Database Basics

Hui Chen [a]

[a]CUNY Brooklyn College, Brooklyn, NY, USA

February 1, 2022

Outline

Table of Contents

Lesson Objectives

- Define the difference between data and information
- Describe what a database is, various types, and why they are valuable assets for decision making
- Explain the importance of database design
- See how modern data bases evolved from file systems
- Understand flaws in file system data management
- Outline the main components of the database system
- Describe the main functions of a database management system (DBMS)

Table of Contents

Data

Characteristics of data

- Ubiquitous (i.e., abundant, global, and everywhere)
- Pervasive (i.e., inescapable, prevalent, and persistent)

Data vs. Information

Data consists of raw facts

- Not yet processed to reveal meaning to the end user
- Building blocks of information

Information results from processing raw data to reveal meaning

- Requires context
- Bed rock of knowledge
- Should be accurate, relevant, and timely

Table of Contents

Concept of Database

Database: Shared, integrated computer structure that stores data

- Databases make data persistent and shareable in a secure way
- Specialized structures that allow computer-based systems to store, manage, and retrieve data very quickly
  - End-user data: raw facts of interest to end user
  - Metadata: data about data, through which the end-user data is integrated and managed
    - Describes data characteristics and relationships

Database management system (DBMS)

- Collection of programs
- Manages the database structure
- Controls access to data stored in the database

Role and Advantages of DBMS

Database management system (DBMS): intermediary between the user and the database

- Enables data to be shared
- Presents the end user with an integrated view of data
- Provides more efficient and effective data management
- Improves sharing, security, integration, access, decision-making, productivity, etc.

Types of DBMS

Categorize along several dimensions

- By number of users
- By location
- By type of data
- By degree of structure

Types of DBMS

Single-user vs. multi-user

- Single-user database: supports one user at a time
  - Desktop database: single-user database on a personal computer
- Multi-user database: supports multiple users at the same time
  - Workgroup databases: supports a small number of users or a specific department
  - Enterprise database: supports many users across many departments

Types of DBMS

Classification by location

- Centralized database: data located at a single site
- Distributed database: data distributed across different sites
- Cloud database: created and maintained using cloud data services that provide defined performance measures for the database

Types of DBMS

Classification by data type

- General-purpose database: contains a wide variety of data used in multiple disciplines
- Discipline-specific database: contains data focused on specific subject areas
- Operational database: designed to support a company's day-to-day operations

Types of DBMS

Analytical database: stores 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 for retrieving, processing, and modeling data from the data warehouse
- Business intelligence: captures and processes business data to generate information that support decision making

Types of DBMS

Databases can be classified to reflect the degree to which the data is structured

- Unstructured data exists in its original (raw) state
- Structured data results from formatting, e.g., structure is applied based on type of processing to be performed
- Semistructured data: processed to some extent
- Extensible Markup Language (XML)
- Represents data elements in textual format

DBMS Examples

DBMS	Single-User	Workgroup	Enterprice	Centralized	Distributed	Operational	Analytical
MS Access	X	X		X	X		
SQL Lite	X	X		X	X		
MS SQL Server	X	X	X	X	X	X	X
IBM DB2	X	X	X	X	X	X	X
MySQL/MariaDB	X	X	X	X	X	X	X
PostgresQL	X	X	X	X	X	X	X
Oracle RDBMS	X	X	X	X	X	X	X

Importance of Database Design

Focuses on design of database structure that will be used to store and manage end-user data

- Well-designed database: facilitates data management and generates accurate and valuable information
- Poorly designed database: causes difficult-to-trace errors that may lead to poor decision making

Table of Contents

From File System Data Processing to DBMS

- Manual file systems: accomplished through a system of file folders and filing cabinets
- Computerized file systems: data processing (DP) specialist created a computer-based system to track data and produce required reports
- file system redux: modern end-user productivity tools, includes spreadsheet programs such as Microsoft Excel

Basic File Terminology

Term	DEFINITION
Data	Raw facts, such as a telephone number, a birth date, a customer name, and a year-to-date (YTD) sales value. Data has little meaning unless it has been organized in some logical manner.
Field	A character or group of characters (alphabetic or numeric) that has a specific meaning. A field is used to define and store data.
Record	A logically connected set of one or more fields that describes a person, pl'a ce, or thing. For example, the fields that constitute a record for a customer might consist of the customer's name, address, phone number, date of birth, credit limit, and unpaid balance.
File	A collection of related records. For example, a file might contain data about the students currently enrolled at Gigantic University.

Problems

Problems with file systems challenge the types of information that can be created from data as well as information accuracy

- Lengthy development times
- Difficulty of getting quick answers
- Complex system administration
- Lack of security and limited data sharing
- Extensive programming

Structural Dependence vs. Independence

Structural Dependence

- Access to a file is dependent on its own structure
- All file system programs are modified to conform to a new file structure

Structural Independence

- Structural independence
- File structure is changed without affecting the application's ability to access the data

Data Dependence vs. Independence

Data dependence

- Data access changes when data storage characteristics change

Data independence

- Data storage characteristics are changed without affecting the program's ability to access the data

Practical significance of data dependence is the difference between logical and physical format

Data Redundancy

Unnecessarily storing the same data at different places

- Islands of information (i.e., scattered data locations)
- Increases the probability of having different versions of the same data

Uncontrolled Data Redundancy

Possible results of uncontrolled data redundancy

- Poor data security
- Data inconsistency
- Data-entry errors
- Data integrity problems

Data Anomalies

When not all of the required changes in the redundant data are made successfully

- Update anomalies
- Insertion anomalies
- Deletion anomalies

Table of Contents

Database Systems

- Logically related data stored in a single logical data repository
  - Physically distributed among multiple storage facilities
  - DBMS eliminates most of file system's data inconsistency, data anomaly, data dependence, and structural dependence problems
- Current generation DBMS software
  - Stores data structures, relationships between structures, and access paths
  - Defines, stores, and manages all access paths and components

Database System Environment

Database system: organization of components that define and regulate the collection, storage, management, and use of data within a database environment

- Hardware
- Software
- People
- Procedures
- Data

Functions of DBMS

- Data dictionary management
  - Data dictionary: stores definitions of data elements and their relationships
- Data storage management
  - Performance tuning ensures efficient performance
- Data transformation and presentation
  - Data is formatted to conform to logical expectations
- Security management
  - Enforces user security and data privacy

Functions of DBMS

- Multi user access control
  - Sophisticated algorithms ensure that multiple users can access the database concurrently without compromising its integrity
- Backup and recovery management
  - Enables recovery of the database after a failure
- Data integrity management
  - Minimizes redundancy and maximizes consistency

Functions of DBMS

- Database access languages and application programming interfaces
  - Query language: lets the user specify what must be done without having to specify how
  - Structured Query Language (SQL): de facto query language and data access standard supported by the majority of DBMS vendors
- Database communication interfaces
- Accept end-user requests via multiple, different network environments

Disadvantages of Database Systems

- Increased costs
- Management complexity
- Maintaining currency
- Vendor dependence
- Frequent upgrade/replacement cycles

Table of Contents

Careers and Terminology

Job Title	Description	Example skills
Database Developer	Create and maintain database-based applications	Programming, database fundamentals, SQL
Database Designer	Design and maintain databases	Systems design, database design, SQL
Database Administrator	Manage and maintain DBMS and databases	Database fundamentals, SQL, vendor courses
Database Analyst	Develop databases for decision support reporting	SQL, query optimization, data warehouses
Database Architect	Design and implementation of database environments (conceptual, logical, and physical)	DBMS fundamentals, data modeling, SQL, hardware knowledge, etc.
Database Consultant	Help companies leverage database technologies to improve business processes and achieve specific goals	Database fundamentals, data modeling, database design, SQL, DBMS, hardware, vendor-specific technologies, etc.
Database Security Officer	Implement security policies for data administration	DBMS fundamentals, database administration, SQL; data security technologies, etc.
Cloud Computing Data Architect	Design and implement the infrastructure for next-generation cloud database systems	Internet technologies, cloud storage technologies. data security, performance tuning, large databases, etc.
Data Scientist	Analyze large amounts of varied data to generate insights, relationships, and predictable behaviors	Data analysis, statistics, advanced mathematics, SQL programming, data mining, machine learning, data visualization

Table of Contents

Summary

- Data consists of raw facts and is usually stored in a database
- Database design defines the database structure
  - Can be classified according to the number of users, location., as well as data usage and structure
- Databases evolved from manual and computerized file systems
  - There are some limitations of file system data management
  - DBMSs were developed to address the file system's inherent weaknesses