

Big Data Systems

2021WI_CS_6545_FR01B

BDS_Handson2

Prepared by:
Mohammadali Rahnama

Student number:
3709515



TABLE OF CONTENTS

TABLE OF CONTENTS	1
LIST OF FIGURES.....	1
LIST OF TABLES.....	1
Task1.....	2
Schema of the table mytable1 (output of show table mytable1)	2
Output of the query: SELECT * FROM mytable1	3
Schema of the table mytable2 (output of show table mytable2)	4
Output of the query: SELECT * FROM mytable2	5
Task2.....	6
Average execution time of the warm runs of Q3, Q4 and Q14 with Stado	6
Average execution time of the warm runs of Q3, Q4 and Q14 with single instance PostgreSQL	8
Speedup of Q3, Q4 and Q14 achieved with Stado against single instance PostgreSQL	10
REFERENCES	11

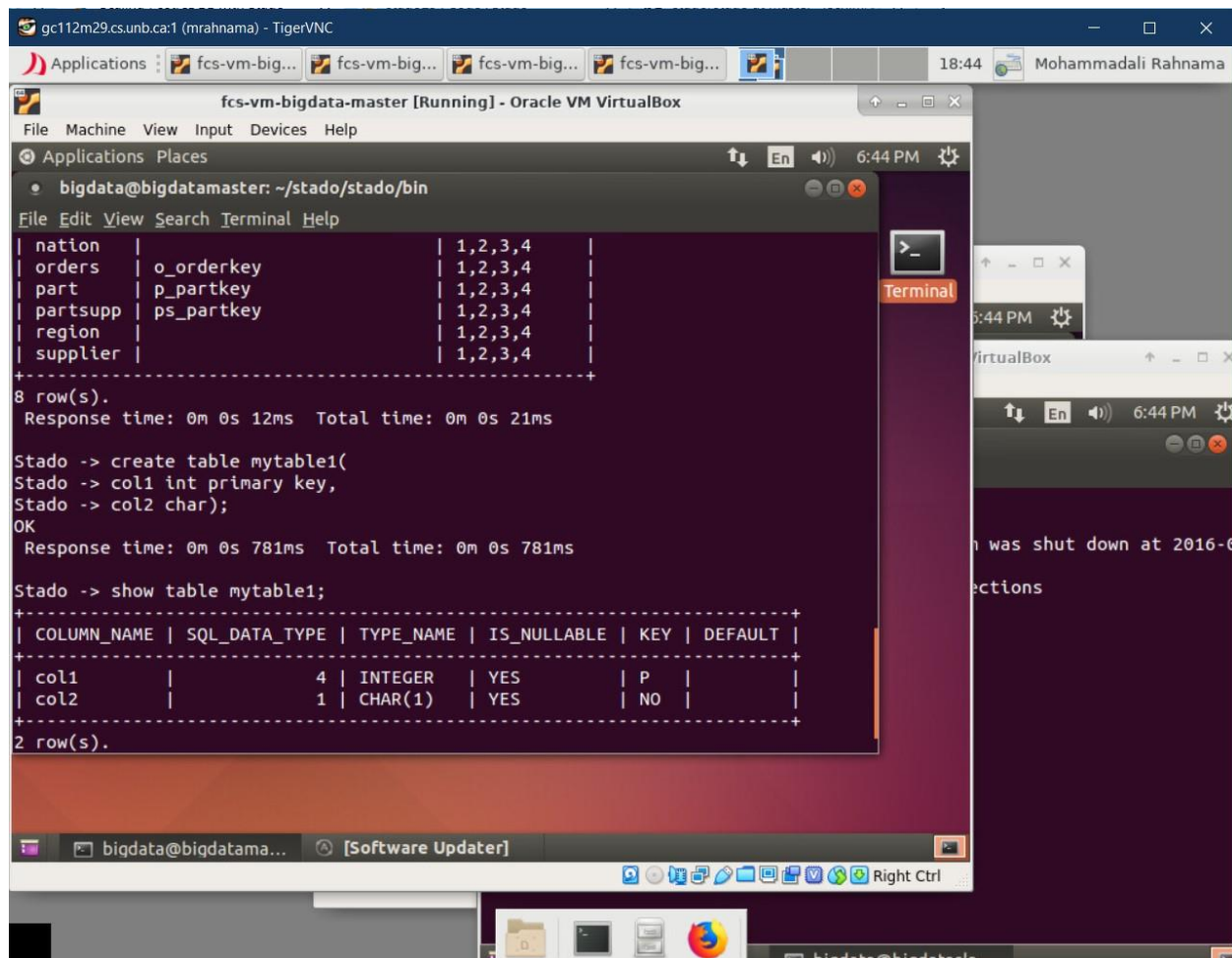
LIST OF FIGURES

Figure 1: SHOW TABLE mytable1.	2
Figure 2: SELECT* FROM mytable1	3
Figure 3: SHOW TABLE mytable2.	4
Figure 4: SELECT* FROM mytable2.....	5
Figure 5: Q4 executed with Stado.	7
Figure 6: Q14 executed with PostgreSQL	9

LIST OF TABLES

Task1

Schema of the table mytable1 (output of show table mytable1)



The screenshot shows a terminal window titled "fcs-vm-bigdata-master [Running] - Oracle VM VirtualBox". The terminal output displays the schema of table mytable1. The first table shows data rows, and the second table shows the table's structure.

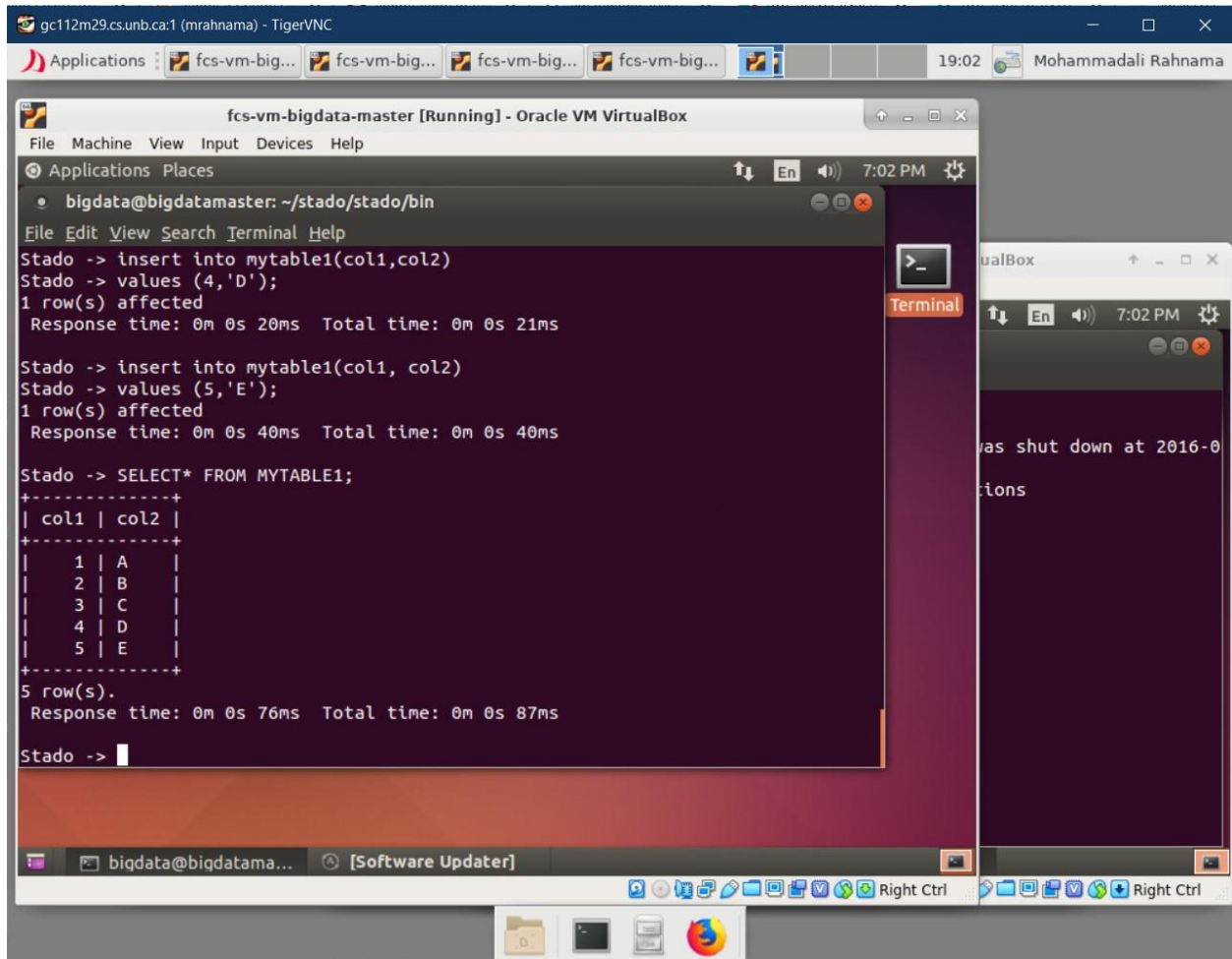
```
bigdata@bigdatamaster: ~/stado/stado/bin
File Edit View Search Terminal Help
+-----+
| nation | o_orderkey | 1,2,3,4 |
| orders | o_orderkey | 1,2,3,4 |
| part    | p_partkey  | 1,2,3,4 |
| partsupp| ps_partkey | 1,2,3,4 |
| region  |            | 1,2,3,4 |
| supplier|            | 1,2,3,4 |
+-----+
8 row(s).
Response time: 0m 0s 12ms Total time: 0m 0s 21ms

Stado -> create table mytable1(
Stado -> col1 int primary key,
Stado -> col2 char);
OK
Response time: 0m 0s 781ms Total time: 0m 0s 781ms

Stado -> show table mytable1;
+-----+
| COLUMN_NAME | SQL_DATA_TYPE | TYPE_NAME | IS_NULLABLE | KEY | DEFAULT |
+-----+
| col1        | 4             | INTEGER   | YES         | P   |         |
| col2        | 1             | CHAR(1)   | YES         | NO  |         |
+-----+
2 row(s).
```

Figure 1: SHOW TABLE mytable1

Output of the query: SELECT * FROM mytable1



```
gc112m29.cs.unb.ca:1 (mrahnama) - TigerVNC
Applications fcs-vm-big... fcs-vm-big... fcs-vm-big... fcs-vm-big... 19:02 Mohammadali Rahnama

fcs-vm-bigdata-master [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places
  • bigdata@bigdatamaster: ~/stado/stado/bin
File Edit View Search Terminal Help
Stado -> insert into mytable1(col1,col2)
Stado -> values (4,'D');
1 row(s) affected
Response time: 0m 0s 20ms Total time: 0m 0s 21ms

Stado -> insert into mytable1(col1, col2)
Stado -> values (5,'E');
1 row(s) affected
Response time: 0m 0s 40ms Total time: 0m 0s 40ms

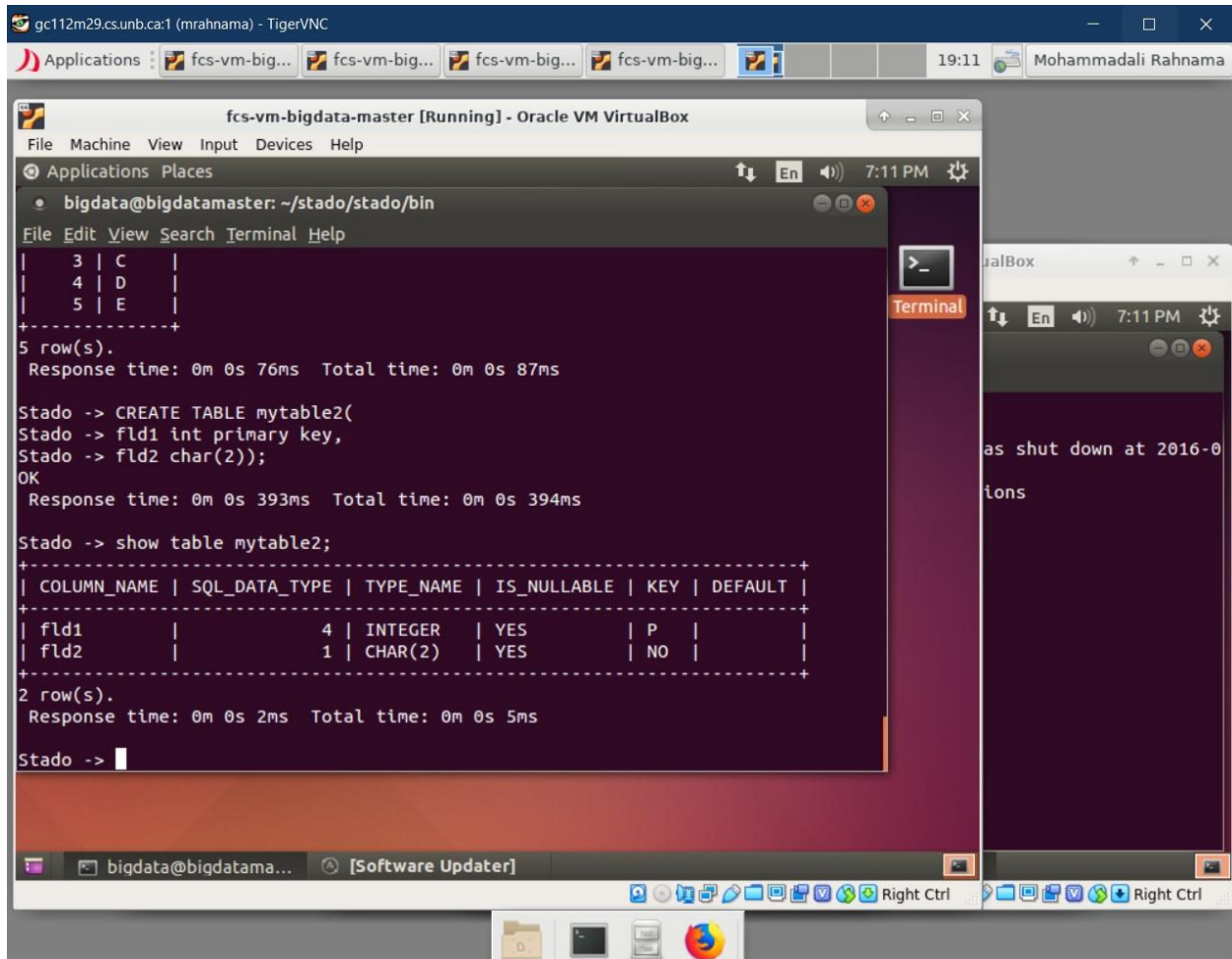
Stado -> SELECT* FROM MYTABLE1;
+-----+
| col1 | col2 |
+-----+
| 1 | A |
| 2 | B |
| 3 | C |
| 4 | D |
| 5 | E |
+-----+
5 row(s).
Response time: 0m 0s 76ms Total time: 0m 0s 87ms

Stado ->

[Software Updater]
```

Figure 2: SELECT* FROM mytable1

Schema of the table mytable2 (output of show table mytable2)



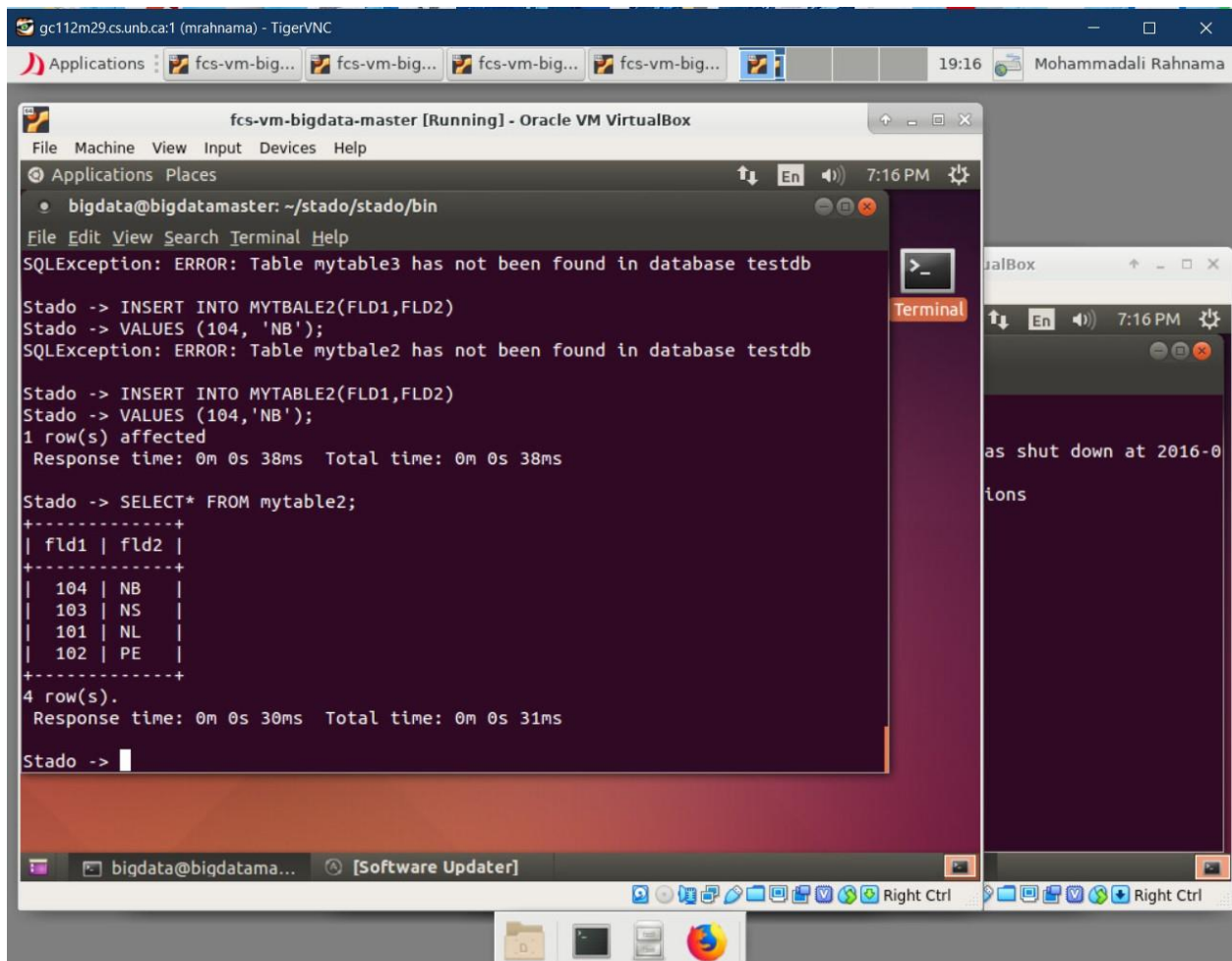
The screenshot shows a terminal window titled "fcs-vm-bigdata-master [Running] - Oracle VM VirtualBox". The terminal is running a Stado database system. The user has entered the command "show table mytable2;", and the output displays the table's schema. The schema is as follows:

COLUMN_NAME	SQL_DATA_TYPE	TYPE_NAME	IS_NULLABLE	KEY	DEFAULT
fld1	4	INTEGER	YES	P	
fld2	1	CHAR(2)	YES	NO	

The terminal also shows the creation of the table "mytable2" with fields "fld1" (integer, primary key) and "fld2" (char(2)).

Figure 3: SHOW TABLE mytable2

Output of the query: SELECT * FROM mytable2



The screenshot shows a TigerVNC window titled "gc112m29.cs.unb.ca:1 (mrahnama) - TigerVNC". The main application is "fcs-vm-bigdata-master [Running] - Oracle VM VirtualBox". The terminal window shows the following commands and output:

```
bigdata@bigdatamaster: ~/stado/stado/bin
File Edit View Search Terminal Help
SQLException: ERROR: Table mytable3 has not been found in database testdb

Stado -> INSERT INTO MYTBAL2(FLD1,FLD2)
Stado -> VALUES (104, 'NB');
SQLException: ERROR: Table mytbale2 has not been found in database testdb

Stado -> INSERT INTO MYTABLE2(FLD1,FLD2)
Stado -> VALUES (104, 'NB');
1 row(s) affected
Response time: 0m 0s 38ms Total time: 0m 0s 38ms

Stado -> SELECT* FROM mytable2;
+-----+
| fld1 | fld2 |
+-----+
| 104 | NB |
| 103 | NS |
| 101 | NL |
| 102 | PE |
+-----+
4 row(s).
Response time: 0m 0s 30ms Total time: 0m 0s 31ms

Stado -> 
```

The terminal output shows a table with 4 rows and 2 columns: fld1 and fld2. The rows are (104, NB), (103, NS), (101, NL), and (102, PE).

Figure 4: SELECT* FROM mytable2

Task2

Average execution time of the warm runs of Q3, Q4 and Q14 with Stado

Runtime results:

Q3:

1st:

Response time: 0m 8s 22ms Total time: 0m 10s 710ms

2nd:

Response time: 0m 7s 776ms Total time: 0m 10s 505ms

3rd:

Response time: 0m 7s 196ms Total time: 0m 10s 3ms

4th:

Response time: 0m 6s 730ms Total time: 0m 9s 391ms

Average execution time: $(10505 + 10003 + 9391) / 3 = 9966.33$

Q4:

1st:

Response time: 0m 4s 40ms Total time: 0m 4s 41ms

2nd:

Response time: 0m 4s 48ms Total time: 0m 4s 48ms

3rd:

Response time: 0m 3s 705ms Total time: 0m 3s 705ms

4th:

Response time: 0m 3s 708ms Total time: 0m 3s 712ms

Average execution time: $(4048 + 3705 + 3712) / 3 = 3821.66$

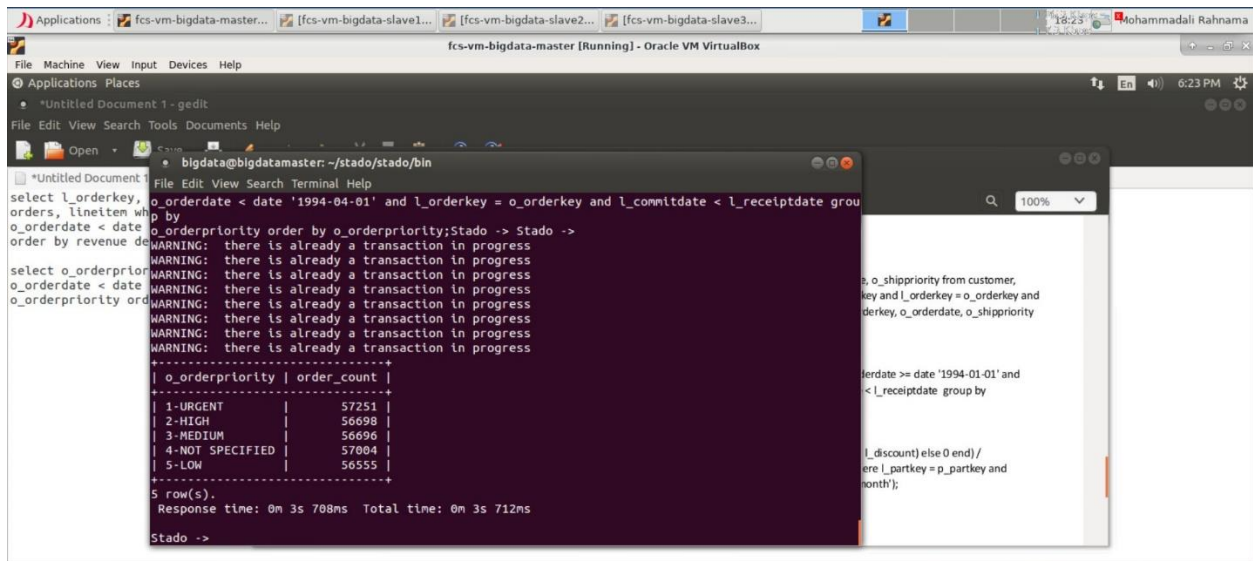


Figure 5: Q4 executed with Stado

Q14:

1st:

Response time: 0m 5s 793ms Total time: 0m 5s 793ms

2nd:

Response time: 0m 2s 323ms Total time: 0m 2s 324ms

3rd:

Response time: 0m 1s 978ms Total time: 0m 1s 978ms

4th:

Response time: 0m 2s 43ms Total time: 0m 2s 45ms

Average execution time: $(2324 + 1978 + 2045) / 3 = 2115.66$

Average execution time of the warm runs of Q3, Q4 and Q14 with single instance PostgreSQL

Runtime results:

Q3:

1st:

Time: 53113.551 ms

2nd:

Time: 18278.821 ms

3rd:

Time: 17244.031 ms

4th:

Time: 18532.079 ms

Average execution time: $(18278.821 + 17244.031 + 18532.079) / 3 = 18018.31$

Q4:

1st:

Time: 15643.706 ms

2nd:

Time: 14743.495 ms

3rd:

Time: 16214.004 ms

4th:

Time: 15437.399 ms

Average execution time: $(14743.495 + 16214.004 + 15437.399) / 3 = 15464.96$

Q14:

1st:

Time: 11093.017 ms

2nd:

Time: 10963.869 ms

3rd:

Time: 10996.896 ms

4th:

Time: 11114.471 ms

Average execution time: $(10963.869 + 10996.896 + 11114.471) / 3 = 11025.07$

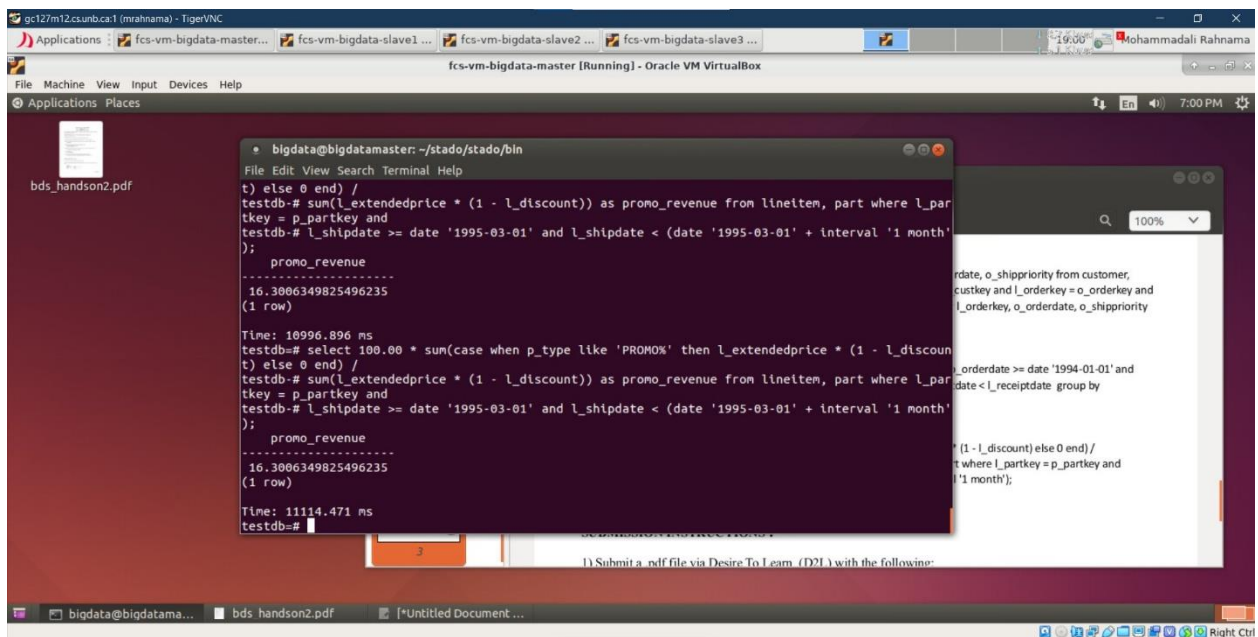


Figure 6: Q14 executed with PostgreSQL

Speedup of Q3, Q4 and Q14 achieved with Stado against single instance PostgreSQL

$$\text{Speedup} = \frac{T_{\text{Serial}}}{T_{\text{Parallel}}} [2]$$

Q3:

Stado Average execution time: 9966.33

PostgreSQL Average execution time: 18018.31

Speedup: $(18018.31/9966.33) = 1.80$

Q4:

Stado Average execution time: 3821.66

PostgreSQL Average execution time: 15464.96

Speedup: $(15464.96/3821.66) = 4.04$

Q14:

Stado Average execution time: 2115.66

PostgreSQL Average execution time: 11025.07

Speedup: $(11025.07/2115.66) = 5.21$

REFERENCES

- [1] Unknown. (2021). Big Data Systems - CS4545/CS6545 Hands-on 2.
- [2] Dr. Suprio Ray. (2021). Parallel Databases. In Big Data Systems (CS4545/CS6545).