Database

Concepts and Components

Scott Peters

https://advancedsqlpuzzles.com

Last updated 12/25/2021

Welcome. Here I have collected all the different database concepts, components, and aspects into one document; formatted into a list of questions with a link to an online resource (which is mostly the Wikipedia article on the subject) covering the subject.

I also provide the links as my intention is to not re-create the answers, as the Wikipedia articles offer a more complete resource then what I can provide.

The summary descriptions I have provided are often lifted with minimal changes from Wikipedia articles, vendor documentation, and various internet sites. I did try and find the best summary description to each question to help give the most concise answer possible.

I would like to thank all the individuals who contributed to the Wikipedia articles, created vendor documentation, or blogs for which this document is created from.

Many of these questions make for great interview questions. And some are for finding tidbits of information that you may not have come across in your database studies.

The latest version of this document can be found at https://advancedsqlpuzzles.com

1. ## What is SQL?

SQL (Structured Query Language) is the language of databases. First appearing in 1974, SQL is a set based, declarative language, strongly typed and cross platform. It is based on relational algebra and tuple calculus. SQL was one of the first commercial languages to use Edgar F. Codd's relational model.

https://en.wikipedia.org/wiki/SQL

2. ## SQL is a declarative language. Describe what this means.

Declarative programming is a programming paradigm— a style of building the structure and elements of computer programs—that expresses the logic of a computation without describing its control flow.

In short, with a declarative language you tell the computer what to do, not how to do it.

https://en.wikipedia.org/wiki/Declarative_programming

3. ## What is meant by a relation in the relational database model?

The relational database model is defined by the relation between a datasets attribute (column) and its tuple (row).

https://en.wikipedia.org/wiki/Relation_(database)

https://en.wikipedia.org/wiki/Relational_model

4. ## What is set theory and how does it relate to SQL?

Set theory is a branch of mathematical logic that studies sets; which informally are collections of objects. SQL is set based and constructed on relational algebra and tuple calculus. Set theory is credited to German mathematician Georg Cantor in the 1870s.

https://en.wikipedia.org/wiki/Set_theory

https://en.wikipedia.org/wiki/Georg_Cantor

5. ## What is the Barber Paradox in set theory?

An interesting paradox in set theory is the Barber Paradox. Consider a group of barbers who shave only those men who do not shave themselves. Suppose there is a barber in this collection who does not shave himself; then by the definition of the collection, he must shave himself. But

no barber in the collection can shave himself. If so, he would be a man who does shave men who shave themselves.

https://en.wikipedia.org/wiki/Barber_paradox

6.   What is relational algebra?

In database theory, relational algebra is a theory that uses algebraic structures with a well-founded semantics for modeling data and defining queries on it.

https://en.wikipedia.org/wiki/Relational_algebra

7.   What is relational calculus?

The relational calculus consists of two calculi, the tuple relational calculus and the domain relational calculus, that are part of the relational model for databases and provide a declarative way to specify database queries.

The relational calculus is similar to the relational algebra, which is also part of the relational model. While the relational calculus is meant as a declarative language which prescribes no execution order on the subexpressions of a relational calculus expression, the relational algebra is meant as an imperative language: the sub-expressions of a relational algebraic expressions are meant to be executed from left-to-right and inside-out following their nesting.

https://en.wikipedia.org/wiki/Relational_calculus

8.   What is tuple calculus?

Tuple calculus is a calculus that was created and introduced by Edgar F. Codd as part of the relational model to provide a declarative database-query language for data manipulation in this data model.

https://en.wikipedia.org/wiki/Tuple_relational_calculus

9.   What is domain calculus?

In computer science, domain relational calculus (DRC) is a calculus that was introduced by Michel Lacroix and Alain Pirotte as a declarative database query language for the relational data model.

https://en.wikipedia.org/wiki/Domain_relational_calculus

10.   Who is E.F. Codd?

Edgar Frank "Ted" Codd (19 August 1923 – 18 April 2003) was an English computer scientist who, while working for IBM, invented the relational model for database management, the theoretical basis for relational databases and relational database management systems.

https://en.wikipedia.org/wiki/Edgar_F._Codd

11. What are E.F. Codd's 12 rules for relational database management systems?

Codd's twelve rules are a set of thirteen rules (numbered zero to twelve) proposed by Edgar F. Codd, a pioneer of the relational model for databases, designed to define what is required from a database management system in order for it to be considered relational.

https://en.wikipedia.org/wiki/Codd%27s_12_rules

12. What is Codd's theorem?

Codd's theorem states that relational algebra and the domain-independent relational calculus queries, two well-known foundational query languages for the relational model, are precisely equivalent in expressive power. That is, a database query can be formulated in one language if and only if it can be expressed in the other.

https://en.wikipedia.org/wiki/Codd%27s_theorem

13. Does SQL have procedural extensions?

Yes, vendors add extensions to their SQL to include procedural programming language functionality, such as control-of-flow constructs.

https://en.wikipedia.org/wiki/SQL#Procedural_extensions

14. What are the different table joins?

Inner, Left Outer, Right Outer, and Cross Join are the common answers.

https://en.wikipedia.org/wiki/Join_(SQL)

But there are also natural joins, semi joins, anti joins, theta joins, equi joins and non equi joins. Each one of these is worthy of their own internet search.

15. Give a real-world example of a self-join.

A good example would be a table that has employee id and a manager id, where you need to find all employees and their associated manager.

https://en.wikipedia.org/wiki/Join_(SQL)#Self-join

16. Give a real-world example of a full outer join.

A good example of a full outer join is doing a comparison on two shopping carts to determine what items exist in both shopping carts, and what items exist only in their respective cart.

https://en.wikipedia.org/wiki/Join_(SQL)#Full_outer_join

17. Give a real-world example of a cross join.

A good example of a cross join is when you have a table of sales rep and the number of sales they have for the day. A sales rep may not have a sale every day. Here you can cross join the distinct dates from the table with the distinct sales employees to create a table of all possible sale dates and sale employees.

https://en.wikipedia.org/wiki/Join_(SQL)#Cross_join

18. What is a correlated sub-query?

In a SQL database query, a correlated subquery is a subquery (a query nested inside another query) that uses values from the outer query. Because the subquery may be evaluated once for each row processed by the outer query, it can be slow.

https://en.wikipedia.org/wiki/Correlated_subquery

19. What is meant by a restricted cartesian product?

Every join is a cartesian product, but inner, right, and left joins are restricted to their join condition.

The best reference I have found for describing how the SQL Server database engine processes a query is "*T-SQL Querying by Itzik Ben-Gan*".

https://www.amazon.com/T-SQL-Querying-Developer-Reference-Ben-Gan/dp/0735685045

20. What is a semi join? What is an anti join?

Semi join uses the IN operator and anti join uses the NOT IN operator.

https://en.wikipedia.org/wiki/Relational_algebra#Joins_and_join-like_operators

21. What is the benefit of using a semi join? What is the benefit of using a anti join?

Readability. The set of data jointed to the IN or NOT IN operator does not need GROUP BY statement and cannot be part of the SELECT statement (known as the queries projection).

https://en.wikipedia.org/wiki/Relational_algebra#Joins_and_join-like_operators

22. What are the different join algorithms?

Hash, Merge-Sort, and Nested. I highly recommend this YouTube video that visually explains the different join algorithms.

https://youtu.be/o1dMJ6-CKzU

23. What is a database execution plan?

An execution plan is a representation of the various steps that are involved in fetching results from the database tables.

https://en.wikipedia.org/wiki/Query_plan

24. What is a query hint?

A query hint is an addition to an SQL statement that instructs the database engine on how to execute the query. For example, a hint may tell the engine to use or not to use an index (even if the query optimizer would decide otherwise).

https://en.wikipedia.org/wiki/Hint_(SQL)

25. What is the logic processing order of a query?

FROM, WHERE GROUP BY, HAVING, SELECT, DISTINCT, ORDER BY, TOP

The best reference I have found for describing how the SQL Server database engine processes a query is "*T-SQL Querying by Itzik Ben-Gan".*

https://www.amazon.com/T-SQL-Querying-Developer-Reference-Ben-Gan/dp/0735685045

It should be noted each database engine (SQL Server, Oracle, MySQL, etc....) will process a query according to its own database engine rules. SQL is a declarative language and not an imperative

language; the underlying process of how it interprets a query is largely arbitrary to the developer (tunning considerations aside).

26. What is the difference between the WHERE clause and the HAVING clause?

The WHERE clause is processed before the GROUP BY, and the HAVING clause is processed after, allowing you to place predicate logic on groupings.

https://en.wikipedia.org/wiki/Where_(SQL)

https://en.wikipedia.org/wiki/Having_(SQL)

27. What is a MERGE statement?

A MERGE statement combines an UPDATE, INSERT and (depending on the RDBMS) DELETE syntax into one statement.

https://en.wikipedia.org/wiki/Merge_(SQL)

28. What is a stored procedure?

A stored procedure is pre-compiled collection of SQL statements and SQL command logic stored in a database. The main purpose of stored procedure is to hide direct SQL queries from the code and improve performance of database operations such as SELECT, UPDATE, and DELETE. Stored Procedures can be cached.

https://en.wikipedia.org/wiki/Stored_procedure

29. What is a trigger?

A database trigger is procedural code that is automatically executed in response to certain events on a particular table or view in a database.

https://en.wikipedia.org/wiki/Database_trigger

30. What is a view?

A view can be thought of as a stored query accessible as a virtual table. It can be used for retrieving data as well as updating or deleting rows. Views allow a preset way to view data from one or more tables.

https://en.wikipedia.org/wiki/View_(SQL)

31. What is the difference between a view and a materialized view?

In a materialized view, indexes can be built on any column. In a normal view, typically it's only possible to exploit indexes on columns that come directly from indexed columns in the base tables.

https://en.wikipedia.org/wiki/Materialized_view

32. What is a table valued function? What is the benefit of a table valued function?

A table-valued function is a user-defined function that returns data of a table type. The result set behaves as a table or view would. The benefit is that the table valued function can be parameterized (a value can be passed to the function). Think of a table valued function as a parameterized view.

https://en.wikipedia.org/wiki/User-defined_function

33. What is the difference between deterministic and non-deterministic functions?

Deterministic functions always return the same result any time they are called with a specific set of input values and given the same state of the database. Nondeterministic functions may return different results each time they are called with a specific set of input values, even if the database state that they access remains the same. For example, the function AVG always returns the same result given the qualifications stated above, but the GETDATE function, which returns the current datetime value, always returns a different result.

https://docs.microsoft.com/en-us/sql/relational-databases/user-defined-functions/deterministic-and-nondeterministic-functions

https://en.wikipedia.org/wiki/Nondeterministic_algorithm

https://en.wikipedia.org/wiki/Deterministic_system

34. What are the DDL, DML and DCL statements? What do these acronyms mean?

DDL is data definition language, DML is data manipulation language, and DCL is data control language.

DDL consists of CREATE, DROP, TRUNCATE and ALTER.

DML consists of SELECT, INSERT, UPDATE, DELETE
DCL consists of GRANT, DENY, REVOKE

https://en.wikipedia.org/wiki/Data_definition_language

https://en.wikipedia.org/wiki/Data_manipulation_language

https://en.wikipedia.org/wiki/Data_control_language

35. What is the difference between the DDL and DML statements?

DDL statements cannot be rolled back and when issued any open transactions are committed.

https://mariadb.com/kb/en/sql-statements-that-cause-an-implicit-commit/

However, every database engine acts slightly different.

https://www.mssqltips.com/sqlservertip/4591/ddl-commands-in-transactions-in-sql-server-versus-oracle/

36. What is the main purpose of common table expressions?

The main purpose is for performing recursion as CTEs are self-referencing objects.

https://en.wikipedia.org/wiki/Hierarchical_and_recursive_queries_in_SQL

37. What are aggregate functions?

An aggregate function performs a calculation on a set of values and returns a single value. Except for COUNT(*), aggregate functions ignore NULL markers. Aggregate functions are often used with the GROUP BY clause of the SELECT statement.

https://docs.microsoft.com/en-us/sql/t-sql/functions/aggregate-functions-transact-sql

38. What are window functions?

Window functions are calculation functions similar to aggregating, but where normal aggregating via the GROUP BY clause combines then hides the individual rows being aggregated. Window functions have access to individual rows and can add some of the attributes from those rows into the result set.

https://en.wikipedia.org/wiki/Window_function_(SQL)

39.   Name some common functions that can be windowed?

The most popular functions to window are COUNT, SUM, ROW_NUMBER, RANK, DENSE_RANK, LAG, and LEAD.

https://en.wikipedia.org/wiki/Window_function_(SQL)

40.   What are the four ranking functions in SQL?

RANK, DENSE_RANK, ROW_NUMBER, NTILE.

https://docs.microsoft.com/en-us/sql/t-sql/functions/ranking-functions-transact-sql

41.   What is the difference between ROLLUP and CUBE?

ROLLUP and CUBE are simple extensions to the SELECT statement's GROUP BY clause. ROLLUP creates subtotals at any level of aggregation needed, from the most detailed up to a grand total. CUBE is an extension similar to ROLLUP, enabling a single statement to calculate all possible combinations of subtotals.

https://docs.microsoft.com/en-us/sql/t-sql/queries/select-group-by-transact-sql

What are the different SET operators?

Set operations allow the results of multiple queries to be combined into a single result set. Set operators include UNION, INTERSECT, and EXCEPT.

https://en.wikipedia.org/wiki/Set_operations_(SQL)

42.   What is the difference between UNION and UNION ALL?

UNION removes duplicates. UNION ALL retains duplicates.

https://en.wikipedia.org/wiki/Set_operations_(SQL)

43.   What is a cursor?

A database cursor is a mechanism that enables traversal over the records in a database. Cursors facilitate subsequent processing in conjunction with the traversal, such as retrieval, addition, and removal of database records. The database cursor characteristic of traversal makes cursors akin to the programming language concept of iterator.

https://en.wikipedia.org/wiki/Cursor_(databases)

44.   What are the control flow statements in SQL?

Keywords for flow control in Transact-SQL include BEGIN and END, BREAK, CONTINUE, GOTO, IF and ELSE, RETURN, WAITFOR, and WHILE.

Control statements are SQL statements that allow SQL to be used in a manner similar to writing a program in a structured programming language. SQL control statements provide the capability to control the logic flow, declare and set variables, and handle warnings and exceptions.

https://en.wikipedia.org/wiki/Transact-SQL#Flow_control

45.   What is a bulk insert?

A bulk insert is a process or method provided by a database management system to load multiple rows of data into a database table.

https://en.wikipedia.org/wiki/Transact-SQL#BULK_INSERT

46.   What are some common SQL coding standards and best practices?

Common SQL standards include placing reserved words in uppercase, object names in lower case or camel case, using two-part naming conventions, adhering to indentation rules, etc.…

Here are two books dedicated to SQL coding standards and best practices that I highly recommend.

"*Oracle PL/SQL Best Practices: Write the Best PL/SQL Code of Your Life*"
https://www.amazon.com/Oracle-PL-SQL-Best-Practices-ebook/dp/B0026OR3NU

"*Joe Celko's SQL Programming Style*"
https://www.amazon.com/Celkos-Programming-Kaufmann-Management-Systems/dp/0120887975

47.   What are the different types of table indexes?

Indexes are used to quickly locate data without having to search every row in a database table every time a table is accessed. The three main types of indexes are clustered, non-clustered and

https://en.wikipedia.org/wiki/Database_index

48.   What is the difference between a clustered and non-clustered index?

- Cluster index is a type of index that sorts the data rows in the table on their key values whereas the non-clustered index stores the data at one location and indices at another location.
- Clustered index stores data pages in the leaf nodes of the index while non-clustered index method never stores data pages in the leaf nodes of the index.
- Cluster index doesn't require additional disk space whereas the non-clustered index requires additional disk space.
- Cluster index offers faster data accessing, on the other hand, non-clustered index is slower.

https://www.guru99.com/clustered-vs-non-clustered-index.html

49.   What is a columnstore index?

Columnstore indexes are the standard for storing and querying large data warehousing fact tables. This index uses column-based data storage.

https://docs.microsoft.com/en-us/sql/relational-databases/indexes/columnstore-indexes-overview

50.   What is a filtered index?

A filtered index is an index which has some condition applied to it (such as the WHERE clause) so that it includes a subset of rows in the table.

https://en.wikipedia.org/wiki/Partial_index

51.   What is a covered index?

A covered index is an index that contains all (and possibly more) the columns you need for your query. Most SQL syntax you will use the INCLUDED syntax to state which columns to include in the index.

https://en.wikipedia.org/wiki/Database_index#Covering_index

52.   What is the difference between an index seek and an index scan?

An index scan is when SQL Server scans the data or index pages to find the appropriate records. A scan is the opposite of a seek, where a seek uses the index to pinpoint the records that are needed to satisfy the query.

https://www.mssqltips.com/sqlservertutorial/277/index-scans-and-table-scans

53. What is a B-Tree structure?

The B-Tree structure provides a database engine with a fast way to move through the table rows based on an index key. A clustered index uses the B-Tree structure.

https://www.sqlshack.com/sql-server-index-structure-and-concepts/
https://en.wikipedia.org/wiki/B-tree

54. What is a database heap?

Heap files are lists of unordered records of variable size.

https://en.wikipedia.org/wiki/Database_storage_structures

55. What is a full table scan?

A full table scan is a scan where each row of the table is read in a sequential (serial) order and the columns encountered are checked for the validity of a condition.

https://en.wikipedia.org/wiki/Full_table_scan

56. What are the different normalization forms?

There are quite a few different normalization forms. At the least, a developer should know 1NF, 2NF and 3NF; any forms about these are generally considered academic in nature.

1NF - A relation is in first normal form if and only if the domain of each attribute contains only atomic (indivisible) values, and the value of each attribute contains only a single value from that domain.

2NF – A relation is in 2NF if it is in 1NF and every non-prime attribute of the relation is dependent on the whole of every candidate key.

3NF - A relation is said to meet 3NF standards if all the attributes are functionally dependent on solely the primary key.

https://en.wikipedia.org/wiki/Database_normalization
https://en.wikipedia.org/wiki/Unnormalized_form
https://en.wikipedia.org/wiki/First_normal_form
https://en.wikipedia.org/wiki/Second_normal_form
https://en.wikipedia.org/wiki/Third_normal_form
https://en.wikipedia.org/wiki/Elementary_key_normal_form
https://en.wikipedia.org/wiki/Boyce%E2%80%93Codd_normal_form

https://en.wikipedia.org/wiki/Fourth_normal_form
https://en.wikipedia.org/wiki/Fifth_normal_form
https://en.wikipedia.org/wiki/Domain-key_normal_form
https://en.wikipedia.org/wiki/Sixth_normal_form

57. What is the benefit of third normal form?

Third normal form reduces the duplication of data, avoid data anomalies, ensure referential integrity, and simplify data management.

https://en.wikipedia.org/wiki/Third_normal_form

58. What is database denormalization?

Denormalization is a strategy used on a normalized database to increase performance. In computing, denormalization is the process of trying to improve the read performance of a database, at the expense of losing some write performance, by adding redundant copies of data or by grouping data.

https://en.wikipedia.org/wiki/Denormalization

59. What is database normalization?

Database normalization is the process of structuring a database, usually a relational database, in accordance with a series of so-called normal forms in order to reduce data redundancy and improve data integrity. It was first proposed by Edgar F. Codd as part of his relational model.

https://en.wikipedia.org/wiki/Database_normalization

60. What is a slowly changing dimension?

A slowly changing dimension (SCD) keeps track of the history of its individual members. The Wikipedia article gives a good overview of SCDs. Note there are several different methods of implementation.

https://en.wikipedia.org/wiki/Slowly_changing_dimension

61. What is a weak entity?

A weak entity is an entity that cannot be uniquely identified by its attributes alone; therefore, it must use a foreign key in conjunction with its attributes to create a primary key. A good example is a table of hotel rooms that join to a hotel table.

https://en.wikipedia.org/wiki/Weak_entity

62. What is an associative entity?

An associative entity is the table that associates two other tables in a many to many relationship. A good example is an author can write several books, and a book can be written by several authors, creating a many to many relationship.

https://en.wikipedia.org/wiki/Many-to-many_(data_model)

https://en.wikipedia.org/wiki/Associative_entity

63. What is referential integrity? What is the benefit of a disabled referential integrity?

Referential integrity requires that if a value of one column references a value of another column, then the referenced value must exist.

The benefit of a disabled referential integrity is that the data model can be imported into a modeling program (such as Microsoft Visio) and the links between tables are maintained and auto created. It is often a best practice to put a disabled referential integrity on a schema design so developers can easily find the joins between tables.

https://en.wikipedia.org/wiki/Referential_integrity

64. What is the difference between OLAP and OLTP?

Online transaction processing (OLTP) captures, stores, and processes data from transactions in real time. Online analytical processing (OLAP) uses complex queries to analyze aggregated historical data from OLTP systems.

https://en.wikipedia.org/wiki/Online_transaction_processing

https://en.wikipedia.org/wiki/Online_analytical_processing

65. What is an OLAP cube?

OLAP cube is a multi-dimensional array of data.

An OLAP Cube is a data structure that allows fast analysis of data according to the multiple dimensions that define a business problem. A multidimensional cube for reporting sales might consist of: Salesperson, Sales Amount, Region, Product, Region, Month, Year.

https://en.wikipedia.org/wiki/OLAP_cube

66. What are the different keys in a database?

Primary Keys, Candidate Keys, Composite Keys, Alternative Keys, Super Keys, Natural Keys, Unique Keys, Surrogate Keys and Foreign Keys.

https://en.wikipedia.org/wiki/Primary_key
https://en.wikipedia.org/wiki/Candidate_key
https://en.wikipedia.org/wiki/Composite_key
https://en.wikipedia.org/wiki/Primary_key#Alternate_key
https://en.wikipedia.org/wiki/Superkey
https://en.wikipedia.org/wiki/Natural_key
https://en.wikipedia.org/wiki/Unique_key
https://en.wikipedia.org/wiki/Surrogate_key
https://en.wikipedia.org/wiki/Foreign_key

67. How do primary keys treat a NULL marker?

Primary Keys do not allow for a NULL marker.

https://en.wikipedia.org/wiki/Primary_key

68. What is an identity field?

An identity column is a column in a database table that is made up of values generated by the database.

https://en.wikipedia.org/wiki/Identity_column

69. What is an Entity-Relationship diagram?

An entity–relationship model describes interrelated things of interest in a specific domain of knowledge.

https://en.wikipedia.org/wiki/Entity%E2%80%93relationship_model

70. What is a dimensional fact model?

The dimensional fact model is an ad hoc and graphical formalism specifically devised to support the conceptual modeling phase in a DW project.

https://en.wikipedia.org/wiki/Dimensional_fact_model

71.   What is a conceptual model?

The goal of the conceptual data model is to establish the entities, their attributes, and their relationships.

https://en.wikipedia.org/wiki/Conceptual_schema

72.   What is a logical data model?

A logical data model defines the structure of the data elements and set the relationships between them.

https://en.wikipedia.org/wiki/Logical_schema

73.   What is a physical data model?

A physical data model describes the database specific implementation of the data model.

https://en.wikipedia.org/wiki/Physical_schema

74.   What is meant by the three-valued logic as it relates to SQL?

SQL uses a three-valued logic: besides true and false, the result of logical expressions can also be unknown.

- True and Unknown equates to Unknown
- False and Unknown equates to False
- Unknown and Unknown equates to Unknown
- True or Unknown equates to True
- False or Unknown equates to Unknown
- Unknown or Unknown equates to Unknown

Unknown could be true or false; if you take each equation and replace unknown with true, and then replace unknown with false, it the output changes between the two, then the equation evaluates too unknown.

https://en.wikipedia.org/wiki/Three-valued_logic#SQL

75.   What is a NULL marker?

NULL is a special marker used in Structured Query Language to indicate that a data value does not exist in the database.

https://en.wikipedia.org/wiki/Null_(SQL)

76.   What is an empty string?

An empty string is a zero-length string field. It should be noted it does not have an associated ASCII value.

Empty string ("") differs from a NULL marker in that the former specifically implies that the value was set to be empty, whereas NULL means that the value was not supplied or is unknown.

An example of a good use for an empty string would be Line2 of an address box that was intentionally left blank as it was not applicable.

https://en.wikipedia.org/wiki/Empty_string

77.   How does a UNIQUE constraint treat a NULL marker?

A UNIQUE constraint allows for a NULL marker unless a NOT NULL constraint is also issued.

https://en.wikipedia.org/wiki/Data_integrity

78.   What is cardinality of a table?

Cardinality refers to the uniqueness of data values contained in a particular column (tuple) of a database table.

https://en.wikipedia.org/wiki/Cardinality_(SQL_statements)

79.   What is a unary relationship? What is a ternary relationship?

Ternary relationship: a relationship type that involves many to many relationships between three tables.

Unary relationship: one in which a relationship exists between occurrences of the same entity set.

https://opentextbc.ca/dbdesign01/chapter/chapter-8-entity-relationship-model

80.   What are all the different data types?

In a RDBMS, each column, local variable, expression, and parameter has a related data type. A data type is an attribute that specifies the type of data that the object can hold. The most common ones are for integer data, character data, monetary data, date, and time data.

https://docs.microsoft.com/en-us/sql/t-sql/data-types/data-types-transact-sql?view=sql-server-ver15

81.   What is the difference between CHAR and VARCHAR datatypes?

CHAR stores the length at a fixed rate and pads spaces to fill any unused space. VARCHAR is able to store string values at a variable length and does not pad unused space, making it take less space.

https://www.geeksforgeeks.org/char-vs-varchar-in-sql/

82.   What is the difference between VARCHAR and NVARCHAR datatypes?

VARCHAR can store non-Unicode variable length characters. NVARCHAR can store both Unicode and Unicode characters (i.e., Japanese, Korean etc.)

https://www.mssqltips.com/sqlservertip/4322/sql-server-differences-of-char-nchar-varchar-and-nvarchar-data-types/

83.   What are all the different types of constraints you can place on a field?

NOT NULL, DEFAULT, CHECK, PRIMARY KEY, FOREIGN KEY, UNIQUE

Constraints are the rules enforced on the data columns of a table. These are used to limit the type of data that can go into a table and ensure the accuracy and reliability of the data in the database.

https://www.tutorialspoint.com/sql/sql-constraints.htm

84.   What is a calendar table?

A calendar table is a permanent table containing a list of dates and various components of those dates. These may be the result of DATEPART operations, time of year, holiday analysis, or any other creative operations we can think of.

A calendar table can save time, improve performance, and increase the consistency of data.

https://www.mssqltips.com/sqlservertip/4054/creating-a-date-dimension-or-calendar-table-in-sql-server/

85. What is a star schema?

The star schema is an approach most widely used to develop data warehouses and dimensional data marts. The star schema consists of one or more fact tables referencing any number of dimension tables.

https://en.wikipedia.org/wiki/Star_schema

86. What is a snowflake schema?

The snowflake schema is represented by centralized fact tables which are connected to multiple dimensions.

https://en.wikipedia.org/wiki/Snowflake_schema

87. What is business intelligence?

Business intelligence (BI) comprises the strategies and technologies used by enterprises for the data analysis of business information.

https://en.wikipedia.org/wiki/Business_intelligence

88. What is a data warehouse?

Data warehousing is the electronic storage of a large amount of information by a business or organization. A data warehouse is designed to run query and analysis on historical data derived from transactional sources for business intelligence and data mining purposes.

https://en.wikipedia.org/wiki/Data_warehouse

89. What is a dimension table?

A dimension is a structure that categorizes facts and measures to enable users to answer business questions. Commonly used dimensions are people, products, place, and time. In general, you can think of dimension tables as the elements you group and pivot on.

https://en.wikipedia.org/wiki/Dimension_(data_warehouse)#Dimension_table

90. What is a fact table?

Fact tables contain measurements of individual business processes. In general, you can think of fact tables as the metrics that you perform mathematical operations on (such as count, sum, avg).

https://en.wikipedia.org/wiki/Fact_table

91.   What is a junk dimension?

A junk dimension is a convenient grouping of typically low-cardinality flags and indicators. By creating an abstract dimension, these flags and indicators are removed from the fact table while placing them into a useful dimensional framework.

https://en.wikipedia.org/wiki/Dimension_(data_warehouse)#Junk_dimension

92.   What is a hierarchy in terms of a dimension?

Typically dimensions in a data warehouse are organized internally into one or more hierarchies. Date is a common dimension, with several possible hierarchies: Days, Weeks, Months, Years, etc.…

https://en.wikipedia.org/wiki/Dimension_(data_warehouse)#Dimension_table

93.   What is a conformed dimension?

Conformed dimensions are dimensions which can be used across any business area.

https://en.wikipedia.org/wiki/Dimension_(data_warehouse)#Conformed_dimension

94.   What is a data mart?

A data mart is a subset of a data warehouse oriented to a specific business line. Data marts contain repositories of summarized data collected for analysis on a specific section or unit within an organization, for example, the sales department.

https://en.wikipedia.org/wiki/Data_mart

95.   What are Bill Inmon's and Ralph Kimball's approach to data warehousing?

Bill Inmon's enterprise data warehouse approach (the top-down design): A normalized data model is designed first. Then the dimensional data marts, which contain data required for specific business processes or specific departments are created from the data warehouse.

Ralph Kimball's dimensional design approach (the bottom-up design): The data marts facilitating reports and analysis are created first; these are then combined to create a broad data warehouse.

https://en.wikipedia.org/wiki/Bill_Inmon

https://en.wikipedia.org/wiki/Ralph_Kimball

https://www.zentut.com/data-warehouse/kimball-and-inmon-data-warehouse-architectures/

96. What is data mining?

Data mining is a process of extracting and discovering patterns in large data sets.

https://en.wikipedia.org/wiki/Data_mining

97. What is an operational data storage?

An operational data store is used for operational reporting and as a source of data for the Enterprise Data Warehouse. It is a complementary element to an EDW in a decision support landscape, and is used for operational reporting, controls and decision making, as opposed to the EDW, which is used for tactical and strategic decision support.

https://en.wikipedia.org/wiki/Operational_data_store

98. What is a hash function?

A hash is a number that is generated by reading the contents of a document or message. Different messages should generate different hash values, but the same message causes the algorithm to generate the same hash value. They are generally used to create a unique key based upon a one or more field inputs.

https://en.wikipedia.org/wiki/Hash_function

99. What is an overloaded function?

Overloading allows you to substitute a subtype value for a formal parameter that is a supertype. This capability is known as substitutability. An example is the COUNT function, where you can pass different data types (number, varchar, etc....) and the routine knows how to handle the data type. Not all RDBMS support overloading, such as SQL Server.

https://docs.oracle.com/en/database/oracle/oracle-database/12.2/adobj/use-of-overloading-in-plsql-with-inheritance.html

100. What is database serialization?

Serialization is the process of converting an object into a stream of bytes to store the object or transmit it to memory, a database, or a file. Its main purpose is to save the state of an object in order to be able to recreate it when needed.

https://en.wikipedia.org/wiki/Serialization

101. What is a database partitioning?

Partitioning is the database process where extremely large tables are divided into multiple smaller parts. By splitting a large table into smaller, individual tables, queries that access only a fraction of the data can run faster because there is less data to scan.

https://en.wikipedia.org/wiki/Partition_(database)

102. What is concurrency control?

Data concurrency is the ability to allow multiple users to affect multiple transactions within a database. Data concurrency allows multiple users to access data at the same time. The ability to offer concurrency is unique to databases.

https://en.wikipedia.org/wiki/Concurrency_control

103. What are isolation levels?

Transactions specify an isolation level that defines the degree to which one transaction must be isolated from modifications made by other transactions. Isolation levels are described in terms of which concurrency side effects, such as dirty reads or phantom reads, are allowed.

The four isolation levels are Serializable, Repeatable Reads, Read Committed, Read Uncommitted.

https://en.wikipedia.org/wiki/Isolation_(database_systems)#Isolation_levels

104. What is a database commit?

A COMMIT statement in SQL ends a transaction within a relational database management system and makes all changes visible to other users.

https://en.wikipedia.org/wiki/COMMIT_(SQL)

105. What is a rollback?

A rollback is an operation which returns the database to some previous state.

https://en.wikipedia.org/wiki/Rollback_(data_management)

106. What is a database checkpoint?

A checkpoint creates a known good point from which a database engine can start applying changes contained in the transaction log during recovery after an unexpected shutdown or crash.
https://docs.microsoft.com/en-us/sql/relational-databases/logs/database-checkpoints-sql-server

107. What are temporal tables?

Temporal tables are a database feature that brings built-in support for providing information about data stored in the table at any point in time rather than only the data that is correct at the current moment in time.

https://en.wikipedia.org/wiki/Temporal_database

https://docs.microsoft.com/en-us/sql/relational-databases/tables/temporal-tables

108. What is meant by ACID compliance?

ACID (atomicity, consistency, isolation, durability) is a set of properties of database transactions intended to guarantee data validity despite errors, power failures, and other mishaps.

https://en.wikipedia.org/wiki/ACID

109. What is a database transaction?

A database transaction symbolizes a unit of work performed within a database management system (or similar system) against a database and treated in a coherent and reliable way independent of other transactions. A transaction generally represents any change in a database.

https://en.wikipedia.org/wiki/Database_transaction

110. What is a record lock?

Record locking is the technique of preventing simultaneous access to data in a database. Their objective is to prevent inconsistent results. The two types of locks are shared and exclusive.

https://en.wikipedia.org/wiki/Record_locking

111. What is a deadlock?

A database deadlock is a situation in which two or more transactions are waiting for one another to give up locks.

https://en.wikipedia.org/wiki/Deadlock

112. What is a dirty read?

A dirty read occurs when a transaction is allowed to read data from a row that has been modified by another running transaction and not yet committed.

https://en.wikipedia.org/wiki/Isolation_(database_systems)#Dirty_reads

113. What is a transaction log?

In the field of databases in computer science, a transaction log (also transaction journal, database log, binary log or audit trail) is a history of actions executed by a database management system used to guarantee ACID properties over crashes or hardware failures.

https://en.wikipedia.org/wiki/Transaction_log

114. Who is Jim Gray and what are some of his contributions to databases?

James Nicholas Gray (1944 – declared dead in absentia 2012) was an American computer scientist who received the Turing Award in 1998 "for seminal contributions to database and transaction processing research and technical leadership in system implementation".

- ACID, an acronym describing the requirements for reliable transaction processing and its software implementation
- Granular database locking
- Two-tier transaction commit semantics
- The five-minute rule for allocating storage
- OLAP cube operator for data warehousing

https://en.wikipedia.org/wiki/Jim_Gray_(computer_scientist)

115. What is change data capture?

Change data capture records insert, update, and delete activity that applies table. This makes the details of the changes available in an easily consumed relational format.

https://en.wikipedia.org/wiki/Change_data_capture

116. What are temporary tables?

A temporary table is a database table that exists temporarily on the database server. They provide a workspace for the intermediate results when processing data within a batch or procedure. There are two types of temporary tables, session and global.

https://en.wikibooks.org/wiki/Structured_Query_Language/Temporary_Table

117. What are the catalog/meta-data/information schema views in a database?

The system catalog consists of tables/views describing the structure of objects such as databases, base tables, views, and indices.

https://en.wikipedia.org/wiki/Database_catalog

118. What organization sets the SQL standards?

SQL was adopted as a standard by the ANSI in 1986 as SQL-86[28] and the ISO in 1987.

https://en.wikipedia.org/wiki/SQL#Standardization_history

119. What are a few popular RDBMS systems?

SQL Server, Oracle, MySQL, DB2, PostgreSQL, SQLite, Teradata and MariaDB are some of the most popular right now.

https://en.wikipedia.org/wiki/List_of_relational_database_management_systems

120. A relational database is one type of database, what are some other types of databases?

Flat File, Hierarchical, Graph, and Relational are a few of the different models. For a complete listing see the Wikipedia article and their relevant links.

https://en.wikipedia.org/wiki/Database_model

121. What are the benefits of a NoSQL database over a RDBMS?

The structure of many different forms of data is more easily handled and evolved with a NoSQL database. NoSQL databases are often better suited to storing and modeling structured, semi-structured, and unstructured data in one database.

https://en.wikipedia.org/wiki/NoSQL

122.   What is a database application?

A database application is a computer program whose primary purpose is entering and retrieving information from a computerized database. Examples include YouTube, CNN, Facebook and Wikipedia.

https://en.wikipedia.org/wiki/Database_application

123.   What is data independence?

Data independence is defined as a property of DBMS that helps you to change the database schema at one level of a database system without requiring to change the schema at the next higher level. Data independence helps you to keep data separated from all programs that make use of it.

https://en.wikipedia.org/wiki/Data_independence

124.   What are sparse columns?

Sparse columns are columns that have an optimized storage for NULL markers. Sparse columns reduce the space requirements for NULL markers at the cost of more overhead to retrieve non-NULL markers.

https://docs.microsoft.com/en-us/sql/relational-databases/tables/use-sparse-columns

125.   What is the difference between row and page compression?

Row compression is a subset of page compression and reduces the amount of space by reducing the metadata overhead, some numeric data types are reduced to save space, and NULL and 0 values take up zero space on a page. Page compression comprises of row compression, prefix compression and dictionary compression.

https://docs.microsoft.com/en-us/sql/relational-databases/data-compression/data-compression

http://blogs.lessthandot.com/index.php/datamgmt/dbprogramming/how-sql-server-data-compression/

126. <u>What is the database engine?</u>

A database engine is defined as the software that recognizes and interprets the SQL commands to access a relational database and interrogate data.

The database engine has two major components: the storage engine and the query processor. The storage engine writes data to and retrieves data from stable media (e.g., disks). The query processor accepts, parses, and executes SQL commands.

https://en.wikipedia.org/wiki/Database_engine

127. <u>What is meant by parsing?</u>

Parsing means examining the characters input and recognizing it as a command or statement by looking through the characters for keywords and identifiers, ignoring comments, arranging quoted portions as string constants, and matching the overall structure to the language syntax making sense of it all. Parsing determines the best way (of all the possible ways) to determine the result, usually with respect to execution time.

https://en.wikipedia.org/wiki/Parsing
https://docs.oracle.com/database/121/TGSQL/tgsql_sqlproc.htm#TGSQL175

128. <u>What is collation in the context of databases?</u>

In database systems, collation specifies how data is sorted and compared in a database. Collation provides the sorting rules, case, and accent sensitivity properties for the data in the database.

For example, when you run a query using the ORDER BY clause, collation determines whether uppercase letters and lowercase letters are treated the same.

https://database.guide/what-is-collation-in-databases/

https://en.wikipedia.org/wiki/Collation

129. <u>What is SQL injection?</u>

SQL injection is a code injection technique used to attack data-driven applications, in which malicious SQL statements are inserted into an entry field for execution

https://en.wikipedia.org/wiki/SQL_injection

You have made it to the end!

If you feel I've left anything out, please contact me at scottpeters1188@outlook.com or through my blog at https://advancedsqlpuzzles.com.