

Due: February 8**Purpose:**

To investigate how mySQL implements queries and to solidify your understanding of how query plans impact performance.

Deliverables:

For part 1, include a table for each join on each query that summarizes the explain output and the SHOW STATUS output. In addition, explain why the query performed as it did, based on the EXPLAIN output and the values you see with SHOW STATUS after running the query.

Details:

Setup: In this assignment, you will create a mySQL database with some tables, load the tables, and investigate how various query plans perform. You should set up the assignment by first doing the following:

- 1) We have created scripts for you to create the tables and load them. These can be found on the CSE server at /projects/class/cse541/assign3, or on the web site
- 2) Copy all files to your own work area. The script.sql file expects the .txt files to be in the same directory as the script itself.
- 3) Start MySQL and run script.sql to create the database and load it.

Tasks: For each of the queries below, write the query 3 different ways:

- 1) Use straight join to force a join order on the tables.
- 2) Let the system optimize the join order by writing the query without specifying a join order.
- 3) Add an index (or indexes) as appropriate to improve the query results. Think about what type of index makes sense (maybe there is no appropriate index). You may need to change the table type to achieve the type of index you would like to use.

Deliverables: For each query below:

- a) Use EXPLAIN to determine how mySQL will execute each version of the query (save this output to turn in with your assignment).
- b) Explain which version of the query should perform better, and why, based on the explain output. What information did you use from the EXPLAIN output to determine this?
- c) Execute the query. Gather statistics such as handler_read_*, sort_rows, sort_scan, select_scan, created_tmp_tables (using SHOW STATUS). Note: these values increase incrementally with each query, so you'll have to subtract after each query to determine the impact of any particular query. Explain what you see.

Queries:

Query 1: Find the employeeId and age of all the managers who have more than 400,000 budget.

Query 2: Find the employees' average age for each state whose employees earn between \$50,000 and \$75,000 and which has departments with a budget of less than 300,000.

Query 3: Find total bonus amount paid out by departments with a total budget greater than \$1,000,000 (that's total budget, across all states).