LESSON 1

DATABASE SYSTEMS

WHY DATABASES?

Why Databases?

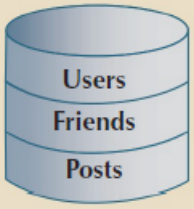
A Day In Susan's Life

See how many databases she interacts with each day

Before leaving for work, Susan checks her Facebook and Twitter accounts

Where is the data about the friends and groups stored?

Where are the "likes" stored and what would they be used for?



On her lunch break, she picks up her prescription at the pharmacy

Where is the pharmacy inventory data stored?

What data about each product will be in the inventory data?

What data is kept about each customer and where is it stored?

After work, Susan goes to the grocery store

Where is the product data stored?

Is the product quantity in stock updated at checkout?

Does she pay with a credit card?

At night, she plans for a trip and buys airline tickets and hotel reservations online

Where does the online travel website get the airline and hotel data from?

What customer data would be kept by the website?

Where would the customer data be stored?

Then she makes a few online purchases

Where are the product and stock data stored?

Where does the system get the data to generate product "recommendations" to the customer?

Where would credit card information be stored?

DATA

vs.

INFORMATION

Data vs. Information

- Consist of raw facts
- stored representations of meaningful objects and events
- *Structured*: numbers, text, dates
- *Unstructured*: images, video, documents

- Information is data that has been processed, organized, structured, or contextualized to provide meaning, relevance, and understanding.
- Information is valuable and helps people make decisions, understand situations, or gain insights. For example, a graph showing the trend of the numbers (e.g., a line chart plotting the data points) provides information about the progression over time.

Data vs. Information

Class Roster

Course: MGT 500 Business Policy

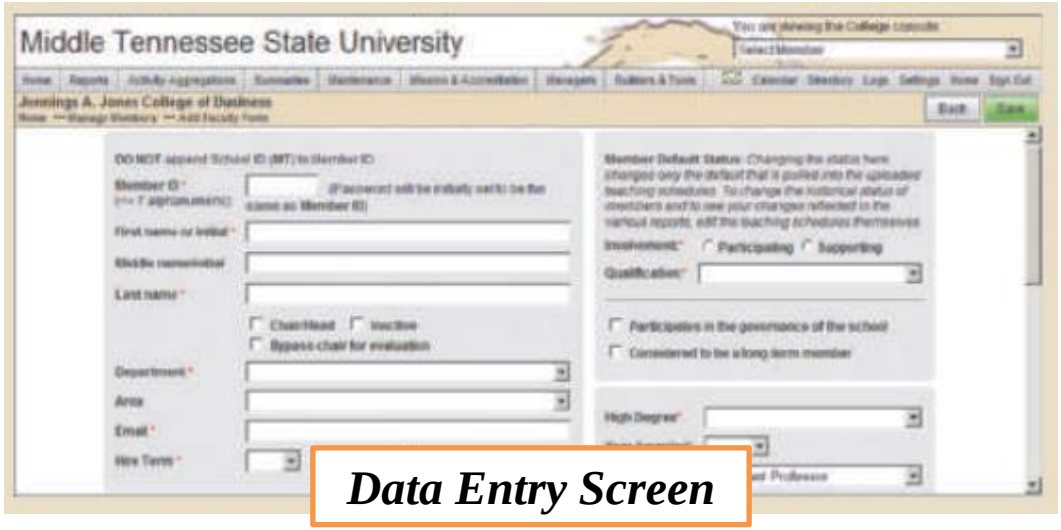
Semester: Spring 200X

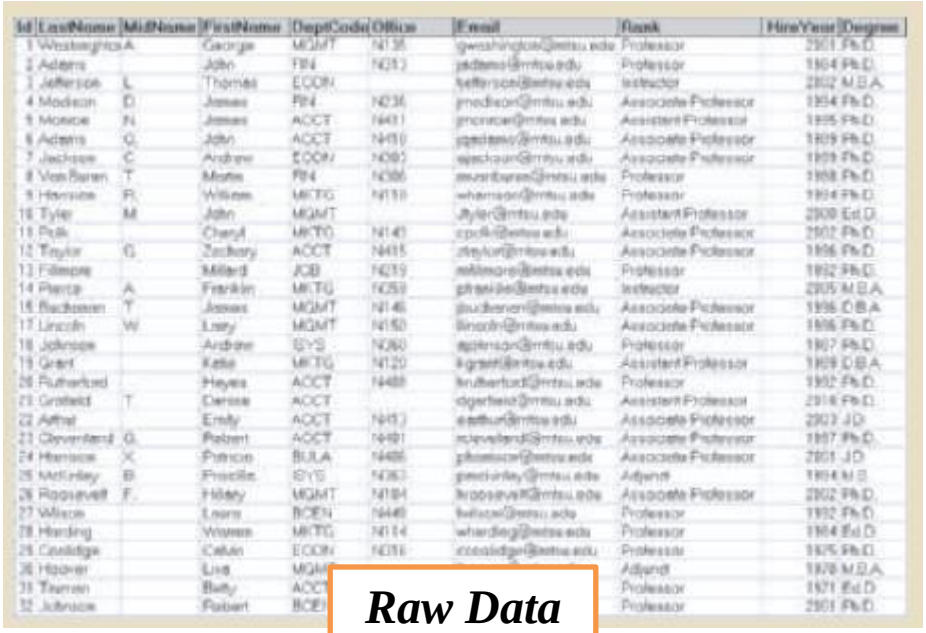
Section: 2

Name	ID	Major	GPA
Baker, Kenneth D.	324917628	MGT	2.9
Doyle, Joan E.	476193248	MKT	3.4
Finkle, Clive R.	548429344	PRM	2.8
Lewis, John C.	551742186	MGT	3.7
McFerran, Debra R.	409723145	IS	2.9
Sisneros, Michael	392416582	ACCT	3.3

Context helps users understand data

Data vs. Information

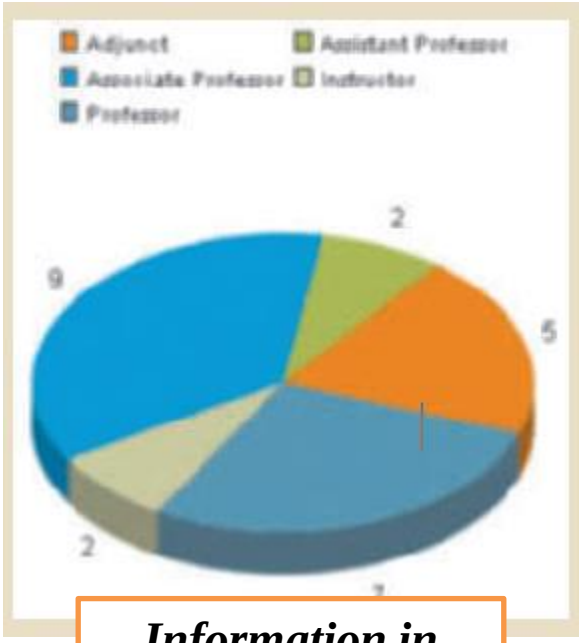




Data vs. Information

Rank	COUNT	%INFS	TOT/COL	%/COL. TOT.	%/COL. FAC.
Adjunct	5	20.00%	23	21.74%	3.27%
Assistant Professor	2	8.00%	28	7.14%	1.31%
Associate Professor	9	36.00%	37	24.32%	5.88%
Instructor	2	8.00%	18	11.11%	1.31%
Professor	7	28.00%	47	14.89%	4.58%

Information in Summary Format



Information in Graphical Format

Data vs. Information

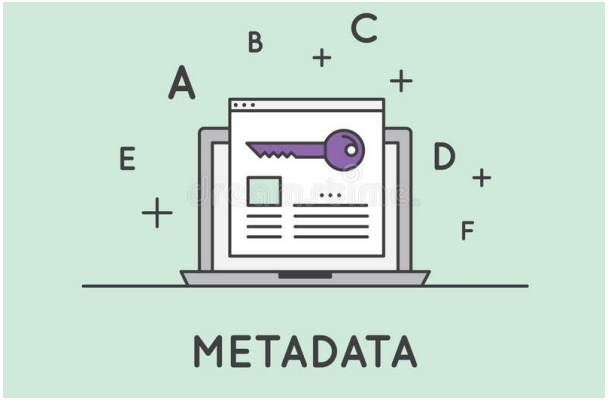
Key Points

- 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

INTRODUCING THE DATABASE

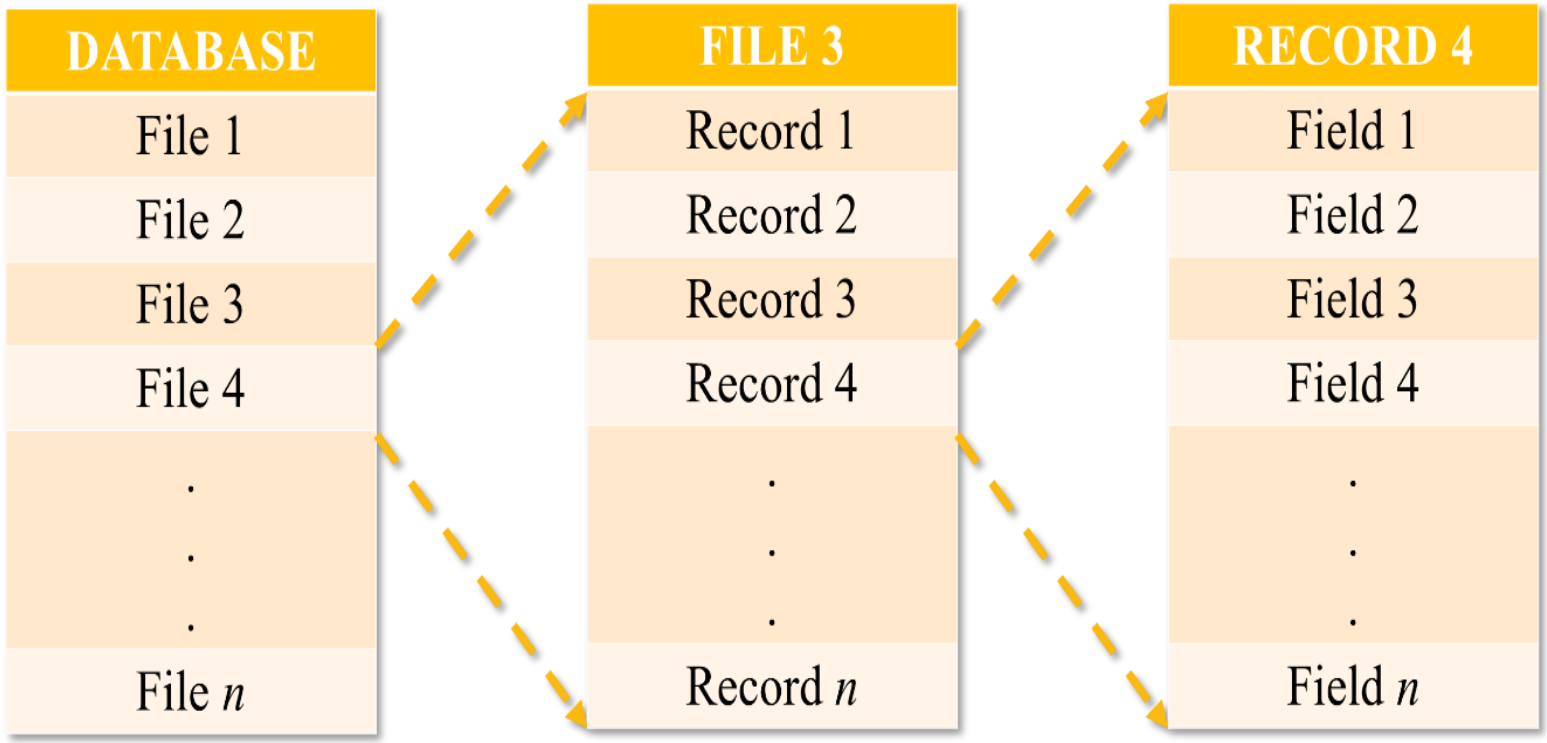
Introducing the Database

- used to store, manipulate, and retrieve data in nearly every type of organization including business, health care, education, government, and libraries.



- a data that describe the properties or characteristics of other data
- also refers to the descriptions of the properties or characteristics of the data, including data types, field sizes, allowable values, and documentation.

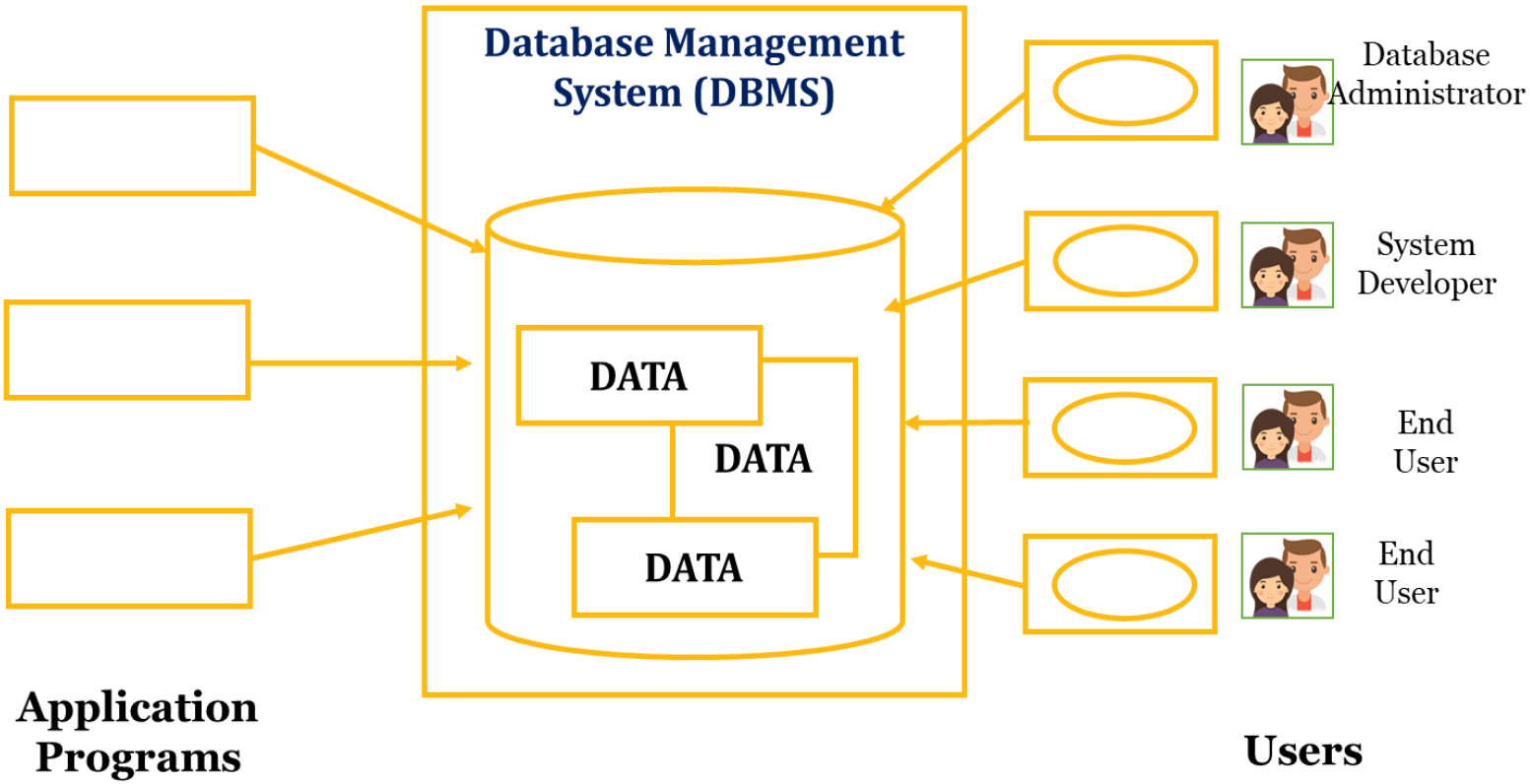
Introducing the Database



Graphical representation of a database

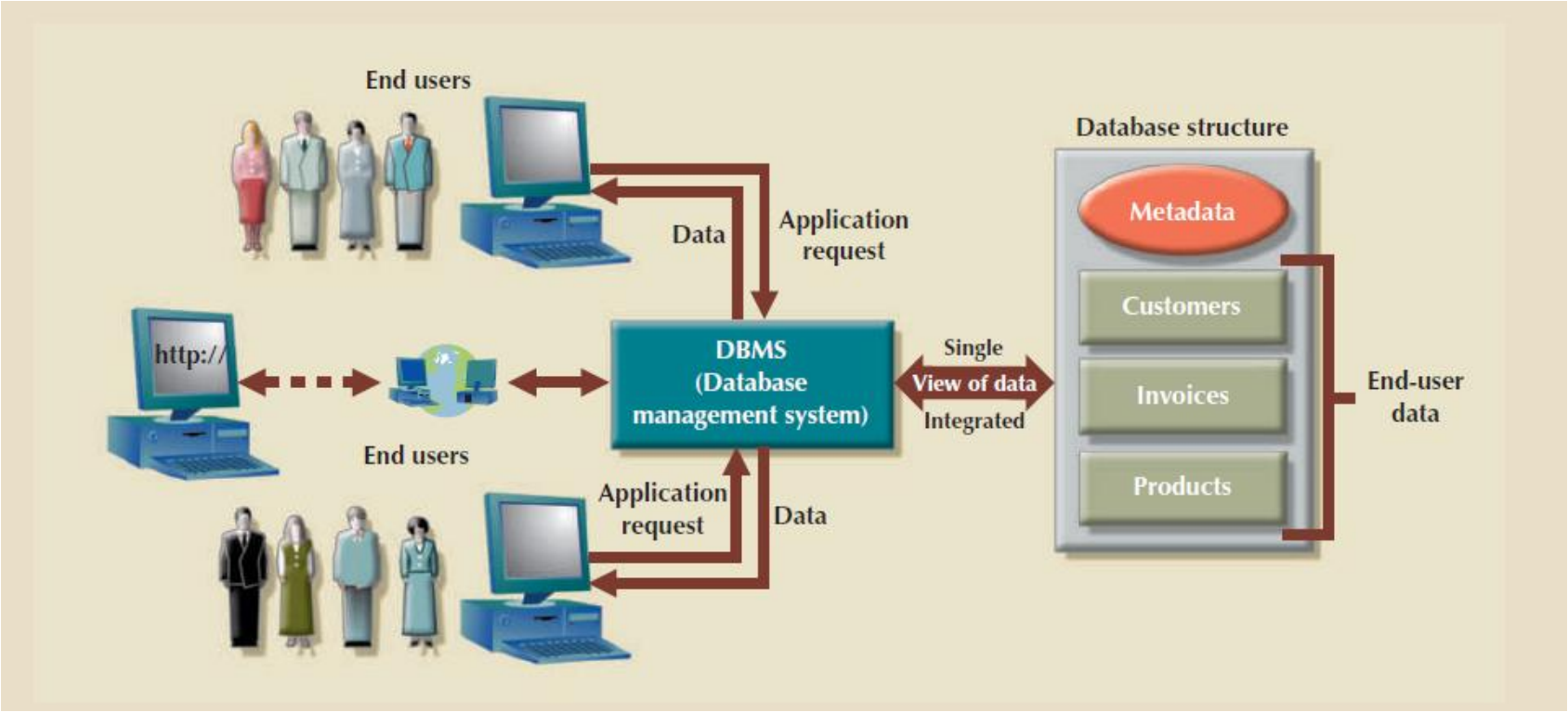
Database Management System (DBMS)

- is a collection of programs that manages the database structure and controls access to the data stored in the database.



Role and Advantages of the DBMS

- serves as the intermediary between the user and the database.
- database structure itself is stored as a collection of files, and the only way to access the data in those files is through the DBMS.
- integrates the many different users' views of the data into a single all-encompassing data repository.



Role and Advantages of the DBMS

- DBMS helps make data management more efficient and effective. In particular, a DBMS provides these advantages:

  - **Improved data sharing**. The DBMS facilitates end-user access to more and better-managed data. This type of access allows users to react quickly to changes in their environment.
  - **Improved data security.** Data security breaches increase as more users access data. Corporations spend a lot of time, effort, and money protecting their data. A DBMS helps enforce data privacy and security policies.
  - **Minimized Data Inconsistency.** Data inconsistency occurs when multiple versions of the same data appear in different places. A properly designed database reduces the likelihood of data inconsistency.

Role and Advantages of the DBMS

- DBMS helps make data management more efficient and effective. In particular, a DBMS provides these advantages:

  - **Improved data access.** The DBMS allows for quick responses to queries. A query is a specific request made to a DBMS for data manipulation.
  - **Improved decision making.** Better data management and access allows for better decision-making. The quality of the output data determines the quality of the output data. Data quality is a thorough approach to ensuring data accuracy, validity, and timeliness. In addition to providing a framework, the DBMS ensures data quality.
  - **Increased end-user productivity.** Because data is readily available, users can make quick, informed decisions that can contribute to making the difference between success and failure in the global economy.

Types of Databases

Based on Number of Users

- **Single-user Database**. A database that can only accommodate a single user at a time.
  - **Desktop Database.** A database for a single user that runs on a personal computer.
- **Multiuser Database.** A database that can accommodate multiple users at the same time.
  - **Workgroup Database.** A multiuser database that typically supports fewer than 50 users or is used by a single department within an organization.
  - **Enterprise Database**. A database that is used throughout the organization and supports a large number of users (more than 50, usually hundreds) from various departments.

Types of Databases

Based on Location

- **Centralized Database**. A database that supports data from a single site.
- **Distributed Database.** A database that supports data that is distributed across multiple sites.
- **Cloud Database**. A database created and maintained using cloud services like Microsoft Azure or Amazon AWS

WHY DATABASE DESIGN IS IMPORTANT

Why Database Design Is Important?

- **Database design** refers to the work on designing a database structure to store and manage end-user information.

Why are there blanks in rows 9 and 10?

How to produce an alphabetical listing of employees?

How to count how many employees are certified in Basic Database Manipulation?

Is Basic Database Manipulation the same as Basic DB Manipulation?

What if an employee acquires a fourth certification? Should we add another column?

ID	ENum	Name	Title	HireDate	Skill1	Skill1Date	Skill2	Skill2Date	Skill3	Skill3Date
1	02345	Brian Oates	DBA	2/14/1997	Basic Database Management	2/14/2004	Advanced Database Manageeneet	2/14/2007	Basic Web Design	8/9/2005
2	08273	Marco Bienz	Analyst	7/28/2008	Basic Web Design	3/8/2011	Advance Process Modeling	8/19/2014		
3	06234	Jasmine Patel	Programmer	8/10/2007	Basic Web Design	8/10/2009	Advanced C# programming	8/10/2009	Basic DB manipulation	1/29/2014
4	03373	Franklin Johnson, Jr.	Purchasing Agent	3/15/2004	Advanced Spreadsheets	6/20/2013				
5	13567	Almond, Robert	Analyst	9/30/2014	Basic Process Modeling	9/30/2016	Basic Database Design	5/23/2017		
6	10282	Richardson, Amanda	Clerk	4/11/2013						
7	09382	Susan Mathis	Database Programmer	8/2/2012	Basic DB Design	8/2/2014	Basic Database Manipulation	8/2/2014	Advanced DB Manipulation	5/1/2015
8	14311	Duong, Lee	Programmer	9/1/2016	Basic Web Design	9/1/2018				
9					Master Database Programming					
10					Basic Spreadsheets					
11	09002	Wade Gaither	Clerk	5/20/2012	Advanced Spreadsheets	5/16/2015	Basic Web Design	5/16/2015		
12	13383	Raymond F. Matthews	Programmer	3/12/2014	Basic C# Programming	3/12/2016				
13	09283	Chavez, Juan	Clerk	7/4/2012						
14	04893	Patricia Richards	DBA	6/11/2006	Advanced Database Management	6/11/2008	Advanced Database Manipulation	9/20/2014		
15	13932	Lee, Megan	Programmer	9/29/2015						

Why Database Design Is Important?

Database design is important because…

An effective data management and the generation of accurate and valuable information are made easier with a well-designed database. The failure of an organization can be caused by the failure of a poorly designed database, which will become a breeding ground for difficult-to-trace errors that will lead to poor decision making—and poor decision making will lead to the failure of an organization.

EVOLUTION OF FILE SYSTEM DATA PROCESSING

Evolution of File System Data Processing

- **Manual File Systems**
  - Data was stored as paper records.
  - Lot of man power involved.
  - Lot of time was wasted (e.g. when searching) therefore, inefficient.

Advantages

- Cannot be destroyed by an accidental power loss
- Hackers cannot hack a manual filing system

Disadvantages

- Summarizing data and writing reports takes a lot of time
- Data duplication
- Lack of security
- Repetition of work
- Slow retrieval of data

Evolution of File System Data Processing

- **Computerized File Systems**
  - create a computer-based system that would track data and produce required reports. Initially, the computer files within the file system were similar to the manual files.

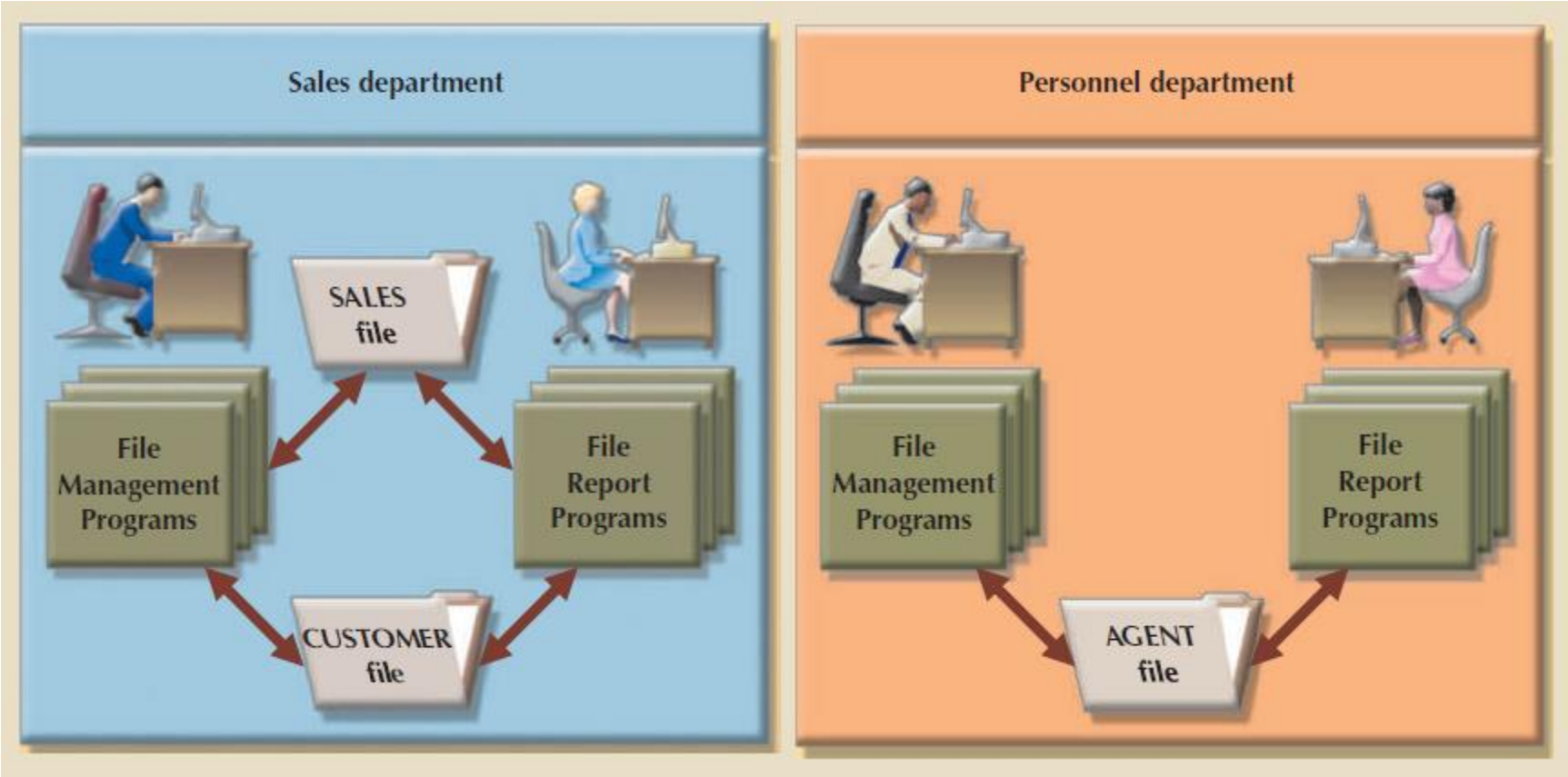
C_NAME	C_PHONE	C_ADDRESS	C_ZIP	A_NAME	A_PHONE	TP	AMT	REN
Alfred A. Ramas	615-844-2573	218 Fork Rd., Babs, TN	36123	Leah F. Hahn	615-882-1244	T1	100.00	05-Apr-2018
Leona K. Dunne	713-894-1238	Box 12A, Fox, KY	25246	Alex B. Alby	713-228-1249	T1	250.00	16-Jun-2018
Kathy W. Smith	615-894-2285	125 Oak Ln, Babs, TN	36123	Leah F. Hahn	615-882-2144	S2	150.00	29-Jan-2019
Paul F. Olowski	615-894-2180	217 Lee Ln., Babs, TN	36123	Leah F. Hahn	615-882-1244	S1	300.00	14-Oct-2018
Myron Orlando	615-222-1672	Box 111, New, TN	36155	Alex B. Alby	713-228-1249	T1	100.00	28-Dec-2018
Amy B. O'Brian	713-442-3381	387 Troll Dr., Fox, KY	25246	John T. Okon	615-123-5589	T2	850.00	22-Sep-2018
James G. Brown	615-297-1228	21 Tye Rd., Nash, TN	37118	Leah F. Hahn	615-882-1244	S1	120.00	25-Mar-2019
George Williams	615-290-2556	155 Maple, Nash, TN	37119	John T. Okon	615-123-5589	S1	250.00	17-Jul-2018
Anne G. Farriss	713-382-7185	2119 Elm, Crew, KY	25432	Alex B. Alby	713-228-1249	T2	100.00	03-Dec-2018
Olette K. Smith	615-297-3809	2782 Main, Nash, TN	37118	John T. Okon	615-123-5589	S2	500.00	14-Mar-2019

C_NAME = Customer name
C_PHONE = Customer phone
C_ADDRESS = Customer address
C_ZIP = Customer zip code

A_NAME = Agent name
A_PHONE = Agent phone
TP = Insurance type
AMT = Insurance policy amount, in thousands of $
REN = Insurance renewal date

Evolution of File System Data Processing

- **Computerized File Systems**



A Simple File System

Evolution of File System Data Processing

- **Computerized File Systems**

Basic File Terminology	
Term	**Definition**
Data	Raw facts, such as a telephone number, a birth date, a student name, and grades. 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, place, or thing. For example, the fields that constitute a record for a student might consist of the student's name, address, phone number, date of birth, program, and unpaid balance.
File	A collection of related records. For example, a file might contain data about the students currently enrolled at Camarines Sur Polytechnic Colleges.

Evolution of File System Data Processing

- **Computerized File Systems**

Advantages

- Information sharing
- Database management
- Storage of data
- Easy access
- Easy information retrieval

Evolution of File System Data Processing

- **Computerized File Systems**

Disadvantages

- **Program-Data Dependence**. All programs maintain metadata for each file they use
- **Duplication of Data**. Different systems/programs have separate copies of the same data
- **Limited Data Sharing**. No centralized control of data
- **Lengthy Development Times**. Programmers must design their own file formats
- **Excessive Program Maintenance**. 80% of information systems budget

PROBLEMS WITH FILE SYSTEM DATA PROCESSING

Problems with File System Data Processing

Problems with Data Dependency

- Each application programmer must maintain his/her own data
- Each application program needs to include code for the metadata of each file
- Each application program must have its own processing routines for reading, inserting, updating, and deleting data
- Lack of coordination and central control
- Non-standard file formats

Problems with File System Data Processing

Problems with Data Redundancy

- Waste of space to have duplicate data
- Causes more maintenance headaches
- The biggest problem:
  - Data changes in one file could cause inconsistencies
  - Compromises in **data integrity.**

Note:

Data integrity is defined as the condition in which all of the data in the database is consistent with the real-world events and conditions. In other words, data integrity means the following:

- **Data is accurate** - there are no data inconsistencies.
- **Data is verifiable** - the data will always yield consistent results.

WHAT IS THE SOLUTION?

Use the **DATABASE** approach

DATABASE SYSTEMS

Database Systems

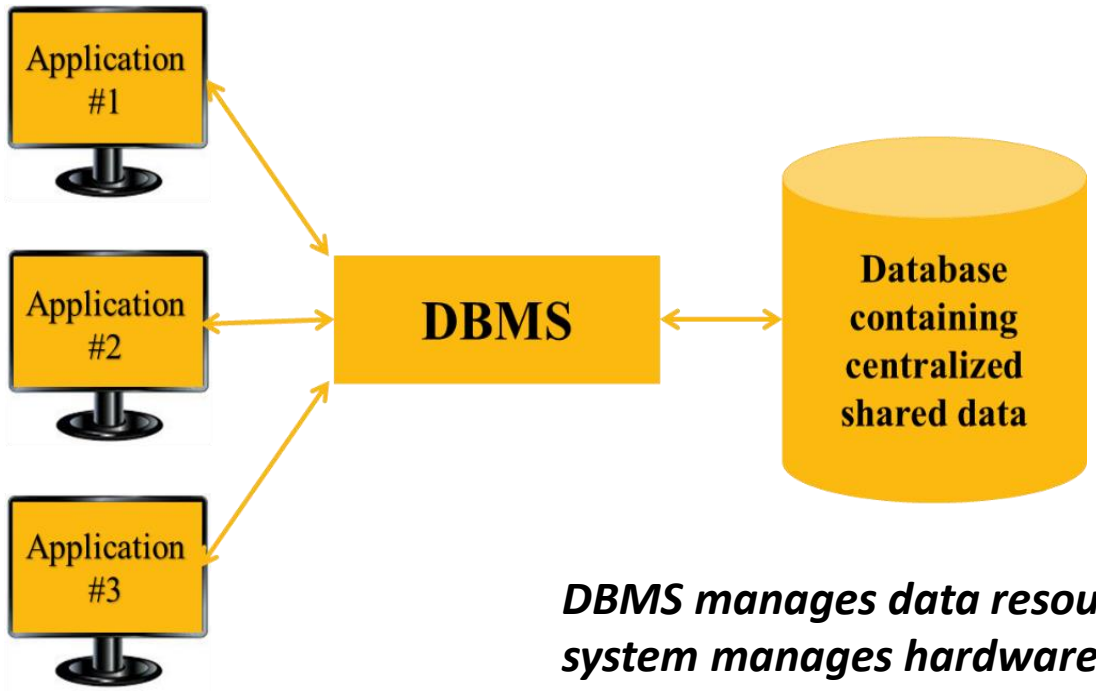
A Database Approach is…

- A central repository of shared data
- Data is managed by a controlling agent
- Stored in a standardized, convenient form

But always remember that this approach requires a ***Database Management System (DBMS)***.

Database Management System (DBMS)

- A software system that is used to create, maintain, and provide controlled access to user databases
- Set of programs to access the data
- contains information about a particular enterprise
- provides an environment that is both convenient and efficient to use.



DBMS manages data resources like an operating system manages hardware resources

Advantages of the Database Approach

- **Program-Data Independence**
  - Metadata stored in DBMS, so applications don't need to worry about data formats
- **Minimal Data Redundancy**
  - Leads to increased data integrity/consistency
- **Improved Data Sharing**
  - Different users get different views of the data

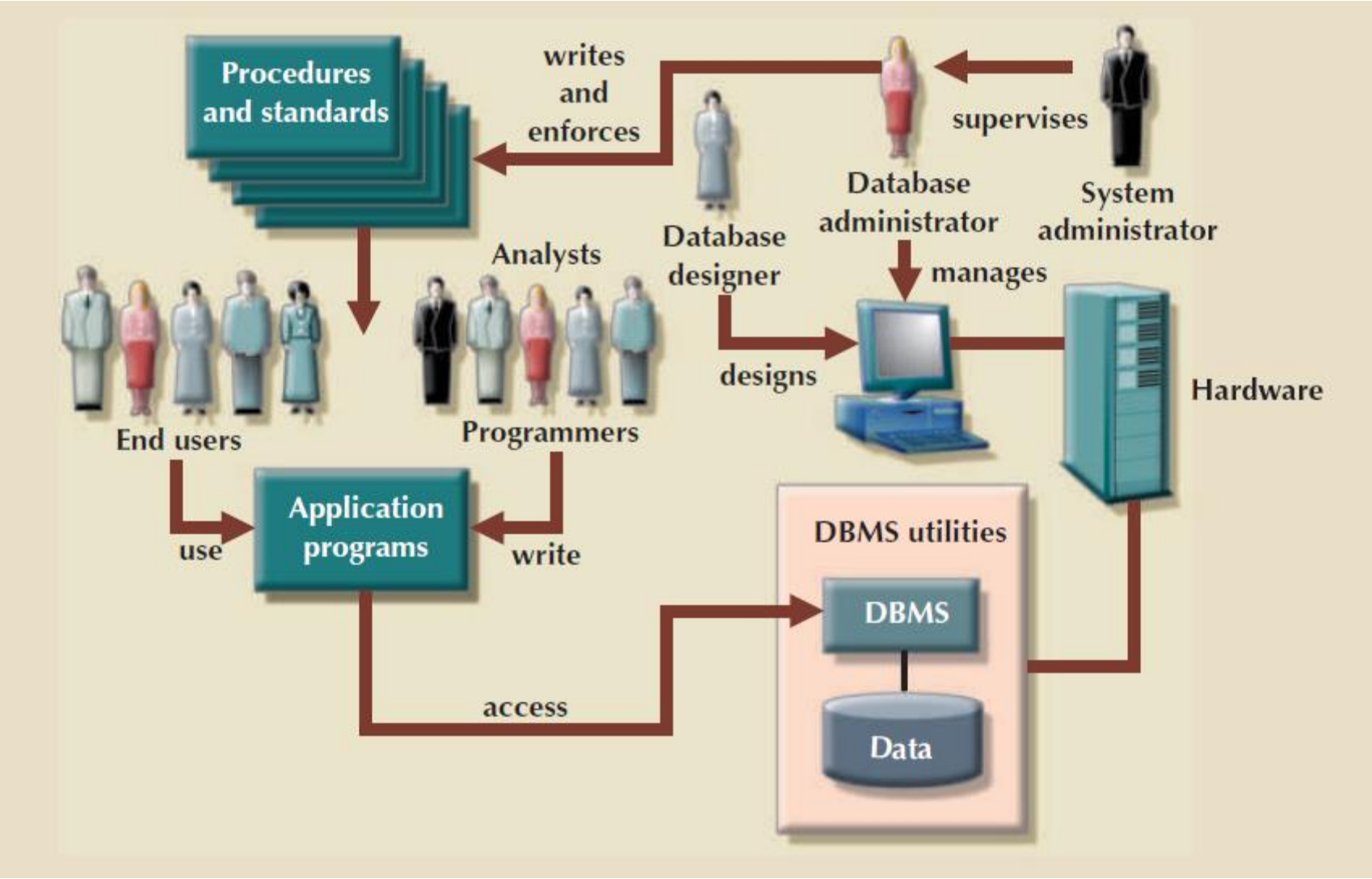
Advantages of the Database Approach

- **Enforcement of Standards**
  - All data access is done in the same way
- **Improved Data Quality**
  - Constraints, data validation rules
- **Better Data Accessibility/ Responsiveness**
  - Use of standard data query language (SQL)
- **Security, Backup/Recovery, Concurrency**
  - Disaster recovery is easier

Costs and Risks of the Database Approach

- New, specialized personnel
- Installation and management cost and complexity
- Conversion costs
- Need for explicit backup and recovery
- Organizational conflict

The Database System Environment

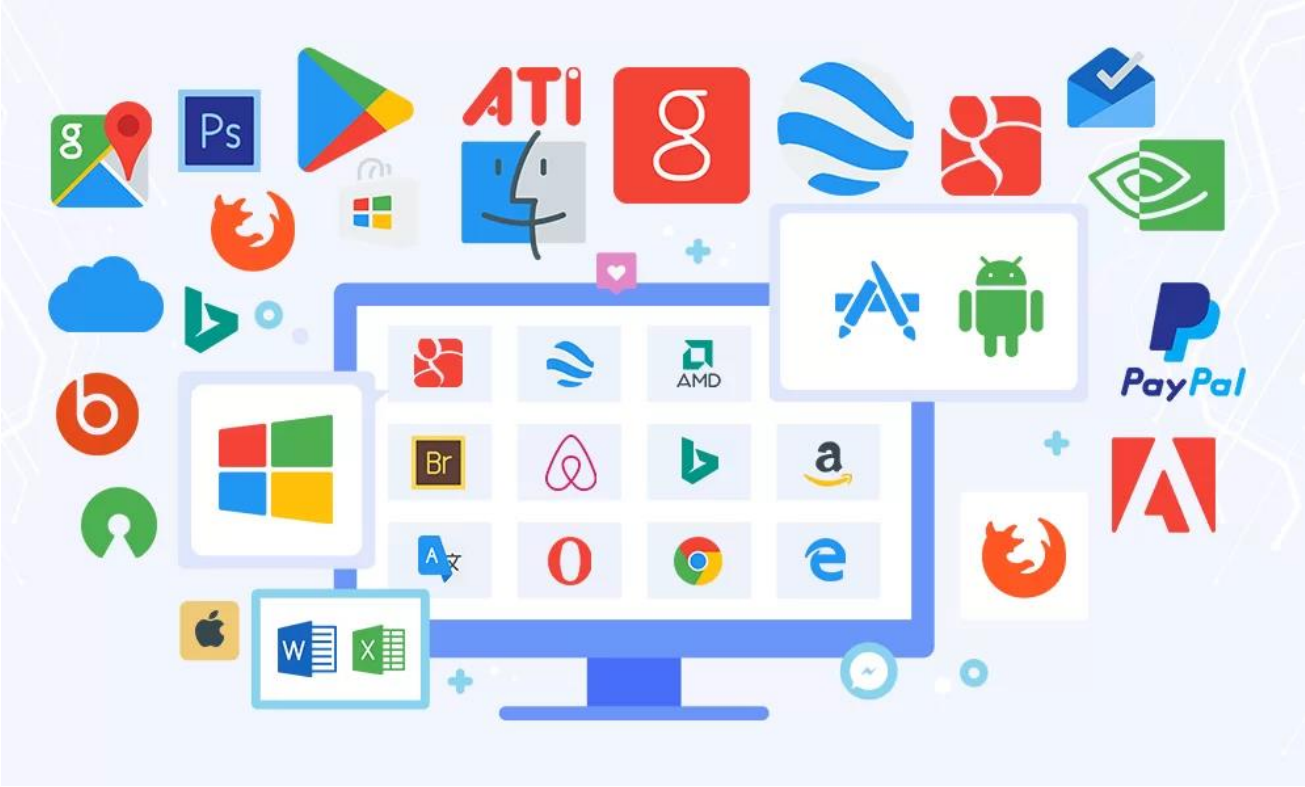


The Database System Environment

- **Hardware.** Hardware refers to all of the system's physical devices, including computers (PCs, tablets, 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).

The Database System Environment

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



The Database System Environment

- **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. Each user type, described next, performs both unique and complementary functions.

System Administrators

Oversee the database system's general operations.

Database Administrators

Manage the DBMS and ensure that the database is functioning properly.

Database Designers

Design the database structure. They are, in effect, the database architects.

System Analyst and Programmers

Design and implement the application programs. They design and create the data-entry screens, reports, and procedures through which end users access and manipulate the database's data.

End Users

The people who use the application programs to run the organization's daily operations.

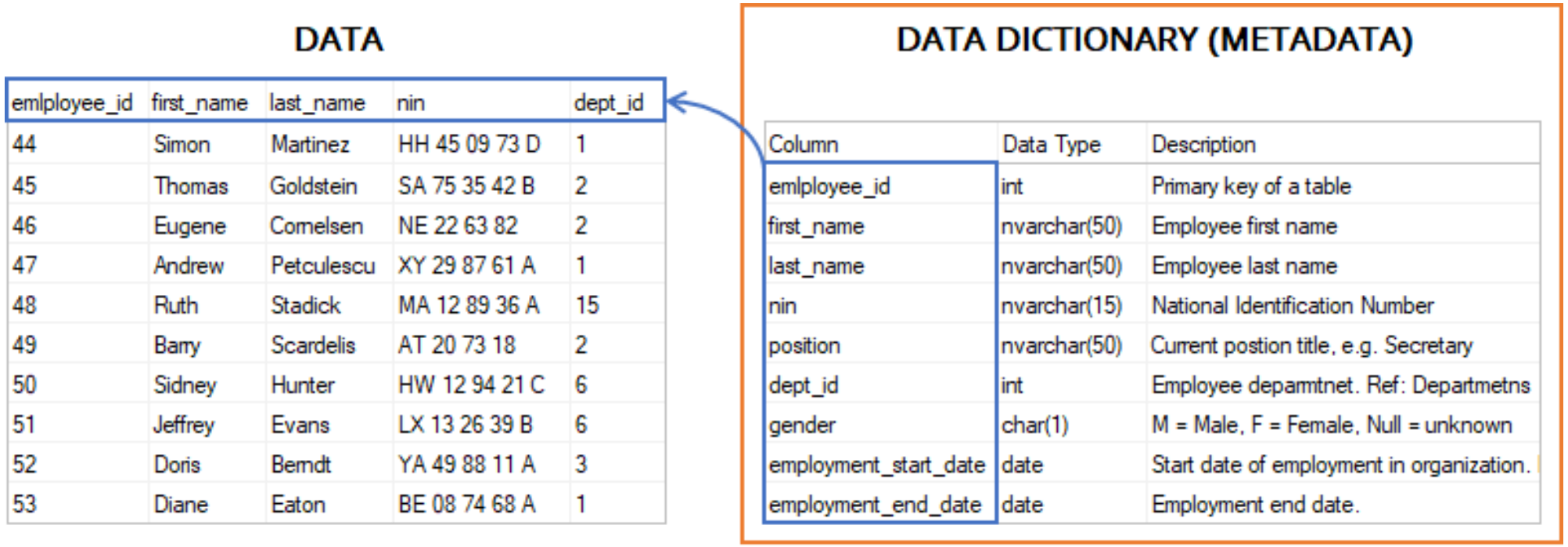
The Database System Environment

- **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.

- **Data.** The raw material from which information is generated, so deciding which data to enter into the database and how to organize it is critical.

DBMS Functions

- **Data Dictionary Management.** A data dictionary in Database Management System (DBMS) can be defined as a component that stores the collection of names, definitions, and attributes for data elements that are being used in a database. The Data Dictionary stores metadata, i.e., data about the database.



emlployee_id	first_name	last_name	nin	dept_id
44	Simon	Martinez	HH 45 09 73 D	1
45	Thomas	Goldstein	SA 75 35 42 B	2
46	Eugene	Cornelsen	NE 22 63 82	2
47	Andrew	Petculescu	XY 29 87 61 A	1
48	Ruth	Stadick	MA 12 89 36 A	15
49	Barry	Scardelis	AT 20 73 18	2
50	Sidney	Hunter	HW 12 94 21 C	6
51	Jeffrey	Evans	LX 13 26 39 B	6
52	Doris	Berndt	YA 49 88 11 A	3
53	Diane	Eaton	BE 08 74 68 A	1

Column	Data Type	Description
emlployee_id	int	Primary key of a table
first_name	nvarchar(50)	Employee first name
last_name	nvarchar(50)	Employee last name
nin	nvarchar(15)	National Identification Number
position	nvarchar(50)	Current postion title, e.g. Secretary
dept_id	int	Employee deparmtnet. Ref: Departmetns
gender	char(1)	M = Male, F = Female, Null = unknown
employment_start_date	date	Start date of employment in organization.
employment_end_date	date	Employment end date.

DBMS Functions

- **Data Storage Management.** Data storage management refers to the process of managing data more effectively. It involves the organization and optimization of data storage to ensure efficient data retrieval, storage, and maintenance. Here are key concepts and techniques related to data storage management in a DBMS:
  - **Data Modeling and Schema Design:** Before data storage can be managed effectively, you need a well-defined data model and schema. The schema defines the structure of the database, including tables, columns, data types, and relationships.
  - **Data Types:** Choosing appropriate data types for columns is crucial. Smaller data types are more space-efficient, but they should be able to accommodate the data you need to store.

DBMS Functions

Continued...

- **Storage Structures:** DBMSs use various storage structures to organize and manage data, such as tables, indexes, and views. Understanding how these structures work and when to use them is essential.
- **File Organization:** Data is stored in files or filegroups. Understanding the file organization method (e.g., heap, clustered index, non-clustered index) is crucial for efficient storage and retrieval.
- **Indexing:** Indexes are used to speed up data retrieval. They require storage space, so you must balance the benefits of indexing with the additional storage costs. Careful selection and maintenance of indexes are essential.
- **Data Compression:** DBMSs often support data compression techniques to reduce storage space while maintaining data integrity. This can significantly reduce storage costs.

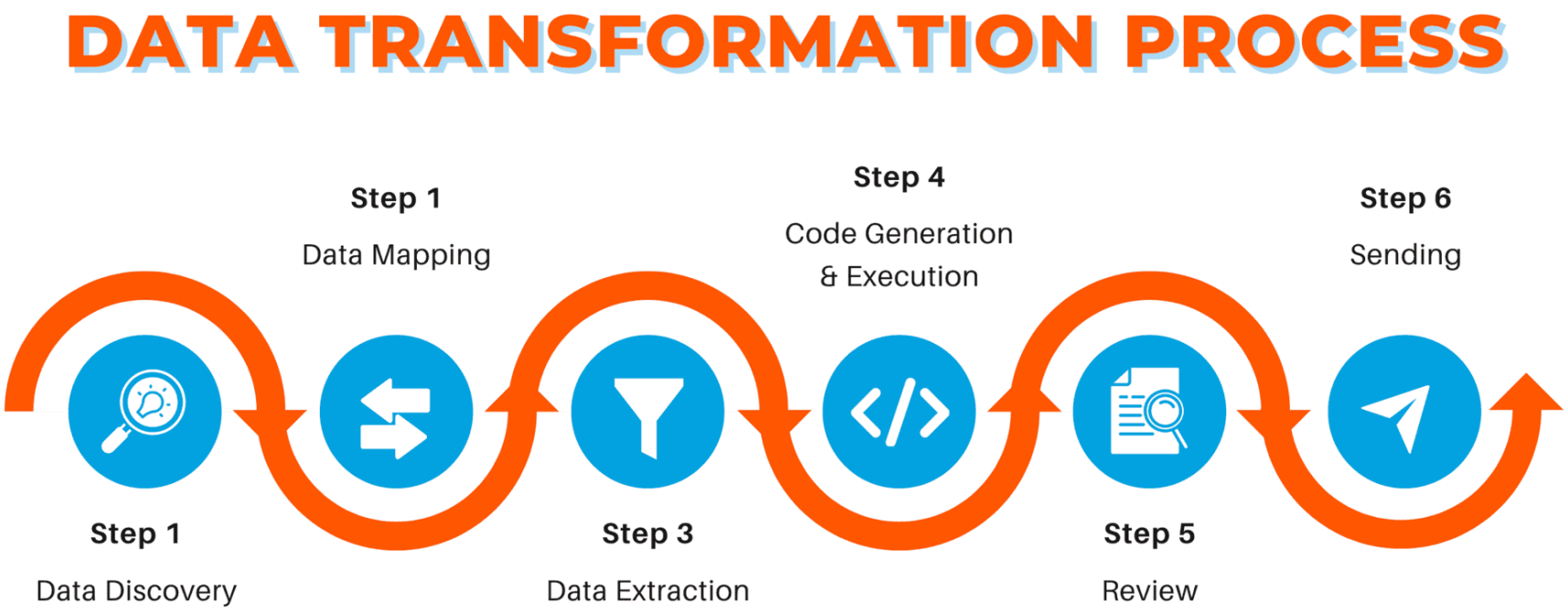
DBMS Functions

Continued...

- **Partitioning:** Partitioning involves dividing a large table into smaller, more manageable pieces called partitions. It improves query performance and makes maintenance tasks more manageable.
- **Normalization and Denormalization:** Normalization involves minimizing data redundancy by organizing data into separate tables, while denormalization combines tables to improve query performance. The choice depends on your specific needs and trade-offs.
- **Data Archiving and Purging:** Implement strategies for archiving or purging old or infrequently accessed data to reduce storage requirements and improve performance.
- **Backup and Recovery:** Establish backup and recovery procedures to safeguard data. This includes storing backups efficiently and ensuring quick recovery in case of data loss.

DBMS Functions

- **Data Transformation and Presentation.** Data transformation is the process of converting data from one format, into another. Transformations typically involve converting a raw data source into a cleansed, validated and ready-to-use format.



DBMS Functions

- **Security Management.** The DBMS enforces user security and data privacy. User access to the database is controlled by security rules that specify which data items each user can access and what data operations they can perform. This is crucial in multiuser databases.
- **Multiuser Access Control.** The DBMS uses sophisticated algorithms to ensure that multiple users can access the database simultaneously without compromising its integrity.
- **Back up and Recovery Management.** The DBMS ensures data integrity and safety by backing up data. Management of recovery from failures such as bad sectors in disks or power outages is called recovery management. This ability is critical to database integrity.

DBMS Functions

- **Data Integrity Management.** Data redundancy is minimized and data consistency is maximized by the DBMS. The data dictionary's data relationships are used to enforce data integrity. Data integrity is critical in transaction-oriented databases.
- **Database Access Languages and Application Programming Interfaces.** The data dictionary's relationships are used to enforce The DBMS provides query language data access. In a query language, the user specifies what to do without specifying how.
- **Database Communication Interfaces.** A modern DBMS accepts user requests from multiple network environments.

Queries?

Carlo Adonis S. San Carlos
Email Address: carlosancarlos@cspc.edu.ph