Database Systems
(2017/07)

[1] A trucking company called ABC-Trucks is responsible for picking up shipments from warehouses of a retail chain called PQR-chains, and delivering the shipments to individual retail store locations of PQR-chains. Currently there are 8 warehouses and 70 retail stores of PQR-chains. A truck may carry several shipments in a single trip, which is identified by a trip#, and delivers each shipment to a particular store. Each shipment is identified by a shipment#, and includes data on shipment volume, weight, the name of the store to be delivered, etc. Trucks have different capacities for both the volumes they can hold and the weights they can carry. The ABC-Trucks company currently has 150 trucks, and a truck makes 3 to 4 trips each week. A database, which will be used by both ABC-Trucks and PQR-chains, is being designed to keep track of truck usage and deliveries. The database will also be used for scheduling trucks to provide timely deliveries to stores. Assume that a single trip can load shipments from several warehouses.

Design an ER schema diagram for the above application. For your convenience, use five entity sets called TRUCKS, WAREHOUSES, STORES, TRIPS, SHIPMENTS. In your ER schema diagram, do not use a multi-valued attribute. (20 points)

[2] Consider the relational database about employees given below.

```
Lives(emp-name, street, city)
Works(emp-name, company-name, salary)
Manages(emp-name, manager-name)
```

Give relational algebra expressions and SQL statements for each of the following queries.

(a) For all employees who work for First Bank Corporation, find employee names, their cities and their manager's names. (10 points)

(b) Find all employees who earn more than every employee of Small Bank Corporation. (10 points)

[3]

(a) Consider Extendible hashing. Briefly describe the procedure for inserting a new record *r*.

(10 points)

(b) Consider an R-tree. Briefly describe the procedure for inserting a new record *r* (also called a data region *r* when we discuss about the R-tree). Here, describe only the case when we are at the root node or interior node. In other words, describe only the first part of the insert algorithm, and need not describe the rest of the algorithm, i.e., the case when we are at the leaf node. (10 points)

[4]

Briefly describe various algorithms for computing natural joins of relations R(X,Y) and S(Y,Z). Here, let us assume the following:

 − The size of relation S is smaller than the size of relation R.
 − The number of available main memory buffers is M.
 − The size of a main memory buffer is the same as the size of a disk block.

You need to provide at least four different algorithms. Simply providing the names of algorithms without algorithm descriptions will not have full credits (20 points)

[5]

(a) Briefly describe "dirty data" and also describe the risk when reading dirty data. (5 points)

(b) Explain what the update lock is.
 Give the compatibility matrix for S(shared), X(exclusive) and U(update) locks. (5 points)

(c) Briefly explain four isolation levels in the SQL standard. You have to describe main characteristics of each isolation level. (10 points)