DATABASES EXAM – 2016

Date: May-31 2016

Docent: Amr Ali-Eldin

Maximum number of points possible: **100**. This exam counts for 60% of your overall grade.

Preliminaries: Please enter your name and student number on every page and use exam paper to answer the questions. The time for the exam is **3** Hours.

If you make any additional assumptions, please state them clearly and indicate their scope.

You can deliver the answers in English or in Dutch.

LIACS, Leiden University

Part One: Entity Relationship Models (20 points)

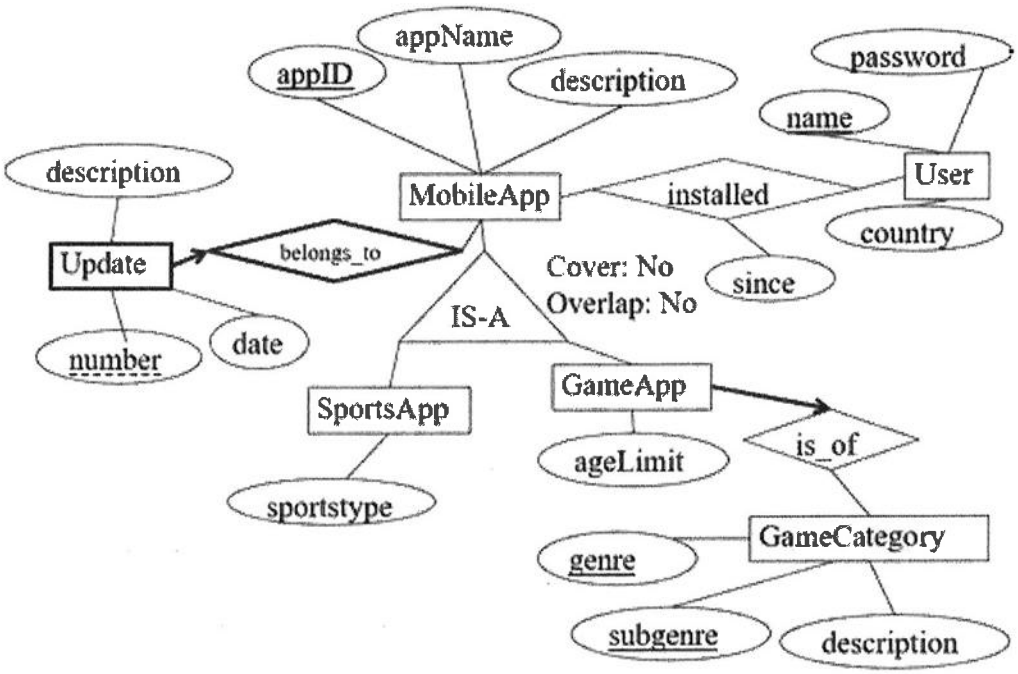
1.1 A music company Xmusic wants to store information about musicians in a database. Model the following data and constraint descriptions in form of an entity relationship diagram (**16 points**):

1. Each musician that records at Xmusic has an SSN, a name, an address, and a phone number. Poorly paid musicians often share the same address, and no address has more than one phone.
2. Each instrument used in songs recorded at Xmusic has a unique identification number, a name (e.g., guitar) and a musical key (e.g., C, B-flat, E-flat).
3. Each album recorded on the Xmusic label has a unique identification number, a title, a copyright date, a format (e.g., CD or MC), and an album identifier.
4. Each song recorded at Xmusic has a title and an author.
5. Each musician may play several instruments, and a given instrument may be played by several musicians.
6. Each album has a number of songs on it, but no song may appear on more than one album.
7. Each song is performed by one or more musicians, and a musician may perform a number of songs.
8. Each album has exactly one musician who acts as its producer. A musician may produce several albums.

1.2 Draw an ER-diagram that captures the following two SQL statements (**4 points**):

```
CREATE TABLE Employees (    ssn     CHAR(11),
                            Name    CHAR(30),
                            mlot    INTEGER,
                            PRIMARY KEY(ssn))
CREATE TABLE Dept_Mgr (     did     INTEGER,
                            dname   CHAR(20),
                            ssn     CHAR(11) NOT NULL,
                            PRIMARY KEY (did),
                            FOREIGN KEY (ssn) REFERENCES Employees,
                            ON DELETE NO ACTION)
```

Part Two: LOGICAL DATABASE DESIGN (15 points)



Imagine, your task is to maintain data of the above ***database***. Translate the ER diagram for ***ER Model above*** to a SQL-DDL Logical Design using the **CREATE TABLE** command.

Part Three: SQL Queries (15 points)

3.1 Consider the following schema:

Suppliers(sid: integer, sname: string, address: string)

Parts(pid: integer, pname: string, color: string)

Catalog(sid: integer, pid: integer, price: real)

The Suppliers relation describes suppliers of parts. The Parts relation contains information about each part. The Catalog relation lists the prices in dollars charged for parts by suppliers. (The keys are underlined: sid is a key for Suppliers, (sid,pid) is a key for Catalog, and pid is a key for Parts.)

Write the following queries in SQL (**10 points**):

1. Find the pnames of parts for which there is some supplier.
2. Find the snames of suppliers who supply every part.
3. For each part, find the sname of the supplier who charges the most for that part.
4. Find the sids of suppliers who supply only red parts.
5. Find the sids of suppliers who supply a red part and a green part.

3.2 Consider the following SQL query on the relation Product(pid, name, price, mfgr):

```
SELECT  *
FROM    Product P
WHERE   NOT (P.price > 500 OR NOT (P.mfgr = 'Sears'))
```

For each of the following tuples in the Product relation, indicate whether it will be selected by the above query (you should answer "yes", "No" or "undefined"). Show how you derived your conclusion **(5 points)**.

(10, Hammer, 400, null):

(11, Hammer, null, Walmart):

(12, Hammer, null, null):

(13, Hammer, 600, null):

Part Four: Relational Algebra (10 points)

4.1 Consider the following relational schema (keys are underlined):

Product (pid, name, price, mfgr), Buys (cid, pid), Customer (cid, cname, age, gender)

a) Write the following query **in relational algebra**: "Find the names of all customers who have purchased all products that are manufactured by Sears." **(3 points)**

b) Write the following query **in relational algebra**: "Find the cids and names of all customers who have purchased the ***most*** expensive product. You can assume that all product prices are unique. **(4 points)**

4.2 Assume we have two relations R(A,B) and S(B,C). All three attributes (A, B, and C) are integer attributes. Assume that Relation R contains the following tuples: (1,2), (2,3), and (3,4). Assume that Relation S contains the following tuples (2,2), (2,3) and (5,1). Recall that a key is minimal.

a) Give an example of an attribute (combination) that cannot be a key for relation S **(1 point)**.

b) How many tuples are in the result of the following relational algebra expression? **(1 point)**

$$\text{PROJECT}_{A}\ (\text{R natural Join S})$$

c) How many keys are there for both relations**(1 point)**.

Part Five: Normal Forms and Functional Dependences (15 points)

5.1- Consider a relation R with five attributes ABCDE. Now assume that R is decomposed into two smaller relations ABC and CDE. Define S to be the relation (ABC Natural Join CDE).

Assume that the above decomposition is lossless join, but not dependency preserving. You do not know any additional information about the decomposition. Which of the following statements can you infer to be *always* true. List all true statements if more than one statement can be inferred to be true (**4 points**):

(1) R = S

(2) $R \subseteq S$

(3) $R \subset S$

(4) $R \supseteq S$

(5) $R \supset S$

(6) none of the above.

5.2 Consider a relation R with six attributes ABCDEF. The following functional dependency, F1 = A → BCDEF, holds over R. Your task is to come up with two other functional dependencies, F2 and F3, which satisfy the following three properties.

1) Neither F2 nor F3 can be inferred from F1 using Armstrong's axioms
2) Relation R with functional dependencies F1 and F2 is in BCNF
3) Relation R with functional dependencies F1, F2, and F3 is in 3NF but not in BCNF

Write down two functional dependencies F2 and F3 that satisfy the above properties (**4 points**).

5.3 Prove the following inference rule for functional dependencies using only Armstrong's axioms:

If P → QR and R → S, then P → QS

Show the steps of your proof, and indicate which of Armstrong's axioms is applied in each step (**3 points**).

5.4 Consider relation R with six attributes ABCDEF and the following functional dependencies: A → CDF, D → A, DF → E. We decompose R into two relations, ABCD and DEF.

a) Is this decomposition a lossless join decomposition? Briefly justify your answer (**2 points**).
b) Is this decomposition a dependency-preserving decomposition? Briefly justify your answer (**2 points**).

Part Six: Transactions Management (15 points)

Consider the following transaction schedule, where time increases from left to right. (cmt stands for commit).

T1:	R(A)	W(A)	R(B)	R(C)							W(B)		cmt.		
T2:					R(A)	R(C)	W(A)					W(C)		cmt.	
T3:								R(C)	R(B)	W(C)					cmt.

Given the interleaved schedule in the figure above:

1. **Draw the precedence graph and decide on basis of that graph whether or not the schedule is conflict serializable.**
2. **Apply Strict 2PL to the schedule above and denote the resulting schedule.**
3. **Draw a waits for graph for the point in time when the first deadlock occurs**

Part Seven: Concepts and Basic Knowledge (10 points)

Please note that multiple answers can be correct.

1. For which operation does a B+ tree index perform better than a hash based index?
    a. Range selection
    b. Insert and Update
    c. Equality selection
    d. None of the above

2. Which of the following statements on ACID properties are correct?
    a. A transaction actions are sometimes partially executed.
    b. After a transaction is committed, the database is always in a consistent state, regardless of the initial state.
    c. Transactions are self-contained and isolated from other transactions.
    d. The effect of committed transaction should persist, even after the system recovers from crashes.

3. A recent trend is to create databases for big data. Which operations should be avoided when dealing with huge tables, because they may result in huge running times:
    a. Joins
    b. Projections
    c. Selections (using B+ Tree index)
    d. Cross Products

4. Stored Procedures are.....
    a. data objects on procedures augmented by time.
    b. programs written in procedural language extension of SQL and executed on the database server.
    c. definitions of relational tables and their attributes that can be requested by the user.
    d. None of the above.

5. In two-phase locking protocol, a transaction obtains locks in .....
    a. The shrinking phase
    b. The growing phase
    c. The running phase
    d. The initial phase

*** END OF EXAM, Success ***