Databases – written exam
Computer Science in English, 2nd year, I

I. 1. Define the *key* from the relational model. **1p**
2. Describe clustered indexes, using an example.

II. The website of the Romanian Society of Ornithology is powered by a relational database that contains data about registered birdwatchers, the observations they collect, and natural areas managed by the society. You are asked to design a part of the database schema and answer some questions using the specified database language. A birdwatcher has: id, first name, last name. A bird species has: latin name, common name, short description, estimated population, genus. A species belongs to a genus; a genus belongs to a family and can correspond to several species; a family can be associated with several genera. Both a genus and a family have only a name. E.g., the *white stork* species (common name) belongs to the *Ciconia* genus, which in turn belongs to the *Ciconiidae* family. A natural area has: id, name; it can be host to several species; a species can live in several natural areas. A birdwatcher's observation has: date, species, and number of observed birds. **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 a PRIMARY KEY, a FOREIGN KEY, and a NOT NULL constraint.

2. Write a query for each of the tasks below, using the specified language: **(2p)**
a. For every birdwatcher with at least 10 observations, find his/her: first name, last name, total number of observations, and total number of observations of species of the *Aquila* genus – in SQL, without views.
b. For every species with an estimated population of less than 1000 birds and that lives in at least one natural area, find its: common name, latin name, and estimated population – in the relational algebra.

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. Rows in a relation in the relational model:
a. are not ordered
b. are not distinct
c. are ordered
d. are distinct
e. none of the above answers is correct.

2-6. Consider the relational schema R[RID, M, N, P, Q, O] with the primary key {RID}, and NOT NULL constraints on columns M and N. R has no other restrictions. Answer questions 2-6 using the legal instance below:

RID	M	N	P	Q	O
10	Allons	Le jour	20	1	10
1	enfants	de	21	1	5
2	de	gloire	22	1	null
11	la	est	23	1	null
3	Patrie,	arrivé!	24	1	2

2. How many records does the query below return when executed on a legal instance of R?

```
SELECT *
FROM R
WHERE M IS NULL
```

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

3. After executing the statement:

```
DROP TABLE R
```

a. both the schema information and the tuples of the table are still stored in the database
b. only the tuples are removed from the table, the schema information is still stored in the database
c. both the schema information and the tuples of the table are removed from the database
d. only the schema information is removed, the tuples are still stored in the table
e. none of the above answers is correct.

4. How many records does the query below return?

```
SELECT M, COUNT(DISTINCT P)
FROM R
```

a. 1
b. 5
c. 4
d. 2
e. none of the above answers is correct.

5. How many records does the query below return?

```
SELECT *
FROM R
WHERE P > ALL (SELECT P
               FROM R)
```

a. 0
b. 5
c. 1
d. 4
e. none of the above answers is correct.

6. How many records does the query below return?

```
SELECT AVG(P)
FROM R
GROUP BY Q
HAVING AVG(P) < 10
```

a. 0
b. 5
c. 2
d. 1
e. none of the above answers is correct.

7. For the relation S[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 S)?

A	B	C
A1	B1	C1
A1	B1	C2
A2	B2	C1

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

8. In an ISAM structure with no overflow chains, *m* primary leaf pages, and *c* children per index page, the cost of the search is:

a. $\sum_{i=1}^{c} \frac{i}{m}$
b. $\prod_{i=1}^{m} \frac{i}{c}$
c. $\log_c m$
d. $\log_2(c/m)$
e. none of the above answers is correct.

9. Of the algorithms below, which one is the best to evaluate a search of the form $C = C_0$, where C is a key?
a. table scan
b. index scan
c. index seek
d. cross join
e. hash join

10. When mapping an *m:n* association to tables, if an association class exists, its attributes:
a. are added to the table that lies on the *m* side of the association
b. are added to the table that lies on the *n* side of the association
c. are added to the link table
d. are not added to any table
e. none of the above answers is correct.

11. Consider the relation R[A, B, C, D, E] with:
- the keys {A, E}, {A, B, C}, {A, B, D};
- the functional dependency {A, C} → {D};
- no repeating attributes.
a. R is 1NF
b. R is 2NF
c. R is 3NF
d. R is BCNF
e. none of the above answers is correct.

12. In a DBMS, the *buffer manager*:
a. automatically drops clustered indexes
b. manages available disk space
c. when necessary, chooses a page in the *buffer pool* for replacement, based on a replacement policy
d. maintains the number of current users for every page in the *buffer pool*
e. none of the above answers is correct.

(0.25p / question)

IV. Let A, B, and C be 3 relations with schemas A[AID, A1, A2, A3], B[BID, B1, B2, B3], C[CID, C1, C2, C3], and E an expression in the relational algebra:

E = $\sigma_{\text{A.AID = B.B1 AND B.BID=C.C2 AND C.C3=5}} (\Pi_{\{\text{A.AID, A.A1, A.A2, B.BID, B.B1, C.C2, C.C3}\}}(\text{A}\times\text{B}\times\text{C}))$

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

1p of