Distributed Database
Exercise 6
Fabio Porto
Winter 2005-2006

Consider the conceptual model diagram, in figure 1:

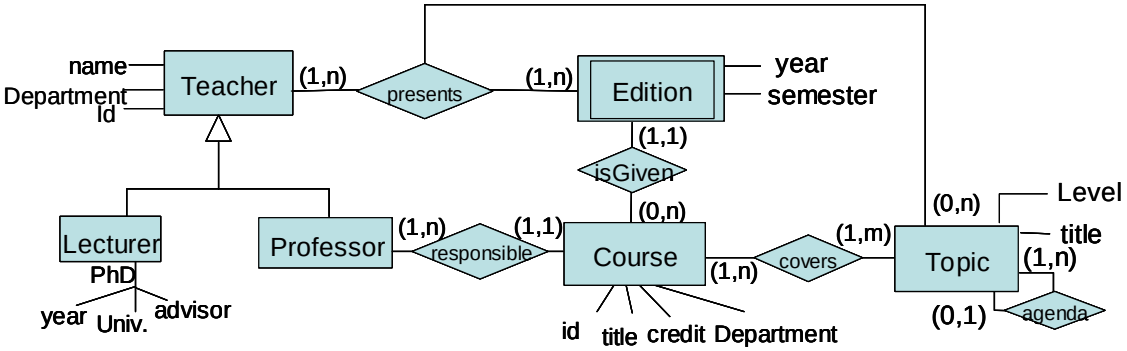

Figure 1: Conceptual model diagram - EER

Legend:

- rectangles represent real-world entities;
- diamonds linking real-world entities represent associations between elements of these entities;
- (a,b) representation in diamond associations specifies that each element of the closest real-world entity participates in a minimum of *a* and a maximum of *b* associations;
- a small line associating a real-world entity with a term represents an attribute;
- the *presents* association represents a ternary association $<e_1,e_2,e_3>$ among real-world entity elements;
- the *agenda* association expresses the Topics (sub-level topics) that are covered in a certain Topic (1^{st}level topic). For instance: Topic: Distributed Databases; Agenda: <Topic: DDB Architecture; Topic: DDB Design>
- the *presents* association refers only to 1^{st} level Topics;

Consider that the diagram represented in figure 1 is implemented in databases physically available in three institutions I={A,B,C}. *A* and *B* have designed and deployed the database as part of a common project in which users in both campi share information agreed on the database design. The database design provides a single view of data with unified constraint enforcement. Institution *C* designed a different logical model and implemented it independently of the others. Sometime later on, *A, B and C*

decided to share (read only) their data by offering applications a common view on top of their databases. Given this scenario, answer the following questions:

1) Identify an adequate distributed database architecture that would be suitable for this scenario. Provide a list of its main components, with a description of their functionality.
2) How would you classify this system in respect to data distribution, autonomy and heterogeneity? Justify your answer.
3) Suppose we want to devise a distributed database design to be deployed by institutions A and B, for the model represented in figure 1. Consider a top-down approach and the adoption of the Relational model for global concepts representation. The following tasks must be accomplished:
    a. Design a centralized relational schema;
    b. Design a dependency graph where nodes are relations and directed edges indicates a dependency between two relations (pk-fk);
    c. Identify relevant simple predicates based on elected application requirements;
    d. Identify fragments;
    e. Apply derived fragmentation and produce DHF;

The distributed database design will initially consider horizontal fragmentation (primary and derived). It shall be based on qualitative requirements as specified by the most relevant applications: the academic service application, the accounting application and the curriculum application. Their data requirements are specified below:

1) Academic application
    a) Manages data in each institution separately. Institution A={dept1,dept2,dept3}; institution B={dept4,dept5};
2) Accounting application
    a) Manages courses based on the number of credits they offer. Short=[1,3] and long=[4,8]
3) Curriculum application
    a) Manages topics according to their level: fundamental, intermediary and advanced;
4) Define a vertical fragmentation strategy based on the following application information:
    a) Queries:

i) Select idteacher, PhDYear,PhDUniv, PhDAdvisor
From Teacher
Where deptid in { }

ii) Select idteacher, name
From Teacher
Where level= ‘X‘

b) Access patterns for qi and qii:

i) qi and qii running frequencies equal to 1;

ii) accesses by site:

I. q_1) acc_A= 15 ; acc_B=35

II. q_2)acc_A=0;acc_B=70

5) Once you have designed your database, implement it in Oracle10g using