**CSCI 4333 Database Design and Implementation**

**Project Stage 4: Database Performance Evaluation**

**(This stage is optional; for extra credit)**

Instructor: Dr. Xiang Lian

Due Date: See the course Web page

**Description**

The fourth stage of the Project consists of the following tasks:

1. Explore bulk loading features of Oracle and MySQL (see the LOAD DATA INFILE statement). Write a program that generates data files to be loaded into database tables, such that tables that participate in the previously defined six queries can be populated with at least 10,000 rows.

***Common pitfall:*** *if a table contains too many duplicate rows, then indexes become much less useful.*

Bulk load data files into the database.

2. Evaluate the performance of the six test queries in Oracle and MySQL on the database instance created in the previous task. Draw a bar graph (e.g., using Excel) that presents queries on the horizontal axis, and their evaluation times in Oracle and MySQL on the vertical axis. A sample graph (with random values) is shown below.



***Common pitfall****: if a query is very fast (e.g., evaluates in 0.000s), then you can either increase the size of the query participating tables or evaluate the query N times in a loop and divide the obtained time measurement by N*.

***Important note:*** *to enable fair comparison, all experiments must be performed on the same machine and under the same environment settings.*

***Important note:*** *while performing experiments, all unrelated activities on a computer should be avoided (e.g., unnecessary programs and services should be terminated and the computer may be disconnected from the Internet). You should monitor CPU and memory loads during the experiments.*

3. Drop all the database indexes and repeat the experiment in Task 2 (*including drawing charts*).

4. Select one of the queries and explore its scalability in Oracle and MySQL implementations. For this experiment, you need to customize your data generator (see task 1) to produce table data of different sizes (e.g., 10,000; 20,000; 30,000; 40,000; and 50,000 rows). Note that you only need to generate data for those tables that are involved in the query. Evaluate the performance of the query in Oracle and MySQL on the database instances of different sizes. Draw a graph (e.g., using Excel) with two scalability curves; a sample graph (with random values) is shown below.



5. Make a conclusion (observations, summary, recommendations, and open questions) based on the conducted experiments.

**Submission**

Please submit an electronic copy of your project solution, including:

(1) student IDs and names of **ALL members in your team**, and

(2) your SQL scripts and database performance evaluation report

to the [Blackboard](https://mycourses.utrgv.edu/). Note that, each team only needs to submit **ONE** version to the Blackboard by one of the team members.

**Grading**

Tasks 1-5 receive 20 points each (100 total, if solved correctly).

If all tasks are accomplished without major errors, you can get **20 extra points**.