DEPARTMENT OF INFORMATION TECHNOLOGY

FACULTY OF MANAGEMENT STUDIES AND COMMERCE

UNIVERSITY OF SRI JAYEWARDENEPURA

ITC 1370
Information Technology for Business

Chapter 04
Data and Databases
Summary

Learning Objectives

Upon successful completion of this chapter, you will be able to:

- Understand the role of data in Information Systems

- Understand the concepts of database and database management systems

- Discuss concepts of Business Intelligence

Data, Information, and Knowledge

Data → Information → Knowledge → Wisdom

- **Data** is raw facts
  - Quantitative – numeric (Rainfall values 100mm, 120mm: Temperature 25C, 27C)
  - Qualitative – descriptive (Customer Satisfaction: Excellent, Good, Bad)
- **Information** is when data is given context and more specific
  - When data is processed to aid in decision making
- **Knowledge** is developed when information has been aggregated and analyzed to make decisions, set policies, and spark innovation
- **Wisdom** is the combination of knowledge and experience
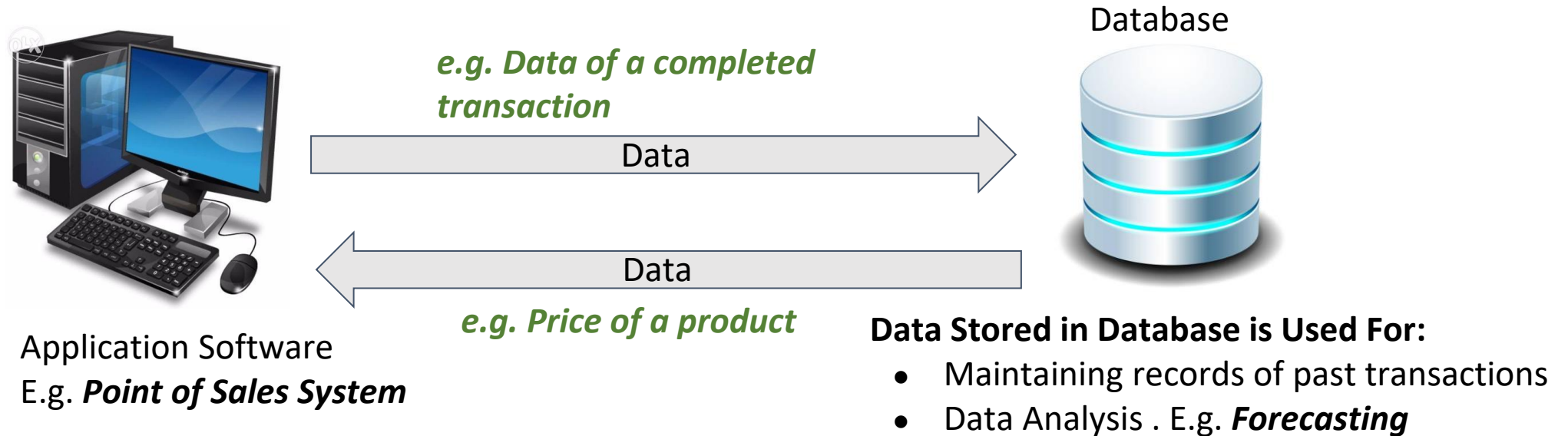 - May take years to develop

Information Systems, Data and Databases

Background of Data and Information Systems

- Behind most of the Information Systems, there is a database to store data persistently (in a structured manner).

- This data is going to be used as (historical) data for future predictions, identifying valuable patterns for prospective sales opportunities, and government's legislative requirements (to be used in external audits) etc…

- Therefore, the manager should possess the skills in understanding, what databases are, how the databases are designed and on how to use them for storing and querying data as well.

Data and Information Systems

e.g. Data of a completed transaction

Data →

Data ←

e.g. Price of a product

Application Software
E.g. **Point of Sales System**

Database

Data Stored in Database is Used For:
- Maintaining records of past transactions
- Data Analysis . E.g. **Forecasting**

Revisiting Previous Lesson

How does a general-purpose software store data?

- General purpose application software stores data in files.

- For example, a letter, a spreadsheet, a schedule, a song, a video, and so on

How does a special purpose software (used in information systems) store data?

In the past, even specialized software that used to use information systems used standard data files to store data.

An information system can have many programs to carry out specific activities. For example, think of software and data in a university system.
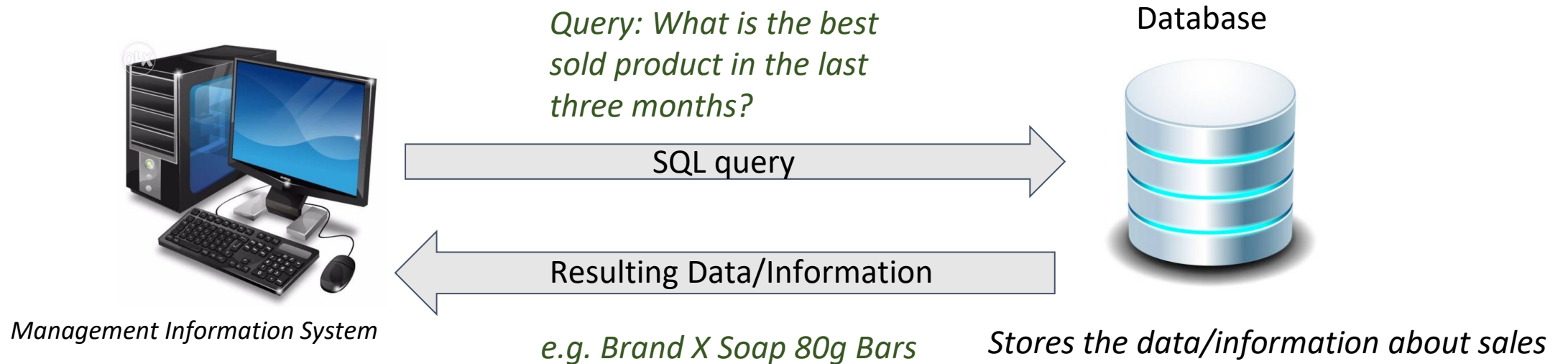
- There are software programs that perform specific activities for students.

- General Administration Branch = Registration Software

- Library = library software

- Medical Center = Health Reporting Software

- Welfare Branch = Welfare Software

- Exam Branch = Exam Software

- Faculty = Faculty Student Software

Information systems use data stored in computer databases to provide needed information.

- Information systems capture data from the organization (internal data) and its environment (external data).
- They store the data items over an extensive period of time in databases.
- When specific information is needed, the appropriate data stores are queried as necessary, and the user receives the resulting information.
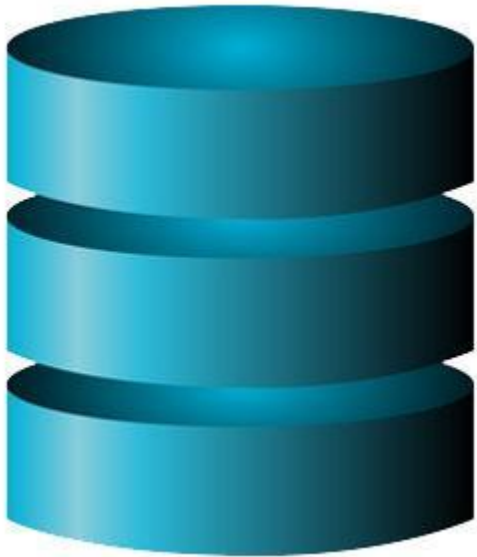
Data Access/Retrieval

Query: What is the best sold product in the last three months?

SQL query →

← Resulting Data/Information

Management Information System

e.g. Brand X Soap 80g Bars

Database

Stores the data/information about sales

What is a database and why use them?

What is a Database?

- A database is an organized collection of logically related information, or data, typically stored electronically in a computer system.

Data is organized into rows, columns and tables to make it easier to find relevant information.

Data gets updated, expanded and deleted as new information is added.

Databases process workloads to create and update themselves, querying the data they contain and running applications against it.

Basic concepts

Files:

A collection of interrelated records

Records:

A collection of fields.

Fields:

Fields formed by the merging of several characters. Ex: Name, Date of Birth

Character (Byte):

This can be a letter, a number, or a special character. This is made up of bits. Ex: A, D, 1

Bit: 0 or 1

Hierarchy	Example
Database	**Employee Database**
File	**Employee Details File**
Record	**Employee Record**
Field	**Employee Name Field**
Byte	01001010 (Letter J in ASCII)
Bit	0

Employee Database

Employee Details File   Training Records File

Salary File

Employee Details File

EMP_NAME	JOB TITLE	DATE EMPLOYED
Alice Carter	Lecturer	31 Mar 2002
Faridah bte Hassan	Sales Manager	9 Aug 2013
Jeffrey Tan	Lecturer	19 Sep 2004
Steve Willis	HR Manager	23 Dec 2005

Employee Record

EMP_NAME	JOB TITLE	DATE EMPLOYED
Jeffrey Tan	Lecturer	19 Sep 2004

Employee Name Field

EMP_NAME
Jeffrey Tan

Before the invention of databases, what was the technique used for storing data?

Flat File Structure

A structure that stores data in a plain text file..

```
patient_id,name,dob,gender,phone,doctor_id,doctor,room
134,Jeff,4-Jul-1993,Male,7876453,01,Dr Hyde,03
178,David,8-Feb-1987,Male,8635467,02,Dr Jekyll,06
....
```

We could also store the data in a spreadsheet which is also a flat file. e.g **Microsoft Excel**

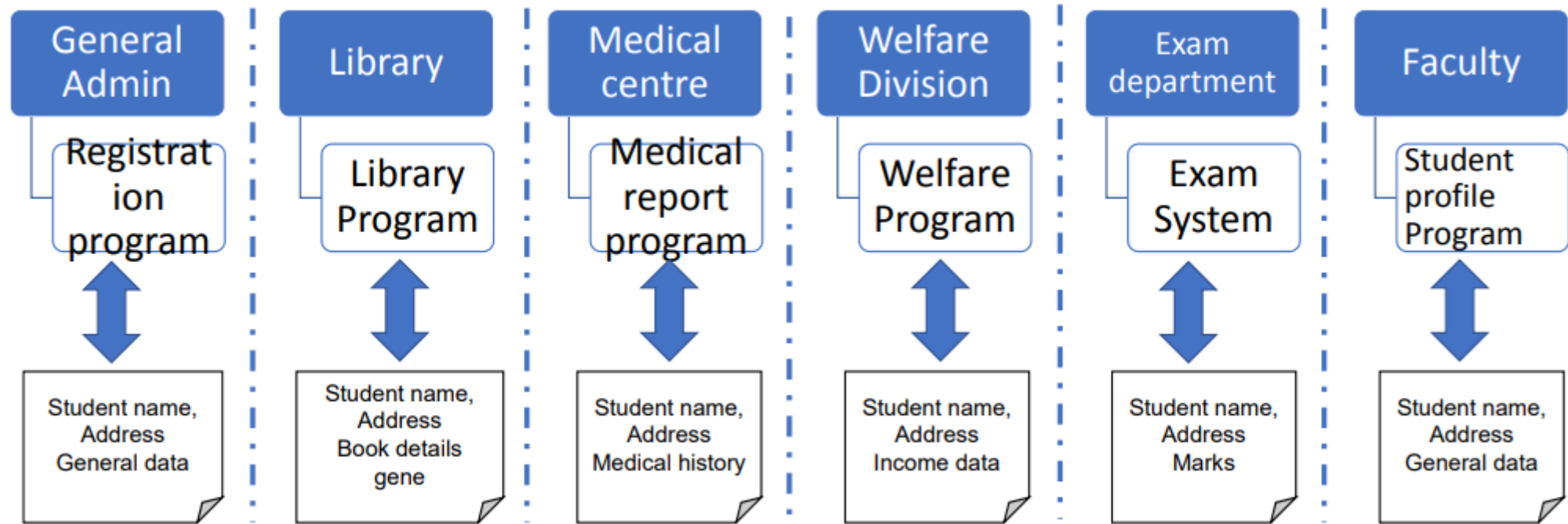| Patient Id	Name	D.o.B	Gender	Phone	Doctor Id	Doctor	Room
134	Jeff	4-Jul-1993	Male	7876453	01	Dr Hyde	03
178	David	8-Feb-1987	Male	8635467	02	Dr Jekyll	06
198	Lisa	18-Dec-1979	Female	7498735	01	Dr Hyde	03
210	Frank	29-Apr-1983	Male	7943521	01	Dr Hyde	03
258	Rachel	8-Feb-1987	Female	8367242	02	Dr Jekyll	06

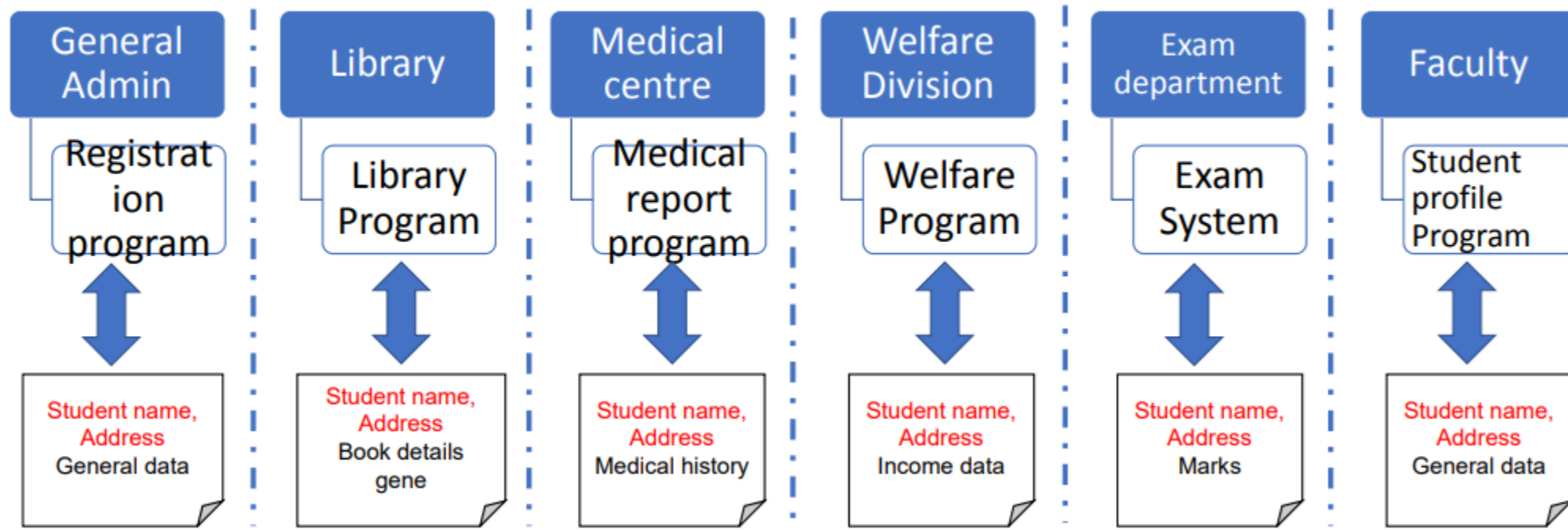# Sample scenario that uses flat files for data storage

Flat File System

- One or more separate (no data transfer) files for each software / program

- These files are created, edited and deleted only by a special program

General Admin	Library	Medical centre	Welfare Division	Exam department	Faculty
Registration program	Library Program	Medical report program	Welfare Program	Exam System	Student profile Program
Student name, Address General data	Student name, Address Book details gene	Student name, Address Medical history	Student name, Address Income data	Student name, Address Marks	Student name, Address General data

Flat File System
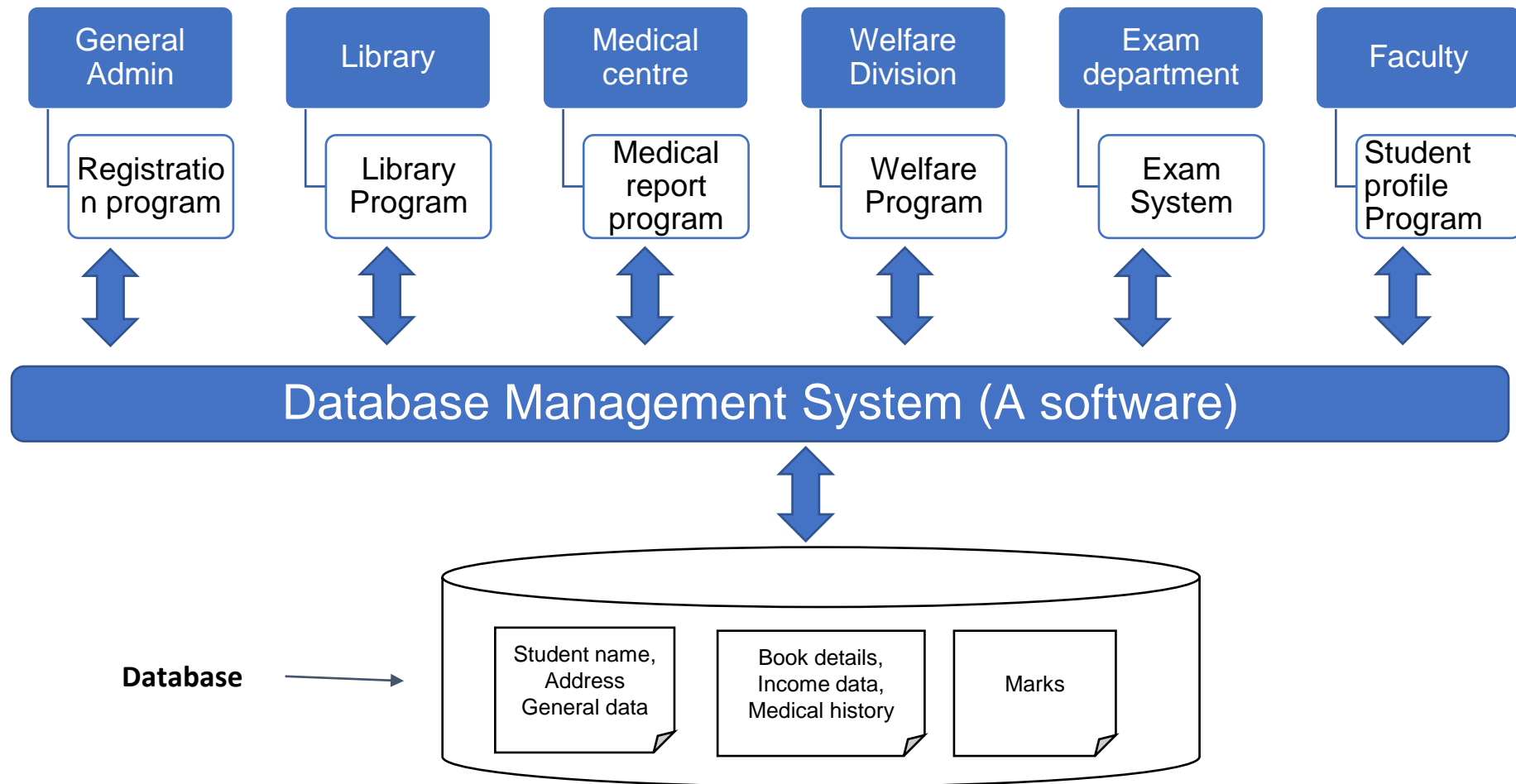
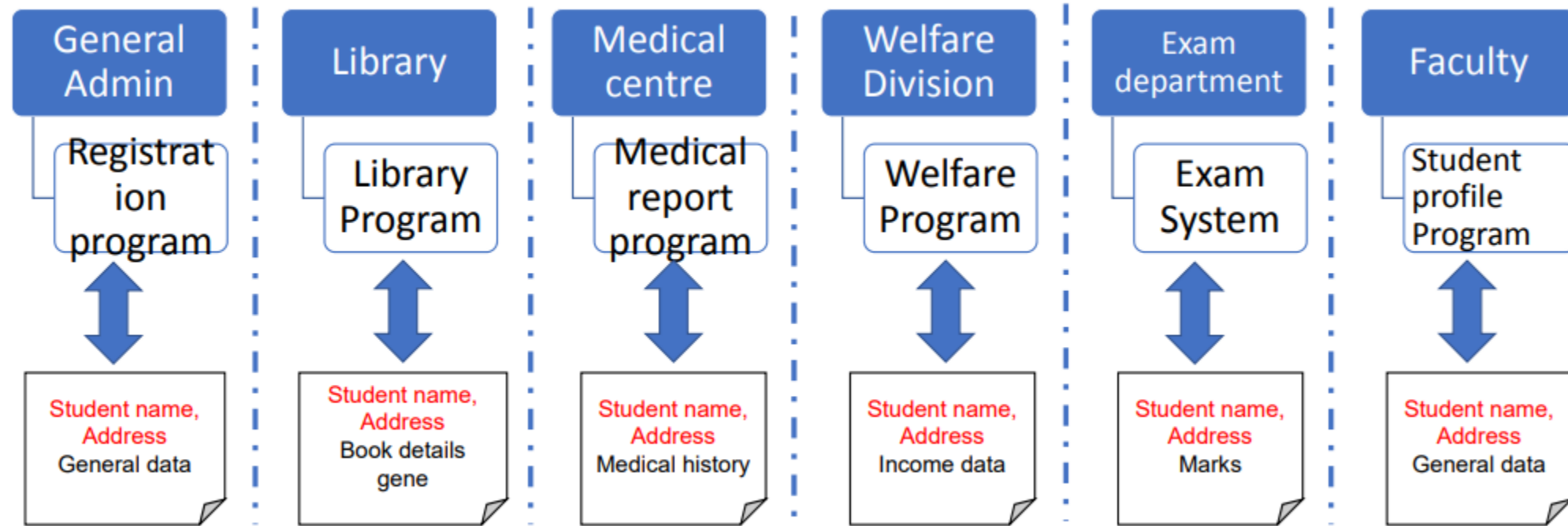- The issues with flat file system…

Disadvantages of Flat File

- Data Redundancy

- Data Integrity Issues

- Inefficient (May have to enter/update data multiple times – wastes time)

- High maintenance cost

- Sharing is coarse

- Weak Security

- Application dependent

We can use databases to overcome these issues, encountered while using flat files.

Compare with database

How is data stored in databases?

- The way the data is stored in databases will depend on the database model.

- There are different database models

- The most popular database model is Relational Database Model

  - In a relational database model, the data are stored in tables which are related to each other using a common field.

Relational Database

- Data is organized into tables (or relations).
- Each table has a set of fields which define the structure of the data stored in the table.
- Data from several tables are tied together (related) using a field that the tables have in common.
- A record is one instance of a set of fields in a table. Think of the records as the rows (or tuple) of the table and the fields as the columns of the table.

Student ID	First name	Last name
52-743965	Charles	Peters
48-209689	Anthony	Sondrup
14-204968	Rebecca	Phillips

ProviderID	Provider name
156-983	UnitedHealth
146-823	Blue Shield
447-784	Carefirst Inc.

Student ID	ProviderID	Type of plan	Start date
52-743965	156-983	HSA	04/01/2016
48-209689	146-823	HMO	12/01/2015
14-204968	447-784	HSA	03/14/2016

A Sample Table in a Relational Database

Different types of database model

- Databases can be organized in many ways by using different models.

- The data model of a database is the logical structure of data items and their relationships.

- There have been several data models. Since 1980s, the **relational data model** has been popular.

History and Evolution of database models

MODEL	FLAT FILES	HIERARCHICAL	NETWORK	RELATIONAL	OBJECT-ORIENTED	MULTI-DIMENSIONAL
YEAR BEGAN	1940s	1960s	1960s	1970s	1980s	1990s
DATA ORGANIZATION	Flat files	Trees	Trees	Tables and relations	Objects	Data cubes, tables and relations, or a combination

Visualization of different data models

Hierarchical model

	(Living) Great apes	

Orangutans — Gorillas — Chimpanzees — Humans

Bornean orangutan — Sumatran orangutan — Eastern gorilla — Western gorilla — Common chimpanzee — Bonobo — Modern humans

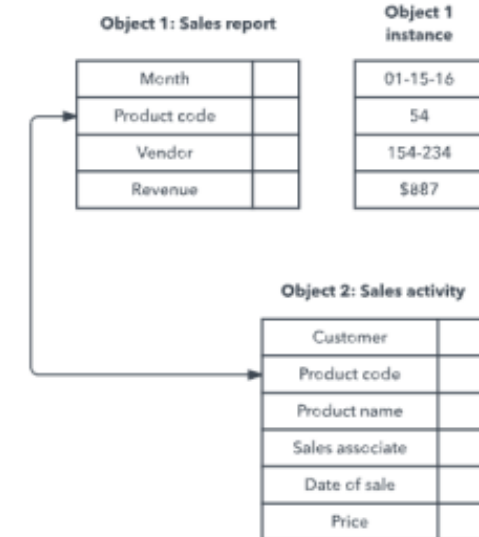**Network model**

Vendor

Employee — Customer — Product

Sales transaction

Object-oriented database

Object 1: Sales report

Month	
Product code	
Vendor	
Revenue	

Object 1 instance

01-15-16
54
154-234
$887

Object 2: Sales activity

Customer	
Product code	
Product name	
Sales associate	
Date of sale	
Price	

Relational model

Student ID	First name	Last name
52-743965	Charles	Peters
48-209689	Anthony	Sondrup
14-204968	Rebecca	Phillips

ProviderID	Provider name
156-983	UnitedHealth
146-823	Blue Shield
447-784	Carefirst Inc.

| Student ID	ProviderID	Type of plan	Start date
52-743965	156-983	HSA	04/01/2016
48-209689	146-823	HMO	12/01/2015
14-204968	447-784	HSA	03/14/2016

and more……

©

How can we create and manage databases?
Is there a software to do that?

What is a DBMS?

- A software for creating and managing databases.

- Used to create, maintain, and access computer databases.

- PC DBMSs include:
  - Microsoft Access, Corel Paradox, Lotus Approach

- For more comprehensive enterprise databases:
  - Oracle Database, IBM DB2
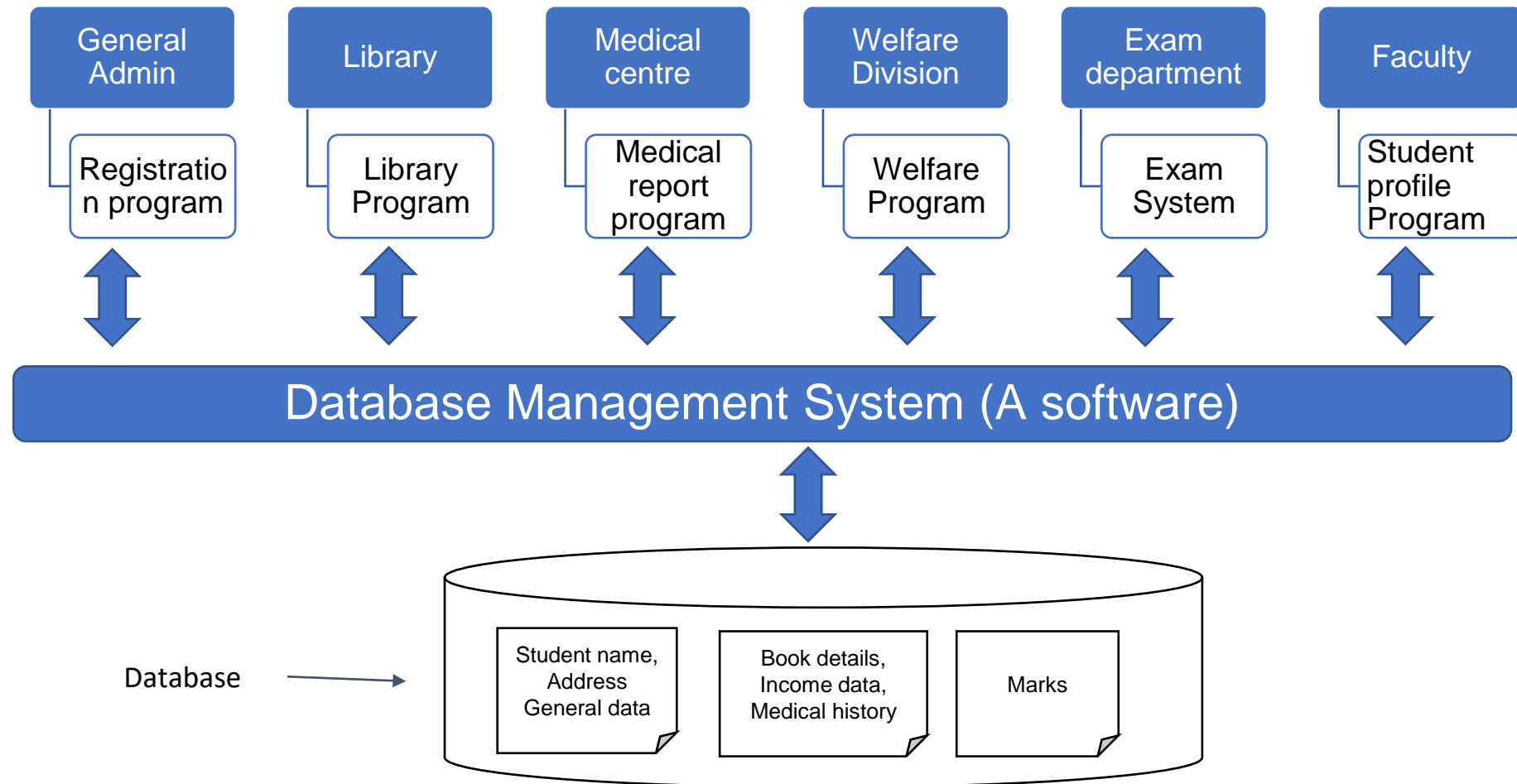
How to Implement Databases?

Designing Databases

- Design is a critical <u>first step</u> in creating a database
  - Understand the goal of how the data of the database will be used.
  - Identify the data needed as part of accomplishing this goal.
  - Identify how the data is related to each other.
  - Identify tables and fields to organize the data.

Overview of Big data and Business Analytics

Big data

- The term refers to such massively large data sets that conventional data processing technologies do not have sufficient power to analyze them.

  - For example, Walmart must process millions customer transactions every hour across the world.
  - Presidential candidates of United states might have to analyze their percentage of supporters and opinions using data in different social media.

- Storing and analyzing that much data is beyond the power of traditional data management tools.

- Understanding and developing the best tools and techniques to manage and analyze these large data sets are a problem that governments and businesses alike are trying to solve.

Business Analytics

Business analytics involves the use of data analysis tools, statistical methods, and predictive modeling to extract valuable insights from business data, enabling informed decision-making and strategic planning.

Source: https://techvidvan.com/tutorials/big-data-applications-across-industries/

Advantages of Business Analytics

- ## Understanding target customers better
  - Big data is used by business today for analyzing sentiments of the target customers and providing them better services to increase the business.

- ## Cutting down in expenditures in various sectors
  - Analysis of such huge volume of data has also helped business in cutting down their expenditures in various sectors wherever possible. Several billions of dollars being saved by improvements in operational efficiency and more.

- ## Increase in operating margins in different sectors
  - Big Data also helps industries in increasing operating margins in different sectors. With the help of Big Data, lot of manual labour can be converted into machine task and this helps in increasing operating margins.

Techniques for Business analytics

- Some of the techniques based on the nature of data
  - Data mining
    - Extracts patterns from large data sets by combining methods from statistics and machine learning, within database management.
  - Machine learning
    - Works with computer algorithms to produce intelligent output based on data
  - Natural language processing
    - The application of computational techniques to the analysis and synthesis of natural language and speech.
  - Statistical Analysis
    - This technique works to collect, organise, and interpret data, within surveys and experiments

    and there are many other techniques…

Where should this data be stored before being taken into analysis

Data Warehouse

- Consists of extracts from one or more of the organization's operational databases

- Allows the data to be copied and stored for analysis
  - Needs to be refreshed as the data changes

- Data is time-stamped when extracted
  - Allows comparisons between different time periods

- Data is standardized
  - All similar fields (e.g., calendar dates) are structured the same
    - Date is MM/DD/YYYY

- Data marts are smaller subsets of data warehouses for specific business problems

ETL: EXTRACT, TRANSFORM, LOAD

Data sources

Databases

CRM/ERP

Web events, etc.

Extract →

Transform

Staging area
(Raw data is converted into a fitting form for a DW)

Load →

Data Warehouse

Prepared data
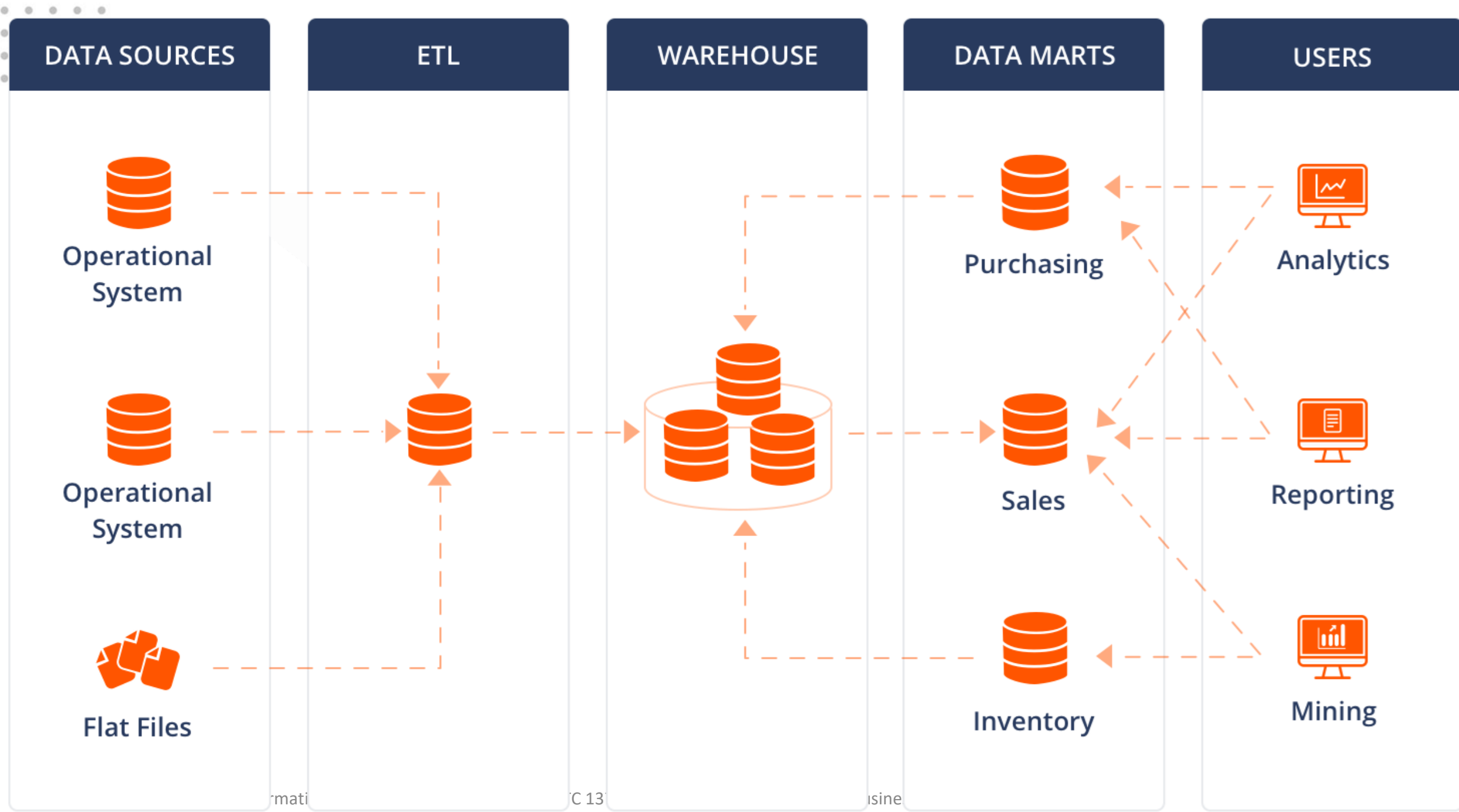
Transmit →

BI Tools

Analytics

Data Warehouse Benefits

- Allows organizations to better understand the data

- Centralized view of data to identify inconsistent data

- Once inconsistencies are resolved, higher quality data is used to make better business decisions

- Data can be analyzed over multiple time periods

- Tools are available to combine data and gain more insight into business operations

What is a data mart?
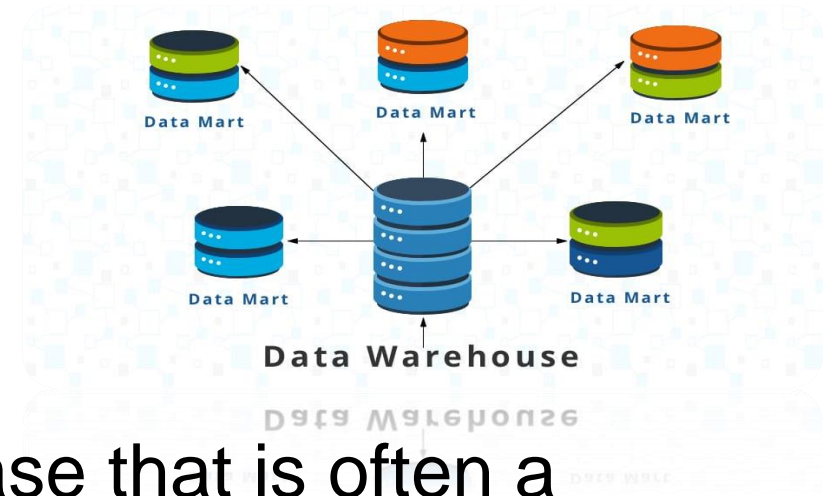
- A data mart is a subject-oriented database that is often a partitioned segment of an enterprise data warehouse.

- The subset of data held in a data mart typically aligns with a particular business unit like sales, finance, or marketing.

- Data marts accelerate business processes by allowing access to relevant information in a data warehouse or operational data store within days, as opposed to months or longer.

- Because a data mart only contains the data applicable to a certain business area, it is a cost-effective way to gain actionable insights quickly.

DATA SOURCES	ETL	WAREHOUSE	DATA MARTS	USERS

Operational System

Operational System

Flat Files

Purchasing

Sales

Inventory

Analytics

Reporting

Mining

ITC 2372: Business Analytics

Current business organizations expect their employees to master the science of analyzing data to find out patterns that will help them to develop business strategies. Analytics as a decision-making approach is been used by big corporations, governments, entrepreneurs, and almost everyone else to generate insights by unearthing patterns and decoding data.

By the time students finish the course, they should be able to,

·   Understand and explain how managers use business analytics to formulate and solve business problems and to support managerial decision-making.
·   Familiar with and conducting the processes needed to develop, report, and analyze business data.
·   Use and apply computer-based tools with business analytics to solve business problems

Summary

- Discussed the role of data in Information Systems

- Discussed the concepts of database and database management systems

- Discussed concepts of Business Intelligence

Activity 1

- Discuss the issues in storing data in flat file systems

Thank You