Sheet3

RELATIONAL ALGEBRA

FUNCTIONAL DEPENDENCIES AND NORMALIZATION

1) This exercise asks you to converting business statements into dependencies. Consider the following relation DiskDrive(serialNumber, manufacturer, model, batch, capacity, retailer). Each tuple in the relation DiskDrive contains information about a disk drive with a unique serialNumber, made by a manufacturer, with a particular model, released in a certain batch, which has a certain storage capacity, and is sold by a certain retailer. For example, the tuple DiskDrive(1978619, WesternDigital, A2235X, 765234, 500, CompUSA) specifies that WesternDigital made a disk drive with serial number 1978619, model number A2235X in batch 765235 with 500GB that is sold by CompUSA.
   Write each of the following dependencies as an FD:
   a) The manufacturer and serial number uniquely identifies the drive
   b) A model number is registered by a manufacturer and hence can't be used by another manufacturer.
   c) All disk drives in a particular batch are the same model.
   d) All disk drives of a particular model of a particular manufacturer have exactly the same capacity.
2) Suppose we have the following requirements for a university database that is used to keep track of students' transcripts:
   a) The university keeps track of each student's name (SNAME), student number (SNUM), social security number (SSSN), current address (SCADDR) and phone (SCPHONE), permanent address (SPADDR) and phone (SPPHONE), birthdate (BDATE), sex (SEX), class (CLASS) (freshman, sophomore, ..., graduate), major department (MAJORDEPTCODE), minor department (MINORDEPTCODE) (if any), and degree program (PROG) (B.A., B.S., ..., Ph.D.). Both ssn and student number have unique values for each student.
   b) Each department is described by a name (DEPTNAME), department code (DEPTCODE), office number (DEPTOFFICE), office phone (DEPTPHONE), and college (DEPTCOLLEGE). Both name and code have unique values for each department.
   c) Each course has a course name (CNAME), description (CDESC), code number (CNUM), number of semester hours (CREDIT), level (LEVEL), and offering department (CDEPT). The value of code number is unique for each course.
   d) Each section has an instructor (INSTUCTORNAME), semester (SEMESTER), year (YEAR), course (SECCOURSE), and section number (SECNUM). Section numbers distinguish different sections of the same course that are taught during the same semester/year; its values are 1, 2, 3, ...; up to the number of sections taught during each semester.
   e) A grade record refers to a student (Ssn), refers to a particular section, and grade (GRADE).

Design a relational database schema for this database application. First show all the functional dependencies that should hold among the attributes. Then, design relation schemas for the database that are each in 3NF. Specify the key attributes of each relation. Note any unspecified requirements, and make appropriate assumptions to make the specification complete.

3) Consider the following relation for published books:
   BOOK (Book_title, Authorname, Book_type, Listprice, Author_affil, Publisher)
   Author_affil refers to the affiliation of the author.
   Suppose the following dependencies exist:
   Book_title → Publisher, Book_type
   Book_type → Listprice
   Author_name → Author-affil
   a) What normal form is the relation in? Explain your answer.
   b) Apply normalization until you cannot decompose the relations further. State the reasons behind each decomposition.
4) Consider the following schema:
   Suppliers(<u>sid</u>: integer, sname: string, address: string)
   Parts(<u>pid</u>: integer, pname: string, color: string)
   Catalog(<u>sid</u>: integer, <u>pid</u>: integer, cost: real)
   The key fields are underlined, and the domain of each field is listed after the field name. Therefore sid is the key for Suppliers, pid is the key for Parts, and sid and pid together form the key for Catalog. The Catalog relation lists the prices charged for parts by Suppliers. Write the following queries in relational algebra:
   a) Find the names of suppliers who supply some red part.
   b) Find the sids of suppliers who supply some red or green part.
   c) Find the sids of suppliers who supply some red part or are at 221 Packer Street.
   d) Find the sids of suppliers who supply some red part and some green part.
   e) Find pairs of sids such that the supplier with the first sid charges more for some part than the supplier with the second sid.
   f) Find the pids of parts supplied by at least two different suppliers.
   g) Find the pids of the most expensive parts supplied by suppliers named Yosemite Sham.
5) Consider the Supplier-Parts-Catalog schema from the previous question. State what the following queries compute:
   a)
   $$\pi_{sname}(\pi_{sid}((\sigma_{color='red'}Parts) \bowtie (\sigma_{cost<100}Catalog)) \bowtie Suppliers)$$
   b)
   $$\pi_{sname}(\pi_{sid}((\sigma_{color='red'}Parts) \bowtie (\sigma_{cost<100}Catalog) \bowtie Suppliers))$$
   c)
   $$(\pi_{sname}((\sigma_{color='red'}Parts) \bowtie (\sigma_{cost<100}Catalog) \bowtie Suppliers)) \cap (\pi_{sname}((\sigma_{color='green'}Parts) \bowtie (\sigma_{cost<100}Catalog) \bowtie Suppliers))$$
   d)
   $$(\pi_{sid}((\sigma_{color='red'}Parts) \bowtie (\sigma_{cost<100}Catalog) \bowtie Suppliers)) \cap (\pi_{sid}((\sigma_{color='green'}Parts) \bowtie (\sigma_{cost<100}Catalog) \bowtie Suppliers))$$
   e)
   $$\pi_{sname}((\pi_{sid,sname}((\sigma_{color='red'}Parts) \bowtie (\sigma_{cost<100}Catalog) \bowtie Suppliers)) \cap (\pi_{sid,sname}((\sigma_{color='green'}Parts) \bowtie (\sigma_{cost<100}Catalog) \bowtie Suppliers)))$$

6) [SQL] For the COMPANY database of Figure 3.5 (the one used in class), specify the following queries in SQL. Show the query results if applied to the database of Figure 3.6.
   a) For each department whose average employee salary is more than $30,000, retrieve the department name and the number of employees working for that department.
   b) Suppose we want the number of male employees in each department rather than all employees.

How to submit the homework assignments?

- Solve the sheet individually without looking up the solution on the Internet. The sheet is to practice; it is a learning tool not an exam.
- Assignments are to be **handwritten**.
- Papers are to be scanned (I like camscanner app). Put all images in a pdf file (camscanner does that for you)
