190/200

Name: Sree Devi Muthukumar
ID#: 11293668

CS 4513 - Database Management - Fall 2013 - Dr. Le Gruenwald
MIDTERM EXAM
(Time and Date: from 1:30 PM – 2:45 PM on 10/7/2013; Maximum Score: 200 points; Closed books, notes and electronic devices)

1) (75 points):
Below is the information about a university database:

- Each professor has a social security number, a name, an age, a rank, and multiple research areas.
- There are two types of professors: tenured and tenure-track.
- For each tenure-track professor, the year in which he or she must apply for tenure is recorded. For each tenured professor, the total number of courses he or she has taught is recorded.
- Each project has a unique project number, a sponsor name (e.g. IBM), a starting date, an ending date, and a budget.
- Each graduate student has a social security number, a name, an age, and a degree program (e.g. M.S, Ph.D.).
- Each project is managed by one professor (known as the project's principal investigator).
- Each project is worked on by one or more professors (known as the project's co-investigators).
- A professor can manage and/or work on multiple projects.
- Each project is worked on by one or more graduate students (known as the project's research assistants).
- When graduate students work on a project, a professor must supervise their work on the project. Graduate students can work on multiple projects, in which case they will have a (potentially different) supervisor for each one.
- Each department has a unique department number, a department name, and a main office.
- Each department has a tenured professor (known as the chairman) who runs the department.
- Professors work in one or more departments, and for each department that they work in, a time percentage is associated with their job.
- Graduate students have one major department in which they are working on their degree.
- Each graduate student has another, more senior graduate student (known as a student advisor) who advises him or her on what courses to take.

Answer the following questions for problem 1:

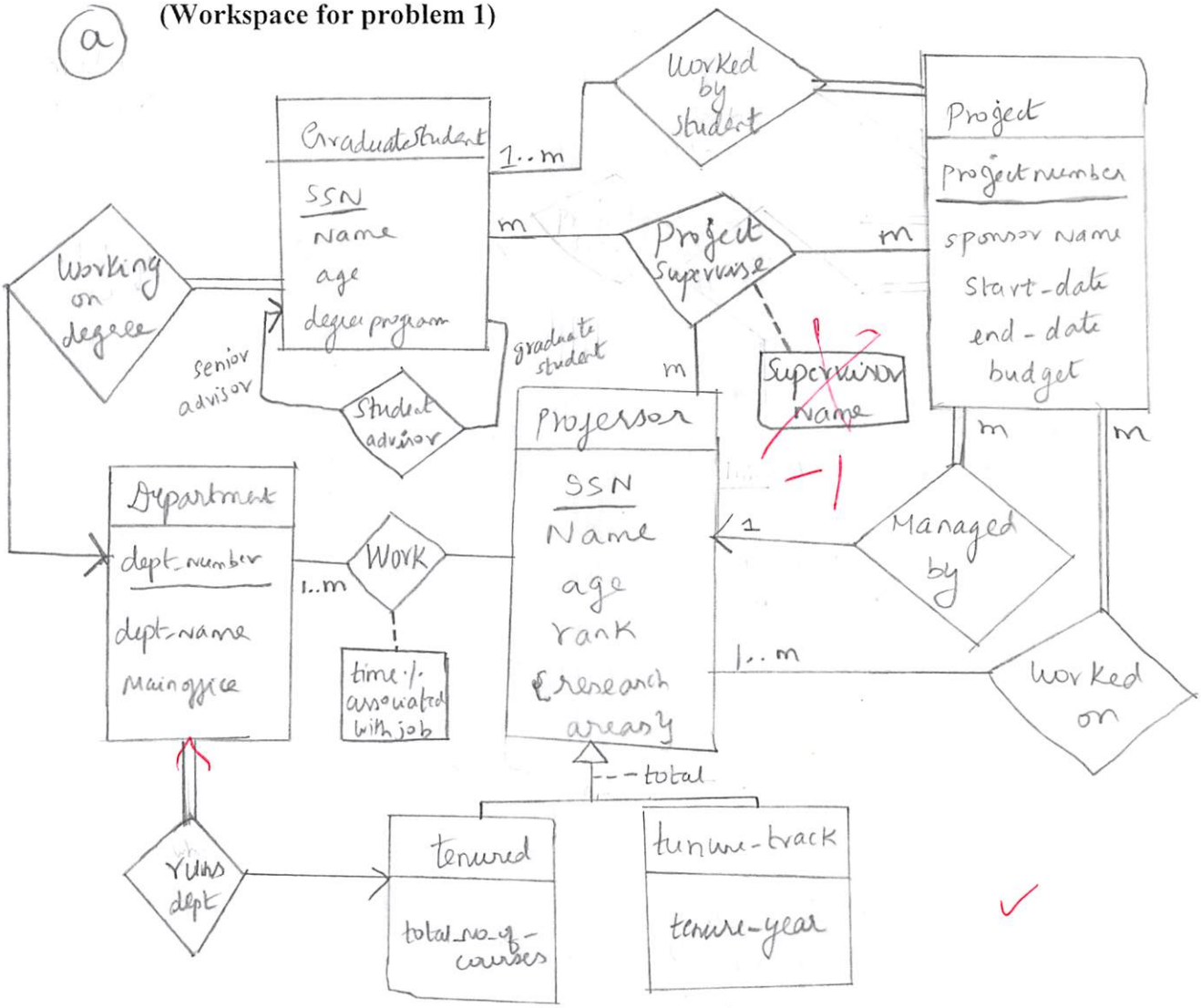
a) Construct an ER diagram that represents the database; **underline all primary keys. No additional attributes are allowed.**

b) How would your design change if the requirement analysis also includes the following:

- The time period consisting of the start date and end date during which a chairman is active is also recorded.

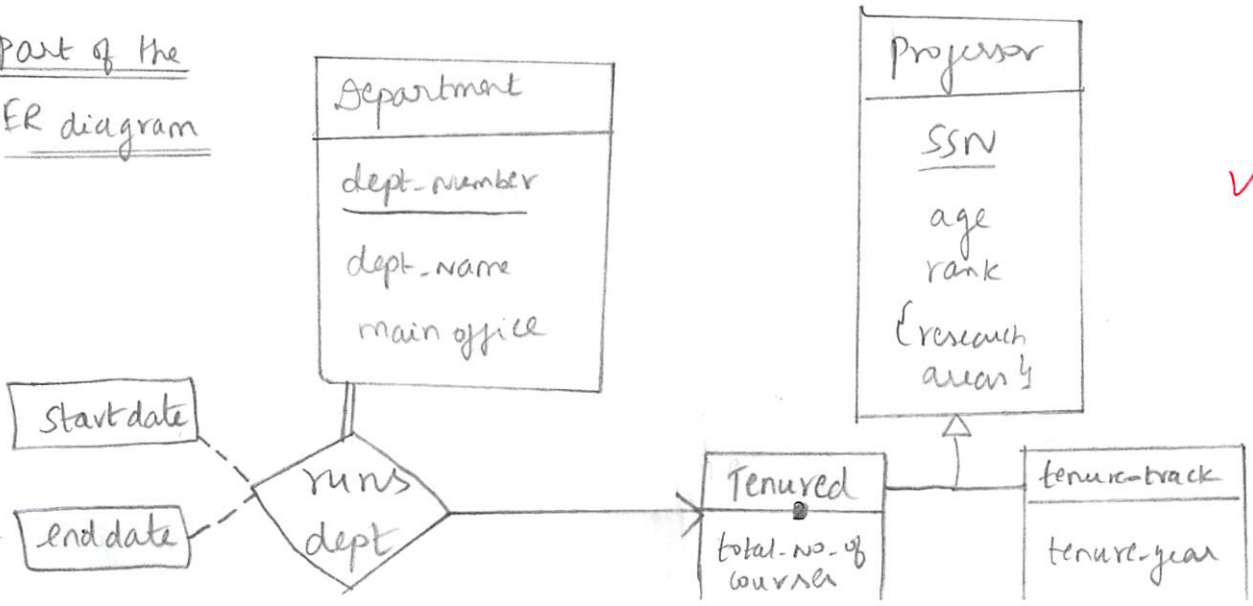
To answer part (b), draw only the part of the ER diagram that reflects the changes (if any).

(a) **(Workspace for problem 1)**

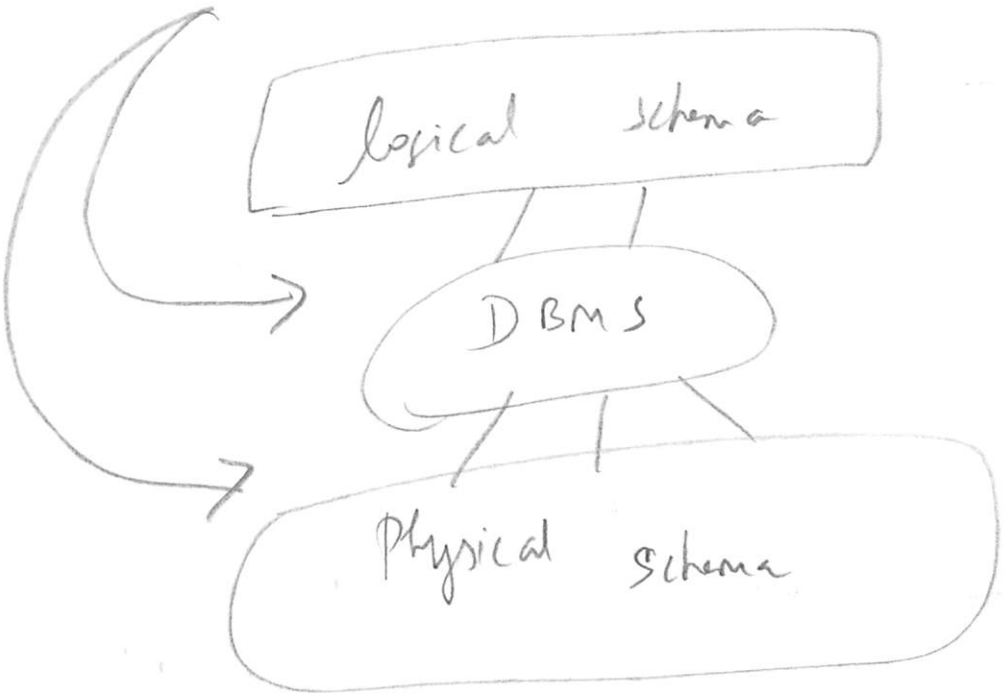


(b) The descriptive attributes namely "start date" and "end date" are added to the "runs dept" relation

Part of the ER diagram



Data independency



2) (20 points): Explain the concept of physical data independence.

Advantages of DBMS

DBMS has data Independence, data constraints, Security, authorization, physical schema, logical schema, view levels, data integrity, data correlation than the file systems.

In file system → datas are in different file formats

✗ there is no data independency.

physical schema → defines how a data is stored (in the file organization, how a record is stored)

logical schema → how data is represented in database, the database schema defined in the metadata (data description) data dictionary.

In DBMS, it is possible to change the physical schema without affecting the logical schema its called physical data Independence

3) (40 points): Given the following relational schemas:

employee(name, ssn, bdate, address, sex, salary, deptnumber)
department(deptname, deptnumber)
dept_location(deptnumber, deptlocation)
project(projname, projnumber, projlocation, deptnumber)
works_on(ssn, projnumber, hours)
dependent(ssn, dependname, dependtgender, dependbdate, relationship)

Using the given relational schemas, do the following:

a) Write an SQL query that lists the names and addresses of all employees who work for the "Computer Science" department.

```
Select name, address
from employee where deptnumber in
(Select deptnumber from department
    where deptname = 'computer Science')
```

✓

b) Write an SQL query that lists the names of employees who have only female dependents (i.e. all dependents of these employees are female).

```
Select ssn from employee where exists
((Select ssn, dependgender from dependent)
    minus
    (Select ssn, dependtgender from dependent)
    where dependtgender = 'Female')
)
```

-5

```
select distinct(name) from employee as e, dependent s
where e.ssn = s.ssn and
s.dependtgender = 'female';
```

4) (65 points): Given the following table:

DEPARTMENT(deptID, budget, managerID)

and the following series of insertion operations:

INSERT INTO DEPARTMENT VALUES (6, 3000, 1001);
INSERT INTO DEPARTMENT VALUES (11, 2500, 1001);
INSERT INTO DEPARTMENT VALUES (7, 3000, 1002);
INSERT INTO DEPARTMENT VALUES (3, 1200, 1003);
INSERT INTO DEPARTMENT VALUES (4, 3000, 1004);
INSERT INTO DEPARTMENT VALUES (15, 1500, 1004);
INSERT INTO DEPARTMENT VALUES (9, 3000, 1005);

Assume that each disk block can store up to 2 tuples of DEPARTMENT. Answer the following questions:

a) Show the contents of table DEPARTMENT when DEPARTMENT is organized as a heap file.

b) Show the contents of table DEPARTMENT and its index on budget when DEPARTMENT is organized as an index-sequential file with the search key being deptID. Use the implementation of index-sequential files discussed in class.

c) Assume that table DEPARTMENT is organized as a sequential file with the search key being deptID. You are asked to build a B+ tree dense index of order 3 for deptID. Show the B+ tree **after each** of the seven insertion operations given above.
Use the B+ tree definition and the B+ tree insertion procedure discussed in class.

(a) Heap file – data records are inserted randomly & free lists are maintained

disk block 1

Header		
6	3000	1001
11	2500	1001

disk block 2

Header		
7	3000	1002
3	1200	1003

disk block 3

Header		
4	3000	1004
15	1500	1004

disk block 4

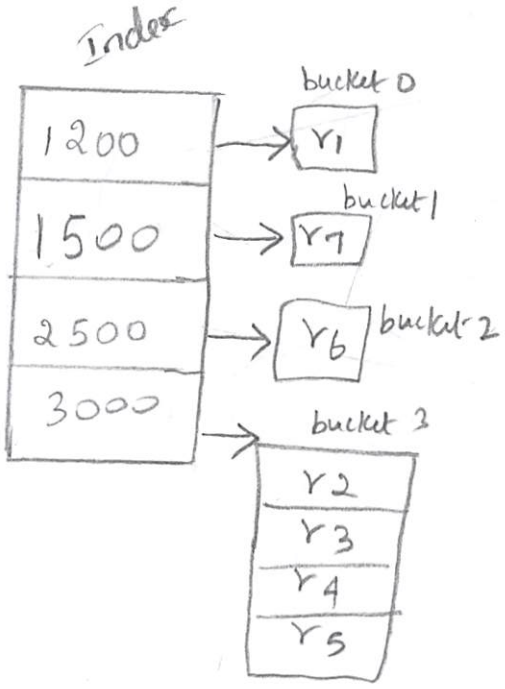
Header		
9	3000	1005

free space information is stored in the header of blocks

(b)

(Additional Workspace for Problem 4)

Index - sequential data file



search key: dept ID

dept ID	budget	manager ID	tuples
3	1200	1003	r1
4	3000	1004	r2
6	3000	1001	r3
7	3000	1002	r4
9	3000	1005	r5
11	2500	1001	r6
15	1500	1004	r7

✓

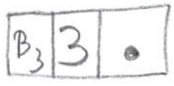
(c) B+tree order n=3

min = $\lceil n/2 \rceil - 1 = 1$

max = $(n-1) = 3-1 = 2$

3, 4, 6, 7, 9, 11, 15

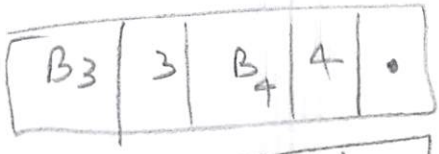
(i) insert 3 → Wrong order of insertion!



-4

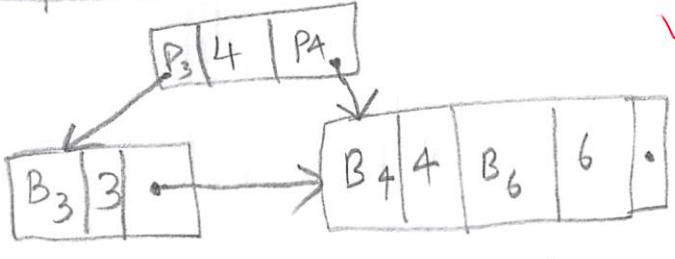
the order of insertions must be the same as the order of insertion operations given in the problem description!

(ii) insert 4

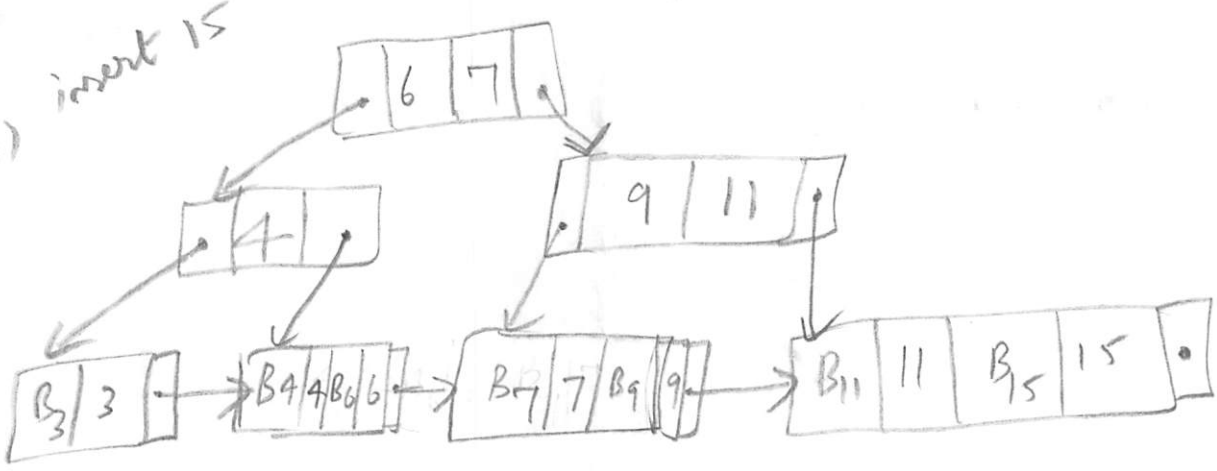


✓

(iii) insert 6

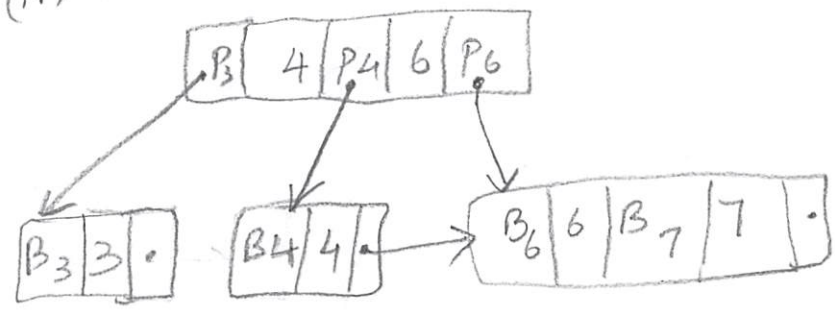


(viii) insert 15

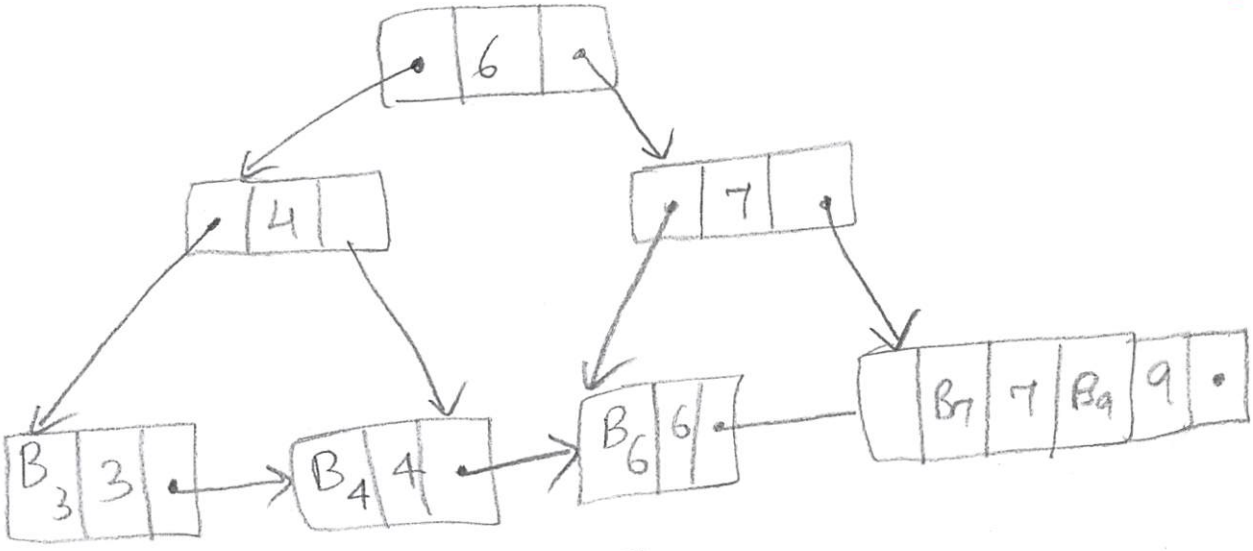


(Additional Workspace for Problem 4)

(iv) insert 7



(vi) insert 9



(vii) insert 11