



ACS-1803

Introduction to Information Systems

Instructor: Trevor Nadeau

Data Management

Lecture Outline 2 - Part 1



Overview

- The importance of Data Management
- The file approach vs. the Database approach
- How Data is Categorized
- Database Design
- Relational Databases
- Data Modeling
- Schemas
- Database Management System
- Business Intelligence, Data Warehouses, Data Marts, and Data Mining

Database Technology

- A collection of related data organized in a way that makes it valuable and useful
- Allows organizations to retrieve, store and analyze information easily
- Is vital to an organization's success in running operations and making decisions

Definitions

- **Database:**
 - Organized collection of data
 - For the most part stored in electronic form
 - Data organized to model relevant aspects of reality
- **Database management system (DBMS):**
 - Group of programs that manipulate the database
 - Provide an interface between the database and its users and other application programs
- **Database administrator (DBA):**
 - Skilled IS professional who directs all activities related to an organization's database

Databases in Action



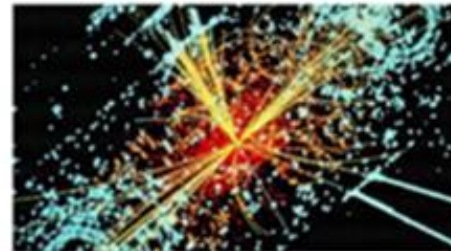
- Reservation systems
 - book flights from multiple airlines, hotel rooms etc.
 - e.g. Amadeus systems
 - Global Distribution System (GDS) founded by Lufthansa, Air France and other partners



- Banking and trading
 - customer data, account information, transactions, ...
 - e.g. London Stock Exchange
 - almost 1 million trades per day



- Embedded databases in cars, airplanes etc.
 - manage configurations and store sensor data
 - e.g. db4o object database used in BMW's Car IT system



- Scientific databases
 - sensor data, classifications (e.g. human genome) as well as data from simulations
 - e.g. LHC Computing Grid
 - LHC experiments at CERN
 - 15 petabytes of data per year



- Geographic Information Systems (GIS)
 - store raster (bitmap) or vector data representing real world objects
 - geospatial query language



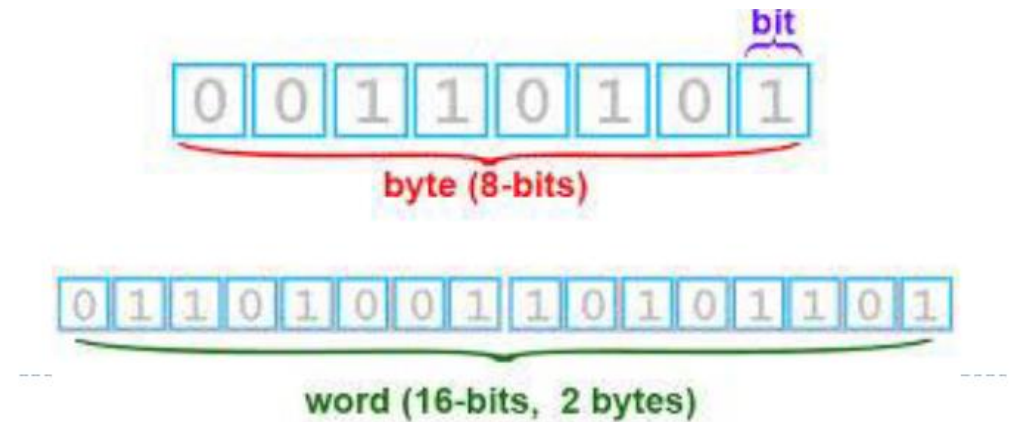
- Many everyday devices contain databases
 - TVs, washing machines, mobile phones, ...
 - e.g. Android phones with SQLite database

Data Management Concepts

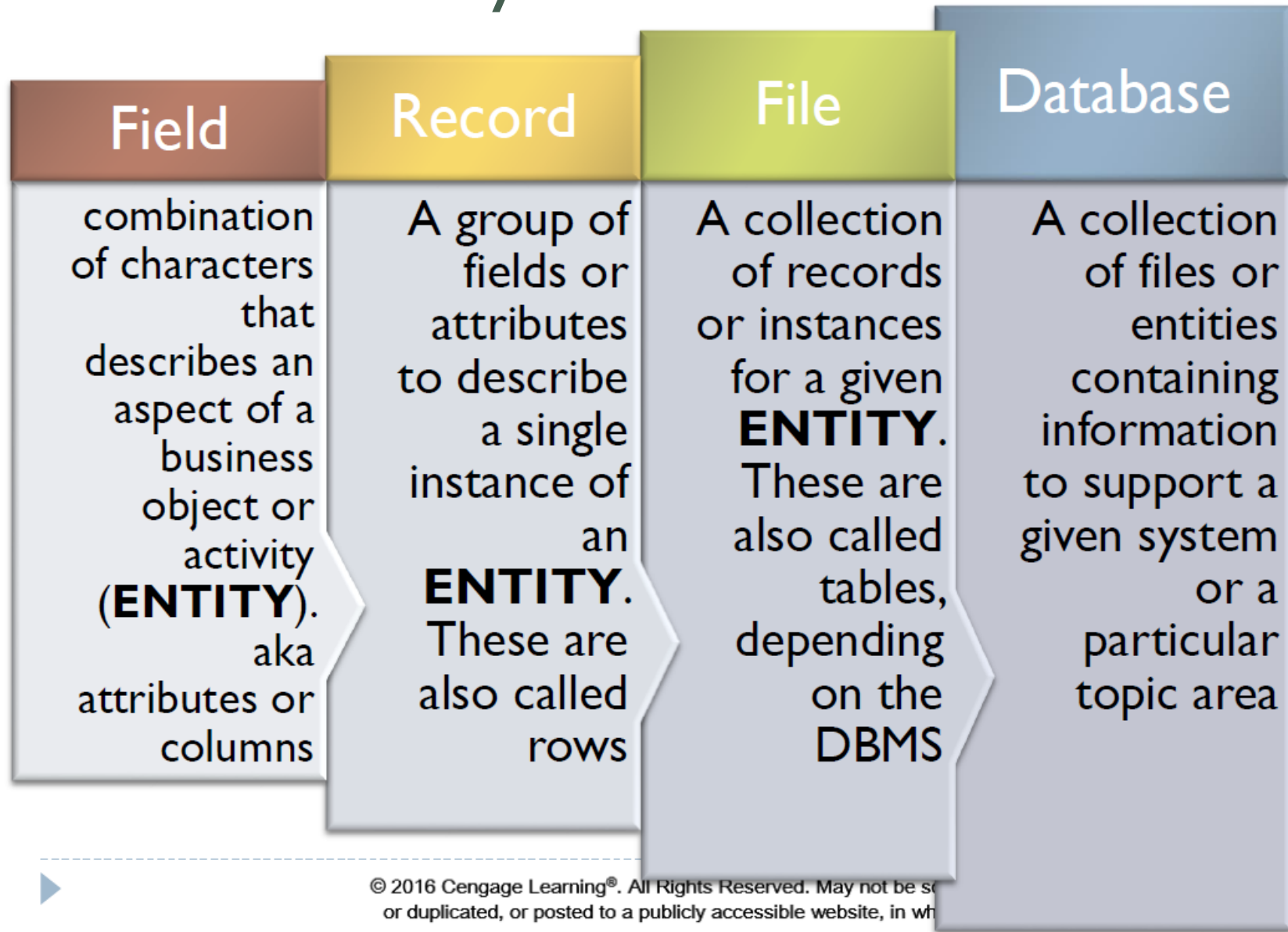
- **Why manage data?**
 - Without data and the ability to process the data: An organization could not successfully complete most business activities
- **Data consists of raw facts**
- **To transform data into useful information:**
 - It must first be organized in a meaningful way i.e. **Database**
- **Database Management System (DBMS)**
 - A Collection of programs that enables users to store, modify, and extract information from a database

The Hierarchy of Data

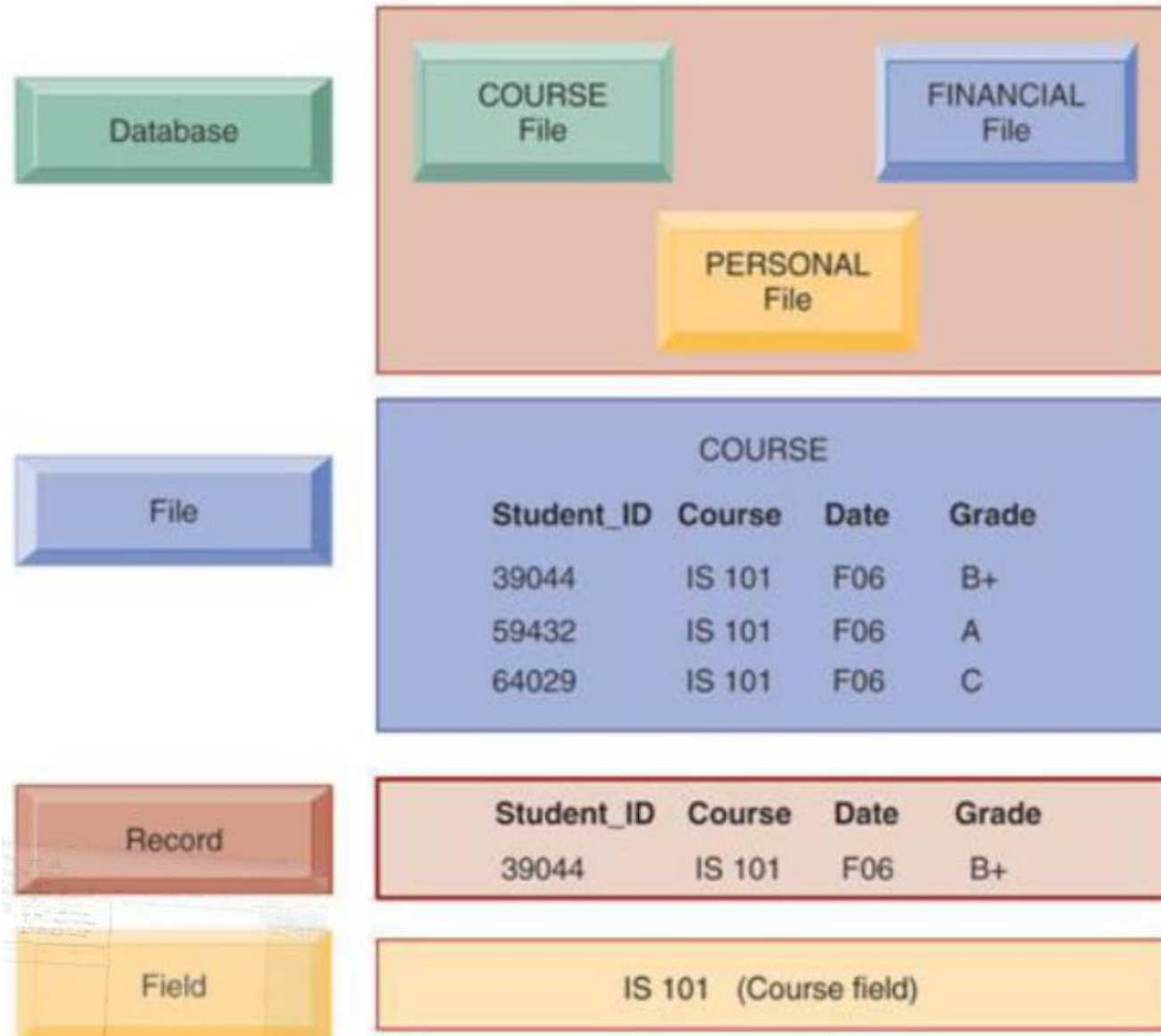
- Bit (a binary digit):
 - Circuit that is either on or off
 - Byte: Typically make up of eight bits
- Character:
 - Basic building block of information
- Field
 - Name, number, or combination of characters that describes an aspect of business object or activity



The Hierarchy of Data



The Hierarchy of Data



Levels of Data within a Database

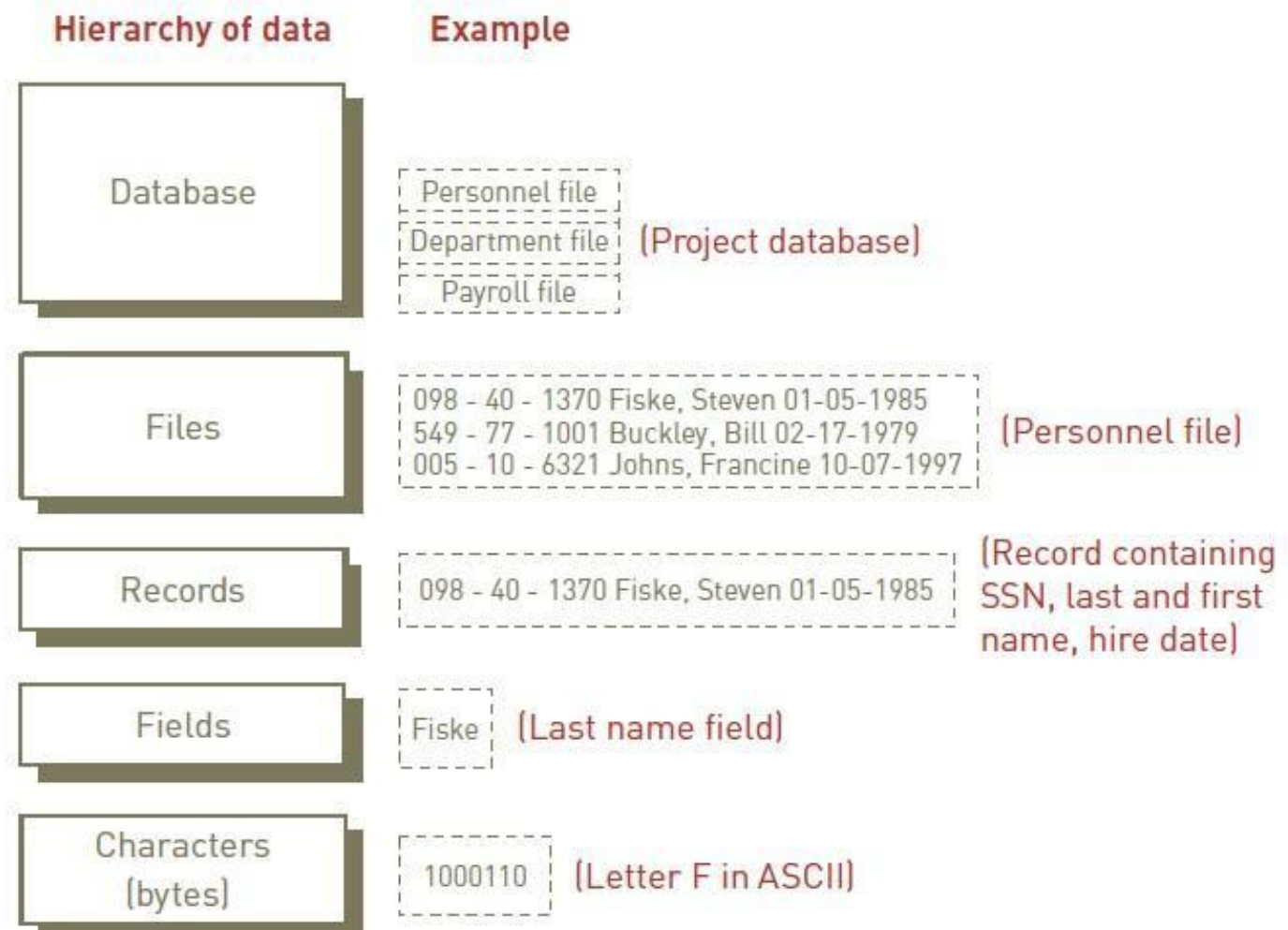
<u>Level</u>	<u>Term</u>	<u>Term Definitions</u>
Highest	Database	A collection of files or entities containing information to support a given system or a particular topic area
	File	A collection of records or instances for a given ENTITY. These are also called tables , depending on the DBMS
	Record	A group of fields or attributes to describe a single instance of an ENTITY. These are also called rows depending on the DBMS
	Field	Individual characteristics about an ENTITY. Fields are also called attributes or columns based on the type of DBMS
	Character	Basic building block of information that defines attributes.
Lowest	Bit	Binary digit that defines a character. Communicates with the circuit technology within a computer through on/off code switches.

Levels of Data within a Database

A database system organizes data in a hierarchy that starts with a data item.

Data items can be **characters**, **numbers**, or **symbols**. Data items can be grouped to form a **field**, or attribute. Related fields or attributes can be grouped to form a single instance of an entity or a **record**.

Related records or multiple instances of a given entity can be collected to form a **file**, or table. A collection of files/ tables or entities containing information to support a given system is organized into a **database**.



Data Management in a Computer

- See Supplementary Notes “Data Management in a Computer”



Data Entities, Attributes, and Items

- **Entity:**
 - Things we store information about. (i.e. **persons, places, objects, events**, etc.)
 - Have **relationships** to **other entities** (i.e. the entity *Student* has a relationship to the entity *Grades* in a University Student database)
 - General class of people, places, or things (objects) for which data is collected, stored, and maintained
- **Attribute:**
 - These are pieces of information (characteristics) about an entity (i.e. Student ID, Name, etc. for the entity *Student*)
- **Data item:**
 - Specific value of an attribute

Data Entities, Attributes, and Items

Employee #	Last name	First name	Hire date	Dept. number
005-10-6321	Johns	Francine	10-07-1997	257
549-77-1001	Buckley	Bill	02-17-1979	632
098-40-1370	Fiske	Steven	01-05-1985	598

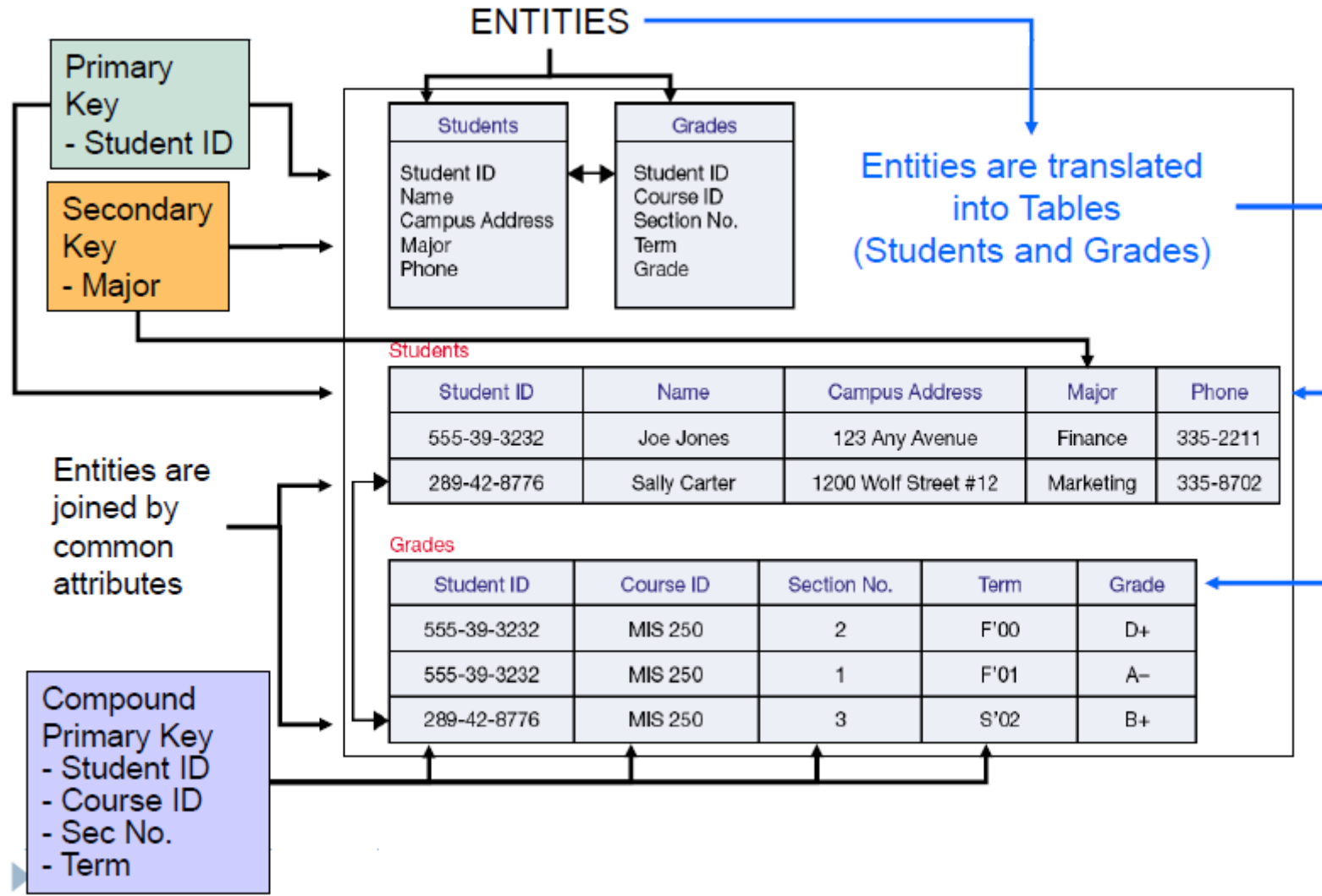
KEY FIELD

ATTRIBUTES (fields)

ENTITIES (records)

The Key field is the Employee Number. The attributes include Last Name, First Name, Hire Date and Department number

Data Entities, Attributes, Items, Keys



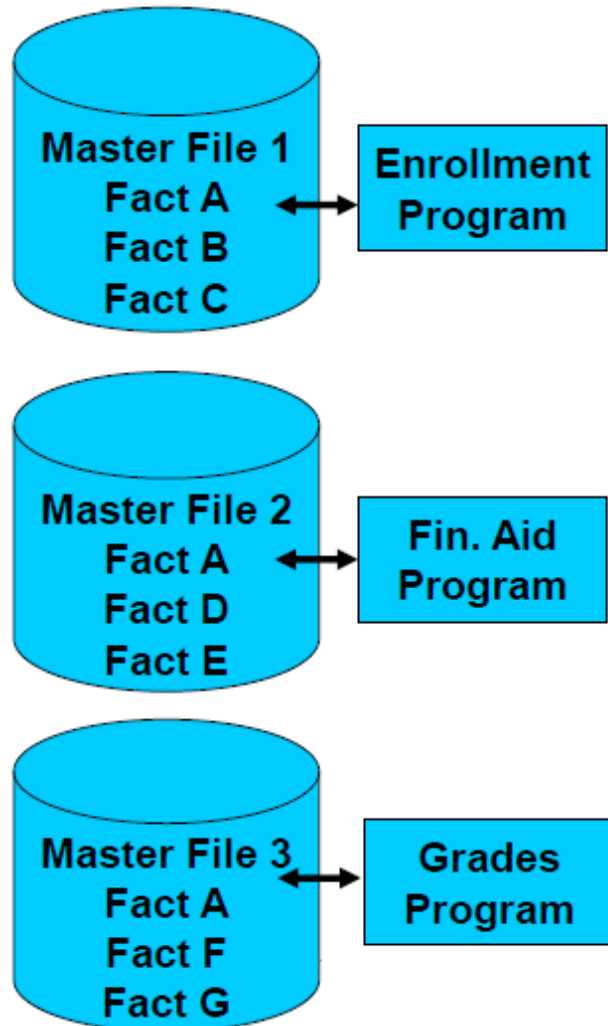
Defining Database – Keys

- **Database Keys**
 - Mechanisms used to identify, select, and maintain one or more records using an application program, query, or report
- **Primary Key**
 - A unique attribute type used to identify a single instance of an entity
- **Compound Primary Key**
 - A unique combination of attribute types used to identify a single instance of an entity
- **Secondary Key**
 - An attribute that can be used to identify one or more records within a table with a given value
- **Foreign Key**
 - An attribute that appears as a non-primary key in one entity (table) and as a primary key attribute in another entity (table)

Data Management and Database Management Systems

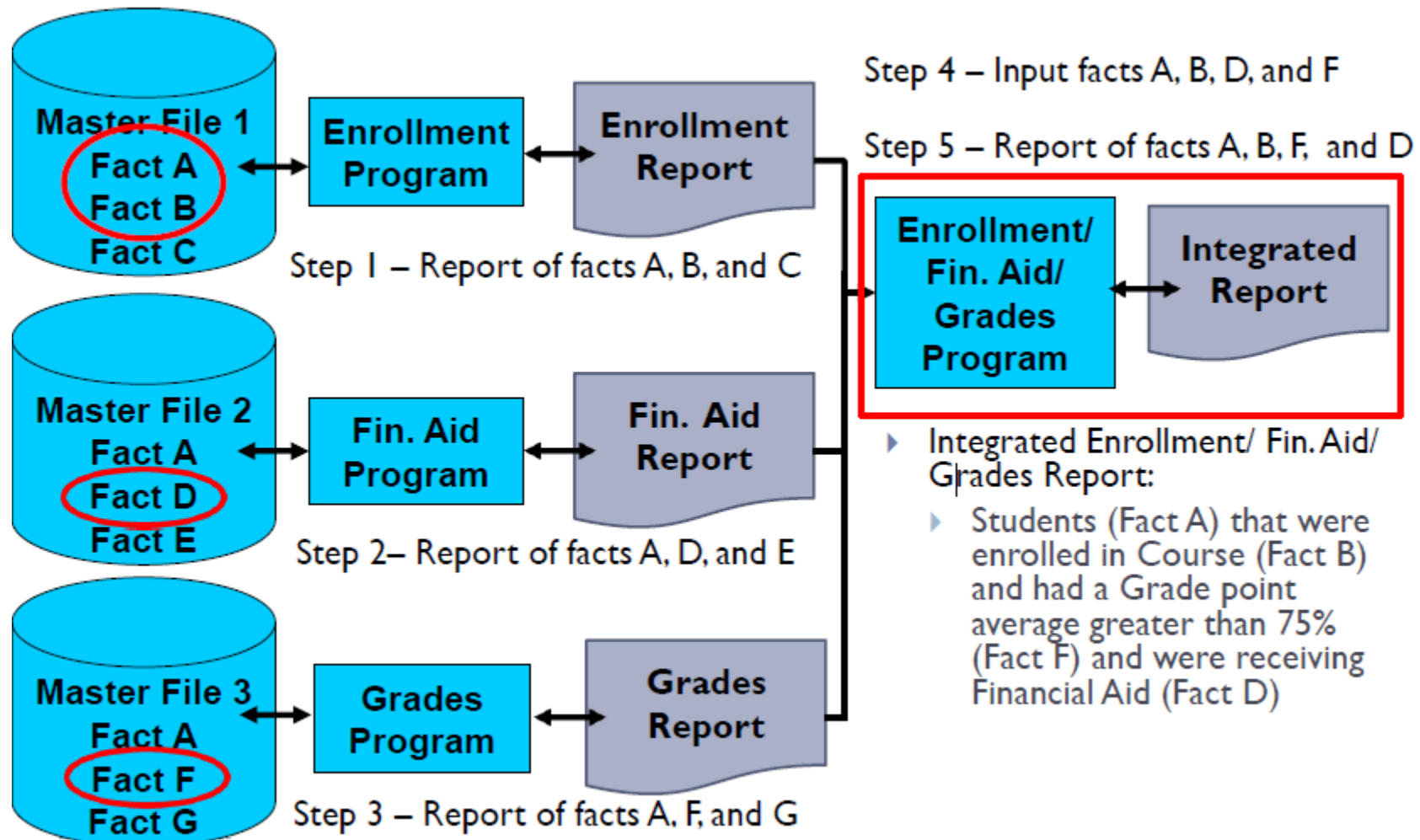
- **Database (Application) System** is the combination of the database, the DBMS, and the application programs that access the database
- **Management Information System (MIS)** is the database system coupled with a set of hardware, software, telecommunications, people, and procedures. Programs which are part of the information system then transform the raw data to useful information

File Approach to Storing Data

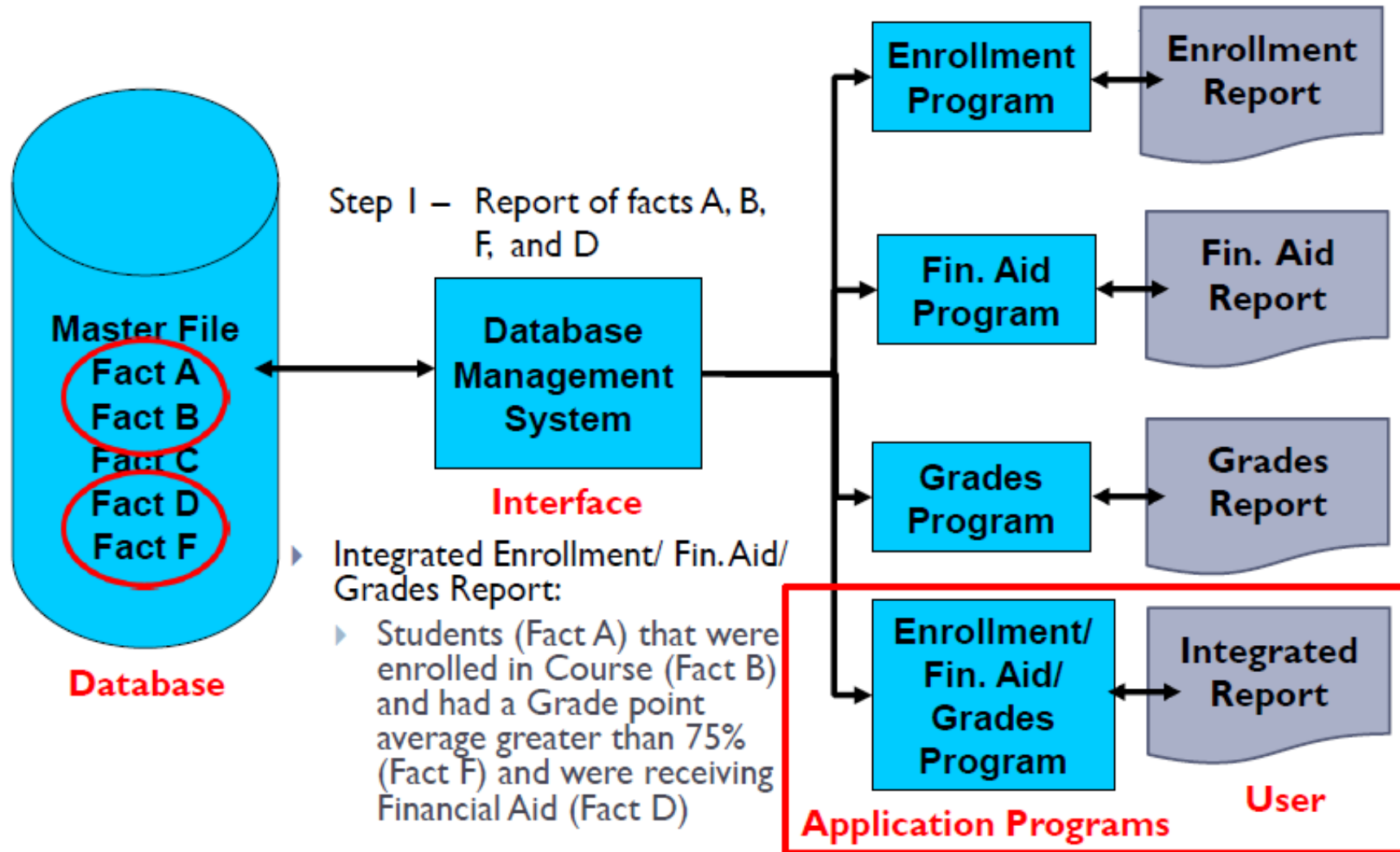


- ▶ This proliferation of master files created problems:
 - ▶ Often the same information was stored in multiple master files.
 - ▶ Made it more difficult to effectively integrate data and obtain an organization-wide view of the data.
 - ▶ Also, the same information may not have been consistent between files.
 - ▶ If a student changed his phone number, it may have been updated in one master file but not another.

File Approach to Storing Data



Database Approach to Storing Data



File Approach Characteristics

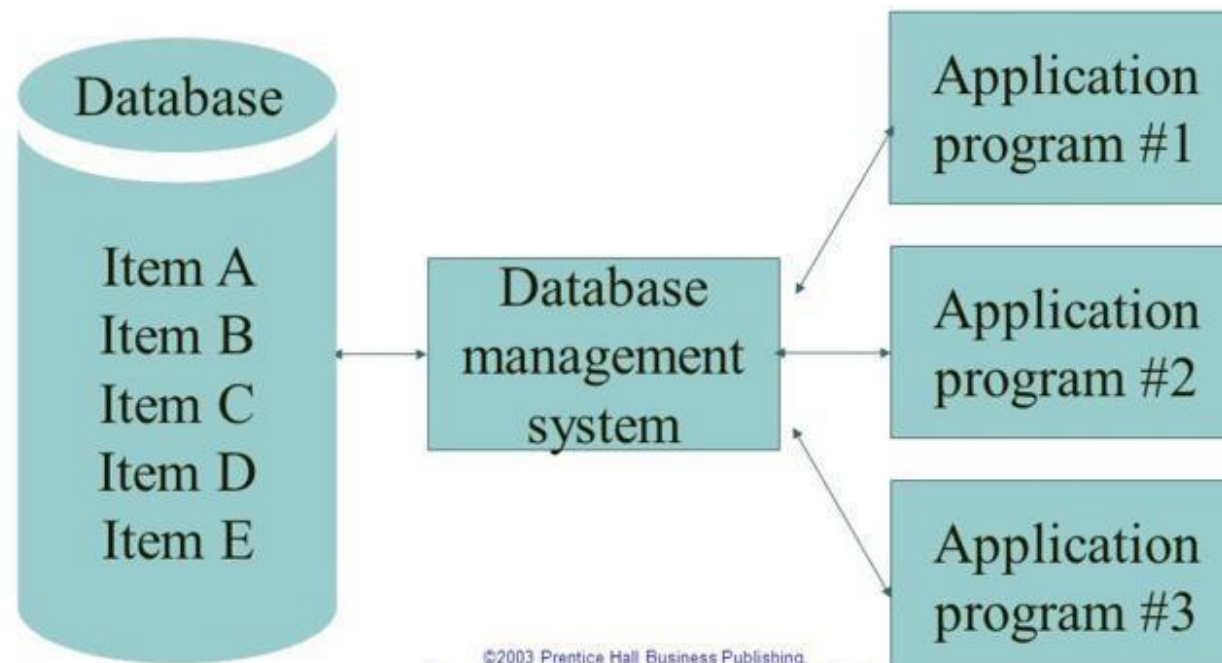
- Each application has its own set of files
- **Inflexibility:** committed to processing particular queries
- **Uncontrolled Redundancy:** If these separate applications need to process the same data there are duplicate copies of the data, waste of valuable storage space.
 - Need to input data to several files
 - Data inconsistency (one fact may have more than one value – various versions may occur).
- **Poor Enforcement of System Standards:** Data names, formats, access restrictions... etc. are not standardized across an organization. This makes modifications difficult and hinders sharing of data.

File Approach Characteristics (con't)

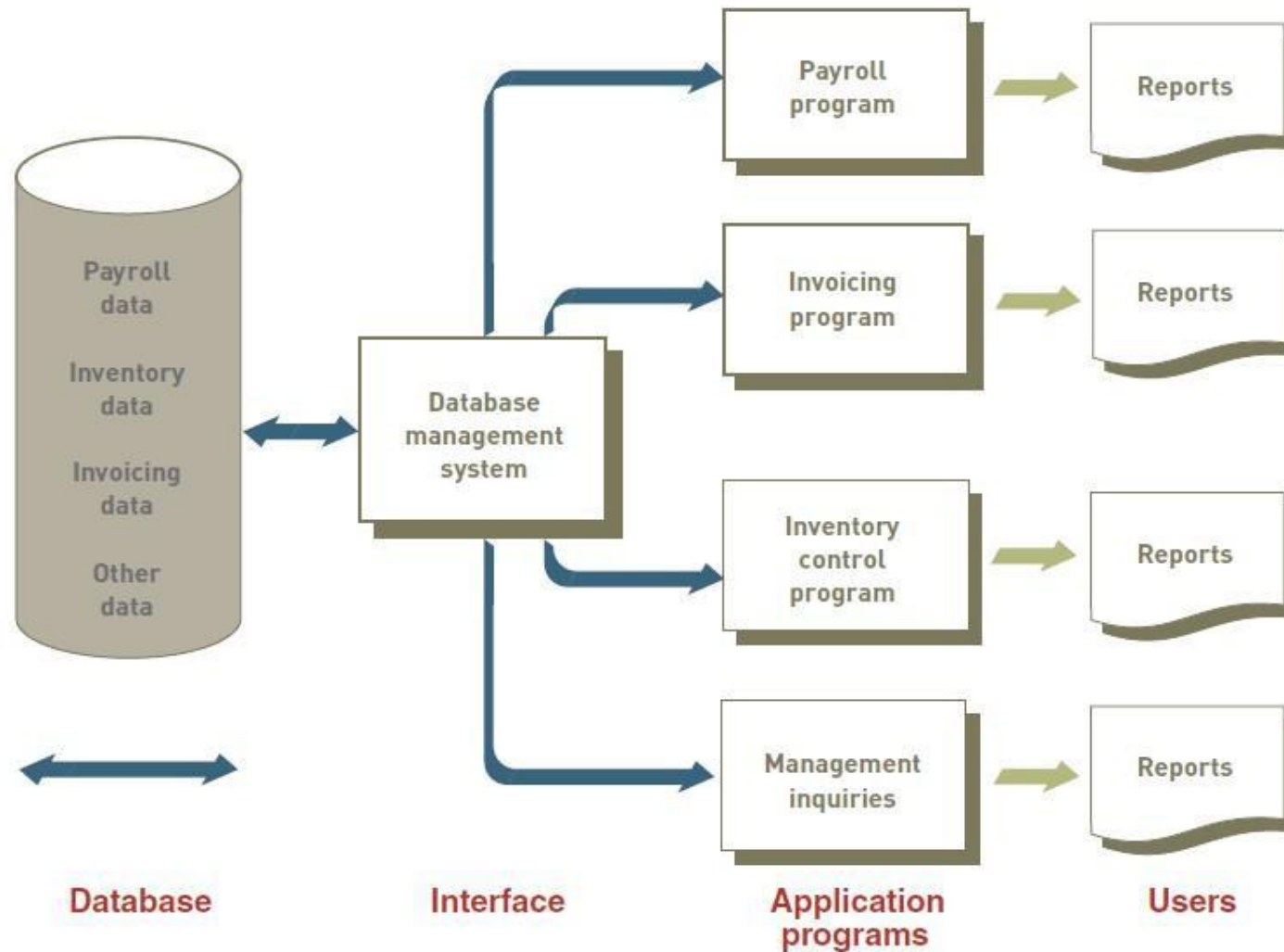
- **Limited Data Sharing:** Each application has its own private file providing little opportunity for users to share existing data. Any new applications would not be able to use existing files leading to low productivity.
- **Program – Data Dependency:** Descriptions of files, records, data items are embedded within application programs. Any modification to data file requires that the application program using that file must also be changes. In other words, program maintenance will be excessive

Database Approach Characteristics

- A centralized Database Management System (DBMS) exists, which handles all data management activities.
- The DBMS does not fragment data into separate files but regards data as being stored in a large conceptual repository – database. The DBMS handles the addition, storage, update, and retrieval of data.



Database Approach - Overview



Database Approach - Advantages

- Improved strategic use of corporate data
- Reduced Data Redundancy
- Improved Data Integrity
- Easier modification and updating
- Data and program independence
- Better access to data and information
- Standardization of data access
- Improved data safeguarding
- Efficient use of resources

Database Approach – Disadvantages

Disadvantages	Explanation
More complexity	DBMSs can be difficult to set up and operate. Many decisions must be made correctly for the DBMS to work effectively. In addition, users have to learn new procedures to take full advantage of a DBMS.
More difficult to recover from a failure	With the traditional approach to file management, a failure of a file affects only a single program. With a DBMS, a failure can shut down the entire database.
More expensive	DBMSs can be more expensive to purchase and operate than traditional file management. The expense includes the cost of the database and specialized personnel, such as a database administrator, who is needed to design and operate the database. Additional hardware might also be required.

File Processing vs Database Approach

File Processing Approach (Old School)	Database Approach (Current)
Storage Media: Sequential tapes or files	<u>D</u> irect <u>A</u> ccess <u>S</u> torage <u>D</u> evice (DASD)
Data: Stored in long sequential files (no relationship with other files)	Stored in tables with relationships with other files
Organization: redundant data in multiple files	Redundant data minimized/ eliminated
Updates: requires multiple updates in many files	Requires few or one update for a data field
Processing: slower query/ faster processing	Faster query/ slower processing

Data Management

End of Lecture 2 Part 1