Final Assignments
in Database Performance

Please choose one of the two proposed assignments, plan and run your performance experiments and write the final report. Please send me the report's PDF file by e-mail (3-6 pages, excluding title page).

Assignment #1: "B*-tree Index Performance"

Goal: Experimentally evaluate B*-tree index performance to answer the following questions.

1. How much can indexes reduce query execution times (vs. sequential scan) in relation to:
   a. table size (number of records)
   b. selectivity of WHERE predicates (percentage of records retrieved)
   c. indexed column data type/size, column expression
2. How accurate are query optimizer cost estimates (vs. real execution time) for index-based query execution plans in relation to:
   a. table size (number of records)
   b. selectivity of WHERE predicates (percentage of records retrieved)
   c. indexed column data type/size, column expression
3. How much do indexes degrade performance of INSERT/UPDATE/DELETE commands (vs. no indexes)
4. How much disk space do indexes consume (vs. table disk size) in relation to:
   a. table size (number of records)
   b. indexed column data type/size, column expression
5. How much time does it take to build a new index in relation to:
   a. table size (number of records)
   b. indexed column data type/size, column expression

Tasks:

1. Design a table containing multiple columns of different data types and sizes (eg. short text, long text, number, date).
2. Implement a script to generate a specified number of synthetic records of random values..
3. Create B*-tree indexes on table columns and on column expressions (eg. `upper(col)`).
4. Implement a set of SQL queries that use the indexes.
5. Implement a set of INSERT/UPDATE/DELETE commands to modify the table.
6. Measure query/statement execution times.
7. Visualize your result in the form of charts.
8. Write the report.

Assignment #2: "Table Join Performance"

Goal: Experimentally evaluate join performance to answer the following questions:

1. How efficient (fast) are different join methods in relation to:
    a. table sizes (number of records)
    b. join operators (eg. "=", like, between)
    c. join column data type/size, column expression
2. How accurate are query optimizer cost estimates (vs. real execution time) for join-based query execution plans in relation to:
    a. join methods (hash join, nested loops, sort merge)
    b. table sizes (number of records)
    c. join operators (eg. "=", like, between)
    d. join column data type/size, column expression
3. How much can indexes improve join performance in relation to:
    a. join methods (hash join, nested loops, sort merge)
    b. table sizes (number of records)
    c. join operators (eg. "=", like, between)
    d. join column data type/size, column expression

Tasks:

1. Design two tables containing multiple columns of different data types and sizes (eg. short text, long text, number, date)
2. Implement a script to generate a specified number of synthetic records of random values, with specified number of distinct values.
3. Implement a set of join SQL queries.
4. Measure query/statement execution times.
5. Visualize your result in the form of charts.
6. Write the report.