1. What is the result of the following boolean expression when \((x, y, z)\) is as specified below? \([3\, \text{pts}]\)

\[
( (x = 10) \text{ OR } (y \text{ IS NOT NULL}) ) \text{ AND } (z > 20)
\]

- \((x = 10, y = 100, z = 30)\) :
- \((\text{NULL}, 100, 10)\) :
- \((\text{NULL}, \text{NULL}, 10)\) :
- \((15, 100, \text{NULL})\) :
- \((10, \text{NULL}, 30)\) :
- \((\text{NULL}, \text{NULL}, \text{NULL})\) :

2. For the relation schema: \(\text{student}(\text{name}, \text{ssn}, \text{majorcode}, \text{gpa})\), where \text{ssn} has been defined to be the primary key, the query:

\[
\text{select ssn, name from student group by ssn;}
\]

is invalid (will return an error), even though it seems to make sense. Why? How would you fix the query without changing the output? \([3\, \text{pts}]\)
3. I need to preserve dependencies, and a BCNF decomposition is not possible. Which of the following
normal forms are still attainable (while preserving dependencies)? Circle your answers. [3 pts]

1NF 3NF 4NF

4. Consider two relations $R(A, B)$ and $S(B, C)$ and a view on them defined as: $V(A, B, C) = R \bowtie S$.
Under what condition(s) can we allow a tuple insertion into $V$, and what would its effect be? [3 pts]

5. Given a relational schema over three attributes, $A$, $B$, and $C$, that has three functional dependencies
$A \rightarrow B, B \rightarrow C$, and $A \rightarrow C$, both the decompositions $(AB, BC)$ and $(AB, AC)$ are in BCNF. Which
one of them is preferable? Why? [3 pts]

6. What does the unknown keyword/functionality mean? Briefly explain why it is needed. [3 pts]

7. Explain what on delete cascade means in the context of referential integrity constraints. [3 pts]

8. What is a prime attribute? Where is it used in normalization? [3 pts]
9. Given a relational schema \( R(A, B, C, D, E) \), with functional dependencies \( AB \rightarrow C \), \( E \rightarrow A \), and \( C \rightarrow E \), is the decomposition into \( R_1(A, B, C) \) and \( R_2(C, D, E) \) “lossless”? Why/why not? [3 pts]

10. The figure shows three different ways to model the entities person, actor, and director. Briefly explain what each one is trying to do, what issues you might face in using them, and which one would be the best option among these. [6 pts]

11. Construct the relational schemas corresponding to the first two E/R models ((i) and (ii)) from above (irrespective of whether they make sense). Underline the primary keys. [4 pts]
12. **Antijoin** is a binary operator similar to a join. Given two relations, \( R(A, B) \) and \( S(B, C) \), the result of the antijoin operation \( R \triangleright S \) has the same schema as \( R \), and contains all tuples of \( R \) that do not have at least one match (on attribute \( B \)) in \( S \). For example, if \( R(A, B) = \{ (1, \alpha), (2, \beta), (3, \gamma) \} \) and \( S(B, C) = \{ (\alpha, 2), (\gamma, 3) \} \), then \( R \triangleright S = \{ (2, \beta) \} \). Express \( R \triangleright S \) using the basic relational operators (\( \sigma, \pi, \times, \cup, - \)). [5 pts]

13. Given the relation schema: \( R(A, B, C, D) \), and FDs on it: [10 pts]

\[
AB \rightarrow C, \ C \rightarrow D, \ D \rightarrow B.
\]

- List all candidate keys.
- Is the relation in BCNF? List one FD that violates it if it is not.
- Decompose the relation into BCNF if it is not already in it.
- Is your decomposition into BCNF dependency-preserving?
- Is the relation (\( R \)) in 3NF? List one FD that violates it if it is not.
The following questions are to be answered on the following simplified *olympics* relational schema, where only the individual events information is stored for one olympics (say Athens 2004).

- Players (player-id, name, countryname, age);
- Events (event-id, name, eventtype);
- Results (player-id, event-id, medal);

*eventtype* can take values: *SWI* (swimming), *ATH* (athletics), *GYM* (gymnastics), etc.

*medal* can take values: *gold*, *silver*, *bronze*.

*age* is an integer. Assume player names are unique.

14. Write a relational algebra expression to find the names of the players who won at least one gold and one silver. [3 pts]

15. Explain what the following expression does. [3 pts]

\[ \pi_{\text{name}}((\pi_{\text{player-id}}(\text{players}) - \pi_{\text{player-id}}(\text{results})) \bowtie \text{Players}) \]

16. Explain what the following expression does. [3 pts]

\[ \pi_{\text{countryname}, \text{eventtype}}(\sigma_{\text{medal} = \text{'gold'}}(\text{players} \bowtie \text{results} \bowtie \text{events})) \div \pi_{\text{eventtype}}(\text{events}) \]

17. Write a relational algebra query to find the names of all the players with the minimum age. Assume player names are unique. [5 pts]
The following questions are to be answered on the following simplified friends relational schema.

Person (name, address);
Friends (name1, name2);

where the second relation stores the information about friends (ie., name1 and name2 are friends). Assume that this is symmetric relationship, and if X and Y are friends, then the second relation contains two tuples (X, Y), and (Y, X).

18. What does the following relational algebra expression do? Specifically, precisely list all types of tuples that will appear in the answer. [3 pts]

\[ \pi_{f1.name1, f2.name2}(\rho_{f1}(friends) \bowtie f1.name2 = f2.name1 \rho_{f2}(friends)) \]

19. If the friends relation contains a tuple (X, Y), under what conditions would the above expression also contain that tuple? [3 pts]

20. How would you change the data in the relations so that the result of the above expression always contains tuple (X, Y) if the friends relation contained tuple (X, Y)? [5 pts]

21. Rewrite the following SQL query so that it does not use a subquery. [3 pts]

```sql
select name
from person p
where exists (select *
    from friends r, person p2
    where r.name1 = p.name and p2.name = r.name2 and p.address = p2.address)
```