Competency 8

Designs and develops database systems to manage data efficiently and effectively
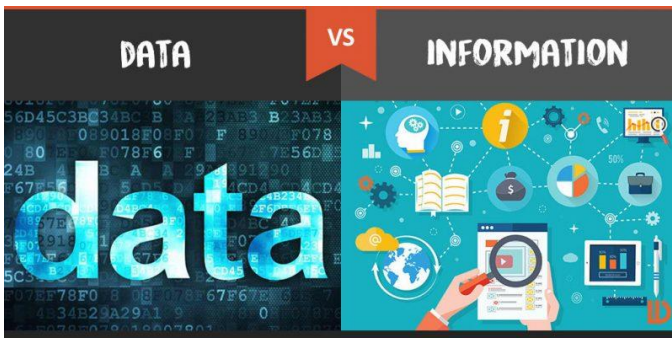
8.1 Learns the basics of information and data, and the need for databases

Time: 2 periods

Learning Outcomes

- Distinguishes data and information
- Defines database
- Lists and briefly describes the database models
- Compares and contrasts database models in terms of their features

Distinguishes data and information

Data can be identified as the inputs into a system that intents to create meaningful information. Information can be defined as the data organized meaningfully. Creation of information can be identified as processing of data. For the purpose of processing there should be instructions to follow. After processing data according to the instructions given, information is produced which is the outcome the processed data. Although information is an output of a process the same information would be a data for another process. Thus, data and information can be identified distinctly through examining inputs and outputs of a particular process.

Data

•**Raw facts**

- Raw data -Not yet been processed to reveal the meaning

•**Building blocks** of information

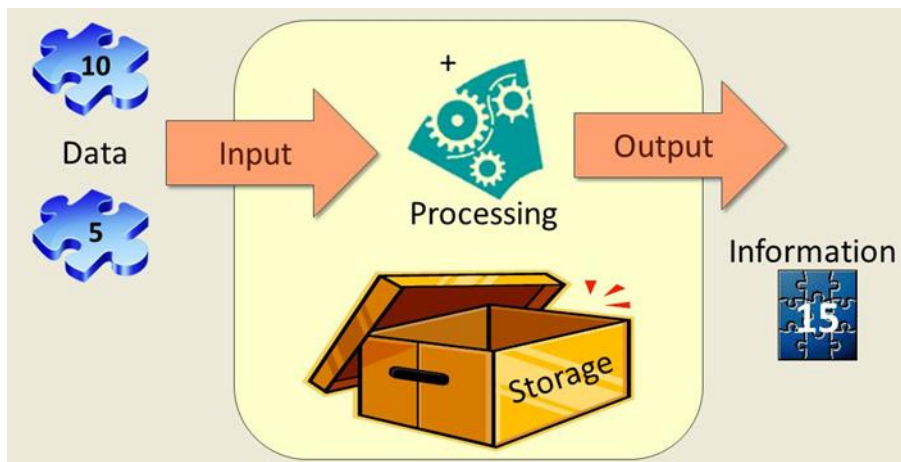•Data **management**

- Data generation, storage, and retrieval

Information

•Produced by **processing** data

•Reveals the **meaning** of data

•Enables **knowledge creation**

•Should be **accurate, relevant, and timely** to enable good decision making

Structured Vs. unstructured data

Structured Data

- Structured data refers to any data that resides in a fixed field within a record or file. This includes data contained in relational databases and spreadsheets.

- This includes defining what fields of data will be stored and how that data will be stored: data type (numeric, currency, alphabetic, name, date, address) and any restrictions on the data input (number of characters; restricted to certain terms such as Mr., Ms. or Dr., M or F).

- Structured data has the advantage of being easily entered, stored, queried and analyzed.

- At one time, because of the high cost and performance limitations of storage, memory and processing, relational databases and spreadsheets using structured data were the only way to effectively manage data.

- Anything that couldn't fit into a tightly organized structure would have to be stored on paper in a filing cabinet.

- Structured data is often managed using Structured Query Language (SQL) – a programming language created for managing and querying data in relational database management systems. Originally developed by IBM in the early 1970s and later developed commercially by Relational Software, Inc. (now Oracle Corporation).
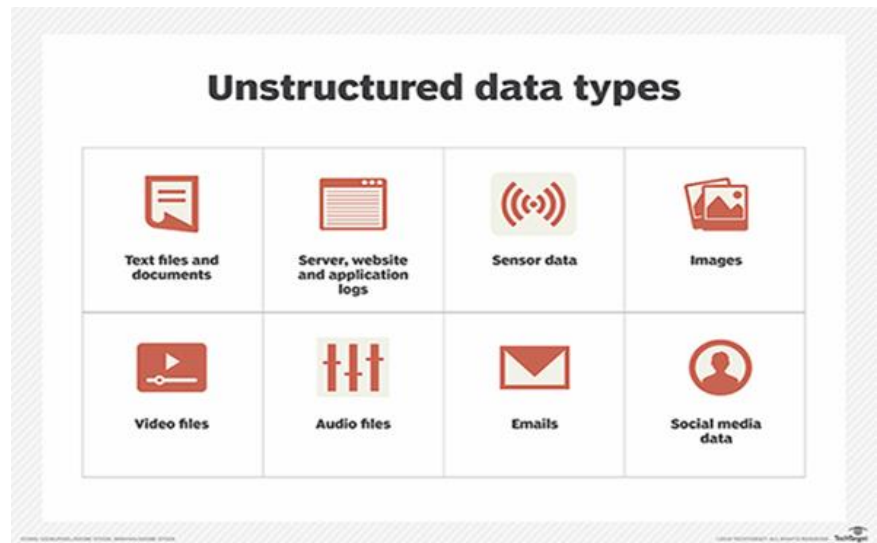
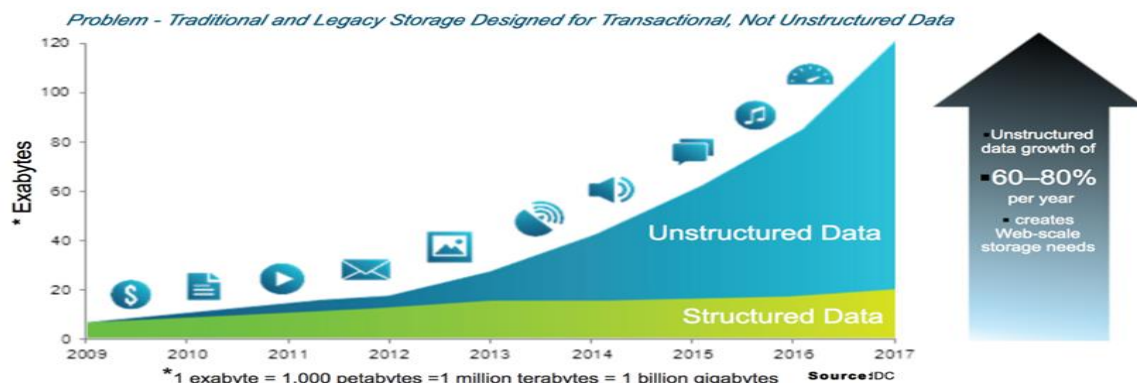| Index number	Name	DOB	House	Address	Contact No
001	Nabeel	2001.03.22		Gampaha	077-3929477
002	Sathsara	2001.11.02		Nawala	071-3925477
003	Dayana	2001.05.12		Madagama	072-3969477
004	Sahan	2001.05.21		Udagama	077-3929487
005	Chamath	2001.12.15		Ja-Ela	071-3929577
006	Anuga	2001.11.18		Negambo	076-3923477

Rows (පේළි)

Columns (තීරු)

Unstructured Data

Unstructured data is all those things that can't be so readily classified and fit into a neat box: photos and graphic images, videos, streaming instrument data, webpages, PDF files, PowerPoint presentations, emails, blog entries, wikis and word processing documents.

Unstructured data types

| Text files and documents | Server, website and application logs | Sensor data | Images |
| Video files | Audio files | Emails | Social media data |

Data Growth

Problem – Traditional and Legacy Storage Designed for Transactional, Not Unstructured Data

Unstructured data growth of 60—80% per year creates Web-scale storage needs

*1 exabyte = 1,000 petabytes =1 million terabytes = 1 billion gigabytes   Source:IDC

	Structured Data	**Unstructured Data**
Characteristics	• Pre-defined data models • Usually text only • Easy to search	• No pre-defined data model • May be text, images, sound, video or other formats • Difficult to search
Resides in	• Relational databases • Data warehouses	• Applications • NoSQL databases • Data warehouses • Data lakes
Generated by	Humans or machines	Humans or machines
Typical applications	• Airline reservation systems • Inventory control • CRM systems • ERP systems	• Word processing • Presentation software • Email clients • Tools for viewing or editing media
Examples	• Dates • Phone numbers • Social security numbers • Credit card numbers • Customer names • Addresses • Product names and numbers • Transaction information	• Text files • Reports • Email messages • Audio files • Video files • Images • Surveillance imagery

Define a database

- Did you use

  - Google to search for required information on the Web?
  - A credit card to buy something?
  - Your bank card to withdraw money from an ATM?
  - Library catalogues to look for books?

- Are these transactions possible without a database?

Virtually all modern business systems rely on databases.

There are some unique problems if not using the database.

❖ **Size of Data**: The small amount of data storing into spreadsheet is fine, however it might turns into a large amount of data then Spreadsheet solution will not work. Even if the size of data records goes into millions then storing data in multiple spreadsheet which will create a problem of speed. It will take you long time to find a record from the multiple spreadsheet files.

❖ **Ease of Updating Data**: Multiple peoples cannot edit the same file on same time. Other peoples must wait until files are available to update which results into wastage of time.

❖ **Accuracy**: When user doing data entry in files then it might be possible to incorrect data due to no validation present like you can enter wrong spelling, wrong dates, and wrong amount. So the Data accuracy is hard to maintain and accuracy is in question.

❖ **Security**: You cannot secure the data in the text files and spreadsheet. Anyone can access the file and read any data present in the file. So storing data will not work with banking, healthcare application, payroll department where privacy is difficult to maintain.

❖ **Redundancy**: The duplication of data can be possible using text files or spreadsheet. Chances of adding multiple copies of data cannot be limited here. This will leads to accuracy issues. Maintaining and updating multiple copies is not an easy task.

❖ **Incomplete Data**: Some of the data is not considered not important, so such data not entered in the file as no validation in place which leads the data integrity is in question.

To prevent above problem associated with storing data in the text file or spreadsheet the database is required.

File Based Systems Vs Databases	
File	**Database**
• Data scattered every where	• Data is integrated
• Sometimes different formats are used	• Data is stored in a common format for all applications to use
• Same data may be stored more than once (duplicated)	• Data duplication is reduced
	• Data can be used simultaneously by multiple users

What is a Database?

Database is a specialized structure that allow computer-based systems to **store, manage** and **retrieve** data very quickly.

A shared, integrated computer structure that stores a collection of:

•**End user data** (raw facts)

•**Meta-data** (data about data)

What is a DBMS?

•DBMS = **Database Management System**

DBMS is a collection of programs that,

- **manage** database structures
- **control** access to data
- **facilitate the sharing** of data among multiple users and applications
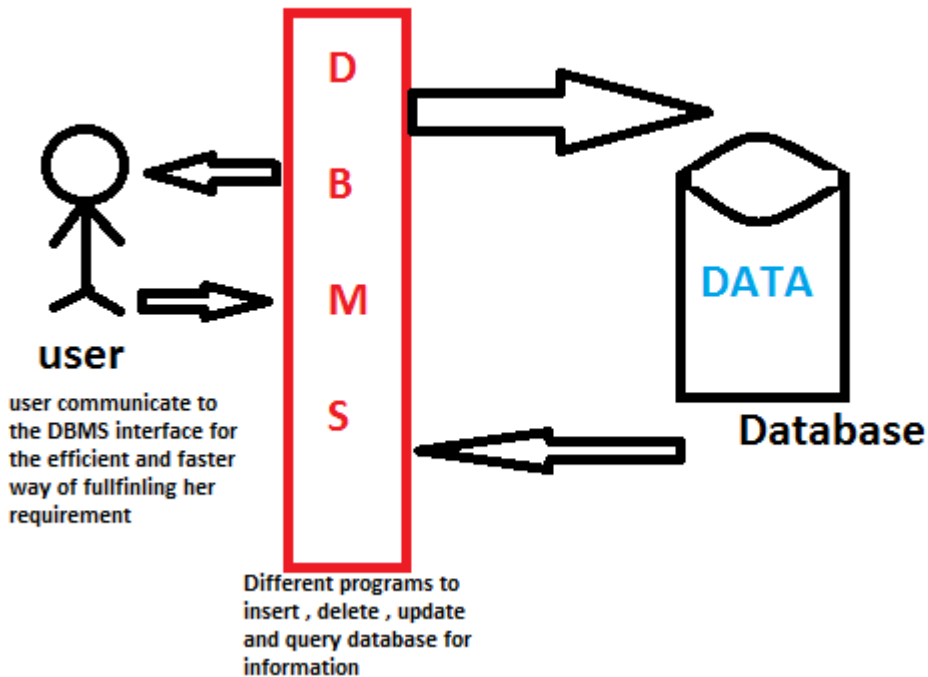- Enable **efficient and effective** data management.

•DB ~ e-filing cabinet

•DBMS helps manage the cabinet's contents
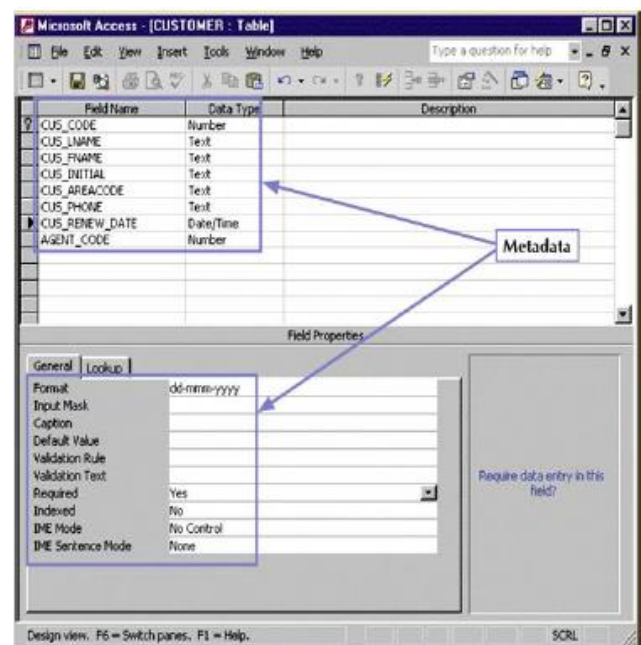
Benefits of a DBMS

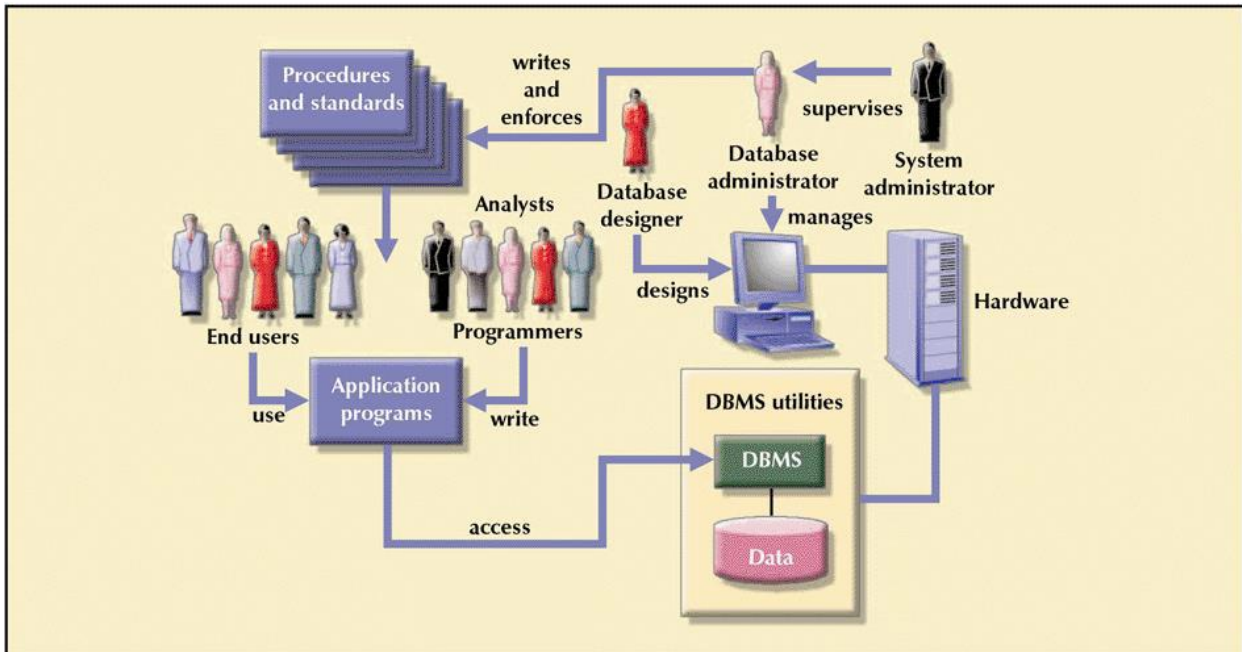The DBMS manages the interaction between the end user and the database.

Benefits of a DBMS

•End users will have **better access** to data

•How is this done?

- provides a **consistent view** to data and operations
- enables **ad-hoc queries**
- reduces data **inconsistencies and errors**

Database Environment

Database Careers

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	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.

Database models

- Flat File Systems/ Flat File Systems
- Hierarchical Model
- Network Model
- Relational Model
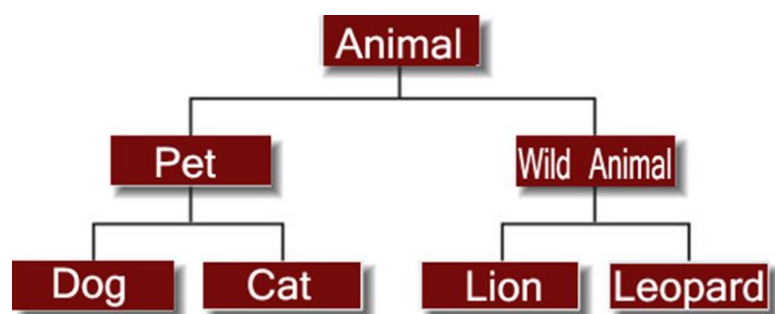- Object Relational Model

Flat File Systems

Flat data model is the first and foremost introduced model and in this all the data used is kept in the same plane. Since it was used earlier this model was not so scientific.

Ex: Excel, Calc

ID	Title	First name	Surname	Address	City	Postcode	Telephone
1	Mr	Tom	Smith	42 Mill Street	London	WE13GW	010344044
2	Mrs	Sandra	Jones	10 Low Lane	Hull	HU237HJ	022344033
2	Mr	John	Jones	10 Low Lane	Hull	HU237HJ	022344033

Hierarchical Model

Hierarchical model has one parent entity with several children entity but at the top we should have only one entity called root. For example, department is the parent entity called root and it has several children entities like students, professors and many more.

Network Model

Network model has the entities which are organized in a graphical representation and some entities in the graph can be accessed through several paths.
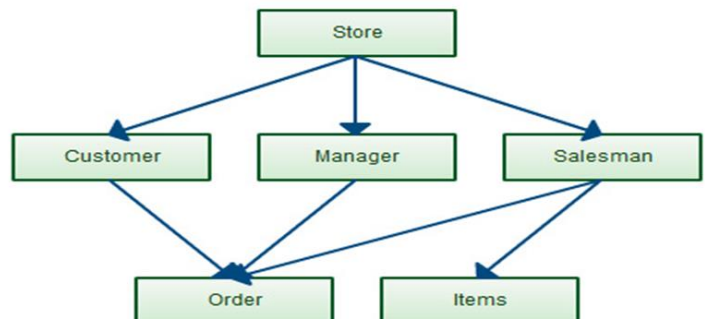
Object Relational Model

Object oriented data model is one of the developed data model and this can hold the audio, video and graphic files. These consist of data piece and the methods which are the DBMS instructions.

Relational Model

Relational model is the most popular model and the most extensively used model. In this model the data can be stored in the tables and this storing is called as relation, the relations can be normalized and the normalized relation values are called atomic values. Each row in a relation contains unique value and it is called as tuple, each column contains value from same domain and it is called as attribute.

An example of the relational model representing a university

Comparison of database models

Criteria	Flat file system	Hierarchical model	Network model	Relational model	Object relational model
Duration	1950 - 1960	1960 -1970 -	1960 - 1970	1970 - to date	1980 – to date
Short Name	DBF	DBMS	DBMS	RDBMS	ODBMS
Used computer programming languages	Assembler, Fortran, COBOL	COBOL, PL1, Fortran	COBOL, PL1, Fortran	SQL, ODBC	Java, C++, Pascal, Python