M.Sc. in Computer Science
Department of Computer Science
University of Sri Jayewardenepura

CSC 542 2.0 Database Systems and Administration

Presented By:
Surani Tissera(PhD)
Department of Computer Science

Course Content

- Introduction to database systems and data models;
- Database Architecture;
- Database administration;
- Entity Relationship (ER) modelling;
- Extended Entity Relationship (EER) Model;
- The relational database model;
- Mapping the ER model to Relational DB;
- Normalization of database tables;
- Introduction to Structured Query Language (SQL);

References

- Carlos Coronel, Steven Morris, and Peter Rob. 2018. Database Systems: Design, Implementation, and Management (13th. ed.). Course Technology Press, Boston, MA, USA.
- Thomas M. Connolly and Carolyn E. Begg. 2014. Database Systems: A Practical Approach to Design, Implementation and Management (6th Edition) (International Computer Science). Addison-Wesley Longman Publishing Co., Inc., USA.
- Ramez Elmasri and Shamkant B. Navathe. 2015. Fundamentals of Database Systems (7th. ed.). Pearson.
- C.J. Date. 2003. An Introduction to Database Systems (8th. ed.). Addison-Wesley Longman Publishing Co., Inc., USA

In this chapter, you will learn:

- The difference between data and information
- What a database is, the various types of databases, and why they are valuable assets for decision making
- The importance of database design
- How modern databases evolved from file systems
- About flaws in file system data management
- The main components of the database system
- The main functions of a database management system (DBMS)

Data Vs Information

Data

- Raw facts; building blocks of information
- Unprocessed information
- Ex: numbers, text, images, sounds, or any other format
- Categories of data:
  - Structured Data: Data is organized in a specific format. Ex:in Database
  - Unstructured Data: Unstructured data doesn't have a predefined format
  - Semi-Structured Data: Partially organized data . Ex:in XML

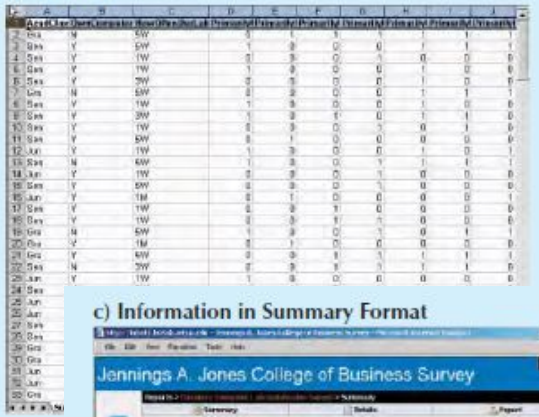
Information

- Meaningful and organized data derived from the raw data
- To reveal meaning, information requires context.
    - For example, an average temperature is 105 degrees.
      Is this in degrees Fahrenheit or Celsius?
      Is this a machine temperature, a body temperature, or an outside air temperature?
- Having access to **accurate**, **relevant**, and **timely** information is key for making good decisions.
- Good decision making is the key to organizational survival in a global environment.

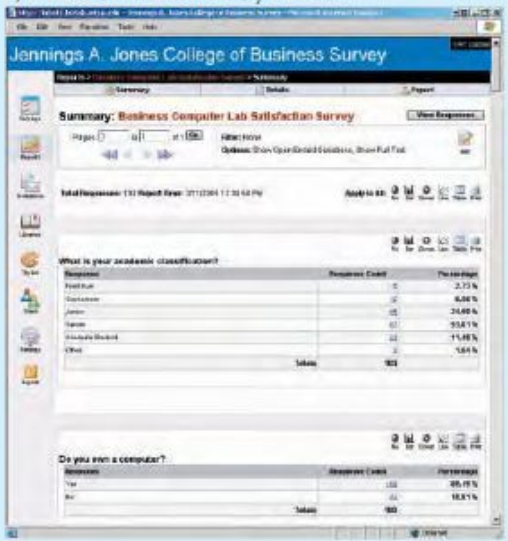
FIGURE 1.1 Transforming raw data into information

a) Initial Survey Screen

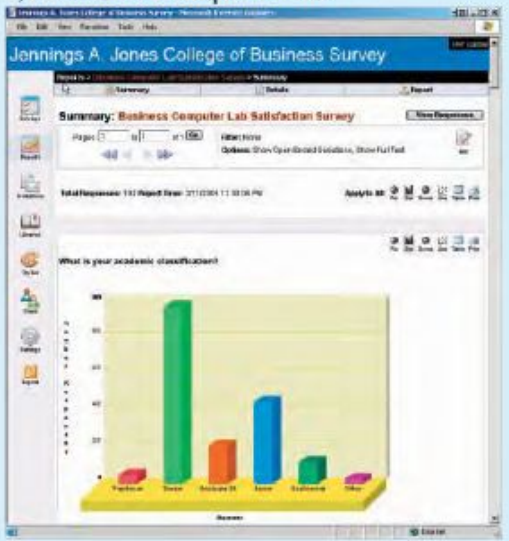
b) Raw Data



c) Information in Summary Format



d) Information in Graphic Format



Introduction to Database

- A database is a **shared, integrated computer structure**.
- It stores a collection of:
  - **End-user data -** raw facts of interest to the end user.
  - **Metadata-** or data about data, through which the end-user data are integrated and managed.

Data Item		Metadata				
Name	Type	Length	Min	Max	Description	Source
Course	Alphanumeric	30			Course ID and name	Academic Unit
Section	Integer	1	1	9	Section number	Registrar
Semester	Alphanumeric	10			Semester and year	Registrar
Name	Alphanumeric	30			Student name	Student IS
ID	Integer	9			Student ID (SSN)	Student IS
Major	Alphanumeric	4			Student major	Student IS
GPA	Decimal	3	0.0	4.0	Student grade point average	Academic Unit

Data management

- Data management is a discipline that focuses on the **proper generation, storage, and retrieval of data.**

- Data management is a core activity for any business, government agency, service organization, or charity.

Introduction to DBMS

- A database management system (DBMS) is a **collection of programs that manages the database structure and controls access to the data** stored in the database.
- A database resembles a very well-organized electronic filing cabinet in which powerful software.
- Possible to share data among multiple applications or users.
- Makes data management more efficient and effective.

Advantage of DBMS

- Improve data sharing.
- Improve data security.
- Better data integration.
- Minimize data inconsistency.
- Improve data access.
- Improve decision making.
- Increase end-user productivity

Compare the features with well-known DBMS

TABLE 1.1 Types of Databases

PRODUCT	NUMBER OF USERS			DATA LOCATION		DATA USAGE		XML
	SINGLE USER	MULTIUSER		CENTRALIZED	DISTRIBUTED	OPERATIONAL	DATA WAREHOUSE	
		WORKGROUP	ENTERPRISE					
MS Access	X	X		X		X		
MS SQL Server	X[3]	X	X	X	X	X	X	X
IBM DB2	X[3]	X	X	X	X	X	X	X
MySQL	X	X	X	X	X	X	X	X*
Oracle RDBMS	X[3]	X	X	X	X	X	X	X

* Supports XML functions only. XML data are stored in large text objects.

[3] Vendor offers single-user/personal DBMS version.

Class Questions

1. What is a DBMS, and what are its functions?
2. What is the role of a DBMS, and what are its advantages? What are its disadvantages?
3. List and describe the different types of databases.
4. What are metadata?
5. Explain why database design is important.
6. What are some basic database functions that a spreadsheet cannot perform?

Class Questions

Given the file structure shown in Figure P1.1, answer Problems

1. How many records does the file contain? How many fields are there per record?
2. What problem would you encounter if you wanted to produce a listing by city? How would you solve this problem by altering the file structure?
3. If you wanted to produce a listing of the file contents by last name, area code, city, state, or zip code, how would you alter the file structure

PROJECT_CODE	PROJECT_MANAGER	MANAGER_PHONE	MANAGER_ADDRESS	PROJECT_BID_PRICE
21-5Z	Holly B. Parker	904-338-3416	3334 Lee Rd., Gainesville, FL 37123	16833460.00
25-2D	Jane D. Grant	615-898-9909	218 Clark Blvd., Nashville, TN 36362	12500000.00
25-5A	George F. Dorts	615-227-1245	124 River Dr., Franklin, TN 29185	32512420.00
25-9T	Holly B. Parker	904-338-3416	3334 Lee Rd., Gainesville, FL 37123	21563234.00
27-4Q	George F. Dorts	615-227-1245	124 River Dr., Franklin, TN 29185	10314545.00
29-2D	Holly B. Parker	904-338-3416	3334 Lee Rd., Gainesville, FL 37123	25559999.00
31-7P	William K. Moor	904-445-2719	216 Morton Rd., Stetson, FL 30155	56850000.00

Traditional(Manual) File Systems

- Historically, manual paper-and-pencil systems
- Organized data using file folders and cabinets
- Effective for small data collections and simple reporting
- Manual file systems are time-consuming and cumbersome
- Challenges with growth and complex reporting

Terminology used in Database Management

Structural Dependence

A file system exhibits **structural dependence**, which means that access to a file is dependent on its structure.

Structural independence exists when it is possible to make changes in the file structure without affecting the application program's ability to access the data.

Structural Dependence

Ex:Adding a customer date-of-birth field to the CUSTOMER file shown in Figure would require the extra works.

1. Reads a record from the original file.
2. Transforms the original data to conform to the new structure's storage requirements.
3. Writes the transformed data into the new file structure.
4. Repeats steps 2 to 4 for each record in the original file

C_NAME	C_PHONE	C_ADDRESS	C_ZIP	A_NAME	A_PHONE	TP	AMT	REN
Alfred A. Ramas	615-844-2573	218 Fork Rd., Babs, TN	36123	Leah F. Hahn	615-882-1244	T1	$100.00	05-Apr-2004
Leona K. Dunne	713-894-1238	Box 12A, Fox, KY	25246	Alex B. Alby	713-228-1249	T1	$250.00	16-Jun-2004
Kathy W. Smith	615-894-2285	125 Oak Ln, Babs, TN	36123	Leah F. Hahn	615-882-2144	S2	$150.00	29-Jan-2005
Paul F. Olowski	615-894-2180	217 Lee Ln., Babs, TN	36123	Leah F. Hahn	615-882-1244	S1	$300.00	14-Oct-2004
Myron Orlando	615-222-1672	Box 111, New, TN	36155	Alex B. Alby	713-228-1249	T1	$100.00	28-Dec-2004
Amy B. O'Brian	713-442-3381	387 Troll Dr., Fox, KY	25246	John T. Okon	615-123-5589	T2	$850.00	22-Sep-2004
James G. Brown	615-297-1228	21 Tye Rd., Nash, TN	37118	Leah F. Hahn	615-882-1244	S1	$120.00	25-Mar-2004
George Williams	615-290-2556	155 Maple, Nash, TN	37119	John T. Okon	615-123-5589	S1	$250.00	17-Jul-2004
Anne G. Farriss	713-382-7185	2119 Elm, Crew, KY	25432	Alex B. Alby	713-228-1249	T2	$100.00	03-Dec-2004
Olette K. Smith	615-297-3809	2782 Main, Nash, TN	37118	John T. Okon	615-123-5589	S2	$500.00	14-Mar-2004

C_NAME = Customer name
C_PHONE = Customer phone
C_ADDRESS = Customer address
C_ZIP = Customer ZIP code

A_NAME = Agent name
A_PHONE = Agent phone
TP = Insurance type
AMT = Insurance policy amount, in thousands of $
REN = Insurance renewal date

Data Dependence

- **Data dependence** exists when changes in the characteristics of data, such as changing a field from integer to decimal, require changes in all the programs that access the file.

- **Data independence** exists when it is possible to make changes in the data storage characteristics without affecting the application program's ability to access the data.

Terminology

Logical data format -How the human being views the data

Physical data format - How the computer must work with the data

Data Redundancy

- Data Redundancy refers to the unnecessary duplication of data in a database, where the same data is stored in multiple locations or tables.
- Data redundancy sets
  - Poor data security
  - Data inconsistency

Data Redundancy -Example

-

	C_NAME	C_PHONE	C_ADDRESS	C_ZIP	A_NAME	A_PHONE	TP	AMT	REN
▶	Alfred A. Ramas	615-844-2573	218 Fork Rd., Babs, TN	36123	Leah F. Hahn	615-882-1244	T1	$100.00	05-Apr-2004
	Leona K. Dunne	713-894-1238	Box 12A, Fox, KY	25246	Alex B. Alby	713-228-1249	T1	$250.00	16-Jun-2004
	Kathy W. Smith	615-894-2285	125 Oak Ln, Babs, TN	36123	Leah F. Hahn	615-882-2144	S2	$150.00	29-Jan-2005
	Paul F. Olowski	615-894-2180	217 Lee Ln., Babs, TN	36123	Leah F. Hahn	615-882-1244	S1	$300.00	14-Oct-2004
	Myron Orlando	615-222-1672	Box 111, New, TN	36155	Alex B. Alby	713-228-1249	T1	$100.00	28-Dec-2004
	Amy B. O'Brian	713-442-3381	387 Troll Dr., Fox, KY	25246	John T. Okon	615-123-5589	T2	$850.00	22-Sep-2004
	James G. Brown	615-297-1228	21 Tye Rd., Nash, TN	37118	Leah F. Hahn	615-882-1244	S1	$120.00	25-Mar-2004
	George Williams	615-290-2556	155 Maple, Nash, TN	37119	John T. Okon	615-123-5589	S1	$250.00	17-Jul-2004
	Anne G. Farriss	713-382-7185	2119 Elm, Crew, KY	25432	Alex B. Alby	713-228-1249	T2	$100.00	03-Dec-2004
	Olette K. Smith	615-297-3809	2782 Main, Nash, TN	37118	John T. Okon	615-123-5589	S2	$500.00	14-Mar-2004

C_NAME = Customer name
C_PHONE = Customer phone
C_ADDRESS = Customer address
C_ZIP = Customer ZIP code

A_NAME = Agent name
A_PHONE = Agent phone
TP = Insurance type
AMT = Insurance policy amount, in thousands of $
REN = Insurance renewal date

Data Integrity

- Data integrity is defined as the condition in which all of the data in the database are consistent with the real-world events and conditions.
- In other words, data integrity means that:
  - Data are accurate—there are no data inconsistencies.
  - Data are verifiable—the data will always yield consistent results.

Data Anomalies

- Modification anomalies
  - Occur when changes must be made to existing records
- Insertion anomalies
  - Occur when entering new records
- Deletion anomalies
  - Occur when deleting records

Update Anomalies

If agent Leah F. Hahn has a new phone number, that number must be entered in each of the CUSTOMER file records in which Ms. Hahn's phone number is shown. In this case, only three changes must be made.

In a large file system, such a change might occur in hundreds or even thousands of records. Clearly, the potential for data inconsistencies is great.

C_NAME	C_PHONE	C_ADDRESS	C_ZIP	A_NAME	A_PHONE	TP	AMT	REN
Alfred A. Ramas	615-844-2573	218 Fork Rd., Babs, TN	36123	Leah F. Hahn	615-882-1244	T1	$100.00	05-Apr-2004
Leona K. Dunne	713-894-1238	Box 12A, Fox, KY	25246	Alex B. Alby	713-228-1249	T1	$250.00	16-Jun-2004
Kathy W. Smith	615-894-2285	125 Oak Ln, Babs, TN	36123	Leah F. Hahn	615-882-2144	S2	$150.00	29-Jan-2005
Paul F. Olowski	615-894-2180	217 Lee Ln., Babs, TN	36123	Leah F. Hahn	615-882-1244	S1	$300.00	14-Oct-2004
Myron Orlando	615-222-1672	Box 111, New, TN	36155	Alex B. Alby	713-228-1249	T1	$100.00	28-Dec-2004
Amy B. O'Brian	713-442-3381	387 Troll Dr., Fox, KY	25246	John T. Okon	615-123-5589	T2	$850.00	22-Sep-2004
James G. Brown	615-297-1228	21 Tye Rd., Nash, TN	37118	Leah F. Hahn	615-882-1244	S1	$120.00	25-Mar-2004
George Williams	615-290-2556	155 Maple, Nash, TN	37119	John T. Okon	615-123-5589	S1	$250.00	17-Jul-2004
Anne G. Farriss	713-382-7185	2119 Elm, Crew, KY	25432	Alex B. Alby	713-228-1249	T2	$100.00	03-Dec-2004
Olette K. Smith	615-297-3809	2782 Main, Nash, TN	37118	John T. Okon	615-123-5589	S2	$500.00	14-Mar-2004

C_NAME = Customer name
C_PHONE = Customer phone
C_ADDRESS = Customer address
C_ZIP = Customer ZIP code

A_NAME = Agent name
A_PHONE = Agent phone
TP = Insurance type
AMT = Insurance policy amount, in thousands of $
REN = Insurance renewal date

Insertion Anomalies

If only the CUSTOMER file existed, to add a new agent, you would also add a dummy customer data entry to reflect the new agent's addition. Again, the potential for creating data inconsistencies would be great.

C_NAME	C_PHONE	C_ADDRESS	C_ZIP	A_NAME	A_PHONE	TP	AMT	REN
Alfred A. Ramas	615-844-2573	218 Fork Rd., Babs, TN	36123	Leah F. Hahn	615-882-1244	T1	$100.00	05-Apr-2004
Leona K. Dunne	713-894-1238	Box 12A, Fox, KY	25246	Alex B. Alby	713-228-1249	T1	$250.00	16-Jun-2004
Kathy W. Smith	615-894-2285	125 Oak Ln, Babs, TN	36123	Leah F. Hahn	615-882-2144	S2	$150.00	29-Jan-2005
Paul F. Olowski	615-894-2180	217 Lee Ln., Babs, TN	36123	Leah F. Hahn	615-882-1244	S1	$300.00	14-Oct-2004
Myron Orlando	615-222-1672	Box 111, New, TN	36155	Alex B. Alby	713-228-1249	T1	$100.00	28-Dec-2004
Amy B. O'Brian	713-442-3381	387 Troll Dr., Fox, KY	25246	John T. Okon	615-123-5589	T2	$850.00	22-Sep-2004
James G. Brown	615-297-1228	21 Tye Rd., Nash, TN	37118	Leah F. Hahn	615-882-1244	S1	$120.00	25-Mar-2004
George Williams	615-290-2556	155 Maple, Nash, TN	37119	John T. Okon	615-123-5589	S1	$250.00	17-Jul-2004
Anne G. Farriss	713-382-7185	2119 Elm, Crew, KY	25432	Alex B. Alby	713-228-1249	T2	$100.00	03-Dec-2004
Olette K. Smith	615-297-3809	2782 Main, Nash, TN	37118	John T. Okon	615-123-5589	S2	$500.00	14-Mar-2004

C_NAME = Customer name
C_PHONE = Customer phone
C_ADDRESS = Customer address
C_ZIP = Customer ZIP code

A_NAME = Agent name
A_PHONE = Agent phone
TP = Insurance type
AMT = Insurance policy amount, in thousands of $
REN = Insurance renewal date

Deletion Anomalies

If you delete the customers Amy B. O'Brian, George Williams, and Olette K. Smith, you will also delete John T. Okon's agent data. Clearly, this is not desirable.

C_NAME	C_PHONE	C_ADDRESS	C_ZIP	A_NAME	A_PHONE	TP	AMT	REN
Alfred A. Ramas	615-844-2573	218 Fork Rd., Babs, TN	36123	Leah F. Hahn	615-882-1244	T1	$100.00	05-Apr-2004
Leona K. Dunne	713-894-1238	Box 12A, Fox, KY	25246	Alex B. Alby	713-228-1249	T1	$250.00	16-Jun-2004
Kathy W. Smith	615-894-2285	125 Oak Ln, Babs, TN	36123	Leah F. Hahn	615-882-2144	S2	$150.00	29-Jan-2005
Paul F. Olowski	615-894-2180	217 Lee Ln., Babs, TN	36123	Leah F. Hahn	615-882-1244	S1	$300.00	14-Oct-2004
Myron Orlando	615-222-1672	Box 111, New, TN	36155	Alex B. Alby	713-228-1249	T1	$100.00	28-Dec-2004
Amy B. O'Brian	713-442-3381	387 Troll Dr., Fox, KY	25246	John T. Okon	615-123-5589	T2	$850.00	22-Sep-2004
James G. Brown	615-297-1228	21 Tye Rd., Nash, TN	37118	Leah F. Hahn	615-882-1244	S1	$120.00	25-Mar-2004
George Williams	615-290-2556	155 Maple, Nash, TN	37119	John T. Okon	615-123-5589	S1	$250.00	17-Jul-2004
Anne G. Farriss	713-382-7185	2119 Elm, Crew, KY	25432	Alex B. Alby	713-228-1249	T2	$100.00	03-Dec-2004
Olette K. Smith	615-297-3809	2782 Main, Nash, TN	37118	John T. Okon	615-123-5589	S2	$500.00	14-Mar-2004

C_NAME = Customer name
C_PHONE = Customer phone
C_ADDRESS = Customer address
C_ZIP = Customer ZIP code

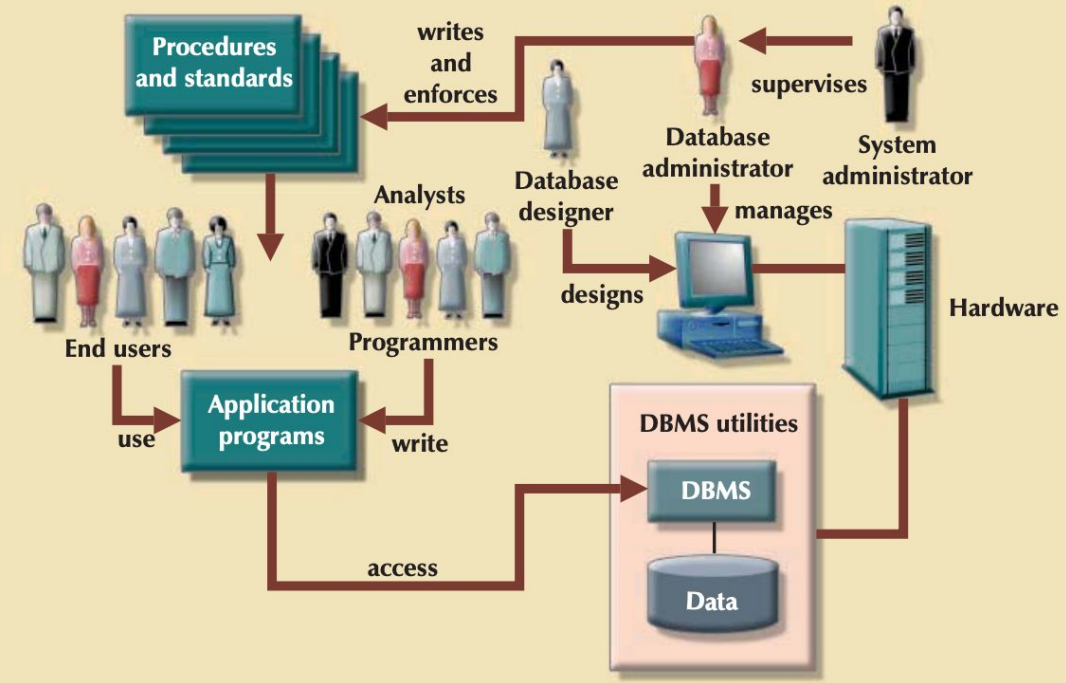
A_NAME = Agent name
A_PHONE = Agent phone
TP = Insurance type
AMT = Insurance policy amount, in thousands of $
REN = Insurance renewal date

The Database System Environment

- Database system is composed of 5 main parts:
    1. Hardware
    2. Software
        - Operating system software
        - DBMS software
        - Application programs and utility software
    3. People
    4. Procedures
    5. Data

The Database System Environment

FIGURE 1.7 The database system environment



DBMS Functions

Performs functions that guarantee integrity and consistency of data

- Data dictionary management
  - defines data elements and their relationships
- Data storage management
  - stores data and related data entry forms, report definitions, etc.
- Data transformation and presentation
  - translates logical requests into commands to physically locate and retrieve the requested data

DBMS Functions

- Security management
  - enforces user security and data privacy within database
- Multi-user access control
  - creates structures that allow multiple users to access the data
- Backup and recovery management
  - provides backup and data recovery procedures

DBMS Functions

- Data integrity management
  - promotes and enforces integrity rules to eliminate data integrity problems
- Database access languages and application programming interfaces
  - provides data access through a query language
- Database communication interfaces
  - allows database to accept end-user requests within a computer network environment

Data Modelling

Data Modeling and Data Models

Data models

- Relatively simple representations, usually graphical, of complex real-world data structures
- Facilitate interaction among the designer, the applications programmer, and the end user

Data Model Basic Building Blocks

- Entity -anything about which data are to be collected and stored
- Attribute -a characteristic of an entity
- Relationship -describes an association among entities
  - –One-to-many (1:M) relationship
  - –Many-to-many (M:N or M:M) relationship
  - –One-to-one (1:1) relationship
- Constraint -a restriction placed on the data

Business Rules

- How do modelers go about modeling data? –By understanding Business Rules!
- It is a **Brief, precise, and unambiguous descriptions of a policies, procedures, or principles within a specific organization**
  - –E.g. a student may take up to 21 credits at a time
  - –E.g. each computer account may only be used by one student
- Any organization that stores and uses data to generate information has business rules (whether they know it or not)
- Business rules are a description of the organization's operations
- They help to create and enforce actions within that organization's environment

Business Rules and Data Modeling

- May identify entities and/or types of relationships
  - E.g. “E.g. each computer account may only be used by one student (the account owner)”
    - identifies the STUDENT and ACCOUNT entities (if we didn’t already have them) and
    - helps to identify that the relationship is 1:1
    - (to fully get that we need another rule how many accounts may a student have?)
- Some business rules don’t impact data modeling (but may impact application development)
  - E.g. students cannot sign up for more than one section of the same course (in same semester (may repeat))

Discovering Business Rules

Sources of Business Rules:

- Company managers
- Policy makers
- Department managers
- Written documentation
  - –Procedures
  - –Standards
  - –Operations manuals
- Direct interviews with end users

Frequently must resolve conflicts between different sources

Translating Business Rules into Data Model Components

Generally, nouns translate into entities
Verbs translate into relationships among entities
Relationships are bi-directional

For example,
the business rule: “a customer may generate many invoices”

- Customer and invoice are nouns(objects) should be represented by their respective entities.
- There is a “generate” relationship between customer and invoice.

The Evolution of Data Models

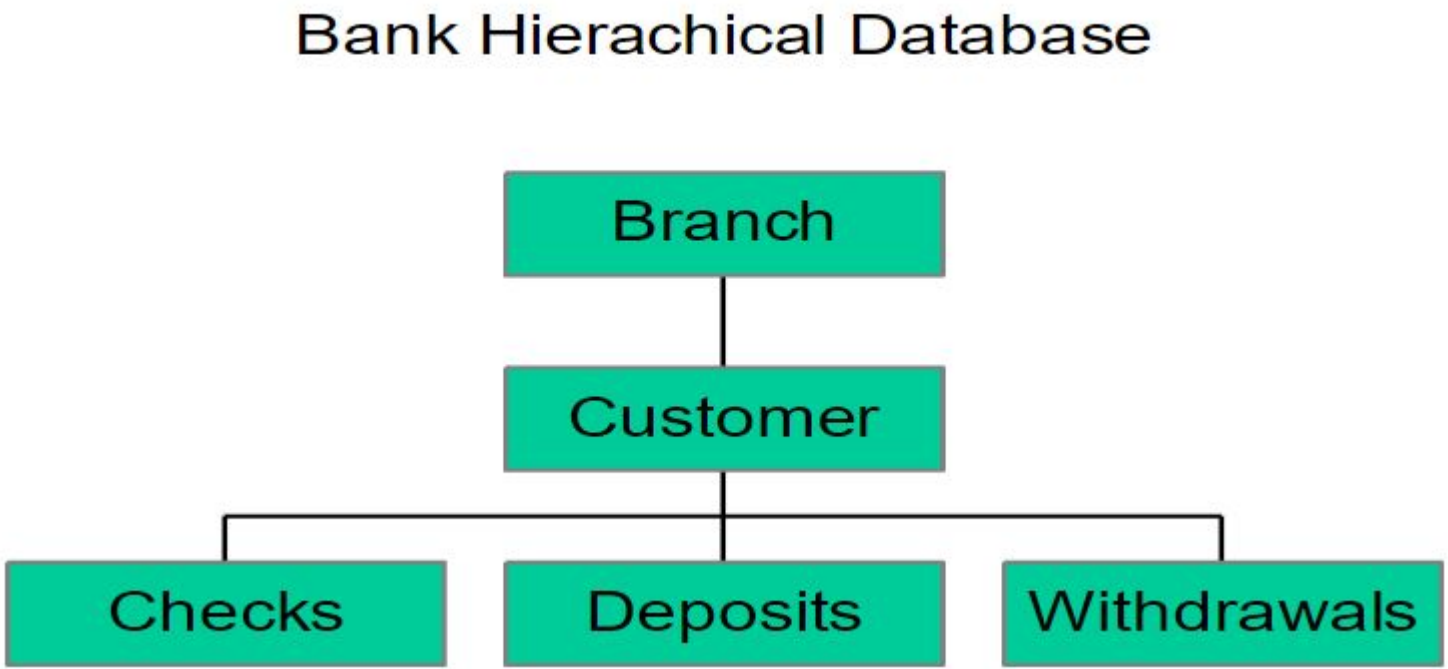
The Evolution of Data Models

- Hierarchical Data Model
- Network Data Model
- Relational Data Model
- Entity relationship Data Model
- Object oriented (OO) Data Model

The Hierarchical Model

- Developed in the 1960s to manage large amounts of data for complex manufacturing projects. (Apollo rocket in 1969)
- Basic logical structure is represented by an upside-down “tree”.
- It contains levels or segments. A segment is the equivalent of a file system’s record type.
- The higher layer is the parent, and the segment beneath is the child.
- Each parent in the hierarchy can have multiple children, but each child has only one parent. It is 1:M relationship.

A Hierarchical Structure

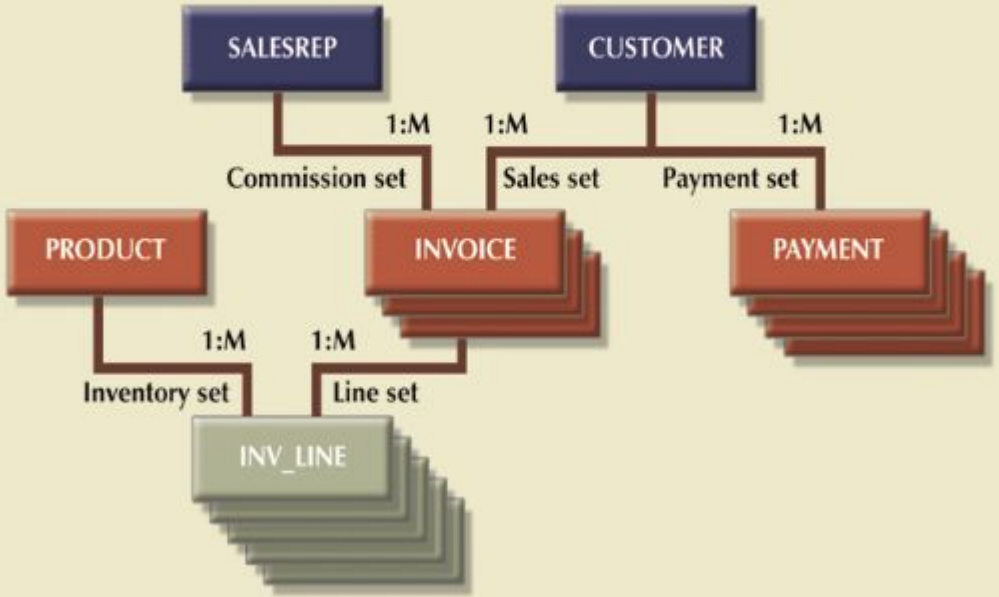


The Network Model

- Created to
  - Represent complex data relationships more effectively than hierarchical model
  - Improve database performance
  - Impose a database standard
- The user perceives the network database as a collection of records in 1:M relationships.
- The network model allows a record to have more than one parent.

The Network Model

FIGURE 2.2 A network data model



The Network Model- Important concepts

- The schema, which is the **conceptual organization of the entire database as viewed by the database administrator**.
- The subschema, which defines the **portion of the database "seen" by the application programs** that actually produce the desired information from the data contained within the database.
- A data management language (DML), which defines the **environment in which data can be managed and to work with the data in the database.**
- A schema data definition language (DDL), which **enables the database administrator to define the schema components**.

The Network Model - Disadvantages

- Too cumbersome
- The lack of ad hoc query capability put heavy pressure on programmers
- Any structural change in the database could produce havoc in all application programs that drew data from the database
- Many database old-timers can recall the interminable information delays

The Relational Model

- This type of model designs the data in the form of rows and columns within a table.
- A relational model uses tables for representing data and in-between relationships.
- This model was initially described by Edgar F. Codd, in 1969.
- The relational data model is the widely used model which is primarily used by commercial data processing applications.

The Relational Model

FIGURE 2.3 Linking relational tables

Database name: Ch02_InsureCo Table name: AGENT (first six attributes)

AGENT_CODE	AGENT_LNAME	AGENT_FNAME	AGENT_INITIAL	AGENT_AREACODE	AGENT_PHONE
501	Alby	Alex	B	713	228-1249
502	Hahn	Leah	F	615	882-1244
503	Okon	John	T	615	123-5589

Link through AGENT_CODE

Table name: CUSTOMER

CUS_CODE	CUS_LNAME	CUS_FNAME	CUS_INITIAL	CUS_AREACODE	CUS_PHONE	CUS_RENEW_DATE	AGENT_CODE
10010	Ramas	Alfred	A	615	844-2573	05-Apr-2006	502
10011	Dunne	Leona	K	713	894-1238	16-Jun-2006	501
10012	Smith	Kathy	W	615	894-2285	29-Jan-2007	502
10013	Olowski	Paul	F	615	894-2180	14-Oct-2006	502
10014	Orlando	Myron		615	222-1672	28-Dec-2006	501
10015	O'Brian	Amy	B	713	442-3381	22-Sep-2006	503
10016	Brown	James	G	615	297-1228	25-Mar-2006	502
10017	Williams	George		615	290-2556	17-Jul-2006	503
10018	Farriss	Anne	G	713	382-7185	03-Dec-2006	501
10019	Smith	Olette	K	615	297-3809	14-Mar-2006	503

The Relational Model

- Reason for popularity is it's powerful and flexible query language.
- Structured Query Language (SQL) allows the user to specify what must be done without specifying how it must be done.

The Entity Relational (ER) Model

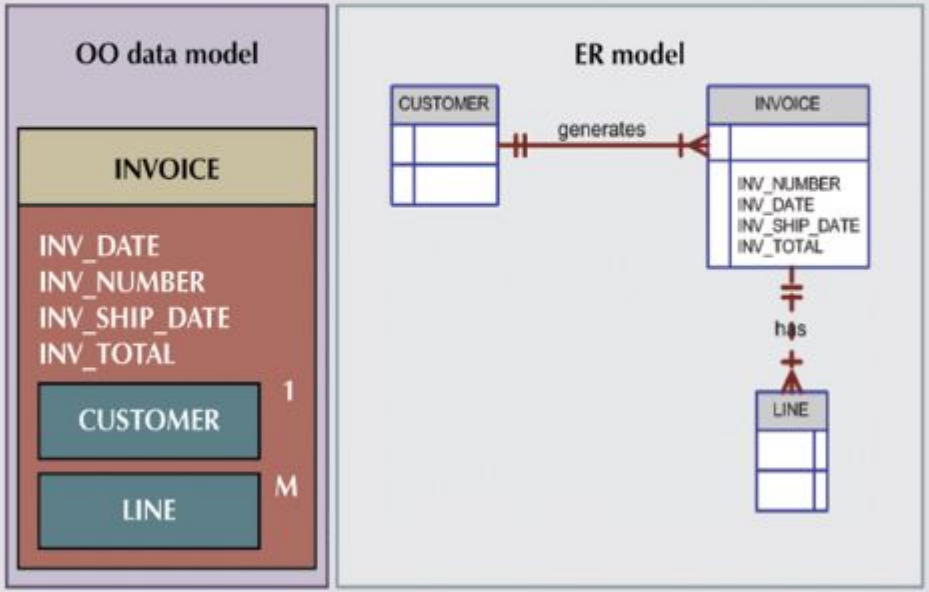
- An ER model is the logical representation of data as objects and relationships among them.
- These objects are known as entities, and relationship is an association among these entities.
- This model was designed by Peter Chen and published in 1976 papers. It was widely used in database designing.

The Object Oriented Model

- Data and their relationships are contained in a single structure which is referred to as an object in this data model.
- In this, real-world problems are represented as objects with different attributes.
- All objects have multiple relationships between them.
- It is a combination of Object Oriented programming and a Relational Database Modeling the object-oriented approach were developed.

The Object Oriented Model

FIGURE 2.7 A comparison of the OO model and the ER model



Degree of Data Abstraction

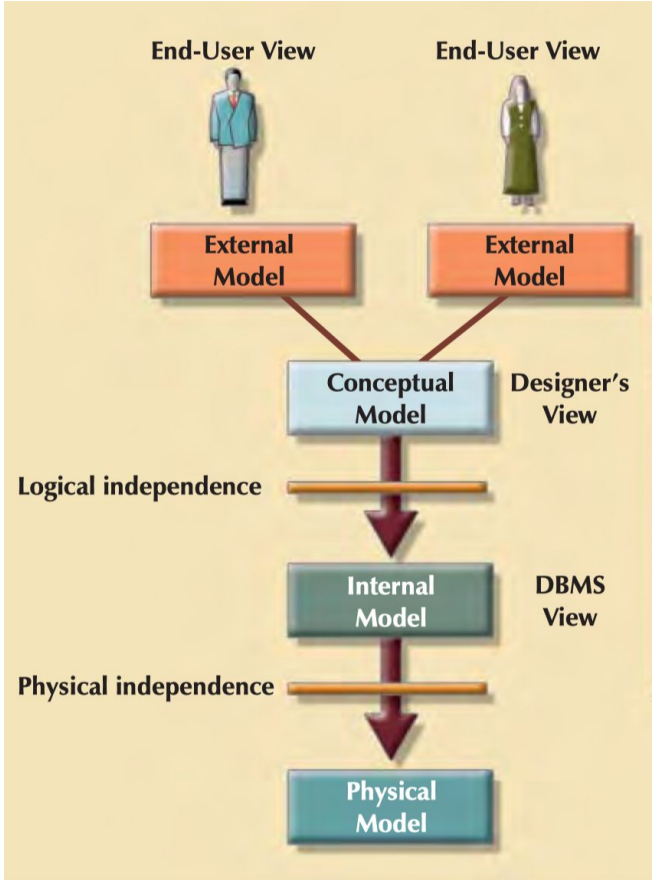
Data abstraction levels

- The American National Standards Institute (ANSI) Standards Planning and Requirements Committee (SPARC) defined a framework for data modeling based on degrees of data abstraction.
- The ANSI/SPARC architecture (as it is often referred to) defines three levels of data abstraction (Three level schema architecture in DBMS):
  - external level
  - conceptual level
  - internal level

Data abstraction levels

- The three-schema architecture divides the database into three-level used to create a separation between the physical database and the user application.

Data abstraction levels



The External Model/ View Model

- This is the highest level of database abstraction.
- This level provides different views of the same database for a specific user or a group of users.
- End users' view of the data environment.
  - End users refers to people who use the application programs to manipulate the data and generate information
  - End user's application has a specific business unit focus. (sales, finance, and marketing)
  - Each business unit is subject to specific constraints and requirements, and each one uses a data subset of the overall data in the organization.
- An external view provides a powerful and flexible security mechanism by hiding the parts of the database from a particular user.

Advantage of the External Model

The use of external views representing subsets of the database has some important advantages:

- It makes it easy to identify specific data required to support each business unit's operations.
- It makes the designer's job easy by providing feedback about the model's adequacy.
- Specifically, the model can be checked to ensure that it supports all processes as defined by their external models, as well as all operational requirements and constraints.
- It helps to ensure security constraints in the database design. Damaging an entire database is more difficult when each business unit works with only a subset of data.
- It makes application program development much simpler.

The Conceptual Model/ Logical level

- This level describes the structure of the whole database.
- It acts as a middle layer between the physical storage and user view.
- It explains what data to be stored in the database, what the data types are, and what relationship exists among those data.
- There is only one conceptual schema per database.
- It acts as a middle layer between the physical storage and user view.
- It explains what data to be stored in the database, what the data types are, and what relationship exists among those data.
- There is only one conceptual schema per database.

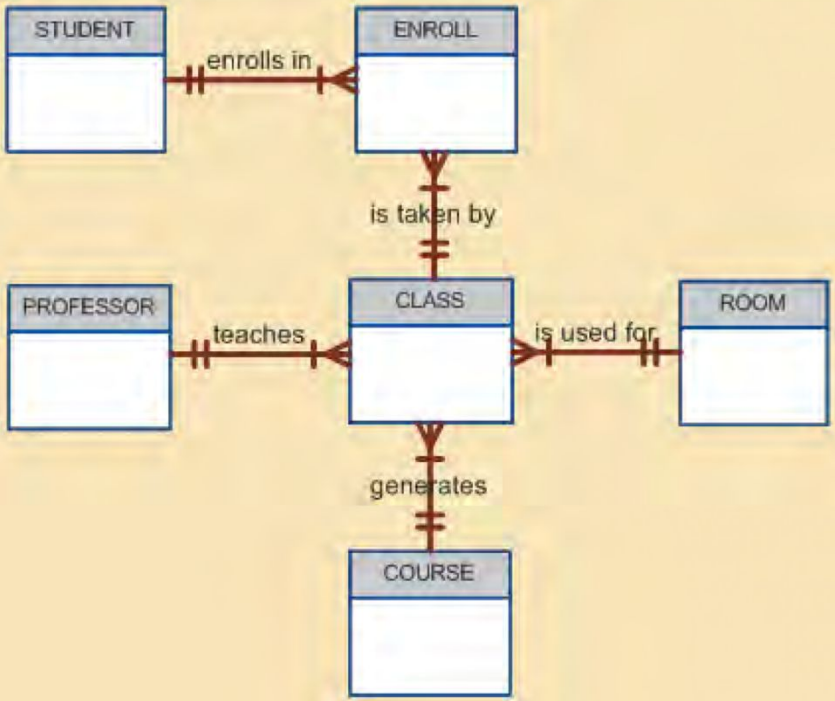
The Conceptual Model

- Graphically represented by an Entity Relational Diagram (ERD) to integrate all external views into a single view.
- Represents a global view of the entire database as viewed by the entire organization
  (entities, relationships, constraints, and processes) into a single global view of the data in the enterprise)
- Also known as a conceptual schema.
- The most widely used conceptual model is the ER model. It illustrates in graphically using ERD (basic database blueprint).
- The conceptual model is independent of both software and hardware.

The Conceptual Model

-

FIGURE 2.8 Conceptual model for Tiny College



The Internal Model/ Physical level

- This is the lowest level of database abstraction.
- It describes how the data is stored in the database and provides the methods to access data from the database.
- It allows viewing the physical representation of the database on the computer system.
- The interface between the conceptual and internal schema identifies how an element in the conceptual schema is stored and how it may be accessed.
- It is one which is closest to physical storage.
- The internal schema not only defines different stored record types, but also specifies what indices exist, how stored fields are represented.

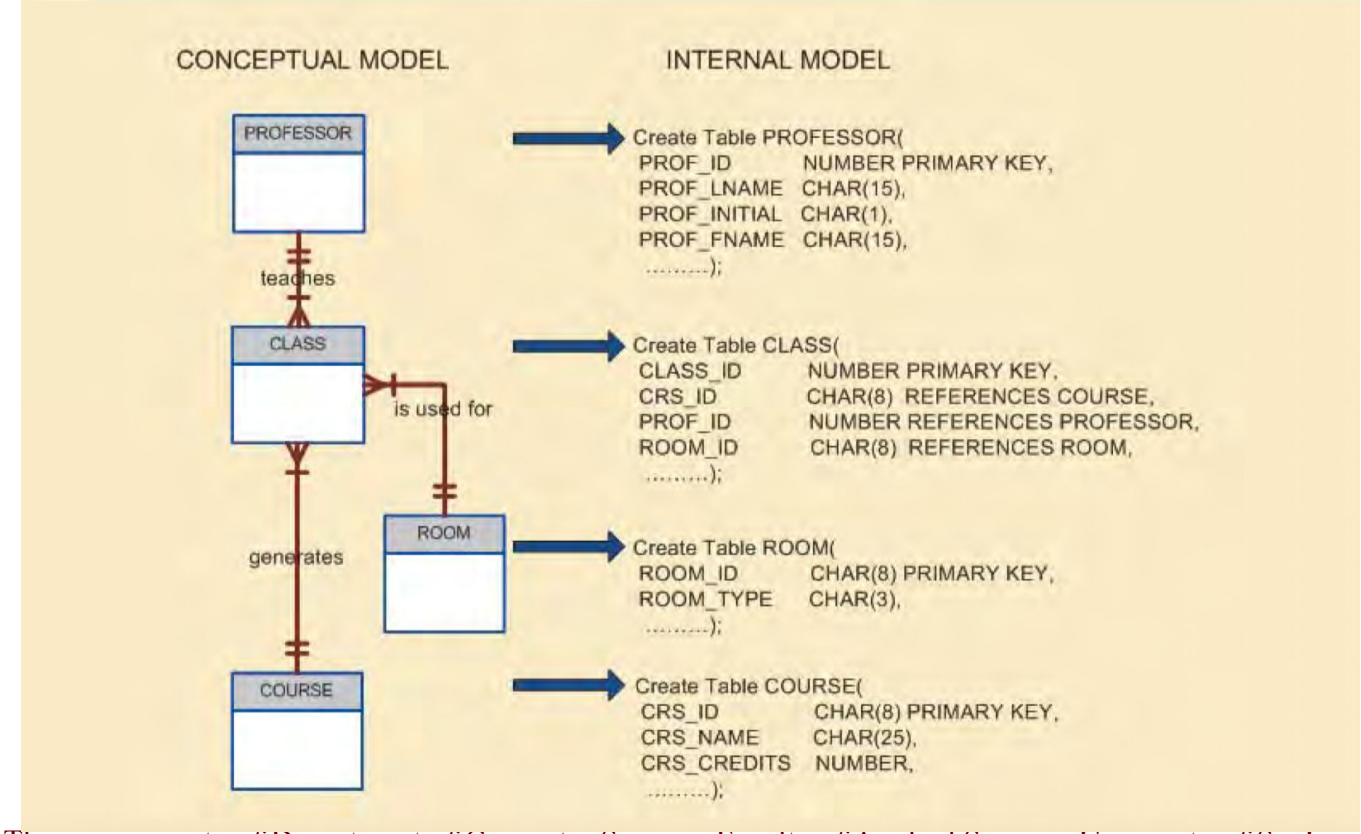
The Internal Model

- Internal model is the representation of the database as “seen” by the DBMS.
- Once a specific DBMS has been selected, the internal model maps the conceptual model to the DBMS.
- The internal model requires the designer to match the conceptual model’s characteristics and constraints to those of the selected implementation model.
- An internal schema depicts a specific representation of an internal model, using the database constructs supported by the chosen database.
  Ex: The entities in the conceptual model are mapped to tables in the relational model.
- The internal schema is expressed using Structured Query Language (SQL), the standard language for relational databases.

The Internal Model

-

FIGURE 2.9 Internal model for Tiny College



The Internal Model

- The internal model is **software-dependent**.
- **logical independence -** if you can change the internal model without affecting the conceptual model.
- The internal model is **hardware-independent.**

The Physical Model

- The physical model operates at the lowest level of abstraction.
- It describing the way data are saved on storage media such as disks or tapes.
- It is software- and hardware- dependent.
- The relational model does not require the designer to be concerned about the data's physical storage characteristics, the implementation of a relational model may require physical-level fine-tuning for increased performance.
- **Physical independence** - When you can change the physical model without affecting the internal model, you have physical independence.

Level of Abstractions

Conceptual/ Internal Mapping

- The Conceptual/ Internal Mapping lies between the conceptual level and the internal level. Its role is to define the correspondence between the records and fields of the conceptual level and files and data structures of the internal level.

External/ Conceptual Mapping

- The external/Conceptual Mapping lies between the external level and the Conceptual level. Its role is to define the correspondence between a particular external and the conceptual view.

Discussion Questions

- What is a relationship, and what three types of relationships exist?
- What is logical independence?
- What is physical independence?
- Explain below terms
  - Database administrator
  - Database designer
  - End user

Next Lesson

- Data Modeling