Model for the written exam
Computer Science in English, 2nd year

I. 1. Define the 2nd normal form. **1p**
2. Describe the *cross join* algorithm.

II. A musical instruments manufacturer relies on a relational database to support its activities. You are asked to design a part of the database schema and answer some questions the management is interested in using the specified database language. The manufacturer produces instruments of different categories. A category has: name (e.g., string, percussion, woodwind, etc), description, and several subcategories. A subcategory belongs to a category and has a name (e.g., violin, piano, contrabass, etc). An instrument currently in stock belongs to a subcategory and has: serial number, date of manufacture, color, and price. A customer has: name, score, and type (e.g., music academy, orchestra, etc). The company takes orders directly from customers, online or by phone. An order is placed by a customer and has: 2 dates – the date when the order was made and the date when it was honored (*null* for unfulfilled orders), a field indicating whether it's been placed online or by phone, and, for each subcategory of instruments in the order, the number of ordered instruments of each color (e.g., an order for 7 red violins, 3 white violins, 2 white pianos, and a yellow contrabass). **4p**

1. a. Draw a database diagram (tables, constraints) for the above data. The schema must be 3NF. (2p)
b. Write a SQL statement that creates a table with at least 2 foreign keys and a CHECK constraint.

2. Write a query for each of the tasks below, using the specified language: (2p)
a. For every orchestra that ordered violins in at least 3 different orders, find the total number of ordered instruments – **in SQL, without views** (CustomerID, TotalNumInstruments).
b. Find all the string instruments that are currently in stock and cost less than 2000 lei – **in the relational algebra** (subcategory name, instrument serial number and price).

III. Choose the correct answer(s) for the following multiple choice questions. Each question has at least one correct answer. Enter the correct answers in the table below. **3p**

1.	2.	3.
4.	5.	6.
7.	8.	9.
10.	11.	12.

1. In a SELECT query in SQL:
a. the SELECT clause can contain arithmetic expressions
b. according to the conceptual evaluation strategy, ORDER BY is evaluated before GROUP BY
c. HAVING can contain row-level qualification conditions
d. DISTINCT eliminates duplicates from the answer set
e. none of the above answers is correct.

2. The natural join operator R_1*R_2 in the relational algebra:
a. returns a relation whose schema contains only the attributes in R_1 that don't appear in R_2
b. returns a relation whose schema contains all the attributes in R_1 and R_2, with common attributes appearing only once
c. returns 2 relation instances
d. is not associative
e. none of the above answers is correct.

3. The ANSI-SPARC architecture for a database system can include:
a. exactly one symbolic structure
b. several conceptual structures
c. several external structures
d. several physical structures
e. none of the above answers is correct.

4. Consider the relation S[A, B, C, D, E, F, G, H] with the key {A, B, C}, the functional dependencies: {F} → {H}, {C} → {E, G}, and no repeating attributes.
a. S is not 1NF
b. S is 2NF
c. S is not BCNF
d. S is 3NF
e. none of the above answers is correct.

5. In a B-tree of order 8:
a. a non-terminal node has at most 8 subtrees
b. a non-terminal node with 7 values has 8 subtrees
c. terminal nodes can be on different levels
d. a node with 7 values has 6 subtrees
e. none of the above answers is correct.

6. Let α, β and γ be subsets of attributes in a relational schema. According to the union rule, if α → β and α → γ, then:
a. γ → αβ
b. βγ → α
c. β → αγ
d. βα → γ
e. none of the above answers is correct.

7. Let RepairLog[RID, MechanicID, RollerCoasterID, RepairTime] be a table with 100.000 records. The table has 2 indexes: a unique clustered index on *RID* and a non-clustered index on *MechanicID*.
Consider the following query:

```
SELECT RID, MechanicID, RepairTime
FROM RepairLog
WHERE MechanicID = 7
```

If the execution plan in SQL Server contains a seek operation on the non-clustered index, it also contains:
a. a scan on the clustered index
b. a scan on the non-clustered index
c. a key lookup on the clustered index
d. a trick on the non-clustered index
e. none of the above answers is correct.

8-10. Consider the relational schema S[FK1, FK2, A, B, C, D, E] with the key {FK1, FK2}. Answer questions 8-10 using the legal instance below:

FK1	FK2	A	B	C	D	E
1	1	a1	b1	c1	7	2
1	2	a_	b3	c1	5	2
1	3	a2	b1	c2	Null	2
2	1	a3	b3	c2	Null	100
2	2	a3	b3	c3	Null	100

8. Consider queries Q_1 and Q_2:
Q_1:
```
SELECT *
FROM S s1 LEFT JOIN S s2 ON s1.FK1 = s2.E
```
Q_2:
```
SELECT DISTINCT *
FROM S s1 INNER JOIN S s2 ON s1.FK1 = s2.E
```
The cardinality of the answer set of Q_i is denoted by $|Q_i|$.
$|Q_1| - |Q_2|$ is:
a. 0
b. 3
c. 10
d. 2
e. none of the above answers is correct.

9. Regarding the functional dependencies of S:
a. at least one of the following dependencies is not satisfied by the instance: {A} → {B}, {FK1, FK2} → {A, B}, {FK1} → {A}
b. by examining the instance, we can conclude that at least one of the following dependencies is specified on the schema S: {A} → {B}, {FK1} → {A, B}, {FK1} → {A}
c. at least two of the following dependencies are not satisfied by the instance: {FK2} → {A, B}, {A} → {E}, {A, B} → {E}, {FK1, FK2} → {E}
d. by examining the instance, we can conclude that at least two of the following dependencies are specified on the schema S: {FK2} → {A, B}, {A} → {E}, {A, B} → {E}, {B} → {C, E}
e. none of the above answers is correct.

10. Consider queries Q_1 and Q_2:
Q_1:
```
SELECT FK2, FK1, COUNT(DISTINCT B)
FROM S
GROUP BY FK2, FK1
HAVING FK1 = 0
```
Q_2:
```
SELECT FK2, FK1, COUNT(C)
FROM S
GROUP BY FK2, FK1
HAVING MAX(E) < 0
```
The cardinality of the answer set of Q_i is denoted by $|Q_i|$.
$|Q_1| - |Q_2|$ is:
a. 0
b. 2
c. 1
d. -1
e. none of the above answers is correct.

11. A secondary index:
a. can contain duplicates
b. cannot contain duplicates
c. can be non-clustered
d. cannot be non-clustered
e. none of the above answers is correct.

12. For the relation R[A, B, C] below, consider the 3 possible projections on 2 attributes: AB[A, B], BC[B, C], and AC[A, C]. How many extra records does AB * BC * AC contain (i.e., records that don't appear in R)?

A	B	C
a1	b2	c1
a1	b1	c2
a2	b1	c1

:
a. 0
b. 1
c. 2
d. 3
e. none of the above answers is correct.

(0.25p / question)

IV. Let P, Q, and R be 3 relations with schemas P[PID, P1, P2, P3], Q[QID, Q1, Q2, Q3, Q4, Q5], R[RID, R1, R2, R3], and E an expression in the relational algebra:

$E = \Pi_{\{P.P2,Q.Q2,\ Q.Q4,\ R.R3\}}(\sigma_{P.PID = Q.Q1\ AND\ Q.QID = R.R2\ AND\ P.P3='Bilbo'\ AND\ Q.Q5 = 100\ AND\ R.R1=7}(P \times Q \times R))$

Optimize E and draw the evaluation tree for the optimized version of the expression. **1p**

1p of