KOREA UNIVERSITY
DATAX LAB

Chapter 16 - Lab

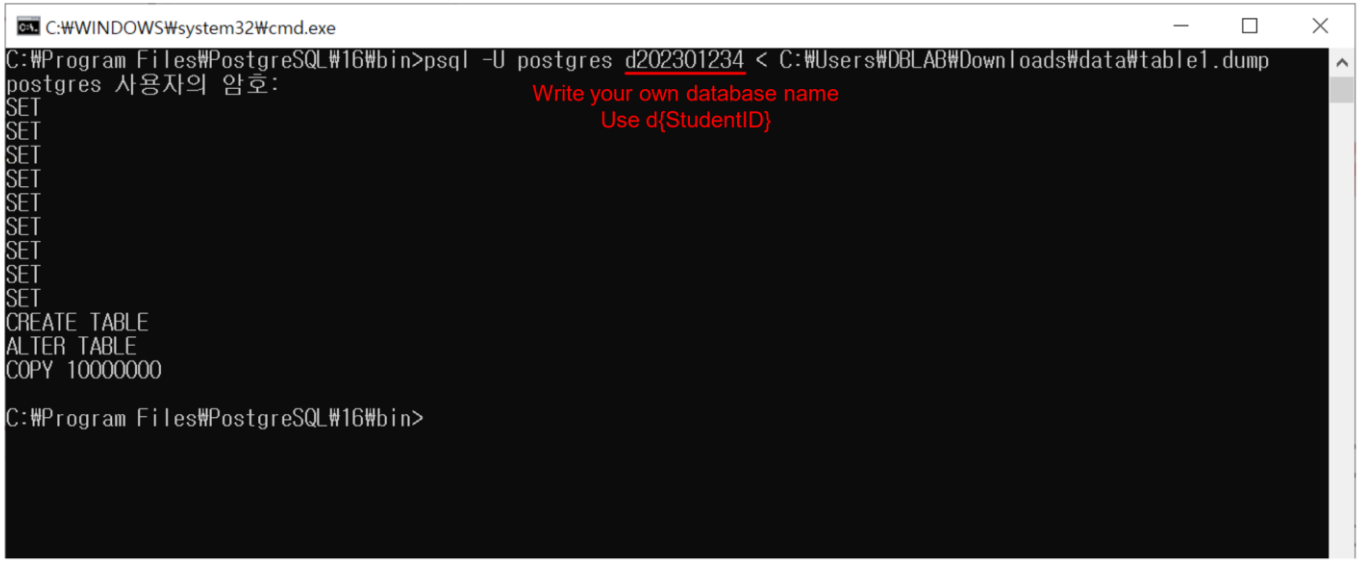
Query Optimization 1

Lab Setup

- Remove table “table1” from PostgreSQL
  - DROP TABLE table1;
- Download the file from blackboard
  - “table1.dump”

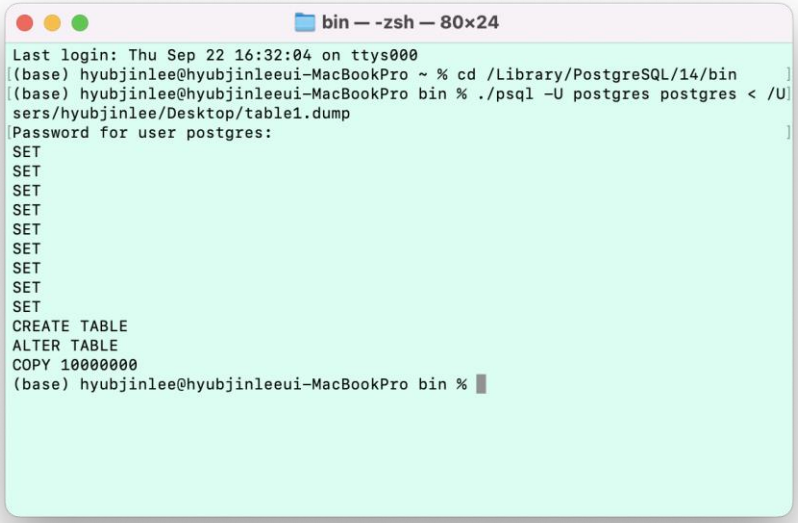
Lab Setup (Windows)

- Open Command Prompt (cmd.exe) and type the following commands:
  1. cd C:\Program Files\PostgreSQL\16\bin
     - This is the default path. If you installed it somewhere else, go to that path.
  2. psql -U postgres d{*StudentID*} < [filepath]\table1.dump
     - For [filepath], type the path where you downloaded “table1.dump”.
  3. Type your own PostgreSQL password



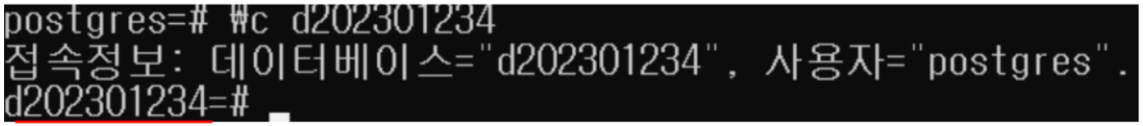
Lab Setup (Max OS X)

- Open Terminal and type the following commands:
  1. cd /Library/PostgreSQL/14/bin
     - This is the default path. If you installed it somewhere else, go to that path.
  2. psql -U postgres d{*StudentID*} < [filepath]\table1.dump
     - For [filepath], type the path where you downloaded “table1.dump”.
  3. Type your own PostgreSQL password



Lab Setup

- Execute PostgreSQL SQL Shell (psql) and login your database
  - Server [localhost]: Press the enter key
  - Database [postgres]: Press the enter key
  - Port [5432]: Press the enter key
  - Username [postgres]: Press the enter key
  - Password for user postgres: Type your own password
  - \c d{StudentID}



Your answers must be displayed along with your student ID.

- Type on psql command line
  - SET enable_bitmapscan=false;
  - SET max_parallel_workers_per_gather=0;

Table Information

- “table1” has 10,000,000 rows
- “table1”’s schema is as follows:

Attribute	Data Type	Cardinality	Features
sorted	integer	2,000,000	Sorted
unsorted	integer	1,986,519	Unsorted
rndm	integer	100,000	Dummy field
dummy	character(40)	1	Dummy field

Exercise 1

- Consider the following query and make corresponding SQL statements, and then show that the results are the same
  - Select “unsorted” from table1 where the “unsorted” value is 967 or 968 or 969 (967~969)
    a. Make an SQL statement using “BETWEEN” and “AND” operator
    b. Make an SQL statement using “IN” operator
    c. Make an SQL statement using “=” and “OR” operator
    d. Make an SQL statement using “UNION ALL” operator

Exercise 2

- Execute exercies1's SQL statements under each of the following conditions, and then compare the execution time of the queries
  a. No index
  b. B-tree index
  c. Hash index

- You must drop the B-tree index before doing the Hash index
- 'EXPLAIN ANALYZE' statement shows the execution time of the query

Lab Setup

- Create two synthetic data tables that has 5,000,000 rows with values between 0 and 500
  - CREATE TABLE pool1(val integer);
  - CREATE TABLE pool2(val integer);
  - INSERT INTO pool1(val) SELECT random()*500 FROM (SELECT generate_series(1,5000000)) as T;
  - INSERT INTO pool2(val) SELECT random()*500 FROM (SELECT generate_series(1,5000000)) as T;

Exercise 3

- Following queries have different syntax but return same result
  a. UNION ALL tables, and then perform aggregation with COUNT function
  b. Perform aggregation with COUNT function on each table, and then aggregate them again with SUM function on the UNION ALL of the aggregated results

- Write the queries and use ‘EXPLAIN ANALYZE’ statement to see how the query execution is actually planned

Exercise 4

- Following queries have different syntax but return same result
  a. SELECT tuple WHERE value is above 250 on each table and then UNION them
  b. UNION two tables and SELECT tuples WHERE value is above 250

- Write the queries and use ‘EXPLAIN ANALYZE’ statement to see how the query execution is actually planned

Homework

- Complete today’s practice exercises
- Write your queries and take screenshots of execution results
- Submit your report on blackboard
  - 10:29, November 12th, 2024
  - **Only PDF files** are accepted
  - **No late submission**

End of Lab