

CS 522 – Advanced Database

Mini Project #3

Due Date: Thursday, May 2, 2018

Background:

The third mini project this semester will be to implement and analyze a distributed database. As for the implementation, you may either simulate a distributed database by implementing each “site” in a different Oracle account, or you may implement them on a different server, if possible. In either case, you must set up a **database link** to make the connections to the remote database.

Design phase

Start with one of your projects from Database I. The requirement for the project is that it must have at least 7 tables.

Select a fictitious scenario and develop a distributed solution around this scenario. Your result must have the following components:

1. At least one table must be solely at one location.
2. At least one table must be completely replicated.
3. At least one table must be fragmented.

Please note that these are the MINIMUM requirements. Your design should require more than one table to be fragmented.

Analysis phase:

You must identify queries for each of the three scenarios above. For each query, determine the cost of performing the local part of this query for every site involved. Also, calculate the communication cost for the global user. Again, you must have a MINIMUM of 3 queries, but you could/should have more.

Implementation phase:

Implement your partitions, one in each group member’s account, or within each DBMS. Implement your queries in a Java program using JDBC. Be sure to indicate where your global user(s) is/are located. You may also want to implement synchronization using PL/SQL code.

Elements of the Final Report

- I. **Introduction** – *provide a review of the centralized database you have chosen to start with.* Also, provide a discussion of the fictitious situation you have devised. Be sure to provide enough information to justify your solutions presented in the following sections.
- II. **Server Site Analysis** – indicate how many different sites are needed and what data will be present at each site. How will you synchronize replicated data?
- III. **Schema** – present your Global Conceptual and External Levels and then describe your fragmentation and allocations schemas. Be sure to justify your solution based on your assumptions presented in section I. *Your Global User must be different from the local sites.*
- IV. **Queries and Query Analysis** – List the queries of your global user(s), first in English, and then in syntactically correct SQL. For each query:
 - A. Determine the optimal execution strategy for the query coming from the global user. You may build B-Trees anywhere you like.
 - B. Determine the communication cost of your solution in part A (in #bytes transferred).
- V. **Implementation:** Create your tables in your Oracle accounts (or in other DBMSs) and submit your code on Western Online. You do not need to include any code in the written report.

Assumption to make

- 1) Assume numbers are stored in unpacked BCD, each character requires one byte, and Dates are 6 bytes long. Use 1024 as your block size.
- 2) As for the number of records for each table, randomly generate a number for each table; the only restriction is that each table must have at least 10,000 records (**VERY IMPORTANT NOTE: do not insert this many records into your real tables!!!**)