Database Systems

Design, Implementation and Management

Chapter 1 Notes

1.1 Why Database?

- Allows companies and businesses to store and access information on topics within the business's internal and external environment.
- Databases are specialized structures that allow computer-based systems to store, manage, and retrieve data very quickly.
- Databases will be a key component underpinning the systems that you use.

1.2 Data vs. Information

- Data: Raw facts
 - Facts have not yet been processed to reveal their meaning.
 - Not very useful in this format.
- Information: Result of processing raw data to reveal its meaning.
 - As simple as organizing data to reveal patterns.
 - To reveal meaning, information requires context.
 - Ex. 'Average temperature reading of 105 degrees'
 - Fahrenheit or Celsius?
 - Machine, body or outside temperature?
- Raw data must be properly formatted for storage, processing, and presentation.
 - AKA Computer Language Yes or No \rightarrow Y/N or 1/0
- Data are the foundation of information, which is the bedrock of knowledge
- Knowledge body of information and facts about a specific subject.
 - Knowledge implies familiarity, awareness, and understanding.
 - 'New' knowledge can be derived from 'old' knowledge.
- Key points:
 - \circ $\;$ Data constitutes the building blocks of information.
 - Information is produced by processing data.
 - \circ $\;$ Information is used to reveal the meaning of data.
 - Accurate, relevant, and timely information is the key to good decision making.
 - o Good decision making is the key to organizational survival in a global environment.
- Data management: Discipline that focuses on the proper generation, storage, and retrieval of data.
 - Core activity.

1.3 Introducing The Database

- Data management usually requires the use of a computer database.
- **Database**: Shared, integrated computer structure that stores:
 - End-user data Raw facts of interest to the end user.
 - **Metadata** (or data about data), through which the end-user data are integrated and managed.
 - Describes the characteristics and set of relationships that links the data found within the database.
 - Ex. Stores information such as the name of each data element, types of value, etc.
 - Information that complements and expands the value and use of the data.
 - Metadata presents a more complete picture of the data in the database.
- Database management system (DBMS): A collection of programs that manage the database structure and controls the access to the data stored in the database.

1.3.1 Role and Advantages of the DBMS

- Serves as the intermediary between the user and database
- DBMS receives all application requests (from end users or application program) and translates them into complex operations to fulfill those requests.
- DBMS hides the translation of the programming language such as Visual Basic .Net, Java, or C# Only showing the end user a single, integrated view of the data in the database.
- Advantages of having a DBMS between the end user and the database:
 - DBMS enables the data in the database to be shared among multiple users or applications.
 - DBMS integrates the many different users' views of the data into a single allencompassing data repository.
- DBMS helps make data management more efficient and effective.
 - Improved data sharing: Helps create an environment where end users have better access to more and better-managed data.
 - Improved data security: Provides a framework for better enforcement of data privacy and security policies, since many users access the data, it increases the risk of data security breaches.
 - Better data integration: Wider access to well-managed data makes it much easier to see how actions in one segment of the company affect other segments.
 - Minimized data inconsistency:
 - Data inconsistency exists when different versions of the same data appear in different places.
 - Probability of data inconsistency is greatly reduced in a properly designed database.
 - Improved data access:

- Produce quick answers to ad hoc queries.
- **Query**: Specific request issued to the DBMS; A question.
- Ad hoc query: Spur-of-the-moment question
- Query result set: The answer sent back by the DBMS.
- Improved decision making:
 - Better managed data and improved data access make it possible to generate better quality information, on which decisions are based.
 - Data quality is necessary in promoting accuracy, validity and timelines of the data.
- Increased end-user productivity:
 - Usable information from the data allows end users to make quick informed decisions.
- Many more unmentioned advantages to using a DBMS.

1.3.2 Types of Databases

- DBMS used to build many different types of databases.
- Each database stores a particular collection of data and is used for a specific purpose.
- Number of Users:
 - Single-user database: Only one user can use it at a time.
 - A single-user database that runs on a personal computer is called a **desktop database**.
 - **Multi-user database**: Supports multiple users at the same time.
 - Workgroup database: Supporting a relatively small number of users (usually fewer than 50), or within a specific department of an organization.
 - Enterprise database: Used by the entire organization and supports more than 50 users, across many departments.
- Location:
 - **Centralized Database**: Supports data located at a single site.
 - **Distributed Database**: Supports data distributed across several different sites.
- Type of data stored on them:
 - General-purpose databases: A wide variety of data used in multiple disciplines.
 - **Discipline-specific databases**: Data focused on specific subject areas.
- Most popular, classifying by how they will be used and the time sensitivity of the information gathered.
 - Operational database/Online transaction processing (OLTP)/Transitional/Production:
 Primarily to support a company's day-to-day operations.
 - Analytical Databases: Storing historical data and business metrics used exclusively for tactical or strategic decision making.
 - Data warehouse: Stores data in a format optimized for decision support.

- Online analytical processing (OLAP): Tools that provide an advanced data analysis environment for retrieving, processing, and modeling data from the data warehouse.
- Business intelligence: Process business data to generate information to support business decision making.

Structure of the data:

- **Unstructured data**: Data that exists in their original (raw) state the format they were collected. Does not contribute to the process that yields information.
- **Structured data**: Result of formatting unstructured data to facilitate storage, use, and the generation of information.
- Semi structured data: Have already been processed to some extent, most data are like this.
- Extensible Markup Language (XML): Special language used to represent and manipulate data elements in a textual format. Used for addressing unstructured and semi structured data.
- Social media: Web and mobile technologies that enable "anywhere, anytime, always on" human interaction.
- Large amounts of data are stored and analyzed for the new "social media".
- NOSQL (Not only SQL): A new generation of database management systems, not based on the traditional model. Usually for the widespread, growing data.

1.4 Why Database Design is Important

- **Database Design:** Activities that focus on the design of the database structure.
- Very crucial aspect of working with databases, having to be designed carefully in order to meet all user requirements.
- Since data is a very valuable asset to an organization, thorough planning to ensure that data are properly used to give the company the most benefit is just as important as proper financial planning to get the best use from its financial resources.
- Since current DBMS's are so easy to use, many computer-savvy business users gain a sense of false security, thus producing flawed, over-simplified structures and producing incomplete results when data is retrieved.
- **Proper database design** requires the designer to identify precisely the database's expected use.
- Integrated data must be decomposed properly to its constituent parts, each part stored in its own table, further, must be implemented so the information can be re-created for the end user later.
- Poorly designed database is likely to become a breeding ground for difficult to trace errors, causing a failure of an organization.
- Too important to be left to luck.

1.5 Evolution of File System Data Processing

 Evolved file system data processing through the understanding of its limitations, creating database designs that avoid the pitfalls of earlier systems. Designer has to understand these problems and how to avoid them.

1.5.1 Manual File Systems

- We used to file manually, through pencil and paper.
- Sorting things by files and folders based on its main use
- Bigger organizations meant more reporting requirements became more complex and tracking data more difficult, it meant using computer for help.

1.5.2 Computerized File Systems

- Data Processing (DP) Specialist is hired to create a computer based system that would track data and produce required reports.
- Manual file systems were slow and cumbersome, needing to hire a DP Specialist.
- Initially, computer files within the file system were similar to the normal files.
- The description of computer files requires a specialized vocabulary, different terminology for each discipline. This allowed for easy, clear communication between practitioners.
- DP Specialists created programs to retrieve, manipulate and present it to the user's request.
- As more and more computerized files were developed, the problems became apparent, having many data files that contained related (often overlapping) data with no means of meaning the data consistently across all of the tiles.
- Through this new way of storing data in computer files was a huge step forward in a business's ability to process data as well as improved tools for manipulating the data to create new information, it required a step between the end user and their data.
- The desire to close the gap between end users and the data led to a split between the ways DP specialists and end users viewed the data when trying to upgrade technologies and techniques.
 - DP Specialists thought of the computer files within the file system were created to be similar to the manual files.
 - End users thought the systems separated the users from the data, having to make more decisions in less time, they found it frustrating having to wait for the data and the DP Specialist creating the program to generate the information.

1.5.3 File System Redux: Modern End-User Productivity Tools

- The user's desire for direct, hands-on access to data helped fuel the adoption of personal computers for business use.
- Users have unfortunately become so adept at working with spreadsheets, they over and misuse it.
- Problems arose when the end users often took the limited data they had access to and placed it in a spreadsheet format, many users making separate copies of the data.

1.6 Problems with File System Data Processing

- Though file systems served very useful, many problems and limitations became evident.
 - Understanding the shortcomings of the file system enables you to understand the development of modern databases.
 - Many of the problems are not unique to file systems. Failure to understand such problems is likely to lead to their duplication in a database environment, even though database technology makes it easy to avoid them.
- Lengthy development times:
 - Even the simplest data-retrieval task requires extensive programming.
 - Old file systems required programmers to specify what to do and how it should be done, now it just requires the what.
- Difficulty of getting quick answers:
 - Needing to write programs to produce reports meant ad hoc queries impossible.
 - If you need the information now, it will serve you no purpose receiving the answers next week or next month.
- Complex system administration:
 - System administration becomes more difficult as the number of files in the system expands.
 - Simple file systems with few files require creating and maintaining several file management programs.
 - Each file must have its own file management programs to edit them.
 - Because ad hoc queries are not possible, the file reporting programs can multiply quickly.
 - Each department in the organization 'owns' its data by creating its own files.
- Lack of security and limited data sharing:
 - Sharing data among multiple geographically dispersed users introduces a lot of security risks.
 - Spreadsheet programs provide basic security options, which aren't always used, isn't sufficient for robust data sharing among users.
 - Even when an attempt is made to improve system and data security, the security devices tend to be limited in scope and effectiveness.
- Extensive programming:
 - Making changes to an existing file structure can be difficult in a file system environment.
 - Any change to a file structure, no matter how minor, forces modifications in all of the programs that use the data in that file.
 - Modifications are likely to produce errors.

1.6.1 Structural and Data Dependence

- A file system exhibits structural dependence, which means that access to a file is dependent on its structure.
 - One change would require all of the file system programs to be modified to conform to the next file structure.

- Conversely, structural independence exists when you can change the file structure without affecting the application's ability to access the data.
- Data dependence: When all data access programs are subject to change when any of the file's data storage characteristic change (that is, changing the data type).
- Conversely, data independence exists when you can change the data storage characteristics without affecting the program's ability to access the data.
- The practical significance of data dependency is the difference between:
 - Logical data format How the human being views the data.
 - **Physical data format** How the computer must work with the data.
 - Any program that accesses a file system's file must tell the computer not only what to do, but how to do it. This made data dependence files extremely cumbersome to the point of view of a programmer.

1.6.2 Data Redundancy

- File system's structure makes it difficult to combine data from multiple sources.
- Lack of security renders the file systems vulnerable to security breaches.
- Organizational structure promotes the storage of the same basic data in different locations.
 Islands of Information.
 - Can happen in a department that share access to a data file and making his or her own copy of the sales data. Usually outdated.
- You only need one correct copy of the file.
- **Data redundancy** exists when the same data are stored unnecessarily at different places.
- Uncontrolled data redundancy sets the stage for:
 - Poor data security:
 - Increases the changes for a copy of the data to be susceptible to unauthorized access.
 - Data Inconsistency:
 - Some files will be different versions, containing outdated or different data.
 - Reports will yield inconsistent results.
- Data that displays data inconsistency are also said to lack data integrity.
 - **Data integrity:** The condition in which all data in the database are consistent with realworld events and conditions.
 - **Data are accurate** There are no data inconsistencies.
 - Data are verifiable The data will always yield consistent results.
- Data anomalies
 - \circ $\;$ Ideally, only one field value is changed, but it usually isn't the case.
 - Could be faced with the prospect of making hundreds of corrections, one for each of the customers served by the agent.
 - Any change in any field value must be correctly made in many places to maintain data integrity.

• **Data Anomaly:** Develops when not all of the required changes in the redundant data are made successfully.

1.6.3 Lack of design and data-modeling skills

- Most users lack proper design and data modeling skills
- Forcing alto of different types of information into a single two-dimensional table structure, which is poor design.
- Data modeling is the most common method of documenting database designs.
- The data model also provides an invaluable resource when maintaining or modifying a database as business requirements change.

1.7 Database Systems

- Consists of logically related data stored in a single logical data repository.
 - Represents a major change in the way end-user data are stored, accessed and managed.
 - The database's DBMS provides numerous advantages over file system managements, eliminating most of the file system's data inconsistency, data anomaly, data dependence, and structural dependence problems.
 - o DBMS is just one of several crucial components of a database system.

1.7.1 The Database System Environment

- Database system refers to an organization of components that define and regular the collection, storage, management, and use of data within a dataset environment.
- Composed of five major parts:

o <mark>Hardware</mark>

- All physical devices, computers, storage devices, printers, network devices, etc.
- Software
 - Most readily identified software is the DBMS itself.
 - Three types of software are needed to make the database function fully:
 - Operating System Software Manages all hardware components
 - **DBMS software** Manages the database within the database system
 - Application programs and utility software Used to access and manipulate data in the DBMS and to manage the computer environment in which data access and manipulate that place.

o <mark>People</mark>

- All users of the database system
 - System administrators Oversee the database system's general operations
 - **Database administrator (DBAs)** Manage the DBMS and ensure that the database is functioning properly.
 - Database designers Design the database structure

- System analyst and programmers Design and implement the application programs
- End users People who use the application programs to run the organization's daily operations.

o Procedures

 The instructions and rules that govern the design and use of the database system

o <mark>Data</mark>

- Covers the collection of facts stored in the database
- Complexity of the managerial structure depends on the organization's size, function, etc.
 - Therefore, database systems can be created at different levels of complexity.
- Database solutions must be cost-effective as well as tactically and strategically effective.
- The database technology already in use is likely to affect the selection of a database system

1.7.2 DBMS Functions

- Performs several important functions that guarantee the integrity and consistency of the data in the database.
 - Most transparent to end users, and achieved only through DBMS.
- Data dictionary management The DBMS stores definitions of the data elements and their relationships (metadata) in a data dictionary.
 - The **data dictionary** is used by the DBMS to look up the required data component structures and relationships, relieving you from having to code complex relationships in each program.
 - Any changes done to the database structure is automatically recorded in the data dictionary (no dependence).
- Data storage management The DBMS creates and manages, the complex structures required for data storage, thus relieving you from the difficult task of defining and programming the physical data characteristics.
 - Data storage management is also important for performance tuning.
 - **Performance tuning** Activities that make the database more run more efficiently in terms of storage and access speed.
 - The DBMS can fulfill database requests concurrently because the DBMS stores the database in multiple physical data files, not having to wait for one disk request to finish before the nest one starts.
- Data transformation and presentation The DBMS transforms entered data to conform to required data structures.
 - Formats the physically retrieved data to make it conform to the user's logical expectations.
- Security Management The DBMS creates a security system that enforces user security and data privacy.

- Security rules determine which users can access the database and which data operations (edit, add, delete) users and perform.
- Multiuser access control To provide data integrity and data consistency, the DBMS uses sophisticated algorithms to ensure that multiple users can access the database concurrently without compromising its integrity.
- Backup and recovery management The DBMS provides backup and data recovery to ensure data safety and integrity.
 - DBMS systems allow the DBA to perform routine and special backup and restore procedures.
 - Critical to preserving the database's integrity.
- Data integrity management The DBMS promotes and enforces integrity rules, thus minimizing data redundancy and maximizing data consistency.
 - Data stored in the dictionary helps enforce data integrity.
 - Data integrity is especially important in transaction-oriented database systems.
- Database access languages and application programming interfaces The DBMS provides data access through a query language.
 - A **query language** is a nonprocedural language one that lets the user specify what must be done without having to specify how.
 - **Structured Query Language (SQL)** is the de facto query language and data access standard supported by the majority of DBMS vendors.
- Database communication interfaces A current generation DBMS accepts end-user requests via multiple, different network environments.

1.7.3 Managing the Database System = A Shift in Focus

- Role of the human component changes from an emphasis on programming (in the file system) to a focus on the broader aspects of managing the organization's data resources and complex database software itself.
- Although the database system yields considerable advantages over previous data management approaches, database systems do carry significant disadvantages.
 - Increased costs Database systems require sophisticated hardware and software and highly skilled personnel.
 - **Management complexity** Database systems interface with many different technologies have significant impact on a company's resource and culture.
 - **Maintaining currency** To maximize the efficiency of the database system, you must keep your system current.
 - **Vendor dependence** Heavy investment in technology and personal training, companies might be reluctant to change database vendors.
 - **Frequent upgrade/Replacement cycles** DBMS vendors frequently upgrade their products by adding new functionality.

1.8 Preparing for your Database Professional Career

- Databases are constantly evolving, to address new challenges such as large databases, semi structured and unstructured data, increasing processing speed, and lowering costs.
- At the end of the book, you'll hopefully be ready to apply your skills to work with the traditional
 OLTP and OLAP as well as cutting-edge, complex database technologies such as:
 - Very large databases (VLDB) Many vendors are addressing the need for databases that support large amounts of data, usually in the petabyte range.
 - Big data databases Products such as Cassandra (Facebook) and BigTable(Google) are using "columnar-database" technologies to support the needs of database applications that manage large amounts of "non-tabular" data.
 - In-memory databases Most major database vendors also offer some type of inmemory database support to address the need for faster database processing.
 - Cloud databases Companies can now use cloud database services to quickly add database systems to their environment while simultaneously lowering the total cost of ownership of a new DBMS.

Summary

- Data are raw facts. Information is the result of processing data to reveal their meaning.
 Accurate, relevant, and timely information is key to good decision making, and good decision making is the key to organizational survival in a global environment.
- Data are usually stored in a database. To implement a database and to manage its contents, you need a database management system (DBMS). The DBMS serves as the intermediary between the user and the database. The database contains the data you have collected "data about data", known as metadata.
- Database design defines the database structure. A well designed database facilitates data management and generates accurate and valuable information. A poorly designed database can lead to decision making, and bad decision making can lead to the failure of an organization.
- Database can be classified according to the number of users supported, where the data are located, the type of data stored, the intended data usage, and the degree to which the data are structured.
- Database evolved from manual and then computerized file system, data are stored in independent files, each requiring its own data management programs. Although this method of data management is largely outmoded, understanding its characteristics makes database design easier to comprehend.
- Some limitations of file system data management are that is requires extensive programming, system administration can be complex and difficult, making changes to existing structures is difficult, and security features are likely to be inadequate. Also, independent files tend to contain redundant data, leading to problems of structural and data dependence.
- Database management systems were developed to address the file system's inherent weakness.
 Rather than depositing data in independent files, a DBMS presents the database to the end user as a single data repository. This arrangement promotes data sharing, thus eliminating the

potential problem of islands of information. In addition, the DBMS enforces data integrity, eliminates redundancy, and promotes data security.