CS/DSA 4513 – Section 001 - Fall 2021 - Dr. Le Gruenwald

GRADED HOMEWORK 3 (Maximum Points: 100 points)

Assigned: 10/6/2021
Group Portion Due: 10/18/2021 at 1:30 PM on Canvas
Individual Portion Due: 10/19/2021 at 11:59 PM on Canvas
No late submission of either the group portion or the individual portion will be accepted.
See the submission instructions at the end of the document. Read the "Group Graded Homework Grading Policy" posted on Canvas

PROBLEM 1:

Group Question:

1.1. To verify whether Azure SQL can check for integrity constraint violations, using the relational database that you have created for Problem 2 in Graded Homework 2, write SQL statements that implement the following queries, run them on Azure SQL, and capture the error messages generated by Azure SQL:
   a) One insertion query that violates the uniqueness of the primary key of a table.
   b) One insertion query that violates the not-null value of the primary key of a table.
   c) One insertion query, one deletion query, and one update query that violate the foreign key constraint of a table.
   d) One retrieval query that violates the domain constraint of an attribute of a table.

1.2. Review the SQL file you have created for Problem 2 in Graded Homework 2, choose one table that should be indexed (assuming all queries are approximately equally frequent), write SQL statement(s) to create an index on that table, and rerun the queries that need to access the table and index. Provide your detailed explanations as to why you chose that table and that search key for indexing, whether that index is primary or secondary, and why you chose those queries to rerun.

Individual Questions:
Each group member will be assigned one version of the following question (asking to consider only query A, B, C, or D being the most frequent) randomly after the group portion of the homework is due (see the submission instructions at the end of the document). You must be ready to answer any of the versions of the following question on your own to submit the individual portion of the homework:

Consider the case when one of the queries from the list below is executed far more frequently than the other queries from Problem 2 (c) in Graded Homework 2:

A) Find the age of the oldest performer who is either named "Hanks" or has acted in a movie named "The Departed".
B) Find the names of all movies that are either a Comedy or have had more than one performer act in them.
C) Decrease the earnings of all directors who directed "Up" by 10%.
D) Delete all movies released in the 70's and 80's (1970 <= release_year <= 1989).

Does this new information change your group's answer for Problem 1 of Graded Homework 3? Provide your detailed explanations as to why it does, or why it does not. If you argue for the change to the group's answer, make sure to explain which new table and/or which new search key should be used for indexing and whether the new index is primary or secondary. Provide the SQL statement(s) to create the new index. There's no need to execute the statement or any queries on Azure SQL.

PROBLEM 2 (Group Question, No Individual Question):

Write **a JAVA program** using JDBC and Azure SQL to implement the following options for the relational database that you have created for Problem 2 in Graded Homework 2:

1. *Insert the pid, pname, (estimated) years_of_experience, and age of a new performer into table Performer. Assume that the value of years_of_experience is not known and will be estimated using the current information in the database as follows: the years_of_experience value should be set to the average of the years_of_experience values for all performers with an age within +/- 10 years of the new performer's age. If there are no such performers, the new performer's years_of_experience should be set to 18 less than his/her age value. After these calculations, the years_of_experience value should be adjusted so that it is at least 0 and no more than the performer's age value.*
2. *Insert the pid, pname, (estimated) years_of_experience, and age of a new performer into table Performer. Assume that the value of years_of_experience is not known and will be estimated using the current information in the database as follows: the years_of_experience value should be set to the average of the years_of_experience values for all performers who have acted in a movie that was directed by a director with a user-given did (director id). If there are no such performers, the new performer's years_of_experience should be set to 18 less than his/her age value. After these calculations, the years_of_experience value should be adjusted so that it is at least 0 and no more than the performer's age value.*
3. *Display the complete information of all performers.*
4. *Quit (exit the program).*

Requirements:

- The program terminates only when the user chooses Option 4.
- Each of Options 1 and 2 must be implemented as a Transact SQL Stored Procedure and the pid, pname, age, and did (for option 2) must be entered as the procedure parameter values at runtime when the procedure is called. No option logic must be handled by the Java program. It must only be used for user input collection, database call to execute the stored procedure, and the result presentation.
- For testing, execute Option 3 once before and after each execution of Options 1 and 2; and execute each of Options 1 and 2 at least three times with different values of pid, pname, and age; and execute Option 4 at least one time to show that your program terminates correctly.
- The Java program and all the Transact SQL Stored Procedures must be commented properly.

PROBLEM 3 (Group Question, No Individual Question):

Given the following relational database table:
ComputerRepairSpecialists (name, certificateNo, city, price_per_hour)

The following insertions are performed on the table *ComputerRepairSpecialists*:

Insert record <Johnson, 11, Yukon, $20>;
Insert record <Black, 33, OKC, $20>;
Insert record <Grant, 22, Norman, $15>,
Insert record <White, 77, OKC, $20>,
Insert record <Chapman, 44, Edmond, $20>,
Insert record <Ford, 66, Enid, $25>,
Insert record <Haas, 99, OKC, $20>,
Insert record <Hougen, 88, Yukon, $25>,
Insert record <Clinton, 55, Tulsa, $25>

Assume each block in the *ComputerRepairSpecialists* file can store up to 3 specialist records and *ComputerRepairSpecialists* is organized as a sequential file with *name* as the ordering field. Do the following:

1. Show the contents (i.e. the data values as well as the associated block/bucket/record addresses) of the file after the last insertion.
2. Assuming that *ComputerRepairSpecialists* is an index-sequential file on the search key *name*, show the contents of the primary index and the secondary index on *price_per_hour* after the last insertion.
3. Assuming that a B-tree index file on *certificateNo* with order 3 is created for *ComputerRepairSpecialists* and no two specialists have the same *certificateNo*, using the definition of the B-tree given in class, show the content of the B-tree index file after the last insertion.

SUBMISSION INSTRUCTIONS:

Group Portion:

- The homework answers (one submission per group) must be submitted to **Canvas by 1:30 PM, 10/18/2021.** The submission includes the following:
    - Solutions for Problem 1: two files: a) the SQL file that shows the required explanations written as in-line comments, the SQL statement(s) for index creation, and the SQL queries that you chose to rerun; and b) The PDF file that shows the screenshots of the Azure SQL creation of the index and the Azure SQL execution of the SQL queries that you chose to rerun. Use the file name convention HW3_Problem1_Group X where X is your group number.
    - Solutions for Problem 2: three files: a) a Java file (extension .java) containing the Java source program; b) an SQL file (extension .sql) containing the Transact SQL Stored Procedures; and c) a PDF file that shows the steps indicating that you have executed the program successfully (the output must be included). Use the file name convention HW3_Problem2_Group X where X is your group number.
    - Solutions for Problem 3: A single PDF file with the typed/diagrammed solution for the Problem 3. No Azure SQL is required to solve Problem 3. Use the file name convention HW3_Problem 3_Group X where X is your group number.
    - Attach to your PDF for Problem 1 a cover page that contains the following information:

      COURSE: CS/DSA-4513 - DATABASE MANAGEMENT
      SECTION: 001
      SEMESTER: FALL 2021
      INSTRUCTOR:
      GRADED HOMEWORK NUMBER: 3
      GROUP NUMBER: <write your group number here>
      GROUP MEMBERS: <list the names of all members here>
      SCORE: <<we will record the total score of your group for both problems 1 and 2 here>>

Individual Portion:

- After the submission deadline of the group portion of this graded homework, and before the submission deadline of the individual portion of this graded homework, you will have to take a quiz on Canvas. The quiz will be open from 2:46 PM, Monday, October 18, 2021 to 11:59 PM, Tuesday, October 19, 2021. The quiz will contain one of the versions of the Individual Question for Problem 1. Once you open the quiz, you will have 60 minutes to submit your answer. You will have to upload one PDF document as your answer. The quiz will also ask you for your feedback on your group members (i.e., the scores you give to each of your group members on the group portion of this graded homework) as outlined in the "Group Graded Homework Grading Policy"

document available on Canvas. If you do not provide the scores for your group members, then we assume that you give the same points (10 points) to each of your group members.

NOTES:

- The instructions for using JDBC and Transact-SQL are available on Canvas.
- If you have questions concerning your Azure SQL account, using JDBC, Transact SQL, or other questions concerning this homework, meet one of your TAs during their office hours over Zoom or email them. Their office hours and contact information are on the Home Page on Canvas.
- Start this project early to avoid last minute system problems. No late submission will be accepted.