I. 1. Define the 3rd normal form.   **1p**
2. Briefly describe a ternary relationship set in the Entity-Relationship model using an example.

II. The website *astronomyforbeginners.ro* is powered by a relational database that contains data about astronomical objects and phenomena. You are asked to design a part of the database schema and answer some questions using the specified database language. A natural satellite has a name and orbits a planet. A planet has a name and belongs to a planetary system, which has formed around a star and can have several planets. A planet can have several natural satellites. A star has a name, age, and metallicity (real number); it belongs to a galaxy and can have a planetary system. A galaxy has a name, diameter, category (*elliptical*, *spiral*, *irregular*), mass, and belongs to a group of galaxies or to a cluster of galaxies. A galaxy can contain a large number of stars. A group of galaxies has a name, diameter, and mass. A cluster of galaxies has a name, mass, and a type - integer number between *1* and *3* as defined by the *Bautz-Morgan classification*. Groups and clusters of galaxies can contain large numbers of galaxies. E.g., the *Sun* is a star in the *Milky Way* galaxy, which in turn belongs to a galaxy group called *Local Group*. *The Earth* is part of the *Sun*'s planetary system, and has one natural satellite, the *Moon*.   **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 and a foreign key.

2. Write a query for each of the tasks below, using the specified language: (2p)
a. Find the name and age of every star in the *Milky Way* galaxy with a planetary system that has:
- at least 5 planets OR
- at least one natural satellite (as stated above, such a satellite belongs to a planet in the planetary system).
  – **in SQL, without views** (StarName, StarAge).
b. Find all galaxies in clusters of type *2* – **in the relational algebra** (cluster name, galaxy name).

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.

1-5. Consider the relational schema S[<u>ID</u>, A, B, C, D, E, F] with the key {ID}. Answer questions 1-5 using the legal instance below:

ID	A	B	C	D	E	F
t1	a1	b2	Și abia plecă bătrânul... Ce mai freamăt, ce mai zbucium!	0	1	0
t2	a1	b2	Codrul clocoti de zgomot și de arme și de bucium,	1	2	1
t3	a1	b3	Iar la poala lui cea verde mii de capete pletoase,	0	3	0
t4	a1	b3	Mii de coifuri lucitoare ies din umbra-ntunecoasă;	2	123	-1
t5	a1	b3	Călăreții umplu câmpul și roiesc după un semn	-1	4	-1

1. When executed on the above instance S:
a. query `SELECT DISTINCT A, B FROM S` returns 2 tuples.
b. query `SELECT * FROM S WHERE B = 'b2' AND B = 'b3'` returns 0 tuples.
c. query `SELECT * FROM S WHERE B = 'b3'`
     `UNION SELECT * FROM S WHERE B = 'b3'`
returns 1 tuple.
d. query `SELECT * FROM S WHERE D >= 0`
     `EXCEPT SELECT * FROM S WHERE E <> 4`
returns 0 tuples.
e. none of the above answers is correct.

2. Consider projections S1[ID, A, B, C, F] and S2[D, E, F]. The result of the natural join S1*S2 contains (column order is not important):
a. only the 5 tuples in S
b. 7 tuples, out of which 5 are the original tuples in S

c. no tuples
d. 9 tuples, out of which 5 are the original tuples in S
e. none of the above answers is correct.

3. How many records does the query below return?
SELECT B, C, COUNT(*)
FROM S
GROUP BY B, C
HAVING D <= 1
a. 5
b. 4
c. 3
d. 2
e. none of the above answers is correct.

4. How many records does the query below return?
SELECT *
FROM S
WHERE C LIKE 'de%'
a. 3
b. 0
c. 5
d. 1
e. none of the above answers is correct.

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

6. A data description model can be used to describe:
a. the structure of the data
b. IF statements
c. the relationships with other data
d. the consistency constraints
e. none of the above answers is correct.

7. In a DBMS, the buffer manager:
a. manages disk space
b. brings pages from the disk into main memory
c. monitors lock requests
d. produces an efficient execution plan for query evaluation
e. none of the above answers is correct.

8. A SELECT statement:
a. can contain a HAVING clause only if it contains a WHERE clause
b. can contain a WHERE clause only if it contains a HAVING clause
c. can contain a HAVING clause only if it contains an ORDER BY clause
d. can contain a GROUP BY clause only if it contains a HAVING clause
e. none of the above answers is correct.

9. Let R[ID1, ID2, ID3, A, B, C, D, E, F, G] be a relational schema with no repeating attributes. The keys of R are {ID1, ID2, ID3} and {A, B}. The following dependencies hold: {ID1, ID2} → {C} and {A} → {G}. R is:
a. 1NF
b. 2NF
c. 3NF
d. BCNF
e. none of the above answers is correct.

10. On a magnetic disk, disk access time includes:
a. lunch time
b. seek time
c. rotational delay
d. transfer time
e. none of the above answers is correct.

11. A sparse index:
a. has one entry for every search key value
b. has one entry for each data page
c. can use alternative 2 for data entries (i.e., pairs of the form <search key value k, data record id rid>)
d. can be dense
e. none of the above answers is correct.

12. The projection operator π in the relational algebra:
a. is a binary operator
b. is a unary operator
c. is a ternary operator
d. is distributive with respect to set-difference
e. none of the above answers is correct.

(0.25p / question)

IV. Let A, B, C, D be 4 relations with schemas A[ID, A1, A2, A3], B[B1, B2], C[C1, C2], D[ID, D1], and E an expression in the relational algebra:

$E = (\sigma_{ID = B1 \text{ AND } A2='exam' \text{ AND } B2='DB'} (A \times B)) * (\sigma_{C1=3} (C \times D))$

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

1p of

is ID=B1 a join condition?
is there anything better than ' * ' natural join?