Topic 1.2: Introducing Databases and the DBMS

What data inconsistency is and why it occurs?

Data inconsistency exists when different and conflicting versions of the same data appear in different places. Data inconsistency creates unreliable information, because it will be difficult to determine which version of the information is correct. (It's difficult to make correct – and timely – decisions if those decisions are based on conflicting information.)

Data inconsistency is likely to occur when there is data redundancy. Data redundancy occurs when the data file/database file contains redundant – unnecessarily duplicated – data. That's why one major goal of good database design is to eliminate data redundancy.

Database— is shared, integrated computer structure that houses:
- End user data (raw facts)
- Metadata (data about data)

1.2.1 The different types of databases.

Databases are classified based on the number of users, database site location, and data use:
- Number of users
  - Single-user
  - Multiuser
  - Workgroup
  - Enterprise
- Database site location
  - Centralized
  - Distributed
- Database use
  - Transactional (production) database
  - Data warehouse database

What are the main components of a database system environment?

The basis of this discussion is Section 1.6.1, THE DATABASE SYSTEM ENVIRONMENT. Figure 1.2, from the text book, provides a good bird's eye view of the components. Note that the system's components are hardware, software, people, procedures, and data.
What is metadata?

Metadata is data about data. That is, metadata define the data characteristics such as the data type (such as character or numeric) and the relationships that link the data. Relationships are an important component of database design. What makes relationships especially interesting is that they are often defined by their environment.

For instance, the relationship between EMPLOYEE and JOB is likely to depend on the organization’s definition of the work environment. For example, in some organizations an employee can have multiple job assignments, while in other organizations – or even in other divisions within the same organization – an employee can have only one job assignment.

The details of relationship types and the roles played by those relationships in data models are defined and described in Chapter 2. Relationships will play a key role in subsequent chapters. You cannot effectively deal with database design issues unless you address relationships.
The differences between transaction databases and data warehouse databases.

A transaction (production) database reflects the daily operations of an organization. As the name implies, a transaction database records transactions such as the sale of a product, the enrollment of a student, or the opening of a checking account. Such transactions are time-critical and must be recorded accurately and immediately. Given their critical role in the day-to-day operations – and given the fact that transaction database data are usually the data source for data warehouse databases – much of the focus on the design of transaction databases.

Data warehouse databases focus primarily on the storage of data to be used as the information basis for tactical and strategic decision making. Such decision making typically requires extensive data manipulation to extract information from data. In short, the data warehouse data are the source for decision support systems.

Data warehouse databases are far less time-critical than transaction databases; the former record the historical data record that will be used to formulate pricing decisions, sales forecasts, market positioning, and so on.

Data warehouse databases typically store complex data derived from many sources. To make it easier to store and retrieve such complex data, the data warehouse database structure ends to be quite different from that of the transaction database. You will see in Chapter 12, “The Data Warehouse,” how database design is affected by the database role.

Concept Check

What data inconsistency is and why it occurs?
What are the different types of databases?
What are the main components of a database system environment?
What is metadata?
What is the differences between transaction databases and data warehouse database?