

CS585
Database Systems
Fall 2013
Exam I

Name: _____

Student ID: _____

	Maximum	Received
Problem 1	20	
Problem 2	25	
Problem 3	25	
Problem 4	15	
Problem 5	15	
Total	100	

1hr 50 minute exam. One 8.5X11 cheat sheet allowed.

1) 20 pts

Indicate whether each of the following statements is true or false (T/F):

T A weak entity set can participate in more than one relationship set

F If an entity set has a relationship with a weak entity set, then that relationship must be a total participation on the weak entity side

T A weak entity set might have derived attributes

F The number of children at each node of the kd -tree depends on the d (number of dimensions)

F Exhaustive search method of splitting MBRs in an R-tree can be expensive but it guarantees the two MBR's found after the split will NOT be overlapping

T Linear cost algorithm for splitting MBRs in an R-tree is relatively inexpensive and may result in overlapping MBR's after the split

F *Distance* between two spatial objects is considered to be a topological relationship in Rose Algebra

F An Object-Relational database (ORDBMS) is an example of a NoSQL type database

T An Object Oriented database (OODBMS) is an example of a NoSQL type database

F A pure relational database (RDBMS) cannot be used to model inheritance

2) 25 pts

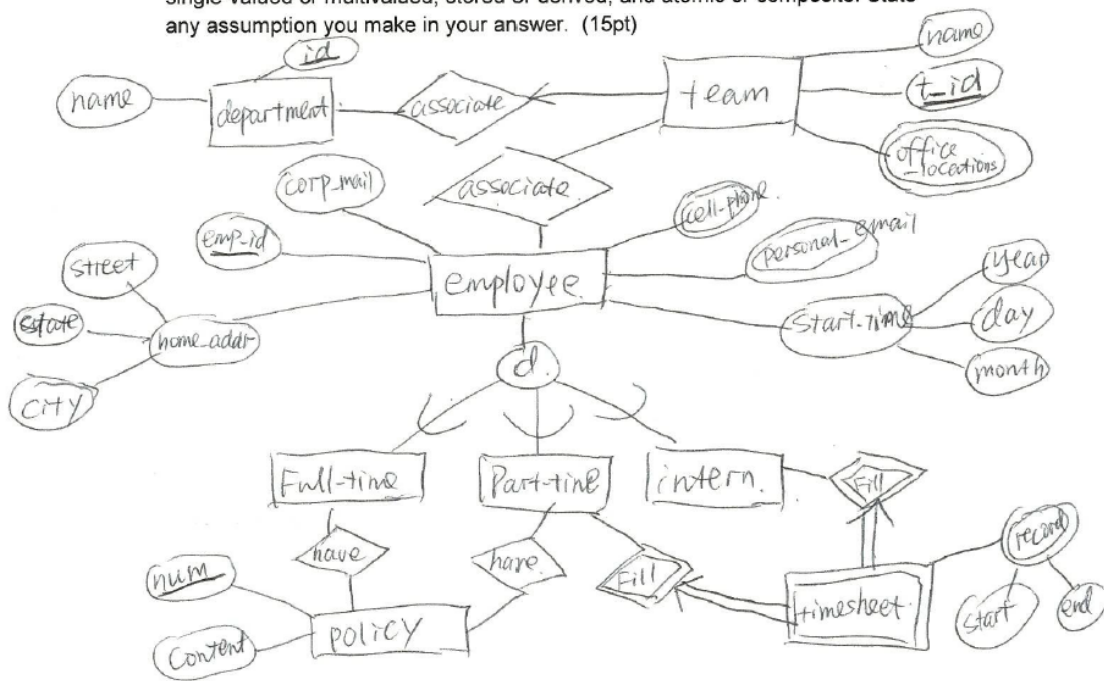
Suppose you want to design a database for your company. It will contain the following information:

1. The company has many departments. Every department has its name and a unique department id.
2. There are teams associated with each department. Every team has its name, team id, and office locations.
3. Every team has its employees. Employees could be associated with multiple teams.
4. Employees have such information as: a unique company-wide employee id, a company email address, personal email address(s), home address, start time, cell phone number(s).
5. Employees should belong to one of the three employee types: full-time, part-time, intern.
6. Full-time employees and part-time employees have insurance policy information, while interns do not have such information. Insurance policies contain a policy number that is used to identify the policy, and the content of the policy.
7. Part-time employees and interns have to fill out timesheets to record their working time. A timesheet contains records of start time and end time. Full-time employees do not need to fill out such timesheets.

- a) Draw a ER/EER diagram to show the conceptual design of the database. Indicate all subclasses, relationships, relationship cardinalities, total participation, attributes, primary keys if necessary. In addition, specify whether each attribute is single-valued or multivalued, stored or derived, and atomic or composite. State any assumption you make in your answer. (15pt)

(Multiple solutions possible depending on your assumptions)

single-valued or multivalued, stored or derived, and atomic or composite. State any assumption you make in your answer. (15pt)



- Assumption:
- ① Each employee has only one policy.
 - ② Team id is unique.
 - ③ Home address contains state, city, street.
 - ④ Start-time of an employee includes year, month, day.
 - ⑤ Policy id is unique.

- b) Map your conceptual design to the pure relational model. (i.e., No Object Oriented or Object Relational). Be sure to identify all integrity constraints.

Department (name, id)

Team (department-id, name, t-id, office-locations)

department-id is a FK^{that} references Department.

department-id is not NULL.

Employee (emp-id, team-id, cell-phones, corp email, personal-emails, employee-type)

Employee_start_time (emp-id, year, day, month)

emp-id is a FK that references Employee

Employee_home (emp-id, street, state, city)

emp-id is a FK that references Employee

Policy (policy-id, emp-id, content) emp-id is a FK that references Employee

Time-sheet_record (emp-id, start-time, end-time)

emp-id is a FK that references Employee

3) 25 pts

Consider the following following table schemas. (25pt)

Company(company_name, year_founded)

HireStats (company_name, number_of_new_hire, year)

In the table company, year_founded stands for the year the company is founded.

HireStats tables record the statistics of the hiring for each company. Company_name in HireStats is a foreign key which references the table Company.

Number_of_new_hire is not null, and is always a positive number.

Primary keys in tables are underlined.

a) Create the two tables (with required constraints). (10pt)

```
Create table company (company_name char(255),  
                      year_founded integer,  
                      primary-key company_name,
```

```
Create table HireStats (company_name char(255),  
                      number-of-new-hire integer  
                      not null,  
                      year integer  
                      primary key (company_name,  
                      year),
```

```
Foreign key (company_name)  
  references company  
  on delete cascade  
  on update cascade  
  check (number-of-new-hire  
         > 0))
```

b) Find the average hiring number for company name "Apple" between year 2011 and 2012. (5pt)

```
select avg(number_of_new_hire) from HireStats where company="Apple" and  
((year=2011) or (year=2012))
```

c) Find the company names that were founded before 2000 and hired more than 60 people in 2012. (5pt)

```
select C.company_name from Company C, HireStats H where C.company_name =  
H.company_name and H.number_of_new_hire>60 and H.year=2012 and  
C.year_founded<2000
```

d) Find the total number of people hired for each company since it was founded. (5pt)

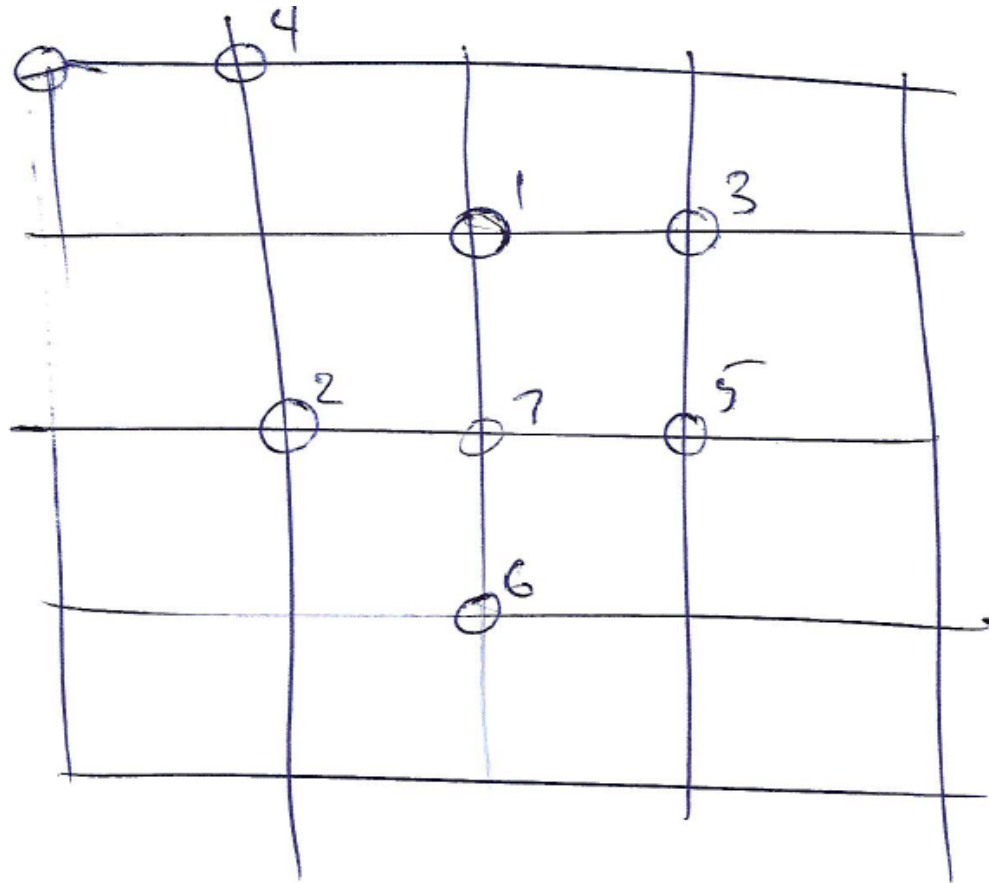
```
select sum(number_of_new_hire) from Company C, HireStats H where  
C.company_name = H.company_name and H.year>=C.year_founded group by  
C.company_name
```

4) 15 pts

The following 6 points are inserted into a spatial DBMS that uses an R-Tree of size (2,4) as an index structure to store points. Show the R-Tree after each insert in the given order. Use the linear time algorithm described in class for splitting MBRs. Show all your work, i.e you will need to draw all 7 R-trees and describe the splitting method you used.

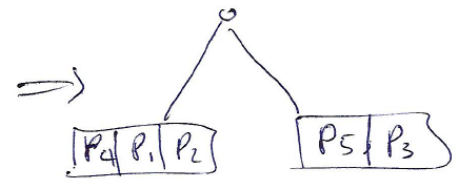
Point #	X	Y
1	2.	3.
2	1.	2.
3	3.	3.
4	1.	4.
5	3.	2.
6	2.	1.
7	2.	2.

(This problem may have multiple solutions if the minimum enlargement is tied.)



(4,5) gives the biggest MBR.

$$\begin{array}{r} \text{MBR1} \\ 4 \\ \hline 1,2 \\ \hline (4,1,2) \end{array} \quad \begin{array}{r} \text{MBR2} \\ 5 \\ \hline 3 \\ \hline (5,3) \end{array}$$



point 6 →



point 7 →



must split MBR1

(4,6) gives the biggest MBR

$$\begin{array}{r} \text{MBR1a} \\ 4 \\ \hline 1,2 \end{array}$$

$$\begin{array}{r} \text{MBR1b} \\ 6 \\ \hline 7 \end{array}$$

⇒



5) 15 pts

Short answer questions

a) Briefly describe two benefits of using views?

1. Base tables can be changed without having to change the applications that use the views. This gives us flexibility in modifying/extending/enhancing the database design without affecting the applications that use the DB.
2. Can be used to provide access to only specific subset of data held in the database. The view can be used as a filter that shows only data that the user is allowed to see, or show aggregation of data without the details of the data.

b) Give examples of 3 types of NoSQL databases and which applications they are used for

Object databases: OODBMS

Applications with complex object models.

Applications where data is found by navigating thru objects

Mobile applications

Applications where there would be a very high overhead on managing foreign keys

Graph databases:

Applications where underlying data model is a graph.

Social networks

Maps

Key-value stores:

Applications where the data elements consist of a key and an associated value or values such as a Dictionary type application

Document stores:

Applications where the data stored consists of various types of documents

Digital asset management systems

Document management systems

- c) Describe how Rose Algebra helps determine the topological relationship between two simple regions

It breaks each regions into their boundary (B) and interior (I).It then determines if there is an intersection between

I1 and I2

I1 and B2

B1 and I2

B1 and B2

Based on the results of the 4 intersections (intersects=1, does not intersect=0), it can determine the topological relationship to be one of the 6 possible relationships:

(0,0,0,0) disjoint,

(1,1,1,1) overlap,

(1,0,0,1) equal,

(0,0,0,1) touch,

(1,1,0,0) or (1,0,1,0) in

(1,1,0,1) or (1,0,1,1,) cover