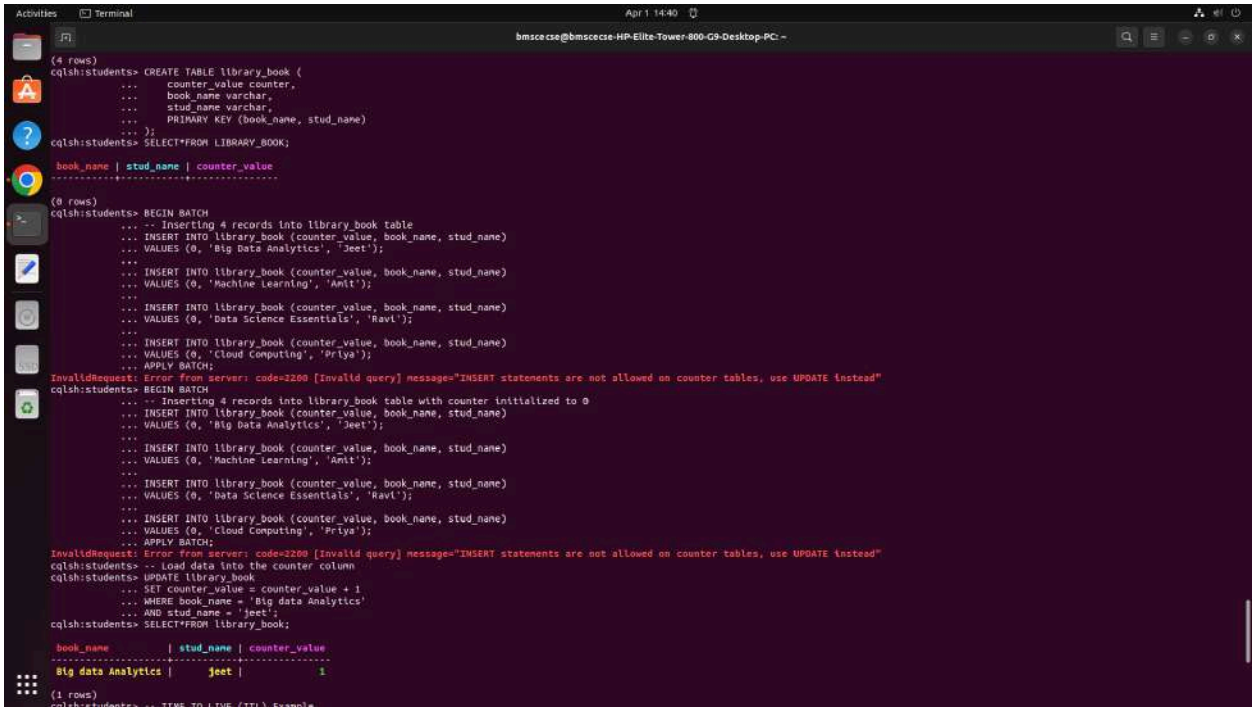

Terminal
Apr 1 14:26
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~
cqlsh:students> BEGIN BATCH
... INSERT INTO Students_Info (Roll_No, StudName, DateOfJoining, last_exan_Percent)
... VALUES (1, 'Asha', '2012-03-12', 79.9);
...
... INSERT INTO Students_Info (Roll_No, StudName, DateOfJoining, last_exan_Percent)
... VALUES (2, 'Krian', '2012-03-12', 89.9);
...
... INSERT INTO Students_Info (Roll_No, StudName, DateOfJoining, last_exan_Percent)
... VALUES (3, 'Tarun', '2012-03-12', 78.9);
...
... INSERT INTO Students_Info (Roll_No, StudName, DateOfJoining, last_exan_Percent)
... VALUES (4, 'Sanrth', '2012-03-12', 90.9);
...
... INSERT INTO Students_Info (Roll_No, StudName, DateOfJoining, last_exan_Percent)
... VALUES (5, 'Simha', '2012-03-12', 67.9);
...
... INSERT INTO Students_Info (Roll_No, StudName, DateOfJoining, last_exan_Percent)
... VALUES (6, 'Rohan', '2012-03-12', 56.9);
... APPLY BATCH;
cqlsh:students> SELECT*FROM students_info;
roll_no | dateofjoining | last_exam_percent | studname
5 | 2012-03-11 18:30:00.000000+0000 | 67.9 | Simha
1 | 2012-03-11 18:30:00.000000+0000 | 79.9 | Asha
2 | 2012-03-11 18:30:00.000000+0000 | 89.9 | Krian
4 | 2012-03-11 18:30:00.000000+0000 | 90.9 | Sanrth
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)
cqlsh: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"
cqlsh: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"
cqlsh: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 | Krian
3 | 2012-03-11 18:30:00.000000+0000 | 78.9 | Tarun
(3 rows)
cqlsh:students> SELECT*FROM students_info WHERE studname="Asha";
SyntaxException: line 1:47 no viable alternative at input ';' (SELECT*FROM students_info WHERE studname=["Ash]a";)
cqlsh: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"
cqlsh: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

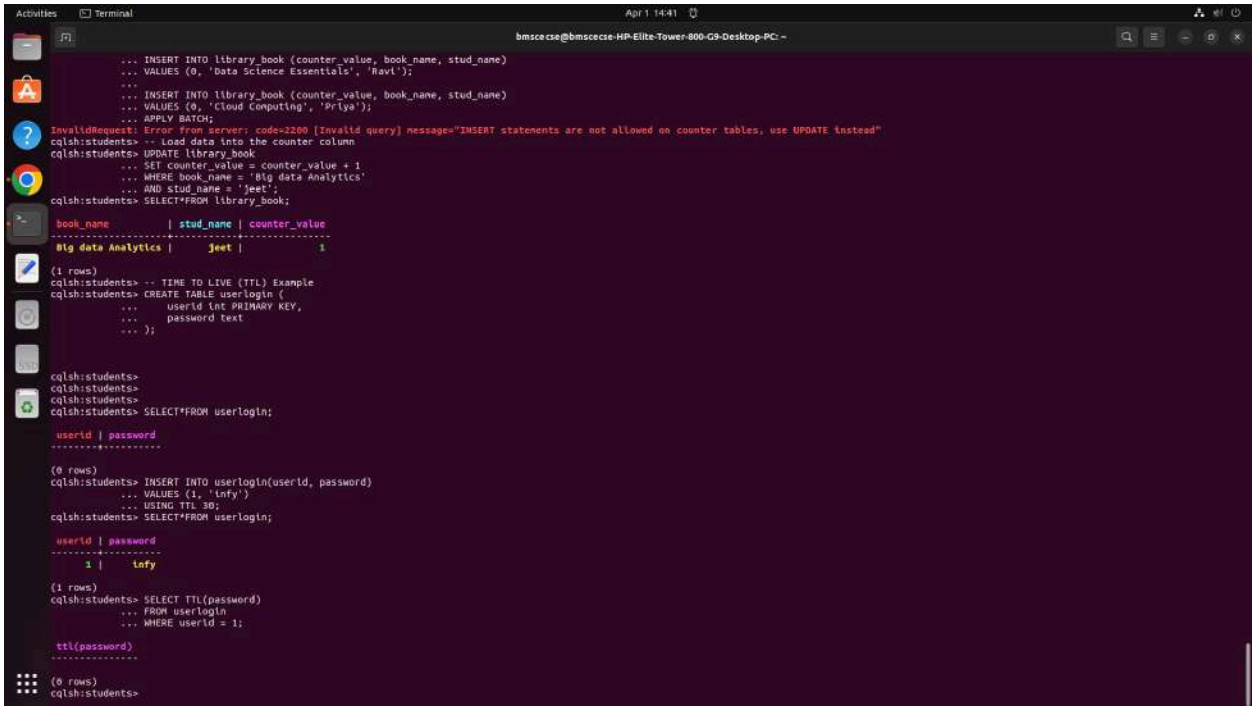

Terminal
Apr 1 14:27
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~
(1 rows)
cqlsh:students> select roll_no as "USN" from students_info;
USN
5
1
2
4
6
3
(6 rows)
cqlsh:students> UPDATE roll_no as "USN" from students_info;
SyntaxException: line 1:15 mismatched input 'as' expecting K_SET (UPDATE roll_no [as] "US...)
cqlsh:students> ALTER TABLE students_info RENAME roll_no TO USN;
cqlsh:students>
cqlsh:students> select*from students_info;
usn | dateofjoining | last_exam_percent | studname
5 | 2012-03-11 18:30:00.000000+0000 | 67.9 | Simha
1 | 2012-03-11 18:30:00.000000+0000 | 79.9 | Asha
2 | 2012-03-11 18:30:00.000000+0000 | 89.9 | Krian
4 | 2012-03-11 18:30:00.000000+0000 | 90.9 | Sanrth
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)
cqlsh:students> UPDATE students_info SET studname='David Sheen' 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 | Simha
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 | Sanrth
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)
cqlsh: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"
cqlsh:students> cqlsh:students> UPDATE students_info SET usn=6 WHERE usn=3;
[2]+ Stopped cqlsh
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ cqlsh
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> use students

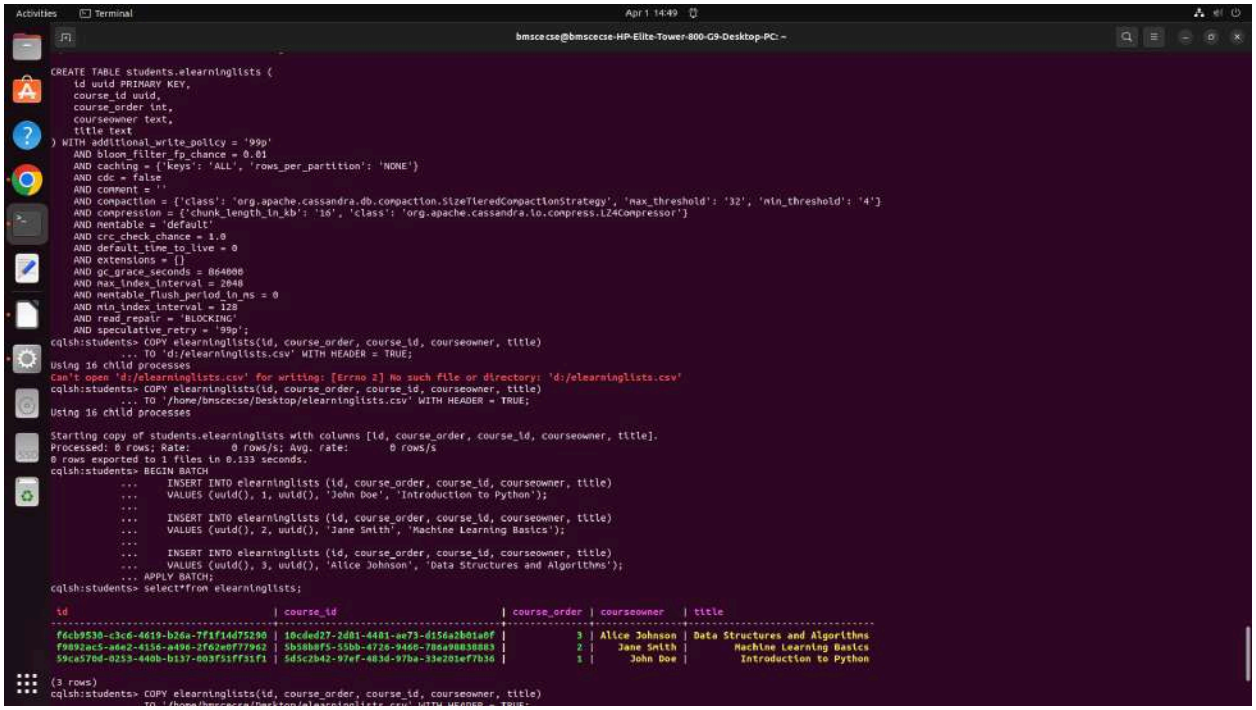

Terminal
Apr 1 14:28
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~
[3]+ Stopped cqlsh
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ cqlsh
Connected to Test Cluster at 127.0.0.1:9042
[cqlsh 6.1.0 | Cassandra 4.1.0 | CQL spec 3.4.6 | Native protocol v5]
Use HELP for help.
cqlsh> USE students;
cqlsh:students> SELECT * FROM students_info;
usn | dateofjoining | last_exam_percent | studname
5 | 2012-03-11 18:30:00.000000+0000 | 67.9 | 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 | Samrth
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)
cqlsh:students> UPDATE students_info SET usn=6 WHERE usn=3;
InvalidRequest: Error from server: code=2200 [Invalid query] message="PRIMARY KEY part usn found in SET part"
cqlsh:students> UPDATE students_info SET usn = 10 WHERE roll_no = 3;
InvalidRequest: Error from server: code=2200 [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=2200 [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 | 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 | Samrth
6 | 2012-03-11 13:00:00.000000+0000 | 78.9 | Tarun
(5 rows)
cqlsh:students> DELETE last_exam_percent from students_info where usn=2;
cqlsh:students> select*fron 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 | null | David Sheen
4 | 2012-03-11 18:30:00.000000+0000 | 90.9 | Samrth
6 | 2012-03-11 13:00:00.000000+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

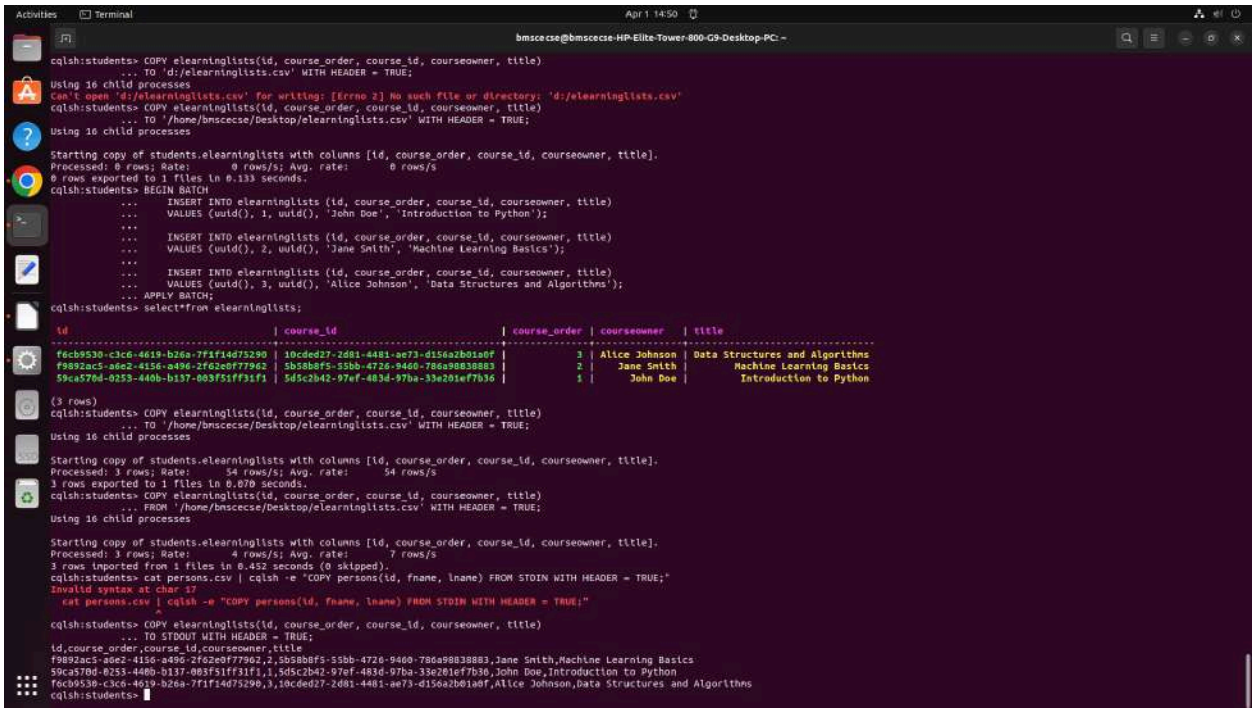

Terminal
Apr 1 14:29
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~
(5 rows)
cqlsh:students> DELETE last_exam_percent from students_info where usn=2;
cqlsh:students> select*from students_info;
usn | dateofjoining | last_exam_percent | studname
5 | 2012-03-11 18:30:00.000000+0000 | 67.9 | Sinha
1 | 2012-03-11 18:30:00.000000+0000 | 79.9 | Asha
2 | 2012-03-11 18:30:00.000000+0000 | null | David Sheen
4 | 2012-03-11 18:30:00.000000+0000 | 90.9 | Samrth
6 | 2012-03-11 13:00:00.000000+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 | Sinha
1 | 2012-03-11 18:30:00.000000+0000 | 79.9 | Asha
4 | 2012-03-11 18:30:00.000000+0000 | 90.9 | Samrth
6 | 2012-03-11 13:00:00.000000+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>
cqlsh:students> select*fron students_info;
usn | dateofjoining | hobbies | last_exam_percent | studname
5 | 2012-03-11 18:30:00.000000+0000 | null | 67.9 | Sinha
1 | 2012-03-11 18:30:00.000000+0000 | null | 79.9 | Asha
4 | 2012-03-11 18:30:00.000000+0000 | null | 90.9 | Samrth
6 | 2012-03-11 13:00:00.000000+0000 | null | 78.9 | Tarun
(4 rows)
cqlsh:students> ALTER TABLE students_info ADD language list<text>;
cqlsh:students> select*fron students_info;
usn | dateofjoining | hobbies | language | last_exam_percent | studname
5 | 2012-03-11 18:30:00.000000+0000 | null | null | 67.9 | Sinha
1 | 2012-03-11 18:30:00.000000+0000 | null | null | 79.9 | Asha
4 | 2012-03-11 18:30:00.000000+0000 | null | null | 90.9 | Samrth
6 | 2012-03-11 13:00:00.000000+0000 | null | null | 78.9 | Tarun


Terminal
Apr 1 14:29
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~
cqlsh:students>
cqlsh:students> select*from students_info;
usn | dateofjoining | hobbies | last_exan_percent | studname
5 | 2012-03-11 18:30:00.000000+0000 | null | 67.9 | Slnha
1 | 2012-03-11 18:30:00.000000+0000 | null | 79.9 | Asha
4 | 2012-03-11 18:30:00.000000+0000 | null | 90.9 | Samrth
6 | 2012-03-11 13:00:00.000000+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_exan_percent | studname
5 | 2012-03-11 18:30:00.000000+0000 | null | null | 67.9 | Slnha
1 | 2012-03-11 18:30:00.000000+0000 | null | null | 79.9 | Asha
4 | 2012-03-11 18:30:00.000000+0000 | null | null | 90.9 | Samrth
6 | 2012-03-11 13:00:00.000000+0000 | null | null | 78.9 | Tarun
(4 rows)
cqlsh:students> UPDATE students_info SET hobbies=hobbies+{'Chess,Table Tennis'} where usn=1;
cqlsh:students> SELECT*FROM STUDENTS_INFO;
usn | dateofjoining | hobbies | language | last_exan_percent | studname
5 | 2012-03-11 18:30:00.000000+0000 | null | null | 67.9 | Slnha
1 | 2012-03-11 18:30:00.000000+0000 | {'Chess,Table Tennis'} | null | 79.9 | Asha
4 | 2012-03-11 18:30:00.000000+0000 | null | null | 90.9 | Samrth
6 | 2012-03-11 13:00:00.000000+0000 | null | null | 78.9 | Tarun
(4 rows)
cqlsh:students> SELECT*FROM student_info where usn=1;
InvalidRequest: Error from server: code=2200 [Invalid query] message="table student_info does not exist"
cqlsh:students> select*from students_info where usn=1;
usn | dateofjoining | hobbies | language | last_exan_percent | studname
1 | 2012-03-11 18:30:00.000000+0000 | {'Chess,Table Tennis'} | null | 79.9 | Asha
(1 rows)
cqlsh:students> update students_info set language=language+['Hindi,English'] where usn=1;
cqlsh:students> select*from students_info;
usn | dateofjoining | hobbies | language | last_exan_percent | studname
5 | 2012-03-11 18:30:00.000000+0000 | null | null | 67.9 | Slnha
1 | 2012-03-11 18:30:00.000000+0000 | {'Chess,Table Tennis'} | ['Hindi,English'] | 79.9 | Asha
4 | 2012-03-11 18:30:00.000000+0000 | null | null | 90.9 | Samrth
6 | 2012-03-11 13:00:00.000000+0000 | null | null | 78.9 | Tarun
(4 rows)
cqlsh:students>


Activities
Terminal
Apr 1 14:40
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~
(4 rows)
cqlsh:students> CREATE TABLE library_book (
... counter_value counter,
... book_name varchar,
... stud_name varchar,
... PRIMARY KEY (book_name, stud_name)
...);
cqlsh:students> SELECT*FROM LIBRARY_BOOK;
book_name | stud_name | counter_value
(0 rows)
cqlsh: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', 'Priya');
... APPLY BATCH;
InvalidRequest: Error from server: code=2200 [Invalid query] message="INSERT statements are not allowed on counter tables, use UPDATE instead"
cqlsh: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', 'Priya');
... APPLY BATCH;
InvalidRequest: Error from server: code=2200 [Invalid query] message="INSERT statements are not allowed on counter tables, use UPDATE instead"
cqlsh:students> -- Load data into the counter column
cqlsh:students> UPDATE library_book
... SET counter_value = counter_value + 1
... WHERE book_name = 'Big data Analytics'
... AND stud_name = 'jeet';
cqlsh:students> SELECT*FROM library_book;
book_name | stud_name | counter_value
Big data Analytics | jeet | 1
(1 rows)

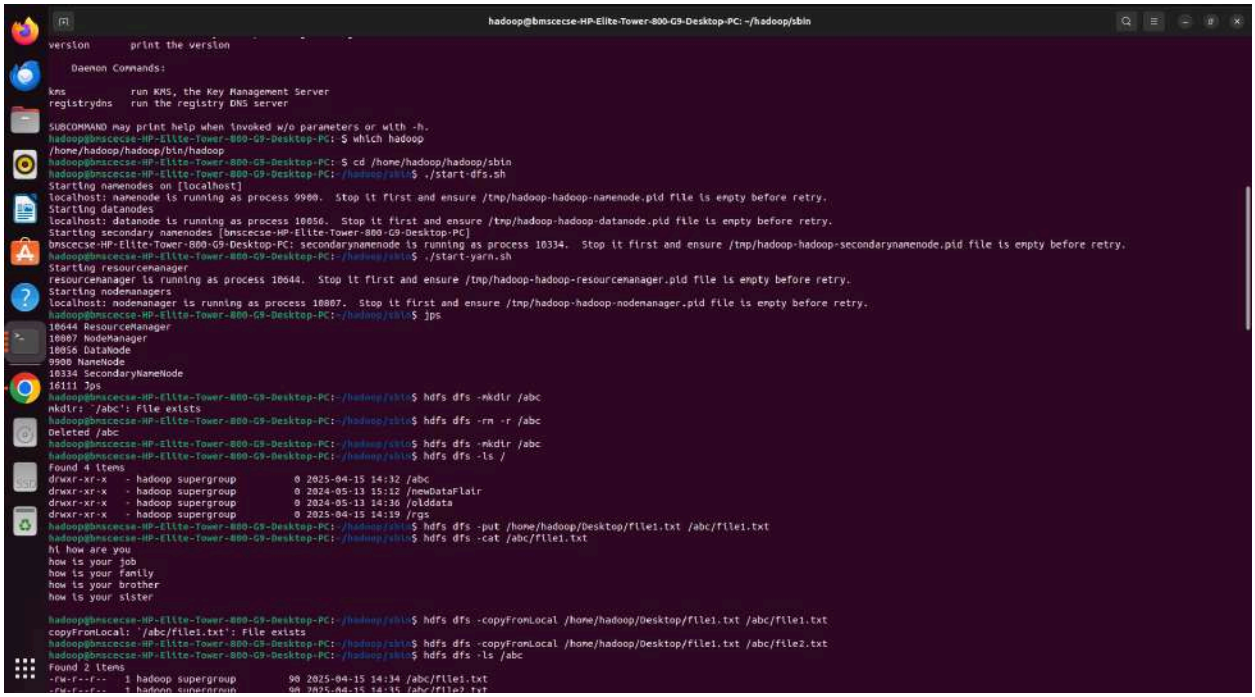

Activities
Terminal
Apr 1 14:41
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~
... INSERT INTO library_book (counter_value, book_name, stud_name)
... VALUES (0, 'Data Science Essentials', 'Ravi');
...
... INSERT INTO library_book (counter_value, book_name, stud_name)
... VALUES (0, 'Cloud Computing',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

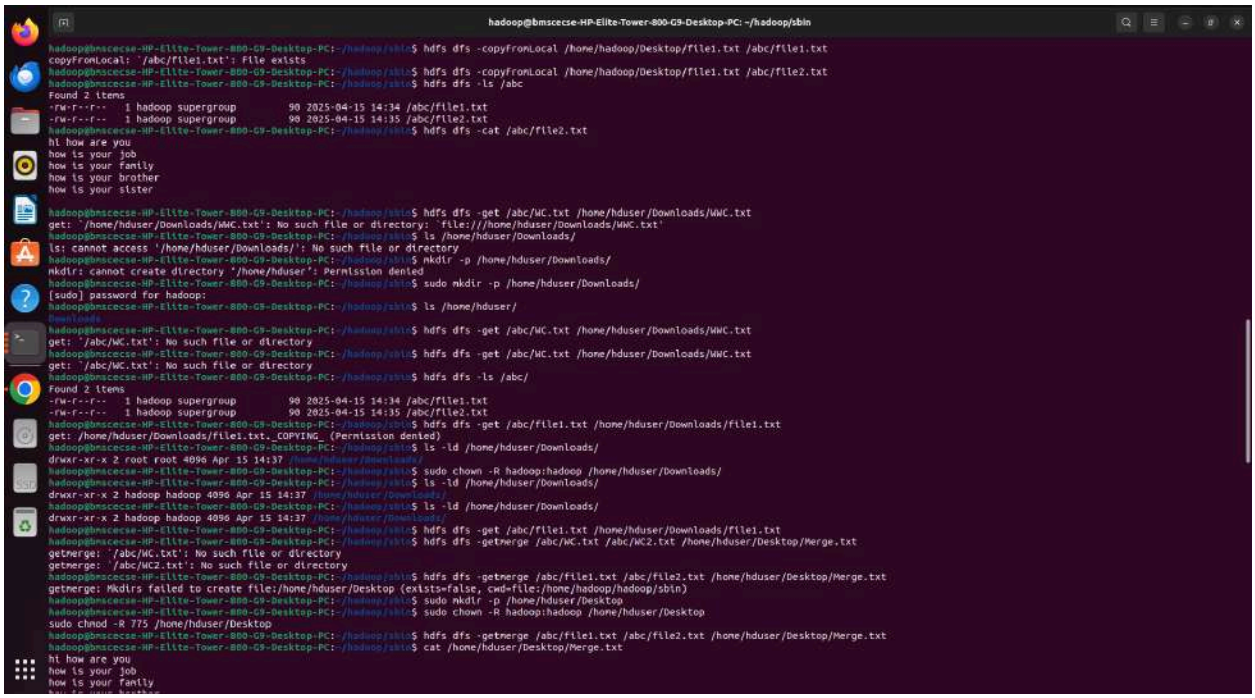

Activities
Terminal
Apr 1 14:49
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~
CREATE TABLE students.elearninglists (
id uuid PRIMARY KEY,
course_id uuid,
course_order int,
courseowner text,
title text
) WITH additional_write_policy = '99p'
AND bloom_filter_fp_chance = 0.01
AND caching = {'keys': '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 compression = {'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 = 2048
AND memtable_flush_period_in_ms = 0
AND min_index_interval = 128
AND read_repair = 'BLOCKING'
AND speculative_retry = '99p';
cqlsh: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'
cqlsh:students> COPY elearninglists(id, course_order, course_id, courseowner, title)
... TO '/home/bmscecse/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.
cqlsh: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;
cqlsh:students> select*from elearninglists;
id | course_id | course_order | courseowner | title
f6cb9530-c3c6-4619-b26a-7f1f14d75290 | 10cded27-2d81-4481-ae73-d156a2b01a0f | 3 | Alice Johnson | Data Structures and Algorithms
f9892ac5-a0e2-4156-a496-2f62e0f77962 | 5b58b0f5-55bb-4726-9460-786a98838883 | 2 | Jane Smith | Machine Learning Basics
59ca570d-0253-440b-b137-083f51ff31f1 | 5d5c2b42-97ef-483d-97ba-33e201ef7b36 | 1 | John Doe | Introduction to Python
(3 rows)
cqlsh:students> COPY elearninglists(id, course_order, course_id, courseowner, title)

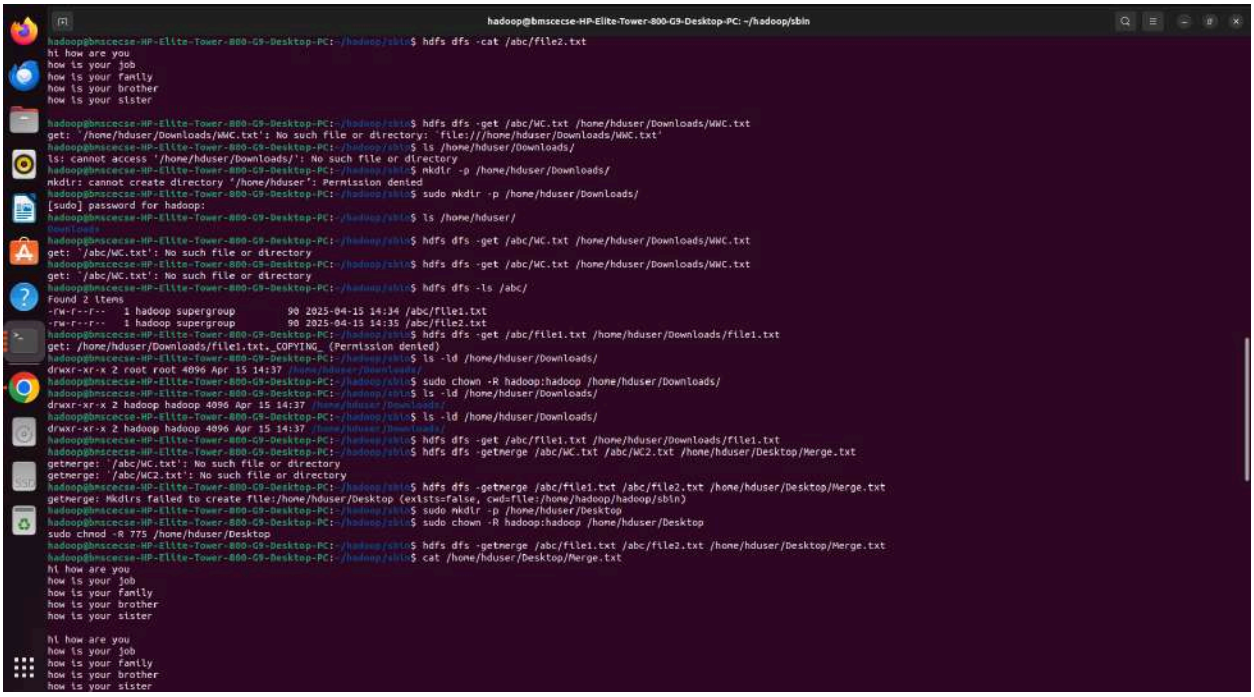

Activities
Terminal
Apr 1 14:50
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~
cqlsh:students> COPY elearninglists(id, course_order, course_id, courseowner, title)
... TO 'd:/elearninglists.csv' WITH HEADER = TRUE;
Using 16 child processes
Can't open 'd:/elearninglists.csv' for writing: [Errno 2] No such file or directory: 'd:/elearninglists.csv'
cqlsh:students> COPY elearninglists(id, course_order, course_id, courseowner, title)
... TO '/home/bmscecse/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.
cqlsh: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;
cqlsh:students> select*from elearninglists;
id | course_id | course_order | courseowner | title
f6cb9530-c3c6-4619-b26a-7f1f14d75290 | 10cded27-2d81-4481-ae73-d156a2b01a0f | 3 | Alice Johnson | Data Structures and Algorithms
f9892ac5-a0e2-4156-a496-2f62e0f77962 | 5b58b0f5-55bb-4726-9460-786a98838883 | 2 | Jane Smith | Machine Learning Basics
59ca570d-0253-440b-b137-083f51ff31f1 | 5d5c2b42-97ef-483d-97ba-33e201ef7b36 | 1 | John Doe | Introduction to Python
(3 rows)
cqlsh:students> COPY elearninglists(id, course_order, course_id, courseowner, title)
... TO '/home/bmscecse/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 exported to 1 files in 0.070 seconds.
cqlsh:students> COPY elearninglists(id, course_order, course_id, courseowner, title)
... FROM '/home/bmscecse/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: 4 rows/s; Avg. rate: 7 rows/s
3 rows imported from 1 files in 0.452 seconds (0 skipped).
cqlsh:students> cat persons.csv | cqlsh -e "COPY persons(id, fname, lname) FROM STDIN WITH HEADER = TRUE;"
Invalid syntax at char 17
cat persons.csv | cqlsh -e "COPY persons(id, fname, lname) FROM STDIN WITH HEADER = TRUE;"
^
cqlsh:students> COPY elearninglists(id, course_order, course_id, courseowner, title)
... TO STDOUT WITH HEADER = TRUE;
id,course_order,course_id,courseowner,title
f9892ac5-a0e2-4156-a496-2f62e0f77962,2,5b58b0f5-55bb-4726-9460-786a98838883,Jane Smith,Machine Learning Basics
59ca570d-0253-440b-b137-083f51ff31f1,1,5d5c2b42-97ef-483d-97ba-33e201ef7b36,John Doe,Introduction to Python
f6cb9530-c3c6-4619-b26a-7f1f14d75290,3,10cded27-2d81-4481-ae73-d156a2b01a0f,Alice Johnson,Data Structures and Algorithms
cqlsh:students>

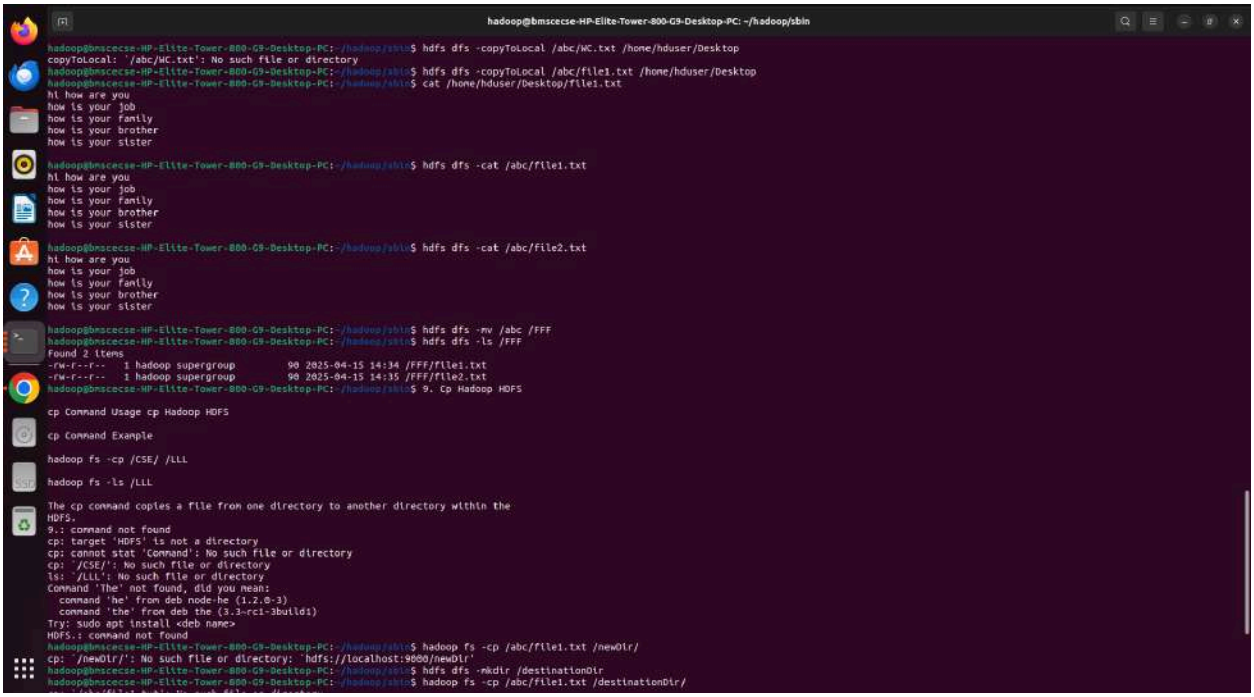
EXPERIMENT-5

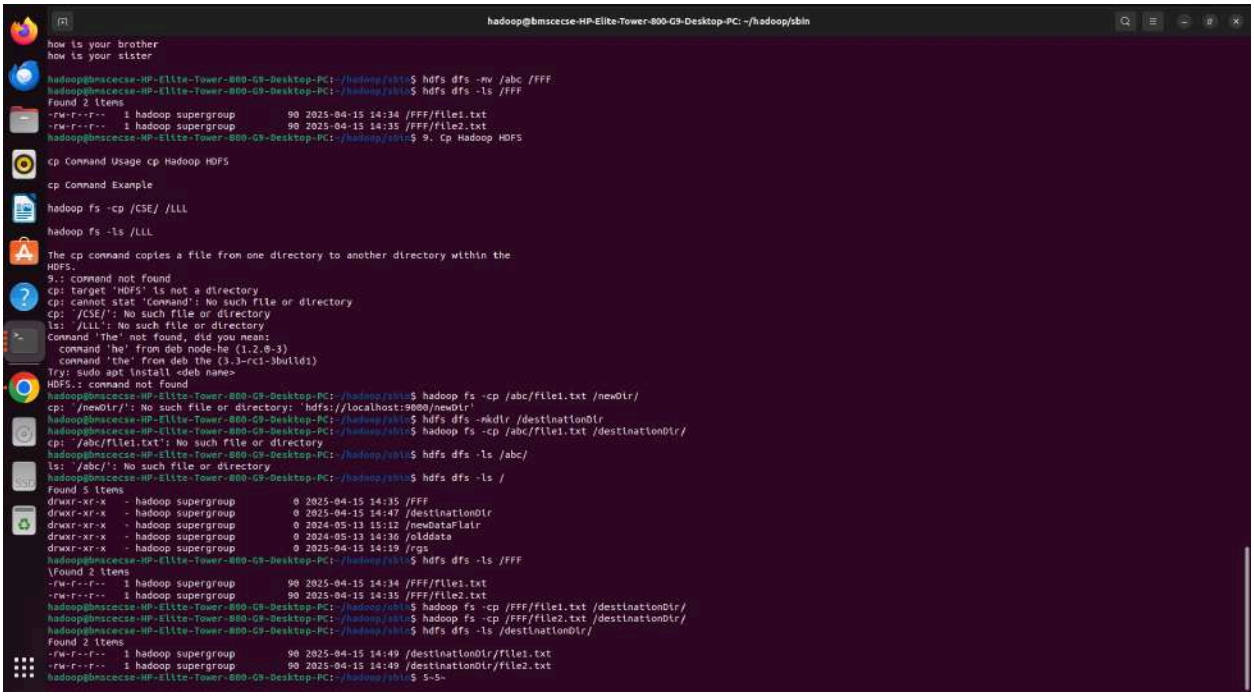
Execution of HDFS Commands for interaction with Hadoop Environment. (Minimum 10 commands to be executed)

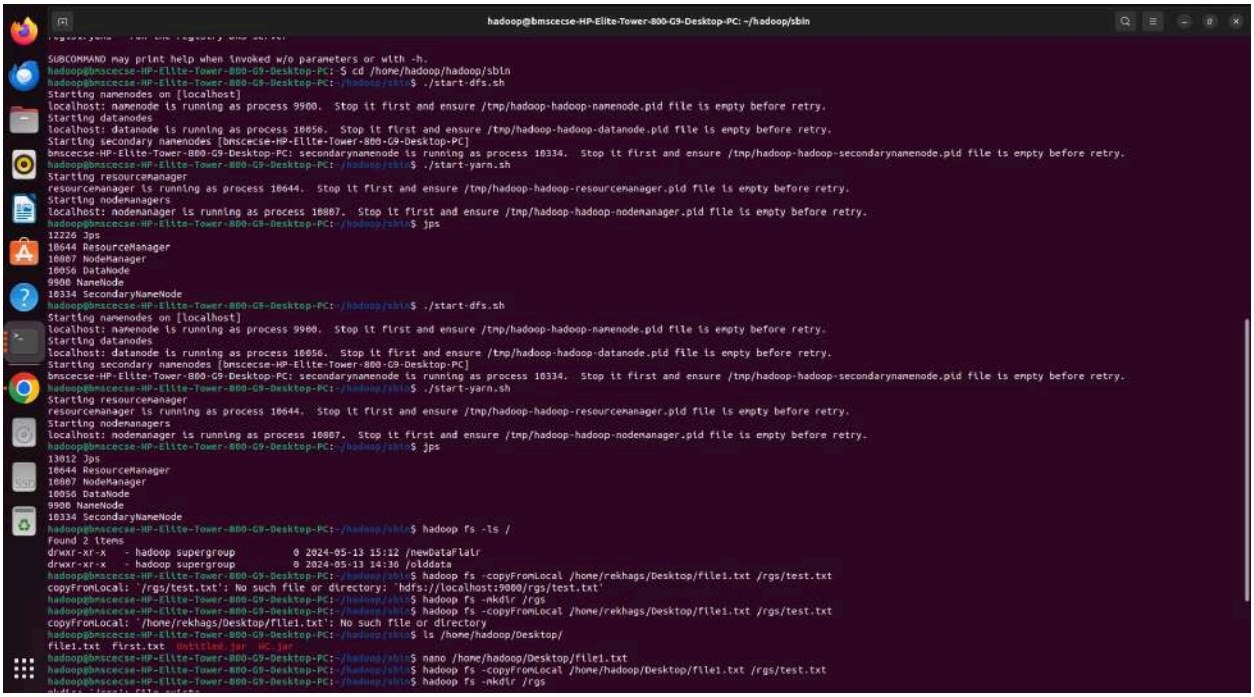


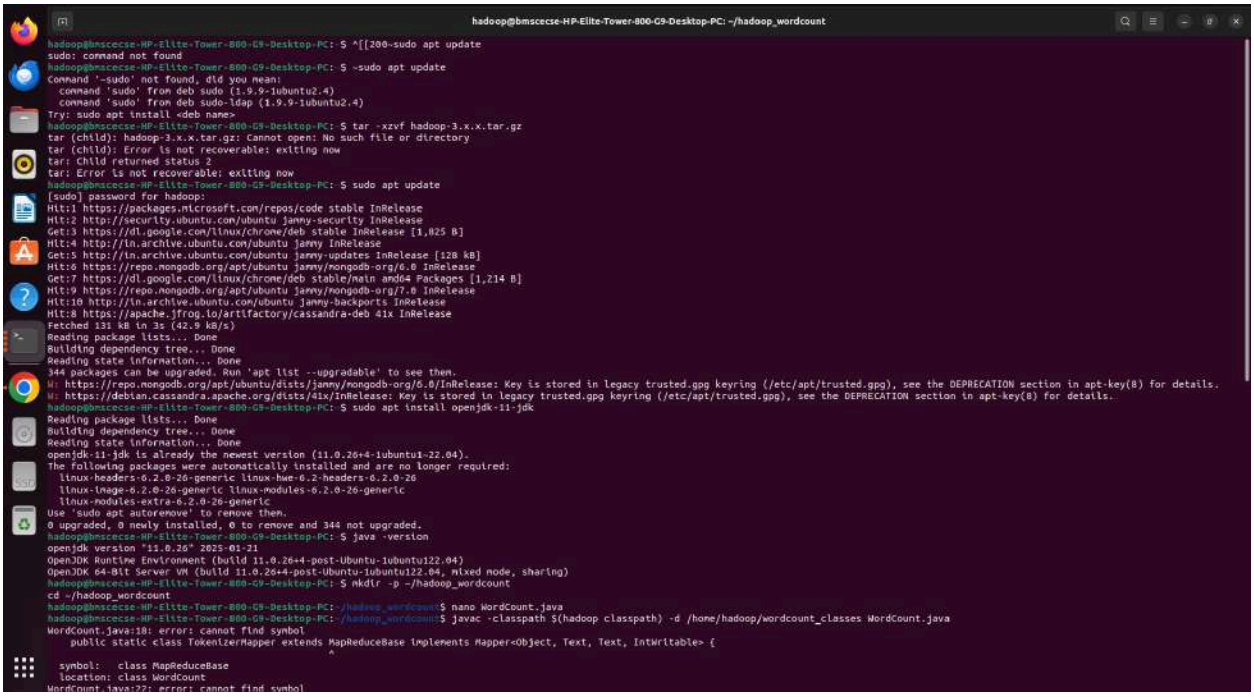



hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -cat /abc/file2.txt
hi how are you
how is your job
how is your family
how is your brother
how is your sister
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -get /abc/WC.txt /home/hduser/Downloads/WWC.txt
get: `/home/hduser/Downloads/WWC.txt': No such file or directory: `file:///home/hduser/Downloads/WWC.txt'
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ ls /home/hduser/Downloads/
ls: cannot access '/home/hduser/Downloads/': No such file or directory
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ mkdir -p /home/hduser/Downloads/
mkdir: cannot create directory '/home/hduser': Permission denied
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ sudo mkdir -p /home/hduser/Downloads/
[sudo] password for hadoop:
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ ls /home/hduser/
Downloads
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -get /abc/WC.txt /home/hduser/Downloads/WWC.txt
get: `/abc/WC.txt': No such file or directory
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -get /abc/WC.txt /home/hduser/Downloads/WWC.txt
get: `/abc/WC.txt': No such file or directory
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -ls /abc/
Found 2 items
-rw-r--r--   1 hadoop supergroup   90 2025-04-15 14:34 /abc/file1.txt
-rw-r--r--   1 hadoop supergroup   90 2025-04-15 14:35 /abc/file2.txt
hadoop@bmscecse-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._COPYING_ (Permission denied)
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ ls -ld /home/hduser/Downloads/
drwxr-xr-x 2 root root 4096 Apr 15 14:37 /home/hduser/Downloads/
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ sudo chown -R hadoop:hadoop /home/hduser/Downloads/
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ ls -ld /home/hduser/Downloads/
drwxr-xr-x 2 hadoop hadoop 4096 Apr 15 14:37 /home/hduser/Downloads/
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ ls -ld /home/hduser/Downloads/
drwxr-xr-x 2 hadoop hadoop 4096 Apr 15 14:37 /home/hduser/Downloads/
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -get /abc/file1.txt /home/hduser/Downloads/file1.txt
hadoop@bmscecse-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@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -getmerge /abc/file1.txt /abc/file2.txt /home/hduser/Desktop/Merge.txt
getmerge: Mkdirs failed to create file:/home/hduser/Desktop (exists=false, cwd=file:/home/hadoop/hadoop/sbin)
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ sudo mkdir -p /home/hduser/Desktop
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ sudo chown -R hadoop:hadoop /home/hduser/Desktop
sudo chmod -R 775 /home/hduser/Desktop
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -getmerge /abc/file1.txt /abc/file2.txt /home/hduser/Desktop/Merge.txt
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ cat /home/hduser/Desktop/Merge.txt
hi how are you
how is your job
how is your family
how is your brother
how is your sister
hi how are you
how is your job
how is your family
how is your brother
how is your sister


hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin
hadoop@bmscecse-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@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -copyToLocal /abc/file1.txt /home/hduser/Desktop
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ cat /home/hduser/Desktop/file1.txt
hi how are you
how is your job
how is your family
how is your brother
how is your sister
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -cat /abc/file1.txt
hi how are you
how is your job
how is your family
how is your brother
how is your sister
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -cat /abc/file2.txt
hi how are you
how is your job
how is your family
how is your brother
how is your sister
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -mv /abc /FFF
hadoop@bmscecse-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@bmscecse-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 'HDFS' is not a directory
cp: cannot stat 'Command': 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 'he' from deb node-he (1.2.0-3)
  command 'the' from deb the (3.3-rc1-3build1)
Try: sudo apt install <deb name>
HDFS.: command not found
hadoop@bmscecse-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@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -mkdir /destinationDir
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hadoop fs -cp /abc/file1.txt /destinationDir/


hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin
how is your brother
how is your sister
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -mv /abc /FFF
hadoop@bmscecse-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@bmscecse-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 'HDFS' is not a directory
cp: cannot stat 'Command': 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 'he' from deb node-he (1.2.0-3)
command 'the' from deb the (3.3~rc1-3build1)
Try: sudo apt install <deb name>
HDFS.: command not found
hadoop@bmscecse-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@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -mkdir /destinationDir
hadoop@bmscecse-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@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hdfs dfs -ls /abc/
ls: `/abc/': No such file or directory
hadoop@bmscecse-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 14:47 /destinationDir
drwxr-xr-x   - hadoop supergroup   0 2024-05-13 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@bmscecse-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@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hadoop fs -cp /FFF/file1.txt /destinationDir/
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hadoop fs -cp /FFF/file2.txt /destinationDir/
hadoop@bmscecse-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@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ S~S~

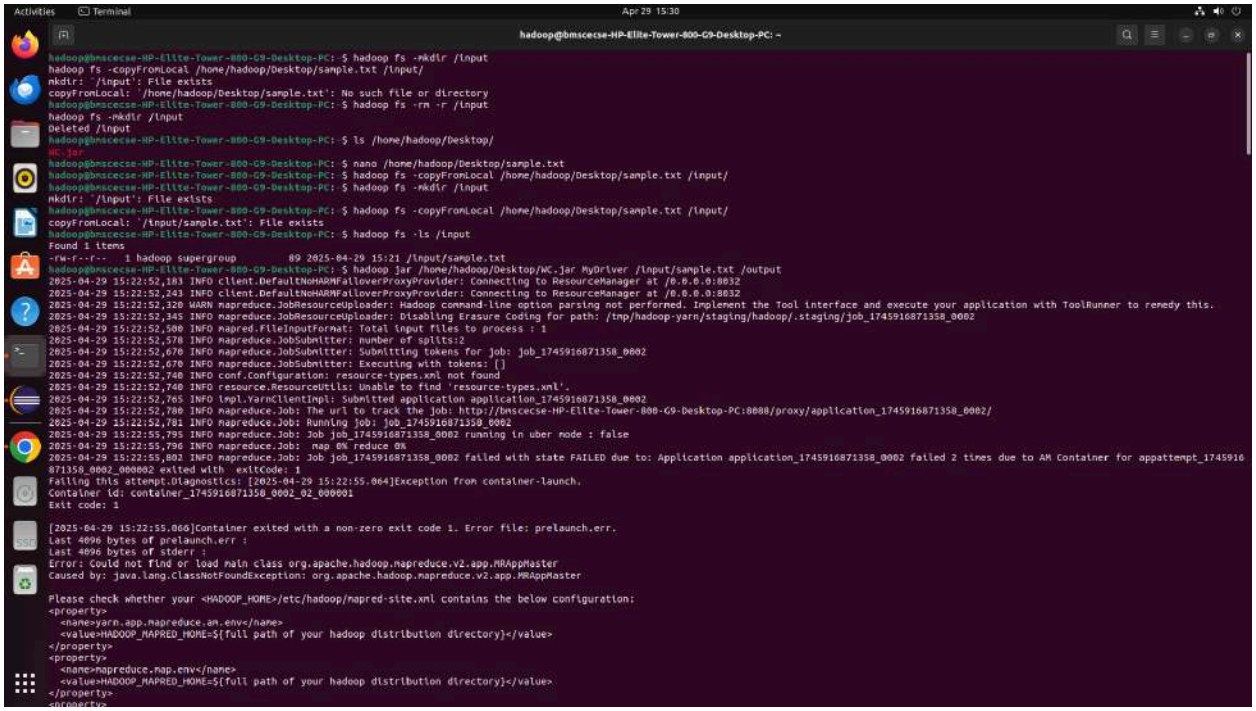

hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop/sbin
SUBCOMMAND may print help when invoked w/o parameters or with -h.
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ cd /home/hadoop/hadoop/sbin
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ ./start-dfs.sh
Starting namenodes 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 [bmscecse-HP-Elite-Tower-800-G9-Desktop-PC]
bmscecse-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@bmscecse-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@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ jps
12226 Jps
10644 ResourceManager
10807 NodeManager
10056 DataNode
9900 NameNode
10334 SecondaryNameNode
hadoop@bmscecse-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 [bmscecse-HP-Elite-Tower-800-G9-Desktop-PC]
bmscecse-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@bmscecse-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@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ jps
13012 Jps
10644 ResourceManager
10807 NodeManager
10056 DataNode
9900 NameNode
10334 SecondaryNameNode
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hadoop 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@bmscecse-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@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hadoop fs -mkdir /rgs
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hadoop fs -copyFromLocal /home/rekhags/Desktop/file1.txt /rgs/test.txt
copyFromLocal: `/home/rekhags/Desktop/file1.txt': No such file or directory
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ ls /home/hadoop/Desktop/
file1.txt first.txt Untitled.jar WC.jar
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ nano /home/hadoop/Desktop/file1.txt
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hadoop fs -copyFromLocal /home/hadoop/Desktop/file1.txt /rgs/test.txt
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop/sbin$ hadoop fs -mkdir /rgs

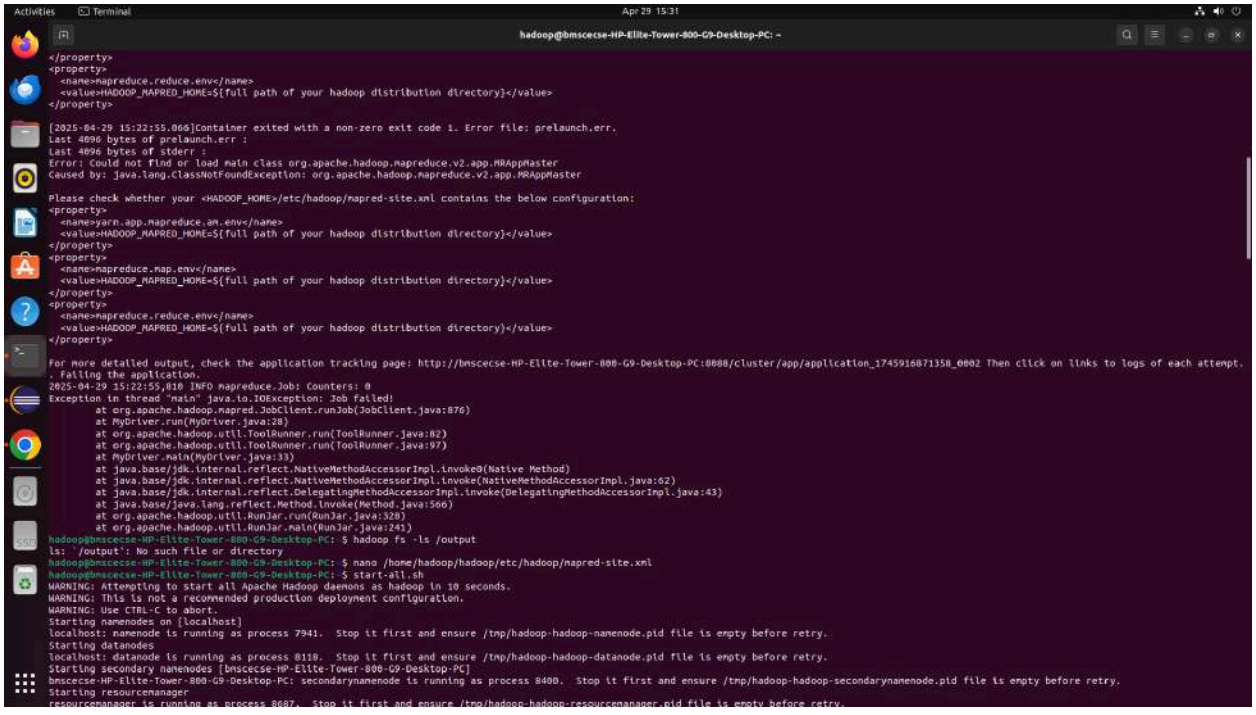

hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~/hadoop_wordcount
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ ^[[200~sudo apt update
sudo: command not found
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ ~sudo apt update
Command '~sudo' not found, did you mean:
command 'sudo' from deb sudo (1.9.9-1ubuntu2.4)
command 'sudo' from deb sudo-ldap (1.9.9-1ubuntu2.4)
Try: sudo apt install <deb name>
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ tar -xzvf hadoop-3.x.x.tar.gz
tar (child): hadoop-3.x.x.tar.gz: Cannot open: No such file or directory
tar (child): Error is not recoverable: exiting now
tar: Child returned status 2
tar: Error is not recoverable: exiting now
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ sudo apt update
[sudo] password for hadoop:
Hit:1 https://packages.microsoft.com/repos/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 https://repo.mongodb.org/apt/ubuntu jammy/mongodb-org/6.0 InRelease
Get:7 https://dl.google.com/linux/chrome/deb stable/main amd64 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:8 https://apache.jfrog.io/artifactory/cassandra-deb 41x InRelease
Fetched 132 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@bmscecse-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 (11.0.26+4-1ubuntu1~22.04).
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@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ java -version
openjdk version "11.0.26" 2025-01-21
OpenJDK Runtime Environment (build 11.0.26+4-post-Ubuntu-1ubuntu122.04)
OpenJDK 64-Bit Server VM (build 11.0.26+4-post-Ubuntu-1ubuntu122.04, mixed mode, sharing)
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ mkdir -p ~/hadoop_wordcount
cd ~/hadoop_wordcount
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop_wordcount$ nano WordCount.java
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/hadoop_wordcount$ javac -classpath $(hadoop classpath) -d /home/hadoop/wordcount_classes WordCount.java
WordCount.java:18: error: cannot find symbol
public static class TokenizerMapper extends MapReduceBase implements Mapper<Object, Text, Text, IntWritable> {
^
symbol: class MapReduceBase
location: class WordCount
WordCount.java:22: error: cannot find symbol

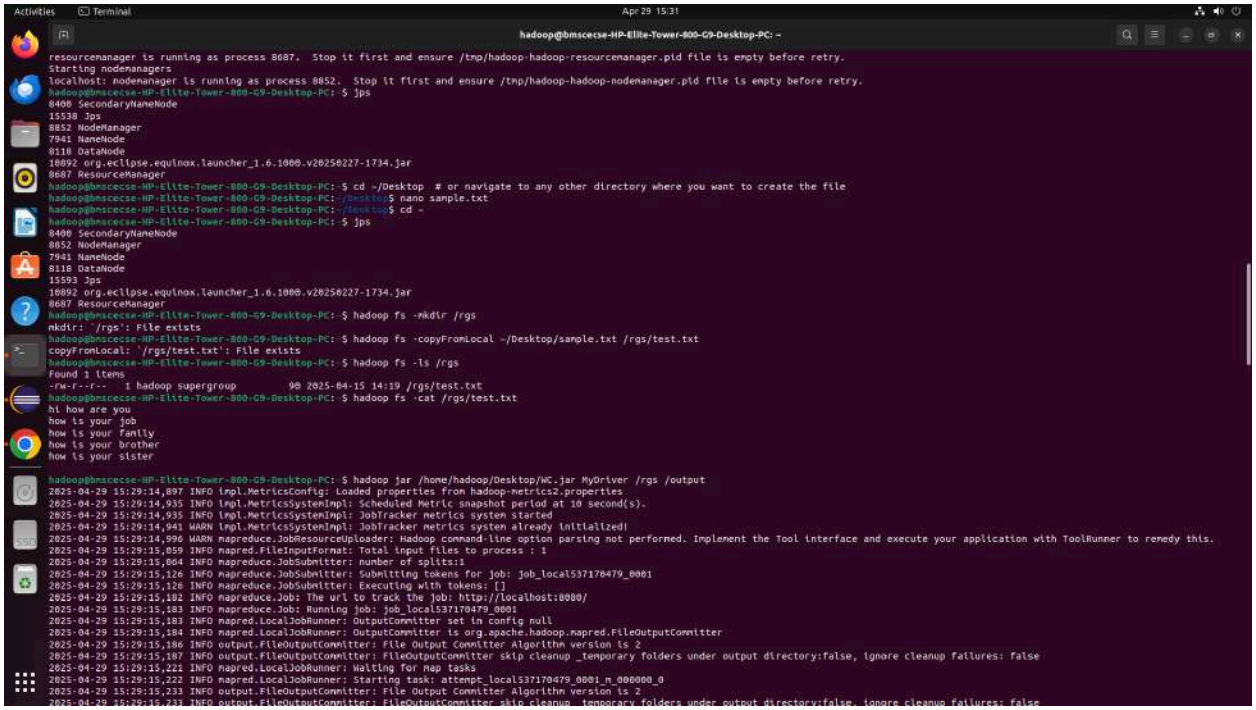
EXPERIMENT-6

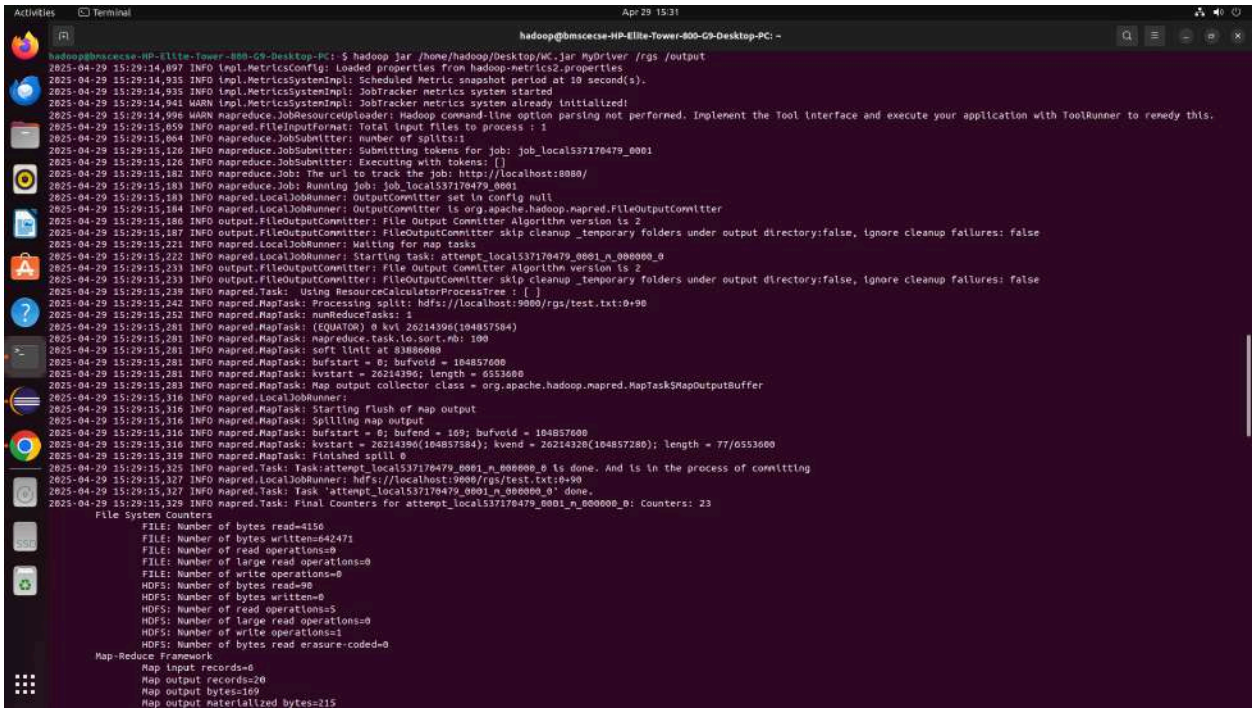
Implement Wordcount program on Hadoop framework

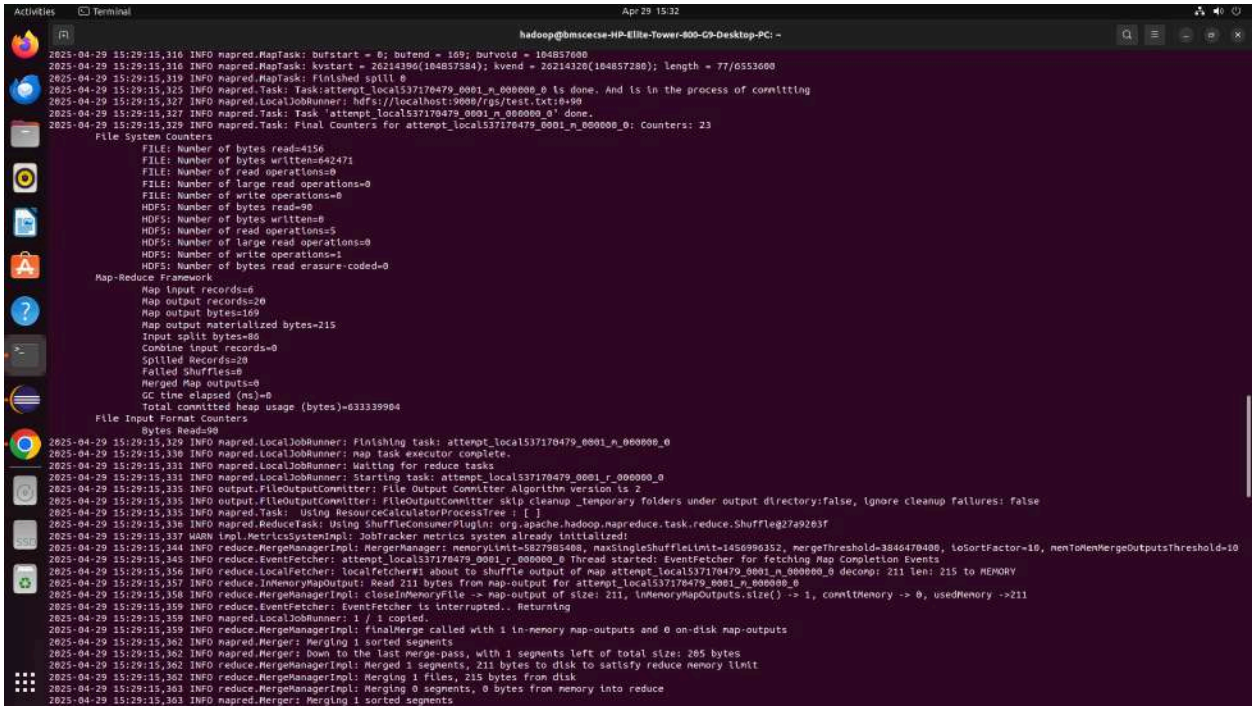
Code with Screenshots:

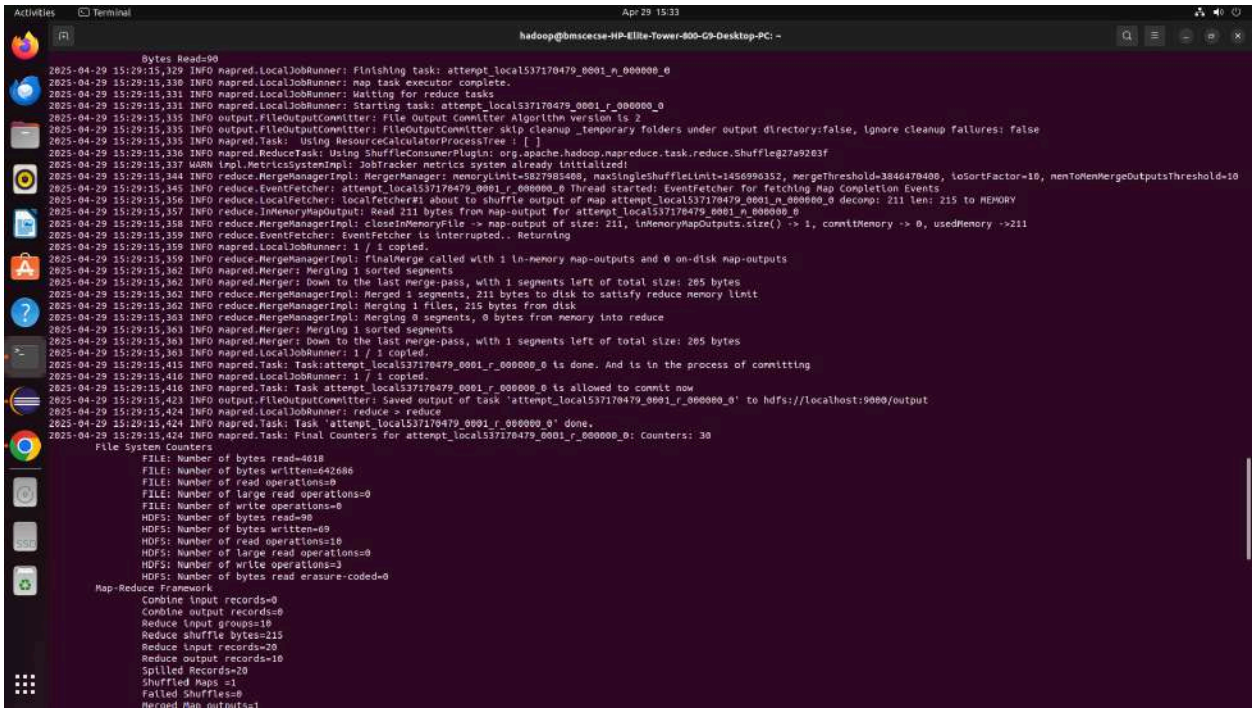


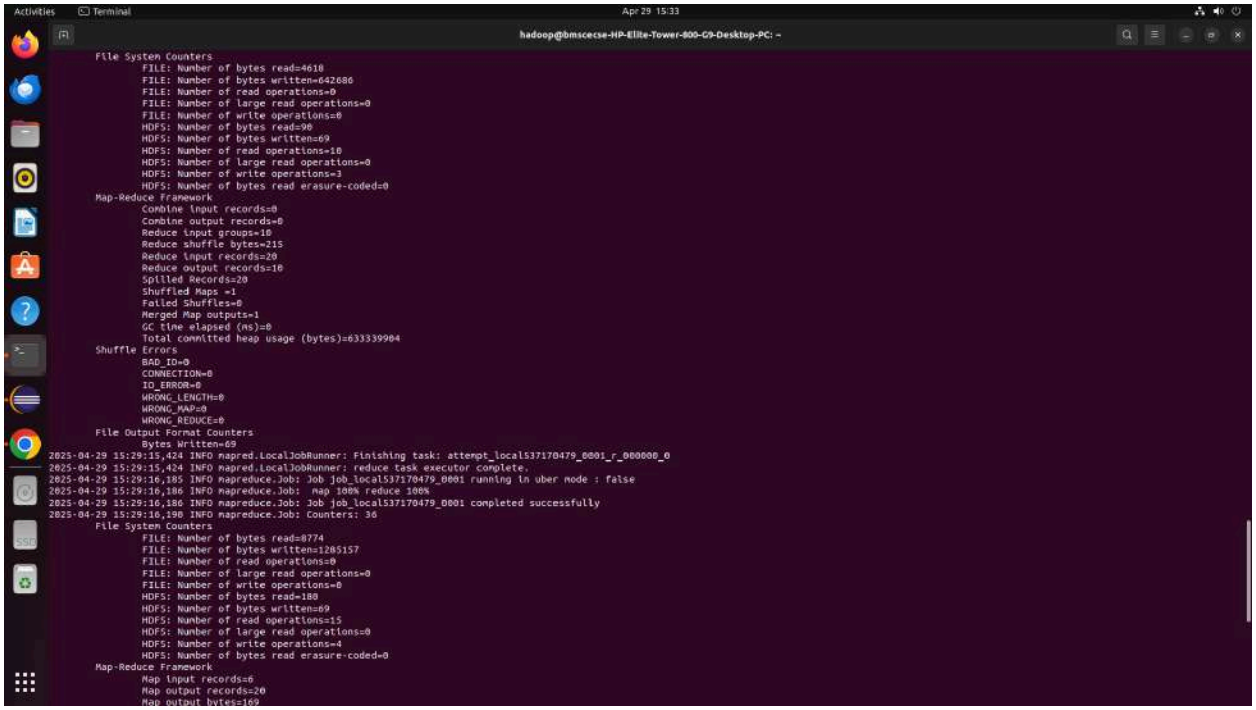


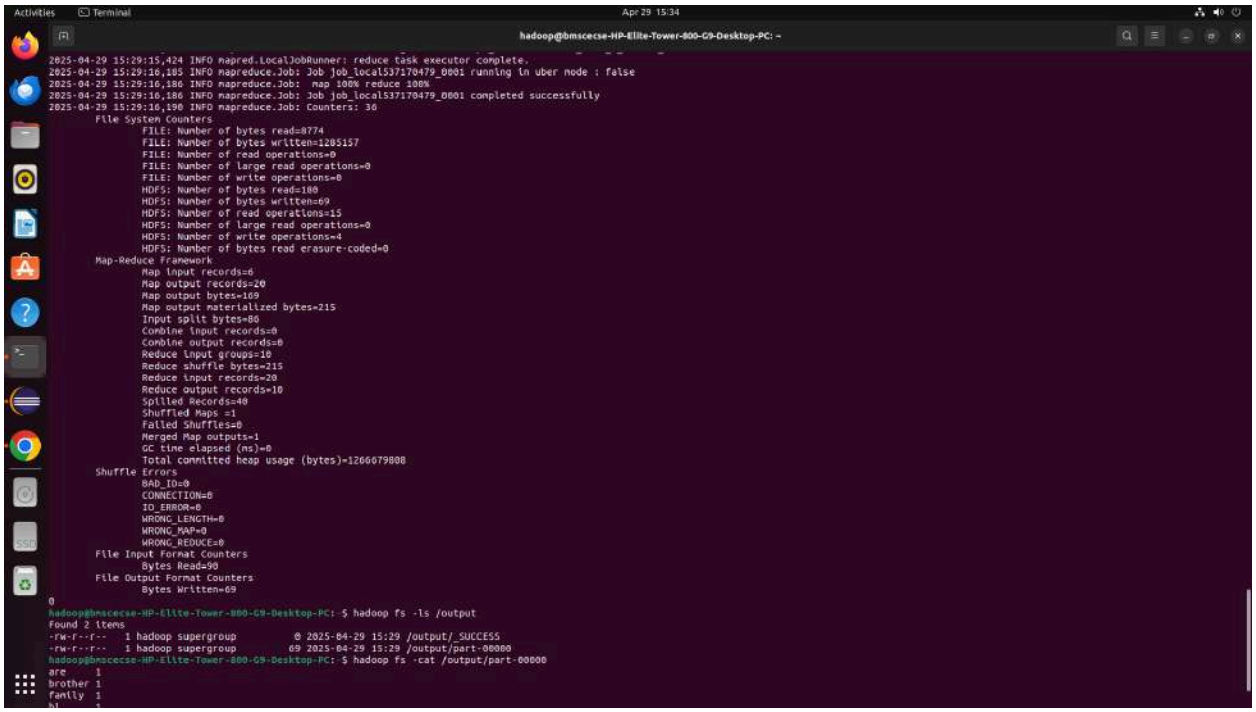

Activities
Terminal
Apr 29 15:31
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~
resourcemanager is running as process 8687. Stop it first and ensure /tmp/hadoop-hadoop-resourcemanager.pid file is empty before retry.
Starting nodemanagers
localhost: nodemanager is running as process 8852. Stop it first and ensure /tmp/hadoop-hadoop-nodemanager.pid file is empty before retry.
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ jps
8400 SecondaryNameNode
15538 Jps
8852 NodeManager
7941 NameNode
8110 DataNode
10892 org.eclipse.equinox.launcher_1.6.1000.v20250227-1734.jar
8687 ResourceManager
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ cd ~/Desktop # or navigate to any other directory where you want to create the file
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~/Desktop$ nano sample.txt
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~/Desktop$ cd ~
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ jps
8400 SecondaryNameNode
8852 NodeManager
7941 NameNode
8110 DataNode
15593 Jps
10892 org.eclipse.equinox.launcher_1.6.1000.v20250227-1734.jar
8687 ResourceManager
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ hadoop fs -mkdir /rgs
mkdir: `/rgs': File exists
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ hadoop fs -copyFromLocal ~/Desktop/sample.txt /rgs/test.txt
copyFromLocal: `/rgs/test.txt': File exists
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ hadoop fs -ls /rgs
Found 1 items
-rw-r--r-- 1 hadoop supergroup 90 2025-04-15 14:19 /rgs/test.txt
hadoop@bmscecse-HP-Elite-Tower-800-G9-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@bmscecse-HP-Elite-Tower-800-G9-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: JobTracker metrics system started
2025-04-29 15:29:14,941 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2025-04-29 15:29:14,996 WARN mapreduce.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 15:29:15,064 INFO mapreduce.JobSubmitter: number of splits:1
2025-04-29 15:29:15,126 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local537170479_0001
2025-04-29 15:29:15,126 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-04-29 15:29:15,182 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-04-29 15:29:15,183 INFO mapreduce.Job: Running job: job_local537170479_0001
2025-04-29 15:29:15,183 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-04-29 15:29:15,184 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapred.FileOutputCommitter
2025-04-29 15:29:15,186 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-04-29 15:29:15,187 INFO output.FileOutputCommitter: FileOutputCommitter 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_local537170479_0001_m_000000_0
2025-04-29 15:29:15,233 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2

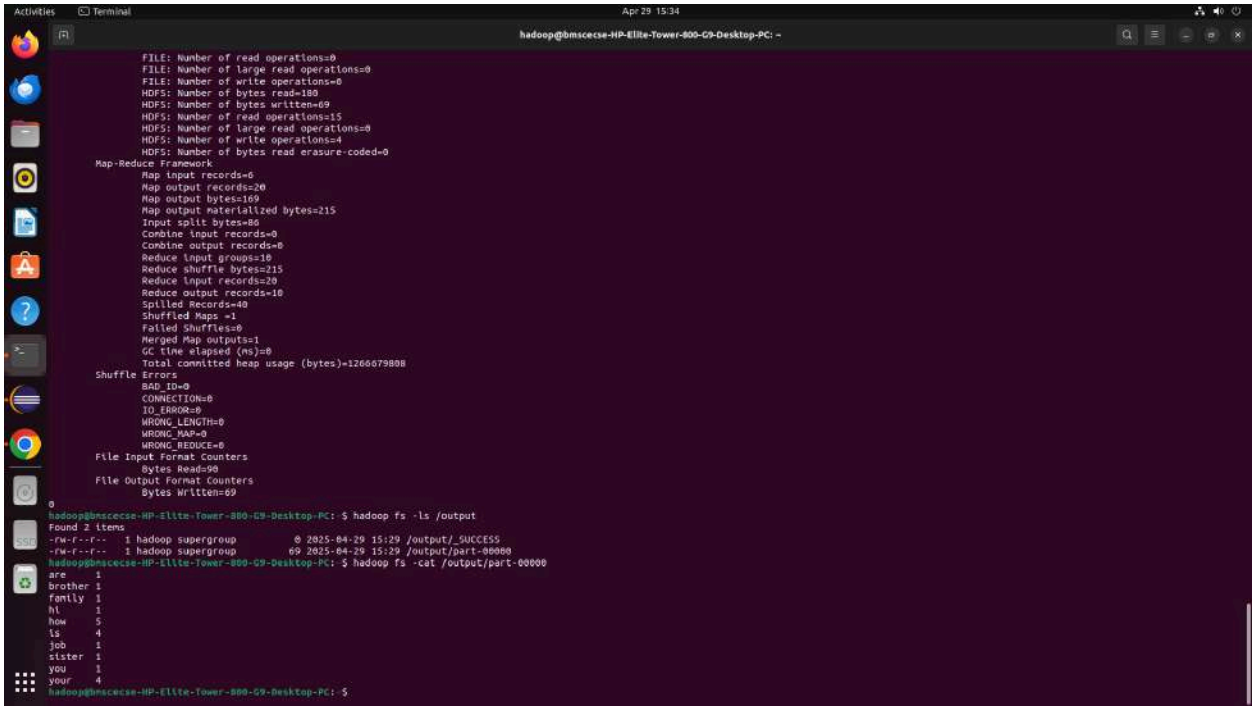

Activities
Terminal
Apr 29 15:31
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ hadoop jar /home/hadoop/Desktop/WC.jar MyDriver /rgs /output
2025-04-29 15:29:14,897 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-04-29 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: JobTracker metrics system started
2025-04-29 15:29:14,941 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2025-04-29 15:29:14,996 WARN mapreduce.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 15:29:15,064 INFO mapreduce.JobSubmitter: number of splits:1
2025-04-29 15:29:15,126 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local537170479_0001
2025-04-29 15:29:15,126 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-04-29 15:29:15,182 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-04-29 15:29:15,183 INFO mapreduce.Job: Running job: job_local537170479_0001
2025-04-29 15:29:15,183 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-04-29 15:29:15,184 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapred.FileOutputCommitter
2025-04-29 15:29:15,186 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-04-29 15:29:15,187 INFO output.FileOutputCommitter: FileOutputCommitter 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_local537170479_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: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-04-29 15:29:15,239 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
2025-04-29 15:29:15,242 INFO mapred.MapTask: Processing split: hdfs://localhost:9000/rgs/test.txt:0+90
2025-04-29 15:29:15,252 INFO mapred.MapTask: numReduceTasks: 1
2025-04-29 15:29:15,281 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
2025-04-29 15:29:15,281 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2025-04-29 15:29:15,281 INFO mapred.MapTask: soft limit at 83886080
2025-04-29 15:29:15,281 INFO mapred.MapTask: bufstart = 0; bufvoid = 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: Spilling map output
2025-04-29 15:29:15,316 INFO mapred.MapTask: bufstart = 0; bufend = 169; bufvoid = 104857600
2025-04-29 15:29:15,316 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214320(104857280); length = 77/6553600
2025-04-29 15:29:15,319 INFO mapred.MapTask: Finished spill 0
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.LocalJobRunner: hdfs://localhost:9000/rgs/test.txt:0+90
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=4156
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=90
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=6
Map output records=20
Map output bytes=169
Map output materialized bytes=215

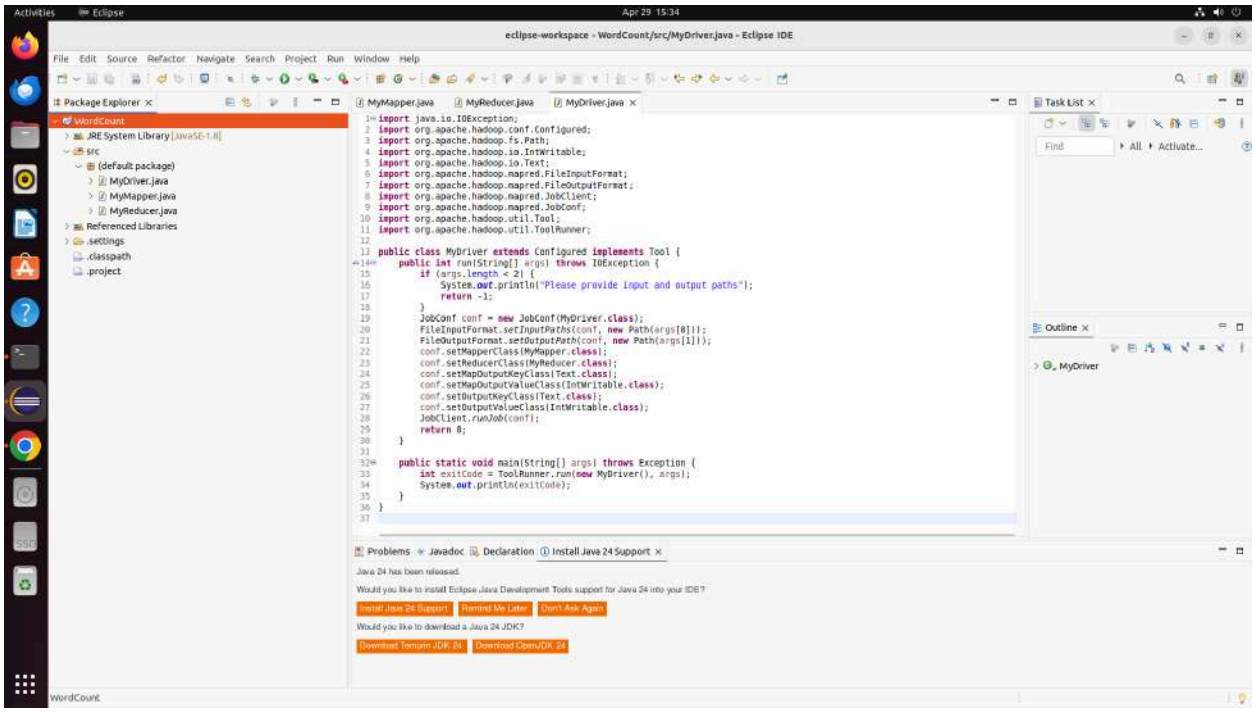

Activities
Terminal
Apr 29 15:32
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~
2025-04-29 15:29:15,316 INFO mapred.MapTask: bufstart = 0; bufend = 169; bufvoid = 104857600
2025-04-29 15:29:15,316 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214320(104857280); length = 77/6553600
2025-04-29 15:29:15,319 INFO mapred.MapTask: Finished spill 0
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.LocalJobRunner: hdfs://localhost:9000/rgs/test.txt:0+90
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=4156
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=90
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=6
Map output records=20
Map output bytes=169
Map output materialized bytes=215
Input split bytes=86
Combine input records=0
Spilled Records=20
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=0
Total committed heap usage (bytes)=633339904
File Input Format Counters
Bytes Read=90
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 output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-04-29 15:29:15,335 INFO output.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.ReduceTask: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@27a9203f
2025-04-29 15:29:15,337 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2025-04-29 15:29:15,344 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=582798540B, maxSingleShuffleLimit=145699635, mergeThreshold=384647040, ioSortFactor=10, memToMemMergeOutputsThreshold=10
2025-04-29 15:29:15,345 INFO reduce.EventFetcher: attempt_local537170479_0001_r_000000_0 Thread started: EventFetcher for fetching Map Completion Events
2025-04-29 15:29:15,356 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map attempt_local537170479_0001_m_000000_0 decomp: 211 len: 215 to MEMORY
2025-04-29 15:29:15,357 INFO reduce.InMemoryMapOutput: Read 211 bytes from map-output for attempt_local537170479_0001_m_000000_0
2025-04-29 15:29:15,358 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of size: 211, inMemoryMapOutputs.size() -> 1, commitMemory -> 0, usedMemory ->211
2025-04-29 15:29:15,359 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning
2025-04-29 15:29:15,359 INFO mapred.LocalJobRunner: 1 / 1 copied.
2025-04-29 15:29:15,359 INFO reduce.MergeManagerImpl: finalMerge called with 1 in-memory map-outputs and 0 on-disk map-outputs
2025-04-29 15:29:15,362 INFO mapred.Merger: Merging 1 sorted segments
2025-04-29 15:29:15,362 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 205 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 15:29:15,363 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce
2025-04-29 15:29:15,363 INFO mapred.Merger: Merging 1 sorted segments

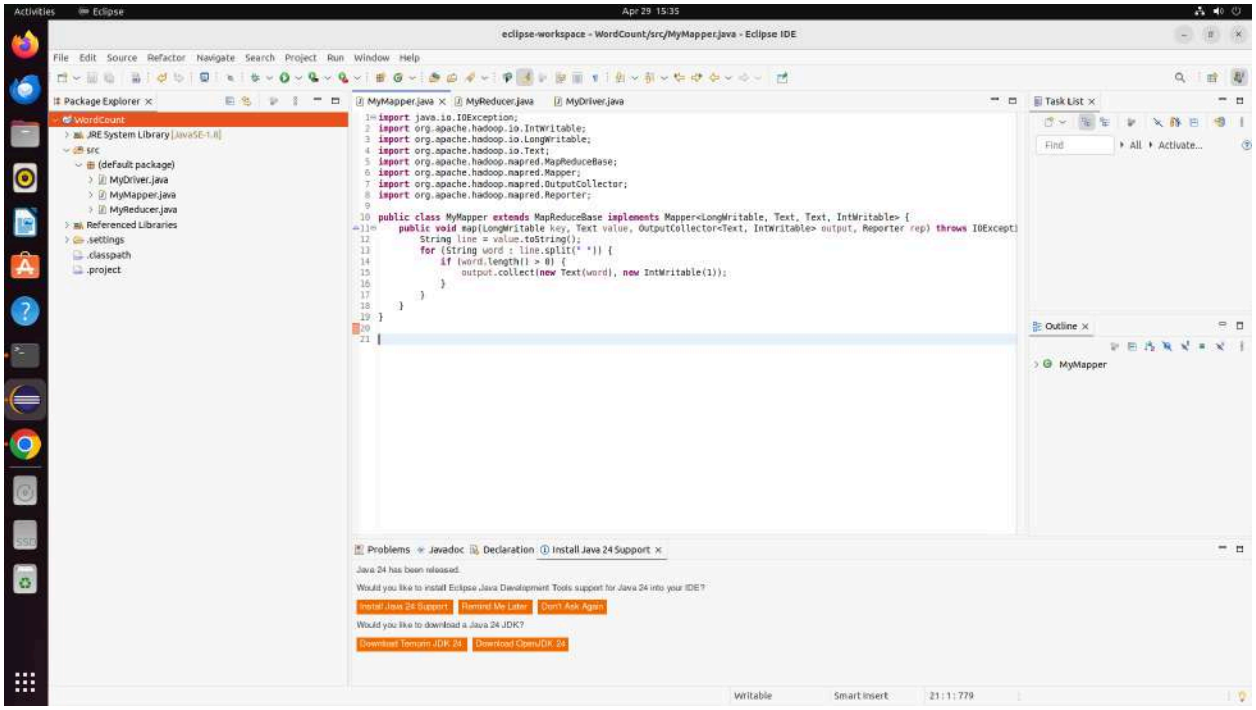

Activities
Terminal
Apr 29 15:33
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~
Bytes Read=90
2025-04-29 15:29:15,329 INFO mapred.LocalJobRunner: Finishing task: attempt_local537170479_0001_m_000000_0
2025-04-29 15:29:15,330 INFO mapred.LocalJobRunner: map task executor complete.
2025-04-29 15:29:15,331 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 output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-04-29 15:29:15,335 INFO output.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.ReduceTask: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@27a9203f
2025-04-29 15:29:15,337 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2025-04-29 15:29:15,344 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=582798540B, maxSingleShuffleLimit=145699635, mergeThreshold=384647040, ioSortFactor=10, memToMemMergeOutputsThreshold=10
2025-04-29 15:29:15,345 INFO reduce.EventFetcher: attempt_local537170479_0001_r_000000_0 Thread started: EventFetcher for fetching Map Completion Events
2025-04-29 15:29:15,356 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map attempt_local537170479_0001_m_000000_0 decomp: 211 len: 215 to MEMORY
2025-04-29 15:29:15,357 INFO reduce.InMemoryMapOutput: Read 211 bytes from map-output for attempt_local537170479_0001_m_000000_0
2025-04-29 15:29:15,358 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of size: 211, inMemoryMapOutputs.size() -> 1, commitMemory -> 0, usedMemory ->211
2025-04-29 15:29:15,359 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning
2025-04-29 15:29:15,359 INFO mapred.LocalJobRunner: 1 / 1 copied.
2025-04-29 15:29:15,359 INFO reduce.MergeManagerImpl: finalMerge called with 1 in-memory map-outputs and 0 on-disk map-outputs
2025-04-29 15:29:15,362 INFO mapred.Merger: Merging 1 sorted segments
2025-04-29 15:29:15,362 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 205 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 15:29:15,363 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce
2025-04-29 15:29:15,363 INFO mapred.Merger: Merging 1 sorted segments
2025-04-29 15:29:15,363 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 205 bytes
2025-04-29 15:29:15,363 INFO mapred.LocalJobRunner: 1 / 1 copied.
2025-04-29 15:29:19,415 INFO mapred.Task: Task:attempt_local537170479_0001_r_000000_0 is done. And is in the process of committing
2025-04-29 15:29:15,416 INFO mapred.LocalJobRunner: 1 / 1 copied.
2025-04-29 15:29:15,416 INFO mapred.Task: Task attempt_local537170479_0001_r_000000_0 is allowed to commit now
2025-04-29 15:29:15,423 INFO output.FileOutputCommitter: Saved output of task 'attempt_local537170479_0001_r_000000_0' to hdfs://localhost:9000/output
2025-04-29 15:29:15,424 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=4018
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 read=90
HDFS: Number of bytes written=69
HDFS: Number of read operations=10
HDFS: Number of large read operations=0
HDFS: Number of write operations=3
HDFS: Number of bytes read erasure-coded=0
Map-Reduce Framework
Combine input records=0
Combine output records=0
Reduce input groups=10
Reduce shuffle bytes=215
Reduce input records=20
Reduce output records=10
Spilled Records=20
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1

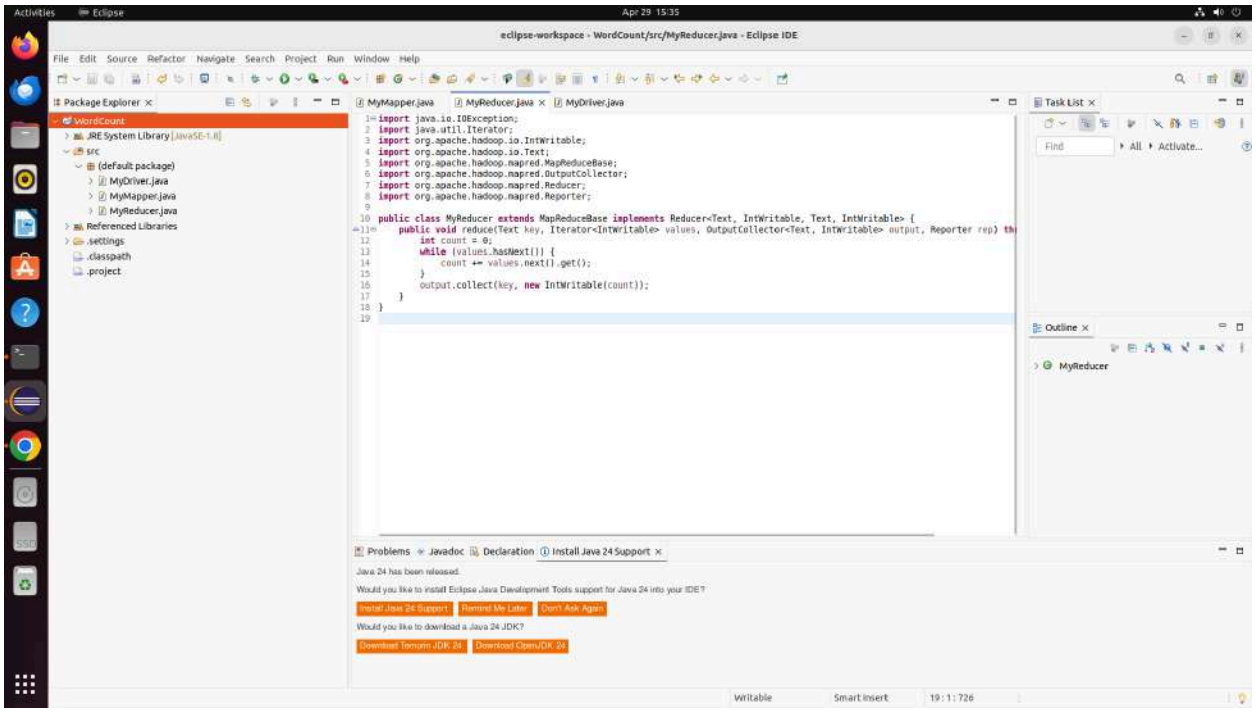

Activities
Terminal
Apr 29 15:33
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~
File System Counters
FILE: Number of bytes read=4610
FILE: Number of bytes written=642686
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=90
HDFS: Number of bytes written=69
HDFS: Number of read operations=10
HDFS: Number of large read operations=0
HDFS: Number of write operations=3
HDFS: Number of bytes read erasure-coded=0
Map-Reduce Framework
Combine input records=0
Combine output records=0
Reduce input groups=10
Reduce shuffle bytes=215
Reduce input records=20
Reduce output records=10
Spilled Records=20
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=0
Total committed heap usage (bytes)=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=69
2025-04-29 15:29:15,424 INFO mapred.LocalJobRunner: Finishing task: attempt_local537170479_0001_r_000000_0
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 100% reduce 100%
2025-04-29 15:29:16,186 INFO mapreduce.Job: Job job_local537170479_0001 completed successfully
2025-04-29 15:29:16,190 INFO mapreduce.Job: Counters: 36
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
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=169

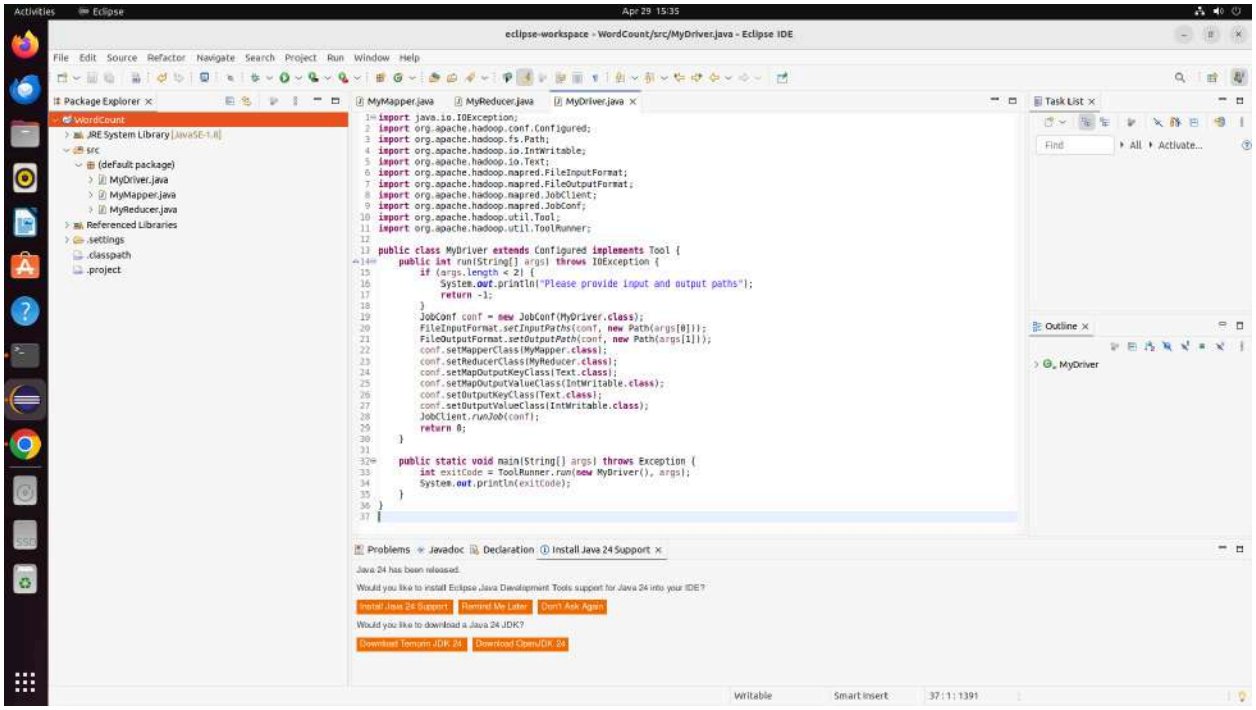

Activities
Terminal
Apr 29 15:34
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~
2025-04-29 15:29:15,424 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-04-29 15:29:16,185 INFO mapreduce.Job: Job job_local537170479_0001 running in uber mode : false
2025-04-29 15:29:16,186 INFO mapreduce.Job: map 100% reduce 100%
2025-04-29 15:29:16,186 INFO mapreduce.Job: Job job_local537170479_0001 completed successfully
2025-04-29 15:29:16,190 INFO mapreduce.Job: Counters: 36
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
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=169
Map output materialized bytes=215
Input split bytes=86
Combine input records=0
Combine output records=0
Reduce input groups=10
Reduce shuffle bytes=215
Reduce input records=20
Reduce output records=10
Spilled Records=40
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=0
Total committed heap usage (bytes)=1266079808
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
Bytes Read=90
File Output Format Counters
Bytes Written=69
0
hadoop@bmscecse-HP-Elite-Tower-800-G9-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@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: $ hadoop fs -cat /output/part-00000
are 1
brother 1
family 1

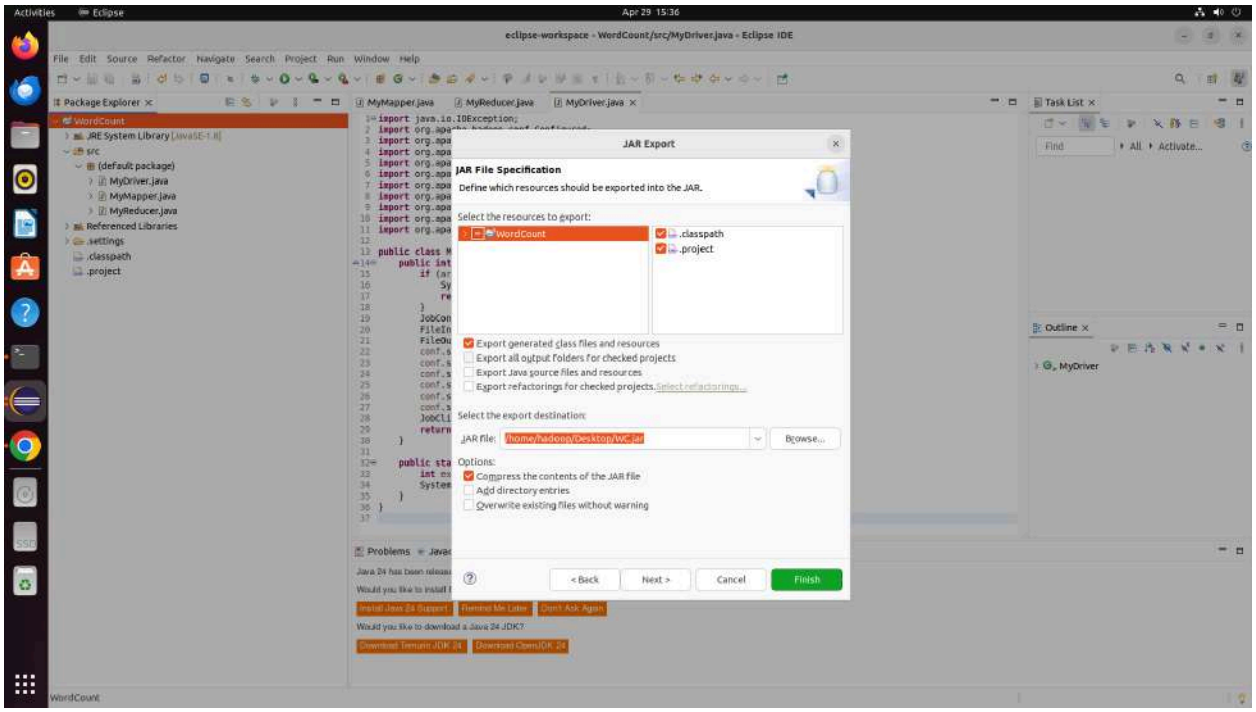

Activities
Terminal
Apr 29 15:34
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC: ~
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=180
HDFS: Number of bytes written=69
HDFS: Number of read operations=15
HDFS: Number 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=6
Map output records=20
Map output bytes=169
Map output materialized bytes=215
Input split bytes=86
Combine input records=0
Combine output records=0
Reduce input groups=10
Reduce shuffle bytes=215
Reduce input records=20
Reduce output records=10
Spilled Records=40
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=0
Total committed heap usage (bytes)=1266679808
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
Bytes Read=90
File Output Format Counters
Bytes Written=69
0
hadoop@bmscecse-HP-Elite-Tower-800-G9-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@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ hadoop fs -cat /output/part-00000
are 1
brother 1
family 1
hi 1
how 5
is 4
job 1
sister 1
you 1
your 4
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$


Activities
Eclipse
Apr 29 15:34
eclipse-workspace - WordCount/src/MyDriver.java - Eclipse IDE
File Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer
WordCount
JRE System Library [JavaSE-1.8]
src
(default package)
MyDriver.java
MyMapper.java
MyReducer.java
Referenced Libraries
.settings
.classpath
.project
MyMapper.java
MyReducer.java
MyDriver.java
import java.io.IOException;
import org.apache.hadoop.conf.Configured;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
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 Configured 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.setReducerClass(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);
}
}
Task List
Find
All
Activate...
Outline
MyDriver
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
WordCount


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
WordCount
JRE System Library [JavaSE-1.8]
src
(default package)
MyDriver.java
MyMapper.java
MyReducer.java
Referenced Libraries
.settings
.classpath
.project
MyMapper.java
MyReducer.java
MyDriver.java
import java.io.IOException;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.MapReduceBase;
import org.apache.hadoop.mapred.Mapper;
import org.apache.hadoop.mapred.OutputCollector;
import org.apache.hadoop.mapred.Reporter;
public class MyMapper extends MapReduceBase implements Mapper<LongWritable, Text, Text, IntWritable> {
public void map(LongWritable key, Text value, OutputCollector<Text, IntWritable> output, Reporter rep) throws IOExcepti
String line = value.toString();
for (String word : line.split(" ")) {
if (word.length() > 0) {
output.collect(new Text(word), new IntWritable(1));
}
}
}
}
Task List
Find
All
Activate...
Outline
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
Smart Insert
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
WordCount
JRE System Library [JavaSE-1.8]
src
(default package)
MyDriver.java
MyMapper.java
MyReducer.java
Referenced Libraries
.settings
.classpath
.project
MyMapper.java
MyReducer.java
MyDriver.java
import java.io.IOException;
import java.util.Iterator;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.MapReduceBase;
import org.apache.hadoop.mapred.OutputCollector;
import org.apache.hadoop.mapred.Reducer;
import org.apache.hadoop.mapred.Reporter;
public class MyReducer extends MapReduceBase implements Reducer<Text, IntWritable, Text, IntWritable> {
public void reduce(Text key, Iterator<IntWritable> values, OutputCollector<Text, IntWritable> output, Reporter rep) th
int count = 0;
while (values.hasNext()) {
count += values.next().get();
}
output.collect(key, new IntWritable(count));
}
}
Task List
Find
All
Activate...
Outline
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
Smart Insert
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
WordCount
JRE System Library [JavaSE-1.8]
src
(default package)
MyDriver.java
MyMapper.java
MyReducer.java
Referenced Libraries
.settings
.classpath
.project
MyMapper.java
MyReducer.java
MyDriver.java
import java.io.IOException;
import org.apache.hadoop.conf.Configured;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
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 Configured 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.setReducerClass(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);
}
}
Task List
Find
All
Activate...
Outline
MyDriver
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
Smart Insert
37 : 1 : 1391


Activities
Eclipse
Apr 29 15:36
eclipse-workspace - WordCount/src/MyDriver.java - Eclipse IDE
File Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer
WordCount
JRE System Library [JavaSE-1.8]
src
(default package)
MyDriver.java
MyMapper.java
MyReducer.java
Referenced Libraries
.settings
.classpath
.project
JAR Export
JAR File Specification
Define which resources should be exported into the JAR.
Select the resources to export:
WordCount
.classpath
.project
Export generated class files and resources
Export all output folders for checked projects
Export Java source files and resources
Export refactorings for checked projects. Select refactorings...
Select the export destination:
JAR file:
/home/hadoop/Desktop/WC.jar
Browse...
Options:
Compress the contents of the JAR file
Add directory entries
Overwrite existing files without warning
< Back
Next >
Cancel
Finish
Outline
MyDriver
Task List
Find
All
Activate...
Problems
Would you like to download a Java 24 JDK?
Download Temurin JDK 24
Download OpenJDK 24
WordCount

EXPERIMENT-7

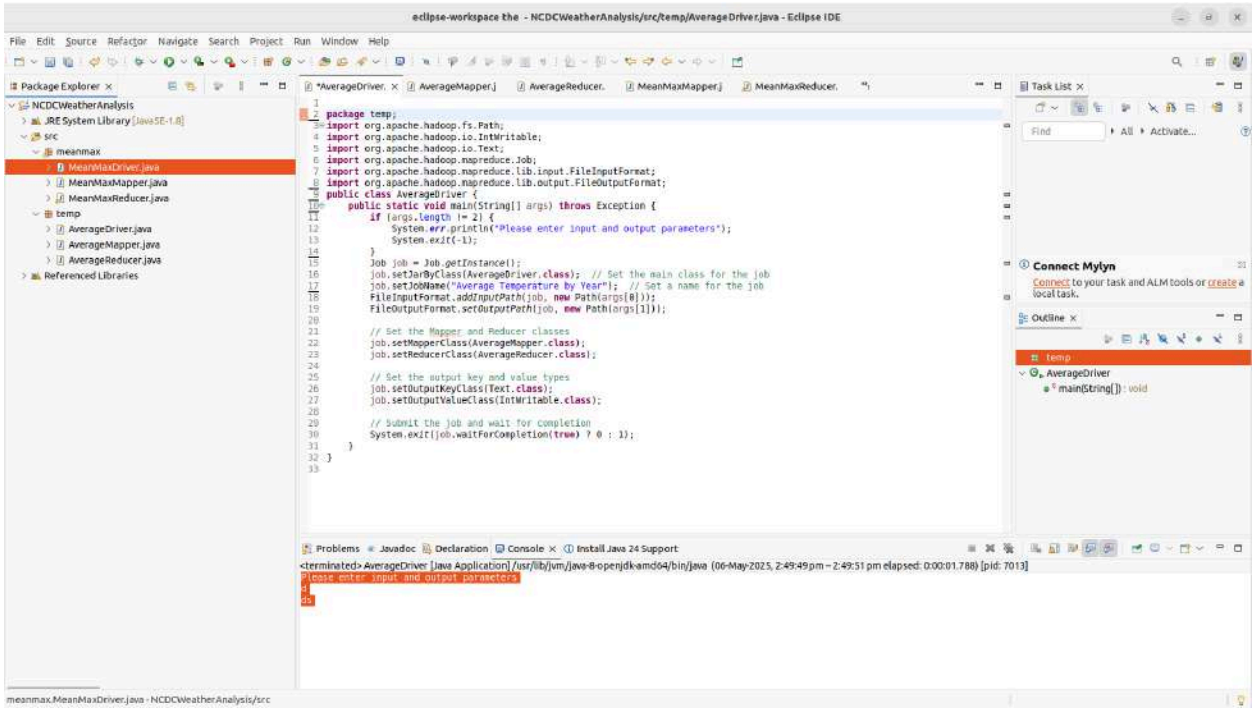
From the following link extract the weather data

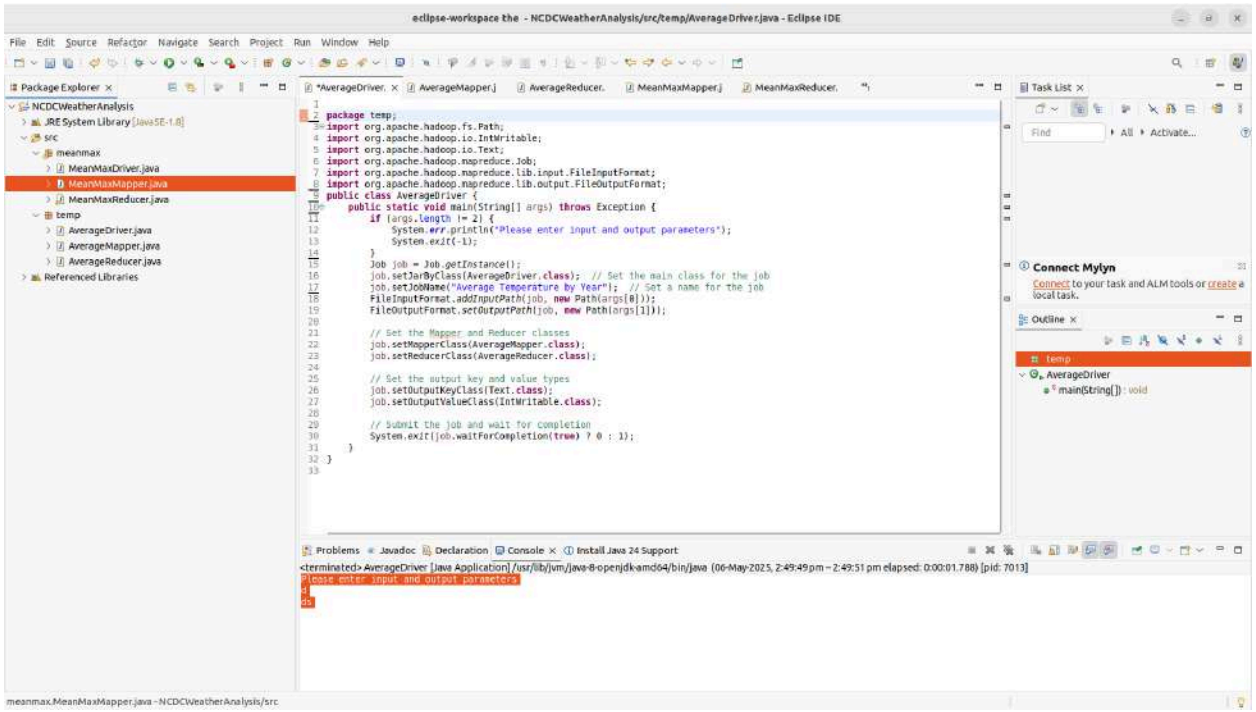
https://github.com/tomwhite/hadoop-book/tree/master/input/ncdc/all

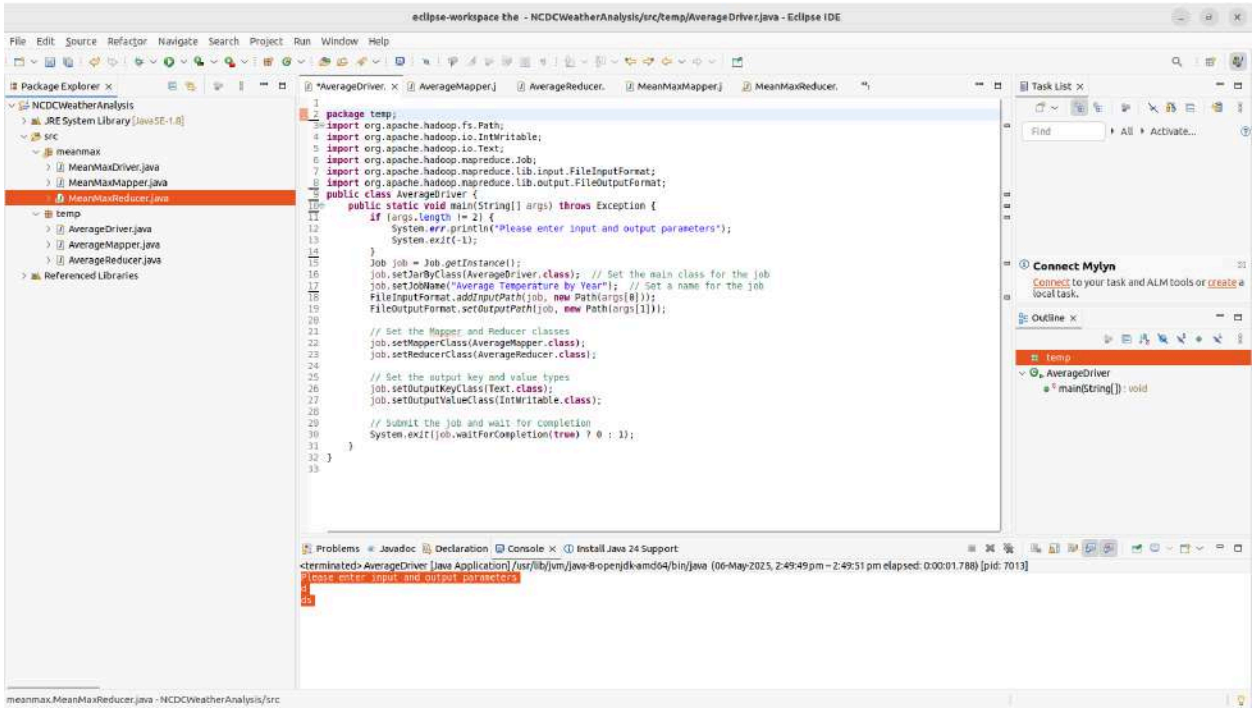
a) Create a MapReduce program to find average temperature for each year from NCDC data set.

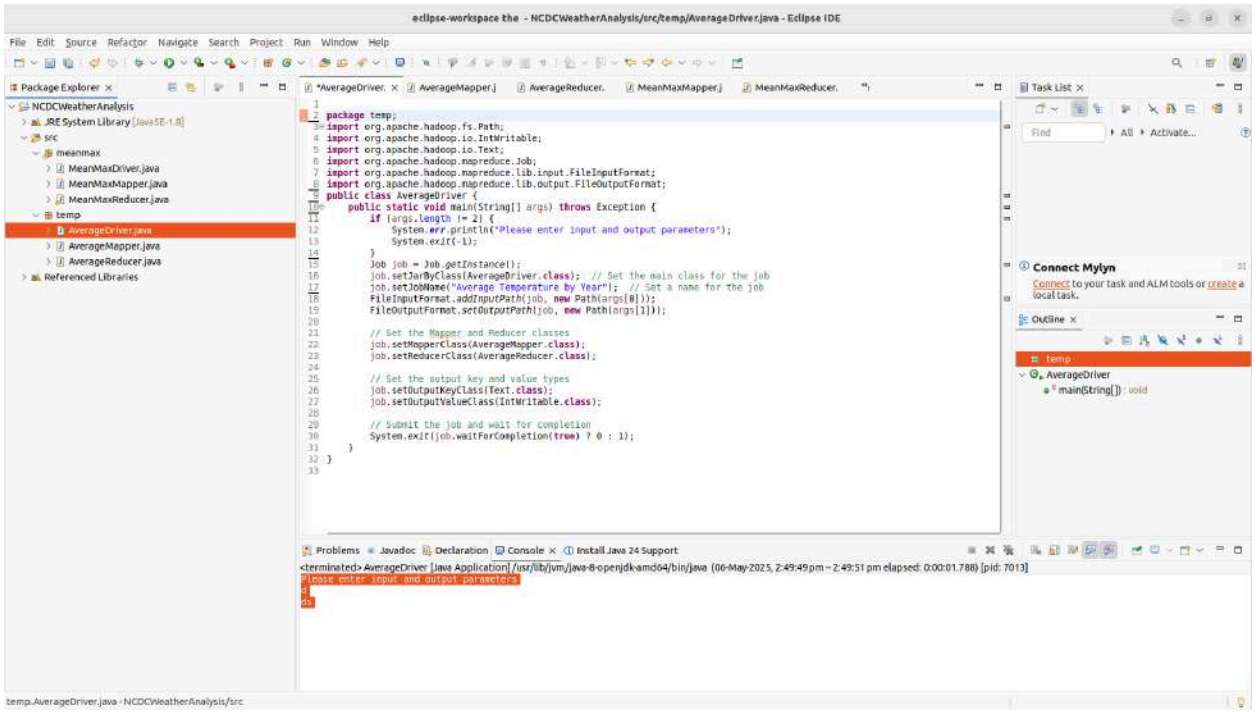
b) find the mean max temperature for every month.

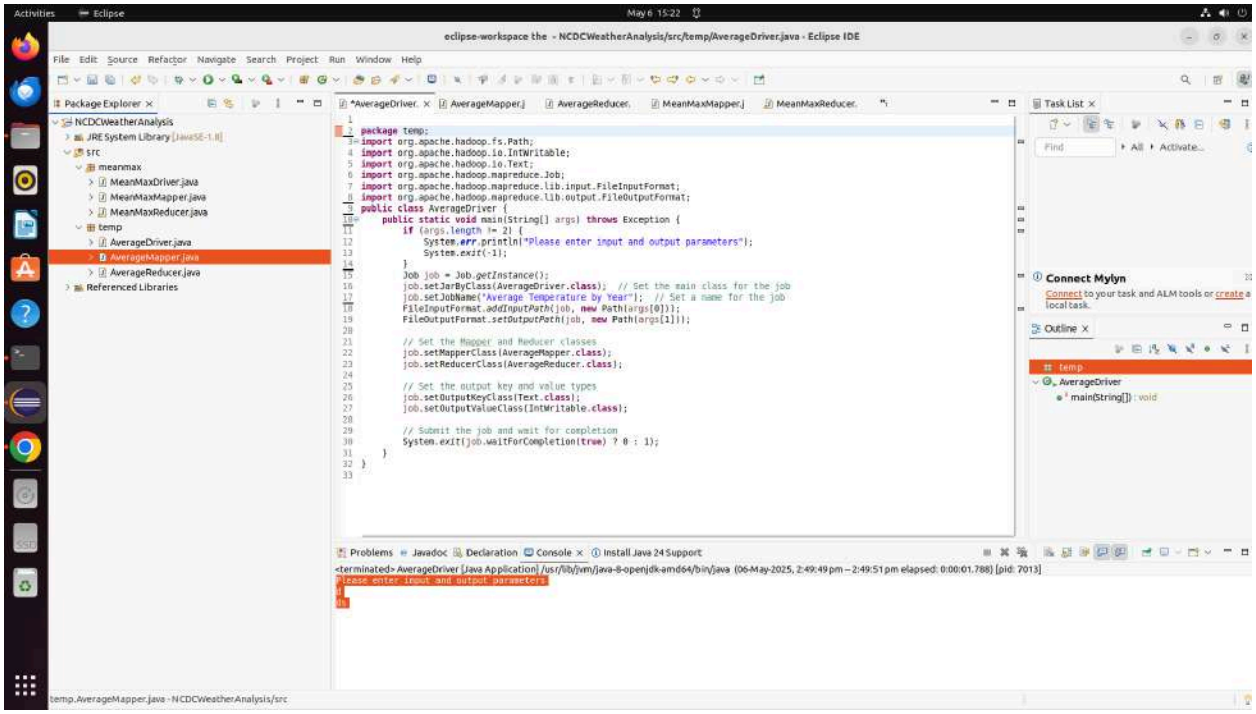
Code with screenshots:

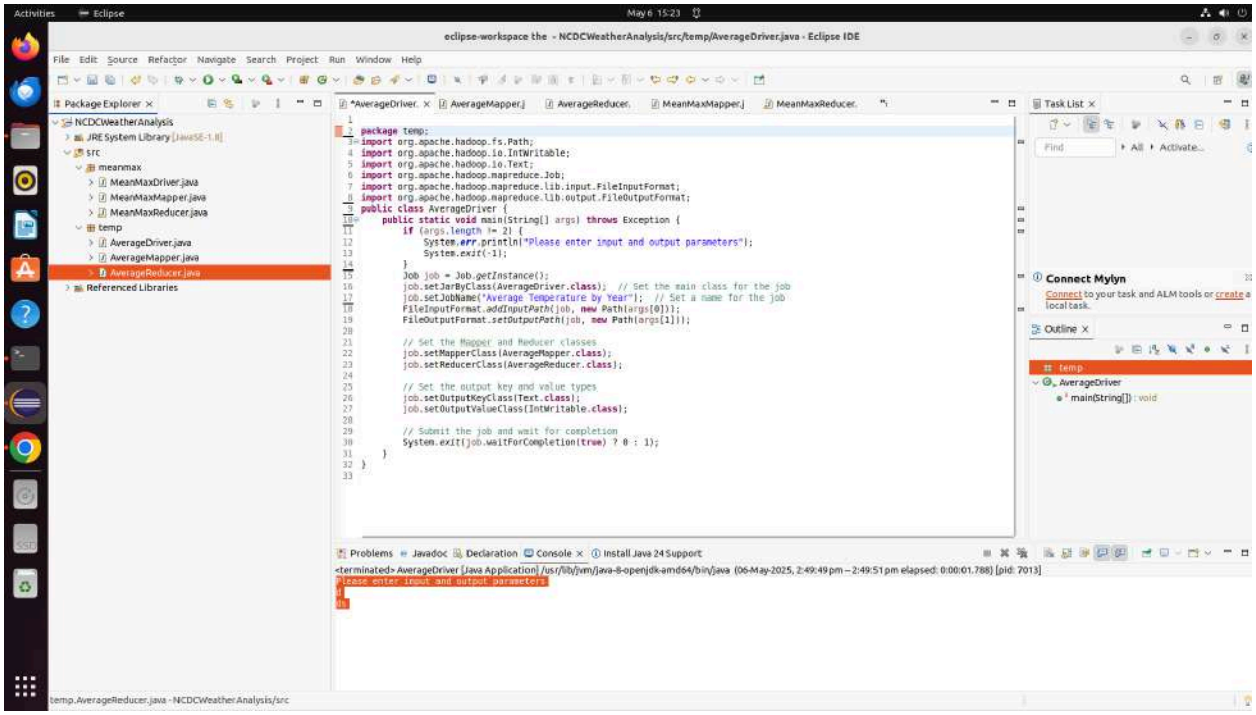

eclipse-workspace the - NCDCWeatherAnalysis/src/temp/AverageDriver.java - Eclipse IDE
File Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer
NCDCWeatherAnalysis
JRE System Library [JavaSE-1.8]
src
meanmax
MeanMaxDriver.java
MeanMaxMapper.java
MeanMaxReducer.java
temp
AverageDriver.java
AverageMapper.java
AverageReducer.java
Referenced Libraries
*AverageDriver.
AverageMapper.j
AverageReducer.
MeanMaxMapper.j
MeanMaxReducer.
package temp;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
public class AverageDriver {
public static void main(String[] args) throws Exception {
if (args.length != 2) {
System.err.println("Please enter input and output parameters");
System.exit(-1);
}
Job job = Job.getInstance();
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);
}
}
Task List
Connect Mylyn
Connect to your task and ALM tools or create a local task.
Outline
temp
AverageDriver
main(String[]) : void
Problems Javadoc 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.MeanMaxDriver.java - NCDCWeatherAnalysis/src

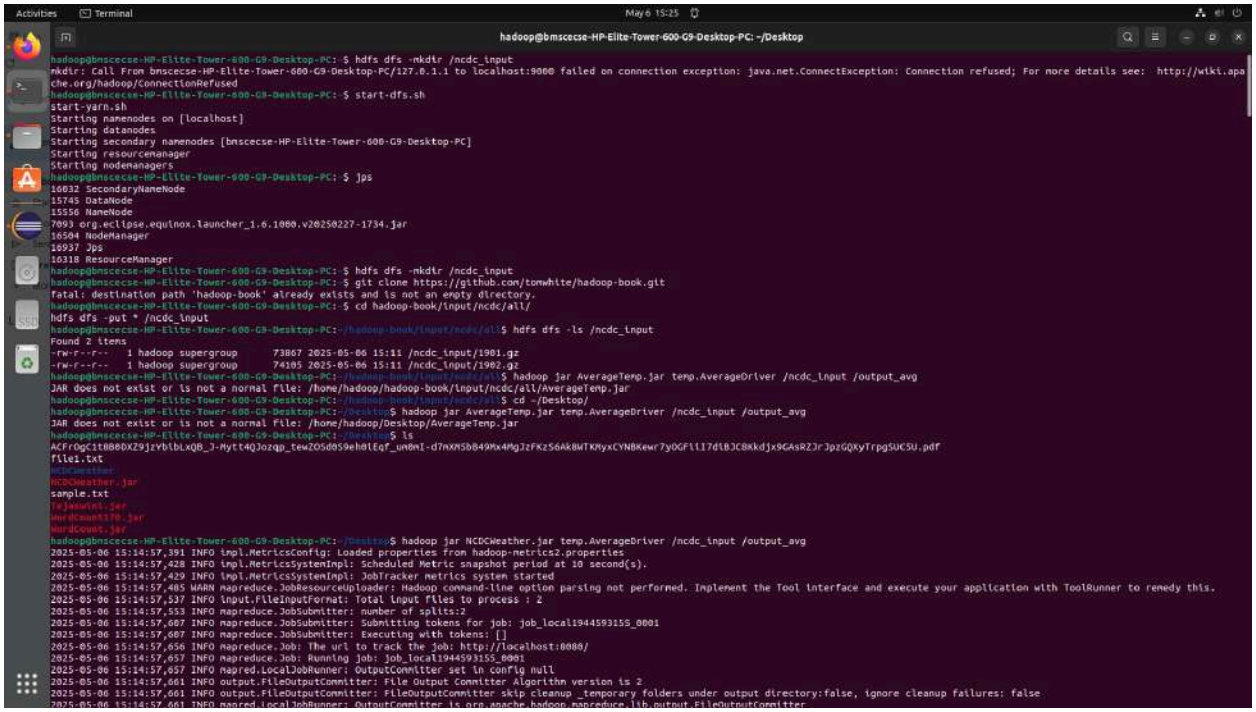

eclipse-workspace the - NCDCWeatherAnalysis/src/temp/AverageDriver.java - Eclipse IDE
File Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer
NCDCWeatherAnalysis
JRE System Library [JavaSE-1.8]
src
meanmax
MeanMaxDriver.java
MeanMaxMapper.java
MeanMaxReducer.java
temp
AverageDriver.java
AverageMapper.java
AverageReducer.java
Referenced Libraries
*AverageDriver.
AverageMapper.j
AverageReducer.
MeanMaxMapper.j
MeanMaxReducer.
package temp;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
public class AverageDriver {
public static void main(String[] args) throws Exception {
if (args.length != 2) {
System.err.println("Please enter input and output parameters");
System.exit(-1);
}
Job job = Job.getInstance();
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);
}
}
Task List
Connect Mylyn
Connect to your task and ALM tools or create a local task.
Outline
temp
AverageDriver
main(String[]) : void
Problems Javadoc 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.MeanMaxMapper.java - NCDCWeatherAnalysis/src

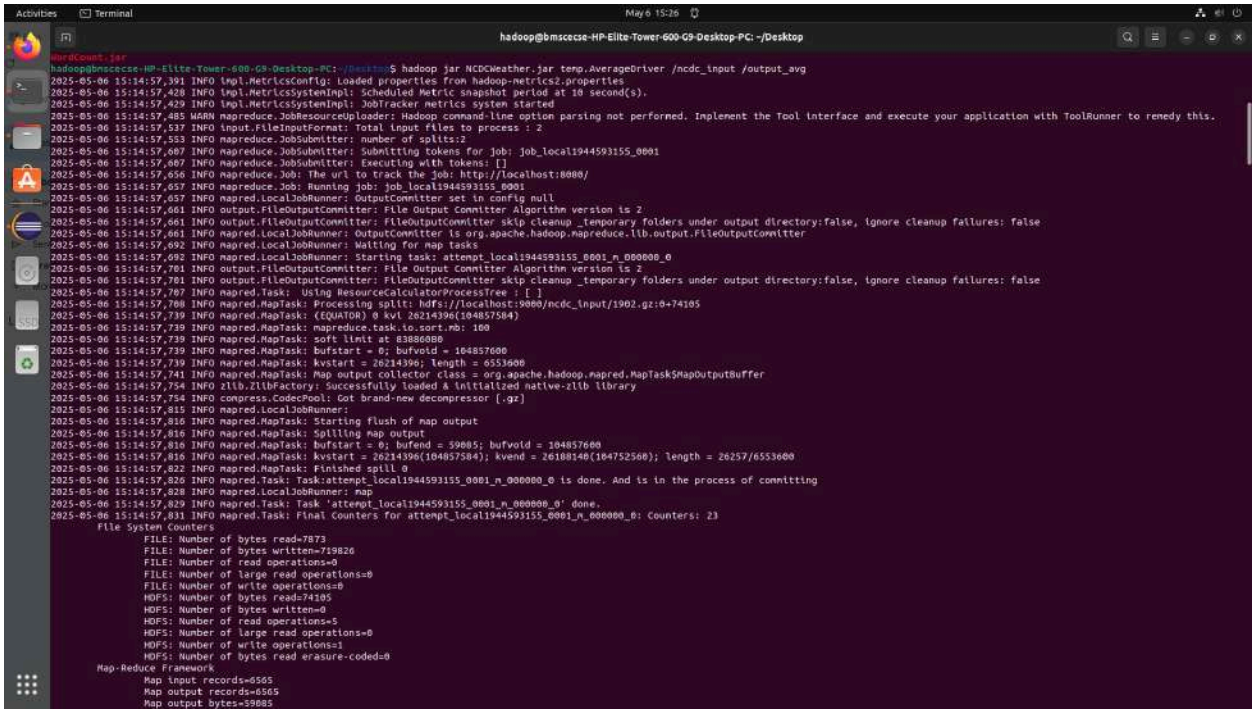

eclipse-workspace the - NCDCWeatherAnalysis/src/temp/AverageDriver.java - Eclipse IDE
File Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer
NCDCWeatherAnalysis
JRE System Library [JavaSE-1.8]
src
meanmax
MeanMaxDriver.java
MeanMaxMapper.java
MeanMaxReducer.java
temp
AverageDriver.java
AverageMapper.java
AverageReducer.java
Referenced Libraries
*AverageDriver.
AverageMapper.j
AverageReducer.
MeanMaxMapper.j
MeanMaxReducer.
package temp;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
public class AverageDriver {
public static void main(String[] args) throws Exception {
if (args.length != 2) {
System.err.println("Please enter input and output parameters");
System.exit(-1);
}
Job job = Job.getInstance();
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);
}
}
Task List
Connect Mylyn
Connect to your task and ALM tools or create a local task.
Outline
temp
AverageDriver
main(String[]) : void
Problems Javadoc 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

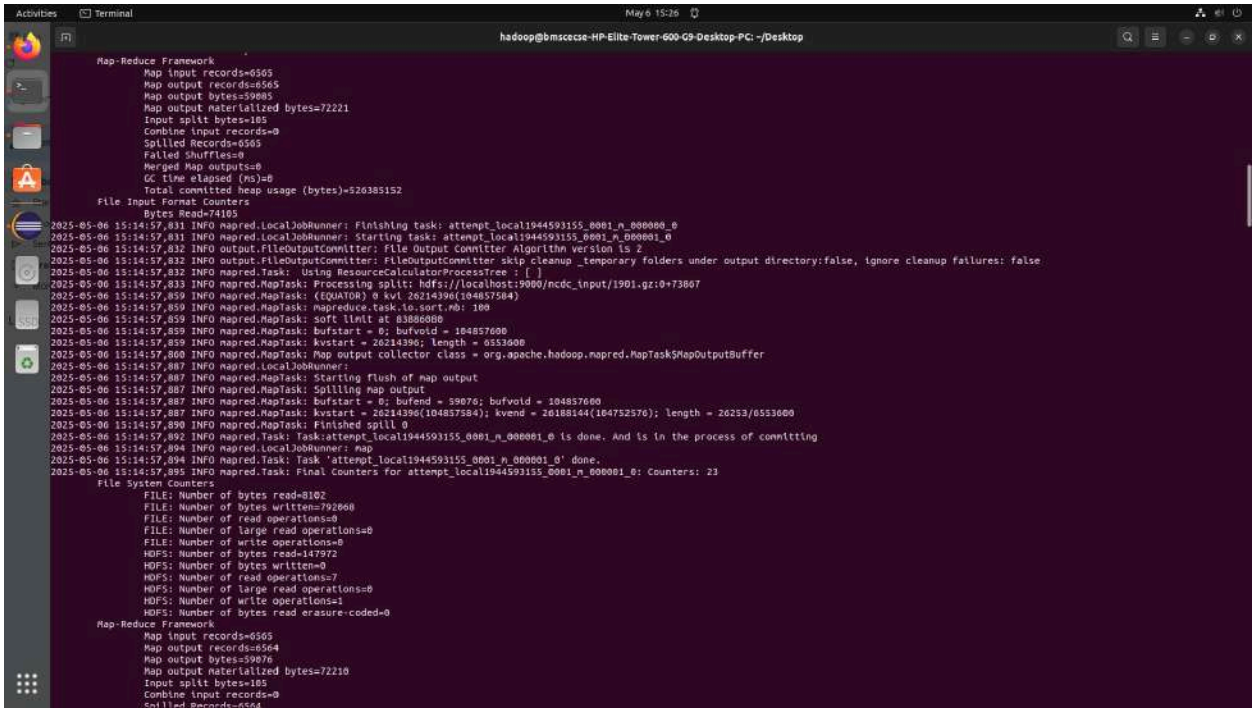

eclipse-workspace the - NCDCWeatherAnalysis/src/temp/AverageDriver.java - Eclipse IDE
File Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer
NCDCWeatherAnalysis
JRE System Library [JavaSE-1.8]
src
meanmax
MeanMaxDriver.java
MeanMaxMapper.java
MeanMaxReducer.java
temp
AverageDriver.java
AverageMapper.java
AverageReducer.java
Referenced Libraries
*AverageDriver.
AverageMapper.j
AverageReducer.
MeanMaxMapper.j
MeanMaxReducer.
package temp;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
public class AverageDriver {
public static void main(String[] args) throws Exception {
if (args.length != 2) {
System.err.println("Please enter input and output parameters");
System.exit(-1);
}
Job job = Job.getInstance();
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);
}
}
Task List
Connect Mylyn
Connect to your task and ALM tools or create a local task.
Outline
temp
AverageDriver
main(String[]) : void
Problems Javadoc 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

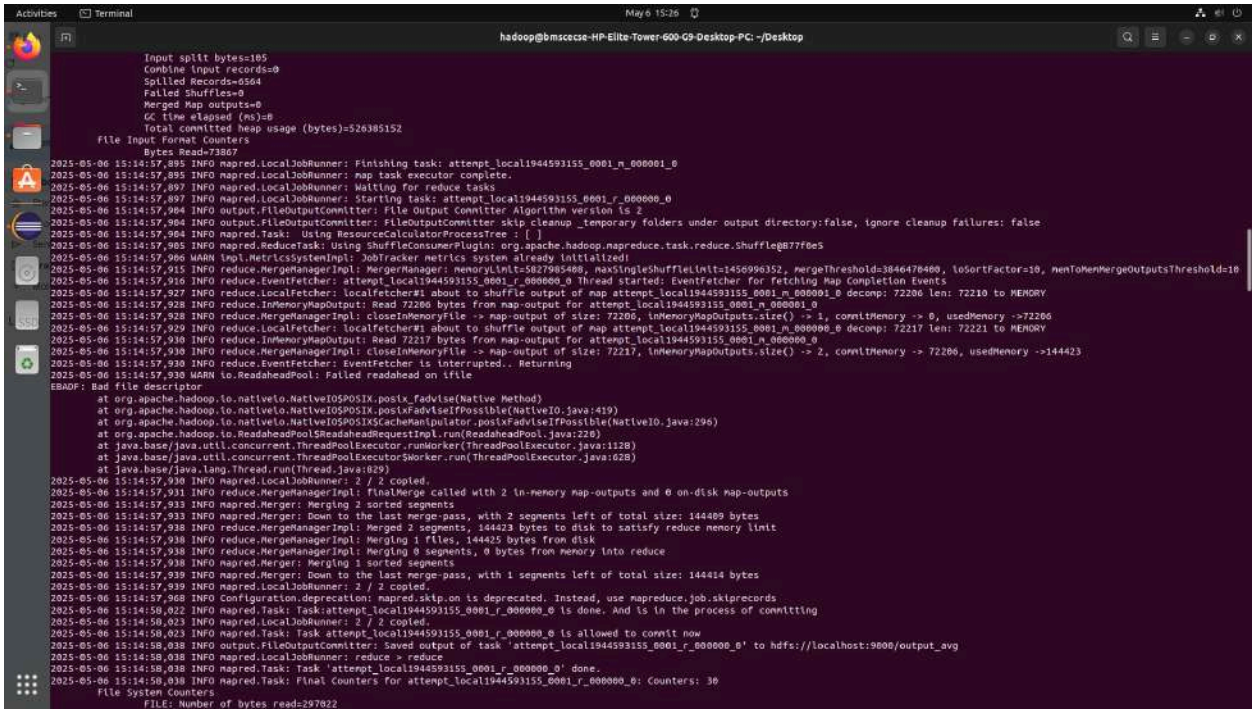

Activities
Eclipse
May 6 15:22
eclipse-workspace the - NCDCWeatherAnalysis/src/temp/AverageDriver.java - Eclipse IDE
File Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer
NCDCWeatherAnalysis
JRE System Library [JavaSE-1.8]
src
meanmax
MeanMaxDriver.java
MeanMaxMapper.java
MeanMaxReducer.java
temp
AverageDriver.java
AverageMapper.java
AverageReducer.java
Referenced Libraries
*AverageDriver.
AverageMapper.j
AverageReducer.
MeanMaxMapper.j
MeanMaxReducer.
Task List
Connect Mylyn
Connect to your task and ALM tools or create a local task.
Outline
temp
AverageDriver
main(String[]) : void
Problems
Javadoc
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.AverageMapper.java - NCDCWeatherAnalysis/src

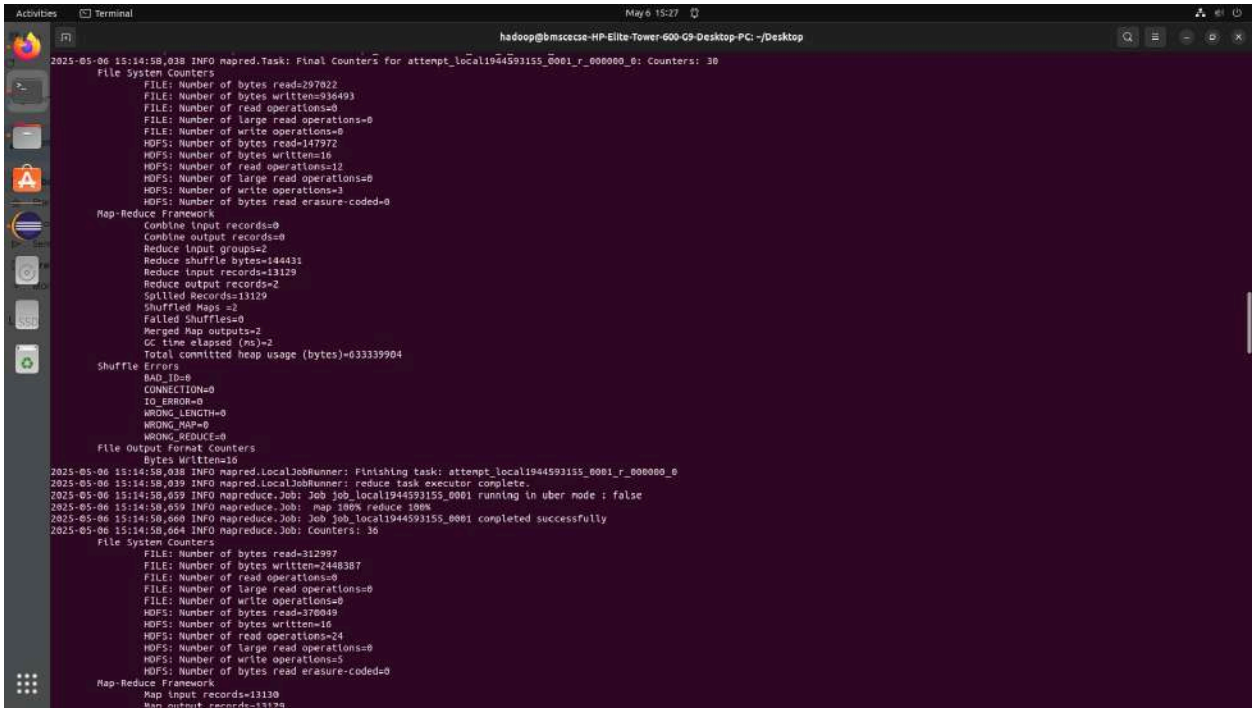

Activities
Eclipse
May 6 15:23
eclipse-workspace the - NCDCWeatherAnalysis/src/temp/AverageDriver.java - Eclipse IDE
File Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer
NCDCWeatherAnalysis
JRE System Library [JavaSE-1.8]
src
meanmax
MeanMaxDriver.java
MeanMaxMapper.java
MeanMaxReducer.java
temp
AverageDriver.java
AverageMapper.java
AverageReducer.java
Referenced Libraries
*AverageDriver.
AverageMapper.j
AverageReducer.
MeanMaxMapper.j
MeanMaxReducer.
Task List
Connect Mylyn
Connect to your task and ALM tools or create a local task.
Outline
temp
AverageDriver
main(String[]) : void
Problems
Javadoc
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.AverageReducer.java - NCDCWeatherAnalysis/src

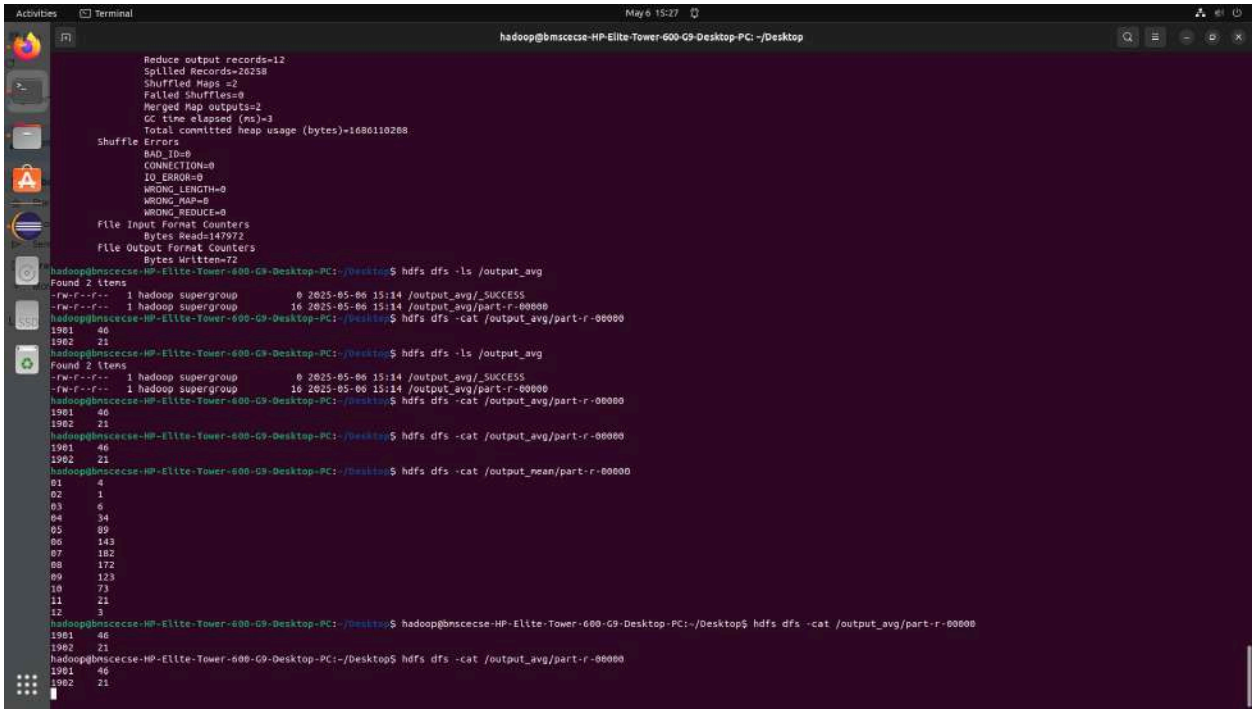

Activities
Terminal
May 6 15:25
hadoop@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC: ~/Desktop
hadoop@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC: $ hdfs dfs -mkdir /ncdc_input
mkdir: Call From bmscecse-HP-Elite-Tower-600-G9-Desktop-PC/127.0.1.1 to localhost:9000 failed on connection exception: java.net.ConnectException: Connection refused; For more details see: http://wiki.apache.org/hadoop/ConnectionRefused
hadoop@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC: $ start-dfs.sh
start-yarn.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [bmscecse-HP-Elite-Tower-600-G9-Desktop-PC]
Starting resourcemanager
Starting nodemanagers
hadoop@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC: $ jps
16032 SecondaryNameNode
15745 DataNode
15556 NameNode
7093 org.eclipse.equinox.launcher_1.6.1000.v20250227-1734.jar
16504 NodeManager
16937 Jps
16318 ResourceManager
hadoop@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC: $ hdfs dfs -mkdir /ncdc_input
hadoop@bmscecse-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@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC: $ cd hadoop-book/input/ncdc/all/
hdfs dfs -put * /ncdc_input
hadoop@bmscecse-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/1902.gz
hadoop@bmscecse-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@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC:~/hadoop-book/input/ncdc/all$ cd ~/Desktop/
hadoop@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC:~/Desktop$ hadoop jar AverageTemp.jar temp.AverageDriver /ncdc_input /output_avg
JAR does not exist or is not a normal file: /home/hadoop/Desktop/AverageTemp.jar
hadoop@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC:~/Desktop$ ls
ACFrOgC1tB80DXZ9jzYblLxQB_J-Mytt4QJozqp_tewZOSd059eh0lEqf_un0nI-d7nXM5b849Mx4MgJzFKzS6Ak8WTKMyxCYNBKewr7yOGFli17dlBJC8kkdjx9GAsRZJrJpzGQXyTrpgSUC5U.pdf
file1.txt
NCDCWeather
NCDCWeather.jar
sample.txt
temjawrit.jar
WordCount10.jar
WordCount.jar
hadoop@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC:~/Desktop$ hadoop jar NCDCWeather.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.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-05-06 15:14:57,429 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-05-06 15:14:57,485 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-05-06 15:14:57,537 INFO input.FileInputFormat: Total input files to process : 2
2025-05-06 15:14:57,553 INFO mapreduce.JobSubmitter: number of splits:2
2025-05-06 15:14:57,607 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1944593155_0001
2025-05-06 15:14:57,607 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-05-06 15:14:57,656 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-05-06 15:14:57,657 INFO mapreduce.Job: Running job: job_local1944593155_0001
2025-05-06 15:14:57,657 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-05-06 15:14:57,661 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-05-06 15:14:57,661 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false


Activities
Terminal
May 6 15:26
hadoop@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC: ~/Desktop
WordCount.jar
hadoop@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC:~/Desktop$ hadoop jar NCDCWeather.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.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-05-06 15:14:57,429 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-05-06 15:14:57,485 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-05-06 15:14:57,537 INFO input.FileInputFormat: Total input files to process : 2
2025-05-06 15:14:57,553 INFO mapreduce.JobSubmitter: number of splits:2
2025-05-06 15:14:57,607 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1944593155_0001
2025-05-06 15:14:57,607 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-05-06 15:14:57,656 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-05-06 15:14:57,657 INFO mapreduce.Job: Running job: job_local1944593155_0001
2025-05-06 15:14:57,657 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-05-06 15:14:57,661 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-05-06 15:14:57,661 INFO output.FileOutputCommitter: FileOutputCommitter 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.output.FileOutputCommitter
2025-05-06 15:14:57,692 INFO mapred.LocalJobRunner: Waiting for map tasks
2025-05-06 15:14:57,692 INFO mapred.LocalJobRunner: Starting task: attempt_local1944593155_0001_m_000000_0
2025-05-06 15:14:57,701 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-05-06 15:14:57,701 INFO output.FileOutputCommitter: FileOutputCommitter 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 15:14:57,708 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 kvi 26214396(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: bufstart = 0; bufvoid = 104857600
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 zlib.ZlibFactory: Successfully loaded & initialized native-zlib library
2025-05-06 15:14:57,754 INFO compress.CodecPool: Got brand-new decompressor [.gz]
2025-05-06 15:14:57,815 INFO mapred.LocalJobRunner:
2025-05-06 15:14:57,816 INFO mapred.MapTask: Starting flush of map output
2025-05-06 15:14:57,816 INFO mapred.MapTask: Spilling map output
2025-05-06 15:14:57,816 INFO mapred.MapTask: bufstart = 0; bufend = 59085; bufvoid = 104857600
2025-05-06 15:14:57,816 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26188140(104752560); length = 26257/6553600
2025-05-06 15:14:57,822 INFO mapred.MapTask: Finished spill 0
2025-05-06 15:14:57,826 INFO mapred.Task: Task:attempt_local1944593155_0001_m_000000_0 is done. And is in the process of committing
2025-05-06 15:14:57,828 INFO mapred.LocalJobRunner: map
2025-05-06 15:14:57,829 INFO mapred.Task: Task 'attempt_local1944593155_0001_m_000000_0' done.
2025-05-06 15:14:57,831 INFO mapred.Task: Final Counters for attempt_local1944593155_0001_m_000000_0: Counters: 23
File System Counters
FILE: Number of bytes read=7873
FILE: Number of bytes written=719826
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=74105
HDFS: Number of bytes written=0
HDFS: Number of read operations=5
HDFS: Number of large read operations=0
HDFS: Number of write operations=1
HDFS: Number of bytes read erasure-coded=0
Map-Reduce Framework
Map input records=6565
Map output records=6565
Map output bytes=59085


Activities
Terminal
May 6 15:26
hadoop@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC: ~/Desktop
Map-Reduce Framework
Map input records=6565
Map output records=6565
Map output bytes=59085
Map output materialized bytes=72221
Input split bytes=105
Combine input records=0
Spilled Records=6565
Failed Shuffles=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_m_000000_0
2025-05-06 15:14:57,831 INFO mapred.LocalJobRunner: Starting task: attempt_local1944593155_0001_m_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 output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-05-06 15:14:57,832 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
2025-05-06 15:14:57,833 INFO mapred.MapTask: Processing split: hdfs://localhost:9000/ncdc_input/1901.gz:0+73867
2025-05-06 15:14:57,859 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
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 83886080
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 = 26214396; length = 6553600
2025-05-06 15:14:57,860 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2025-05-06 15:14:57,887 INFO mapred.LocalJobRunner:
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; bufend = 59076; bufvoid = 104857600
2025-05-06 15:14:57,887 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26188144(104752576); length = 26253/6553600
2025-05-06 15:14:57,890 INFO mapred.MapTask: Finished spill 0
2025-05-06 15:14:57,892 INFO mapred.Task: Task:attempt_local1944593155_0001_m_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_m_000001_0' done.
2025-05-06 15:14:57,895 INFO mapred.Task: Final Counters for attempt_local1944593155_0001_m_000001_0: Counters: 23
File System Counters
FILE: Number of bytes read=8102
FILE: Number of bytes written=792860
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=0
HDFS: Number of read operations=7
HDFS: Number of large read operations=0
HDFS: Number of write operations=1
HDFS: Number of bytes read erasure-coded=0
Map-Reduce Framework
Map input records=6565
Map output records=6564
Map output bytes=59076
Map output materialized bytes=72210
Input split bytes=105
Combine input records=0


Activities
Terminal
May 6 15:26
hadoop@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC: ~/Desktop
Input split bytes=105
Combine input records=0
Spilled Records=6564
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=0
Total committed heap usage (bytes)=526385152
File Input Format Counters
Bytes Read=73867
2025-05-06 15:14:57,895 INFO mapred.LocalJobRunner: Finishing task: attempt_local1944593155_0001_m_000001_0
2025-05-06 15:14:57,895 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: File Output Committer Algorithm version is 2
2025-05-06 15:14:57,904 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-05-06 15:14:57,904 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
2025-05-06 15:14:57,905 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@877f0e5
2025-05-06 15:14:57,906 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2025-05-06 15:14:57,915 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=5827985408, maxSingleShuffleLimit=1456996352, mergeThreshold=3846470400, ioSortFactor=10, memToMemMergeOutputsThreshold=10
2025-05-06 15:14:57,916 INFO reduce.EventFetcher: attempt_local1944593155_0001_r_000000_0 Thread started: EventFetcher for fetching Map Completion Events
2025-05-06 15:14:57,927 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map attempt_local1944593155_0001_m_000001_0 decomp: 72206 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 15:14:57,928 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 decomp: 72217 len: 72221 to MEMORY
2025-05-06 15:14:57,930 INFO reduce.InMemoryMapOutput: Read 72217 bytes from map-output for attempt_local1944593155_0001_m_000000_0
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 reduce.EventFetcher: EventFetcher is interrupted.. Returning
2025-05-06 15:14:57,930 WARN io.ReadaheadPool: Failed readahead on ifile
EBADF: Bad file descriptor
at org.apache.hadoop.io.nativeio.NativeIOSPOSIX.posix_fadvise(Native Method)
at org.apache.hadoop.io.nativeio.NativeIOSPOSIX.posixFadviseIfPossible(NativeIO.java:419)
at org.apache.hadoop.io.nativeio.NativeIOSPOSIXSCacheManipulator.posixFadviseIfPossible(NativeIO.java:296)
at org.apache.hadoop.io.ReadaheadPool$ReadaheadRequestImpl.run(ReadaheadPool.java:220)
at java.base/java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1128)
at java.base/java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:628)
at java.base/java.lang.Thread.run(Thread.java:829)
2025-05-06 15:14:57,930 INFO mapred.LocalJobRunner: 2 / 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 15:14:57,938 INFO reduce.MergeManagerImpl: Merged 2 segments, 144423 bytes to disk to satisfy reduce memory limit
2025-05-06 15:14:57,938 INFO reduce.MergeManagerImpl: Merging 1 files, 144425 bytes from disk
2025-05-06 15:14:57,938 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce
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 / 2 copied.
2025-05-06 15:14:57,968 INFO Configuration.deprecation: mapred.skip.on is deprecated. Instead, use mapreduce.job.skiprecords
2025-05-06 15:14:58,022 INFO mapred.Task: Task:attempt_local1944593155_0001_r_000000_0 is done. And is in the process of committing
2025-05-06 15:14:58,023 INFO mapred.LocalJobRunner: 2 / 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 output.FileOutputCommitter: Saved output of task 'attempt_local1944593155_0001_r_000000_0' to hdfs://localhost:9000/output_avg
2025-05-06 15:14:58,038 INFO mapred.LocalJobRunner: reduce > reduce
2025-05-06 15:14:58,038 INFO mapred.Task: Task 'attempt_local1944593155_0001_r_000000_0' 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: Number of bytes read=297022

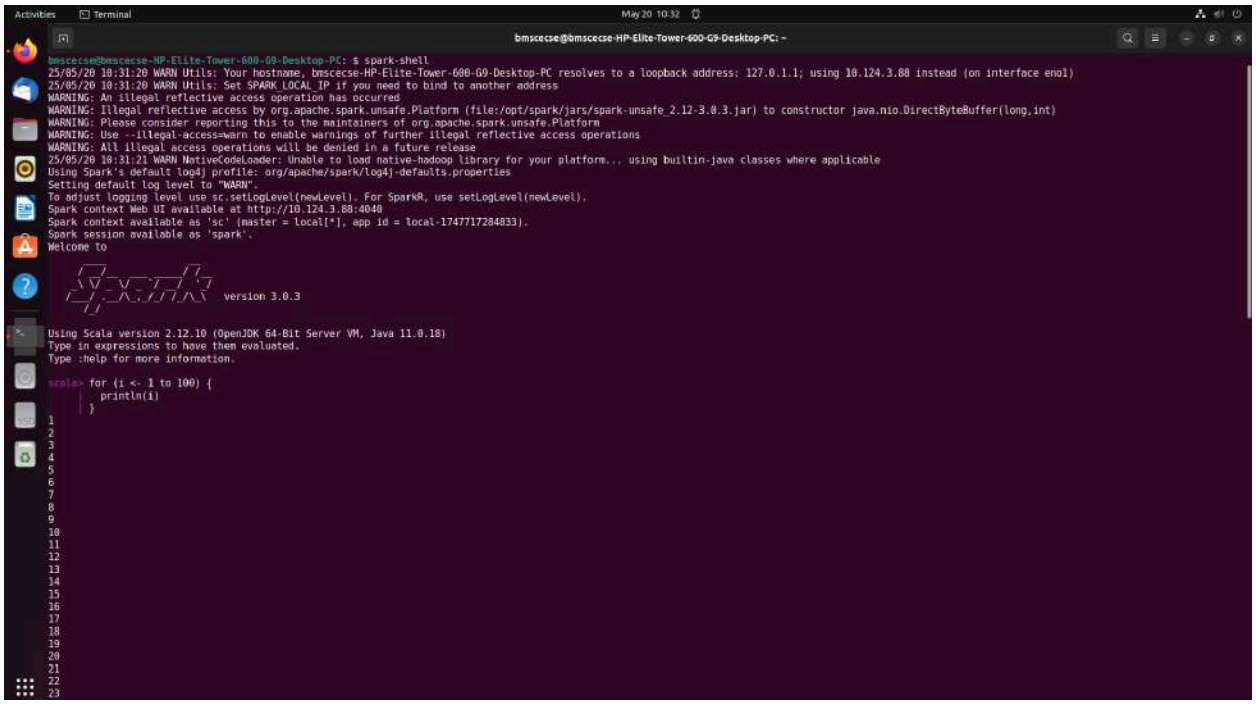

Activities
Terminal
May 6 15:27
hadoop@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC: ~/Desktop
2025-05-06 15:14:58,038 INFO mapred.Task: Final Counters for attempt_local1944593155_0001_r_000000_0: Counters: 30
File System Counters
FILE: Number of bytes read=297022
FILE: Number of bytes written=936493
FILE: 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
Reduce shuffle bytes=144431
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=16
2025-05-06 15:14:58,038 INFO mapred.LocalJobRunner: Finishing task: attempt_local1944593155_0001_r_000000_0
2025-05-06 15:14:58,039 INFO mapred.LocalJobRunner: reduce task executor complete.
2025-05-06 15:14:58,659 INFO mapreduce.Job: Job job_local1944593155_0001 running in uber mode : false
2025-05-06 15:14:58,659 INFO mapreduce.Job: map 100% reduce 100%
2025-05-06 15:14:58,660 INFO mapreduce.Job: Job job_local1944593155_0001 completed successfully
2025-05-06 15:14:58,664 INFO mapreduce.Job: Counters: 36
File System Counters
FILE: Number of bytes read=312997
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=370049
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=13130

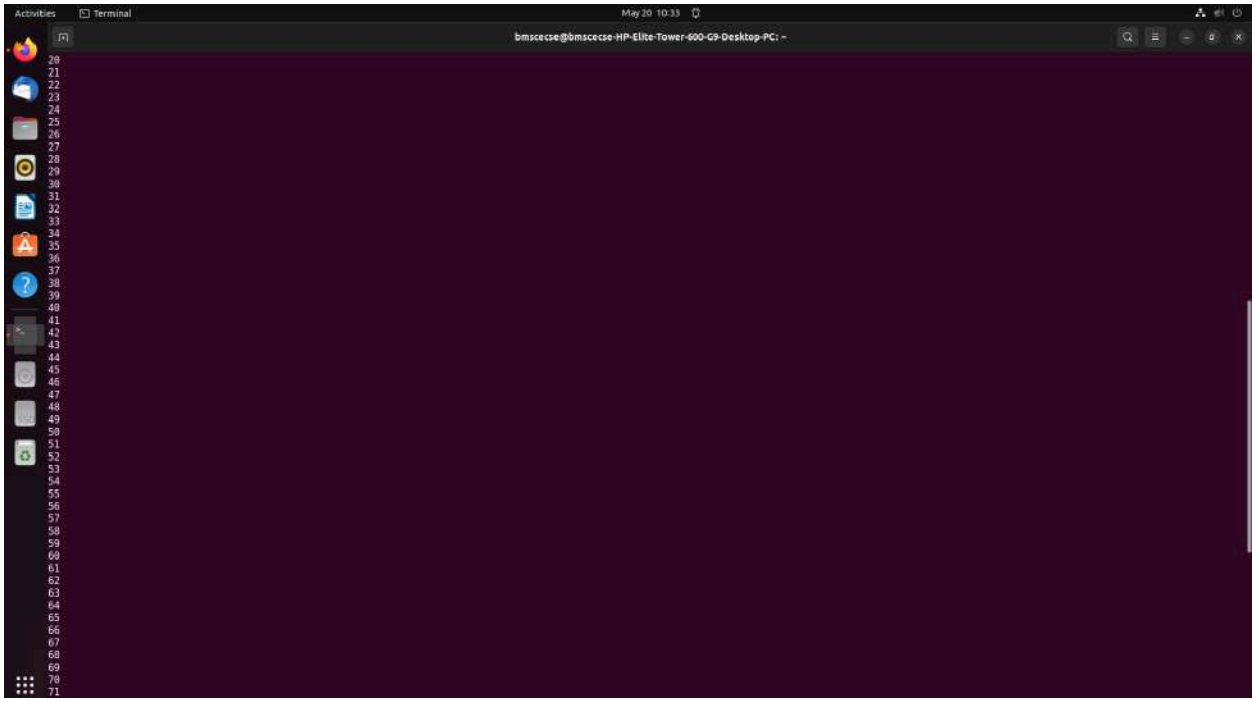

Activities
Terminal
May 6 15:27
hadoop@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC: ~/Desktop
Reduce output records=12
Spilled Records=26258
Shuffled Maps =2
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=3
Total committed heap usage (bytes)=1686110208
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
Bytes Read=147972
File Output Format Counters
Bytes Written=72
hadoop@bmscecse-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@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC:~/Desktop$ hdfs dfs -cat /output_avg/part-r-00000
1901 46
1902 21
hadoop@bmscecse-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@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC:~/Desktop$ hdfs dfs -cat /output_avg/part-r-00000
1901 46
1902 21
hadoop@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC:~/Desktop$ hdfs dfs -cat /output_avg/part-r-00000
1901 46
1902 21
hadoop@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC:~/Desktop$ hdfs dfs -cat /output_mean/part-r-00000
01 4
02 1
03 6
04 34
05 89
06 143
07 182
08 172
09 123
10 73
11 21
12 3
hadoop@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC:~/Desktop$ hadoop@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC:~/Desktop$ hdfs dfs -cat /output_avg/part-r-00000
1901 46
1902 21
hadoop@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC:~/Desktop$ hdfs dfs -cat /output_avg/part-r-00000
1901 46
1902 21

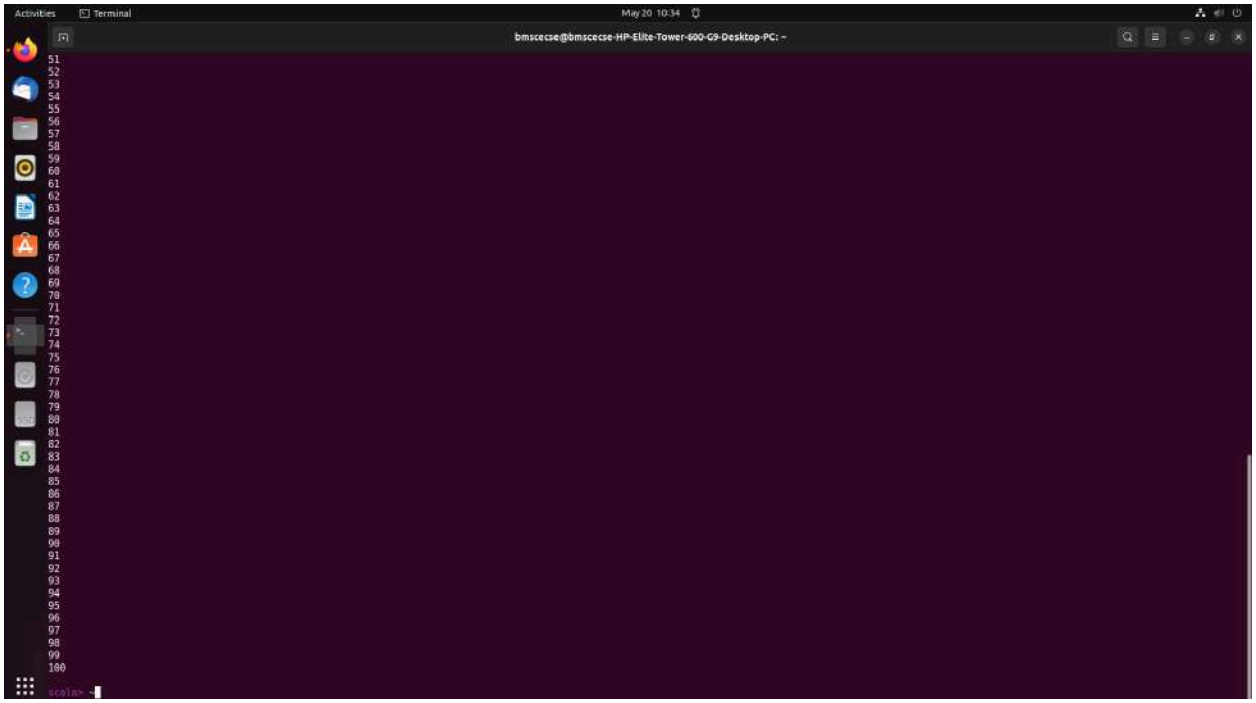
EXPERIMENT-8

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

Code with screenshots:




Activities
Terminal
May 20 10:33
bmscecse@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC: ~


Activities
Terminal
May 20 10:34
bmscecse@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC: ~
scala>

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:

```
bmscecse@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ spark-shell
25/05/20 15:32:38 WARN Utils: Your hostname, bmscecse-HP-Elite-Tower-800-G9-Desktop-PC resolves to a loopback address: 127.0.1.1
; using 10.124.2.8 instead (on interface eno1)
25/05/20 15:32:38 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.spark.unsafe.Platform (file:/opt/spark/jars/spark-unsafe_2.12-3.0.3.jar) to con
structor 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 wh
ere applicable
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Spark context Web UI available at http://10.124.2.8:4040
Spark context available as 'sc' (master = local[*], app id = local-1747735361481).
Spark session available as 'spark'.
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /___/ .__/\_,_/_/ /_/\_\   version 3.0.3
      /_/

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

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

scala>

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

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

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

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

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

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

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

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

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  |
+------+
```