CMPT 354 Fall 2019
Database Systems
Martin Ester

Sample Final Exam

Problem 1: RA queries (60 marks)

Consider the following schema of a Company database:

```
Employees(eid: integer, ename: string, address: string,
          supereid: integer)
Departments(did: integer, dname: string)
Projects(pid: integer, pname: string, did: integer)
Works_on(eid: integer, pid: integer, hours: integer)
```

Each Employee has a supervisor (another Employee) referenced by his/her supereid. Projects are uniquely assigned to a Department. The Works_on relation records which Employee works on which Project for how many hours a week.

Formulate each of the following queries in relational algebra (RA):

a) For each Employee, find their name and the name of their supervisor.

b) Find the eids of Employees who work on a project of every Department, i.e. find the eids of Employees who work for at least one project of every Department.

c) Find the pid of Projects for which at least two different Employees work.

Problem 2: SQL queries (60 marks)

Consider the schema of the Company database from Problem 1.

```
Employees(eid: integer, ename: string, address: string,
          supereid: integer)
Departments(did: integer, dname: string)
Projects(pid: integer, pname: string, did: integer)
Works_on(eid: integer, pid: integer, hours: integer)
```

Formulate each of the following queries in SQL (both, standard SQL and MS SQL notation are acceptable):

a) For each Employee, find their name and the name of their supervisor.

b) Find the eids of Employees who work on a project of every Department, i.e. find the eids of Employees who work for at least one project of every Department.

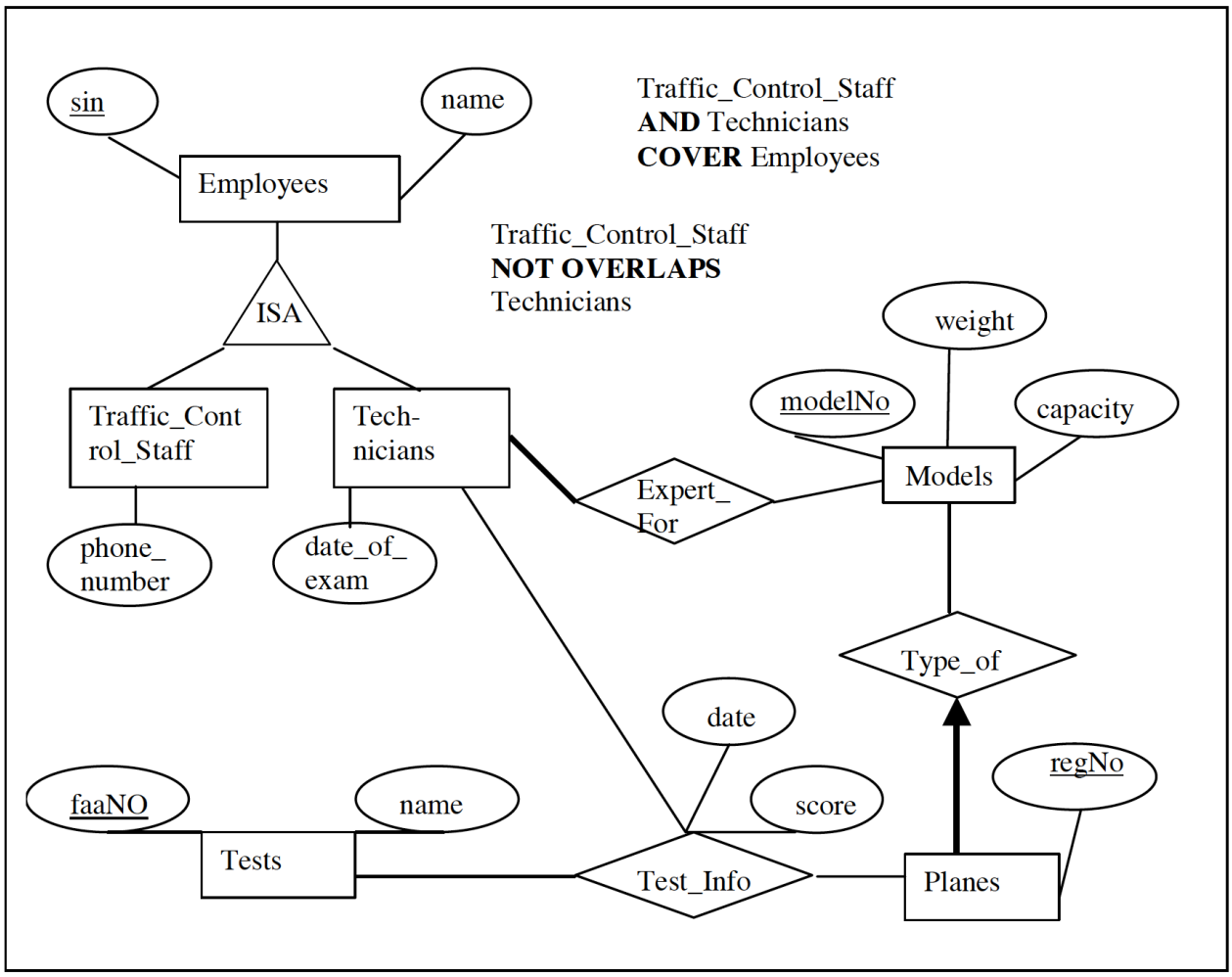
c) Find the total number of hours worked per project in the Marketing Department, and the pid, for all projects for which that number is at least 100.

Problem 3: ER diagram to relational schema (100 marks)

Translate the following ER diagram of an Airport database into a relational schema. Present the SQL statements to create all the tables of your relational schema.

Your relational schema should satisfy the following two design criteria:
·The number of tables should be minimal.
·As many integrity constraints from the ER diagram as possible should be expressed.

Note: Use INTEGER as domain for all the numerical attributes, VARCHAR(10) for all the text attributes, and dateTime for attribute date. You do not need to specify the actions for violations of foreign key constraints such as ON DELETE CASCADE.



Problem 4: Database design theory (80 marks)

Consider a table R with the following set of attributes and set of functional dependencies:

R(A, B, C, D, E)

A, B —> C
C —> D
D —>B
D —> E

a) What is the closure of A, B under the given functional dependencies?

What is the closure of C under the given functional dependencies?

What is the only primary key of R? Why?

b) Design a sequence of (binary) decompositions that transforms the table R into a set of tables that are all in BCNF. For every decomposition,

- state the input table to be decomposed,
- list the primary key of the input table,
- list a FD A1, . . ., An → B1, . . ., Bm that violates BCNF,
- add all attributes to the right side of the FD that are (indirectly) functionally dependent from A1, . . ., An, and
- list the attribute sets of the two tables resulting from the decomposition.

Finally, list the set of tables that result from your sequence of decompositions and underline their primary keys.

Problem 5: Multiple Choice Questions (100 marks)

Answer the following questions, which have ONE correct answer. You get 5 marks if you provide the correct answer and no wrong answer. **Write down a single answer, e.g., A or B, on the left of the question.**

(1) What is the number of values of one attribute of one record in a relational table?

A: at most one B: exactly one C: one or two D: any number

(2) A table in the relational data model is a . . . of tuples.

A: bag B: list C: queue D: set

(3) Which of the following is not a schema in a relational DBS:

A: conceptual schema	B: external schema	C: internal schema	D: physical schema

(4) What is the definition of atomicity of a transaction?

A: all B: none C: some D: all or none actions of a transaction are performed.

(5) A DBMS uses locks to ensure the . . . of transactions.

A: efficiency B: durability C: consistency D: isolation

(6) A DBMS executes transactions concurrently to improve the

A: throughput of the system B: response time of transactions C: security D:safety

(7) Which of the following SQL statements does not modify the instance of a table?

A: INSERT B: DELETE C: ALTER TABLE D: UPDATE

(8) An SQL DELETE statement can delete

A: one tuple	B: many tuples from one table	C: many tuples from two tables	D: many tuples from any tables

(9) What is the runtime complexity of an eqality query in a B-tree that indexes *n* records?

A: O(1) B: O(log *n*) C: O(*n*) D: O(*n* log*n*)

(10) Besides B-trees, what method is commonly used as index in a DBMS?

A: hashing B: heap C: queue D: sorting

(11) What is the runtime complexity of an equality query, e.g. an SQL query SELECT * FROM Students WHERE sid=1234567, in a table with *n* records without index?

A: O(1) B: O(log *n*) C: O(*n*) D: O(*n* log*n*)

(12) An SQL assertion is associated with

A: an attribute B: a table C: a database D: none of those

(13) Every SQL assertion can be implemented equivalently through

A: a CHECK constraint	B: multiple CHECK constraints	C: a trigger	D: multiple triggers

(14) A data cube consists of

A: dimensions	B: measures	C: dimensions and measures	D: none of the two

(15) The following is not an OLAP query:

A: DRILL DOWN B: ORDER BY C: PIVOT D: ROLL-UP

(16) How many SQL (GROUP BY) queries need to be executed to compute all cells of a data cube with d dimensions?

A: 1 B: d C: d^2 D: 2^d

(17) Suppose that we have a classification problem with two classes "+" and "–" and four different training datasets, where each example is represented through its class label. Which of the following datasets has the smallest entropy?

A: ++++- B: +++- C: ++++-- D: +++---

(18) We say that a decision tree overfits if it has

A: high training accuracy and high test accuracy B: high training accuracy and low test accuracy C: low training accuracy and high test accuracy D: low training accuracy and low test accuracy

(19) An association rule has to have

A: minimum confidence B: minimum support C: minimum confidence and minimum support D: none of the two

(20) Suppose that the following are the frequent 3-itemsets in a given transaction database: ABC, ABD, ABE, ACD, ACE, ADE, BCD, BCE, BDE. Which of the following 4-itemsets cannot be frequent?

A: ABCD, B: ABCE, C: ABDE, D: BCDE