

PG_DWH TASK 6

Legal Notice:

This document contains privileged and/or confidential information and may not be disclosed, distributed or reproduced without the prior written permission of EPAM®.

```
Query Query History

1  CREATE TABLE SALES_INFO
2  id INTEGER,
4  category VARCHAR(1),
5  ischeck BOOLEAN,
6  eventdate DATE
7 );

Data Output Messages Notifications

CREATE TABLE

Query returned successfully in 31 msec.
```

Table created, table partitioned by eventdate.

```
Query Query History
 id INTEGER,
     category VARCHAR(1),
     ischeck BOOLEAN,
     eventdate DATE
     ) PARTITION BY RANGE (eventdate);
     CREATE TABLE sales_info_2021 PARTITION OF sales_info
     CREATE TABLE sales_info_2022 PARTITION OF sales_info
         FOR VALUES FROM ('2022-01-01') TO ('2023-01-01');
     CREATE TABLE sales_info_2023 PARTITION OF sales_info
FOR VALUES FROM ('2023-01-01') TO ('2024-01-01');
16 ~
     CREATE TABLE sales_info_2024 PARTITION OF sales_info
     CREATE TABLE sales_info_2025 PARTITION OF sales_info
Data Output Messages Notifications
CREATE TABLE
Query returned successfully in 47 msec.
```

```
CREATE OR REPLACE FUNCTION partition_sales_info() RETURNS trigger

A $ $ $
BEGIN

IF (new.eventdate >= '2022-01-01'::DATE AND new.eventdate < '2023-01-01'::DATE) THEN

INSERT INTO sales_info_2022 VALUES (new.*);

ELSIF (new.eventdate >= '2021-01-01'::DATE AND new.eventdate < '2022-01-01'::DATE) THEN

INSERT INTO sales_info_2021 VALUES (new.*);

VELSIF (new.eventdate >= '2023-01-01'::DATE AND new.eventdate < '2024-01-01'::DATE) THEN

INSERT INTO sales_info_2023 VALUES (new.*);

INSERT INTO sales_info_2024 VALUES (new.*);

ELSIF (new.eventdate >= '2023-01-01'::DATE AND new.eventdate < '2025-01-01'::DATE) THEN

INSERT INTO sales_info_2024 VALUES (new.*);

ELSIF (new.eventdate >= '2023-01-01'::DATE AND new.eventdate < '2026-01-01'::DATE) THEN

INSERT INTO sales_info_2025 VALUES (new.*);

ELSIF (new.eventdate >= '2023-01-01'::DATE AND new.eventdate < '2026-01-01'::DATE) THEN

INSERT INTO sales_info_2025 VALUES (new.*);

ELSIF (new.eventdate >= '2023-01-01'::DATE AND new.eventdate < '2026-01-01'::DATE) THEN

INSERT INTO sales_info_2025 VALUES (new.*);

ELSIF (new.eventdate >= '2023-01-01'::DATE AND new.eventdate < '2026-01-01'::DATE) THEN

INSERT INTO sales_info_2025 VALUES (new.*);

ELSIF (new.eventdate >= '2023-01-01'::DATE AND new.eventdate < '2026-01-01'::DATE) THEN

INSERT INTO sales_info_2025 VALUES (new.*);

ELSIF (new.eventdate >= '2023-01-01'::DATE AND new.eventdate < '2026-01-01'::DATE) THEN

INSERT INTO sales_info_2025 VALUES (new.*);

ELSIF (new.eventdate >= '2023-01-01'::DATE AND new.eventdate < '2026-01-01'::DATE) THEN

INSERT INTO sales_info_2025 VALUES (new.*);

ELSIF (new.eventdate >= '2023-01-01'::DATE AND new.eventdate < '2026-01-01'::DATE) THEN

INSERT INTO sales_info_2025 VALUES (new.*);

ELSIF (new.eventdate >= '2023-01-01'::DATE AND new.eventdate < '2026-01-01'::DATE) THEN

INSERT INTO sales_info_2025 VALUES (new.*);

ELSIF (new.eventdate >= '2023-01-01'::DATE AND new.eventdate < '2026-01-01'::DATE AND new.eventdate < '2026-01-01'::DATE AND new.eventdate < '2026-01-01'::DATE AND new.eventdate <
```

Function created for partitioning.

Trigger to use the function partition_sales_info on sales_info table , it will be called before inserting row in the table.

```
50
51 V CREATE TRIGGER partition_sales_info_trigger
52 BEFORE INSERT ON sales_info
53
54 FOR EACH ROW
55 WHEN (pg_trigger_depth() < 1)
56 EXECUTE PROCEDURE partition_sales_info();
57
58
```

I got an error before used WHEN condition , it was inserting the same row over and over , idk why but I found solution by using WHEN condition here .

Then inserted all rows.

```
SET max_stack_depth = 7680

SET max_stack_depth = 7680

INSERT INTO sales_info(id, category, ischeck, eventdate)

SELECT id,

(ARRAY['A','B','C','D','E','F','J','H','I','J','K'])[((RANDOM())*10)::INTEGER] AS category,

((1*(RANDOM())::INTEGER)<1) AS ischeck,

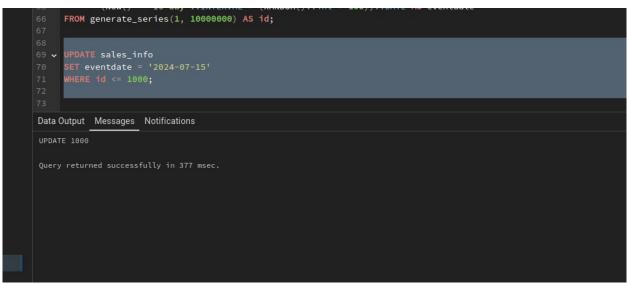
(NOW() - '10 day'::INTERVAL * (RANDOM()::int * 100))::DATE AS eventdate

FROM generate_series(1, 10000000) AS id;

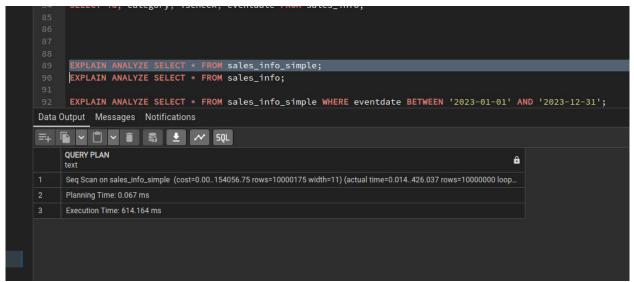
Data Output Messages Notifications

INSERT 0 10000000

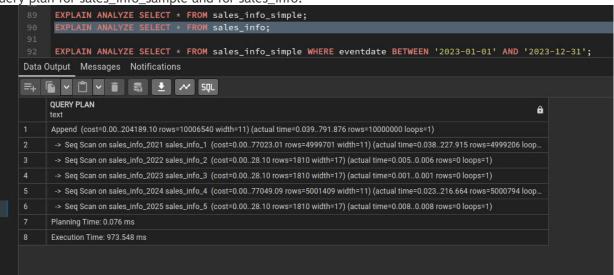
Query returned successfully in 8 secs 270 msec.
```



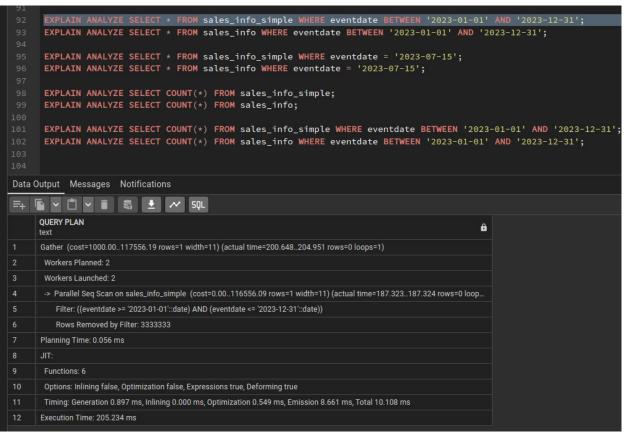
Update some rows.



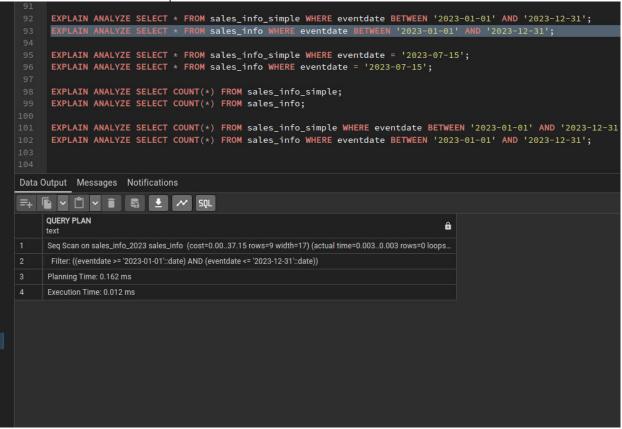
Query plan for sales_info_sample and for sales_info.



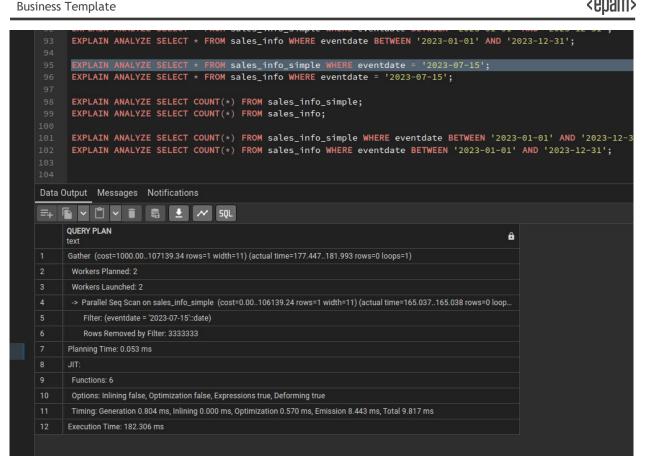
As we can see it uses Sequential Scan for whole table which is not partitioned, in second table it scanes separately in each partition.

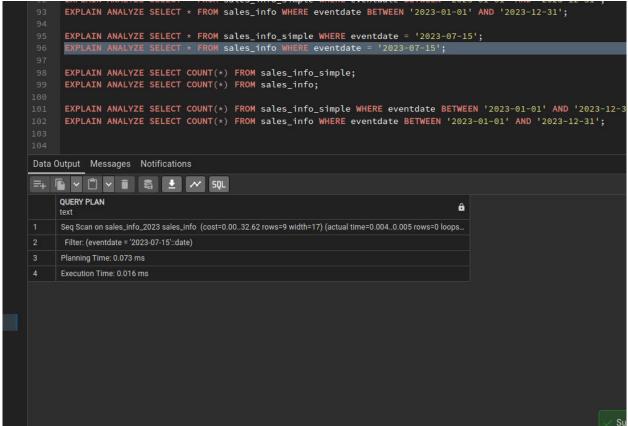


Plan with date fillter in not partitioned table.

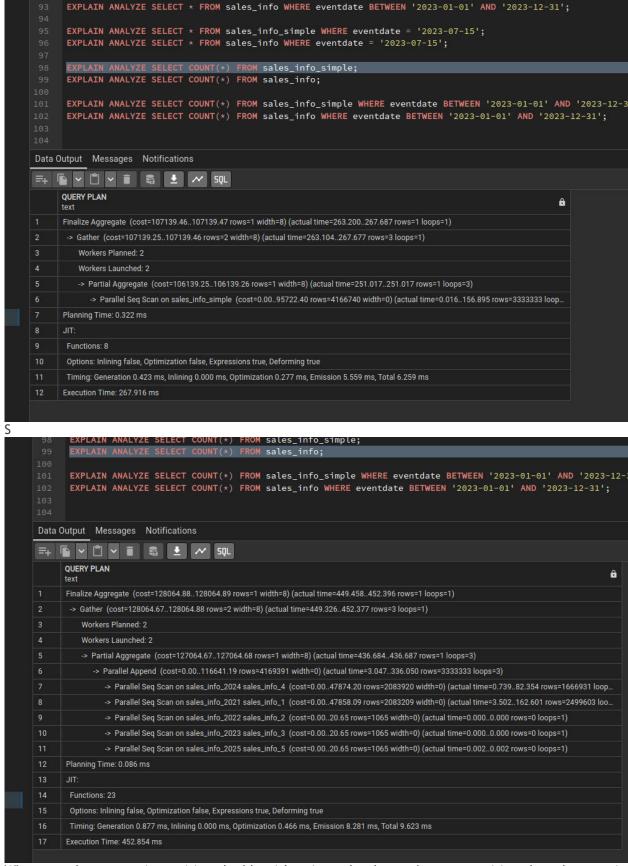


For partitioned one we can see that its only scan on the one partition where the date is included , in not partitioned table it's need to check whole table.



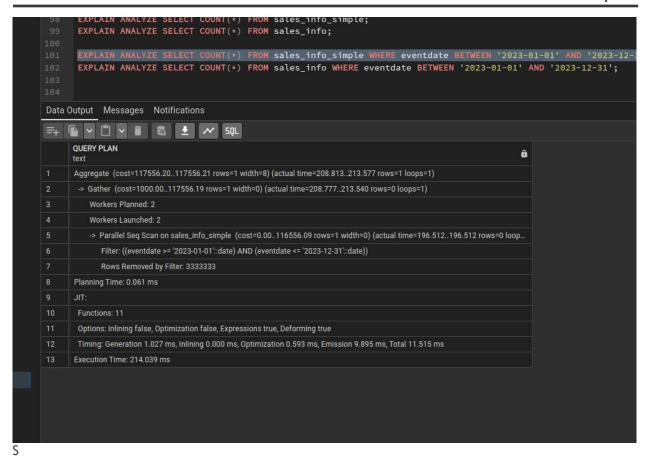


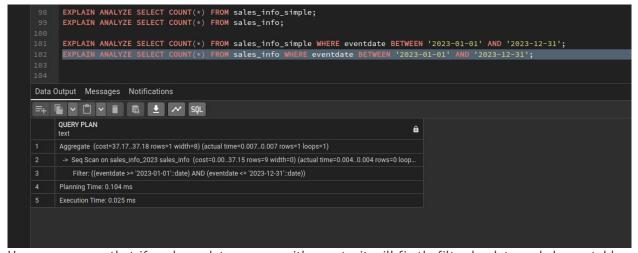
Same thing here, we can see that partitioned table is way efficient.



When we select count in partitioned table, it's going to be slower than not partitioned one because in not partitioned table it goes only through one table, but in partitioned one each partition is separate table.







Here we can see that if we have dates , even with count , it will firstly filter by date and choose table to go though , so as we see it's perfectly optimizes our query.

New partition created.

Task_2
Table created ,declarative partitions created for each year and categories.

```
Ø postgres/postgres@DWH

        □
        □
        ✓
        ✓
        ✓
        No limit
        →
        □
        ▶
        ▶
        □
        □
        □
        ○
        □
        □
        ○
        □
        ○
        □
        ○
        ○
        □
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
        ○
Query Query History
         FOR VALUES IN ('DEFAULT');
58 V CREATE TABLE SALES_INFO_DP_2023_CAT_A_B PARTITION OF SALES_INFO_DP_2023
61 v CREATE TABLE SALES_INFO_DP_2023_CAT_C_D_E_F PARTITION OF SALES_INFO_DP_2023
         FOR VALUES IN ('C', 'D', 'E', 'F', null);
64 V CREATE TABLE SALES_INFO_DP_2023_DEFAULT PARTITION OF SALES_INFO_DP_2023
       FOR VALUES IN ('DEFAULT');
68 V CREATE TABLE SALES_INFO_DP_2024_CAT_A_B PARTITION OF SALES_INFO_DP_2024
71 v CREATE TABLE SALES_INFO_DP_2024_CAT_C_D_E_F PARTITION OF SALES_INFO_DP_2024
74 V CREATE TABLE SALES_INFO_DP_2024_DEFAULT PARTITION OF SALES_INFO_DP_2024
79 V INSERT INTO SALES_INFO_DP(id, category, ischeck, EventDate)
          ,('{"A","B","C","D","E","F"}'::text[])[((
        RANDOM())*10)::INTEGER] category
         ,((1*(RANDOM())::INTEGER)<1) ischeck
            ,(NOW() - '10 day'::INTERVAL * (RANDOM()::int * 100))::
85 DATE EventDate
         FROM generate_series(1,10000000) id;
Data Output Messages Notifications
```

```
FROM generate_series(1,10000000) id;

88 VUPDATE SALES_INFO_DP

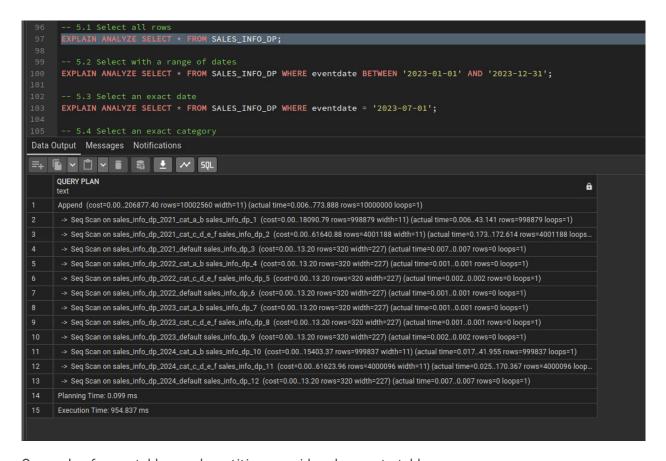
SET category = 'B'

WHERE category = 'A' AND eventdate >= '2021-01-01' AND eventdate < '2022-01-01';

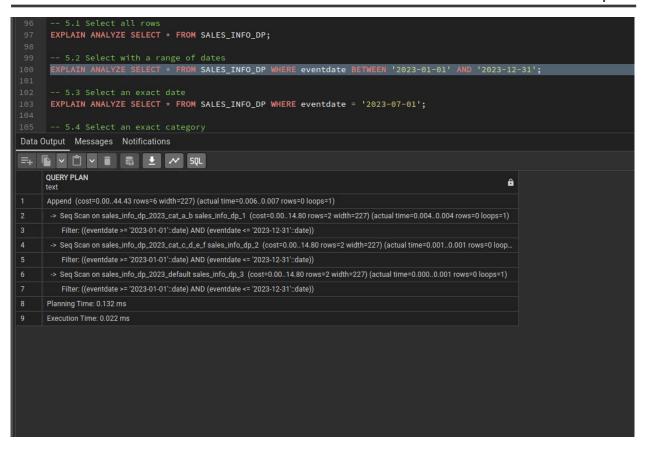
Data Output Messages Notifications

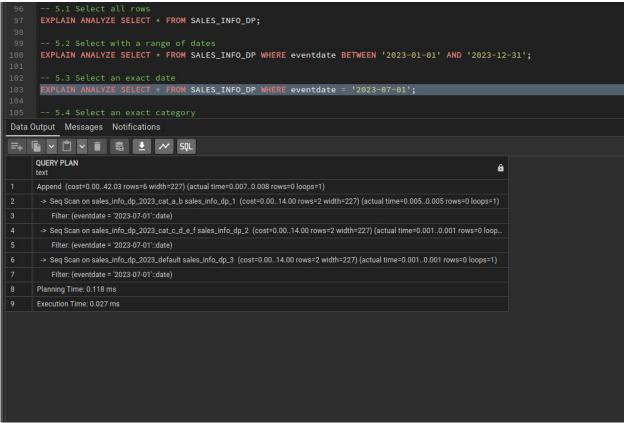
UPDATE 499863
```

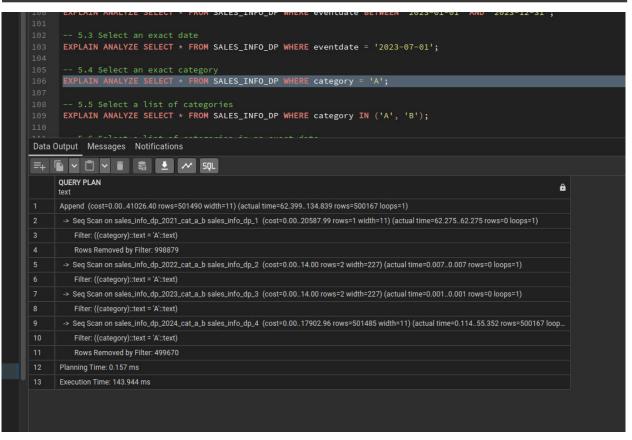
Updated some records.

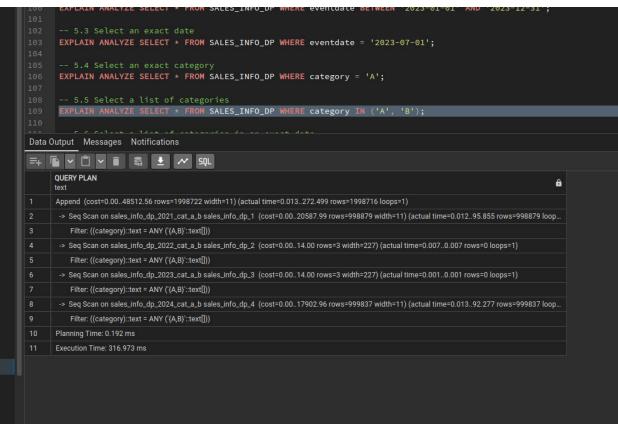


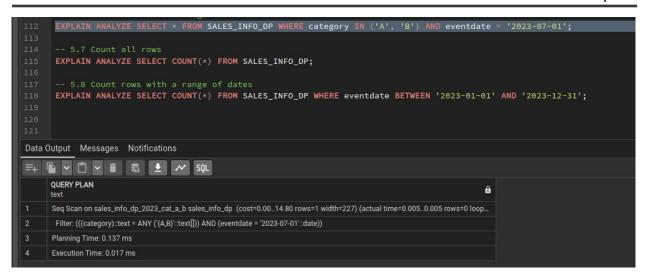
Query plan for our table , each partitions considered separate table.

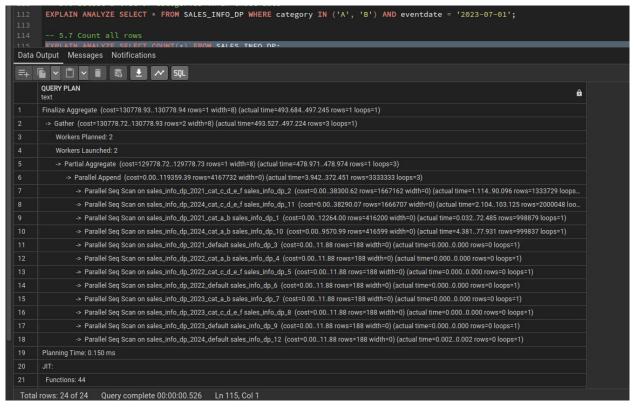




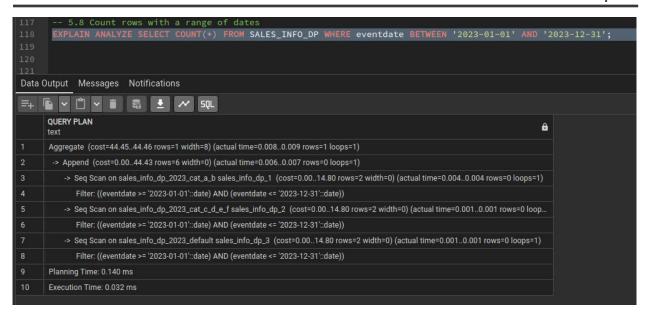












Task 3

```
SET max_parallel_workers_per_gather = 4;

129
130
131
EXPLAIN ANALYZE SELECT * FROM SALES_INFO;

133
134
EXPLAIN ANALYZE SELECT * FROM SALES_INFO_DP;

135
136
EXPLAIN ANALYZE SELECT * FROM SALES_INFO_SIMPLE;

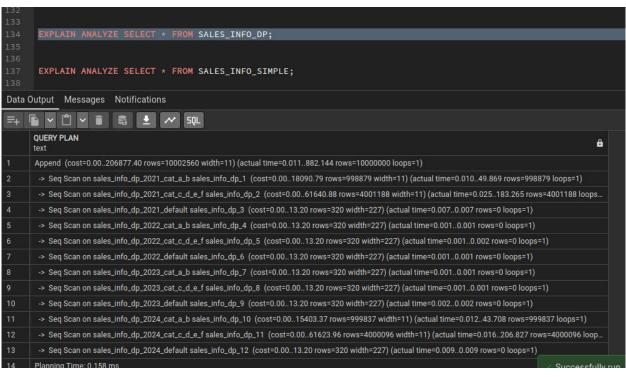
138

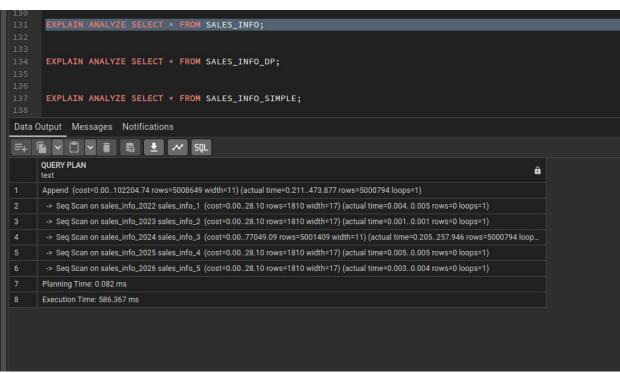
Data Output Messages Notifications

Explain Analyze Select * FROM SALES_INFO_SIMPLE;

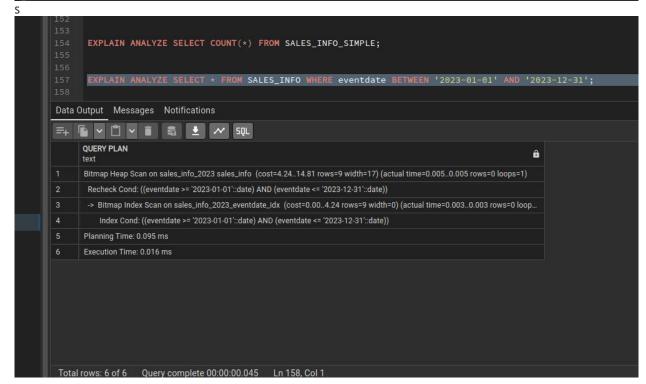
1 Seq Scan on sales_info_simple (cost=0.00.154056.75 rows=10000175 width=11) (actual time=0.060.455.260 rows=10000000 loop...

2 Planning Time: 0.034 ms
3 Execution Time: 637.447 ms
```

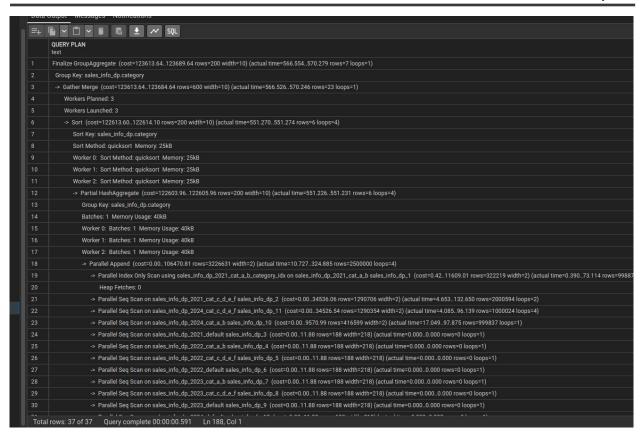




```
FRUM SALES_INFU_DF;
       EXPLAIN ANALYZE SELECT * FROM SALES_INFO_SIMPLE;
        EXPLAIN ANALYZE SELECT * FROM SALES_INFO ORDER BY eventdate;
       EXPLAIN ANALYZE SELECT * FROM SALES_INFO_DP ORDER BY eventdate;
        EXPLAIN ANALYZE SELECT * FROM SALES_INFO_SIMPLE ORDER BY eventdate;
Data Output Messages Notifications
=+ 🖺 v 📋 v 🛊 👼 👱 🚜 SQL
      QUERY PLAN
      Append (cost=1.04..144307.72 rows=5008034 width=11) (actual time=0.050..617.866 rows=5000794 loops=1)
       -> Index Scan using sales_info_2022_eventdate_idx on sales_info_2022 sales_info_1 (cost=0.15..71.30 rows=1810 width=17) (actual time=0.005..0.005 rows=0 loops
       -> Index Scan using sales_info_2023_eventdate_idx on sales_info_2023 sales_info_2 (cost=0.15..71.30 rows=1810 width=17) (actual time=0.002..0.002 rows=0 loops=
       -> Index Scan using sales_info_2024_eventdate_idx on sales_info_2024 sales_info_3 (cost=0.43..118982.34 rows=5000794 width=11) (actual time=0.043..417.184 row
       -> Index Scan using sales_info_2025_eventdate_idx on sales_info_2025 sales_info_4 (cost=0.15..71.30 rows=1810 width=17) (actual time=0.008..0.008 rows=0 loops=
       -> Index Scan using sales_info_2026_eventdate_idx on sales_info_2026 sales_info_5 (cost=0.15..71.30 rows=1810 width=17) (actual time=0.004..0.004 rows=0 loops=
      Planning Time: 0.159 ms
      Execution Time: 723.178 ms
```



S



Selecting count of row, as wee see it's goes through all partitions in parrallel after creating index.

