CS430/630 – Homework 2
Released February 24, Due March 11
50 points (5/100 of final grade)

Instructions: The homework is due on Wed, March 11, 20:30:00 (BEFORE CLASS). All submissions must be in digital form. Create a folder “HW2” under your main folder for the course, and place two script files named Q1.sql and Q2.sql there. Ensure that the files are not readable by “others” (using chmod o-r filename) and that the files belong to the group CS630-1G and are readable by the group (chmod g+r filename). DO NOT CHANGE PERMISSIONS FOR ANY OF THE DIRECTORIES (ESPECIALLY THE 630 DIRECTORY IN YOUR HOMEDIR)!

YOU MUST ENSURE THAT THE SCRIPT CAN CORRECTLY EXECUTE ON THE DBS3 SERVER. ANSWERS THAT DO NOT RUN ON DBS3 WILL NOT RECEIVE ANY POINTS!

One-line comments in the script are recommended, in order to make the homework more readable.

Question 1 (30 points)
Consider a database schema with three relations:

Customers (cid:integer, cname:string, age:integer, zipcode:integer)
Vehicles (vin:integer, manufacturer:string, year:integer, seats:integer)
Sales (cid:integer, vin:integer, price:integer)

The database contains customer and vehicle information for a car dealership. The description is as follows: customers are uniquely identified by cid and they have as attributes name, age and zipcode of residence. Vehicles are uniquely identified by vin. Each vehicle has a manufacturer company (assume no two companies have the same name, this is a constraint that your data is guaranteed to have, there is no need to enforce it), production year and number of seats (how many persons fit in the car). The Sales relation provides information about car purchases, including price of purchase for each sale. Primary keys are underlined.

Provide SQL statements for the following:

(a) Write SQL declarations for creating the schemas. Include necessary key constraints.
(b) Find the names of customers who bought some vehicle costing more than $20,000.
(c) Find the manufacturers whose vehicles were never bought from zipcode 02125. List each manufacturer once. Exclude manufacturers for which there is no vehicle sold.
(d) Find the age of customers who never purchased a 6-seat vehicle (include customers who never purchased any vehicle).
(e) Find the manufacturers whose vehicles were purchased from all zipcodes.
(f) Find the year of fabrication for the vehicle(s) with the lowest price.
(g) Find the average sale price of a vehicle.
(h) Find the total dollar amount spent by each customer (listed by cname) who bought at least five vehicles.
(i) Find the zipcodes from which there are purchases of vehicles produced by all manufacturers. List each zipcode once.
(j) Find the earliest fabrication year among vehicles costing more than $30,000 for each manufacturer that had at least 100 vehicles sold (at any price).
(k) Find the average price of vehicles for each manufacturer that sells only vehicles costing $100,000 or more (implies they sold at least one such vehicle).

(l) Find the name(s) of customer(s) who bought the most vehicles costing $30,000 or more (note that there may be ties).

(m) [630 students only] Find the most common fabrication year (i.e., with most occurrences) among vehicles purchased from zipcode 12345.

(n) [630 students only] Find the zipcode(s) that generated the highest total amount of sales (refers to dollar amount) among vehicles that have the second-most-recent fabrication year.

Question 2 (20 points)

Consider a database schema with three relations:

Movies (movie_id:integer, title:string, year:integer, studio:string)
Actors (actor_id:integer, name:string, nationality:string)
StarsIn (actor_id:integer, movie_id:integer, character:string)

Provide SQL statements for the following:

(a) Write SQL declarations for creating the schemas. Include necessary key constraints.
(b) Find the title and studio of movies starring actor ‘Tom Hanks’.
(c) Find the names of actors of ‘US’ nationality.
(d) Find the nationalities of actors that star in some movie for all producing studios (another way to phrase this is “find the nationalities of actors that worked for all studios”). Ensure that a nationality appears only once in the result.
(e) Find for each year the number of distinct actors that played a character that starts with letter “G” and has at least three letters in the character name.
(f) Find the movie titles that are produced by “Universal” studio and in which there are at least ten actors starring.
(g) [630 students only] Find the nationality (or nationalities) best-represented (i.e., nationality of most actors) among actors that starred in movies produced in year 2019.