

Review

CSC230 Database Technologies for Analytics

09 November 2016

1. In his Turing Award lecture, Michael Stonebraker explained that he wanted to create a database management system that could store more than words and numbers. What other kind of data did he think was important?
-

Stonebraker and his collaborators built a system that could work with geometric data. The software that they created could be used in geographic information systems and computer-aided systems.

2. The largest value that can be stored in a field whose type **INTEGER** is 2,147,483,647. Why is the upper bound not an easier to remember value like 999,999,999 or 1,000,000,000 or 1,000,000,000,000?
-

The computer stores numerical values in binary (base 2) form.

The computer allocates 32 bits for integer variables. With 32 binary digits, it is possible to represent values in the range -2^{31} to $2^{31} - 1$.

3. Computability, complexity, and correctness are three important concepts and subjects of study in computer science.

Which of these three “C” words best applies in each of the following cases?

- (a) With some sorting algorithms, a computer will execute a number of instructions that is roughly proportional to N^2 , where N is the number of items in the list that is being sorted. With other sorting algorithms, a computer will execute a number of instructions that is roughly proportional to $N \log N$.
- (b) There are computer programs that control vehicles, power plants, and robotic surgical instruments. The software engineers who write these programs bear responsibility for the protection of lives and property.

- (c) Alan Turing proved the impossibility of writing a program that can check any other program to determine whether or not its execution with a given input will terminate in finite time.
-

- (a) The study of complexity is the study of the efficiency of algorithms.
(b) Efforts to prove the correctness of computer programs are, of course, especially important in the design of software whose failure could cause the loss of life or property.
(c) Turing proved that the Halting Problem is non-computable. The study of computability is an effort to determine which questions can be answered by computer programs, and which cannot.

4. What are the advantages of using a file system as a database?

- ubiquitous, already available, no separate purchase
- no special expertise required

5. What are the disadvantages of using a file system as a database?

- searching is likely to be slow
- simultaneous modifications by two or more people may produce undesired results
- interrupted operations can produce undesired results
- searching and sorting in new ways could be very difficult

6. An ACID database management system has four properties:

- Atomicity
- Consistency
- Isolation
- Durability

Which of these four properties applies in each of the following cases?

- (a) The system can retrieve/recover data in any circumstances, even after the failure of hardware.
 - (b) The system does not allow to one operation to see the partial (incomplete) results of another operation.
 - (c) The system does not allow any operation that does not respect constraints that define the integrity of the database. For example, it forbids the deletion of a row in one table when a row in another table contains a reference to the first row.
 - (d) The system erases/reverses all effects of any operation that it is unable to complete.
-

- (a) Durability.
- (b) Isolation.
- (c) Consistency.
- (d) Atomicity.

7. What is “brittle” software?

Brittle software works well for the one purpose for which it was designed but not for purposes whose importance its users discover later.

8. What are some advantages of using a declarative language?

- we can construct a query without any knowledge of how the data is represented and organized in the computer’s memory
- we can compose more reliable software—a query is unlikely to harbor some subtle bug that manifests itself only rarely—the query will likely produce the desired result everytime or fail in an obvious on the first try

9. What, according to Philip Greenspun, does a caveman programmer need to know about relational databases?

A relational database is a spreadsheet that many people can modify at the same time.

10. What is the value of an index in an RDBMS?

Indices can be used to greatly improve performance.

11. (a) Give an example of a SQL statement that is part of the Data Definition Language?
(b) Give an example of a SQL statement that is part of the Data Manipulation Language?

-
- (a) **CREATE TABLE**
(b) **SELECT, INSERT**

12. Philip Greenspun identifies an alternative to relationship databases and a drawback to that alternative.
(a) What is the alternative?
(b) What is the drawback?

-
- (a) Object databases are an alternative to relational databases.
(b) Programmers create applications that draw upon the services that a database management systems provide. If the system manages a relational database, then the programmers interact with the database only through SQL. Bugs are unlikely to damage the database. If the system manages an object database, then the programmers manipulate the database in a deeper way that makes damage to the database a more probable consequence of programming errors.