Database Notes :: Chapter 1 :: Powerpoint

Main ideas:

- Data vs information
- What is a database / different types / Why they're useful
- Importance of database design
- How modern database evolved from file systems
- Flaws in file system data management
- Main components of the database system
- Database management system (DBMS) main functions

Why Databases:

- Characteristics of data in today's world:
    - **Ubiquitous** → abundant / global / everywhere.
    - **Pervasive** → unescapable, prevalent, persistent.
- 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.**

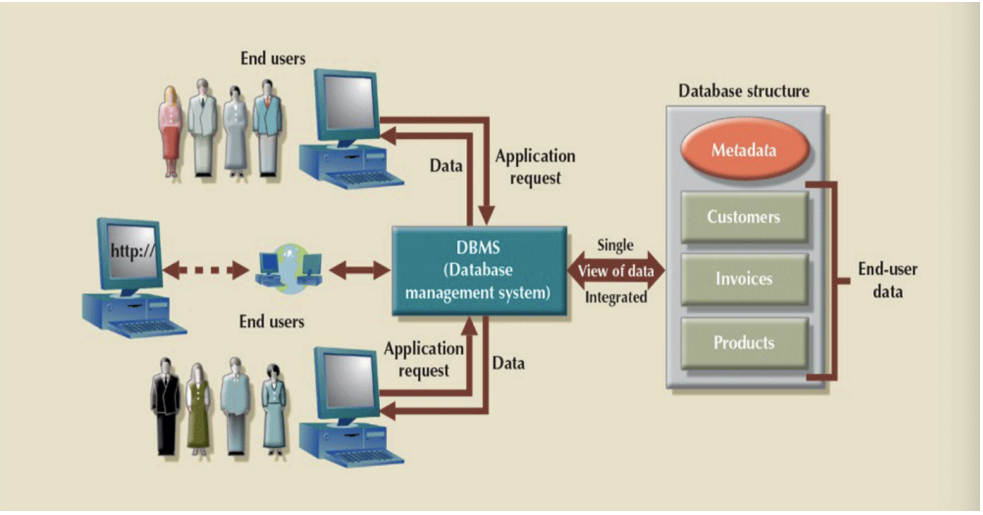
Data versus Information:

- **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**
    - Bedrock of knowledge
    - Should be **accurate, relevant, and timely.**
    - Key to good decision making → key to business survival in a global market.
- Example of transforming raw data into information:
    - Data entry screen → Raw Data → information in summary format → info graph.

Database Intro:

- **Database: Shared, integrated computer structure that stores data.**
    - **End-user data:** raw facts of interest to the end user.
    - **Metadata:** data about data, through which the end-user is integrated and managed. Describes data characteristics and relationships.
- **Database Management System (DBMS):** intermediary between the user and the database.
    - Collection of programs
    - Manages the database structure
    - Controls access to the data stored in the database.
    - Enables data to be shared
    - Presents the end user with an integrated view of the data.

- Provides more efficient and effective data management.
- Improves sharing, security, integration, access, decision-making, productivity, ect.
- **Manages the interaction between the end user and the database.**



Types of Databases:

- Classification by user size:
  - **Single-user database:** supports **one** user at a time.
    - **Desktop database:** single-user database on a personal computer.
  - **Multiuser database:** supports multiple users at the same time.
    - **Workgroup database:** supports a small number of users or a specific department.
    - **Enterprise database:** supports many users across many departments.
- 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.
    Examples: **Microsoft Azure, Amazon AWS**: Third party vendors that provide defined performance measures: **data storage capacity, required throughput,** and **availability**.
    Do not always specify underlying infrastructure to implement it
- Classification by data type:
  - **General-purpose database:** contains a wide variety of data used in multiple disciplines. ex: census database or ProQuest database (which contains media articles for a variety of topics).
  - **Discipline-specific:** contains data focused on specific subject areas.
    Used mainly for academic or research purposes within a small set of disciplines.
  - **Operational database:** Designed to support a company's day to day operations.

    - **Analytical database:** stores historical data and business metrics used exclusively for tactical or strategic decision making.
        - This analysis requires **data massaging** aka **data manipulation** to produce information on which to base decisions on.
        - Analytical database comprised of two main components:
            - **Data warehouse:** contains historical data obtained from the operational databases as well as data from other external sources.
            - **Online Analytical Processing Front End (OLAP):** Set of tools that work together to provide an advanced **data analysis environment** for **retrievin**g, **processing**, and **modeling** data from the data warehouse. Evolved into its own discipline: **Business Intelligence**.
            - **Business intelligence:** captures and **processes** business data to generate information that support decision making.
- Databases can be classified by the degree to which data is structured.
    - **Unstructured Data**: data exists in its original (**raw**) state.
    - **Structured Data:** results from **formatting** unstructured data to **facilitate storage, use** and **generation of information**. Structure is applied based on the type of processing to be performed on data.
    - **Semistructured data:** Processed to some extent. Most data user encounters are best classified as semistructured.
    - Unstructured and semistructured data storage and management needs are addressed by databases known as **XML** databases.
    - **Extensible Markup Language (XML):** Represents and manipulates data elements in textual format.
    - **Some data may be considered unstructured for some types of processing but structured for other types of processing.**

<u>Why Database Design is Important:</u>

- **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**.

<u>Evolution of File System Data Processing:</u>

- Types of File Systems:
    - **Manual File Systems:** 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. ex: spreadsheet programs like Microsoft Excel.
- Basic File Terminology:

    - **Data:** raw facts. ex: phone number, birth data, name, ect. Data has little meaning unless it has been organized in some logical manner.
    - **Field:** character or group of characters (alphanumeric) 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, place or thing.
    - **File:** Collection of related records. ex: may contain data about students enrolled at a particular school.

Problems with File System Data Processing:

- **Lengthy Development Times:** Even the simplest data retrieving task requires extensive programming. With older file systems, a specification must be done and how to do it. Modern databases use a **non-procedural data manipulation language** that allows the user to specify what must be done without specifying how.
- **Difficulty of getting answers:** The need to write programs to reduce even the simplest reports makes ad hoc (when necessary) queries impossible.
- **Complex System Administration:** System administration becomes more difficult as the number of files in the system expands. Regardless of file number, required to create and maintain several file management programs. Each file must have its own file management program for users to add, modify and delete records, to list the file content and generate reports.
- **Lack of security and Limited Data Sharing:** Sharing data among geographically dispersed users introduce a lot of security risks. Spreadsheet programs have insufficient security guarantees for robust data sharing among users.
- **Extensive Programming:** Changes to existing file structure in file system env requires programs that reads, transforms, writes and repeats preceding steps in the original file.
- **Structural Dependence and Redundancy**

Structural and Data Dependence:

- **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:** Exists when file structure is changed without affecting the application's ability to access the data.
- Practical Significance of Data Dependence:
    - **Logical Data Format:** How the human being views the data.
    - **Physical Data Format:** How the computer must work with the data.
    - **Practical significance of data dependence** is the difference between the logical data format and the physical data format. Any program that accesses a file system format file must tell the computer not only what to do but also how to do it.
      i.e. Each program must contain lines that specify the opening file type, its record specifications, and its field definitions.

Data dependence makes the file system extremely cumbersome from the point of view of a programmer and database manager.

Data Redundancy:

- File system structure makes it difficult to combine data from multiple sources.
- The organizational structure promotes the storage of the **same basic data** in different locations.
- 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**
    - **Data redundancy:** different versions of the same data.
- Possible results of uncontrolled data redundancy:
    - **Poor Data Security:** Multiple copies of data increases the chances for a copy of data to be susceptible to unauthorized access.
    - **Data Inconsistency:** Exists when different and conflicting versions of the same data appear in different places. i.e when trying to update information, you update in one file and not the other, leading to different versions of the same data, some of which may not be correct. Reports will **yield inconsistent results** depending on what version of the data is used.
    - **Data-entry Errors:** accidental eros when entering data are more likely to occur when made in several different files.
    - **Data Integrity Problems:** May be able to enter data (i.e. phone number and name) for an object (i.e. a sales agent) that does not exist.
    - Data entry error yields the same kind of data integrity problems.

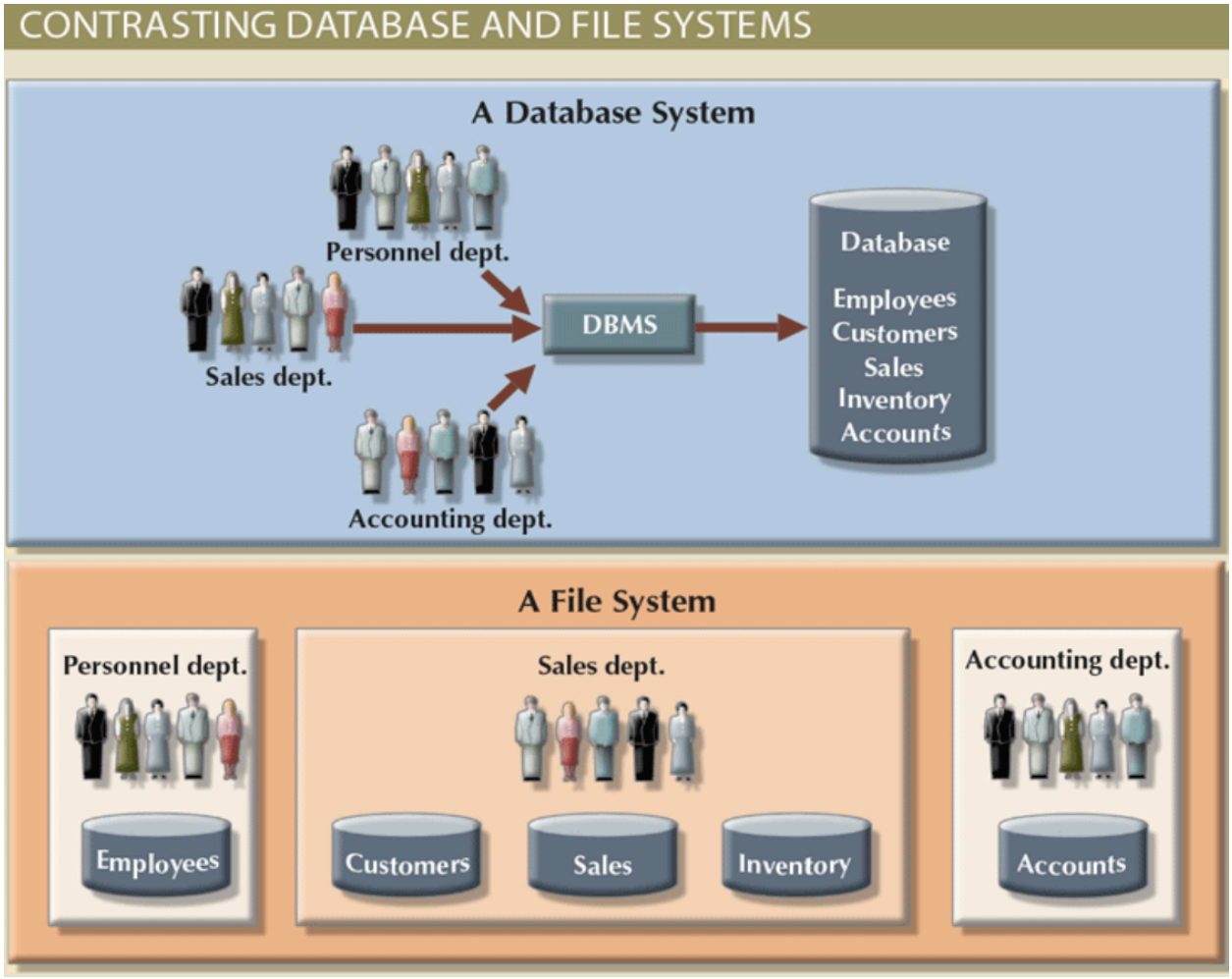
Data Anomalies:

- Develop when not all of the required changes in the redundant data are made successfully.
- Types of data anomalies:
    - **Update Anomalies**
    - **Insertion Anomalies**
    - **Deletion Anomalies**

Database Systems:

- Logically related data stored in a single logical data repository.
    - Physically distributed among multiple storage facilities.
    - **Database Management System (DBMS)** eliminates most of the file system's data inconsistency, data anomaly, data dependence, and structural dependence problems.
- Current generation DBMS software:
    - Stores **data structures**, the **relationships between structures**, and the **access paths to those structures** all in a **central location.**

- Takes care of **defining**, **storing**, and **managing** all required access paths to those components.
- DBMS is **just one of several crucial components** of a database system.
- DBMS is like the database systems heart.



Database System Environment:

- **Database System:** Organization of components that **define** and **regulate** the **collection**, **storage**, **management**, and **use** of **data** within a database environment.
- Database system composed of **five** major parts: **Hardware**, **Software**, **People**, **Procedures**, and **Data**.
    - **Software: Three** types of software needed to make the database system function fully:
        - **Operating System Software:** manages all hardware components and makes it possible for all other software to run on the computers.
        - **DBMS Software:** manages the database within the database system. ex: Microsoft SQL server, Oracle Corporation's Oracle, Oracles' MySQL and IBM's DB2.
        - **Application Programs and Utilities Software:**

            - **Application Programs:** are used to access and manipulate data in the DBMS and manage the computer's env in which data access and manipulation takes place.
            - **Utilities:** the software tools used to help manage the database systems computer components.
    - **People:** All users of the database system. **Five types of users:**
        - **System Administrators, Database administrators, database designers, system analyst and programmer,** end **end users.**
    - **Procedures:** Instructions and rules that govern the design and use of database systems.
    - **Data:** Collection of facts stored in the database.

DBMS Functions:

- DBMS performs several functions that guarantee **integrity** and **consistency** of the database.
- **Data Dictionary Management:** Stores definitions of data elements and their relationships.
- **Data Storage Management** Performance tuning to ensure efficient performance.
- **Data Transformation and Presentation:** Data formatted to conform to logical expectations.
- **Security Management:** Enforces user security and data privacy.
- **Multiuser 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.
- Database access languages and application programming interfaces:
    - **Query Language:** lets the user specify what must be done without specifying how.
    - **Structured Query Language (SQL):** de facto query language and data access standard supported by the majority of DBMS vendors.
- **Database Communication Interfaces:** Accepts end-user requests via multiple, different network environments.

Disadvantages of Database Systems:

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

Careers in Database Use:

TABLE 1.3	DATABASE CAREER OPPORTUNITIES	
JOB TITLE	DESCRIPTION	SAMPLE SKILLS REQUIRED
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	QL, 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

Summary:

- Data consists of raw facts 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 limitations of file system data management
    - DBMS' were developed to address the file system's inherent weaknesses.

Chapter 2 :: Powerpoint

Learning Objectives:

-