

Database Systems Overview

1. What database models do you know?

- Hierarchical model – tree-like structure.
- Network / graph model – uses graphs
- Relational – the most used one. Uses tables and relations between them.
- Object-oriented – uses C# like objects.

2. Which are the main functions performed by a Relational Database Management System (RDBMS)

The main functions performed by a RDBMS are management of the data. This includes creating, altering, deleting tables and relationships between them (database schema), adding, hanging, deleting, searching and retrieving of data stored in the tables. It also supports the SQL language.

3. Define what is "table" in database terms.

- Table represents the structure the data will be stored. Table stores information organized in rows and columns.
- Row represents information about one record.
- Column represents a piece of information about certain record. Column has name and type.

4. Explain the difference between a primary and a foreign key.

- Primary key (Composite primary key) is a column(s) of the table that uniquely identifies the rows (records). Primary key is unique - meets only one time in certain column.
- The purpose of Foreign key is to avoid data duplicates in table columns. The repeating data is separate in new table as each record has primary key used in the master table. Foreign key is not unique - can be used many times in certain column (usually number instead strings e.g.).

5. Explain the different kinds of relationships between tables in relational databases.

- One-to-many (or many-to-one) - a single record in the first table has many corresponding records in the second table.
- Many-to-many - records in the first table have many corresponding records in the second one and vice versa. Implemented through additional table.
- One-to-one - a single record in a table corresponds to a single record in the other table. Used to model inheritance between tables

6. When is a certain database schema normalized? What are the advantages of normalized databases?

Normalization of the relational schema removes repeating data. The advantages are that by having less data repeated the overall storage space consumed by the database is less.

7. What are database integrity constraints and when are they used?

Database integrity constraints are rules that ensure that the data entered in the database is valid (by setting those validations rules in advance). For example we can constrain an age to not be less than 0 and greater than 150. When we try to save incorrect data into the DB we will get an error instead of invalid data.

8. Point out the pros and cons of using indexes in a database.

- *Pros:*
 - Faster lookup for results in certain column or group of columns (using index structures such as B-Trees or Hash Indexes to speed up searching of values).
 - Instead of scanning the entire table for the results.
- *Cons:*
 - Slower writes - adding / deleting records in indexed tables is slower.
 - May cause the system to restructure the index of structure (Hash Index, B-Tree, etc), which can be very computationally expensive.
 - Takes up more disk space - stores more data.

9. What's the main purpose of the SQL language?

- DDL - Data Definition Language:
 - Creating, altering, deleting tables and other objects in the database.
- DML - Data Manipulation Language:
 - Searching, retrieving, inserting, modifying and deleting table data (rows).

10. What are transactions used for? Give an example.

- Used of competitive data access.
- Sequence of operations executing as a single unit.
- Can be rolled back if they are not completed properly.

11. What is a NoSQL database?

- Use document-based model (non-relational)
- Data stored as documents
- Single entity (document) is a single record

- Documents do not have a fixed structure

12. Explain the classical non-relational data models.

- A non-relational database is a database that does not incorporate the table/key model that relational database management systems (RDBMS) promote.
- These kinds of databases require data manipulation techniques and processes designed to provide solutions to big data problems that big companies face.
- The most popular emerging non-relational database is called NoSQL (Not Only SQL).

13. Give few examples of NoSQL databases and their pros and cons.

MongoDB, Redis, CouchDB are the most famous NoSQL Databases. The pros and cons of using NoSQL are the following:

- *Pros:*
 - Mostly open source.
 - Horizontal scalability. There's no need for complex joins and data can be easily shared and processed in parallel.
 - Support for Map/Reduce. This is a simple paradigm that allows for scaling computation on cluster of computing nodes.
 - No need to develop fine-grained data model – it saves development time.
 - Easy to use.
 - Very fast for adding new data and for simple operations/queries.
 - No need to make significant changes in code when data structure is modified.
 - Ability to store complex data types in a single item of storage.
- *Cons*
 - Immaturity. Still lots of rough edges.
 - Possible database administration issues. NoSQL often sacrifices features that are present in SQL solutions “by default” for the sake of performance. Memory is usually heavily used.
 - No indexing support (Some solutions like MongoDB have indexing but it's not as powerful as in SQL solutions).
 - No ACID (Some solutions have just atomicity support on single object level).
 - Bad reporting performance.
 - Complex consistency models (like eventual consistency). NoSQL vendors are trying to make their solutions as fast as possible and consistency is most typical trade-off.
 - Absence of standardization. No standard APIs or query language. It means that migration to a solution from different vendor is more costly. Also there are no standard tools (e.g. for reporting)