

# Database Management

---

## Nature of Data

## Chapter 1

# Overview

- What is a database (revisited)
- The *data asset*
- Database characteristics
- Database applications

# What is a database, exactly?

---

A database is a set of data that has a regular structure and is organized in such a way that a computer can easily retrieve the desired results.

- Note: in this course we are dealing with structured data
- Text data, audio, video considered unstructured data



# What is a database management system DBMS (Revisited)?

## Software package

- Create, organize, manage a database
- Oracle in this course

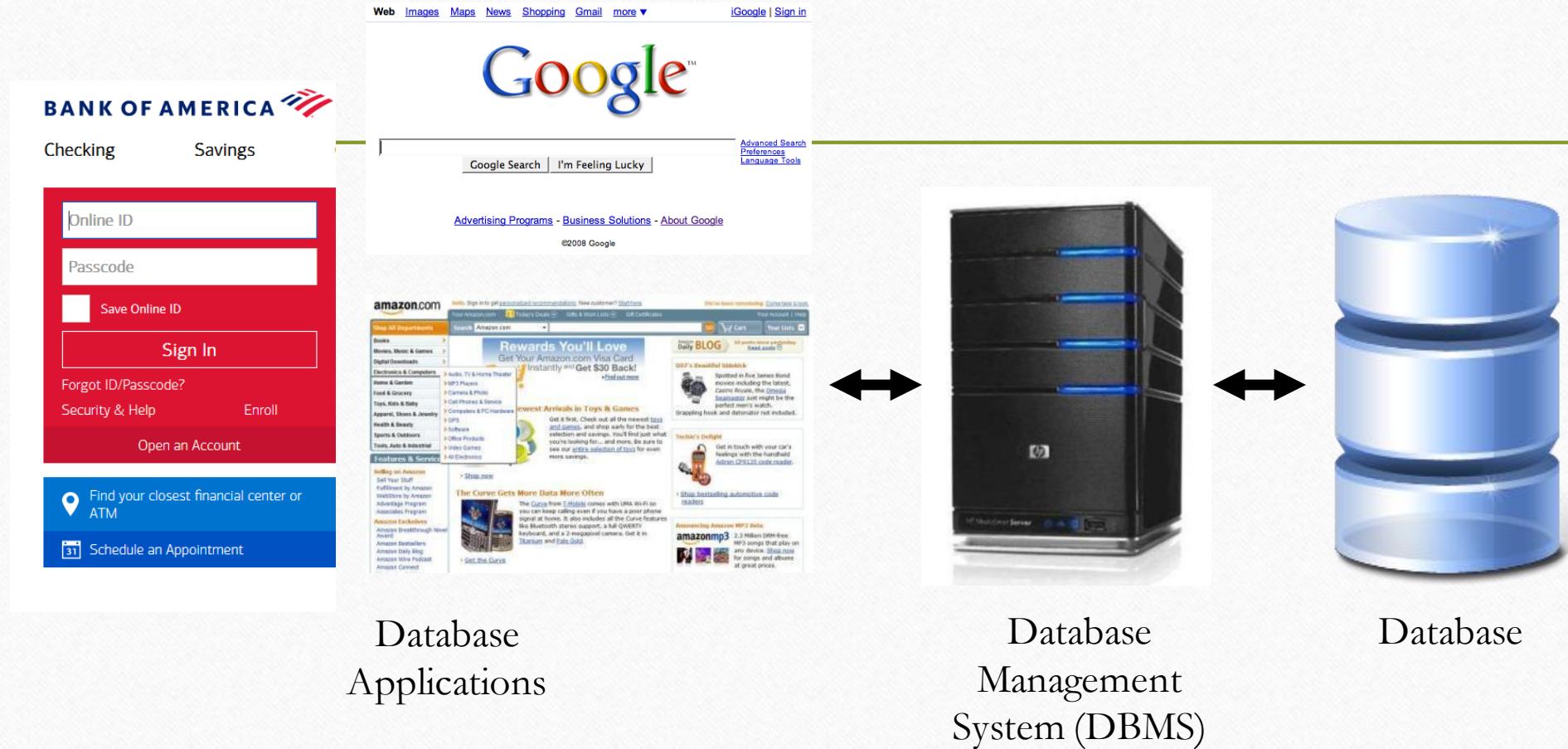
## Tools

- Querying, handling multiple users
- SQL

## Issues

- Design for effective use
- Privacy, security, ethics

# Database System (Revisited)



Important point: “raw” data must be protected.



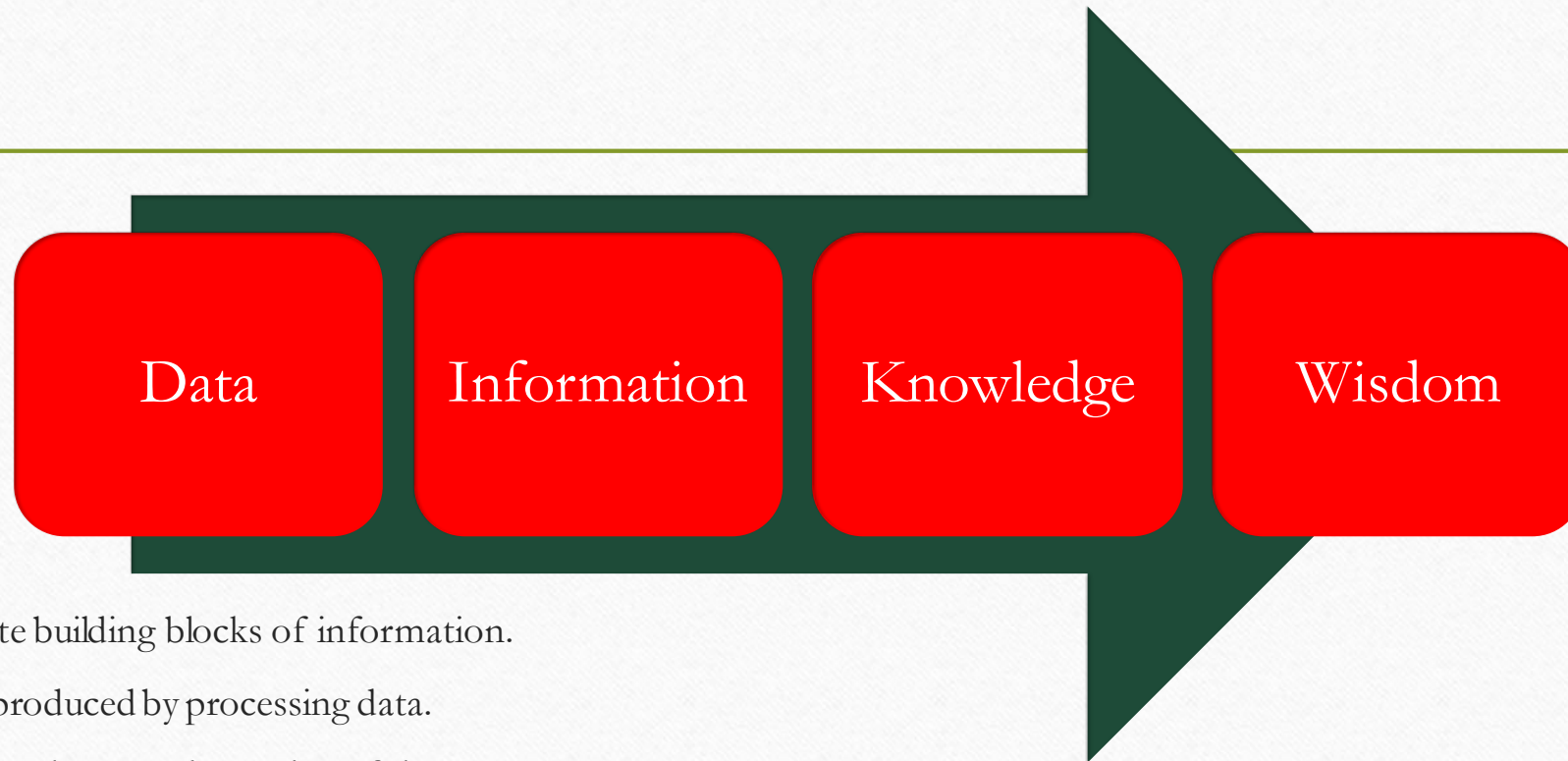
# Uses of DBMS (revisited)

## Data

- Archive data
  - Why?

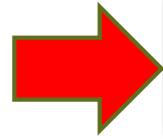
---
- Transactions (OLTP – Online Transaction Processing)
  - Point of sale (POS)
  - Banking transaction
  - Securities Trading
- Analysis (OLAP – Online Analytic Processing)
  - Business Intelligence/Data Mining
  - Assessment versus prediction / trend analysis
- What other sources / uses of data?

# This course is about data. What is data?



- Data constitute building blocks of information.
- Information produced by processing data.
- Information used to reveal meaning of data.
- Accurate, relevant, and timely information key to good decision making.
- Good decision-making key to organizational survival in a global environment.

# Overview



What is a database (revisited)

The data asset

Database characteristics

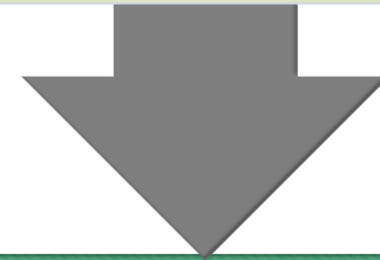
Database applications



# The Data Asset: Learning Objectives

Understand how **growth of data resources**, cheaper and faster computing, and effective tools for working with data *are collectively enabling a new age of decision making*

**Internal** and **external** sources of data



Know key terms and understand the importance of technologies associated with data organization and management

Database  
management  
systems

Data  
warehouses

Business  
intelligence and  
data analytics

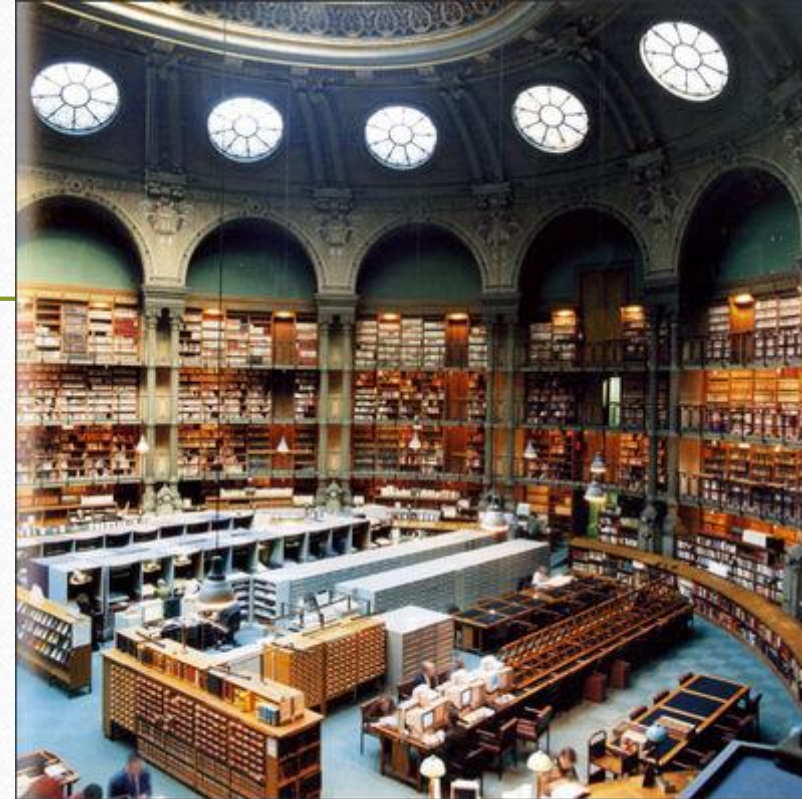
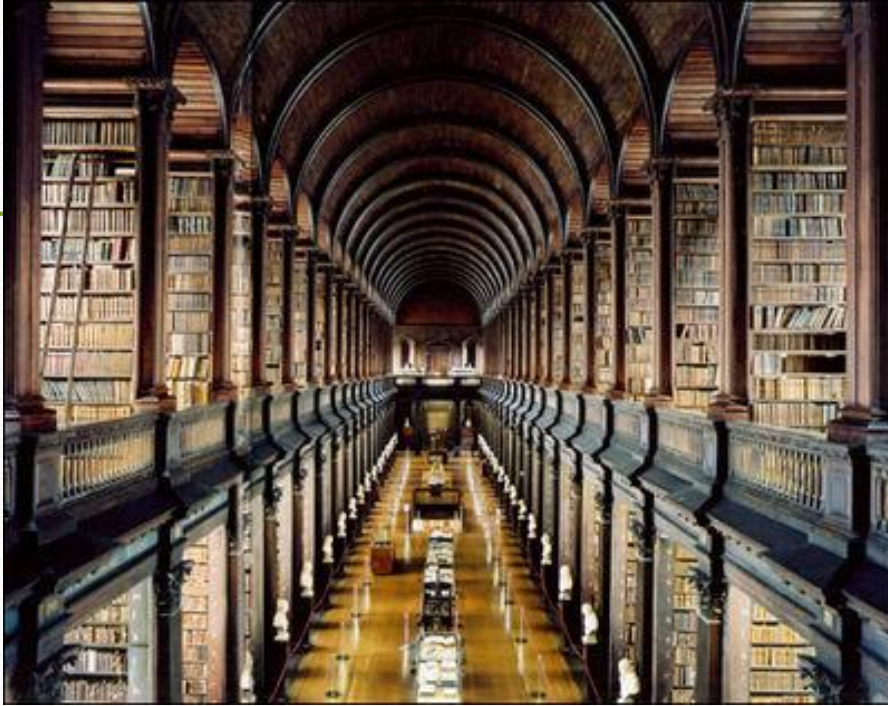
# The Data Asset: Sizing the Corporation

bit (binary digit)	1 or 0	Basic unit of computing, on/off switch, magnetic polarity
Byte (B)	8 bits	1 Byte = 1 character, 01000001 = A, 01100001 = a
Kilobyte (KB)	1,000 Bytes	1 Kilobytes = 1 paragraph of text
Megabyte (MB)	1,000 KB	1 Megabyte = 200 pages of text
Gigabyte (GB)	1,000 MB	1 Gigabyte = 10 yards of books on a shelf
Terabyte (TB)	1,000 GB	10 Terabytes = print collections of the U.S. Library of Congress Cost?
Petabyte (PB)	1,000 TB	1 Petabyte = 5 billion pages of text 2 Petabytes = entire contents of all U.S. academic research
Exabyte (EB)	1,000 PB	2 Exabytes = total volume of information generated in 1999
Zettabyte (ZB)	1,000 EB	500 Exabytes = the entire Internet in 2009. 2019, 2020?
Yottabyte	1,000 ZB	2021?

Question: Where are companies today wrt size of data asset?



# Sizing Data



The Library of Congress is the largest library in the world, with more than 147 million items on approximately 838 miles (1348 kilometers) of bookshelves. The collections include more than 33 million books and other print materials, 3 million recordings, 12.5 million photographs, 5.4 million maps, 6 million pieces of sheet music and 64.5 million manuscripts.



10 Terabytes = print collections of the U.S. Library of Congress. We can hold the digital version in our hand.



# Wal-Mart: Data Asset

---

- **Really Big Data At Walmart: Real-Time Insights From Their 40+ Petabyte Data Cloud**

- Tagline: “Walmart – the world’s biggest retailer with over 20,000 stores in 28 countries, is in the process of building the world’ biggest private cloud, to process 2.5 petabytes of data every hour.”

Over 200 streams of internal and external data, including 40 petabytes of recent **transactional data**, can be **modelled**, **manipulated** and **visualized**.

E.g. “... during Halloween, sales analysts were able to see in real-time that although a particular novelty cookie was very popular in most stores, there were two stores where it wasn’t selling at all.”

Ref. [Please read] <https://www.forbes.com/sites/bernardmarr/2017/01/23/really-big-data-at-walmart-real-time-insights-from-their-40-petabyte-data-cloud/#74bdf6556c10>

# The New Know (T. May) “New Know”

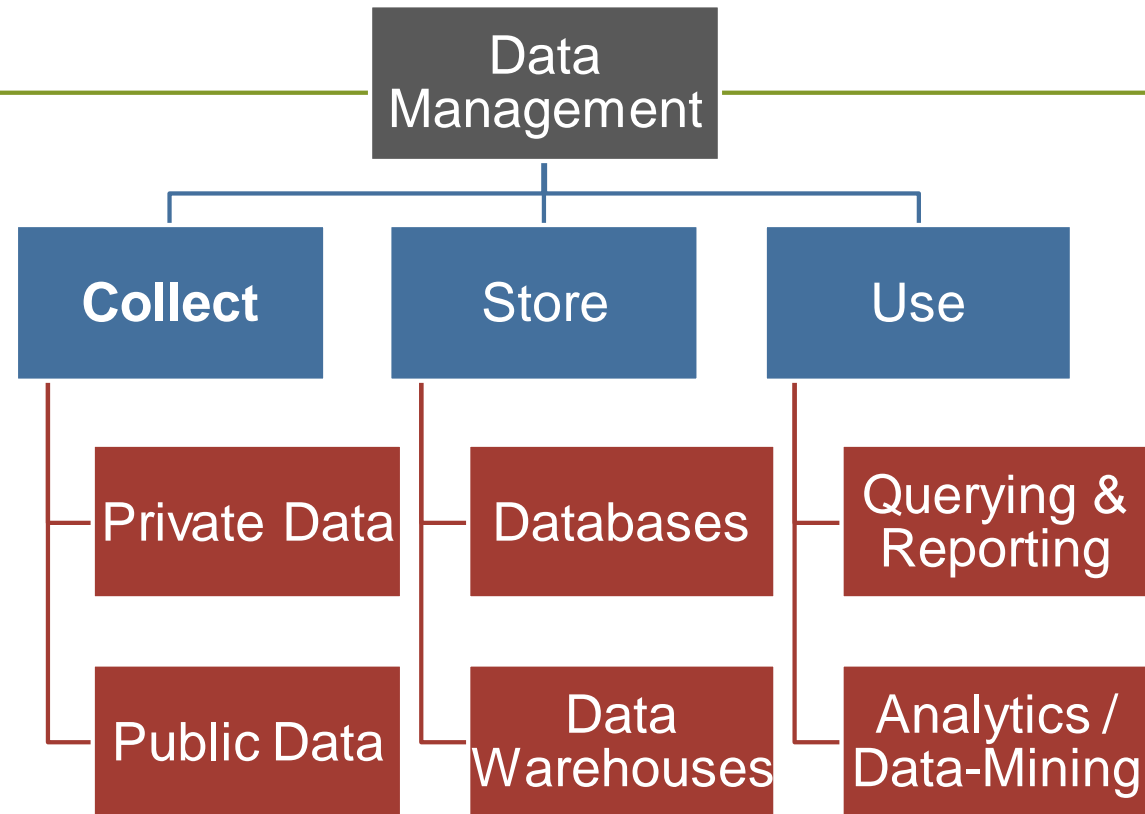
## What to know?

- **Knowns**
- **Known unknowns**
- **Unknown unknowns**

## What products? Who buys them?

- **If a consumer stopped buying one item, then, that customer would stop going into the store. What is the item?**
- **[Unknown unknown, but would be useful to know.]**

# This Course





# Data Types and Sources

---

- **Private** (Internal)
  - collected by organization
- **Sources**
  - **Transaction Processing Systems (TPS)**
    - Record business transactions (e.g. register sales, banking transaction, product returns)
  - **Enterprise software – CRM (Customer Relationship Management), SCM (Supply Chain Management), and ERP (Enterprise Resource Planning)**
    - CRM -- additional customer data beyond conventional purchase transactions
    - SCM and ERP -- every aspect of value chain

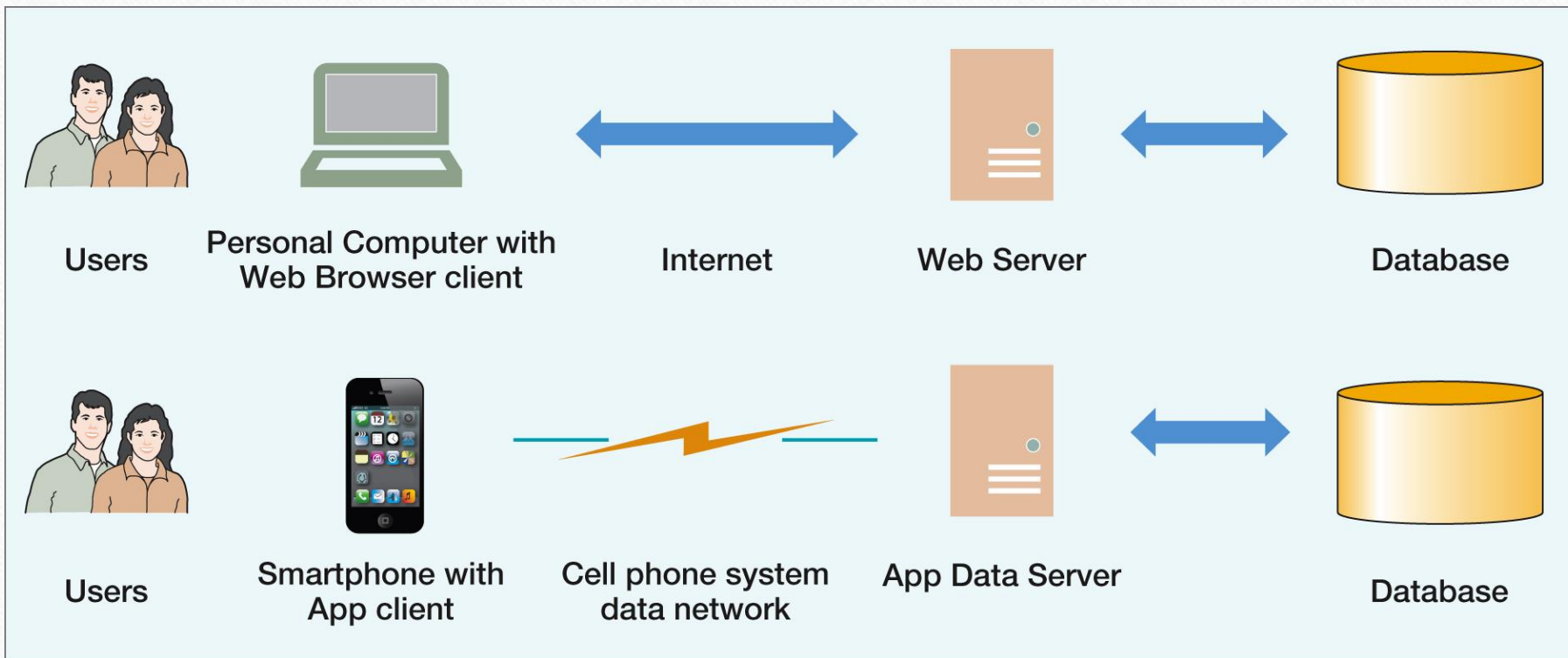
# Data Types and Sources

---

- **Public** (External)
  - collected by someone else
  - free / purchase
  - not source of competitive advantage by itself
  - data quality can be an issue
- Sources
  - Industry partners
  - Data aggregators
  - Public sources (e.g., census data, public records)

# The Internet and Mobile Device World

## Client-Server Architecture (Text. Chapter 1)



1-17



# Overview



- What is a database (revisited)
- The data asset
- Database characteristics
- Database applications

# Characteristics of Databases [Terminology]

- The purpose of a **database** is to help people track things of interest to them.
- Data is stored in **tables**, which have rows and columns like a spreadsheet. A database may have multiple tables, where each table stores data about a different thing.
- Each row in a table stores data about an occurrence or **instance** of the thing of interest.
- A database stores **data** and represents **relationships**.

[Note: Will be revisited with respect to relational databases. A relational database (implemented in Oracle tables, also called relations. We will design and implement a relational database in this course.)]



# Figure 1-3

## The STUDENT and CLASS Tables

The STUDENT table

This row stores the data for Sam Cooke

StudentNumber	LastName	FirstName	EmailAddress
1	Cooke	Sam	Sam.Cooke@OurU.edu
2	Lau	Marcia	Marcia.Lau@OurU.edu
3	Harris	Lou	Lou.Harris@OurU.edu
4	Greene	Grace	Grace.Greene@OurU.edu
(New)			

The CLASS table

This column stores the ClassName for each class

ClassNumber	ClassName	Term	Section
10	CHEM 101	2017-Fall	1
20	CHEM 101	2017-Fall	2
30	CHEM 101	2018-Spring	1
40	ACCT 101	2017-Fall	1
50	ACCT 101	2018-Spring	1



# Data in Tables

The STUDENT table

The CLASS table

The GRADE table  
—but who do these  
grades belong to?

STUDENT			
StudentNumber	LastName	FirstName	EmailAddress
1	Cooke	Sam	Sam.Cooke@OurU.edu
2	Lau	Marcia	Marcia.Lau@OurU.edu
3	Harris	Lou	Lou.Harris@OurU.edu
4	Greene	Grace	Grace.Greene@OurU.edu
(New)			

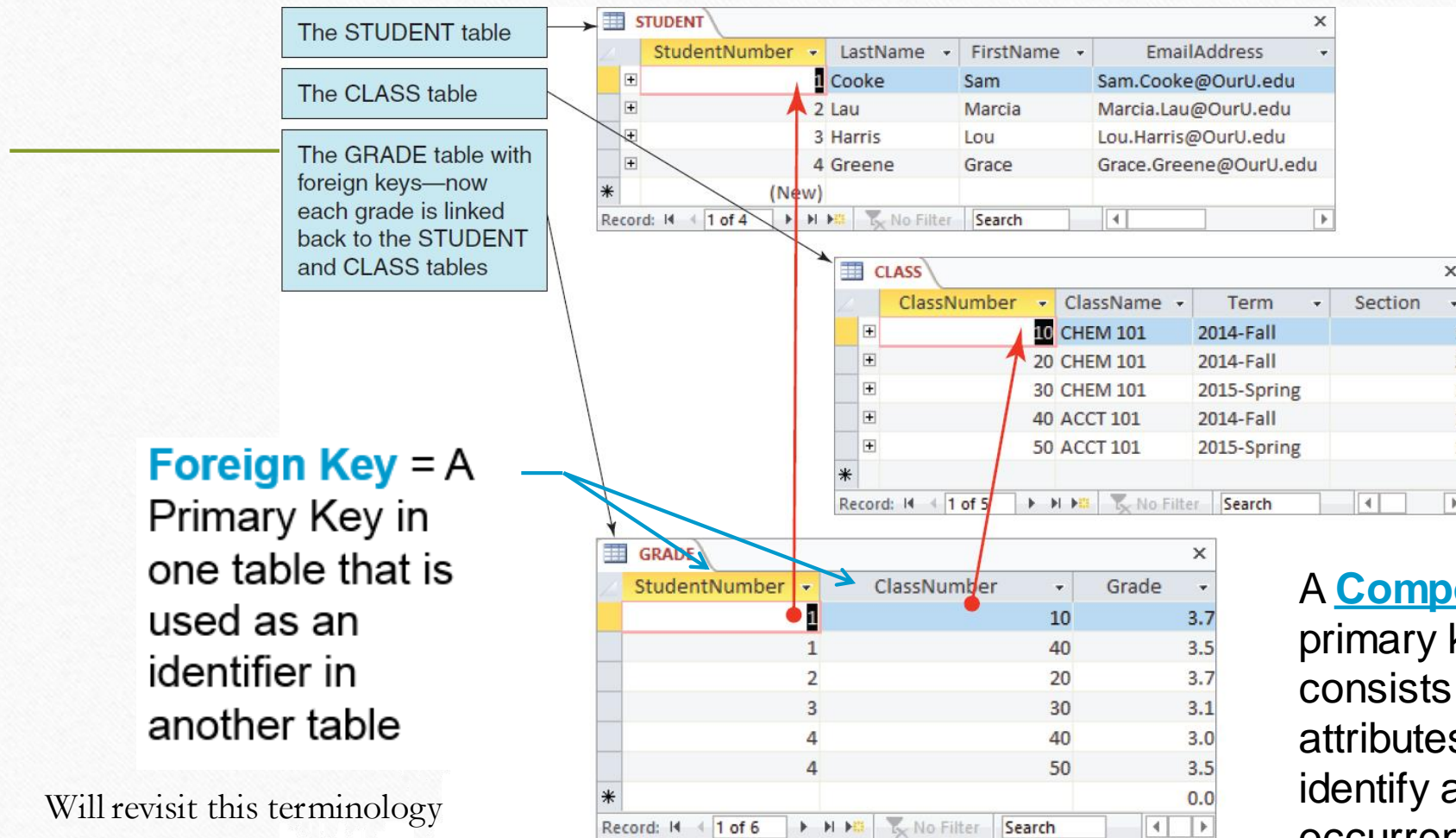
CLASS			
ClassNumber	ClassName	Term	Section
10	CHEM 101	2014-Fall	1
20	CHEM 101	2014-Fall	2
30	CHEM 101	2015-Spring	1
40	ACCT 101	2014-Fall	1
50	ACCT 101	2015-Spring	1

GRADE	
Grade	
3.7	
3.5	
3.7	
3.1	
3.0	
3.5	
0.0	

A **Primary Key** is a unique identifier field within a table.

A **Surrogate Key** is a primary key field that is automatically assigned by the computer. An example is the StudentNumber field in the STUDENT table above.

# Key Characteristic of Databases: Related Tables



A **Composite Key** is a primary key that consists of 2 or more attributes that uniquely identify an entity occurrence.

1-22



# Databases: Creation of Information

---

- **Data** = recorded facts and figures
- **Information** = derived from data
- Databases record data, so we can produce information from the data.
  - The data on STUDENTs, CLASSes, and GRADEs could produce information about each student's GPA.



# Overview

---

What is a database (revisited)

---

---

The data asset

---

---

Database characteristics

---

---

 Database applications

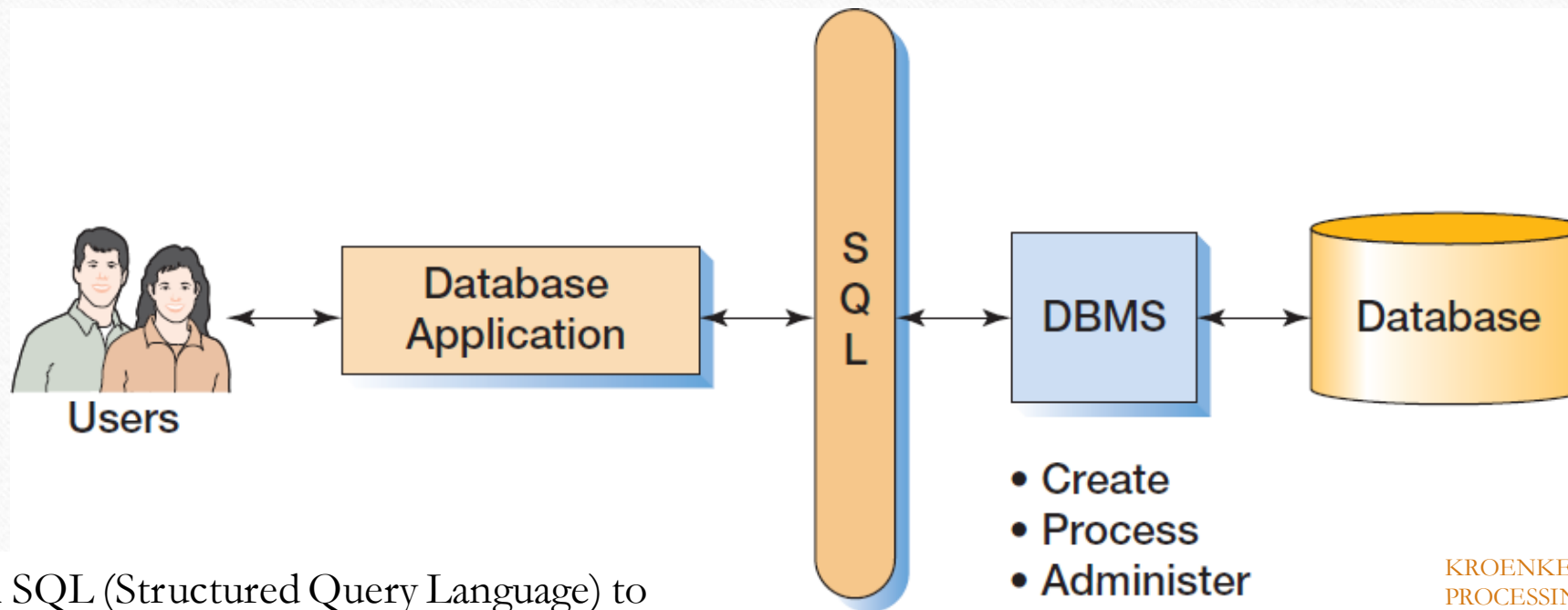
---

## Figure 1-7: Example Database Applications

• **Note.** Databases can be of different size and for a different number of users.

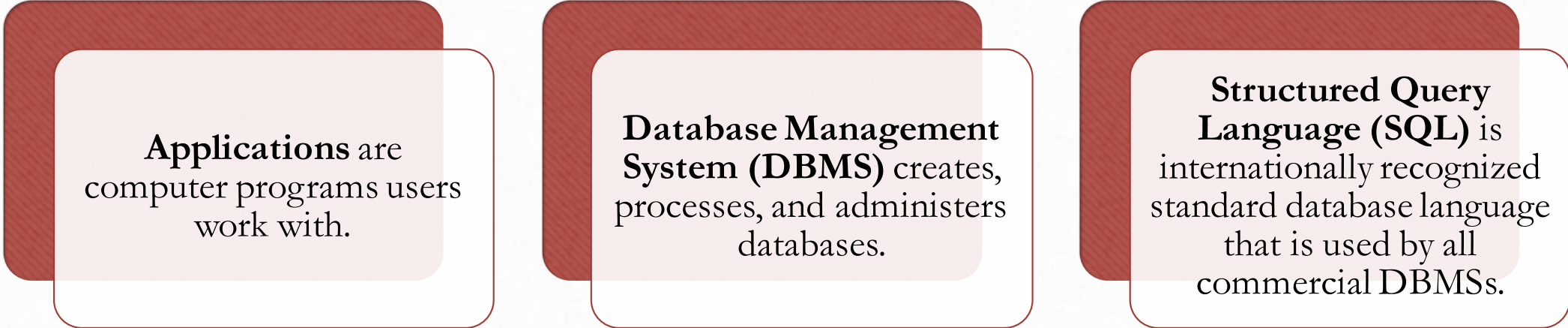
Application	Example Users	Number of Users	Typical Size	Remarks
Sales Contact Manager	Salesperson	1	2,000 rows	Products such as GoldMine and Act! Are database centric
Patient appointment (doctor, dentist)	Medical office	15 to 50	100,000 rows	Vertical market software vendors incorporate databases into their software products
Customer relationship management (CRM)	Sales, marketing, or customer service departments	500	10 million rows	Major vendors such as Microsoft and Oracle PeopleSoft Enterprise build applications around the database
Enterprise resource planning (ERP)	An entire organization	500	10 million+ rows	SAP uses a database as a central repository for ERP data.
E-commerce site	Internet users	Possibly millions	1 billion+ rows	Drugstore.com has a database that grows at the rate of 20 million rows per day!
Digital dashboard	Senior managers	500	100,000 rows	Extractions, summaries, and consolidations of operational databases.
Data mining	Business analysts	25	100,000 to millions+	Data are extracted, reformatted, cleaned, and filtered for use by statistical mining tools.

# Components of a Database System with SQL



We will learn SQL (Structured Query Language) to run queries against a relational database.





**Applications** are computer programs users work with.

**Database Management System (DBMS)** creates, processes, and administers databases.

**Structured Query Language (SQL)** is internationally recognized standard database language that is used by all commercial DBMSs.

# Applications, DBMS, and SQL

# Applications

## Basic Functions of Application Programs

Create and process forms

Process user queries

Create and process reports

Execute application logic

Control the application itself

# The DBMS (Database Management System)

Functions of a DBMS
Create database
Create tables
Create supporting structures (e.g., indexes)
Modify (insert, update, or delete) database data
Read database data
Maintain database structures
Enforce rules
Control concurrency
Perform backup and recovery



# Conclusion

- Data (asset)
  - Internal / external data
  - Use in decision making
    - (What kind today and beyond?)
- Database Management Systems
  - First need to design database well
  - Next topic
- Know: Functions and uses of databases
- Appreciate: Challenges in design, creation, use