Mark as done

Part 1: Follow the instructions for loading the large relations file provided on the moodle page.

Then do the following (no submission required):

NOTE: *Execute the queries for this part using pgadmin3 rather than psql, since you get a much better looking output, and you can even see query trees in the Explain tab.*

1. Before going further, run the ANALYZE command:
   - ANALYZE collects statistics about the contents of tables in the database, and stores the results in the system catalog. Subsequently, the query planner uses these statistics to help determine the most efficient execution plans for queries.
   - With no parameter, ANALYZE examines every table in the current database. With a parameter, ANALYZE examines only that table. It is further possible to give a list of column names, in which case only the statistics for those columns are collected
2. **NOTE**: PostgreSQL uses a technique called bitmap-index-scan, which is used for secondary index access to multiple tuples. It scans the secondary index, and creates a bitmap with 1 bit per page in the file, and sets the page bit to 1 if the index shows a relevant tuple is in that page. After finishing the index scan the bitmap is complete, and pages are now scanned in physical order (using the bitmap heap scan operator) skipping pages whose bit is 0. For each retrieved page, all tuples are scanned, and the original selection condition has to be rechecked, since the page may contain tuples that did not satisfy the original condition (as few as 1 tuple may satisfy the condition). Bitmap index scans are useful only when multiple tuples are retrieved using a secondary index.
3. Run each of the following queries to find the time taken, and use the explain feature to find the plan used for each of the queries. By studying the plans, try to figure out why each query either ran fast or ran slowly. No need to submit anything for this part of the assignment, just see what plans are generated).
   1. select * from student where ID = '1000'
   2. select * from section where course_id = 'CS-101'
   3. select * from takes natural join student;
   4. select * from takes natural join student where ID = '1000';
   5. select * from student, instructor where student.id = instructor.id and student.id = '1000'
   6. select * from student, instructor where upper(student.id) = upper(instructor.id) and student.id = '1000'
4. Learn about pg_stats, by reading the manual (https://www.postgresql.org/docs/current/static/view-pg-stats.html) Then run the following queries to see some of the statistics and other catalog information
   1. select * from pg_stats where tablename = 'student'
   2. select * from pg_indexes where tablename not like 'pg_%'
5. Try out the graphical explain plan feature of pgadmin3. Enter a query, then on the top menu go to Query > Explain (or Explain Analyze) to see the plan.

Last modified: Tuesday, 3 September 2024, 3:54 PM

Jump to...

‹ Previous Activity

Next Activity ›