CS585
Database Systems
Summer 2005
Final Exam

Name: _____
Student ID: _____

	Maximum	Received
Problem 1	28	
Problem 2	12	
Problem 3	15	
Problem 4	20	
Problem 5	15	
Problem 6	10	

Note: Students can bring a one-sided 8.5”x11” sheet of notes to the exam.

1) Short answer questions
   a. 4pts
      Give two reasons why a standard RDBMS is not suitable to store and retrieve large quantities of spatial data.

   b. 6 pts
      For each of the indexing structures below, give an example of a query (e.g. in SQL) for which the index structure is optimal.
      - B+ Tree

      - KDB Tree

      - R+ Tree

c. 4 pts

State whether each of the following is defined or declared within an XML schema:

- Types _____
- Elements _____
- Attributes _____
- Attribute groups _____

d. 4 pts

When designing an XML document, state two reasons why you would declare information as an attribute as opposed to an element.

e. 2 pts

Describe how a distributed DBMS could be more economical than a central one.

f. 2 pts
Give two reasons why you would replicate the data at local nodes in a distributed database.

g. 2 pts
Give two reasons why you would NOT replicate the data at local nodes in a distributed database.

h. 4 pts
While searching for an object in an R-tree, one may have to go thru the entire set of data in the worst case. Explain why. How does an R^+-Tree address this problem?

2) 12 pts
Write an XML schema which is equivalent to the DTD element declarations below. Make your attributes reusable if possible.

```
<!ELEMENT SalesOrder (Customer,Item+,Contract?,Promotion*)>
<!ELEMENT Customer (#PCDATA)>
<!ELEMENT Item (#PCDATA)>
<!ELEMENT Contract (#PCDATA)>
<!ELEMENT Promotion (#PCDATA)>
<!ATTLIST SalesOrder Quatity CDATA #REQUIRED>
```

3) 15 pts
Given the XML documents users.xml, items.xml, and bids.xml with the abbreviated set of data showing the XML format of the instances,

```
<items>
  <item_tuple>
    <itemno>1001</itemno>
    <description>Red Bicycle</description>
    <offered_by>U01</offered_by>
    <start_date>1999-01-05</start_date>
    <end_date>1999-01-20</end_date>
    <reserve_price>40</reserve_price>
  </item_tuple>
  <!-- !!! Snip !!! -->

<users>
  <user_tuple>
    <userid>U01</userid>
    <name>Tom Jones</name>
    <rating>B</rating>
  </user_tuple>
  <!-- !!! Snip !!! -->

<bids>
  <bid_tuple>
    <userid>U02</userid>
    <itemno>1001</itemno>
    <bid>35</bid>
    <bid_date>1999-01-07</bid_date>
    </bid_tuple>
  <bid_tuple>
  <!-- !!! Snip !!! -->
```

a- Write an XQuery expression to find cases where a user with a rating worse (alphabetically, greater) than "C" is offering an item with a reserve price of more than 1000. For each case create a warning that includes user name, rating, and

reserve price.

b- Write an XQuery expression to find the userid, name, number of bids, and average bid, for each user who has placed a bid, in order by userid.

4) 20 pts
Given the relations EMP, ASG, and PROJ

EMP

ENO	ENAME	TITLE	SALARY
E1	J. Doe	Elec. Eng	95,000
E2	M. Smith	Syst. Anal.	105,000
E3	A. Lee	Mech Eng.	85,000
E4	J. Miller	Programmer	125,000
…			

ASG

ENO	PNO	RESP	DUR
E1	P1	Manager	12
E2	P2	Analyst	24
E3	P3	Analyst	6
…			10

PROJ

PNO	PNAME	BUDGET	Due Date
P1	Instrumentation	1,000,000	12-10-07
P2	Database Dev	2,000,000	10-1-06
P3	CAD/CAM	1,500,000	10-1-07
…			

And the query

```
SELECT    EMP.ENO
FROM      EMP  E, ASG  A, PROJ  P
WHERE     A.RESP = "Manager"
          AND E.SALARY > 100,000.
          AND P.BUDGET < 1,500,000.
          AND P.PNO = A.PNO
          AND A.ENO = E.ENO
```

And the following assumptions:
- Relations are all of the same size
- Salaries are between 50,000 and 150,000
- Budgets are between 100,000 and 2,000,000
- 10% of employees are managers

a- Specify an efficient order of operations to execute the above query if all data

was held centrally. You can either present a query tree or a relational algebra expression. (State any other assumptions you make)

b- Given the following distribution of relations across the network, give an

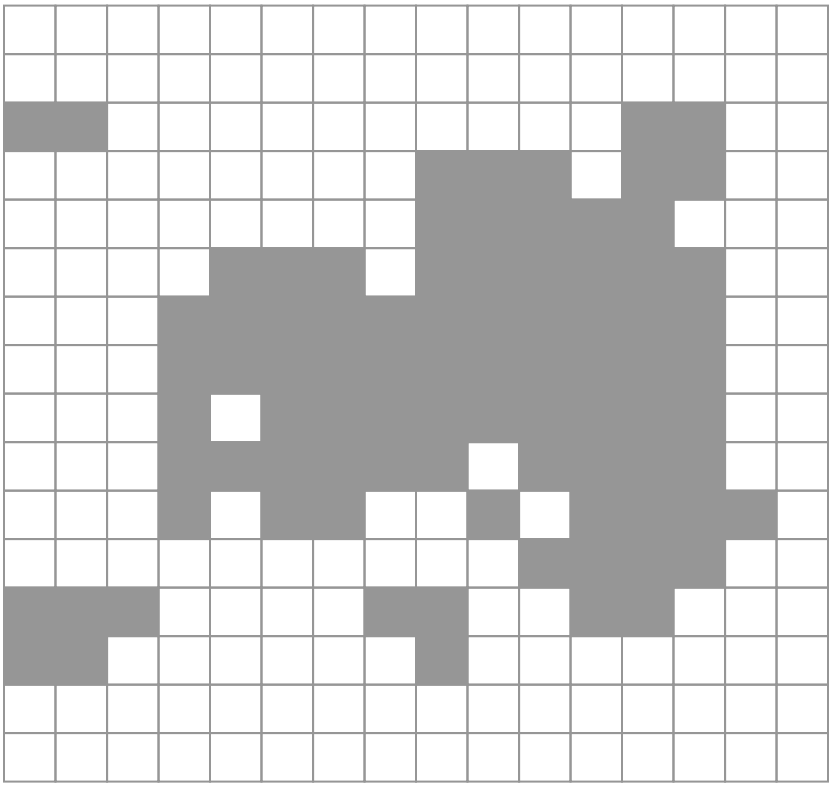
efficient order of operations to execute the same query. Use a query tree and show all transfer of data from site to site.
- EMP tuples are equally split between sites 1 and 2. Site 1 gets all ENO>200
- ASG tuples are equally split between sites 3 and 4. Site 3 gets all ENO>200
- PROJ tuples are equally split between sites 5 and 6. Site 5 gets all PNO>50

c- To improve the efficiency of the above query, if you were allowed to replicate

one of the relations once--in other words you could have had another copy of a relation either at one site or split across a number of sites--which relation would you replicate and how (i.e. would you keep the entire copy at one site or would you split it and if so how would you split it, and at which site(s) would you keep it)

5) 15 pts
Show how the gray region below can be represented thru a region Quad Tree.

a- Mark region indexes on the diagram
b- Show the resulting index of the region using a Quad tree



6) 10 pts
Consider the following 2-dimensional cube. Each cell defines the sales forecast in

Mill dollars for each region over a period of 5 months.

	Jan	Feb	March	April	May
Europe	20	10	20	10	10
N. America	10	20	30	20	20
S. America	10	10	20	10	10
Africa	10	20	20	10	10
Asia	20	10	20	10	20

a- Identify measure attributes and dimension attributes

b- Compute the corresponding prefix sum cube

c- Using the results from part b compute the “percentage of all sales for the months of February thru April that are expected from North and South America”. Show all your work.

Additional pages