

VISVESVARAYA TECHNOLOGICAL UNIVERSITY

“JnanaSangama”, Belgaum -590014, Karnataka.



LAB REPORT on

BIG DATA ANALYTICS (20CS6PEBDA)

Submitted by

Prathik Vittal (1BM19CS117)

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

May-2022 to July-2022

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 **Prathik Vittal(1BM19CS117)**, 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 2022. The Lab report has been approved as it satisfies the academic requirements in respect of a **Big Data Analytics - (20CS6PEBDA)** work prescribed for the said degree.

Dr. Antara Roy Choudhury
Assistant Professor
Department of CSE
BMSCE, Bengaluru

Dr. Jyothi S Nayak
Professor and Head
Department of CSE
BMSCE, Bengaluru

Index Sheet

Sl. No.	Experiment Title	Page No.
1	Students Database using Cassandra	
2	Library Database using Cassandra	
3	Students Database using MongoDB	

Course Outcome

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

Lab-1:

prathikvittal2508@Prathiks-MacBook-Pro ~ % cqlsh

/usr/local/Cellar/cassandra/4.0.4/libexec/bin/cqlsh.py:460: DeprecationWarning: Legacy execution parameters will be removed in 4.0. Consider using execution profiles.

/usr/local/Cellar/cassandra/4.0.4/libexec/bin/cqlsh.py:490: DeprecationWarning: Setting the consistency level at the session level will be removed in 4.0. Consider using execution profiles and setting the desired consistency level to the EXEC_PROFILE_DEFAULT profile.

Connected to Test Cluster at 127.0.0.1:9042

[cqlsh 6.0.0 | Cassandra 4.0.4 | CQL spec 3.4.5 | Native protocol v5]

Use HELP for help.

```
cqlsh> CREATE KEYSPACE Students WITH REPLICATION = { 'class': 'SimpleStrategy',  
'replication_factor': 1};
```

```
cqlsh> DESCRIBE KEYSPACES
```

```
assignment  system_auth      system_traces  
students    system_distributed system_views  
system      system_schema    system_virtual_schema
```

```
cqlsh> use students;
```

```
cqlsh:students> CREATE TABLE Student_info(  
    ... RollNo int PRIMARY KEY,  
    ... Name text,  
    ... DateOfJoin timestamp,  
    ... LastExamPercent double,  
    ... );
```

```
cqlsh:students> DESCRIBE TABLE Student_info;
```

```
CREATE TABLE students.student_info (  
    rollno int PRIMARY KEY,  
    dateofjoin timestamp,  
    lastexampercent double,  
    name 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 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';

```

```

cqsh:students> BEGIN BATCH INSERT INTO Student_info(RollNo, Name, DateOfJoin,
LastExamPercent) VALUES(1,'Puneeth Kumar','2021-03-12',95.5) APPLY BATCH;

```

```

cqsh:students> BEGIN BATCH INSERT INTO Student_info(RollNo, Name, DateOfJoin,
LastExamPercent) VALUES(2,'Pankaj Gupta','2019-05-22',92.3) APPLY BATCH;

```

```

cqsh:students> BEGIN BATCH INSERT INTO Student_info(RollNo, Name, DateOfJoin,
LastExamPercent) VALUES(3,'Preetham','2017-01-19',89.9) APPLY BATCH;

```

```

cqsh:students> SELECT * FROM Student_info;

```

rollno	dateofjoin	lastexampercent	name
1	2021-03-11 18:30:00.000000+0000	95.5	Puneeth Kumar
2	2019-05-21 18:30:00.000000+0000	92.3	Pankaj Gupta
3	2017-01-18 18:30:00.000000+0000	89.9	Preetham

(3 rows)

```

cqsh:students> SELECT * FROM Student_info WHERE RollNo IN (1,2);

```

rollno	dateofjoin	lastexampercent	name
1	2021-03-11 18:30:00.000000+0000	95.5	Puneeth Kumar
2	2019-05-21 18:30:00.000000+0000	92.3	Pankaj Gupta

(2 rows)

cqlsh:students> CREATE INDEX ON Student_info(Name);

cqlsh:students> SELECT * FROM Student_info WHERE Name='Preetham';

rollno	dateofjoin	lastexampercent	name
3	2017-01-18 18:30:00.000000+0000	89.9	Preetham

(1 rows)

cqlsh:students> SELECT RollNo, Name, LastExamPercent FROM Student_info LIMIT 2;

rollno	name	lastexampercent
1	Puneeth Kumar	95.5
2	Pankaj Gupta	92.3

(2 rows)

cqlsh:students> SELECT RollNo, LastExamPercent as "Score" FROM Student_info;

rollno	Score
1	95.5
2	92.3
3	89.9

(3 rows)

cqlsh:students> UPDATE Student_info SET LastExamPercent=97 WHERE RollNo=1;

```
cqlsh:students> SELECT * FROM Student_info;
```

rollno	dateofjoin	lastexampercent	name
1	2021-03-11 18:30:00.000000+0000	97	Puneeth Kumar
2	2019-05-21 18:30:00.000000+0000	92.3	Pankaj Gupta
3	2017-01-18 18:30:00.000000+0000	89.9	Preetham

(3 rows)

```
cqlsh:students> DELETE FROM Student_info WHERE RollNo=2;
```

```
cqlsh:students> SELECT * FROM Student_info;
```

rollno	dateofjoin	lastexampercent	name
1	2021-03-11 18:30:00.000000+0000	97	Puneeth Kumar
3	2017-01-18 18:30:00.000000+0000	89.9	Preetham

(2 rows)

```
cqlsh:students> DELETE LastExamPercent FROM Student_info where RollNo=1;
```

```
cqlsh:students> SELECT * FROM Student_info
```

... ;

rollno	dateofjoin	lastexampercent	name
1	2021-03-11 18:30:00.000000+0000	null	Puneeth Kumar
3	2017-01-18 18:30:00.000000+0000	89.9	Preetham

(2 rows)

```
cqlsh:students> SELECT RollNo,LastExamPercent as "Score" FROM Student_info;
```

rollno	Score
--------	-------

```
1 | null
3 | 89.9
```

(2 rows)

```
cqlsh:students> UPDATE Student_info SET LastExamPercent=97 WHERE RollNo=1;
cqlsh:students> SELECT * FROM Student_info;
```

rollno	dateofjoin	lastexampercent	name
1	2021-03-11 18:30:00.000000+0000	97	Puneeth Kumar
3	2017-01-18 18:30:00.000000+0000	89.9	Preetham

(2 rows)

```
cqlsh:students> CREATE TABLE Project_Details(
    ... pid int,
    ... pname text,
    ... stud_name text,
    ... duration int,
    ... PRIMARY KEY(pid,pname)
    ... );
cqlsh:students> DESCRIBE TABLE Project_Details;
```

```
CREATE TABLE students.project_details (
    pid int,
    pname text,
    duration int,
    stud_name text,
    PRIMARY KEY (pid, pname)
) WITH CLUSTERING ORDER BY (pname ASC)
    AND 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 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> BEGIN BATCH INSERT INTO Project_Details(pid,pname,duration,stud_name)
VALUES(1,'LMS',3,'Puneeth Kumar') INSERT INTO Project_Details(pid,pname,duration,stud_name)
VALUES(2,'Colab',4,'Preetham') APPLY BATCH;

```

```

cqlsh:students> SELECT * FROM Project_Details;

```

```

pid | pname | duration | stud_name

```

```

-----+-----+-----+-----

```

```

1 | LMS | 3 | Puneeth Kumar
2 | Colab | 4 | Preetham

```

(2 rows)

```

cqlsh:students> SELECT * FROM Project_Details WHERE PID=2;

```

```

pid | pname | duration | stud_name

```

```

-----+-----+-----+-----

```

```

2 | Colab | 4 | Preetham

```

(1 rows)

```
cqlsh:students> CREATE INDEX ON Project_Details(stud_name);
```

```
cqlsh:students> SELECT * FROM Project_Details WHERE stud_name='Puneeth Kumar';
```

pid	pname	duration	stud_name
-----	-------	----------	-----------

1	LMS	3	Puneeth Kumar
---	-----	---	---------------

(1 rows)

```
cqlsh:students> UPDATE Project_Details SET duration=5 WHERE pid=1 and pname='LMS';
```

```
cqlsh:students> SELECT * FROM Project_Details;
```

pid	pname	duration	stud_name
-----	-------	----------	-----------

1	LMS	5	Puneeth Kumar
2	Colab	4	Preetham

(2 rows)

```
cqlsh:students> DELETE duration FROM Project_Details where pid=2 and pname='Colab';
```

```
cqlsh:students> SELECT * FROM Project_Details;
```

pid	pname	duration	stud_name
-----	-------	----------	-----------

1	LMS	5	Puneeth Kumar
2	Colab	null	Preetham

(2 rows)

```
cqlsh:students>
```

Lab-2:

prathikvittal2508@Prathiks-MacBook-Pro ~ % cqlsh

/usr/local/Cellar/cassandra/4.0.4/libexec/bin/cqlsh.py:460: DeprecationWarning: Legacy execution parameters will be removed in 4.0. Consider using execution profiles.

/usr/local/Cellar/cassandra/4.0.4/libexec/bin/cqlsh.py:490: DeprecationWarning: Setting the consistency level at the session level will be removed in 4.0. Consider using execution profiles and setting the desired consistency level to the EXEC_PROFILE_DEFAULT profile.

Connected to Test Cluster at 127.0.0.1:9042

[cqlsh 6.0.0 | Cassandra 4.0.4 | CQL spec 3.4.5 | Native protocol v5]

Use HELP for help.

```
cqlsh> CREATE KEYSPACE Library WITH REPLICATION = { 'class': 'SimpleStrategy',  
'replication_factor': 1};
```

```
cqlsh> USE Library;
```

```
cqlsh:library> CREATE TABLE library_info (  
    ... stud_id int,  
    ... stud_name text,  
    ... book_name text,  
    ... book_id int,  
    ... date_of_issue timestamp,  
    ... counter_value counter,  
    ... PRIMARY KEY (stud_id, stud_name, book_name, book_id, date_of_issue)  
    ... );
```

```
cqlsh:library> DESCRIBE TABLE Library_Info;
```

```
CREATE TABLE library.library_info (  
    stud_id int,  
    stud_name text,  
    book_name text,  
    book_id int,  
    date_of_issue timestamp,  
    counter_value counter,  
    PRIMARY KEY (stud_id, stud_name, book_name, book_id, date_of_issue)
```

```
) WITH CLUSTERING ORDER BY (stud_name ASC, book_name ASC, book_id ASC, date_of_issue ASC)
```

```
AND 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 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:library> Library_Info SET Counter_value=Counter_value+1 where Stud_Id=1 and  
Stud_Name='Pankaj Gupta' and Book_name='BDA' and Book_id=111 and Date_Of_Issue='2021-03-15';
```

```
SyntaxException: line 1:0 no viable alternative at input 'Library_Info' ([Library_Info]...)
```

```
cqlsh:library> library_info SET Counter_value=Counter_value+1 where Stud_Id=1 and  
Stud_Name='Pankaj Gupta' and Book_name='BDA' and Book_id=111 and Date_Of_Issue='2021-03-15';
```

```
SyntaxException: line 1:0 no viable alternative at input 'library_info' ([library_info]...)
```

```
cqlsh:library> UPDATE Library_Info SET Counter_value=Counter_value+1 where Stud_Id=1 and  
Stud_Name='Pankaj Gupta' and Book_name='BDA' and Book_id=111 and Date_Of_Issue='2021-03-15';
```

```
cqlsh:library> UPDATE Library_Info SET Counter_value=Counter_value+1 where Stud_Id=2 and  
Stud_Name='Priyanka' and Book_name='OOMD' and Book_id=112 and Date_Of_Issue='2021-02-12';
```

```
cqlsh:library> UPDATE Library_Info SET Counter_value=Counter_value+1 where Stud_Id=112 and  
Stud_Name='Ashwin' and Book_name='BDA' and Book_id=1123 and Date_Of_Issue='2021-01-18';
```

```
cqlsh:library> SELECT * FROM Library_Info;
```

stud_id	stud_name	book_name	book_id	date_of_issue	counter_value
---------	-----------	-----------	---------	---------------	---------------

stud_id	stud_name	book_name	book_id	date_of_issue	counter_value
1	Pankaj Gupta	BDA	111	2021-03-14 18:30:00.000000+0000	1
2	Priyanka	OOMD	112	2021-02-11 18:30:00.000000+0000	1
112	Ashwin	BDA	1123	2021-01-17 18:30:00.000000+0000	1

(3 rows)

```
cqlsh:library> UPDATE Library_Info SET Counter_value=Counter_value+1 where Stud_Id=112 and Stud_Name='Aswin' and Book_name='BDA' and Book_id=1123 and Date_Of_Issue='2021-01-18';
```

```
cqlsh:library> cqlsh:Library> SELECT * FROM Library_Info;
```

SyntaxException: line 1:0 no viable alternative at input 'cqlsh' ([cqlsh]...)

```
cqlsh:library> SELECT * FROM Library_Info;
```

stud_id	stud_name	book_name	book_id	date_of_issue	counter_value
1	Pankaj Gupta	BDA	111	2021-03-14 18:30:00.000000+0000	1
2	Priyanka	OOMD	112	2021-02-11 18:30:00.000000+0000	1
112	Ashwin	BDA	1123	2021-01-17 18:30:00.000000+0000	1
112	Aswin	BDA	1123	2021-01-17 18:30:00.000000+0000	1

(4 rows)

```
cqlsh:library> UPDATE Library_Info SET Counter_value=Counter_value+1 where Stud_Id=112 and Stud_Name='Ashwin' and Book_name='BDA' and Book_id=1123 and Date_Of_Issue='2021-01-18';
```

```
cqlsh:library> SELECT * FROM Library_Info;
```

stud_id	stud_name	book_name	book_id	date_of_issue	counter_value
1	Pankaj Gupta	BDA	111	2021-03-14 18:30:00.000000+0000	1
2	Priyanka	OOMD	112	2021-02-11 18:30:00.000000+0000	1
112	Ashwin	BDA	1123	2021-01-17 18:30:00.000000+0000	2
112	Aswin	BDA	1123	2021-01-17 18:30:00.000000+0000	1

(4 rows)

```
cqlsh:library> COPY
Library_Info(Stud_Id,Stud_Name,Book_Name,Book_Id,Date_Of_Issue,Counter_value) TO
'g:\libraryInfo.csv';
```

Using 7 child processes

Starting copy of library.library_info with columns [stud_id, stud_name, book_name, book_id, date_of_issue, counter_value].

Processed: 4 rows; Rate: 13 rows/s; Avg. rate: 6 rows/s

4 rows exported to 1 files in 0.652 seconds.

```
cqlsh:library> CREATE TABLE Library_Info_Import( Stud_Id int, Counter_value counter, Stud_Name text,
Book_Name text, Book_Id int, Date_Of_Issue timestamp, PRIMARY
KEY(Stud_Id,Stud_Name,Book_Name,Book_Id,Date_Of_Issue));
```

```
cqlsh:library> COPY
Library_Info_Import(Stud_Id,Stud_Name,Book_Name,Book_Id,Date_Of_Issue,Counter_value) FROM
'g:\libraryInfo.csv';
```

Using 7 child processes

Starting copy of library.library_info_import with columns [stud_id, stud_name, book_name, book_id, date_of_issue, counter_value].

Processed: 4 rows; Rate: 3 rows/s; Avg. rate: 5 rows/s

4 rows imported from 1 files in 0.771 seconds (0 skipped).

```
cqlsh:library> SELECT * FROM Library_Info_Import;
```

stud_id	stud_name	book_name	book_id	date_of_issue	counter_value
1	Pankaj Gupta	BDA	111	2021-03-14 18:30:00.000000+0000	1
2	Priyanka	OOMD	112	2021-02-11 18:30:00.000000+0000	1
112	Ashwin	BDA	1123	2021-01-17 18:30:00.000000+0000	2
112	Aswin	BDA	1123	2021-01-17 18:30:00.000000+0000	1

(4 rows)

```
cqlsh:library>
```

Lab-3:

prathikvittal2508@Prathiks-MacBook-Pro ~ % mongosh

Current Mongosh Log ID: 629cf050a8a12ff9e092e50b

Connecting to:

mongodb://127.0.0.1:27017/?directConnection=true&serverSelectionTimeoutMS=2000&appName=mongosh+1.4.2

Using MongoDB: 5.0.7

Using Mongosh: 1.4.2

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

The server generated these startup warnings when booting:

2022-06-04T02:21:44.003+05:30: Access control is not enabled for the database. Read and write access to data and configuration is unrestricted

2022-06-04T02:21:44.003+05:30: Soft rlimits for open file descriptors too low

Warning: Found ~/.mongorc.js, but not ~/.mongoshrc.js. ~/.mongorc.js will not be loaded.

You may want to copy or rename ~/.mongorc.js to ~/.mongoshrc.js.

test> show dbs

admin 40.00 KiB

config 36.00 KiB

local 72.00 KiB

mydb 192.00 KiB

sampledb 76.00 KiB

test> use Student

switched to db Student

Student> show collections

Student> db.student_info.insertMany([{"RollNo":10, Name:"XYZ", Age:21, ContactNo:9876787675, EmailId:"xyz.cs18@bmsce.ac.in"}, {"RollNo":11, Name:"ABC", Age:21, ContactNo:9886786675, EmailId:"abc.cs18@bmsce.ac.in"}, {"RollNo":12, Name:"DEF", Age:20, ContactNo:8876737674, EmailId:"def.cs18@bmsce.ac.in"}])

```
{
  acknowledged: true,
  insertedIds: {
    '0': ObjectId("629cf06fa8a12ff9e092e50c"),
    '1': ObjectId("629cf06fa8a12ff9e092e50d"),
    '2': ObjectId("629cf06fa8a12ff9e092e50e")
  }
}
```

```
Student> db.student_info.find()
```

```
[
  {
    _id: ObjectId("629cf06fa8a12ff9e092e50c"),
    RollNo: 10,
    Name: 'XYZ',
    Age: 21,
    ContactNo: 9876787675,
    EmailId: 'xyz.cs18@bmsce.ac.in'
  },
  {
    _id: ObjectId("629cf06fa8a12ff9e092e50d"),
    RollNo: 11,
    Name: 'ABC',
    Age: 21,
    ContactNo: 9886786675,
    EmailId: 'abc.cs18@bmsce.ac.in'
  },
  {
    _id: ObjectId("629cf06fa8a12ff9e092e50e"),
    RollNo: 12,
    Name: 'DEF',
    Age: 20,
    ContactNo: 8876737674,
```



```

    EmailId: 'def.cs18@bmsce.ac.in'
  }
]
Student> db.student_info.update({RollNo:10},{ $set:{EmailId:"xyznew.cs18@bmsce.ac.in"}})
DeprecationWarning: Collection.update() is deprecated. Use updateOne, updateMany, or bulkWrite.
{
  acknowledged: true,
  insertedId: null,
  matchedCount: 1,
  modifiedCount: 1,
  upsertedCount: 0
}
Student> db.student_info.find()
[
  {
    _id: ObjectId("629cf06fa8a12ff9e092e50c"),
    RollNo: 10,
    Name: 'XYZ',
    Age: 21,
    ContactNo: 9876787675,
    EmailId: 'xyznew.cs18@bmsce.ac.in'
  },
  {
    _id: ObjectId("629cf06fa8a12ff9e092e50d"),
    RollNo: 11,
    Name: 'ABC',
    Age: 21,
    ContactNo: 9886786675,
    EmailId: 'abc.cs18@bmsce.ac.in'
  },
  {
    _id: ObjectId("629cf06fa8a12ff9e092e50e"),

```

```

RollNo: 12,
Name: 'DEF',
Age: 20,
ContactNo: 8876737674,
EmailId: 'def.cs18@bmsce.ac.in'
}
]
Student> db.student_info.replaceOne({RollNo:11},{RollNo:11, Name:"FEM", Age:21,
ContactNo:9886786675, EmailId:"abc.cs18@bmsce.ac.in"})
{
  acknowledged: true,
  insertedId: null,
  matchedCount: 1,
  modifiedCount: 1,
  upsertedCount: 0
}
Student> db.student_info.find()
[
  {
    _id: ObjectId("629cf06fa8a12ff9e092e50c"),
    RollNo: 10,
    Name: 'XYZ',
    Age: 21,
    ContactNo: 9876787675,
    EmailId: 'xyznew.cs18@bmsce.ac.in'
  },
  {
    _id: ObjectId("629cf06fa8a12ff9e092e50d"),
    RollNo: 11,
    Name: 'FEM',
    Age: 21,
    ContactNo: 9886786675,

```

```
EmailId: 'abc.cs18@bmsce.ac.in'
},
{
  _id: ObjectId("629cf06fa8a12ff9e092e50e"),
  RollNo: 12,
  Name: 'DEF',
  Age: 20,
  ContactNo: 8876737674,
  EmailId: 'def.cs18@bmsce.ac.in'
}
]
Student> db.student_info.drop()
true
Student> db.student_info.find()

Student>
```