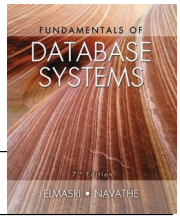
Database System Concepts



Chapter 1

REVISION 1

Terms (GNU AGPLv3)

Chapter 1 | Review Questions

1. Define the following terms:

1. **Data**
   Data is referred as known facts that can be recorded and have some implicit meaning.
2. **Database**
   Database is a collection of related data.
3. **DBMS**
   A DBMS is a computerized system that enables user to create and maintain a database.
4. **Database system**
   Database system referes to the combination of a database, a DBMS and application that uses the data.
5. **Database catalog**
   Database catalog referes to the directory that stores metadata like table structures and constrains.
6. **Program-data independence**
   Program data independence refers to the ability where we can change the database structure without affecting the programs that use it.
7. **User view**
   User view refers to the part where an user interacts with the database.
8. **DBA**
   DBA or database administrator refers to the role where they assign the roles of other database users.
9. **End user**
   End users are the people whose jobs require access to the database for querying, updating, and generating reports; the database primarily exists for their use.
10. **Canned transaction**
    Canned transaction refers to the standard types of queries and updates which are used constantly.
11. **Deductive database system**
    Deductive database system is a system where data is stored with logical reasoning capabilities.
12. **Persistent object**
    Persistent object refers to the object which remains stored, even after closing the application that created the object.
13. **Meta-data**
    Meta data is the information that is stored in the DBMS catalog, where structure of all files, file-type and constrains are stored.
14. **Transaction-processing application**
    Transaction processing application refers to the real time applications that needs to update the data frequently.

2. What four main types of actions involve databases? Briefly discuss each.

Four types of actions are,

1. **Create:** We need to insert new information.
2. **Read:** We may need to read the existing data.
3. **Update:** We may need to update some information in our database.
4. **Delete:** We may also need to delete some entry from our database if necessary.

3. Discuss the main characteristics of the database approach and how it differs from traditional file systems.

Main characteristics of database includes central file processing. Here all applications can use the central database for data. Besides data is stored as program independent manner. So any changes in the DBMS or the application won't break the data.

In the other way, traditional file systems use multiple files for storing data. So duplicates of the same data may occur. And we also have to separately write our applications for accessing each of these files, which may make the entire process more time-consuming.

4. What are the responsibilities of the DBA and the database designers?

DBA or database administrator refers to the role where they assign the roles of other database users. They usually maintain the whole database system by allocting each users roles.

In the other hand, **database designers** usually design the whole data scheme and relations concept which will be used by the database DDL. As it is a bit complicated to change the system after once creating the data definition structure, so this process requires a lot of domain knowledge.

5. What are the different types of database end users? Discuss the main activities of each.

End users are the people whose jobs require access to the database for querying, updating, and generating reports; the database primarily exists for their use. Some of them are discussed below:

1. **Causal end users:** Causal end users access the database occasionally. But they may need different types of data each time.
2. **Naive end users:** Naive users are users who constantly update or query the database. They don't have to have much knowledge about the database.
3. **Sophisticated end users:** Sophisticated end users refers to the people who thoroughly study the database, so that they can implement it in their application.
4. **Standalone users:** Standalone users refers to the personal databases by using ready-made programs that provide user-friendly interface.

6. Discuss the capabilities that should be provided by a DBMS.

A DBMS should provide following capabilities:

1. Data definition (DDL)
2. CRUD operation
3. Data integrity and constrains auto checkup
4. Security and user roles
5. Concurrency control
6. Transaction Management
7. Backup and restore support
8. Performance and optimization
9. Data sharing and appropriate language support
10. Scalability and the ability to deploy in network

7. Discuss the differences between database systems and information retrieval systems.

Database system	Information retrieval system
Manages structure data	Data may be unstructured
Needs SQL or predefined rule	Keyword based search
Uses SQL for generating the same results	Uses ranking system for retrieving information
Uses tables, constrains	Uses files for linking
Focuses on accuracy	Focuses on precision
Can be used for data storing	Can be used for information recommendation system
For example, MySQL, MongoDB, Oracle DB	For example, Google or bing or any modern search engine

8. Identify some informal queries and update operations that you would expect to apply to the database shown in Figure 1.2.

Some informal queries that I may apply are,

1. Getting the grade of a particular student
2. Generating a report of a grade sheet of all students with same class
3. Listing the prerequisites of a particular course (if required)
4. List all instructor and courses of a particular class

9. What is the difference between controlled and uncontrolled redundancy? Illustrate with examples.

In a **controlled redundancy system** instead of duplicating a data in RDBMS, we use reference to the parent data, so that we don't need to store multiple value of the same data. On the other hand, in **uncontrolled redundancy** same data is copied multiple times, leading to much more storage and no auto update support.

For example, in a library management system, we have to put the book name on two tables, one bookshelf and one another is to the loan table. But if controlled redundancy is used then, changing the title of the book will update the other tables as well. On the other hand if uncontrolled redundancy is used then, we may need to manually update the name of the book in all tables, which may lead to further problems.

10. Specify all the relationships among the records of the database shown in Figure 1.2.

Here among multiple tables, we can construct some relations like,

1. *Course_number* in prerequisite table refers to the *Course_number* of section table.
2. *Prerequisite_number* in prerequisite table refers to the *Course_number* of section table.
3. *Student_number* in grade report refers to the *Student_number* in Student table.
4. *Section_number* in grade report refers to the *Section_number* in Section table.
5. *Course_number* in section table refers to the *Course_number* of section table.

11. Give some additional views that may be needed by other user groups for the database shown in Figure 1.2.

Some additional views that we can implement are,

1. Grade report per student
2. Instructors list per class
3. All courses list per year or class

12. Cite some examples of integrity constraints that you think can apply to the database shown in Figure 1.2.

Some constrains that we can apply to the database are,

1. Year can be between 1 and 8
2. Course number must have 6 alphanumeric digit
3. Semester should be either “Fall” or “Spring”
4. Major must be “CS” or “MATH” or in a set of values.
5. Name’s length should be within a certain range.
6. Credit hour’s value should be between 0 and an integer, like 4.

13. Give examples of systems in which it may make sense to use traditional file processing instead of a database approach.

Some examples where file processing instead of a databse approach is much efficient are,

1. **Media storage:** In media storage, generally files don’t change that much frequently. And mostly they are stored as BLOB in the databases. So in this case file storage is much more efficient.
2. **Single user application:** In single user applications like portable applications, concurrency is not generally used, so we can use single files instead of database. In this way we can also back up our data more accurately.

14. Consider Figure 1.2.

a. If the name of the ‘CS’ (Computer Science) Department changes to ‘CSSE’ (Computer Science and Software Engineering) Department and the corresponding prefix for the course number also changes, identify the columns in the database that would need to be updated.

If the department name changes then we have to update the *Department* column of COURSE table. And we also have to update the Course_number in COURSE, SECTION and PREREQUISITE table.

b. Can you restructure the columns in the COURSE, SECTION, and PREREQUISITE tables so that only one column will need to be updated?

We can implement primary key in COURSE table and corresponding foreign key in SECTION and PREREQUISITE table, so that we only need to update the COURSE table.

"Simplicity is the easiest path to true beauty."
–Seishuu Handa, Barakamon