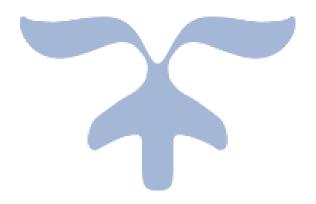


HANDSON 2

Big Data Systems

Soheil Shirvani - 3720505



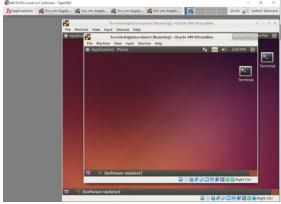
Contents

Lunching Stado Parallel database	2
Steps for lunching Stado Parallel:	2
Task 1	4
1.a Create a partitioned table partab	4
1.b Insert the tuples (with given values) into partab	5
2.a Create a replicated table reptab.	6
2.b Insert the tuples (with given values) into reptab	7
Results Only:	8
Schema of your table partab (output of: show table partab)	8
Output of the query: SELECT * FROM partab	8
Schema of your table reptab (output of: show table reptab)	8
Output of the query: SELECT * FROM reptab	8
Task 2	9
Run 3 query on Stado as parallel database	9
Query 1:	9
Query 3:	10
Query 6:	11
Run 3 query on PostgreSQL as Single database.	12
Query 1:	12
Query 3:	13
Query 6:	14
Results Only	15
Average execution time of the warm runs of Q1, Q3 and Q6 with Stado	15
Average execution time of the warm runs of Q1, Q3 and Q6 with Single DB	15
Speedup of Q1, Q3 and Q6 achieved with Stado against single instance PostgreSQL	15

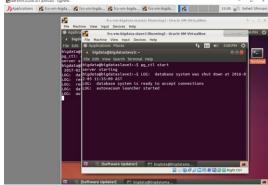
Lunching Stado Parallel database

Steps for lunching Stado Parallel:

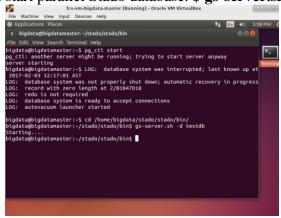
1. Here first we login into 4 different VMs. 1 is master VM and 3 Slave VMs. To do so we open all the VMs and login.



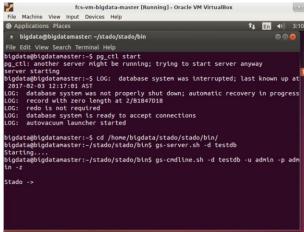
2. Start single PostgreSQL database on all the VMs. \$ pg_ctl start



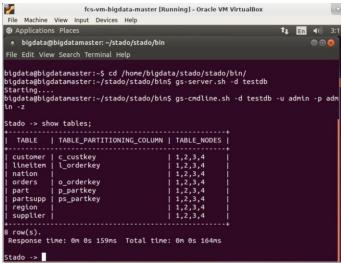
3. On master VM we will start parallel Stado database. \$ gs-server.sh -d testdb



4. We will start Stado SQL client to connect to the parallel database. **\$ gs-cmdline.sh -d testdb -u admin -p admin -z**



5. Then we are abale to see the tables in the database or create and insert in it.



Task 1

1.a Create a partitioned table partab

To Create a partitioned table named partab with 2 column including (col1, Integer) and (col2, Char) and partition key of (col1) we use the query below:

CREATE TABLE partab (col1 INTEGER, col2 CHAR(15)) PARTITIONING KEY col1 ON ALL; And response is:

```
Stado -> CREATE TABLE partab (col1 INTEGER, col CHAR(15)) PARTITIONING KEY col1
ON ALL;
OK
Response time: 0m 1s 275ms Total time: 0m 1s 276ms
```

So the table is created. And Show tables here include the new table:

```
Stado -> show tables;
           | TABLE_PARTITIONING_COLUMN | TABLE_NODES
   TABLE
 customer | c custkey
                                          1,2,3,4
 lineitem | l orderkey
                                          1,2,3,4
 nation
                                          1,2,3,4
 orders
           o orderkey
                                          1,2,3,4
             p partkey
 part
                                          1,2,3,4
 partab
           | col1
             ps partkey
 partsupp |
                                          1,2,3,4
 region
                                          1,2,3,4
 supplier |
                                          1,2,3,4
9 row(s).
 Response time: Om Os 5ms Total time: Om Os 7ms
```

1.b Insert the tuples (with given values) into partab

We will Insert 6 predefined tuples into our partab table that we created before. We use the command Below:

INSERT INTO partab VALUES ('2','I'), ...

```
• bigdata@bigdatamaster: ~/stado/stado/bin
                                                                                   00
File Edit View Search Terminal Help
1 row(s) affected
Response time: Om Os 7ms Total time: Om Os 7ms
Stado -> INSERT INTO partab VALUES (3,'B');
 row(s) affected
Response time: Om Os 5ms Total time: Om Os 5ms
Stado -> INSERT INTO partab VALUES (4,'G');
1 row(s) affected
Response time: 0m 0s 7ms Total time: 0m 0s 7ms
Stado -> INSERT INTO partab VALUES (5,'Y');
1 row(s) affected
Response time: Om Os 8ms Total time: Om Os 8ms
Stado -> INSERT INTO partab VALUES (6,'0');
 row(s) affected
 Response time: Om Os 4ms Total time: Om Os 4ms
Stado -> INSERT INTO partab VALUES (7,'R');
 row(s) affected
 Response time: Om Os 5ms Total time: Om Os 5ms
```

After Inserting all the tuples we can query our table and check if the tuples are in the table using the command below and see the result:

SELECT * FROM partab

Table Actually have all the rows (7,'R') is not displayed and it was inserted twice. So all the insertion worked perfectly fine.

2.a Create a replicated table reptab.

To Create a replicated table named reptab with 2 column including (fld1, Integer) and (fld2, Char(2)) we use the query below:

CREATE TABLE reptab (fld1 INTEGER, fld2 CHAR(2)) REPLICATED;

And response is:

```
Stado -> CREATE TABLE reptab (fld1 INTEGER, fld2 CHAR(2)) REPLICATED;
OK
Response time: Om Os 106ms Total time: Om Os 106ms
```

So the table is created. And Show tables here include the new table:

```
Stado -> show tables;
  TABLE | TABLE PARTITIONING COLUMN | TABLE NODES
 customer | c custkey
                                       1,2,3,4
 lineitem | l_orderkey
                                         1,2,3,4
                                         1,2,3,4
 nation
           o orderkey
 orders
                                         1,2,3,4
           | p partkey
                                         1,2,3,4
 part
 partab
           | col1
                                         1,2,3,4
 partsupp | ps partkey
                                         1,2,3,4
 region
                                         1,2,3,4
 reptab
                                         1,2,3,4
 supplier
                                         1,2,3,4
10 row(s).
Response time: Om Os 16ms Total time: Om Os 17ms
```

2.b Insert the tuples (with given values) into reptab

We will Insert 6 predefined tuples into our reptab table that we created before. We use the command Below:

INSERT INTO reptab VALUES ('2', 'I'), ...

```
Stado -> INSERT INTO reptab VALUES (7,'NB');

1 row(s) affected
Response time: 0m 0s 18ms Total time: 0m 0s 19ms

Stado -> INSERT INTO reptab VALUES (8,'NL');

1 row(s) affected
Response time: 0m 0s 12ms Total time: 0m 0s 12ms

Stado -> INSERT INTO reptab VALUES (9,'NS');

1 row(s) affected
Response time: 0m 0s 13ms Total time: 0m 0s 15ms

Stado -> INSERT INTO reptab VALUES (10, 'PE');

1 row(s) affected
Response time: 0m 0s 11ms Total time: 0m 0s 11ms

Stado ->
```

After Inserting all the tuples we can query our table and check if the tuples are in the table using the command below and see the result:

SELECT * FROM reptab

As we can see all the tuples were inserted into the reptab successfully.

Results Only:

1. Partitioned

CREATE TABLE partab (col1 INTEGER, col2 CHAR) PARTITIONING KEY col1 ON ALL;

Schema of your table partab (output of: show table partab)

Output of the query: SELECT * FROM partab

2. Replicated

Schema of your table reptab (output of: show table reptab)

Output of the query: SELECT * FROM reptab

Task 2

Here we are going to run 3 queries from TPC-H benchmark and find the response time of database in 2 modes Parallel and Single. I ignored the first run of each query (cold) and made a note of the response time of 2 next query execution for both modes.

Run 3 query on Stado as parallel database.

Query 1: After running 1 time (cold run) the Second result is: **Time: 13s 689ms**

```
| Stado -> select l_returnflag, l_linestatus, sum(l_quantity) as sum_qty, sum(l_extendedprice) as sum_base_price, sum(l_extendedprice * (1-l_discount)) as sum_disc_price, sum(l_extendedprice * (1-l_discount) as sum_disc_price, sum(l_extendedprice * (1-l_discount)) as sum_disc_price, sum(l_extendedprice * (1-l_discount) as sum_disc_price, sum(l_extendedprice) as avg_price, avg(l_discount) as avg_disc, count(*) as count_order from linetlene where l_shipdate <= (date '1998-12-01' - interval '110 days') group by l_returnflag, l_linestatus order by l_returnflag, l_linestatus; waRNING: there is already a transaction in progress warning: warning warning
```

Third run is: Time: 13s 961ms

```
| Stado -> select L_returnflag, L_tinestatus, sun(l_quantity) as sun_qty, sun(l_extendedprice) as sun_base_price, sun(l_extendedprice * (1-l_discount)) as sun_disc_price, sun(l_extendedprice) as sun_price, avg(l_discount) as sun_disc_price, sun(l_extendedprice) as sun_price, avg(l_discount) as sun_disc_price, sun_disc_price | sun
```

Query 3: After running 1 time (cold run) the Second result is: **Time: 10s 451ms**

Stado -> SELECT L_ORDERKEY, SUM(L_EXTENDEDPRICE*(1-L_DISCOUNT)) AS REVENUE, O_ORDERDATE, O_SHIPPRIORITY FROM CUSTOMER, ORDERS, LINEITEM WHERE C_MKTSEGMENT IN ('AUTOMOBIL E') AND C_CUSTKEY = O_CUSTKEY AND L_ORDERKEY = O_ORDERKEY AND O_ORDERDATE < date '1995-03-19' AND L_SHIPDATE > date '1995-03-19' GROUP BY L_ORDERKEY, O_ORDERDATE, O_SHIP PRIORITY ORDER BY REVENUE DESC, O_ORDERDATE;

3585	9800 1095	.4/52	1994-12-11	<u> </u>	0	
4302	2085 1091	.9898	1995-02-22	2	0	
3452	2033 1076	.0937	1994-12-01	l I	0	
7557	7798 1066	.9282	1995-01-28	3	0	
9988	3964 1066	.7136	1995-01-01	1 1	0	
442	2437 1059	.9072	1994-12-01	l 1	0	
6450	6420 1059	.9072	1995-02-19)	0	
8086	0036 1050	.6652	1995-03-12	·	0	
3445	601 1044	.6432	1995-01-06	5	0	
6834	311 1037	.9440	1995-03-06	5	0	
10804	1000 1029	.6891	1995-01-26)	0	
5012	2578 989	.7888	1994-12-18	3	0	
7674	1112 977	.7034	1995-02-02	<u> </u>	0	
5613	3284 963	.3096	1994-12-36)	0	
4182	2179 924	.4992	1994-11-25	5	0	
7676	915	.8496	1995-02-27	7	0	
2204	743 906	.7872	1995-03-15	5	0	
7231	1653 900	.9637	1994-12-21	l I	0	
9097	7634 883	.9502	1995-01-14	!	0	
9418	3727 873	.9905	1994-12-01		0	
22451 rov		724	T-4-1 44	0- 40- 454		
Response	time: 0m 7s	/34MS	Total time	e: 0m 10s 451ms		

Third run is: Time: 10s 459ms

	9988964	1066.7136	1995-01-01	0
	442437	1059.9072	1994-12-01	0
	6456420	1059.9072	1995-02-19	0
	8080036	1050.6652	1995-03-12	0
	3445601	1044.6432	1995-01-06	0
	6834311	1037.9440	1995-03-06	0
	10804000	1029.6891	1995-01-20	0
	5012578	989.7888	1994-12-18	0
	7674112	977.7034	1995-02-02	0
	5613284	963.3096	1994-12-30	0
	4182179	924.4992	1994-11-25	0
	7670117	915.8496	1995-02-27	0
	2204743	906.7872	1995-03-15	0
	7231653	900.9637	1994-12-21	0
	9097634	883.9502	1995-01-14	0
	9418727	873.9905	1994-12-01	0
24	151 row(s).			
		Om 7s 815ms	Total time: 0m 10s	459ms

Query 6:

After running 1 time (cold run) the Second result is: Time: 2s 219ms

Stado -> SELECT SUM(L_EXTENDEDPRICE*L_DISCOUNT) AS REVENUE FROM LINEITEM WHERE L_SHIPDATE >= date '1995-01-01' AND L_SHIPDATE < (date '1995-01-01' + interval '1 year') A
ND L_DISCOUNT BETWEEN .04 - 0.01 AND .04 + 0.01 AND L_QUANTITY < 24;

Third run is: Time: 1s 909ms

Run 3 query on PostgreSQL as Single database.

First we stop the Stado and go to PostgreSQL command line to do the queries.

```
Stado -> ^C
bigdata@bigdatamaster:-/stado/stado/bin$ cd
bigdata@bigdatamaster:-$ psql testdb
psql (9.4.0)
Type "help" for help.

testdb=#
```

Query 1:

After running 1 time (cold run) the Second result is: Time: 40s 248ms

testdb=# select l_returnflag, l_linestatus, sum(l_quantity) as sum_qty, sum(l_extendedprice) as sum_base_price, sum(l_extendedprice * (1-l_discount)) as sum_disc_price, sum(l_extendedprice) as avg_price, avg(l_discount) as avg_disc, count(*) as avg_qty, avg(l_extendedprice) as avg_price, avg(l_discount) as avg_disc, count(*) as count_order from lineitem where l_shipdate <= (date '1998-12-01' - interval '110 days') group by l_returnflag, l_linestatus;

Third run is: **Time: 40s 788ms**

Query 3: After running 1 time (cold run) the Second result is: **Time: 16s 948ms**

testdb=# SELECT L_ORDERKEY, SUM(L_EXTENDEDPRICE*(1-L_DISCOUNT)) AS REVENUE, O_ORDERDATE, O_SHIPPRIORITY FROM CUSTOMER, ORDERS, LINEITEM WHERE C_MKTSEGMENT IN ('AUTOMOBIL E') AND C_CUSTKEY = O_CUSTKEY AND L_ORDERKEY = O_ORDERKEY AND O_ORDERDATE < date '1995-03-19' AND L_SHIPDATE > date '1995-03-19' GROUP BY L_ORDERKEY, O_ORDERDATE, O_SHIP PRIORITY ORDER BY REVENUE DESC, O_ORDERDATE;

******		1		Č.
1824643	380013.3687	1995-03-06	[)
1088995	375667.8475	1995-03-11	[)
6669831	375492.0278	1995-02-07	6)
4710272	375188.6517	1995-03-17	6)
3778628	373654.2087	1995-02-25	()
4736678	372807.1864	1995-03-03	[)
1062786	372672.5509	1995-02-10	6)
6680197	369266.7219	1995-02-13	6)
10350439	369187.7220	1995-02-16	1 6)
8257025	368258.2766	1995-03-16	1 6)
4496835	367877.7452	1995-03-18	6)
7205063	366929.7907	1995-02-26	6)
1102534	366089.2778	1995-03-15	6)
7166048	365379.9582	1995-03-02	6)
Time: 16948.2	201 ms			

Third run is: Time: 16s 187ms

4710272	375188.6517	1995-03-17	1	0
3778628	373654.2087	1995-02-25	1	0
4736678	372807.1864	1995-03-03		0
1062786	372672.5509	1995-02-10	1	0
6680197	369266.7219	1995-02-13	1	0
10350439	369187.7220	1995-02-16	1	0
8257025	368258.2766	1995-03-16	1	0
4496835	367877.7452	1995-03-18	1	0
7205063	366929.7907	1995-02-26	1	0
1102534	366089.2778	1995-03-15	1	0
7166048	365379.9582	1995-03-02	1	0
Time: 16187.	535 ms			

Query 6:

After running 1 time (cold run) the Second result is: Time: 32s 14ms

testdb=# SELECT SUM(L_EXTENDEDPRICE*L_DISCOUNT) AS REVENUE FROM LINEITEM WHERE L_SHIPDATE >= date '1995-01-01' AND L_SHIPDATE < (date '1995-01-01' + interval '1 year') A
ND L_DISCOUNT BETWEEN .04 - 0.01 AND .04 + 0.01 AND L_QUANTITY < 24;

Third run is: Time: 31s 888ms

```
revenue
164347732.3953
(1 row)
Time: 31888.667 ms
```

Results Only

Average execution time of the warm runs of Q1, Q3 and Q6 with Stado

 $Q1 \left\{ \begin{matrix} warn \ run \ 1: \ 13s \ 689ms: 13689ms \\ warn \ run \ 2: \ 13s \ 961ms: 13961ms \end{matrix} \right. \rightarrow Avg = 13825ms$

 $Q3 \begin{tabular}{l} \textit{Warn run 1}: 10s\ 451ms: 10451ms \\ \textit{Warn run 2}: 10s\ 459ms: 10459ms \\ \end{tabular} \rightarrow \textit{Avg} = 10455ms$

 $Q6 \begin{cases} warn \, run \, 1: \, 2s \, 219ms: 2219ms \\ warn \, run \, 2: \, 1s \, 909ms: 1909ms \end{cases} \rightarrow Avg = 2064ms$

Average execution time of the warm runs of Q1, Q3 and Q6 with Single DB

 $Q1 \left\{ \begin{matrix} warn \ run \ 1: \ 40s \ 248ms: 40248ms \\ warn \ run \ 2: \ 40s \ 788ms: 40788ms \end{matrix} \right. \rightarrow Avg = 40518ms$

Q3 $\begin{cases} warn \, run \, 1 : \, 16s \, 948ms : 16948ms \\ warn \, run \, 2 : \, 16s \, 187ms : 16187ms \end{cases} \rightarrow Avg = 16568ms$

 $Q6 \begin{cases} warn \, run \, 1: \, 32s \, 14ms: 32014ms \\ warn \, run \, 2: \, 31s \, 888ms: 31888ms \end{cases} \rightarrow Avg = 31951ms$

Speedup of Q1, Q3 and Q6 achieved with Stado against single instance PostgreSQL

$$Q1 \, Speedup = \frac{40518}{13825} = 2.93$$

$$Q3 Speedup = \frac{16568}{10455} = 1.584$$

$$\textit{Q6 Speedup} = \frac{31951}{2064} = 15.48$$