EPAM Systems, RD Dep., RD Dep.

POSTGRESQL DB FOR DWH AND ETL BUILDING

PostgreSQL Data Access and PostgreSQL Optimizer

CONTENTS

1. READING THE PLAN

1.1 TASK 1 – TABLE WITHOUT INDEX

Task Result: Read the plan and describe what happened and why with screenshots where needed. Check the difference between plans of EXPLAIN, EXPLAIN ANALYZE and EXPLAIN (ANALYZE, BUFFERS) command.

1. Create table test_index:

```
CREATE TABLE labs.test_index_plan (
    num       float NOT NULL,
    load_date timestamptz NOT NULL
);
```

2. Fill the table with a lot of test data:

```
INSERT INTO labs.test_index(num, load_date)
SELECT random(), x
FROM generate_series('2017-01-01 0:00'::timestamptz,
    '2021-12-31 23:59:59'::timestamptz, '10 seconds'::interval) x;
```

3. Check the plan of the select (twice at least, is any difference in plans?). Disable the parallel query planning If it needed:

```
SET max_parallel_workers_per_gather = 0;

SELECT *
FROM labs.test_index_plan
WHERE load_date BETWEEN '2021-09-01 0:00' AND '2021-10-31
11:59:59'
ORDER BY 1;
```

1.2 TASK 2 – ADDING INDEX

Task Result: Read the plan and describe what happened and why with screenshots where needed. Check the difference between plans of EXPLAIN, EXPLAIN ANALYZE and EXPLAIN (ANALYZE, BUFFERS) command.

1. Create B-Tree Index on test_index_plan table for load_date column.
2. Check the plan of the select (twice at least, is any difference in plans?). Disable the parallel query planning If it needed:

```
SET max_parallel_workers_per_gather = 0;

SELECT *
FROM labs.test_index_plan
WHERE load_date BETWEEN '2021-09-01 0:00' AND '2021-10-31
     11:59:59'
ORDER BY 1;
```

3. What can be done to query to use INDEX ONLY SCAN method?
4. DROP B-tree Index from test_index_plan table and create BRIN index on test_index_plan table for load_date column. Check the plan of the select (twice at least, is any difference in plans?).

```
SELECT *
FROM labs.test_index_plan
```

```
 WHERE load_date BETWEEN '2021-09-01 0:00' AND '2021-10-31
11:59:59'
 ORDER BY 1;
```

2. ADDING DATA WITH INSERT AND COPY

2.1 TASK 3 BULK INSERT

Task Result: Provide queries where needed. Describe what happened and why with screenshots where needed.

1. Create new table:

```
CREATE TABLE labs.test_inserts (
    num       float NOT NULL,
    load_date timestamptz NOT NULL
);
```

2. Add B-Tree index on the table test_inserts on load_date column.
3. INSERT into test_inserts by using:

```
SELECT num, load_date
FROM labs.test_index_plan;
```

4. Create new table*:

```
CREATE TABLE emp (
    empno        NUMERIC(4)  NOT NULL CONSTRAINT emp_pk PRIMARY KEY,
    ename        VARCHAR(10) UNIQUE,
    job          VARCHAR(9),
    mgr          NUMERIC(4),
    hiredate     DATE
);
```

5. Rewrite INSERT statements to more efficient way, run it:

```
INSERT INTO emp VALUES (1,'SMITH','CLERK',13,'17-DEC-80');
INSERT INTO emp VALUES (2,'ALLEN','SALESMAN',6,'20-FEB-81');
INSERT INTO emp VALUES (3,'WARD','SALESMAN',6,'22-FEB-81');
INSERT INTO emp VALUES (4,'JONES','MANAGER',9,'02-APR-81');
INSERT INTO emp VALUES (5,'MARTIN','SALESMAN',6,'28-SEP-81');
INSERT INTO emp VALUES (6,'BLAKE','MANAGER',9,'01-MAY-81');
INSERT INTO emp VALUES (7,'CLARK','MANAGER',9,'09-JUN-81');
INSERT INTO emp VALUES (8,'SCOTT','ANALYST',4,'19-APR-87');
INSERT INTO emp VALUES (9,'KING','PRESIDENT',NULL,'17-NOV-81');
INSERT INTO emp VALUES (10,'TURNER','SALESMAN',6,'08-SEP-81');
INSERT INTO emp VALUES (11,'ADAMS','CLERK',8,'23-MAY-87');
INSERT INTO emp VALUES (12,'JAMES','CLERK',6,'03-DEC-81');
INSERT INTO emp VALUES (13,'FORD','ANALYST',4,'03-DEC-81');
INSERT INTO emp VALUES (14,'MILLER','CLERK',7,'23-JAN-82');
```

*please do not delete emp table after this module.

2.2 TASK 4 COPY COMMAND

Task Result: Provide queries where needed. Describe what happened and why with screenshots where needed.

1. Use COPY Command to export your test_index_plan table into csv file:

```
COPY labs.test_index_plan TO
'\path_to_file\test_index_plan.csv' DELIMITER ',' CSV HEADER;
```

   Change command to export column load_date from test_index_plan with quotes and num column without quotes.

2. Use COPY to export data from test_index_plan table into csv file 'test_index_plan_short.csv' where load_date between '2021-09-01 0:00' AND '2021-09-01 11:59:59'
3. Create new table:

```
CREATE TABLE labs.test_copy (
      num       float NOT NULL,
      load_date timestamptz NOT NULL
);
```

4. Add B-Tree index on the table test_inserts on load_date column.
5. COPY into test_copy by using test_index_plan.csv file.

2.3 TASK 5 UPSERT

Task Result: Provide queries where needed. Describe what happened and why with screenshots where needed.

1. Add into emp table following information in one UPSERT statement:

1	SMITH	MANAGER	13	01-DEC-21
14	KELLY	CLERK	1	01-DEC-21
15	HANNAH	CLERK	1	01-DEC-21
11	ADAMS	SALESMAN	8	01-DEC-21
4	JONES	ANALIST	9	01-DEC-21