CHAPTER ONE

Introduction to Database Systems

By Aster A.

Introduction

- A ***database*** is a collection of related data.
  - ***Data is a known facts that can be recorded and that have implicit meaning***. E.g. consider the ***names, telephone numbers, and addresses*** of the people you know.
  - You may have recorded this data in an ***indexed address book*** or ***stored it on a hard drive***, using a personal ***computer and software such as Microsoft Access or Excel***.
- This ***collection of related data*** with an implicit meaning is a database.

Introduction…

- ***A database has the following implicit properties:***
  - A database ***represents some aspect of the real world***, sometimes called the ***miniworld or the universe of discourse*** (UoD).
    - ✓***Changes to the miniworld*** are reflected in the database.
  - A ***database*** is ***logically coherent collection of data*** with some ***inherent meaning***.
    - ✓A ***random assortment of data cannot correctly be referred to as a database***.
- A database is ***designed, built, and populated*** with ***data for a specific purpose.***
  - It has an ***intended group of users and some preconceived applications in which these users are interested***.

Introduction…

- A database may be ***generated and maintained manually or it may be computerized***.
- ***For example***, a ***library card catalog is a database*** that may be ***created and maintained manually***.
- A ***computerized database*** may be created and maintained either by a ***group of application programs*** written specifically for that task or by a ***database management system.***

DBMS (Database Management System)

- A ***Database Management System (DBMS):***
  - is a ***software package*** designed to ***store and manage databases***.
  - is a ***collection of programs*** that enables users to ***create and maintain a database***.
  - is a ***general-purpose software system*** that facilitates the processes of ***defining, constructing, manipulating, and sharing*** databases among various users and applications.
- ***Example:*** Access, My-SQL, SQL, Oracle etc.

DBMS (Database Management System)…

- The ***DBMS*** also ***protect the database and maintaining*** it over a long period of time.
- Protection includes ***system protection against hardware or software malfunction*** (or crashes) and ***security protection against unauthorized or malicious access.***

DBMS (Database Management System)…

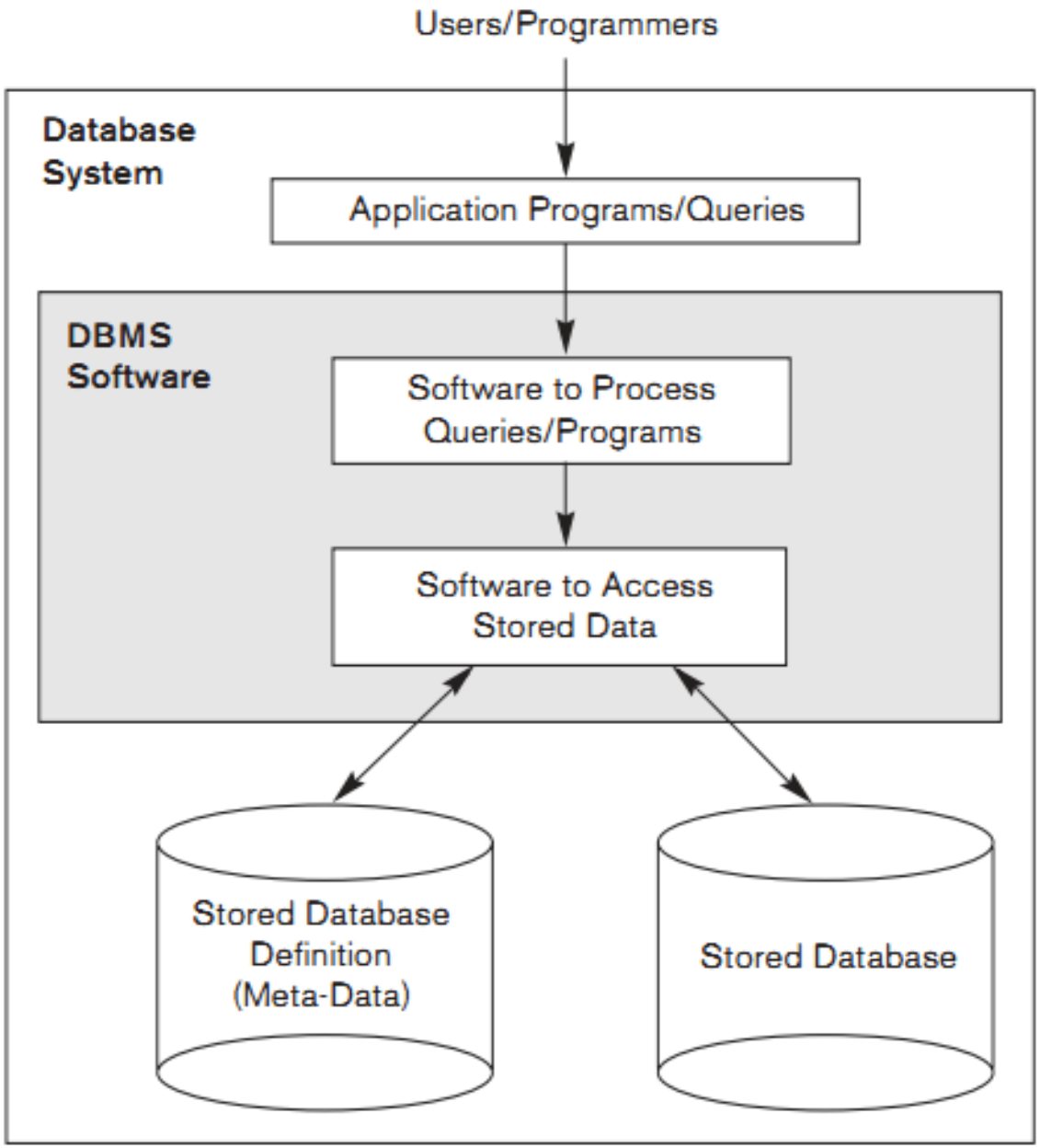
- A typical large ***database may have a life cycle of many years***, so the ***DBMS must be able to maintain the database system*** by allowing the system to evolve as ***requirements change over time.***
- ➢ It is not absolutely necessary to use ***general-purpose DBMS software*** to implement a computerized database.
- ➢ We could write our own set of programs to ***create and maintain the database***, in effect creating our own ***special-purpose DBMS software***
-

DBMS (Database Management System)…

- In either case whether we use a general-purpose DBMS or not—we usually have to deploy a considerable amount of complex software.
- In fact, most DBMSs are very complex software systems.

➢ The ***database and DBMS software*** together a **Database System**.

- The ***DB Application*** is simply a ***program that interacts with the DB at some point in its execution.***



- ***A simplified Database System Environment.***

Data Management Approaches

- ***Data Management (keeping your data records)can be done***:
  - ***Manual Approach***
  - ***File-Based Approach***
  - ***Database Approach***

- ***Currently, all methods of data handling are in use to some extent.***

1. Manual File Handling Systems

- The ***primitive and traditional*** way of information handling where ***cards and papers*** are used for the purpose. This approach:
  - requires ***intensive human labour***
  - ***Events and objects are written on files (paper)***
  - ***Files are labeled and stored in one or more cabinets.***
  - The cabinets could be ***kept in safe places for security purpose.***
  - ***Insertion and retrieval*** is done by ***searching*** first for the ***right cabinet*** then for the ***right file, then the information.***
  - One could have ***indexing system*** to facilitate access to the data.

➢ Example , Personal Calendar

- We might start by building a file with the following structure:

What	Day	Time	with_whom	Where
Lunch	10/24	1pm	Abebe	Bole
Shop	10/26	9am	Ayele	Piassa
Dinner	10/26	6PM	Elfinesh	Café

- This calendar is easy to deal with.

- **Let us also build our address book**
  - Abebe A. 0911164743
  - Seble K. 0912150844
  - Ayele L. 0913170841
  - Elfinesh W. 0911169673
  - Ermias H. 0912151043
  - Worku A. 0911199904
  - Worku D. 0912620410
  - Abebe D. 0914556768

Limitations of Manual File Handling

A) Problem of Data Organization

- Suppose we want to cancel one of our appointments.
- ***Two conceptual "entities"*** **-- address and calendar --** with a relationship between them, linking people in the ***calendar to their contact information.***

B) Problem of Efficiency

- ***Size of your personal address book is probably less than one hundred*** entries, but ***there are things we'd like to do quickly and efficiently.***
  - "Give me all appointments on 10/28"
  - "When am I next meeting Dr. Dawit?"
- ***What would happen if you were using a business calendar with hundreds of thousands of entries?***

Limitations of Manual File Handling …

C) Prone to error

D) Difficult to update, retrieve, integrate

E) You have the data but it is difficult to compile the information

F) Significant amount of duplication of data

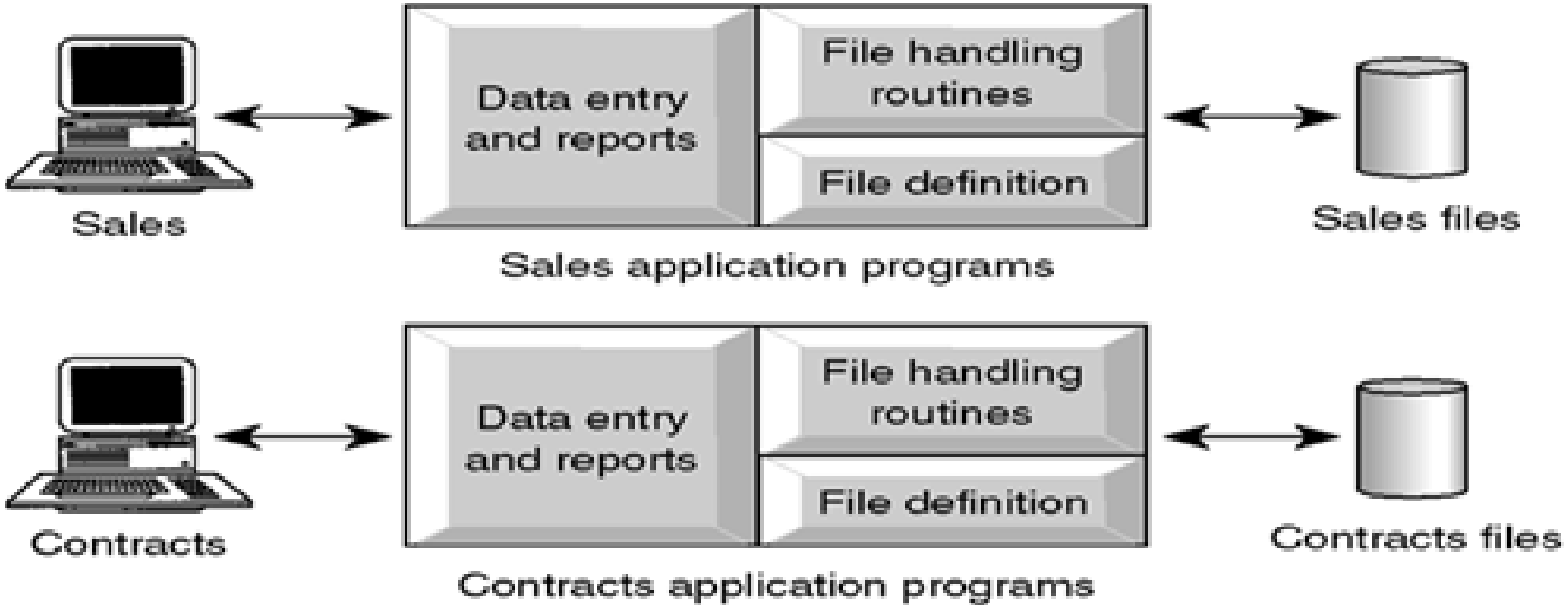
G) Cross referencing is difficult

- ***What is the solution*** to ***overcome the problems or difficulties inherited in the Manual File Handling Approaches:***

- Two ***computerized approaches*** evolved to overcome this limitations
  - **File Based Approach → Decentralised**
  - **Database Approach→ Centralised**

2. File-Based Approach

- File ***based systems were an early attempt*** to ***computerize the manual filing system***.
  - This approach is a ***decentralized computerized data handling method*** - to ***develop a program or a number of programs for each different application***.
  - Since every ***application defines and manages its own data***, the system is ***subjected to serious data duplication problem***



Sales Files

Property_for_Rent(Property Number, Street, Area, City, Post Code, Property Type, Number of Rooms, Monthly Rent, Owner Number)

Owner(Owner Number, First Name, Last Name, Address, Telephone Number)

Renter(Renter Number, First Name, Last Name, Address, Telephone Number, Preferred Type, Maximum Rent)

Contracts Files

Lease(Lease Number, Property Number, Renter Number, Monthly Rent, Payment Method, Deposit, Paid, Rent Start Date, Rent Finish Date, Duration)

Property_for_Rent(Property Number, Street, Area, City, Post Code, Monthly Rent)

Renter(Renter Number, First Name, Last Name, Address, Telephone Number)

Limitations of File-Based systems

- **Data Redundancy (Duplication of data)**
  - *Same data is held by different programs*
  - *Wasted space (Uncontrolled duplication of data)*
- **Separation and isolation of data**
  - ***Each program maintains its own set of data.***
  - ***Users of one program may be unaware of potentially useful data held by other programs.***
  - ***Limited data sharing***
  - ***Special codes for different queries***

Limitations of File-Based systems…

- **Data Inconsistency and confusion**

  Eg. ***Consider an organization***

  - ***Personnel Department stores details relating to each member.***
  - ***Payroll Department stores salaries of each staff***

- There are ***potentially different values and/or different formats for the same item.***
- ***Consider the duplication of data between the payroll and personnel departments.***
- If an ***employee moves house and the change of address is communicated only to personnel*** and ***not to payroll, the person's pay slip will be sent to the wrong address***.

Limitations of File-Based systems…

- **Data Dependence**
  - ***File structure is defined in the program code*** and is ***dependent on the application programming language***.
  - Each ***application program must have its own processing routines*** for ***reading, inserting, updating and deleting*** data.
- ***Incompatible file formats (Lack of Data Sharing and Availability)***
  - ***Programs are written in different languages, and so cannot easily access each others files***.
    - **Eg.** ***personnel writes in C***
      payroll writes in COBOL

Limitations of File-Based systems…

- Suppose ***payroll wants to know when a certain employee will be on pension***.
- It is ***difficult to share this information unless we have another application programmer*** to write ***software to convert the files to some common format.***

➢ **Poor Security and Administration**

- ***Unauthorized people may access the data***.

➢ **Anomalies**

1) Modification Anomalies:

- A ***problem experienced*** when one or more data value is ***modified on one application program*** but ***not one others containing the same data set.***

2) Deletion Anomalies:

- A ***problem encountered*** where ***one record set is deleted from one application*** but ***remain untouched in other application programs***.

3) Insertion Anomalies:

- A ***problem experienced*** whenever there is ***a new data item to be recorded,*** and the ***recording is not made in all the applications.***

3. Database Approach

- The limitation of the ***file based can be attributed to two factors:***
  - ***Definition of data is embedded in the application programs,*** rather than being ***stored separately and independently.***
  - There is ***no control over the access and manipulation of data*** beyond that imposed by the application programs.
- To become more effective, a ***new approach was required*** by the name ***Database Approach***.
  - What emerged were the ***database and database management systems.***
  - A ***single repository of data*** is maintained

3. Database Approach…

- A ***Database*** is a ***shared collection of logically related*** data ***designed to meet the information needs of an organization- (Centralized System).***
  - ***Logically related data comprises entities, attributes, and relationships*** of an ***organization's information***.
- ***System Catalog (Data Dictionary or Metadata)*** provides the ***description of the data to enable program–data independence.***
- The ***database contains not only*** the ***database itself*** but also a ***complete definition or description of the database***.

3. Database Approach…

- This definition is stored in the ***system catalog***, which contains information such as the ***structure of each file, type and storage format of each data item and various constraints on the data.***.
- The ***information stored is called metadata*** and it ***describes the structure of the primary database***.

DBMS Functions

- A **DBMS** performs several important functions that guarantee the ***integrity and consistency*** of the data in the database.
- Most of those functions are ***transparent to end users***, and most can be achieved only through the ***use of a DBMS***.

DBMS Functions…

- They 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***
- ***Database Communication Interfaces***

1. Data Dictionary Management

- The ***DBMS stores definitions of the data elements*** and ***their relationships (metadata) in a data dictionary***.
- In turn, ***all programs that access the data in the database*** work through the DBMS.
- The ***DBMS uses the data dictionary*** to look up the ***required data component structures*** and relationships.
- Additionally, ***any changes made in a database structure*** are ***automatically recorded in the data dictionary.***
- In other words, the ***DBMS provides data abstraction***, and it ***removes structural and data dependence from the system***.

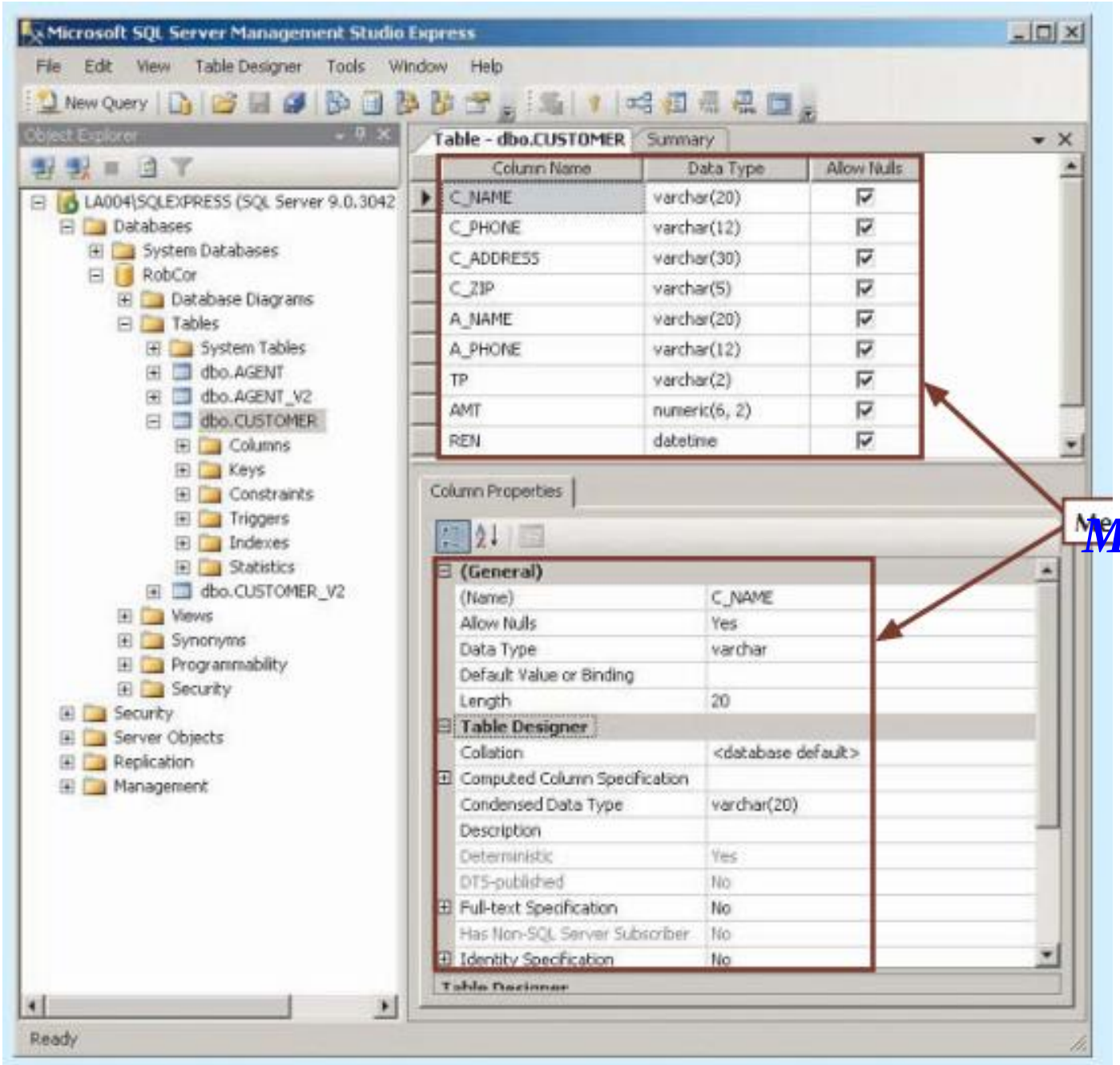


Fig: Illustrating metadata with Microsoft SQL Server Express

➢ A useful ***data dictionary system*** should ***store and manage the following types of information:***

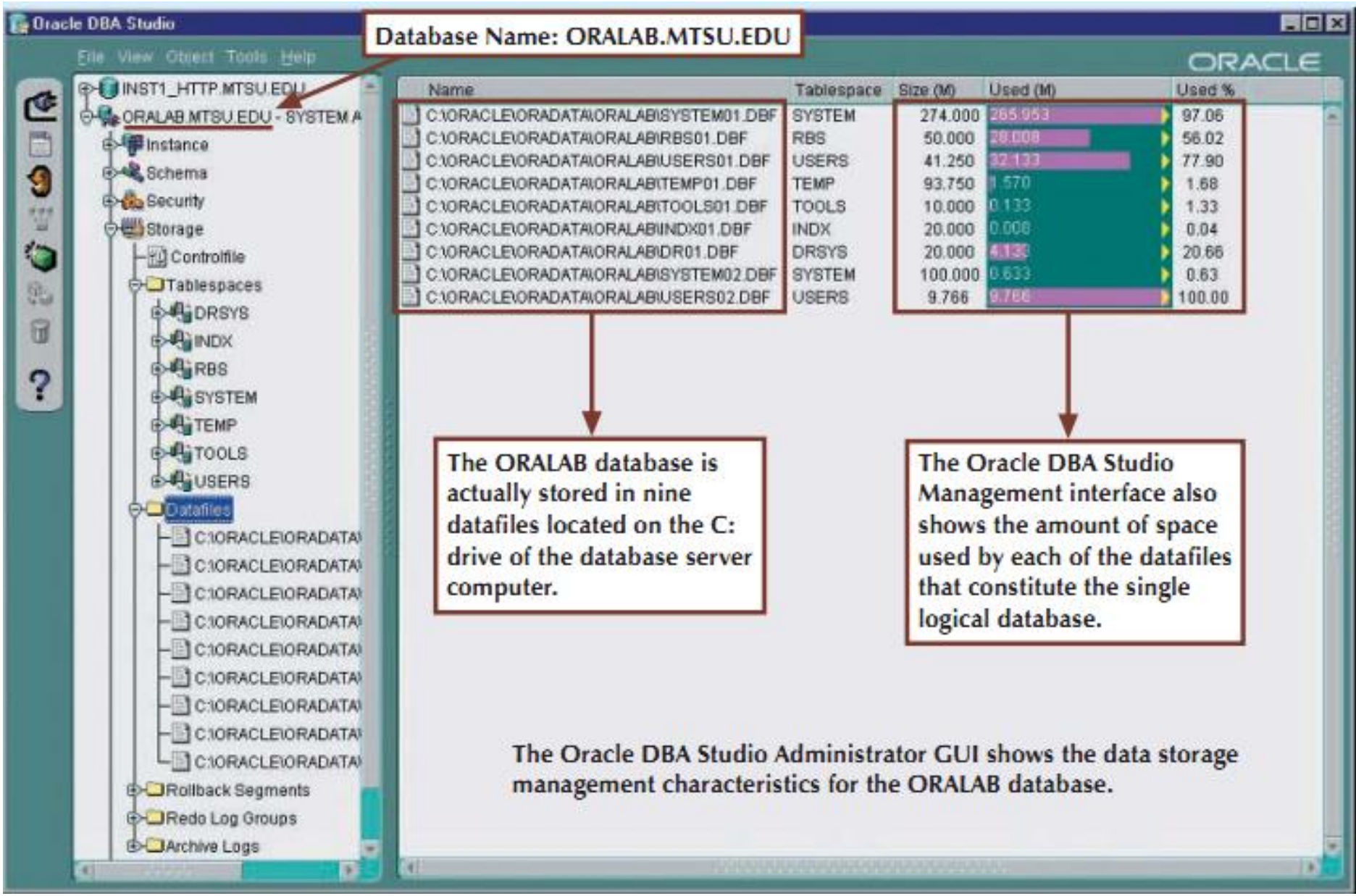
- ***Descriptions of the schemas of the database system***.
- Detailed information on ***physical database design***, such as ***storage structures, access paths, and file*** and ***record sizes***.
- Descriptions of the types of ***database users, their responsibilities, and their access rights***.
- High-level descriptions of the ***database transactions and applications*** and of the ***relationships of users to transactions.***
- The relationship between ***database transactions and the data items*** referenced by them.
- ***Usage statistics*** such as ***frequencies of queries and transactions*** and ***access counts to different portions of the database.***

2. Data Storage Management

- The DBMS ***creates and manages the complex structures required for data storage.***
- A modern DBMS provides ***storage not only for the data***, but also 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.
- ***Data storage management*** is also important for database performance tuning.

2. Data Storage Management…

- **Performance tuning** relates to the activities that make the ***database perform more efficiently in terms of storage and access speed.***
- Although the user sees the database as a single data storage unit, the ***DBMS actually stores the database in multiple physical data files.***
- See the following figure. Such data files may even be ***stored on different storage media.***
- Therefore, the ***DBMS doesn't have to wait for one disk request*** to ***finish before the next one starts.***
- In other words, the ***DBMS can fulfill database requests concurrently.***



Illustrating data storage management with Oracle

3. 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 the ***integrity of the database.***

4. Backup and Recovery Management

- The DBMS provides ***backup and data recovery*** to ensure ***data safety and integrity***.
- Current DBMS systems provide special utilities that allow the DBA to ***perform routine and special backup and restore procedures.***
- ***Recovery Management*** deals with the ***recovery of the database after a failure***, such as a ***bad sector in the disk or a power failure***.
- Such ***capability is critical to preserving the database's integrity***.

5. Data Integrity Management

- The DBMS promotes and enforces integrity rules, thus ***minimizing data redundancy and maximizing data consistency***.
- The ***data relationships stored*** in the ***data dictionary*** are used to ***enforce data integrity***.
- Ensuring data integrity is especially important in ***transaction-oriented database systems***.

6. 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 it is to be done.***
- Structured Query Language (SQL) is the de facto query language and data access standard supported by the majority of DBMS vendors.
-

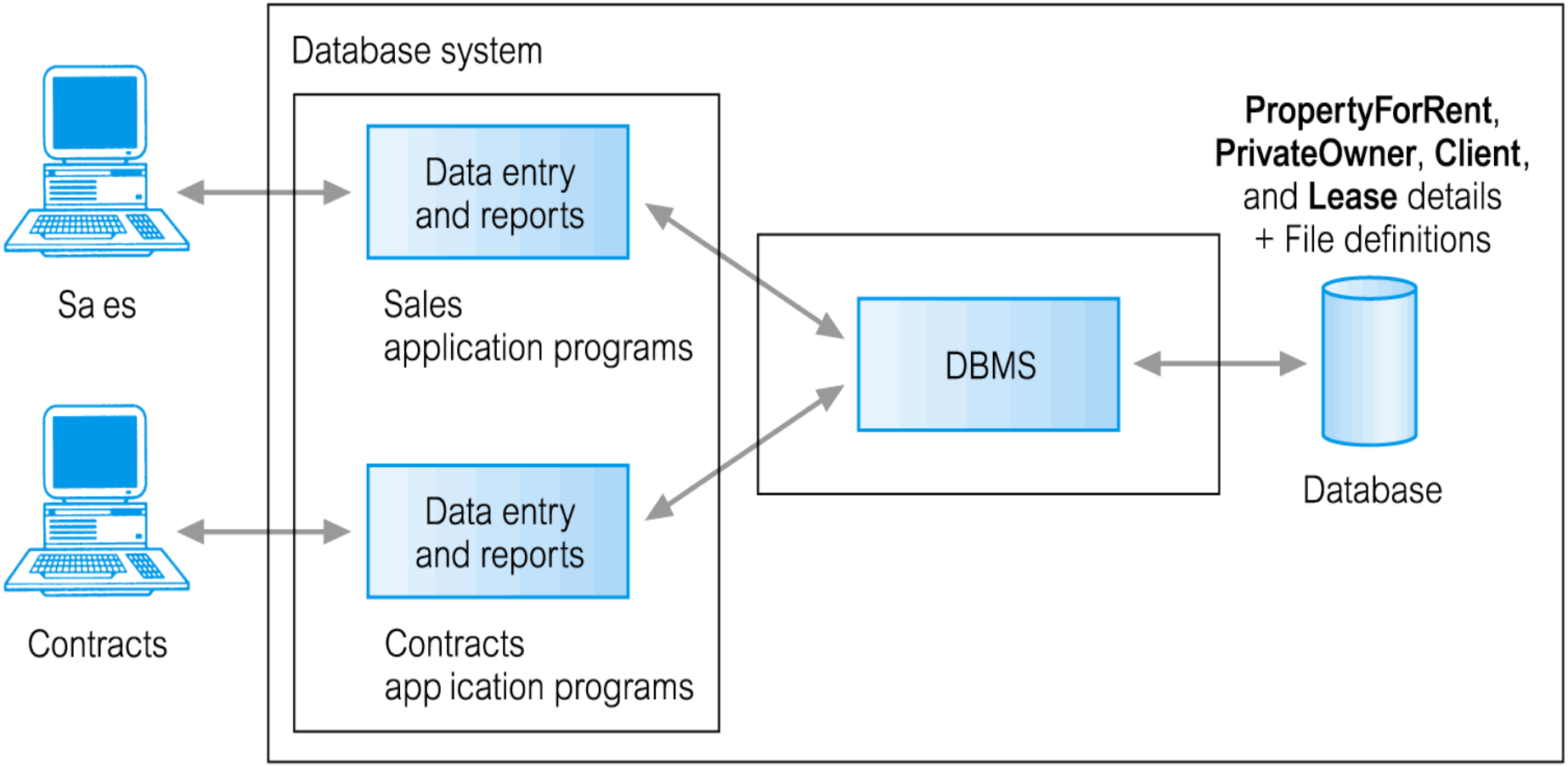
- The ***DBMS*** also provides ***application programming interfaces*** to procedural languages such as COBOL, C, Java, Visual Basic.NET, and C#.
- The DBMS also provides ***administrative utilities*** used by the DBA and the database designer to ***create, implement, monitor, and maintain the database***.

6. Database Communication Interfaces

- Current-generation DBMSs ***accept end-user requests*** via ***multiple, different network environments***.
- ***For example,*** the DBMS might provide ***access to the database*** via the ***Internet*** through the use of ***Web browsers*** such as ***Mozilla Firefox or Microsoft Internet Explorer***.

- In this environment, ***communications can be accomplished in several ways:***

- End users can generate answers to ***queries by filling in screen forms*** through their preferred Web browser.
- The ***DBMS can automatically publish predefined reports on a Website***.
- The DBMS can ***connect to third-party systems*** to distribute information via ***e-mail or other productivity applications.***



PropertyForRent (propertyNo, street, city, postcode, type, rooms, rent, ownerNo)
PrivateOwner (ownerNo, fName, lName, address, telNo)
Client (clientNo, fName, lName, address, telNo, prefType, maxRent)
Lease (leaseNo, propertyNo, clientNo, paymentMethod, deposit, paid, rentStart, rentF nish)

- Figure: Database Approach

Advantages of Database Systems

- **Data can be shared**: two or more users can access and use same data instead of storing data in redundant manner for each user
- **Improved Data Accessibility -** By using structured query languages, the users can easily access data without programming experience.
- **Redundancy can be reduced**: Isolated data is integrated in database to decrease the redundant data stored at different applications
- **Quality data can be maintained:** the different integrity constraints in the database approach will maintain the quality leading to better decision making.
- **Inconsistency can be avoided:** controlled data redundancy will avoid inconsistency of the data in the database to some extent.

Advantages of Database Systems…

- **Transaction support can be provided**: basic demands of any transaction support systems are implanted in a full scale DBMS.
- **Integrity can be maintained**: Data at different applications will be integrated together with additional constraints to facilitate shared data resource
- **Security measures can be enforced**: the shared data can be secured by having different levels of clearance and other data security mechanisms.
- **Improved decision support**: the database will provide information useful for decision making
- **Standards can be enforced:** the different ways of using and dealing with data by different units of the organization can be balanced and standardized by using database approach.

Advantages of Database Systems…

- **Less Labor**: Unlike the other data handling methods, data maintenance will not demand much resource
- **Centralized information control:** Since relevant data in the organization will be stored at one repository, it can be controlled and managed at the central level.
- **Data Independence** - Applications insulated from how data is structured and stored

Limitations and Risk of DB Approach

- Although the database system yields considerable ***advantages over previous data management approaches***, ***database systems do carry significant disadvantages***. For example:

1) Increased Costs.

- Database systems require sophisticated ***hardware and software and highly skilled personnel.***
- The ***cost of maintaining*** the ***hardware, software, and personnel required to operate and manage a database system can be substantial***.
- ***Training, licensing, and regulation compliance costs*** are often overlooked when database systems are implemented.

Limitations and Risk of DB Approach…

2) Management Complexity.

- Database systems ***interface with many different technologies*** and have a significant impact on a company's resources and culture.
- The ***changes introduced by the adoption of a database*** system must be properly managed to ***ensure that they help advance the company's objectives.***
- Given the fact that database systems hold crucial company data that are accessed from multiple sources, security issues must be assessed constantly.

Limitations and Risk of DB Approach…

3) Maintaining Currency

- To maximize the efficiency of the database system, you must keep your system current.
- Therefore, you must ***perform frequent updates and apply the latest patches and security measures to all components.***
- Because database technology advances rapidly, ***personnel training costs tend to be significant.***

4) Vendor Dependence

- Given the heavy ***investment in technology and personnel training,*** companies might be reluctant to change database vendors.

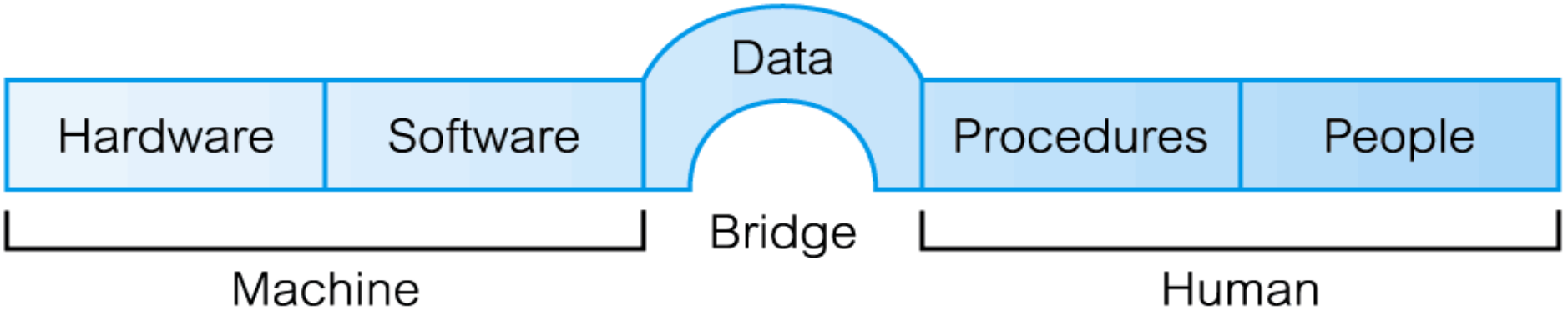
- As a consequence, vendors are less likely to offer pricing point advantages to existing customers, and those customers might be limited in their choice of database system components.

5) Frequent Upgrade/Replacement Cycles

- DBMS vendors frequently ***upgrade their products by adding new functionality.***
- Such new features often come bundled in new upgrade versions of the software.
- Some of these versions require ***hardware upgrades***.
- Not only do the ***upgrades themselves cost money***, but it also ***costs money to train database users and administrators*** to properly use and manage the new features.

Database System Environment

- The term database system refers to an organization of components that define and regulate the collection, storage, management, and use of data within a database environment.
- From a general management point of view, the database system is composed of the ***five major parts: hardware, software, people, procedures, and data***.



Database System Environment…

1) Hardware

- Hardware refers to all of the ***system's physical devices***; for example, ***computers (PCs, workstations, servers, and supercomputers), storage devices, printers, network devices (hubs, switches, routers, fiber optics),*** and other devices (automated teller machines, ID readers, and so on).

Database System Environment…

2) Software

- Although the most readily identified software is the ***DBMS itself***, to make the database system function fully, ***three types of software are needed***: ***operating system software, DBMS software, and application programs and utilities***.

➢ ***Operating system software*** manages all ***hardware components and makes it possible for all other software to run on the computers.***

➢ ***Examples of operating system*** software include ***Microsoft Windows, Linux, Mac OS, UNIX, and MVS***.

➢ ***DBMS software*** manages the ***database within the database system.***

Database System Environment…

- Some examples of DBMS software include ***Microsoft's SQL Server, Oracle Corporation's Oracle, Sun's MySQL, and IBM's DB2***.

➢ ***Application programs and utility software*** are used to ***access and manipulate data*** in the DBMS and to ***manage the computer environment*** in which data access and manipulation take place.

- ***Application programs*** are most commonly used to ***access data found within the database to generate reports***, ***tabulations, and other information*** to facilitate decision making.
- ***Utilities*** are the software tools used to help ***manage the database system's computer components***.

Database System Environment…

3) People

- This component includes all ***users of the database system***.
- On the basis of ***primary job functions***, ***five types of users can be identified in a database system***:
  - **System Administrators**
  - **Database Administrators**
  - **Database Designers**
  - **System Analysts and Programmers**
  - **End Users**
- Each user type, described below, performs both unique and complementary functions.

Database System Environment…

3.1 System Administrators oversee the database system's general operations.

3.2 Database Administrators, also known as DBAs, manage the DBMS and ensure that the database is functioning properly.

- In a database environment, the ***primary resource is the database itself***, and the ***secondary resource is the DBMS and related software***.

Database System Environment…

- Administering these resources is the responsibility of the database administrator (DBA).

➢The DBA is responsible for:

- Authorizing access to the database,
- Coordinating and monitoring its use, and
- Acquiring software and hardware resources as needed.

➢The DBA is accountable for problems such as

- Security breaches and poor system response time

Database System Environment…

3.3 Database Designers design the database structure.

- They are, in effect, the database architects.
- If the database design is poor, even the best application programmers and the most dedicated DBAs cannot produce a useful database environment.
- Because organizations strive to optimize their data resources, the database designer's job description has expanded to cover new dimensions and growing responsibilities.

Database System Environment…

A) Logical Designers

- Concerned with identifying the ***data, entities and attributes, relationships and constraints***
- Need a complete knowledge of the organization's ***data and business rules***

B) Physical Designers

- Decide how the logical database design is to be physically realized
- Map the logical database design into a set of tables
- Select specific storage structures and access methods
- Design security measures

Database System Environment…

3.4 System Analysts and Programmers design and implement the ***application programs.***

- ***Design and create the data entry screens, reports, and procedures*** through which ***end users access and manipulate the database's data.***
- For example work based on the requirement specified by the ***system analyst.***
- Each program contains statements that request the DBMS to perform some operation on the database – ***retrieving, inserting, updating and deleting data***

Database System Environment…

3.5 End users are the people who use the application programs to run the organization's daily operations.

- For example, ***salesclerks, supervisors, managers, and directors are all classified as end users***.
- High-level end users employ the information obtained from the database to make tactical and strategic business decisions.
- End users are differentiated by the way they expect to interact with the system.

Database System Environment…

- End Users can be either ***Native users or Sophisticated users***

➢ **Naive Users**
  - Unaware of the DBMS and the DB
  - Depend on the simplicity of the GUI

➢ **Sophisticated Users**
  - Familiar with the structure of the database and the facilities of the DBMS
  - May use high-level query languages (SQL) to perform the required operation
  - May even write application program for their own use

Database System Environment…

4) Procedures

- Procedures are the ***instructions and rules that govern the design and use of the database system***.
- Procedures are a critical, although occasionally forgotten, component of the system.
- Procedures play an important role in a company because they enforce the standards by which business is conducted within the organization and with customers.
- Procedures are also used to ensure that there is an organized way to ***monitor and audit both the data that enter the database and the information*** that is generated through the use of those data.

Database System Environment…

5. Data

- The word data covers the collection of facts stored in the database.
- Because data are the raw material from which information is generated, the determination of what data are to be entered into the database and how those data are to be organized is a vital part of the database designer's job.

Reading Assignment

1. Read about **Database development life cycle and phases**