Lecture 2: Database Systems

BADM/ACCY 352

Spring 2017

Instructor: Yi Yang, PhD

Last lecture

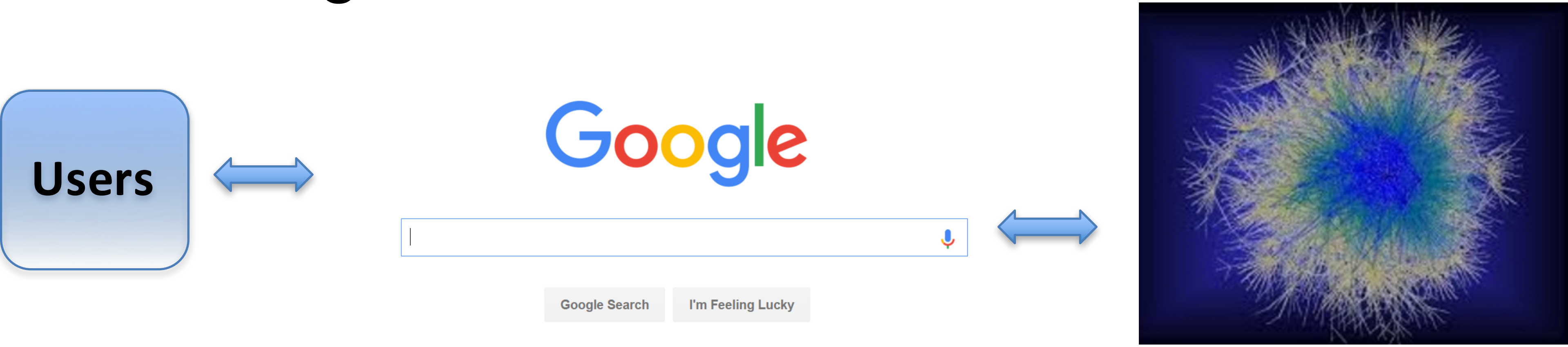
- Course Overview

This lecture

- Database systems overview
- Relational database

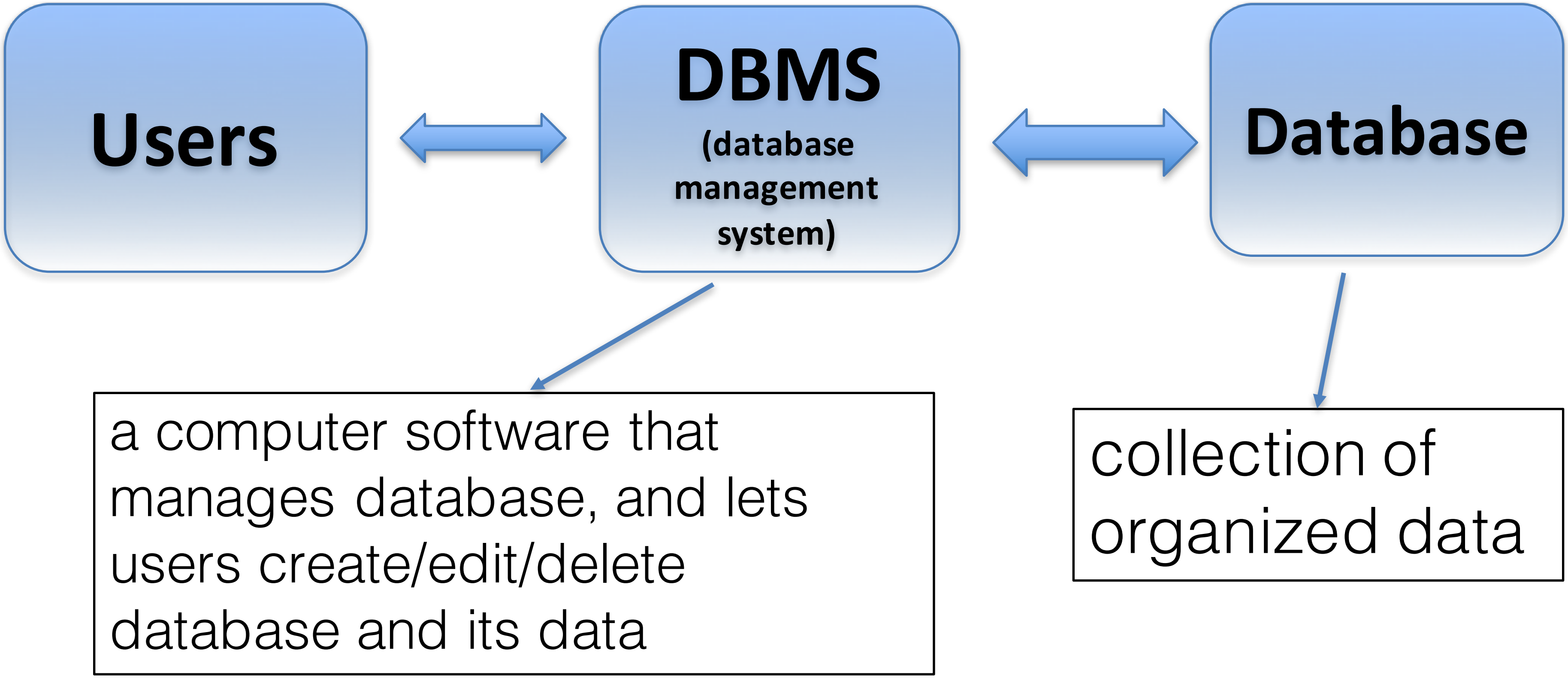
Open Question:

- Is Google a database?



- Is contact book a database?

Database and DBMS



Users

DBMS
(database management system)

Database

a computer software that manages database, and lets users create/edit/delete database and its data

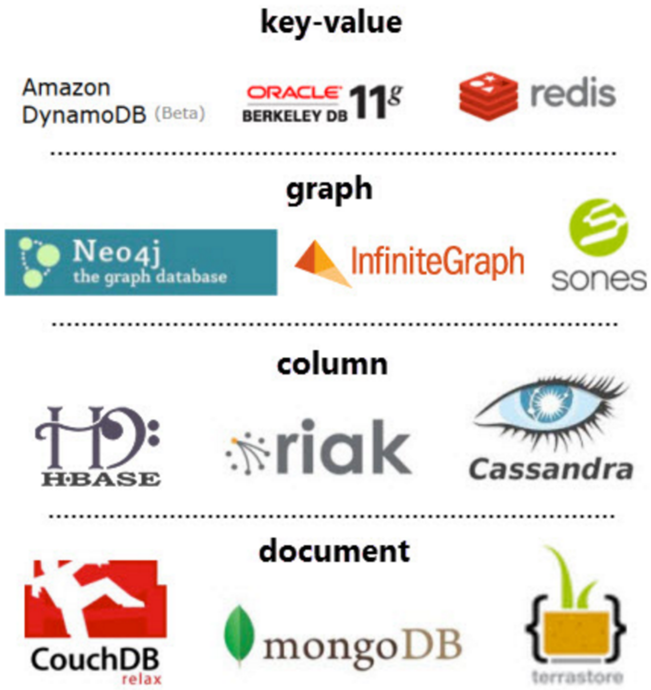
collection of organized data

We often call a database *X database* if it's managed by DBMS *X*.

Types of DBMS by Data Model

- Data model describe the structure of a database
- Relational Data Model: SQL
- Non-Relational Data Model: NoSQL

ORACLE
Microsoft SQL Server
PostgreSQL
MySQL
SYBASE
IBM DB2.



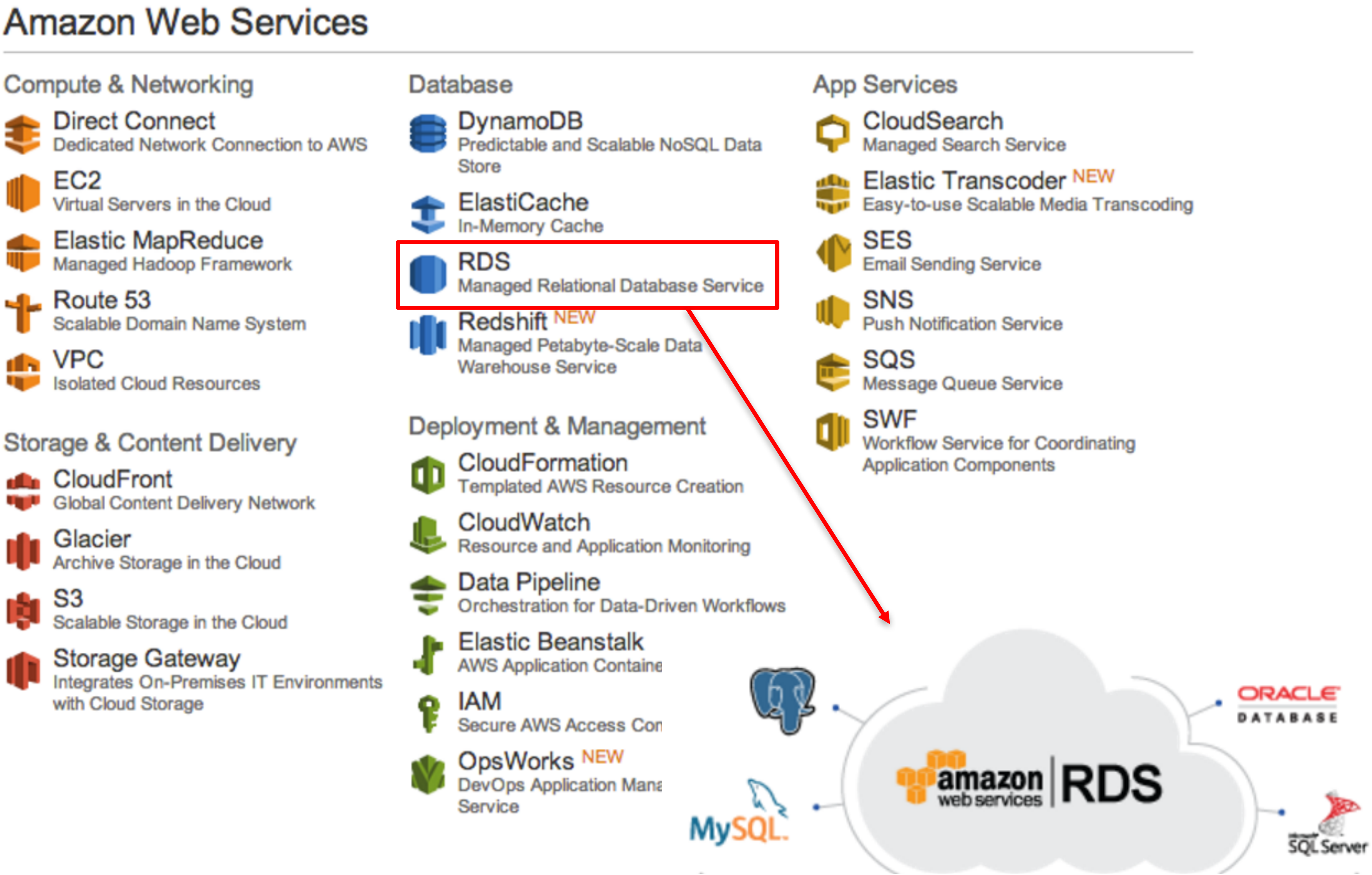
Types of DBMS by Location

- Centralized:
  - Supports data located at a single site
- Distributed:
  - Supports data distributed across several sites
- Cloud:
  - services provided by third-party vendors, such as Amazon Web Service
- Embedded:
  - tightly integrated with an application software

Advantage of using cloud database

- Scalability
  - Very easy to add resources
- Elasticity
  - Don’t have to accommodate periodic spikes in network traffic
- Cost saving
  - You pay what you use

Perfect solution for start-up companies and enterprises.



Amazon Web Services

Compute & Networking

Direct Connect
Dedicated Network Connection to AWS

EC2
Virtual Servers in the Cloud

Elastic MapReduce
Managed Hadoop Framework

Route 53
Scalable Domain Name System

VPC
Isolated Cloud Resources

Storage & Content Delivery

CloudFront
Global Content Delivery Network

Glacier
Archive Storage in the Cloud

S3
Scalable Storage in the Cloud

Storage Gateway
Integrates On-Premises IT Environments with Cloud Storage

Database

DynamoDB
Predictable and Scalable NoSQL Data Store

ElastiCache
In-Memory Cache

RDS
Managed Relational Database Service

Redshift NEW
Managed Petabyte-Scale Data Warehouse Service

Deployment & Management

CloudFormation
Templated AWS Resource Creation

CloudWatch
Resource and Application Monitoring

Data Pipeline
Orchestration for Data-Driven Workflows

Elastic Beanstalk
AWS Application Containe

IAM
Secure AWS Access Con

OpsWorks NEW
DevOps Application Mana
Service

App Services

CloudSearch
Managed Search Service

Elastic Transcoder NEW
Easy-to-use Scalable Media Transcoding

SES
Email Sending Service

SNS
Push Notification Service

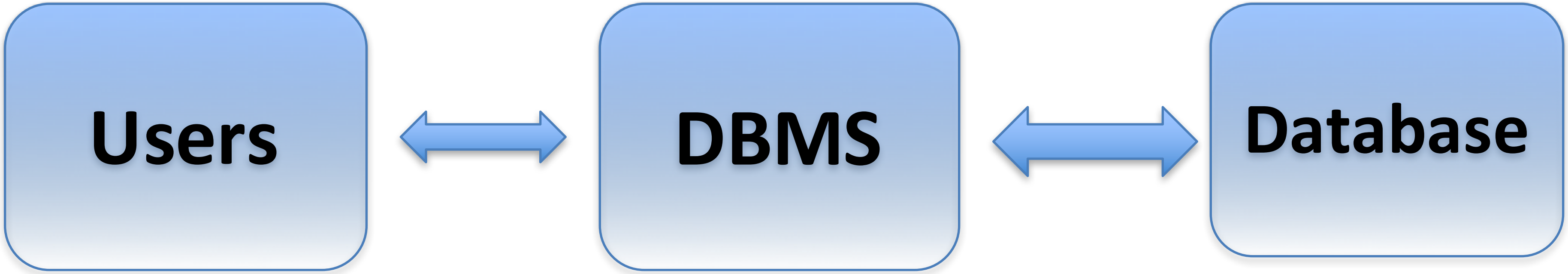
SQS
Message Queue Service

SWF
Workflow Service for Coordinating Application Components

- https://aws.amazon.com/solutions/case-studies/

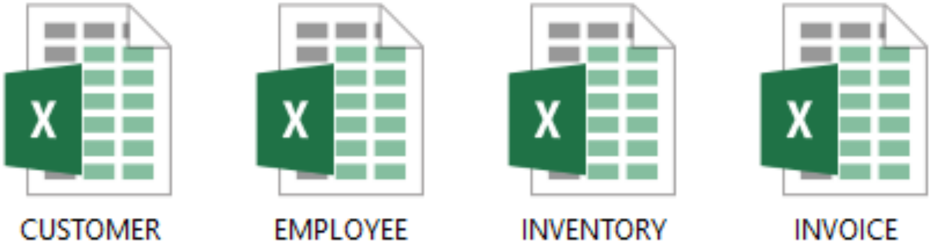
Summary

- Types of DBMS by Data Model
- Types of DBMS by Location

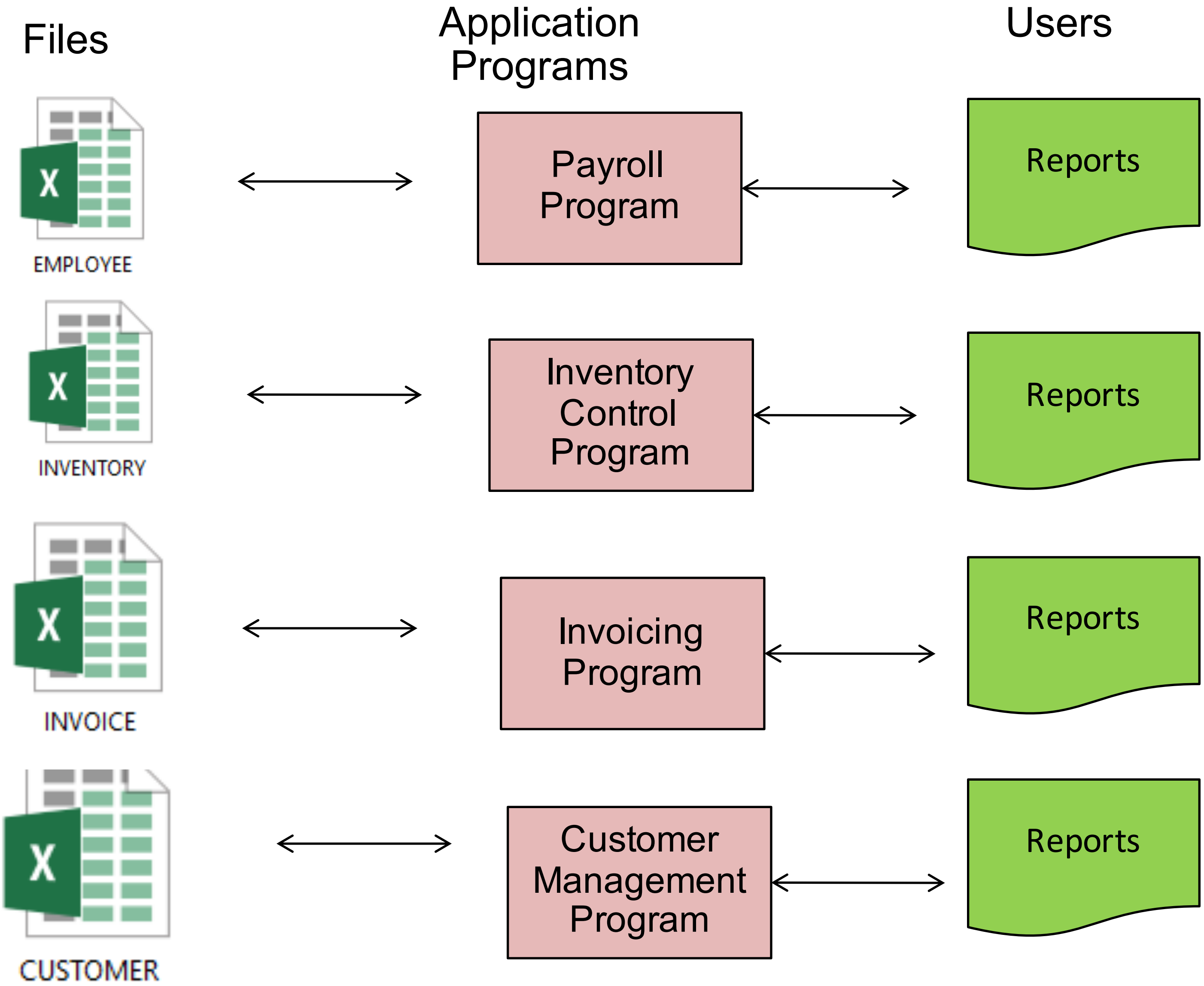


Data Management

- If you are the manager of a local grocery store, how would you manage the store operation data, including Employee, Customer, Inventory, and Invoice?



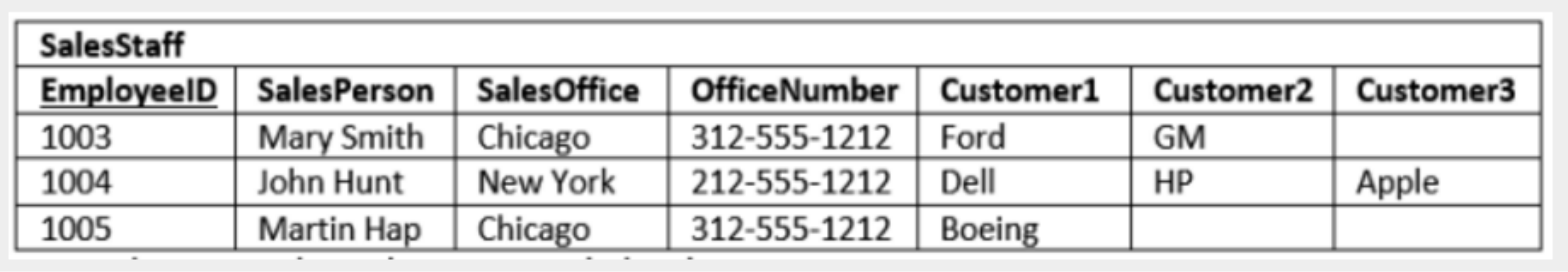
Traditional File Approach



Disadvantages

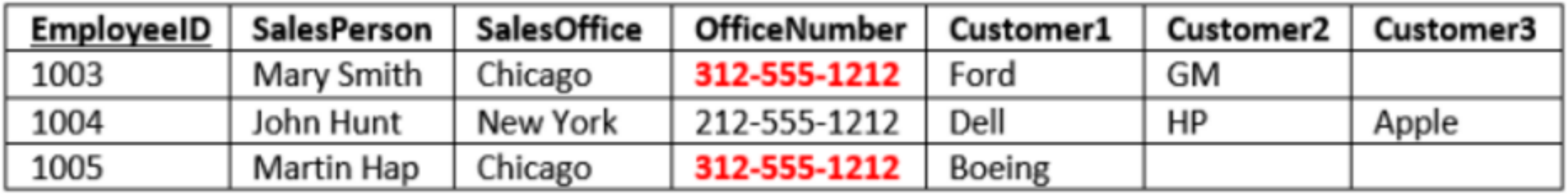
- Limited data sharing
- Difficulty of getting quick answers
- Poor data security
- Data redundancy -> data inconsistency

Data redundancy -> data inconsistency



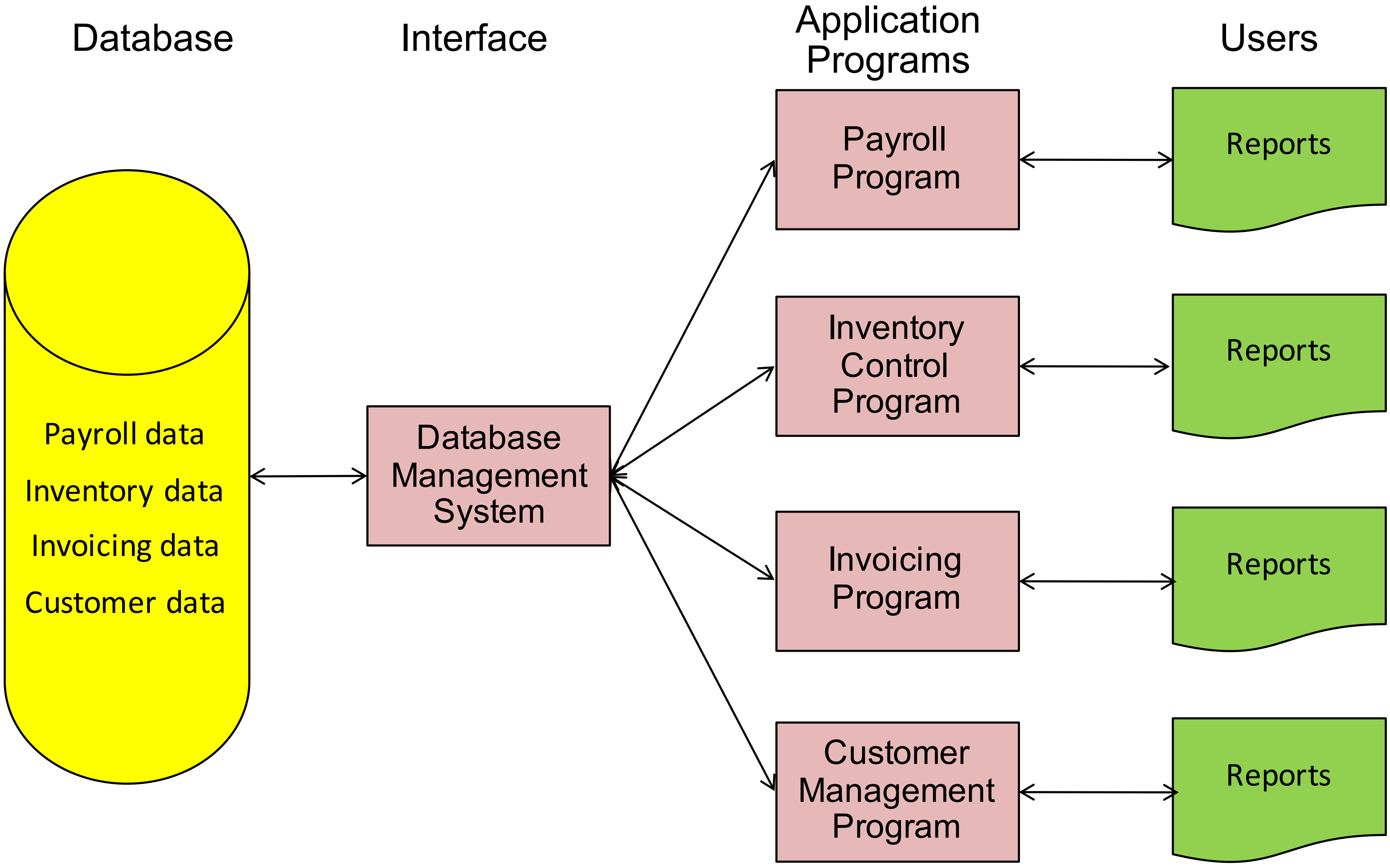
SalesStaff						
EmployeeID	SalesPerson	SalesOffice	OfficeNumber	Customer1	Customer2	Customer3
1003	Mary Smith	Chicago	312-555-1212	Ford	GM	
1004	John Hunt	New York	212-555-1212	Dell	HP	Apple
1005	Martin Hap	Chicago	312-555-1212	Boeing		

if the office number changes, then there are multiple updates that need to be made.



EmployeeID	SalesPerson	SalesOffice	OfficeNumber	Customer1	Customer2	Customer3
1003	Mary Smith	Chicago	312-555-1212	Ford	GM	
1004	John Hunt	New York	212-555-1212	Dell	HP	Apple
1005	Martin Hap	Chicago	312-555-1212	Boeing		

Database Approach



Pros and Cons of Database Approach

Pros:

- data sharing
- fast and efficient query language
- minimal data redundancy -> data consistency
- Security

Cons:

- vendor dependence & frequent upgrade

  (Oracle RDBMS, IBM DB2, MS Access, MS SQL Server, MySQL)

iClicker question

Based on data model, database systems can be categorized as relational database (such as Oracle, MySQL, etc) or non-relational database (such as MongoDB, Neo4j, etc).

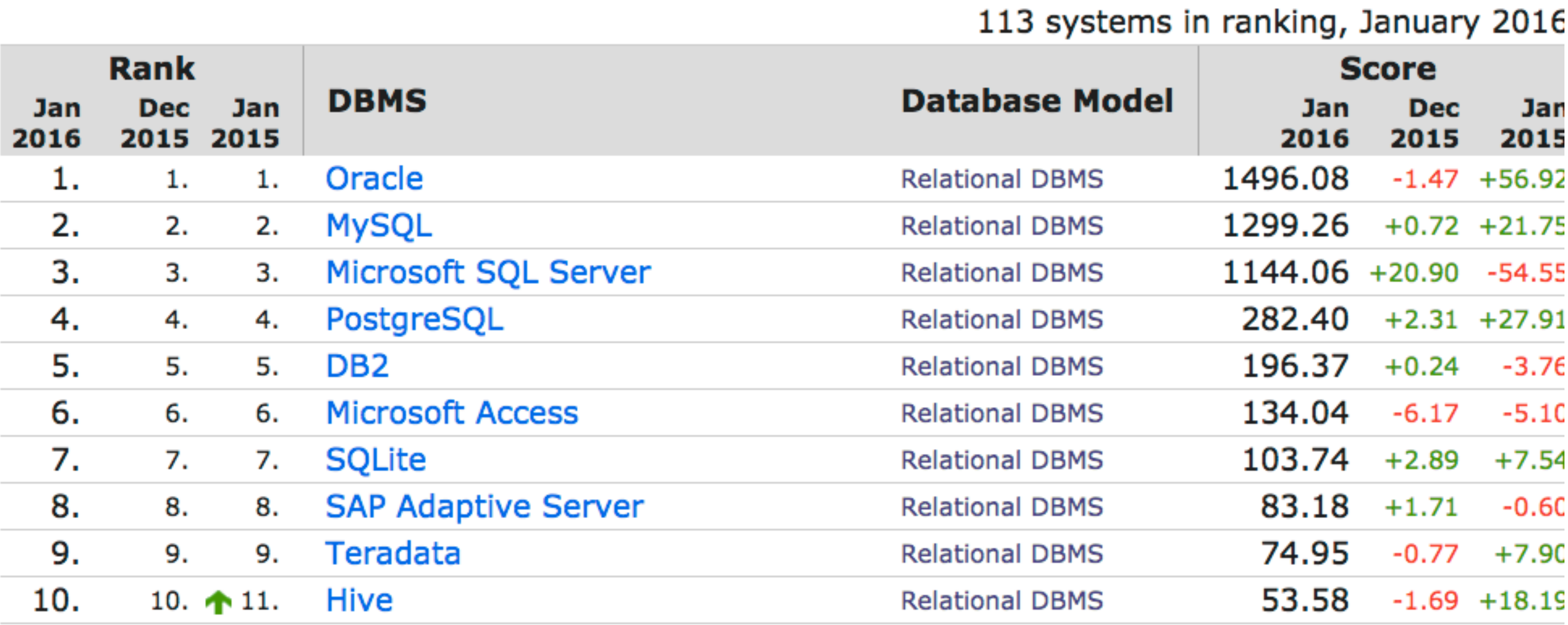
A. True

B. False

Let’s start the journey of data modeling!

Relational Database

- It is a database whose organization is based on the relational model of data.
- The DBMS used to maintain a relational database is known as a relational DBMS (RDBMS)
- All relational database systems use SQL as the language for querying and maintaining the database.



113 systems in ranking, January 2016

Rank			DBMS	Database Model	Score		
Jan 2016	Dec 2015	Jan 2015			Jan 2016	Dec 2015	Jan 2015
1.	1.	1.	Oracle	Relational DBMS	1496.08	-1.47	+56.92
2.	2.	2.	MySQL	Relational DBMS	1299.26	+0.72	+21.75
3.	3.	3.	Microsoft SQL Server	Relational DBMS	1144.06	+20.90	-54.55
4.	4.	4.	PostgreSQL	Relational DBMS	282.40	+2.31	+27.91
5.	5.	5.	DB2	Relational DBMS	196.37	+0.24	-3.76
6.	6.	6.	Microsoft Access	Relational DBMS	134.04	-6.17	-5.10
7.	7.	7.	SQLite	Relational DBMS	103.74	+2.89	+7.54
8.	8.	8.	SAP Adaptive Server	Relational DBMS	83.18	+1.71	-0.60
9.	9.	9.	Teradata	Relational DBMS	74.95	-0.77	+7.90
10.	10.	↑ 11.	Hive	Relational DBMS	53.58	-1.69	+18.19

Relational Database Popularity Ranking

Relational Model

- It models data into one or more related tables of columns and rows, with a unique key identifying each row.

- A relational database contains a group of related tables.

Table

- Two-dimensional structure composed of rows and columns.
- It’s a collection of related data.

- Table is called Relation
  - Relation is not Relationship (between tables)
- Row is called record.
- Column is called attribute.

More on Tables

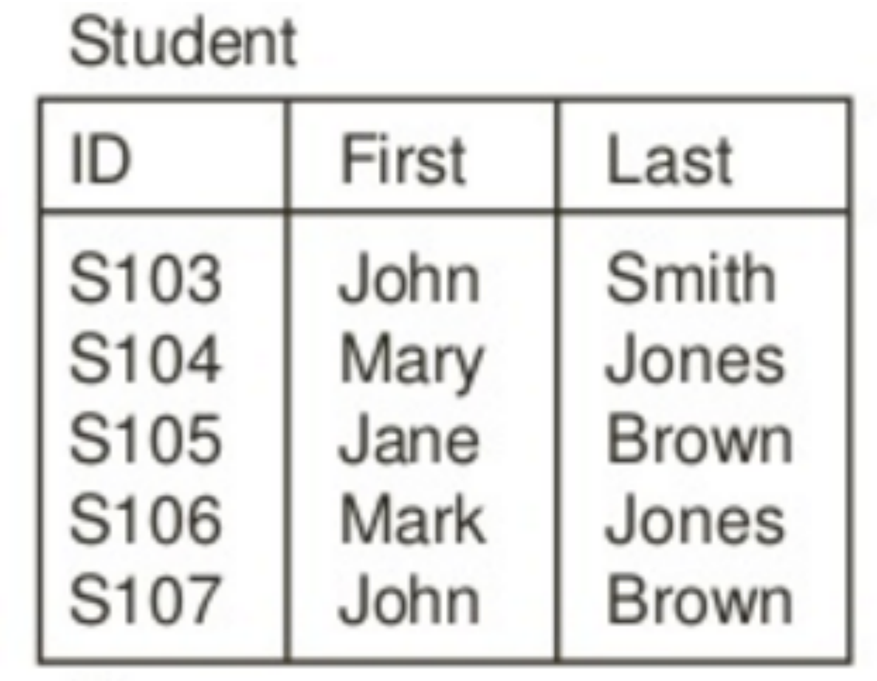
- Each table should have an attribute (column) or combination of attributes that uniquely identifies each record (row)

- Each column/row intersection represents a single data value (could be empty)

- All values in a attribute must conform to the same data format. Some examples are number, text, date, currency

- Table is a logical construct—not necessarily how data are actually stored physically

A Table (relation) Example

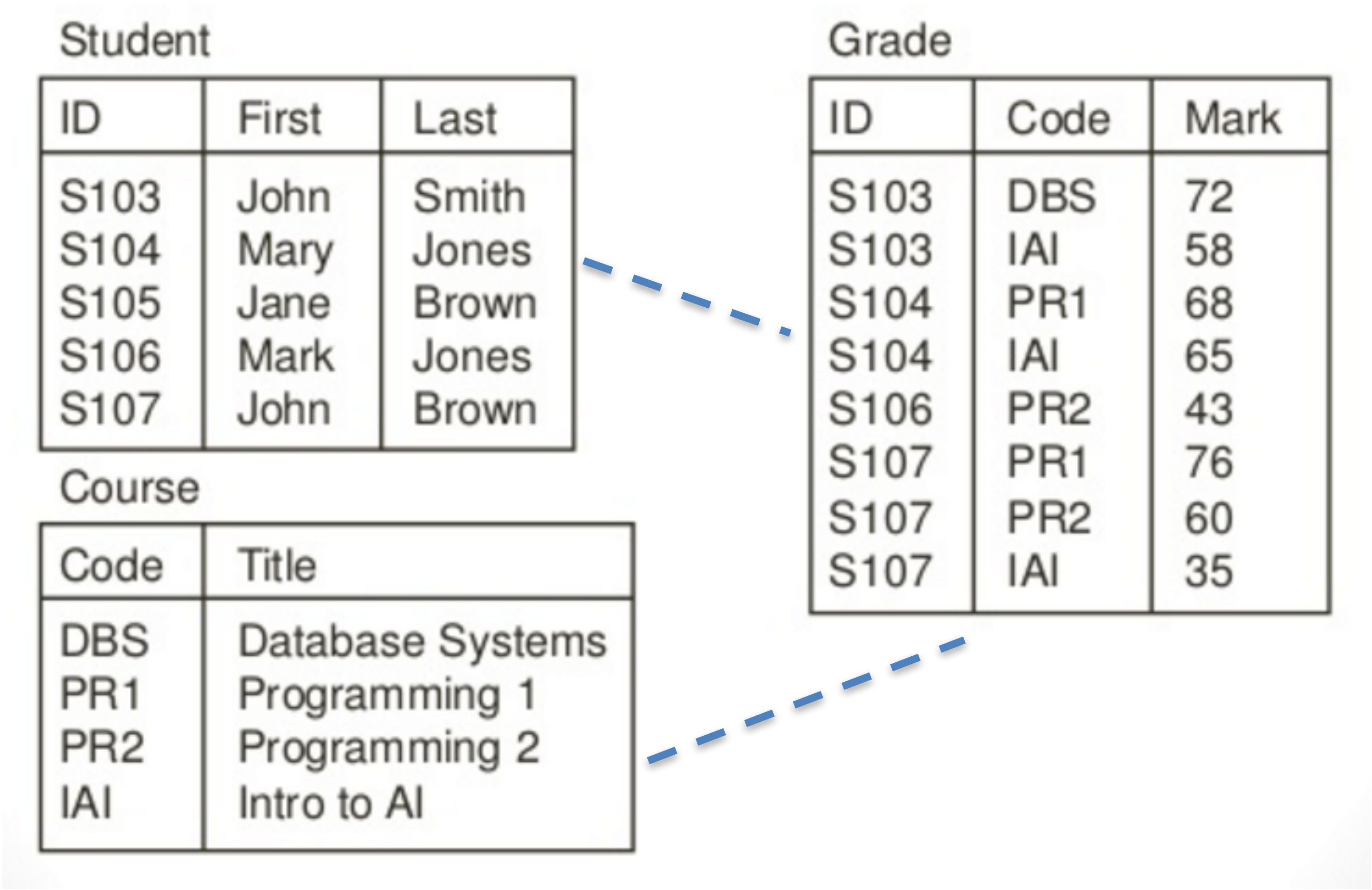


Student

ID	First	Last
S103	John	Smith
S104	Mary	Jones
S105	Jane	Brown
S106	Mark	Jones
S107	John	Brown

How many attributes?
How many records?
Which attribute uniquely identify each record?

A Relational Database Example



Student

ID	First	Last
S103	John	Smith
S104	Mary	Jones
S105	Jane	Brown
S106	Mark	Jones
S107	John	Brown

Course

Code	Title
DBS	Database Systems
PR1	Programming 1
PR2	Programming 2
IAI	Intro to AI

Grade

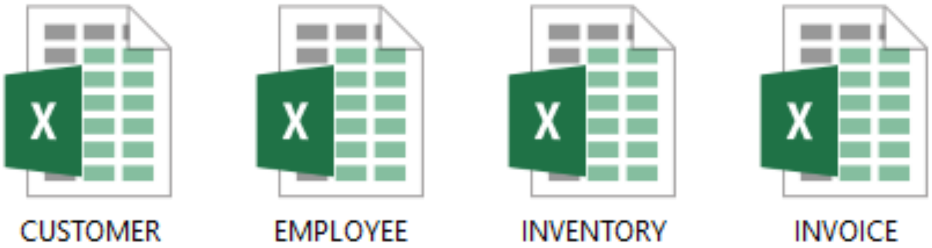
ID	Code	Mark
S103	DBS	72
S103	IAI	58
S104	PR1	68
S104	IAI	65
S106	PR2	43
S107	PR1	76
S107	PR2	60
S107	IAI	35

Is an Excel Spreadsheet a Database?

- ***No***
- handle less data
- not allowing multi-user editing
- less secure
- lacks of enforcement to minimize data redundancy and ensure consistency
- cannot define relationships among tables

Transforming Excels to Database

- If you are the manager of a local Walmart store, how would you manage the store operation data, including Employee, Customer, Inventory, and Invoice?



Summary

- Database systems
  - Understand different types of databases
  - Understand the advantages of using a database to manage data

- Relational database
  - Understand relational model