❖ 1-1Why Databases?

❖ 1-2Data Versus Information

Data – Raw facts, or facts that have not yet been processed to reveal their meaning to the end user.

Information – The result of processing raw data to reveal its meaning. Information consists of transformed data and facilitates decision making.

Knowledge – The body of Info. And facts about a specific subject. Knowledge implies familiarity, awareness, and understanding of info. As it applies to an environment. A key characteristic is that new knowledge can be derived from old knowledge.

- ✓ Data constitutes the building blocks of information.
- ✓ Information is produced by processing data.
- ✓ Information is used to reveal the meaning of data.
- ✓ Accurate, relevant, and timely information is the key to good decision making.
- ✓ Good decision making is the key to organizational survival in a global environment.

Data Management – Process that focuses on data collection, storage, and retrieval. Common data management functions include: addition, deletion, modification, and listing.

❖ 1-3Introducing The Database

Database – A shared, integrated computer structure that houses a collection of related data. A database contains 2 types of data. End-user data(raw facts) and metadata.

Metadata – Data about data: data about data characteristics and relationships. See also data dictionary.

Database management system(DBMS) – Collection of programs that manage the database structure and controls access to the data stored in the database.

❖ 1-3aRole and Advantages of The DBMS

- ✓ Improved data sharing.
- ✓ Improved data security.
- ✓ Better data integration.
- ✓ Minimized **Data inconsistency** – (exists when the same data is in different places)
- ✓ Improved data access.

Query- A Question. Specific request issued to the DBMS for data manipulation.

Ad hoc query – "Spur-of-the-Moment" Question

Query result set – An Answer. Sent from DBMS to the application.

- ✓ Improved Decision making. **Data quality** – comprehensive approach to promoting accuracy, validity, and timeliness of the data.
- ✓ Increased End-User productivity.

❖ 1-3bTypes of Databases

- ✓ databases can be classified by the number of users supported, where the data is located, the type of data stored, the intended data usage, and the degree to which the data is structured.

Classified by # of Users.
Single-user database – Only supports 1 user at a time. Single-user database that runs a person computer is a **Desktop database**.
Multiuser database – Supports multiple users.
-Workgroup database- Multiuser database that supports 50> users, or a specific department within an organization.
-Enterprise database - Multiuser database that supports 50+(usually hundred) users across many departments.
Classified by Location.
Centralized database – Supports data at a single site.
Distributed database - Supports data at several sites.
Cloud database – created and maintained using cloud services, such as Microsoft Azure or Amazon AWS, for storage only.
Classified by Type of Info.
General-purpose database – Wide variety of data.
Discipline-specific database – Data of specific subjects.
Classified by Use/Time Sensitivity.(Day-to-day)
-Operational database – Supports Company's day-to-day operations, Primarily. Also known as **Online transaction processing (OLTP) database/Transactional database/Production database**.
-Analytical database – Primarily storing historical data for strategic decisions.
--Data warehouse – Specialized. Stores data in a format for optimized decision making.
Online analytical processing (OLAP) – set of tools that work together to provide an advanced data analysis environment for retrieving, processing, and modeling data from the data warehouse.
Business intelligence – Specific Discipline. comprehensive approach to capture and process business data with the purpose of generating information to support business decision making.
Classified by Degree of data structure.
Unstructured data - data that exists in its original (raw) state.the format in which it was collected.
Structured data – result of formatting unstructured data to facilitate storage, use, and the generation of information.
Semistructured data – processed to some extent.

Extensible Markup Language (XML) – special language used to represent and manipulate data elements in a textual format.
XML database – supports the storage and management of semistructured XML data.

Social media – web and mobile technologies that enable "anywhere, anytime, always on" human interactions
NoSQL – new generation of database designed to handle the unprecedented volume of data, variety of data types and structures, and velocity of data operations that are characteristic of these new business requirements.

❖ **1-4**Why Database Design is Important
Database design – activities that focus on the design of the database structure that will be used to store and manage end-user data.

❖ **1-5**Evolution of file system data processing
❖ **1-5a**Manual File Systems (Paper Filing Cabinets)
❖ **1-5b**Computerized File Systems
Data processing (DP) specialist - hired to create a computer-based system that would track data and produce required reports.
Data- Raw Facts (Ex: Telephone # or Name)
Field-Used to Define/Store Data. A character or group of characters (alphabetic or numeric) that has a specific meaning.
Record- A logically connected set of one or more fields that describes a person, place, or thing.(Ex. Customer name, address, phone number, date of birth, credit limit, and unpaid balance.)
File- A collection of related records.
❖ **1-5c**File System Redux: Modern End-User Productivity Tools

---

❖ **1-6** Problems With File System Data Processing
  ✓ Lengthy development times. Difficulty getting quick answers. Complex system administration. Lack of security and data sharing. Extensive Programming.
❖ **1-6a**Structural and Data Dependence
Structural dependence – access to a file is dependent on its structure.
Structural independence – you can change the file structure without affecting the application's ability to access the data.
Data dependence – all data access programs are subject to change when any of the file's data storage characteristics change
Data independence - exists when you can change the data storage characteristics without affecting the program's ability to access the data.
Logical data format – how human beings views the data
Physical data format – how the computer must work with the data.
❖ **1-6b**Data Redundancy
Islands of information – storage of the same basic data in different locations.
Data redundancy – Same data is stored unnecessarily at different places.

- ✓ Poor data security. Data inconsistency. Data-Entry Errors. Data integrity problems.

Data integrity – Data that displays data inconsistency. Accurate(No Inconsistencies) and Verifiable(always yields same result)

❖ **1-6c**Data Anomalies

Data anomaly - not all of the required changes in the redundant data are made successfully.

- ✓ Caused by Updating info., Inserting new info., Deleting old info.

❖ **1-7**Database Systems

❖ **1-7a**The Database System Environment

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

- ✓ 5 Components:
- ✓ hardware(Computers),
- ✓ software(Operating System Software[windows], DBMS Software[Oracle's MySQL], Application Programs and Utility Software[To access for data manipulation & help manage database system's computer components]),
- ✓ people (system administrators, database administrators, database designers, system analysts and programmers[&End Users]),
- ✓ procedures(Instructions)
- ✓ data(what data is important/how to organize it)

❖ **1-7b**DBMS Functions

- ✓ Functions include: data dictionary management, data storage management, data transformation and presentation, security management, multiuser access control, backup and recovery management, data integrity management, database access languages and application programming interfaces, and database communication interfaces
- ✓ -data dictionary management- stores definitions of the data elements and their relationships (metadata) in a **data dictionary** – Used to look up structures/relationships.
- ✓ -data storage management- storage not only for the data but for related data-entry forms or screen definitions, report definitions, data validation rules, procedural code, structures to handle video and picture formats, and so on. Important for **Performance tuning** – activities that make the database perform more efficiently in terms of storage and access speed
- ✓ -data transformation and presentation- transforms entered data to conform to required data structures, also to the user's logical expectations.
- ✓ security management- security system that enforces user security and data privacy.
- ✓ multiuser access control - sophisticated algorithms to ensure that multiple users can access the database concurrently without compromising its integrity.
- ✓ backup and recovery management- ensure data safety and integrity.
- ✓ data integrity management- promotes and enforces integrity rules.

- ✓ database access languages and application programming interfaces- DBMS provides data access through a **query language** - nonprocedural language—one that lets the user specify what must be done without having to specify how. **Structured query language (SQL)** – The Standard for query language and data access. also provides application programming interfaces to procedural languages, like Java.
- ✓ database communication interfaces- DBMS accepts end-user requests via multiple, different network environments, like Google Chrome or Firefox, and can automatically publish predefined reports on a website, or through Email.

❖ **1-7c**Managing The Database System: A Shift in Focus

- ✓ Disadvantages: Increased Costs. Management complexities(with User Accesses/Security), Maintaining Currency(Updates/Patches/Security), Vendor Dependence(Limited/Pricy), Frequent upgrade/replacement cycles(New features in new upgrade versions of the software/Reqs new Hardware/ new training for new features)

❖ **1-8**Preparing For Your Database Professional Career

-*Very Large Databases (VLDB)*- support large amounts of data, usually in the petabyte range. 1000TB

- *Big Data databases*- "columnar-database" technologies to support the needs of database applications that manage large amounts of "nontabular" data.

- *In-memory database*- store most of their data in primary memory (RAM) rather than in slower secondary storage (hard disks).

- *Cloud databases*- same as local DBMS, but accessible through the Internet.

❖ SUMMARY

- Data consists of raw facts. Information is the result of processing data to reveal its meaning. Accurate, relevant, and timely information is the key to good decision making, and good decision making is the key to organizational survival in a global environment.
- Data is 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 and "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 poor decision making, and poor decision making can lead to the failure of an organization.
- Databases can be classified according to the number of users supported, where the data is located, the type of data stored, the intended data usage, and the degree to which the data is structured.
- Databases evolved from manual and then computerized file systems. In a file system, data is 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 it 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 weaknesses. 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.

- Knowledge of database technologies leads to many career opportunities in the ever-expanding IT industry. There is a variety of specialization within the database arena for a wide range of skills and expertise.