

VISVESVARAYA TECHNOLOGICAL UNIVERSITY

“JnanaSangama”, Belgaum -590014, Karnataka.



LAB REPORT on

BIG DATA ANALYTICS (20CS6PEBDA)

Submitted by

HIMANI BM (1BM19CS062)

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 **HIMANI BM(1BM19CS062)**, 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.

ANTARA ROY CHOUDURY
Designation
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	cassandra	

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

```

1  cqlsh> create keyspace employee_space WITH REPLICATION = {'class' : 'SimpleStrategy','replication_factor':2};
2
3  CREATE TABLE employee_space.employee_info (emp_id int PRIMARY KEY,emp_name text,designation text,date_of_joining timestamp,salary float,dept_name text);
4
5  cqlsh> begin batch INSERT INTO employee_space.employee_info(emp_id,emp_name,designation,date_of_joining,salary,dept_name) VALUES(1,'Damodar','Manager','2022-01-24',100000,'Marketing');
6  ... apply batch;
7  cqlsh> begin batch INSERT INTO employee_space.employee_info(emp_id,emp_name,designation,date_of_joining,salary,dept_name) VALUES(2,'Mahalaxmi','Accountant','2021-01-24',200000,'Accounts');
8  ... INSERT INTO employee_space.employee_info(emp_id,emp_name,designation,date_of_joining,salary,dept_name) VALUES(3,'Mahesh','Manager','2021-03-24',500000,'Marketing');
9  ... INSERT INTO employee_space.employee_info(emp_id,emp_name,designation,date_of_joining,salary,dept_name) VALUES(4,'Nidhi','Administrator','2021-05-24',500000,'Administration');
10 ... INSERT INTO employee_space.employee_info(emp_id,emp_name,designation,date_of_joining,salary,dept_name) VALUES(5,'Rahul','Administrator','2009-05-24',200000,'Administration');
11 ... apply batch;
12
13 cqlsh> use employee_space;
14
15 cqlsh:employee_space> select * from employee_info;
16
17 emp_id | date_of_joining          | dept_name | designation | emp_name | salary
18 -----+-----+-----+-----+-----+-----
19      5 | 2009-05-23 18:30:00.000000+0000 | Administration | Administrator | Rahul | 2e+06
20      1 | 2022-01-23 18:30:00.000000+0000 | Marketing | Manager | Damodar | 1e+05
21      2 | 2021-01-23 18:30:00.000000+0000 | Accounts | Accountant | Mahalaxmi | 2e+05
22      4 | 2021-05-23 18:30:00.000000+0000 | Administration | Administrator | Nidhi | 5e+05
23      3 | 2021-03-23 18:30:00.000000+0000 | Marketing | Manager | Mahesh | 5e+05
24
25 (5 rows)
26
27 cqlsh:employee_space> update employee_info set emp_name='Radha' where emp_id=1;
28
29 cqlsh:employee_space> update employee_info set dept_name='Development' where emp_id=1;
30
31 cqlsh:employee_space> select * from employee_info;
32
33 emp_id | date_of_joining          | dept_name | designation | emp_name | salary
34 -----+-----+-----+-----+-----+-----
35      5 | 2009-05-23 18:30:00.000000+0000 | Administration | Administrator | Rahul | 2e+06
36      1 | 2022-01-23 18:30:00.000000+0000 | Development | Manager | Radha | 1e+05
37      2 | 2021-01-23 18:30:00.000000+0000 | Accounts | Accountant | Mahalaxmi | 2e+05
38      4 | 2021-05-23 18:30:00.000000+0000 | Administration | Administrator | Nidhi | 5e+05
39      3 | 2021-03-23 18:30:00.000000+0000 | Marketing | Manager | Mahesh | 5e+05
40
41 (5 rows)
42
43 cqlsh:employee_space> alter table employee_info add projects set<text>;
44
45 cqlsh:employee_space> update employee_info set projects=projects+{'Web development','machine learning'} where emp_id=2;
46
47 cqlsh:employee_space> select * from employee_info;
48
49 emp_id | date_of_joining          | dept_name | designation | emp_name | projects | salary
50 -----+-----+-----+-----+-----+-----+-----
51      5 | 2009-05-23 18:30:00.000000+0000 | Administration | Administrator | Rahul | null | 2e+06
52      1 | 2022-01-23 18:30:00.000000+0000 | Development | Manager | Radha | null | 1e+05
53      2 | 2021-01-23 18:30:00.000000+0000 | Accounts | Accountant | Mahalaxmi | {'Web development', 'machine learning'} | 2e+05
54      4 | 2021-05-23 18:30:00.000000+0000 | Administration | Administrator | Nidhi | null | 5e+05
55      3 | 2021-03-23 18:30:00.000000+0000 | Marketing | Manager | Mahesh | null | 5e+05
56
57 (5 rows)
58 cqlsh:employee_space> update employee_info set projects=projects+{'Web development','machine learning','cybersecurity'} where emp_id=5;
59
60 cqlsh:employee_space> select * from employee_info;
61
62 emp_id | date_of_joining          | dept_name | designation | emp_name | projects | salary

```

```

63 emp_id | date_of_joining          | dept_name    | designation | emp_name | projects                                     | salary
64 -----+-----+-----+-----+-----+-----+-----
65      5 | 2009-05-23 18:30:00.000000+0000 | Administration | Administrator | Rahul | {'Web development', 'cybersecurity', 'machine learning'} | 2e+06
66      1 | 2022-01-23 18:30:00.000000+0000 | Development    | Manager      | Radha | null | 1e+05
67      2 | 2021-01-23 18:30:00.000000+0000 | Accounts       | Accountant   | Mahalaxmi | {'Web development', 'machine learning'} | 2e+05
68      4 | 2021-05-23 18:30:00.000000+0000 | Administration | Administrator | Nidhi | null | 5e+05
69      3 | 2021-03-23 18:30:00.000000+0000 | Marketing      | Manager      | Mahesh | null | 5e+05
70
71 (5 rows)
72
73 cqlsh:employee_space> INSERT INTO employee_space.employee_info(emp_id,emp_name,designation,date_of_joining,salary,dept_name) VALUES(6,'Harshitha','Manager','2022-01-24',100000
74
75 cqlsh:employee_space> select * from employee_info;
76
77 emp_id | date_of_joining          | dept_name    | designation | emp_name | projects                                     | salary
78 -----+-----+-----+-----+-----+-----+-----
79      5 | 2009-05-23 18:30:00.000000+0000 | Administration | Administrator | Rahul | {'Web development', 'cybersecurity', 'machine learning'} | 2e+06
80      1 | 2022-01-23 18:30:00.000000+0000 | Development    | Manager      | Radha | null | 1e+05
81      2 | 2021-01-23 18:30:00.000000+0000 | Accounts       | Accountant   | Mahalaxmi | {'Web development', 'machine learning'} | 2e+05
82      4 | 2021-05-23 18:30:00.000000+0000 | Administration | Administrator | Nidhi | null | 5e+05
83      6 | 2022-01-23 18:30:00.000000+0000 | Marketing      | Manager      | Harshitha | null | 1e+05
84      3 | 2021-03-23 18:30:00.000000+0000 | Marketing      | Manager      | Mahesh | null | 5e+05
85
86 (6 rows)
87
88 cqlsh:employee_space> select * from employee_info;
89
90 emp_id | date_of_joining          | dept_name    | designation | emp_name | projects                                     | salary
91 -----+-----+-----+-----+-----+-----+-----
92      5 | 2009-05-23 18:30:00.000000+0000 | Administration | Administrator | Rahul | {'Web development', 'cybersecurity', 'machine learning'} | 2e+06
93      1 | 2022-01-23 18:30:00.000000+0000 | Development    | Manager      | Radha | null | 1e+05
94      2 | 2021-01-23 18:30:00.000000+0000 | Accounts       | Accountant   | Mahalaxmi | {'Web development', 'machine learning'} | 2e+05
95      4 | 2021-05-23 18:30:00.000000+0000 | Administration | Administrator | Nidhi | null | 5e+05
96      3 | 2021-03-23 18:30:00.000000+0000 | Marketing      | Manager      | Mahesh | null | 5e+05
97

```

LAB 2

```
1  cqlsh> create keyspace library_space WITH REPLICATION={'class':'SimpleStrategy','replication_factor':2};
2
3  cqlsh> use library_space;
4
5  cqlsh:library_space> create table library_info(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_id));
6
7  cqlsh:library_space> update library_info set counter_value=counter_value+1 where stud_id=1 and stud_name='abc' and book_name='book1' and book_id=11 and date_of_issue='2022-01-01 18:30:00.000000+0000';
8
9  cqlsh:library_space> update library_info set counter_value=counter_value+1 where stud_id=2 and stud_name='def' and book_name='book2' and book_id=12 and date_of_issue='2022-03-01 18:30:00.000000+0000';
10
11 cqlsh:library_space> update library_info set counter_value=counter_value+1 where stud_id=3 and stud_name='ghi' and book_name='book3' and book_id=13 and date_of_issue='2022-05-01 18:30:00.000000+0000';
12
13 cqlsh:library_space> update library_info set counter_value=counter_value+1 where stud_id=4 and stud_name='jkl' and book_name='book4' and book_id=14 and date_of_issue='2022-07-01 18:30:00.000000+0000';
14
15 cqlsh:library_space> update library_info set counter_value=counter_value+1 where stud_id=5 and stud_name='mno' and book_name='book5' and book_id=15 and date_of_issue='2022-09-01 18:30:00.000000+0000';
16
17 cqlsh:library_space> select * from library_info;
18
19 stud_id | stud_name | book_name | book_id | date_of_issue | counter_value
20 -----+-----+-----+-----+-----+-----
21      5 |      mno |      book5 |      15 | 2022-09-29 18:30:00.000000+0000 |          1
22      1 |      abc |      book1 |      11 | 2022-01-29 18:30:00.000000+0000 |          1
23      2 |      def |      book2 |      12 | 2022-03-29 18:30:00.000000+0000 |          1
24      4 |      jkl |      book4 |      14 | 2022-07-29 18:30:00.000000+0000 |          1
25      3 |      ghi |      book3 |      13 | 2022-05-29 18:30:00.000000+0000 |          1
26
27 (5 rows)
28
29 cqlsh:library_space> update library_info set counter_value=counter_value+1 where stud_id=5 and stud_name='mno' and book_name='book5' and book_id=15 and date_of_issue='2022-09-01 18:30:00.000000+0000';
30
31 cqlsh:library_space> select * from library_info;
```

```
33 stud_id | stud_name | book_name | book_id | date_of_issue | counter_value
34 -----+-----+-----+-----+-----+-----
35      5 |      mno |    book5 |      15 | 2022-09-29 18:30:00.000000+0000 |          2
36      1 |      abc |    book1 |      11 | 2022-01-29 18:30:00.000000+0000 |          1
37      2 |      def |    book2 |      12 | 2022-03-29 18:30:00.000000+0000 |          1
38      4 |      jkl |    book4 |      14 | 2022-07-29 18:30:00.000000+0000 |          1
39      3 |      ghi |    book3 |      13 | 2022-05-29 18:30:00.000000+0000 |          1
40
41 (5 rows)
42
43 cqlsh:library_space> copy library_info(stud_id,stud_name,book_name,book_id,date_of_issue,counter_value) to '/home/bmscecse/Desktop/bda.csv';
44
45 Using 11 child processes
46
47 Starting copy of library_space.library_info with columns [stud_id, stud_name, book_name, book_id, date_of_issue, counter_value].
48 Processed: 5 rows; Rate:      45 rows/s; Avg. rate:      45 rows/s
49 5 rows exported to 1 files in 0.121 seconds.
50
51 cqlsh:library_space> create table library_info_copy(stud_id int,counter_value counter,stud_name text,book_name text,book_id int,date_of_issue timestamp,PRIMARY KEY(stud_id,stu
52
53 cqlsh:library_space> copy library_info_copy(stud_id,stud_name,book_name,book_id,date_of_issue,counter_value) from '/home/bmscecse/Desktop/new.csv';
54
55 Using 11 child processes
56
57 Starting copy of library_space.library_info_copy with columns [stud_id, stud_name, book_name, book_id, date_of_issue, counter_value].
58 Processed: 5 rows; Rate:      8 rows/s; Avg. rate:     12 rows/s
59 5 rows imported from 1 files in 0.406 seconds (0 skipped).
60
61 cqlsh:library_space> select * from library_info where counter_value=2 allow filtering;
62
63 stud_id | stud_name | book_name | book_id | date_of_issue | counter_value
64 -----+-----+-----+-----+-----+-----
65      5 |      mno |    book5 |      15 | 2022-09-29 18:30:00.000000+0000 |          2
```