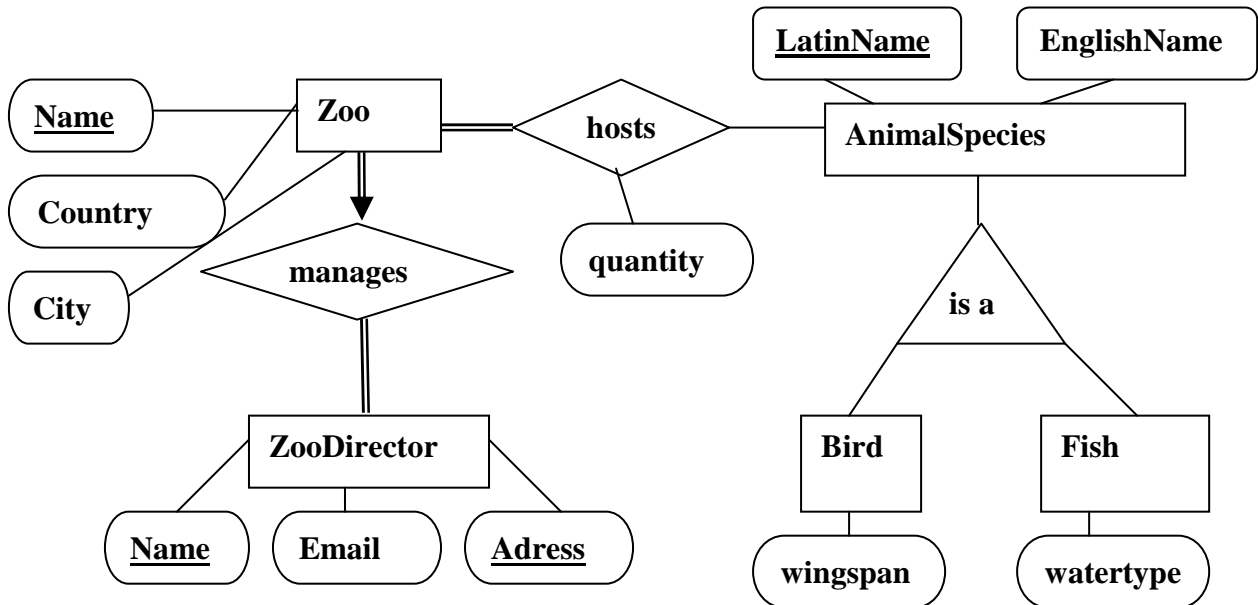


Part I: Entity Relationship Diagrams and SQL (40/100 Pt.)

Q.1.1) Translate the following E-R Schema to SQL-DDL tables using the CREATE Table Statement and check constraints, if needed:



Q.1.2) Consider you have to work with a database for the *e-Olympics*, a competition on computer (sport) games. The database you can use has the following schema:

Player (playerID, RealName, e-mail, age)
Game (GameName, Category, Highscore)
Played (playerID, GameName, Score, Time)
Medals (GameName, playerID, MedalType)

Write the following eight queries in SQL:

- Find GameNames for all games played by the players with RealName "John Lee"
- Find RealName of player(s) who won the gold medal for the game with name "Car Racing"
- Find playerID of all players who won at least one silver but not a gold medal
- Find playerID of all players who won at least one silver and at least one gold medal
- Find the playerIDs and realName of the youngest player, who won a gold medal
- Find the number of different GameNames for each category
- Find out all RealNames in the database which are not unique
- Find out all playerIDs and GameNames for played games in which the player reached the highscore

Part II: Schema Normalization (20/100)

Given a database with Schema:

R: ABCDEFGHIJK, and a set of eight functional dependencies:

- (a) $AB \rightarrow CDEFGHIJ$
- (b) $CD \rightarrow E$
- (c) $C \rightarrow F$
- (d) $EF \rightarrow J$
- (e) $J \rightarrow HI$
- (f) $J \rightarrow G$
- (g) $G \rightarrow HI$
- (h) $HI \rightarrow B$

Q.2.1) Find a key for the schema given the functional dependences

Q2.2) Which of the following three functional dependences is implied? Whenever the functional dependency is implied, prove this using Armstrong's Axioms, Union and/or Decomposition Rule:

- (a) $CD \rightarrow HI$
- (b) $CHI \rightarrow F$
- (c) $CDEFGHIJ \rightarrow AB$

Q2.3) Compute a lossless join decomposition of this schema in BCNF

Q2.4) Is this decomposition also dependency preserving?

Q2.5) Describe the difference between a 3NF and a BCNF.

Part III: Transaction management 20/100

Q.3.1) Describe ACID properties of transactions

Q.3.2) What is a locking protocol and why is it used? Describe the Two-Phase Locking (2PL) and the Strict Two-Phase Locking (Strict 2PL) protocol. What can you say about the schedules allowed by Strict 2PL protocol?

Q.3.3) Examine the schedule given below. There are four transactions, T1, T2, T3, and T4.

T1				R(A)	W(A)								Commit
T2								R(B)		W(B)			Commit
T3	R(A)		W(A)								R(B)	W(B)	Commit
T4		R(B)				R(A)	W(B)		W(A)				Commit

- Draw the precedence graph for this schedule.
- What is the equivalent serialization order for this schedule? If no order is possible, then state 'none'.
- Apply strict two-phase locking protocol to this schedule.
- Draw the 'waits-for' graph based on strict 2PL protocol, just after the deadlock.

Part IV: Questions (40/100)

- Q.4.1) Describe the three-tier architecture of a database-backed internet application and typical software tools used in each layer. What is the advantage of this architecture in comparison to thick client architecture?
- Q.4.2) Describe at least two problems related caused by redundancy in databases. What can be done to reduce redundancy in the database design phase?
- Q.4.3) XML and HTML are two formats used to exchange data in the internet. Explain briefly the main differences between XML and HTML and what they have in common.
- Q.4.4) Consider, your task is to tune the speed of a database. You can use three types of indexes and/or file organizations:

- (I1) Clustered B+Tree Index
- (I2) Unclustered B+Tree Index
- (I3) Sorted File
- (I4) Clustered Hash Index
- (I5) Unclustered Hash Index

Choose the appropriate index to speed up queries in the following situations and give reasons for your choice. It is possible to use more than one index/file-organization:

- (a) The database of the education administrative center contains for each bachelor student the courses she/he has passed. The number of courses is smaller than 20 and each course is described by a course ID. A frequent query is to ask for all courses a student has passed.
- (b) A scientific database contains data of endangered fish species observed in the river Danube. The total number of observations records is about 1000000. The observations are stored in a sorted file (sorted by time the observation was made). Frequently, researchers ask which species have been observed in a certain range of the river (e.g. all observations between Budapest and Belgrade).
- (c) The IP-Address database contains a large number of IP addresses and the address of their owners. A workload analysis has shown that by far the most frequent query is to find out the owner of an IP-address is (his/her address). Another, less frequent, query is to find out the IP address of an owner.

- Q.4.5) What is the main difference between an operational query language (relational algebra) and a declarative query language (e.g. SQL)? Given a schema $S(a,b)$ $T(b,c,d)$. Which of the following queries are equivalent?

Case1:

(SQL) `SELECT a,b FROM S, T WHERE S.b = T.b`

(RA) $S \bowtie T$

Case2:

(SQL) SELECT a,b,b,c,d FROM S, T

(RA) $S \times T$

Case 3:

(SQL) SELECT DISTINCT a FROM S,T WHERE T.b = S.b AND S.b='green'

(RA3) $\pi_a(\sigma_{b=\text{green}}(S \bowtie T))$

Case 4:

(SQL)

(SELECT DISTINCT b FROM S) INTERSECT (SELECT DISTINCT b FROM T)

(RA) $\pi_b(S) - (\pi_a(S) - \pi_b(S))$

Note: ' \bowtie ' stands for the natural join operation

Q.4.6) List at least five advantages of using a relational database management system as compared to storing the data in a table in one big ASCII file? Are there also disadvantages?

Q.4.7) Mark the correct statements about terms used in the context of databases. Multiple answers can be right

Q.4.8.1) Data Mining ...

- (a) is deductive system that estimates general rules based on given data
- (b) is a tool used to find patterns in large amounts data
- (c) can be used to compute the transitive closure for a set of relations

Q.4.8.2) A Relational Algebra Tree ...

- (a) is a unique representation of a SQL query
- (b) defines the way a query is processed
- (c) is a graphical representation of a relational algebra query
- (d) is often equivalent to many other RA trees

Q.4.8.3) A DTD is used ...

- (a) to check whether an XML file is valid or not
- (b) to check whether an XML file is well-formed or not
- (c) to define application specific constraints on the syntax of an XML file.
- (d) to specify the graphical representation of an XML file

Success!