CSCI585 - Database Systems
Spring 2008

Midterm Exam

Please Complete the Following:

Name	
Student ID #	
Location (if remote)	

Notes:

1. ***Duration of the exam is <u>1 hour and 15 minutes</u>***
2. ***Total number of pages including this page: 21***
3. ***You may not have enough time to finish the entire exam!***
   Look over the entire questions and answer the easy ones first.

For grader use only:

	Maximum	Received
Problem 1	12	
Problem 2	14	
Problem 3	12	
Problem 4	18	
Problem 5	10	
Problem 6	20	
Problem 7	14	
Total	100	

Problem 1: (12 Points)

Indicate whether each of the following statement is true or false (T/F): (1.5 points each)

_T___ 1. Entity sets are weak when their key attributes come from other classes to which they are related.

_F_ 2. All ternary relationship can always be reduced to two or three binary relationships.

_F__ 3. The degree of a relation instance is its number of tubles.

_T__ 4. Foreign keys can be null

5. Define the following terms

Domain Constraint: is to specify that each attribute A must be an atomic value from the domain dom(A) such as varchar2, number

Entity Integrity Constraint: is a rule (or check) that is enforced by the DBMS to ensure that incorrect data is not stored in the database.

6. Define each property in A-C-I-D principle

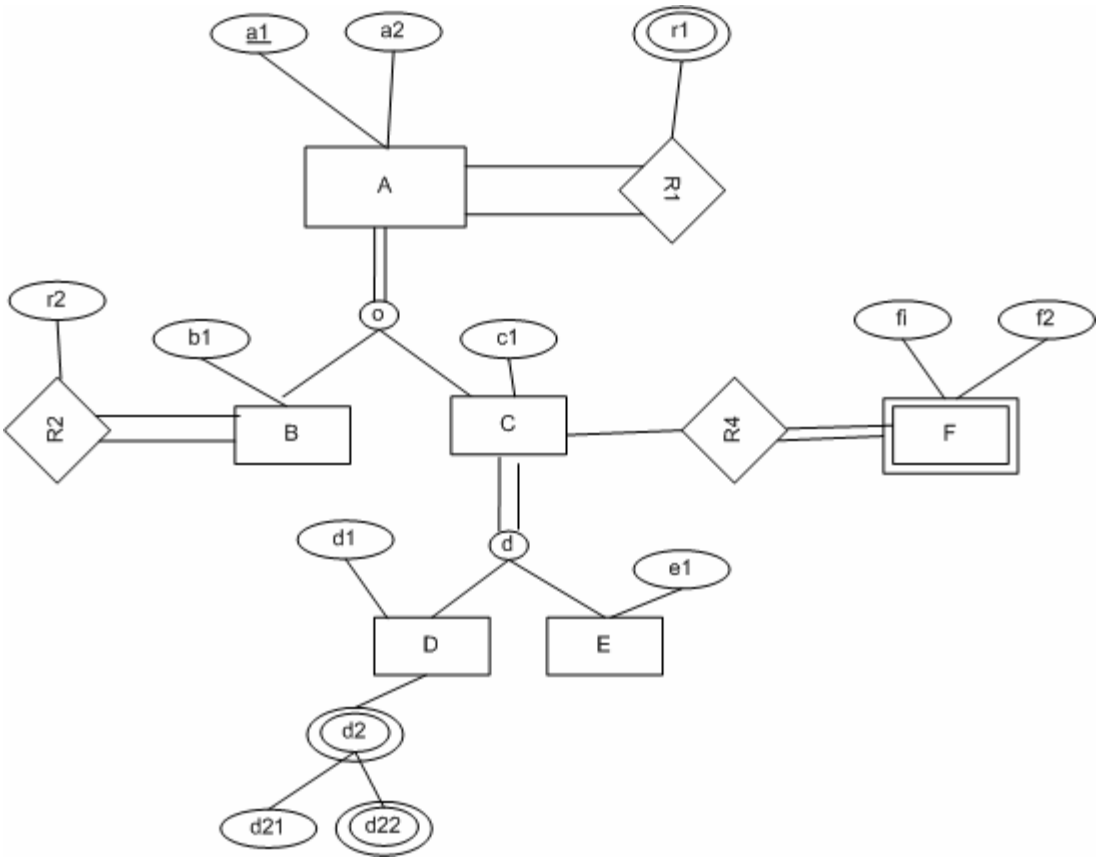
Atomicity, Consistency, Isolation, Durability

_F__ 7. Indexing spatial data with an R+-tree is always advantageous when compared with indexing by an R-tree.

_T _ 8. A PR-Quadtree can become extremely unbalanced. This may defeat the objective of indexing.

Problem 2: (14 points)

Reduce the given EER diagram to relations using pure relational model (i.e., No Object Oriented or Object Relational). Be sure to identify all primary and foreign keys.



Solution; **A [a1, a2] , R1 [A.a1,A.a1,r1]**
B **[A.a1, b1] , R2 [A.a1, A.a1, r2]**
C **[A.a1, c1], R4 [A.a1, f1]**
D[**A.a1, d1 ,d21], D2 [DA.a1,d22]**
E **[A.a1, e1]**
F **[f1, f2]**

Problem 3: (12 points)

There are 3 tables describing the student, department and student department relationship. The primary key of each table is the attribute(s) underlined. Other attributes are not necessarily unique.

Student

SID	SName	GPA	Age
001	Jane	3.7	22
002	Tommy	2.8	27
003	Mike	3.2	30
004	Julio	3.0	25
005	Ken	3.0	23

Student_Department

SID	DID	StartDate	EndDate
001	D1	1/29/08	null
002	D2	5/10/05	null
003	D1	1/10/06	null
003	D2	5/14/07	null
004	D3	2/1/04	null
005	D3	2/1/04	null
005	D2	5/20/06	null

Department

DID	DName
D1	CSCI
D2	EE
D3	BA

For each query, write the output and give a brief description of the query

(a) (4 Points)

```
select s.sname, d.dname, sd.startdate
from student s, student_department sd, department d
where s.SID=sd.SID and d.DID = sd.DID
order by 1,3
```

Jane,CSCI,1/29/0008
Julio,BA,2/1/0004
Ken,BA,2/1/0004
Ken,EE,5/20/0006
Mike,CSCI,1/10/0006
Mike,EE,5/14/0007
Tommy,EE,5/10/0005

Description: Retrieve the student name, department and startdate (for each department) of each student

(b) (4 Points)

```
SELECT DID,Dname
FROM department
WHERE DID not IN (SELECT DID FROM student_department
GROUP BY did HAVING Count(*) > 2)
```

Output D1 CSCI
D3 BA

Description: Retrieve the did and name of the departments where total registration of students less than 3

(c) (4 Points)

```
select s.SNAME, s.gpa, d.DNAME
from student s,  student_department sd, department d
where s.SID= sd.SID and d.did=sd.did
and s.gpa= (select min(gpa) from student)
```

Output: Tommy,2.8,EE
Description: Retreive the gpa,department name , student name who has min gpa among all students.

Problem 4: (18 points)

Consider the relational conceptual database schema below for keeping track of shows of artists in galleries:

ARTIST (SSN, ArtistName, Category, Age, Rating)
ARENA (ANo, AreanaName, Address)
PERFORM (SSN, ANo, StartDate, Duration)

Here, ARTIST contains the SSN, artist name, category, age and rating (1 to 10) of the artist. ARENA contains a tuple for each arena #, arena name, and the address of the arena. PERFORM keeps the relation between artists and arenas.

(a) Retrieve the name and age of each artist who performed in the arena "Staples Center". Order your results by increasing age (4 points)

```
select a.artistname,a.age
from artist a, arena ar, perform p
where a.ssn=p.ssn and p.ano=ar.ANO
and ar.ARENANAME='Staples Center'
order by age asc
```

(b) Find the average age of artists who performed at least 3 times in "Staples Center" (4 points)

```
select avg(age)
from artist a
where a.ssn in (
select ssn
from arena ar, perform p
where ar.ANO=p.ano and ar.ARENANAME='a1'
group by ssn
having count(ar.ano) >=3
)
```

(c) For all artists, retrieve name and age of the artist, the arena name where the artist performed max duration of time. (5 points)

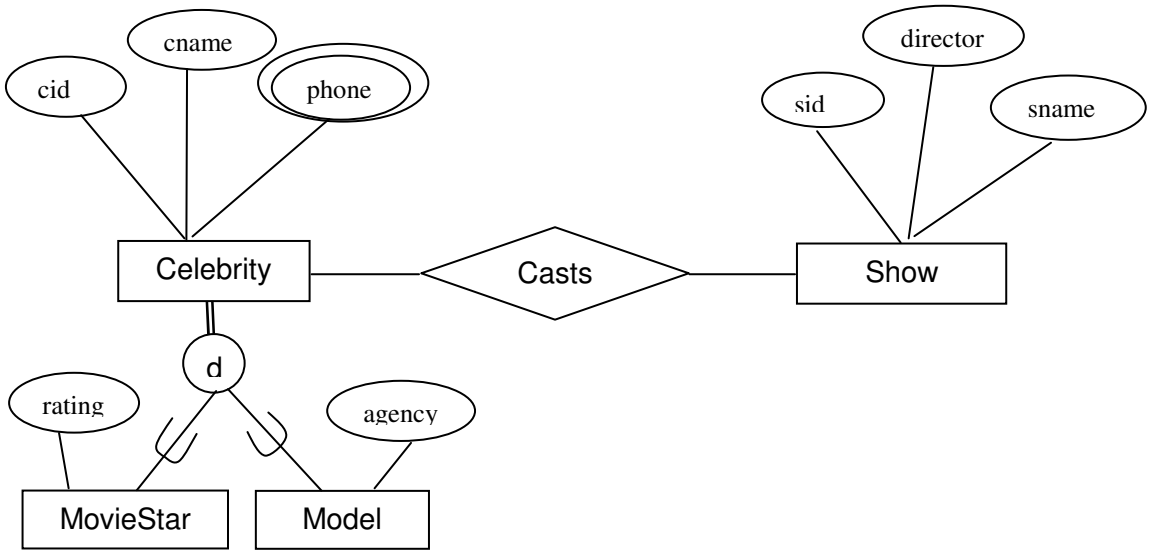
```
select a.artistname,age, ar.arenaname, duration
from artist a, arena ar, perform p
where a.ssn=p.ssn and p.ano=ar.ANO
and p.duration = (select max(duration) from perform pp where
pp.ssn=a.ssn)
```

(d) Find the artists whose rating is more than the average of the all artist. Retrieve artist's name, rating, arena name that they perform, start date of their performance. (5 points)

```
select a.artistname,rating , ar.arenaname,p.startdate
from artist a, arena ar, perform p
where a.ssn=p.ssn and p.ano=ar.ANO
and a.rating > (select avg(rating) from artist)
```

Problem 5: (10 points)

Consider the following Extended ER diagram and the corresponding reduction (Note: attributes are not shown in the diagram).




```
CREATE TYPE Celebrity AS OBJECT (cid NUMBER, Cname VARCHAR2(20),
phone phone_type)
  NOT FINAL;
CREATE TYPE MovieStar UNDER Celebrity (rating  NUMBER);
CREATE TYPE Model UNDER Celebrity (agency VARCHAR2(20));

CREATE TABLE Show (sid NUMBER, Sname VARCHAR2 (20), director
VARCHAR2 (20));

CREATE TABLE CelebrityTable OF Celebrity (PRIMARY KEY(cid));
CREATE TABLE Casts(cid NUMBER, sid NUMBER, PRIMARY KEY(cid,sid));
```

(a) When creating Celebrity table, "phone phone_type" is used as an attribute. Fill in the blanks for creating Phone_Type (3 points)

CREATE or REPLACE TYPE PHONE_TYPE .....

Solution: CREATE or REPLACE TYPE PHONE_TYPE AS VARRAY(10) OF VARCHAR2(20);

Fill in the blanks of below insert statement for celebrity table with the values cid=12, cname='Julio' , phone1=213-555-5555, phone2=213-666-6666. Note that phone is multi valued.

INSERT INTO Celebrity VALUES…….

Solution: INSERT INTO Celebrity VALUES (12,'Julio',Phone_Type('213-555-5555','213-666-6666');

(b) Write an SQL3 statement to find name, phone number(s), rating of movie stars who has played in show "Lost" (3 points)

```
select cid , phone, TREAT(VALUE(s) AS MovieStar).rating
from celebrityTable ce, Casts c, Show s
where VALUE(s) IS OF (MovieStar)
and ce.cid=c.cid  and s.sid=c.sid and s.sname='Lost'
```

(c) Write an SQL3 statement to find show name , agency and name of models who has cast in at least two shows. (4 points)

```
select s.sname , ,ce.cname TREAT(VALUE(s) AS Model).agency
from celebrityTable ce, Casts c, Show s
where VALUE(s) IS OF (MODEL)
and ce.cid=c.cid  and s.sid=c.sid and ce.cid in (select c2.cid from
casts c2 group by c2.cid having count(*)>2)
```

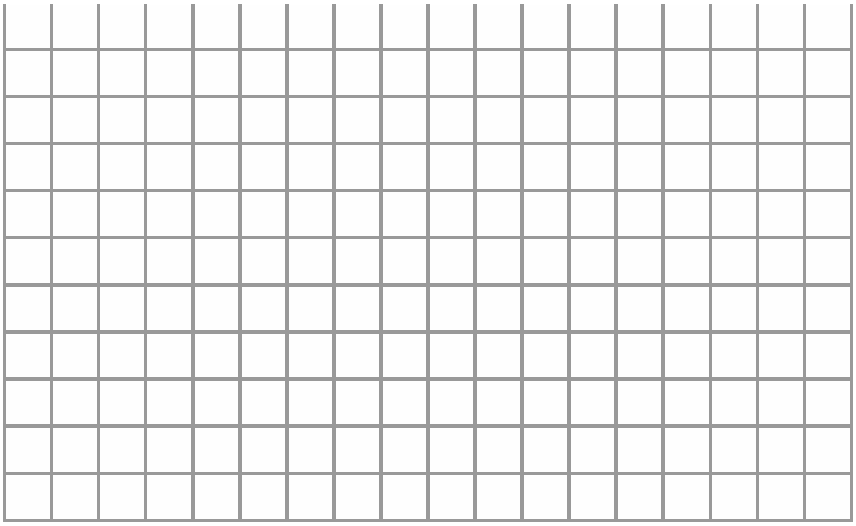
Problem 6: (20 points)

Assume the following table for the position of some of shapes in a painting which is stored in a database with spatial support.

ID	Shape	Spatial Data
S1	Circle	center: (2,2), radius:1
S2	Oval	center: (1.5, 5), r_1:0.5, r_2:1
S3	Hexagon	(3,5), (4,6), (5,6), (6,5), (5,3), (4,3)
S4	Line	(6,2), (7,3)
S5	Line	(6,1), (8,2)

Assume that the shapes are inserted in to the table in the ascending order of ID (i.e., S1, S2, S3, S4, S5). Also assume that (m,M)=(1,2).

(a) Draw the R-Tree index generated for the above table after each insertion. In other words, you should draw five R-Trees. Use "Quadratic" method to split the R-Tree Nodes. Show your computation for the intermediate steps (e.g., how to pick seeds and entries for each group). You can use the following chart to draw the shapes. (12 points)

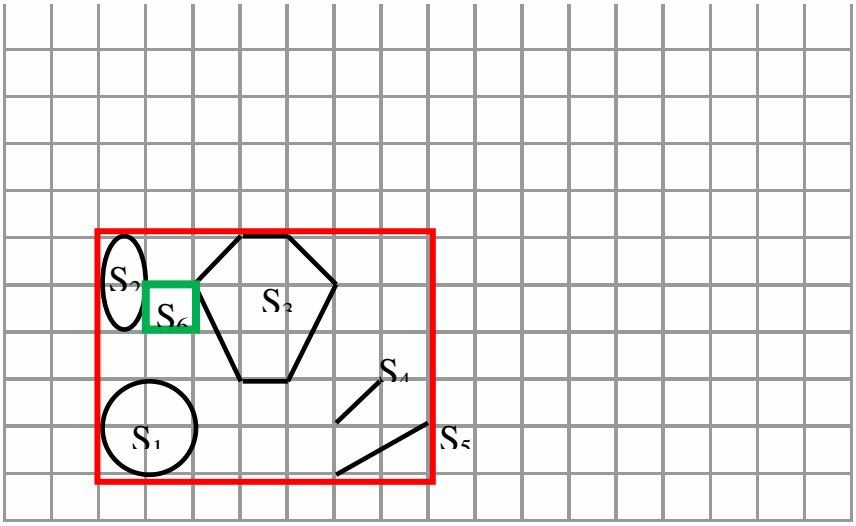


(b) Suppose the following new shape is inserted. Update the R-Tree to include this new shape. (5 points)

ID	Shape	Spatial Data
S6	Square	(2,4), (2,5), (3,5), (3,4)

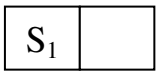
(c) In order to search for an object in R-Tree, you might have to go through several rectangles or the entire database in the worst case. Explain why? How does R+-Tree address this problem? (3 points)

Solution:
(a)

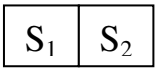


The resulting index will be:

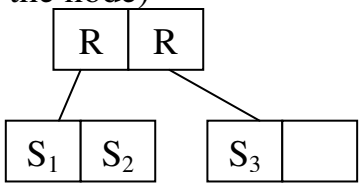
1) Insert the first one into the empty tree:



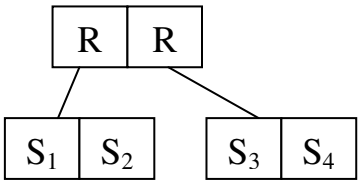
2) Insert the second one into the previous one (still have room):



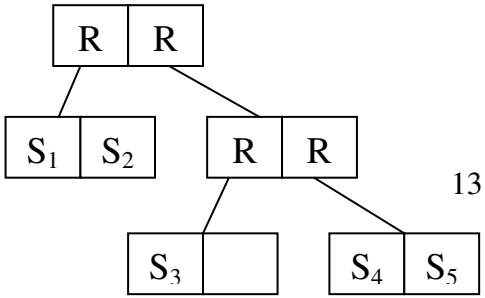
3) Insert the third one into the previous one (don't have room, need to split the node)



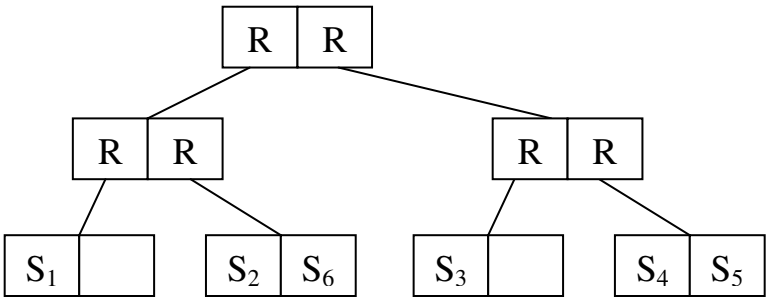
4) Insert the 4th one into the previous one (still have room):



5) Insert the 5th one into the previous one (don't have room, need to split the node)



(b)



(c)

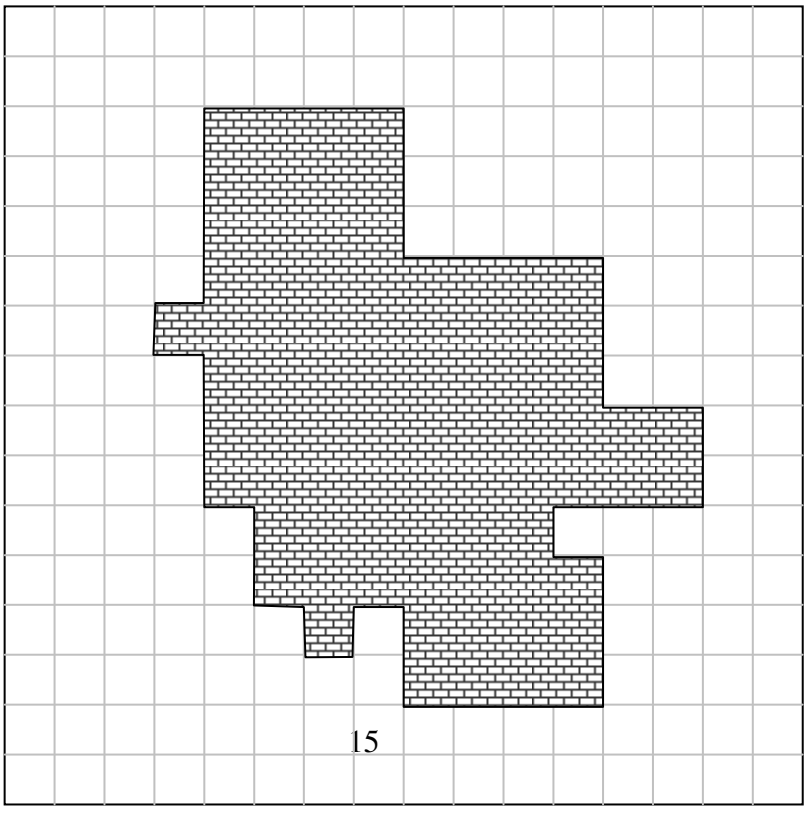
1) In R-Tree, the bounding rectangles could overlap each other, and an object is only associated with one bounding rectangle.

2) In R+-Tree, the space is decomposed into disjoint cells. Moreover, if a node overlaps with several rectangles, it would appear in all overlapping rectangles.

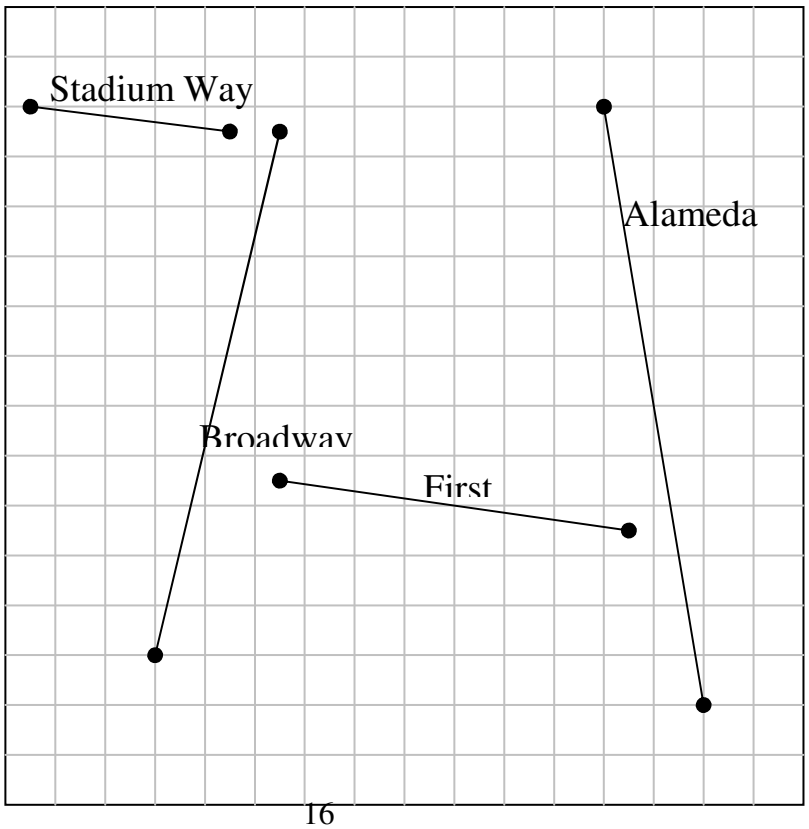
Problem 7: (14 points)

In this problem, you would need to have the knowledge of Quad Tree:

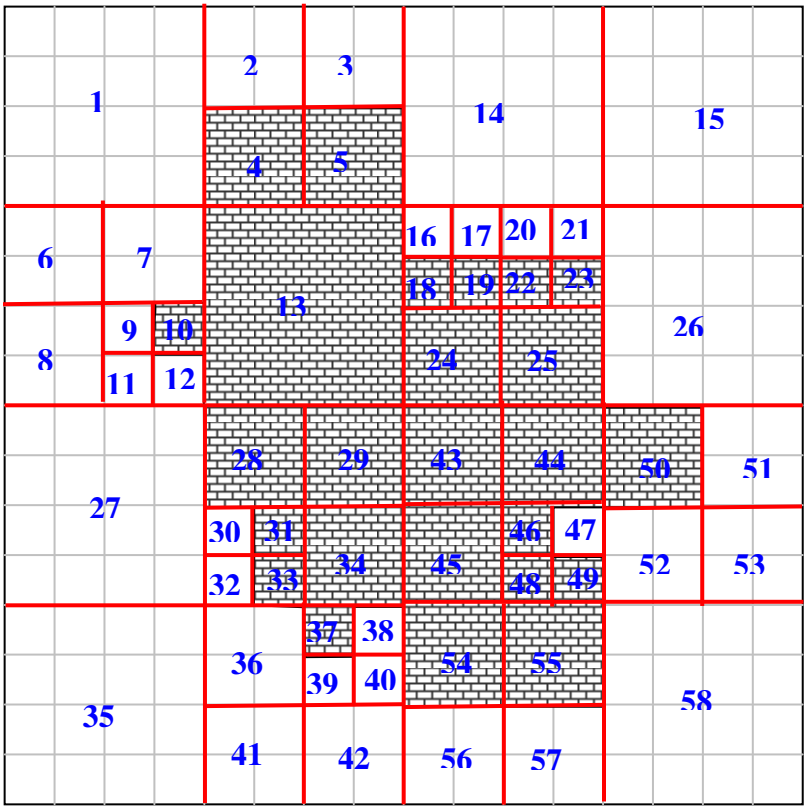
a) Please draw the resulting index of the following region by using Region QuadTree. (6 points)



b) Assume the following table for the position of some of the streets in Los Angeles is stored in a database with spatial support. Assume the streets are inserted in to the database in alphabetical order (from A to Z). Decompose the following region based on the criteria of PM1 Quadtree. (8 points)



Solution
(a)


1
2
3
14
15
4
5
6
7
16
17
20
21
18
19
22
23
13
9
10
26
8
24
25
11
12
28
29
43
44
50
51
27
30
31
46
47
32
33
34
45
48
49
52
53
37
38
36
54
55
39
40
58
35
41
42
56
57

(b)