

## Technical Papers Database Questions: Database Part I

**Examrace Placement Series** prepares you for the toughest placement exams to top companies.

### RDBMS Concepts

#### What is database?

A database is a logically coherent collection of data with some inherent meaning, representing some aspect of real world and which is designed, built and populated with data for a specific purpose.

#### What is DBMS?

It is a collection of programs that enables user to create and maintain a database. In other words it is general-purpose software that provides the users with the processes of defining, constructing and manipulating the database for various applications.

#### What is a Database system?

The database and DBMS software together is called as Database system.

#### Advantages of DBMS?

Redundancy is controlled.

Unauthorised access is restricted.

Providing multiple user interfaces.

Enforcing integrity constraints.

Providing backup and recovery.

#### Disadvantage in File Processing System?

Data redundancy & inconsistency.

Difficult in accessing data.

Data isolation.

Data integrity.

Concurrent access is not possible.

Security Problems.

Describe the three levels of data abstraction?

The are three levels of abstraction:

Physical level: The lowest level of abstraction describes how data are stored.

Logical level: The next higher level of abstraction, describes what data are stored in database and what relationship among those data.

View level: The highest level of abstraction describes only part of entire database.

Define the “integrity rules”

There are two Integrity rules.

Entity Integrity: States that Primary key cannot have NULL value

Referential Integrity: States that Foreign Key can be either a NULL value or should be Primary Key value of other relation.

What is extension and intension?

Extension-

It is the number of tuples present in a table at any instance. This is time dependent.

Intension-

It is a constant value that gives the name, structure of table and the constraints laid on it.

What is System R? What are its two major subsystems?

System R was designed and developed over a period of 1974 – 79 at IBM San Jose Research Center. It is a prototype and its purpose was to demonstrate that it is possible to build a Relational System that can be used in a real life environment to solve real life problems, with performance at least comparable to that of existing system.

Its two subsystems are

Research Storage

System Relational Data System.

How is the data structure of System R different from the relational structure?

Unlike Relational systems in System R

Domains are not supported

Enforcement of candidate key uniqueness is optional

Enforcement of entity integrity is optional

Referential integrity is not enforced

What is Data Independence?

Data independence means that the application is independent of the storage structure and access strategy of data. In other words, The ability to modify the schema definition in one level should not affect the schema definition in the next higher level.

Two types of Data Independence:

Physical Data Independence: Modification in physical level should not affect the logical level.

Logical Data Independence: Modification in logical level should affect the view level.

NOTE: Logical Data Independence is more difficult to achieve

What is a view? How it is related to data independence?

A view may be thought of as a virtual table, that is, a table that does not really exist in its own right but is instead derived from one or more underlying base table. In other words, there is no stored file that direct represents the view instead a definition of view is stored in data dictionary.

Growth and restructuring of base tables is not reflected in views. Thus the view can insulate users from the effects of restructuring and growth in the database. Hence accounts for logical data independence.

What is Data Model?

A collection of conceptual tools for describing data, data relationships data semantics and constraints.

What is E-R model?

This data model is based on real world that consists of basic objects called entities and of relationship among these objects. Entities are described in a database by a set of attributes.

What is Object Oriented model?

This model is based on collection of objects. An object contains values stored in instance variables with in the object. An object also contains bodies of code that operate on the object. These bodies of code are called methods. Objects that contain same types of values and the same methods are grouped together into classes.

What is an Entity?

It is a 'thing' in the real world with an independent existence.

What is an Entity type?

It is a collection (set) of entities that have same attributes.

What is an Entity set?

It is a collection of all entities of particular entity type in the database.

What is an Extension of entity type?

The collections of entities of a particular entity type are grouped together into an entity set.

What is Weak Entity set?

An entity set may not have sufficient attributes to form a primary key, and its primary key compromises of its partial key and primary key of its parent entity, then it is said to be Weak Entity set.

What is an attribute?

It is a particular property, which describes the entity.

What is a Relation Schema and a Relation?

A relation Schema denoted by  $R(A_1, A_2, A_n)$  is made up of the relation name  $R$  and the list of attributes  $A_i$  that it contains. A relation is defined as a set of tuples. Let  $r$  be the relation which contains set tuples  $(t_1, t_2, t_3, \dots, t_n)$ . Each tuple is an ordered list of  $n$ -values  $t = (v_1, v_2, \dots, v_n)$ .

What is degree of a Relation?

It is the number of attribute of its relation schema.

What is Relationship?

It is an association among two or more entities.

What is Relationship set?

The collection (or set) of similar relationships.

What is Relationship type?

Relationship type defines a set of associations or a relationship set among a given set of entity types.

What is degree of Relationship type?

It is the number of entity type participating.

What is DDL (Data Definition Language)?

A data base schema is specifies by a set of definitions expressed by a special language called DDL.

What is VDL (View Definition Language)?

It specifies user views and their mappings to the conceptual schema.

What is SDL (Storage Definition Language)?

This language is to specify the internal schema. This language may specify the mapping between two schemas.

What is Data Storage-Definition Language?

The storage structures and access methods used by database system are specified by a set of definition in a special type of DDL called data storage-definition language.

What is DML (Data Manipulation Language)?

This language that enable user to access or manipulate data as organised by appropriate data model.

Procedural DML or Low level: DML requires a user to specify what data are needed and how to get those data.

Non-Procedural DML or High level: DML requires a user to specify what data are needed without specifying how to get those data.

What is DML Compiler?

It translates DML statements in a query language into low-level instruction that the query evaluation engine can understand.

What is Query evaluation engine?

It executes low-level instruction generated by compiler.

What is DDL Interpreter?

It interprets DDL statements and record them in tables containing metadata.

What is Record-at-a-time?

The Low level or Procedural DML can specify and retrieve each record from a set of records. This retrieve of a record is said to be Record-at-a-time.

What is Set-at-a-time or Set-oriented?

The High level or Non-procedural DML can specify and retrieve many records in a single DML statement. This retrieve of a record is said to be Set-at-a-time or Set-oriented.

What is Relational Algebra?

It is procedural query language. It consists of a set of operations that take one or two relations as input and produce a new relation.

What is Relational Calculus?

It is an applied predicate calculus specifically tailored for relational databases proposed by E. F. Codd. E. g. Of languages based on it are DSL ALPHA, QUEL.

How does Tuple-oriented relational calculus differ from domain-oriented relational calculus

The tuple-oriented calculus uses a tuple variables i.e., variable whose only permitted values are tuples of that relation. E. g. QUEL

The domain-oriented calculus has domain variables i.e., variables that range over the underlying domains instead of over relation. E. g. ILL, DEDUCE.

What is normalization?

It is a process of analysing the given relation schemas based on their Functional Dependencies (FDs) and primary key to achieve the properties

Minimizing redundancy

Minimizing insertion, deletion and update anomalies.

What is Functional Dependency?

A Functional dependency is denoted by  $X \rightarrow Y$  between two sets of attributes  $X$  and  $Y$  that are subsets of  $R$  specifies a constraint on the possible tuple that can form a relation state  $r$  of  $R$ . The constraint is for any two tuples  $t_1$  and  $t_2$  in  $r$  if  $t_1[X] = t_2[X]$  then they have  $t_1[Y] = t_2[Y]$ . This means the value of  $X$  component of a tuple uniquely determines the value of component  $Y$ .

When is a functional dependency  $F$  said to be minimal?

Every dependency in  $F$  has a single attribute for its right hand side.

We cannot replace any dependency  $X \rightarrow A$  in  $F$  with a dependency  $Y \rightarrow A$  where  $Y$  is a proper subset of  $X$  and still have a set of dependency that is equivalent to  $F$ .

We cannot remove any dependency from  $F$  and still have set of dependency that is equivalent to  $F$ .

What is Multivalued dependency?

Multivalued dependency denoted by  $X \twoheadrightarrow Y$  specified on relation schema  $R$ , where  $X$  and  $Y$  are both subsets of  $R$ , specifies the following constraint on any relation  $r$  of  $R$ : If two tuples  $t_1$  and  $t_2$  exist in  $r$  such that  $t_1[X] = t_2[X]$  then  $t_3$  and  $t_4$  should also exist in  $r$  with the following properties

$$t_3[X] = t_4[X] = t_1[X] = t_2[X]$$

$$t_3[Y] = t_1[Y] \text{ and } t_4[Y] = t_2[Y]$$

$$t_3[Z] = t_2[Z] \text{ and } t_4[Z] = t_1[Z]$$

where  $[Z = (R - (X \cup Y)) ]$

What is Lossless join property?

It guarantees that the spurious tuple generation does not occur with respect to relation schemas after decomposition.

What is 1 NF (Normal Form)?

The domain of attribute must include only atomic (simple, indivisible) values.

What is Fully Functional dependency?

It is based on concept of full functional dependency. A functional dependency  $X \twoheadrightarrow Y$  is full functional dependency if removal of any attribute A from X means that the dependency does not hold any more.

What is 2NF?

A relation schema R is in 2NF if it is in 1NF and every non-prime attribute A in R is fully functionally dependent on primary key.

What is 3NF?

A relation schema R is in 3NF if it is in 2NF and for every FD  $X \twoheadrightarrow A$  either of the following is true

X is a Super-key of R.

A is a prime attribute of R.

In other words, if every non prime attribute is non-transitively dependent on primary key.

What is BCNF (Boyce-Codd Normal Form)?

A relation schema R is in BCNF if it is in 3NF and satisfies an additional constraint that for every FD  $X \twoheadrightarrow A$ , X must be a candidate key.

What is 4NF?

A relation schema R is said to be in 4NF if for every Multivalued dependency  $X \twoheadrightarrow Y$  that holds over R, one of following is true

X is subset or equal to (or)  $XY = R$ .

X is a super key.