Homework 5

Due December 15, 2015 by 23:59 EDT

Graders:

Rohit Shivaji Bhangale
rsb432@nyu.edu

Tarang Dawer
td1098@nyu.edu

Instructions

1. Please follow the instructions for the homework below exactly.

2. If you have any questions regarding the homeworks, please email one or both of the graders for the course. Their names and email addresses are listed above as well as on the course website. They are not authorized to change the homework in any way without checking with me first, but they are there to help you if you need it, including to meet in person by appointment.

3. Submit your homework electronically to NYU classes by the due date and time. A typed document is much preferred to a scanned, handwritten one, but either is acceptable. If you submit a scanned, handwritten document, is must be written neatly and legibly.

4. Show *all* of your applicable work!

5. Students are required to do their own work and turn in their own assignments.

6. Be sure to follow the academic integrity rules linked to on the course website.

Assignments

1. **Instructions**:

   First review the instruction about how to use MySQL we uploaded before in NYUClass resources. Pay attention to how to execute SQL scripts, and how to output your results to a file.

   Then look at the database depicted in Homework05Relation.pdf. It describes tables for Company, Shipper, Customer, Vendor, Warehouse, Product, Stored, Orders, ProductUsage, Item, and Contains. The tables are almost the same as your solution to Homework 2, except for the definition of Shipment Payment in question (h).

   Please also read the script Homework05Script.sql carefully. It defines and creates the sample database and has placeholders for putting in your solutions. You need to produce the queries in part 3 and put your solutions in Homework05Script.sql.

   For each query of Part 3, *unless stated otherwise*,
   - a. Sort the results in ascending order
   - b. Remove duplicates from the answer

   Thus, for example, assuming you are going to select a and b, you should actually use:

   SELECT DISTINCT a, b …
   ORDER BY a ASC, b ASC;

   You may use intermediate tables while producing your answers. In order to run your queries without getting errors, please use TEMP1, TEMP2... as your intermediate table names. The graders will explicitly DROP them before your queries so that their old values, if any, will not create problems.

   After filling your solutions in Homework05Script.sql, you need then run your script on MySQL using the instructions on how to pipe the output into a file. You should name this output file Homework05Spool.txt, and you will need to hand in both this output file and your SQL script.

2. **Files Including in this Homework:**
   - a. This file: Homework05.pdf
   - b. Homework05Relation.pdf, a Visio implementation of the database.
   - c. Homework05Script.sql, a script that will produce the databse in MySQL and also contains placeholders to put your solutions to Part 3.

3. Queries:

Produce queries for the following questions and put your answers in
Homework05Script.sql.

a. Produce table Answer01 (WAREHOUSENAME, PRODUCTNAMETOTAL QUANTITY) from
Table ORDER, which produces possible aggregations based on subsets of
{WAREHOUSENAME,PRODUCTNAME} using ROLLUP.

b. Produce table Answer02 (WAREHOUSENAME, PRODUCTNAME, TOTAL QUANTITY)
from Table ORDER, which produces possible aggregations based on subsets of
{PRODUCTNAME, WAREHOUSENAME} using ROLLUP.

c. Produce table Answer03 (PRODUCTUSAGEVALUE, PRODUCTNAME, PRODUCTID)
after inserting a ProductUsage record : ProductUsageValue is 90, ProductName is XBOX
and ProductID is 2651-01-2126.

d. Produce table Answer05(PRODUCTNAME, PRODUCTID, PRICE, WEIGHTINLB, EIN)
after deleting Product records for which ProductName is Books.

e. Produce table Answer06(ITEMID, ITEMSIZE, SECONDARY) after updating an Item record
: change ItemSize from 9 to 19 where ItemID is 0016.

4. **A question on Physical Design**

You are given a disk containing 2^{42} bytes, the largest easily available now. The disk is organized into blocks each containing 2^{12} bytes, which is the default for NTFS. On the disk, we will store a file whose records are of length 200 bytes each. Among the fields of the file there is a field F of length 30 storing the primary key of the record.

a. (i) Using F as the primary key, design a B+ tree to serve as an index to this file by specifying the value of the parameter m as we did in class.

(ii) What would be the number of leaves in the "narrowest" and the "widest" possible tree which has 3 levels, if we wanted to use such a tree? (A tree consisting of the root only has one level, so you need to go down to the level of "grandchildren" of the root.)

(iii) What is the smallest and the largest number of pointers that can come "out" from the leaves of a tree of 3 levels?

b. Assume that the file can be clustered. What is the smallest and the largest number of records that a block of the file can have, assuming we want to use the space as efficiently as possibly while building the file using the process we had in class? Ignore the trivial case that the file has only a very small number of records.

For this question it is particularly important to show all your work and not just to state the answer with a brief sketch of the derivation.

Your solution has to be submitted as a PDF file. You may handwrite it neatly and scan it. There are free PDF scanners available on campus.

5. **What to Submit:**

A file YourNnumber.zip containing:

a. Homework05Script.sql
b. Homework05Spool.txt
c. Homework06Physical.pdf