CS 5513 - Spring 2022- Homework 1
Assigned: 2/16/2022
Due: 3/2/2022 by 2:30 PM to the class website on Canvas
Maximum Points: 75 points (without the BONUS Problem) or 85 points (with the BONUS Problem)

Notes:
- **Homework answers must be typed and submitted by 2:30 PM on the due date to the class website on Canvas.**
- **Homework is individual work; it must be done by you only, no collaboration with anyone else is allowed.**
- **Late homework will be accepted until 11:59 PM on the date following the due date with 5% (of the maximum points) penalty. Late homework submitted after this time will not be graded.**

Problem 1 (20 points): Consider the following relations for a relational database:

MovieTheater(theater_name, number_of_seats, address)
Employee(employee_ID, name, salary, theater_name)
Movie(movie_title, year_made, duration)
MovieShowsIn(movie_title, year_made, theater_name, ticket_cost)

Assume that Baker Movies, Inc. has three movie theaters: "Baker Movies LA" in Los Angeles, "Baker Movies SF" in San Francisco and "Baker Movies NYC" in New York City. Each theater is the site (node) of a distributed database. In addition to this, assume that every employee works in at most one movie theater.

Regarding the design of the distributed database, assume the following:
- The *Employee* relation is fragmented horizontally by *theater_name*, such that the information of an employee is stored at the site corresponding to the movie theater where he/she works.
- The relation *MovieTheater* is replicated at both sites: Baker Movies NYC and Baker Movie SF.
- The relations *Movie* and *MovieShowsIn* are stored at Baker Movies LA.

Describe a good strategy for processing each of the following distributed queries:
  a) A customer submits a query to the Baker Movies LA site to find the average ticket cost at each movie theater.
  b) A manager at Baker Movies SF submits a query to this site to find the names of the employees working at the movie theaters that are showing the movie titled "Spider-Man: No Way Home."

Problem 2 (20 points): Investigate the distributed database aspects of Oracle 21c and answer the following questions (a) and (b):

  a) How to create a distributed database in Oracle 21c and how to write a distributed query/transaction in this system? <u>Illustrate your answers to this question (a) by solving the following example problem using Oracle 21c:</u>

Assume that we have 3 different machines called gpel7, gpel8, and gpel9 under the domain ou.edu, and assume that we do not have permission problems. The distributed database consists of 4 relational tables: *Customer(cust_id, name, address), Product(prod_id, manufacturer, price), Transaction(trans_id, cust_id, date) and BoughtInTransaction(prod_id, trans_id, quantity)*. Assuming that the *Customer* table is in gpel7, the *Product* table is in gpel8, and the *Transaction* and *BoughtInTransaction* tables are in gpel9, write a query to find the customer names and average number of products that each customer has purchased.

Note: **for the question (a), you do not need to run your solution on Oracle 21c; you just need to write the Oracle 21c commands (step-by-step) for your solution**. **Simply writing the Oracle's general commands without applying them to solve the given example problem is NOT the answer.**

b) Describe the commit protocol that Oracle 21c uses to handle the commitment of distributed transactions. You must explain the protocol <u>in detail</u>.

Note that for each of the questions (a-b) for Problem 2, you will need to provide your detailed answer, references and soft copies of the references that support your answer. Without the references and soft copies of the references, your answer will not be graded and will be given a grade of zero point.

Problem 3 (35 points): Find a suitable business or organization with a requirement for distributed sharing of data. The company does not necessarily need to have a distributed database system at present, or even have a current requirement for such a system. The objective is to demonstrate how the company's requirements could best be represented in a distributed database. Suitable companies may include travel agents, estate agents, banks, insurance companies, retail chains, supermarket chains, video shops, etc. You are asked to produce a report for the analysis and design of your proposed distributed database system that includes the following:
   a) A detailed analysis of the company's data and query requirements including where each query is initiated, what its frequency is, and so on. At least 3 queries must be given.
   b) A suitable replication scheme for the system.
   c) A suitable fragmentation scheme for the system.
   d) A suitable query processing strategy for EACH query.

Note that for each of the questions (b-d) above for Problem 3, you will need to provide <u>your solutions AND detailed justifications based on the analysis you gave in Question (a).</u> The application you choose for this homework problem must not come from the textbooks / lecture notes / homework assignments / class projects / exams / examples / qualifying exams used for CS 5513 at the University of Oklahoma.

BONUS PROBLEM (optional) (10 points) (your answer to this bonus problem will be graded only if you also completed ALL the required problems 1, 2, and 3 in this homework assignment):
Describe the Locking Mechanism that Oracle 21c uses to maintain consistent data in a multiuser database environment. You must explain the mechanism <u>in detail.</u>

Note that for the BONUS problem, you will need to provide your detailed answer, references and soft copies of the references that support your answer. Without the references and soft copies of the references, your answer will not be graded and will be given a grade of zero point.

Notes:

- <u>References:</u> for all homework problems, if you have used any references to solve them, list the references in a separate section titled "References" in your homework paper and submit the soft copy of the papers/book chapter pages that you have referenced to the class website on Canvas when you submitted your homework paper.

- <u>Homework submission</u>: all files must be submitted as PDF files to the class website on Canvas. The file name convention for your homework answer file is your first name followed by your last named-HW1 (Example: JohnDoe-HW1). The file name convention for the PDF file storing a complete workshop/conference/journal paper or storing the cited pages of a book listed in your reference list is your first name followed by your last name-HW1-Reference# where # is the order number of the reference in your reference list (Example: JohnDoe-HW1-Reference5). **Submit each file as an individual PDF file; DO NOT ZIP YOUR FILES.**