CSCC43 Introduction to Databases (Summer 2009)

Assignment 4 - Database Design

Due (Electronic Copy): Wednesday, Aug. 05, at 11:59pm
(You do not need to submit hard copy for this assignment)

Total Marks: 100

Weight: 10% of your final grade

The last 48 hours before the deadline are considered a silent period.

Student 1

Name (Last, First) : ______________________________

Student Number : ______________________________

Student 2

Name (Last, First) : ______________________________

Student Number : ______________________________

UTSC login name: ______________________________

Problem 1 (40 points)

University of Toronto (UofT) has hired you to design a database for their student registration system. They provided you with the following relevant information:

- Every student has a unique student number, a name (first name and last name), a date of birth, an address and one or more phone numbers (home phone number, cell phone number). There are two categories of students: undergraduate students and graduate students.
- All the students are enrolled in programs, and a student cannot enroll in more than one program. For every student, the University records the year when the student enrolled in his/her program. A program has a name which is unique among all the programs at UofT, and is offered by one department.
- Each department at UofT has a unique name, an address and a phone number. All the employees at UofT have a unique employee number, a SIN, a name (first name and last name), an address and one or more phone numbers (home phone number, work phone number). An employee works in one department only.
- There are three types of employees at UofT: researchers, teachers and supporting staff. Researchers are supervising graduate students, and they are aliated with a research group. Teachers teach undergraduate courses, and supporting sta are in charge with administrative tasks.
- Undergraduate students enroll in courses in their program of study. A course has a unique course number and a title. Courses can have prerequisite courses associated with them. A semester is identifed by a year and a season (F(all), W(inter) and S(ummer)). A course is taught in a semester by a teacher, and there are semesters when a course is not offered. When a course is offered in a semester, it can have assigned one or more graduate students as teaching assistants. Students can enroll in more courses in one semester, and we need to record the final grade a student obtains for a course in one semester. There are semesters when a student does not enroll in a course.

For this problem, you are asked to hand in an E/R diagram for the UofT database described above.

- Structure your solution in following way as described in class: (1) identifying main concepts, (2) structuring requirements around main concepts, (3) adding entity sets, relationship sets and attributes to the schema, (4) adding constraints to the schema, (5) show the final schema with all the elements and constraints.
- Include a write-up, explaining the constraints that cannot be captured in E/R diagram.

- The above description is complex, may leave details out and include unnecessary and/or contradictory information. You will need to make some assumptions about the problem for missing or inconsistent information.

Sample Solution

See file er-sol.pdf for a sample solution to this problem.

Problem 2 (10 points)

Consider a database schema as described below

- University(uid, uname, city, country, president), which records the id and name of a university, the city and coutnry where it is located, and the name of its current President.
- Faculty(fid, uid, fname, dean), which records the id and name of a faculty, the name of its current Dean, and the id of university this faculty belongs to.
- Professor(eno, name, dob, rank), which records the employee number and name of a professor, his/her date of birth and current rank.
- AppointedBy(eno, fid, sdate, edate, salary), which records the appointment of a professor in a faculty, and the start, end date and annual salary of the appointment.

2.1 (5 points)

Is the design of this schema a good one? If you think the design is a good one, indicate which normal form it satisfies and justify your answer; if you think the design is a bad one, provide a better one, and justify your answer.

Sample Solution

Notice that all the non-trivial FDs are specified by the key constraints in the database schema. In other words, let F be a set of FD as shown below (corresponding to the key constraints declared in the schema):

- uid → uname,city,country,president
- fid,uid → fname,dean
- eno → name,dob, rank
- eno,fid → sdate,edate,salary

then F^+ is the set of all the non-trivial FDs are specified in the schema.

Therefore, the design of the database schema is in BCNF with respect to F. This is because we can verify that for any given relation schema R, the LHS of any FD in F^+ over R is a superkey for R.

2.2 (5 points)

Now consider following three new rules (which may not be true):
- The unveristy name within a city in a country is always unique
- A university cannot have two faculty with the same name
- Every professor is appointed by exactly one faculty at a time.

Rewrite each of these rules as a FD. And *given these three new FDs* consider again is the design of this schema a good one? If you think the design is a good one, indicate which normal form it satisfies and justify your answer; if you think the design is a bad one, provide a better one, and justify your answer.

Sample Solution

The three rules correspond to following three FDs:
- country,city,uname → uid
- uid,fname → fid
- eno,sdate,edate → fid

then F^+ is the set of all the non-trivial FDs are specified in the schema.

Given these two FDs, the design of the database schema is still in BCNF. This is because the LHS of each of the newly introduced FD is a superkey for the corresponding relation schema.

Problem 3 (30 points)

Consider a relation schema R with the set of attributes U = (I,J,K,D,E,L,G,H) and the set of functional dependencies F = {I → E, ID → JE, IK → E, E → J, JG → L, JE → D, JDH → E, L → I, D → H, KD → I }.

3.1 (25 points)

Find a minimal cover of the relation R and provide a 3NF lossless decomposition of the relation R. To get full marks, you have to explain all the steps.

Sample Solution

Find Minimal Cover

Step 1: Make RHS of each FD into a single attribute

G = (F - {ID → JE}) ∪ {ID → J, ID → E}

Step 2: Eliminate redundant attributes from LHS

- ID → J
  $(I)^+{}_F$ = IJDEH
  $(D)^+{}_F$ = DH
  D is eliminated from ID → J
  G = (G - {ID → J}) ∪ {I → J}

- ID $\rightarrow$ E
  $(I)^+{}_F$ = IJDEH
  $(D)^+{}_F$ = DH
  D is eliminated from ID $\rightarrow$ E
  G = (G - {ID $\rightarrow$ E}) $\cup$ {I $\rightarrow$ E}

- IK $\rightarrow$ E
  $(I)^+{}_F$ = IJDEH
  $(K)^+{}_F$ = K
  K is eliminated from IK $\rightarrow$ E
  G = (G - {IK $\rightarrow$ E}) $\cup$ {I $\rightarrow$ E}

- JG $\rightarrow$ L
  $(J)^+{}_F$ = J
  $(G)^+{}_F$ = G
  nothing is eliminated from JG $\rightarrow$ L
  G is unchanged

- JE $\rightarrow$ D
  $(J)^+{}_F$ = J
  $(E)^+{}_F$ = JDEH
  J is eliminated from JE $\rightarrow$ D
  G = (G - {JE $\rightarrow$ D}) $\cup$ {E $\rightarrow$ D}

- KD $\rightarrow$ I
  $(K)^+{}_F$ = K
  $(D)^+{}_F$ = DH
  nothing is eliminated from KD $\rightarrow$ I
  G is unchanged

- JDH $\rightarrow$ E
  $(JD)^+{}_F$ = JDEH
  $(DH)^+{}_F$ = DH
  $(JH)^+{}_F$ = JH
  H is eliminated from JDH $\rightarrow$ E
  G = (G - {JDH $\rightarrow$ E}) $\cup$ {JD $\rightarrow$ E}

- JD $\rightarrow$ E
  $(J)^+{}_F$ = J
  $(D)^+{}_F$ = DH
  nothing is eliminated from JDH $\rightarrow$ E
  G is unchanged

So after Step 2, G = {I $\rightarrow$ E, I $\rightarrow$ J, E $\rightarrow$ J, JG $\rightarrow$ L, E $\rightarrow$ D, JD $\rightarrow$ E, L $\rightarrow$ I, D $\rightarrow$ H, KD $\rightarrow$ I }

Step 2: Delete redundant FDs

- Can I $\rightarrow$ E be removed?
  $(\mathrm{I})^+{}_{F-\{I\rightarrow E\}}$ = IJ
  No.

- Can I $\rightarrow$ J be removed?
  $(\mathrm{I})^+{}_{F-\{I\rightarrow J\}}$ = IJDEH
  Yes.

- Can E $\rightarrow$ J be removed?
  $(\mathrm{E})^+{}_{F-\{E\rightarrow J\}}$ = DEH
  No.

- Can JG $\rightarrow$ L be removed?
  $(\mathrm{JG})^+{}_{F-\{JG\rightarrow L\}}$ = JG
  No.

- Can E $\rightarrow$ D be removed?
  $(\mathrm{E})^+{}_{F-\{E\rightarrow D\}}$ = JE
  No.

- Can JD $\rightarrow$ E be removed?
  $(\mathrm{JD})^+{}_{F-\{JD\rightarrow E\}}$ = JDH
  No.

- Can L $\rightarrow$ I be removed?
  $(\mathrm{L})^+{}_{F-\{L\rightarrow I\}}$ = L
  No.

- Can D $\rightarrow$ H be removed?
  $(\mathrm{D})^+{}_{F-\{D\rightarrow H\}}$ = D
  No.

- Can KD $\rightarrow$ I be removed?
  $(\mathrm{KD})^+{}_{F-\{KD\rightarrow I\}}$ = KDH
  No.

Finally, the minimal cover is G = {I $\rightarrow$ E, E $\rightarrow$ J, JG $\rightarrow$ L, E $\rightarrow$ D, JD $\rightarrow$ E, L $\rightarrow$ I, D $\rightarrow$ H, KD $\rightarrow$ I }

3NF Decomposition

Step 1: Partition the minimal cover into sets such that the LHS of all elements are the same.

- G_1 = {I → E}
- G_2 = {E → J, E → D}
- G_3 = {JG → L}
- G_4 = {JD → E}
- G_5 = {L → I}
- G_6 = {D → H}
- G_7 = {KD → I }

Step 2: For each G_i form schema R_i=(U_i, G_i).

- R_1 = (IE, {I → E})
- R_2 = (JDE, {E → J, E → D})
- R_3 = (JLG, {JG → L})
- R_4 = (JDE, {JD → E})
- R_5 = (IL, {L → I})
- R_6 = (DH, {D → H})
- R_7 = (IKD, {KD → I })

Note: R_2 and R_4 can be combined. ***Step 3: If no R_i is a superkey of R, add schema R0 = (U0,), where U0 is a key of R.***

- R_0 = (IKG, {}), or, (KLG, {}), (JKG, {}), (KDG, {})

3.2 (5 points)

Check if all the resulting relations are in BCNF. If you find a schema that is not, provide a lossless BCNF decomposition of it. To get full marks, you have to explain all the steps.

Sample Solution

The resulting relations are all in BCNF; therefore no decompositions are required.

Problem 4 (20 points)

Consider a relation schema R with the set U = {L,M,N,O,P,Q,S,R} of attrbutes and the set of functional dependencies F = S → M, LM → O, OP → N, SQ → R, MO → P, LR → Q, NP → L, LO → R, SQ → O. Prove that the decomposition R_1(L,M,N,O), R_2(N,O,P,Q), R_3(P,Q,S,R) of this schema is in BCNF.

Sample Solution

The decomposition is in BCNF if R_1(L,M,N,O), R_2(N,O,P,Q), R_3(P,Q,S,R) are all in BCNF w.r.t. F_1, F_2 and F_3, where F_1, F_2 and F_3 are restrictions of F to R_1, R_2 and R_3 respectively. (Recall the restriction of F to R_i is the set of all FDs in F^+ that include only attributes of R_i).

So in order to check if R_i is in BCNF, we need to first obtain F_i, which requires to compute F^+. To compute F^+, we could, for each $\gamma \subseteq R$, compute γ^+; and for each $S \subseteq \gamma^+$, output a FD $\gamma \rightarrow S$. However, for a schema of 8 attributes, we there are 2^8 subsets of R, and we need to computer attribute set closure for 2^8 - 1(empty set) - 1 (R itself) = 254 times!

A better way is to obtain F_i directly without computing F^+. To do that, we compute the closure for subsets of R_i instead, which requires only 2^4 - 1(empty set) - 1 (R itself) = 14 computations for each R_i.

For R_1(L,M,N,O)

- Closures of attribute sets:
  - $(L)^+{}_F$ = L
  - $(M)^+{}_F$ = M
  - $(N)^+{}_F$ = N
  - $(O)^+{}_F$ = O
  - $(LM)^+{}_F$ = LMNOPQR
  - $(LN)^+{}_F$ = LN
  - $(LO)^+{}_F$ = LOR
  - $(MN)^+{}_F$ = MN
  - $(MO)^+{}_F$ = LMNOPQR
  - $(NO)^+{}_F$ = NO
  - $(LMN)^+{}_F$ = LMNOPQR
  - $(LMO)^+{}_F$ = LMNOPQR
  - $(LNO)^+{}_F$ = LNOR
  - $(MNO)^+{}_F$ = LMNOPQR
- nontrial FDs in F_1:
  - LM $\rightarrow$ NO
  - MO $\rightarrow$ LN

The LHS of both FDs is a candidate key for R_1(L,M,N,O); so R_1(L,M,N,O) is in BCNF.

For R_2(N,O,P,Q)

- Closures of attribute sets:
  - $(N)^+{}_F$ = N
  - $(O)^+{}_F$ = O
  - $(P)^+{}_F$ = P
  - $(Q)^+{}_F$ = Q
  - $(NO)^+{}_F$ = NO
  - $(NP)^+{}_F$ = LNP
  - $(NQ)^+{}_F$ = NQ
  - $(OP)^+{}_F$ = LNOPQR
  - $(OQ)^+{}_F$ = OQ
  - $(PQ)^+{}_F$ = PQ
  - $(NOP)^+{}_F$ = LNOPQR
  - $(NOQ)^+{}_F$ = NOQ

- $(NPQ)^+{}_F$ = LNPQ
- $(OPQ)^+{}_F$ = LNOPQR

- nontrial FDs in F_1:
  - OP → NQ

The LHS of both FDs is a candidate key for R_2(N,O,P,Q); so R_2(N,O,P,Q) is in BCNF.

For R_3(P,Q,S,R)

- Closures of attribute sets:
  - $(P)^+{}_F$ = P
  - $(Q)^+{}_F$ = Q
  - $(S)^+{}_F$ = MS
  - $(R)^+{}_F$ = R
  - $(PQ)^+{}_F$ = PQ
  - $(PS)^+{}_F$ = MPS
  - $(PR)^+{}_F$ = PR
  - $(QS)^+{}_F$ = LMNOPQSR
  - $(QR)^+{}_F$ = QR
  - $(SR)^+{}_F$ = MSR
  - $(PQS)^+{}_F$ = LMNOPQSR
  - $(PQR)^+{}_F$ = PQR
  - $(PSR)^+{}_F$ = PSR
  - $(QSR)^+{}_F$ = LMNOPQSR
- nontrial FDs in F_1:
  - QS → PR

The LHS of both FDs is a candidate key for R_3(P,Q,S,R); so R_3(P,Q,S,R) is in BCNF.

Submission instructions

You can work in groups of 2 people, with only one submission per group.

The electronic submission must contain the following files (with these exact names):

1. **group.txt** - text file, with information about all members of the group:
   - Last name:
   - First name:
   - Student number:
   - Electronic submission account:
2. **a4.<extension>** - your solution to this assignment
   - you can use one of the acceptable formats (**.ps, .pdf or plain ASCII file**); we prefer Postscript or PDF submissions (you may lost points for clarity and readability reasons). Submissions in other format will not be marked.
   - It is very important that you include in this file the information from group.txt as well (suggestion: include it in the first page of your file).
3. Check the assignments section on the course website for using *submit* command.

Suggestions

You can use any editor you like, or you are familiar with. If you want to edit your assignment in LaTeX, you can find some useful information on the course website. You can use the example file provided there as a reference. If you choose to copy some examples from there, please acknowledge this in your assignment.