ACS-1803 Introduction to Information Systems

Instructor: Kevin Robertson

Data Management

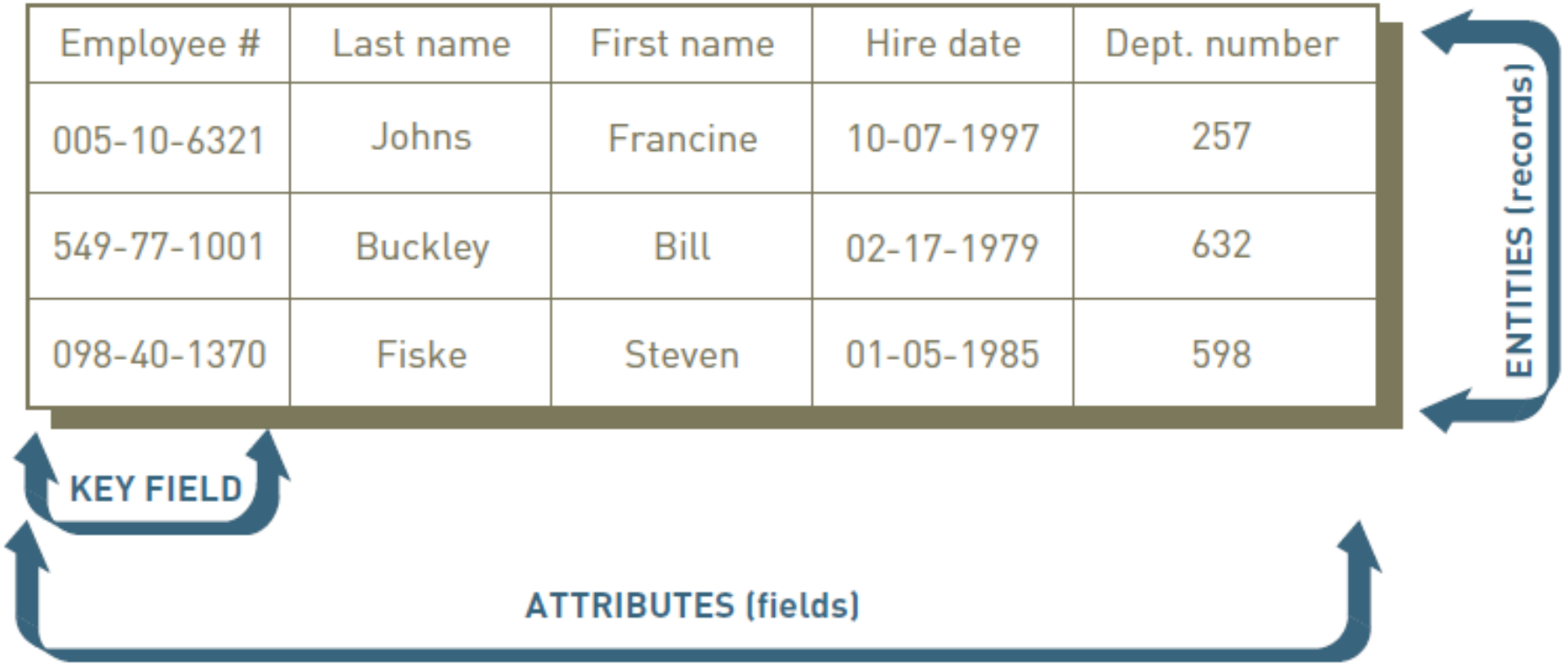
Part 2

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

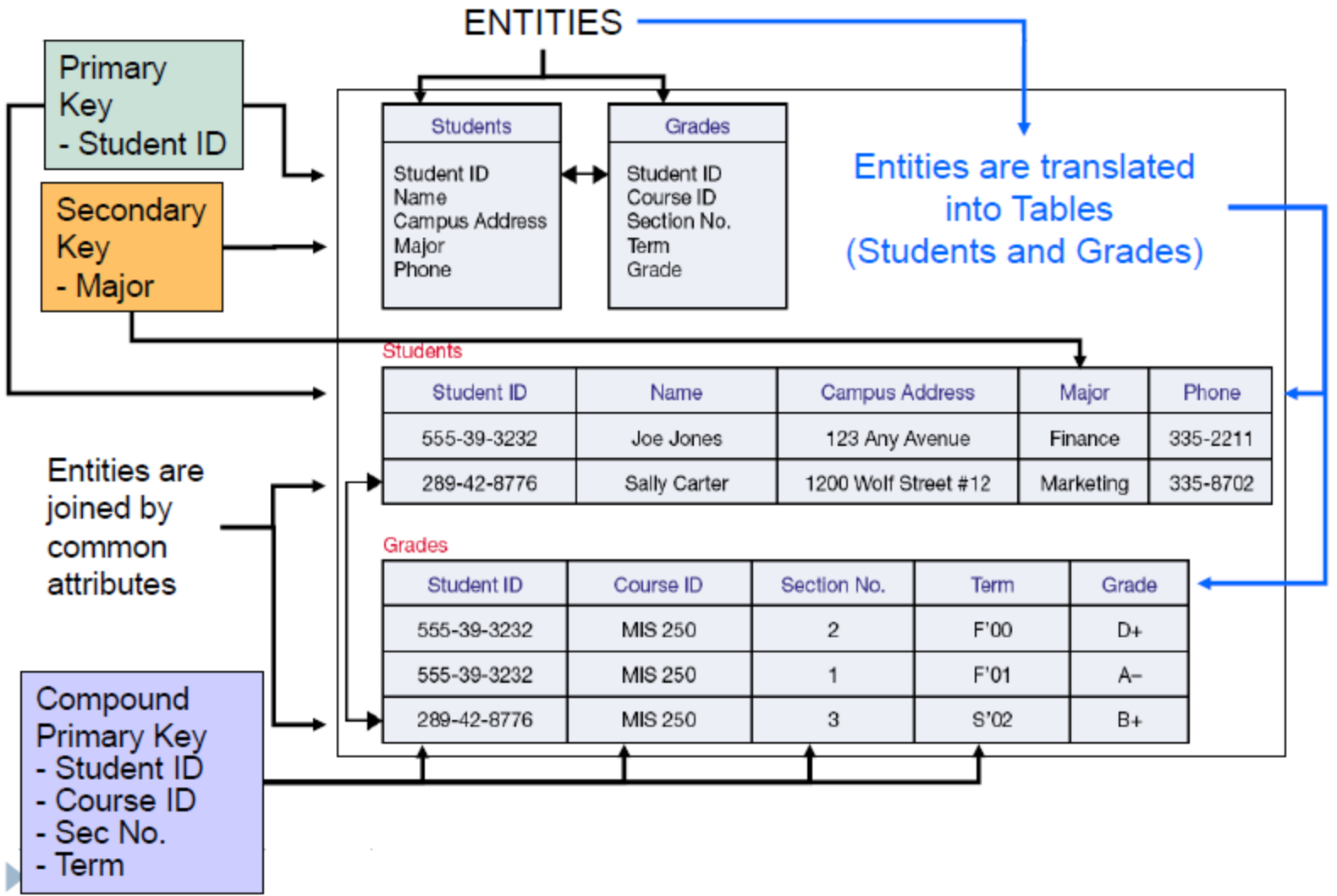
Data Entities, Attributes, and Items

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



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

Data Entities, Attributes, Items, Keys



Data Entities, Attributes, and Items

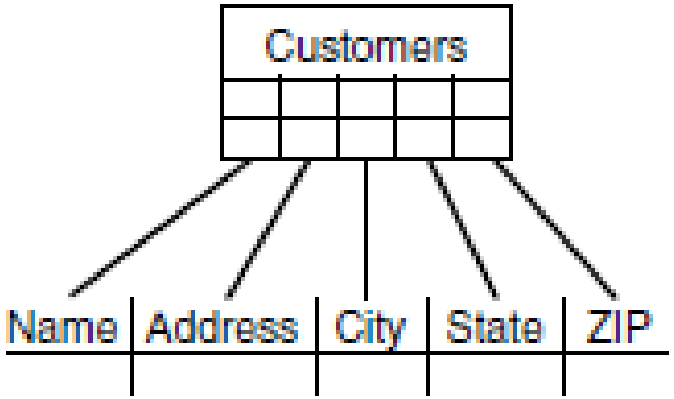
- Normalization
  - Process of streamling complex groups of data to:
    - Minimize redundant data elements.
    - Minimize awkward many-to-many relationships.
    - Increase stability and flexibility.

- Referential integrity rules
  - Used by relational databases to ensure that relationships between coupled tables remain consistent.
  - For example: when one table has a foreign key that points to another table, you may not add a record to the table with foreign key unless there is a corresponding record in the linked table.

Data Entities, Attributes, and Items

Columns

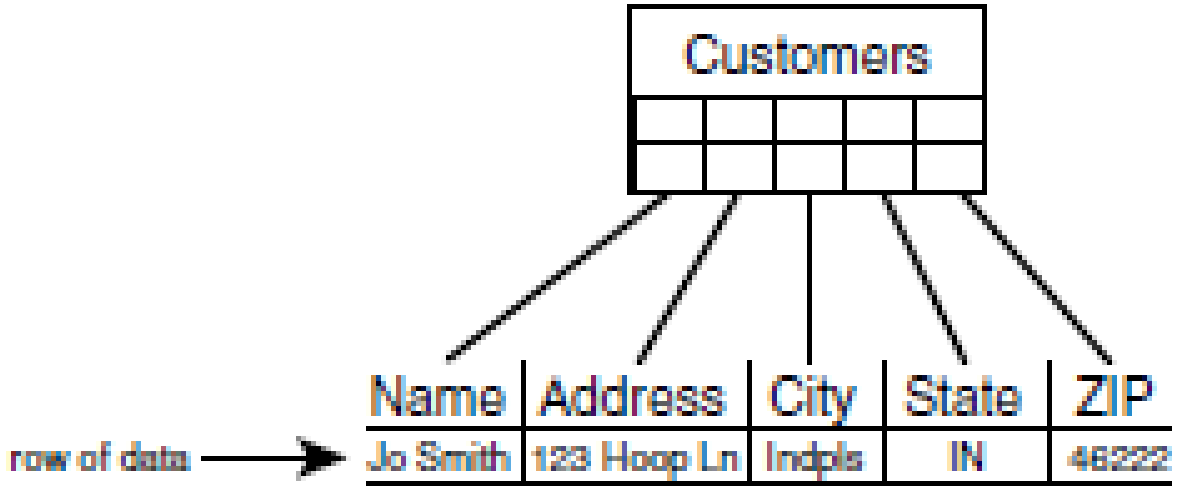
A *column*, or *field*, is a specific category of information that exists in a table. A column is to a table what an attribute is to an entity. In other words, when a business model is converted into a database model, entities become tables and attributes become columns. A column represents one related part of a table and is the smallest logical structure of storage in a database. Each



Data Entities, Attributes, and Items

Rows

A *row* of data is the collection of all the columns in a table associated with a single occurrence. Simply speaking, a row of data is a single record in a table. For example, if there are 25,000 book titles with which a bookstore deals, there will be 25,000 records, or rows of data, in the book titles table once the table is populated. The number of rows within the table will obviously change as books' titles are added and removed. See Figure 1.7 for an illustration of a



Data Entities, Attributes, and Items

Three types of table relationships that can be derived are as follows:

- One-to-one—One record in a table is related to only one record in another table.
- One-to-many—One record in a table can be related to many records in another table.
- Many-to-many—One record in a table can be related to one or more records in another table, and one or more records in the second table can be related to one or more records in the first table.

Data Entities, Attributes, and Items

One-to-one

Table 1	
ID	VALUE
1	A
2	B
3	C

Table 2	
FK	VALUE
1	A
2	B
3	C

One-to-many

Table 1	
ID	VALUE
1	A
2	B
3	C

Table 2	
FK	VALUE
1	A
1	A
2	B

Many-to-many

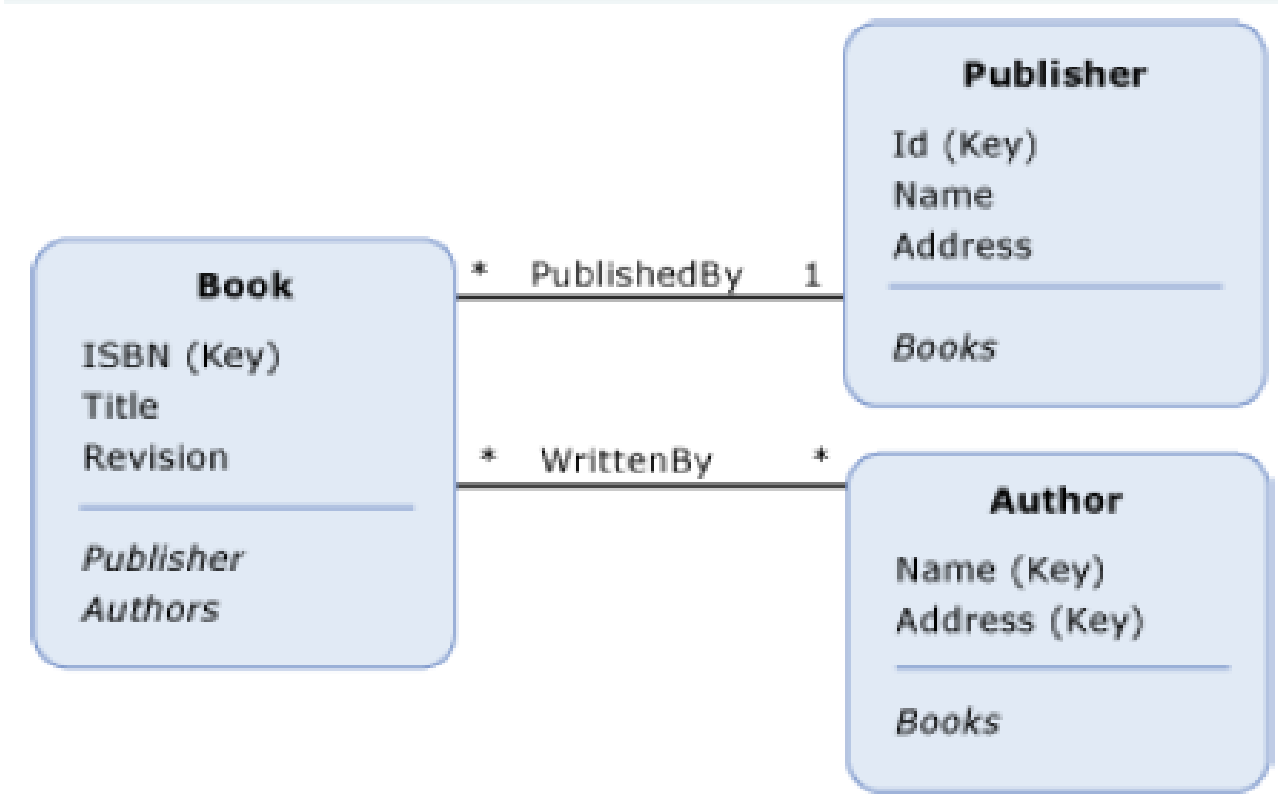
Table 1	
ID	VALUE
1	A
1	A
2	B

Table 2	
FK	VALUE
2	B
1	A
1	A

Designing Databases – Data Model

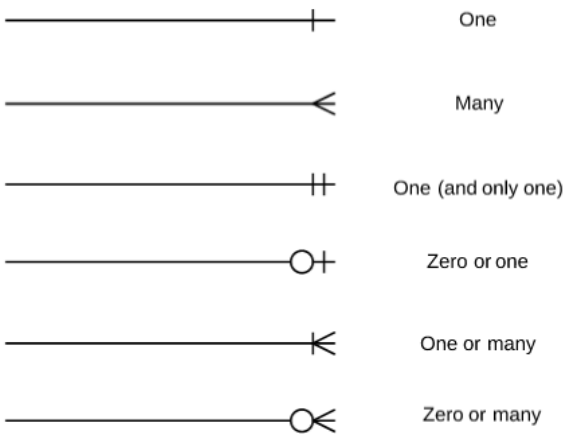
- **Data Model**
  - A map or diagram that represents entities and their relationships
  - Used by Database Administrators and System Designers to design tables with their corresponding associations
  - Modelling tools help the designers construct and validate models before building the information system – therefore improving quality and reducing long terms operational costs

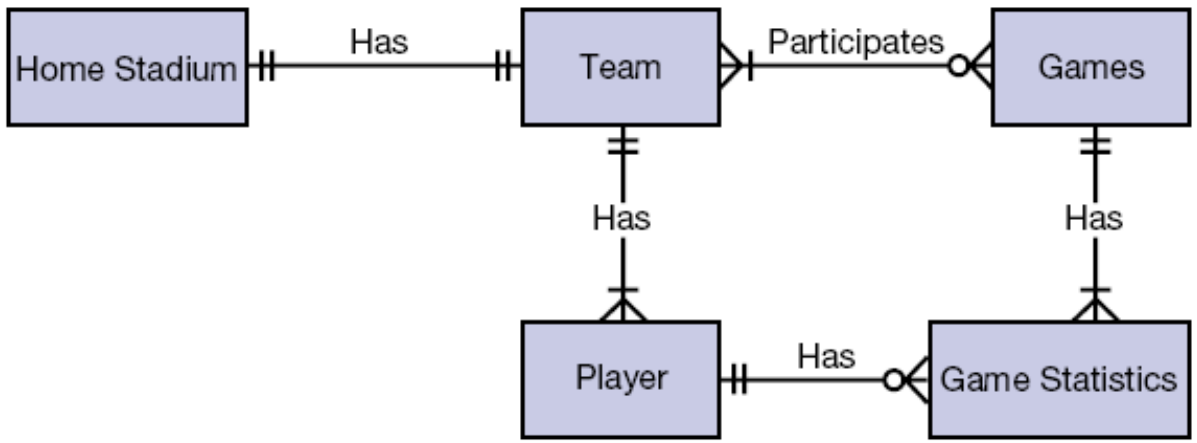
Designing Databases – Data Model



Designing Databases - Associations

- Entity Relationship Diagram (ERD)
  - Diagramming tool used to express entity relationships
  - Very useful in developing complex databases
- Example
  - Each Home Stadium has a Team (One-to-One)
  - Each Team has Players (One-to-Many)
  - Each Team participates in Games
  - For each Player and Game there are Game Statistics



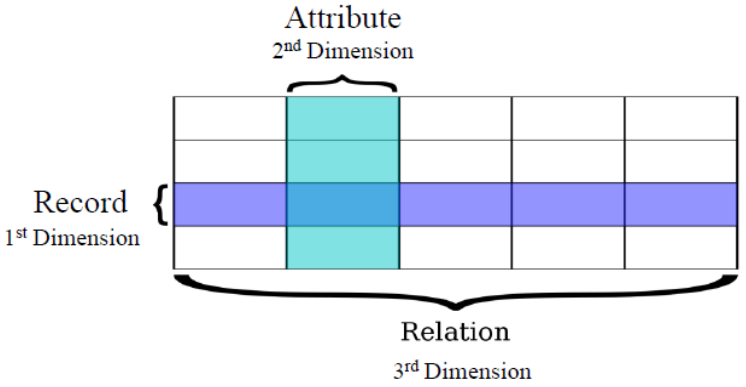


Relational Database

- Data set up as a centralized collection of tables
- Tables are linked by common columns of data
- Tables are designed together to minimize repetition
- This is the most common database structure
- This is the one use in microcomputer databases and many larger ones too.

The Relational Model

- The most common type of database model used today in organizations
- Is a three-dimensional model compared to the traditional two-dimensional database models
  - Rows (first-dimension)
  - Columns (second-dimension)
  - Relationships (third-dimension)
- The third-dimension makes this model so powerful because any row of data can be related to any other row or rows of data



Relational Model – Example 1

customer-name	*social-security*	*customer-street*	*customer-city*	*account-number*
Johnson	192-83-7465	Alma	Palo Alto	A-101
Smith	019-28-3746	North	Rye	A-215
Johnson	192-83-7465	Alma	Palo Alto	A-201
Jones	321-12-3123	Main	Harrison	A-217
Smith	019-28-3746	North	Rye	A-201

account-number	*balance*
A-101	500
A-201	900
A-215	700
A-217	750

One-to-One

Relational Model – Example 2

Department Records

Department No	Dept Name	Location	Dean
Dept A			
Dept B			
Dept C			

One-to-Many

Instructor Records

Instructor No	Inst Name	Title	Salary	Dept No
Inst 1				Dept A
Inst 2				Dept B
Inst 3				Dept C
Inst 4				Dept A

Figure 3.12 With the relational model, we represent these two entities, department and instructor, as two separate tables and capture the relationship between them with a common column in each table.

Data Modeling and Database Characteristics

- When building a database, an organization must consider:
  - ***Content***: What data should be collected and at what cost?
  - ***Access***: What data should be provided to which users and when?
  - ***Logical structure***: How should data be arranged so that it makes sense to a given user?
  - ***Physical organization***: Where should data be physically located?

Data Modeling

- Building a database requires two types of designs:
  - Logical design:
    - Abstract model of how data should be structured and arranged to meet an organization's information needs
  - Physical design:
    - Starts from the logical database design and fine-tunes it for performance and cost considerations
  - Planned data redundancy:
    - Done to improve system performance so that user reports or queries can be created more quickly

Data Modeling (continued)

- Data Model:
  - Diagram of data entities and their relationships
- Enterprise data modeling:
  - Starts by investigating the general data and information needs of the organization at the strategic level
- Entity-relationship (ER) diagrams:
  - Data models that use basic graphical symbols to show the organization of and relationships between data

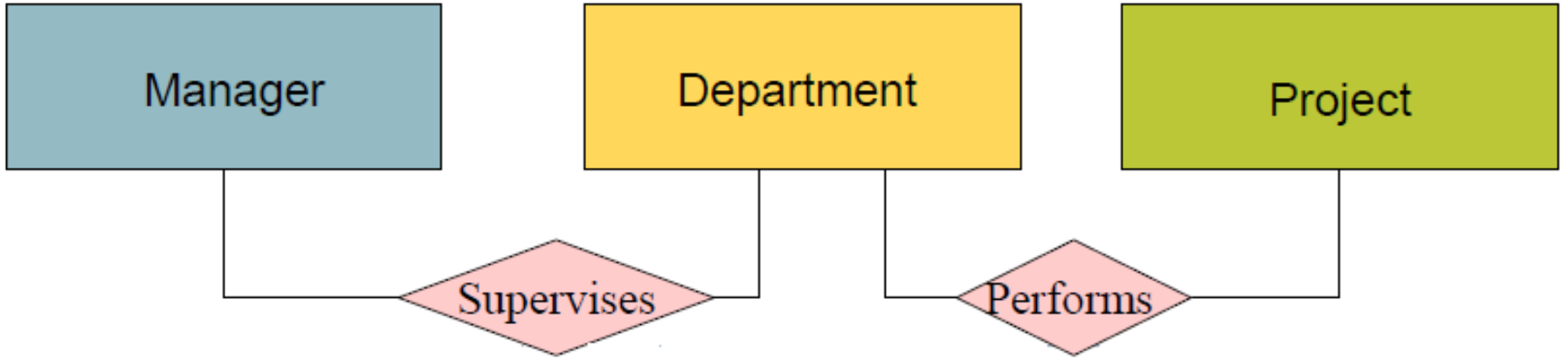
Data Modeling Illustration

First, list data fields

- SSN
- Last Name
- First name
- Hire Date
- Dept. Number
- Dept. name
- Project Number
- Description

Next, identify the relationships and design the tables

Manager	Dept	Project
• SSN •L_Name •F_Name •Hire Date •Dept. Number	•DeptNumber •DeptName	•ProjNumber •ProjDescript



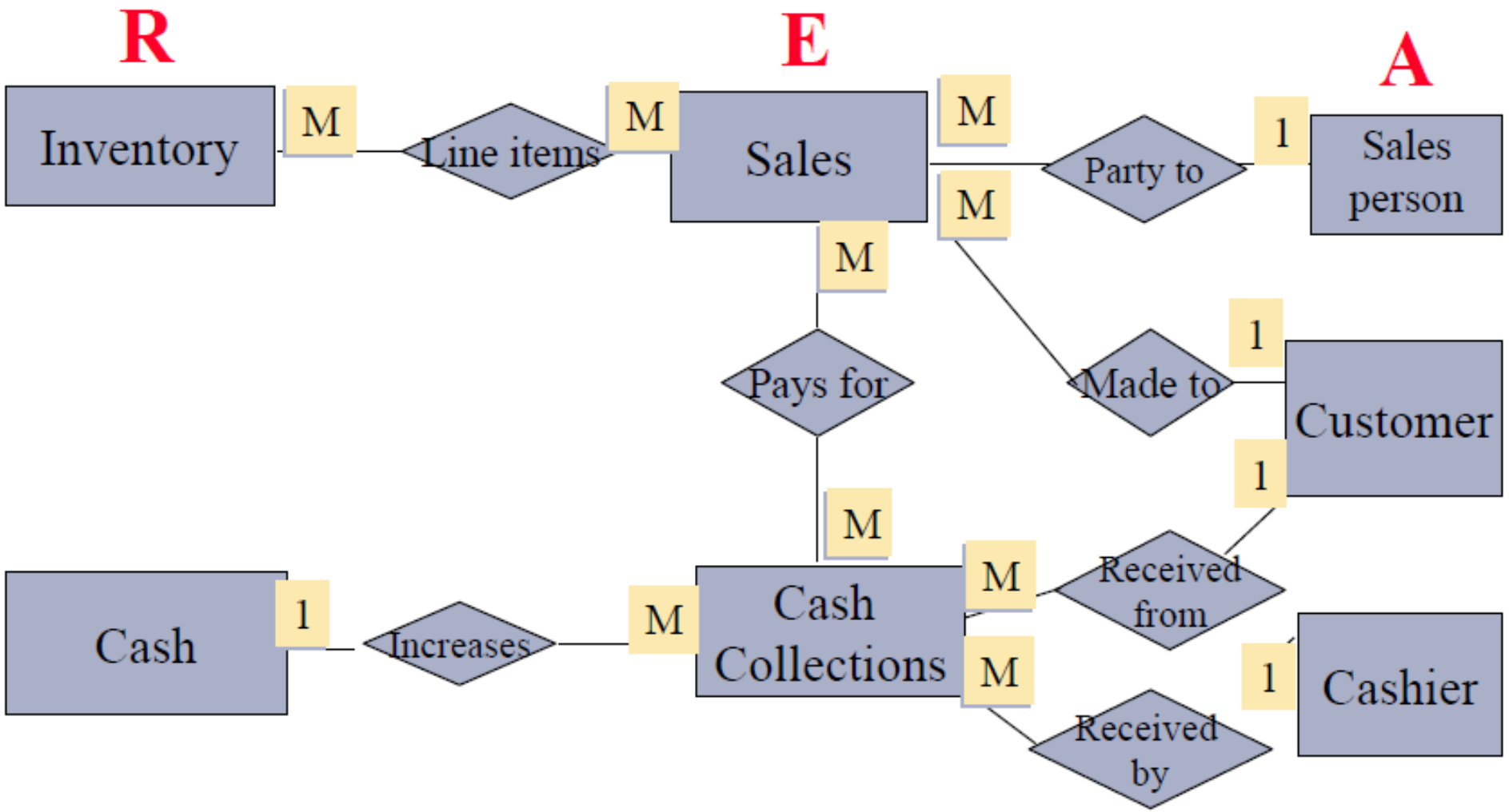
REA Model

- Database where the entities model specific business **RESOURCES**, **EVENTS** and **AGENTS** is called a Resource Event Agent model
- This model is useful in designing databases for various business information systems (e.g. modern accounting systems)
- REA treats the accounting system as a virtual representation of the actual business. In other words, it creates computer objects that directly represent real-world-business objects.

REA Model (continued)

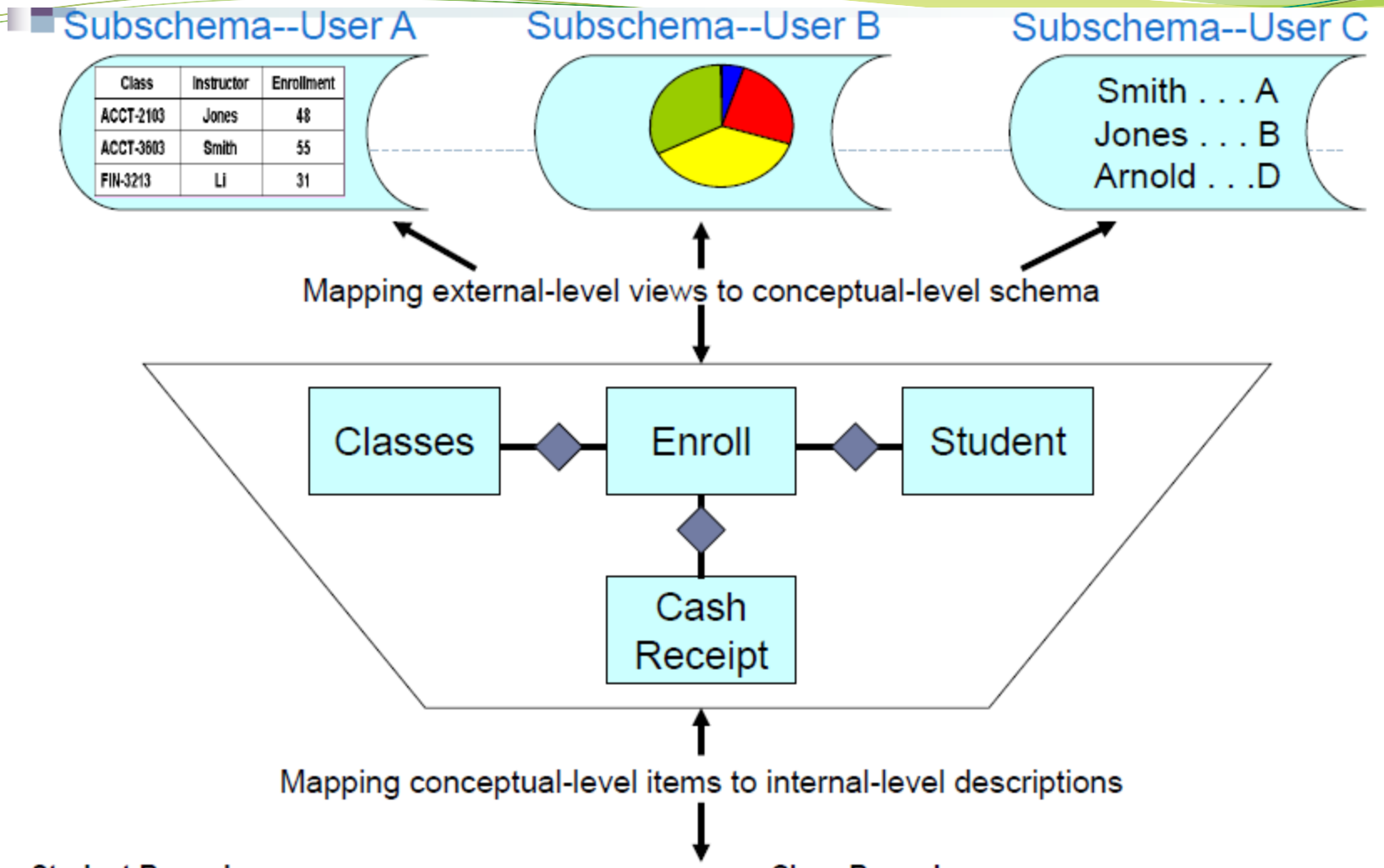
- The real objects included in the REA model are:
  - goods, services or money, i.e., RESOURCES
  - business transactions or agreements that affect resources, i.e., EVENTS
  - people or other human agencies (other companies, etc.), i.e., AGENTS
- These objects contrast with conventional accounting terms such as asset or liability, which are less directly tied to real-world objects. For example, a conventional accounting asset such as goodwill is not an REA resource.

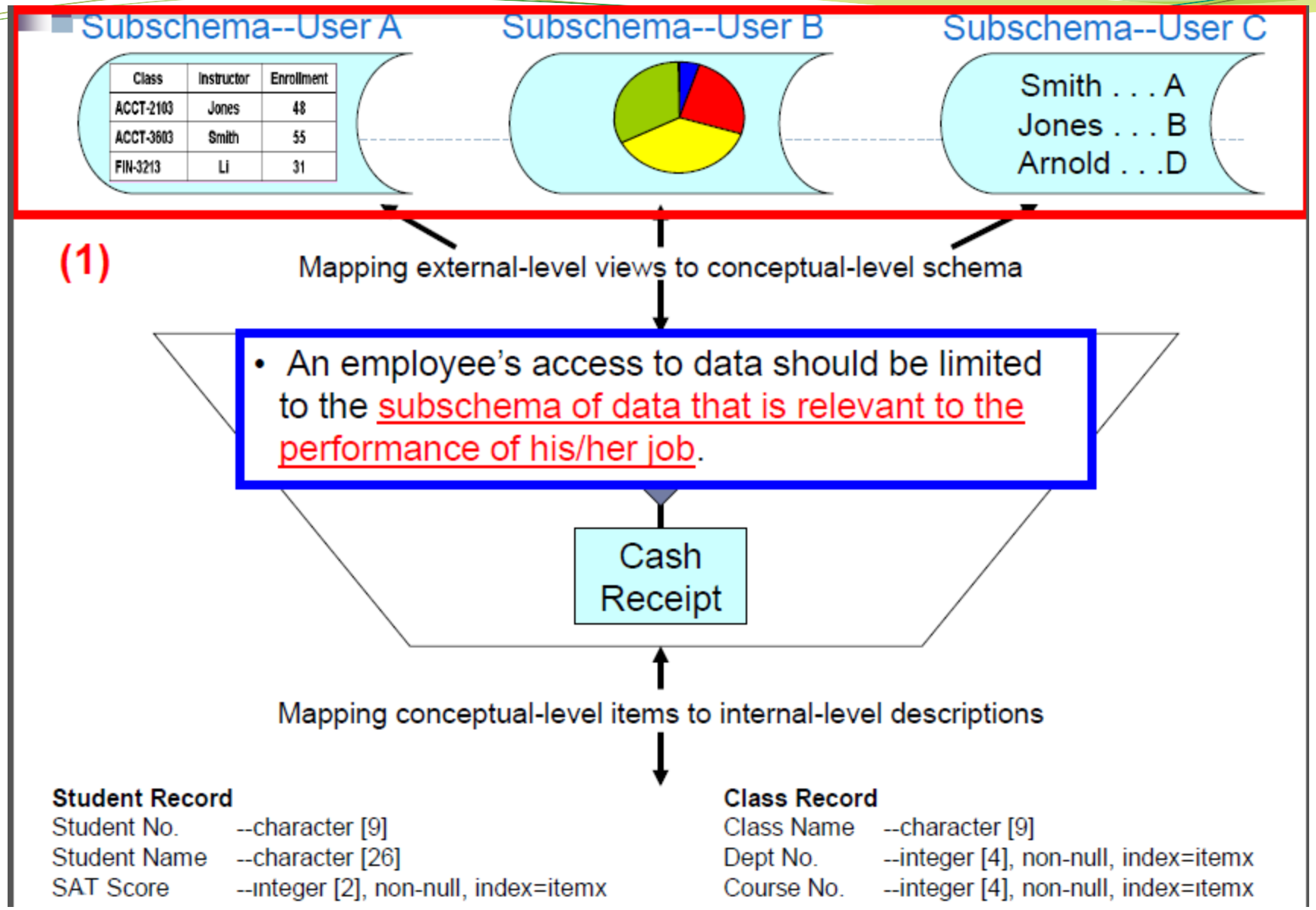
An REA Data Model Example

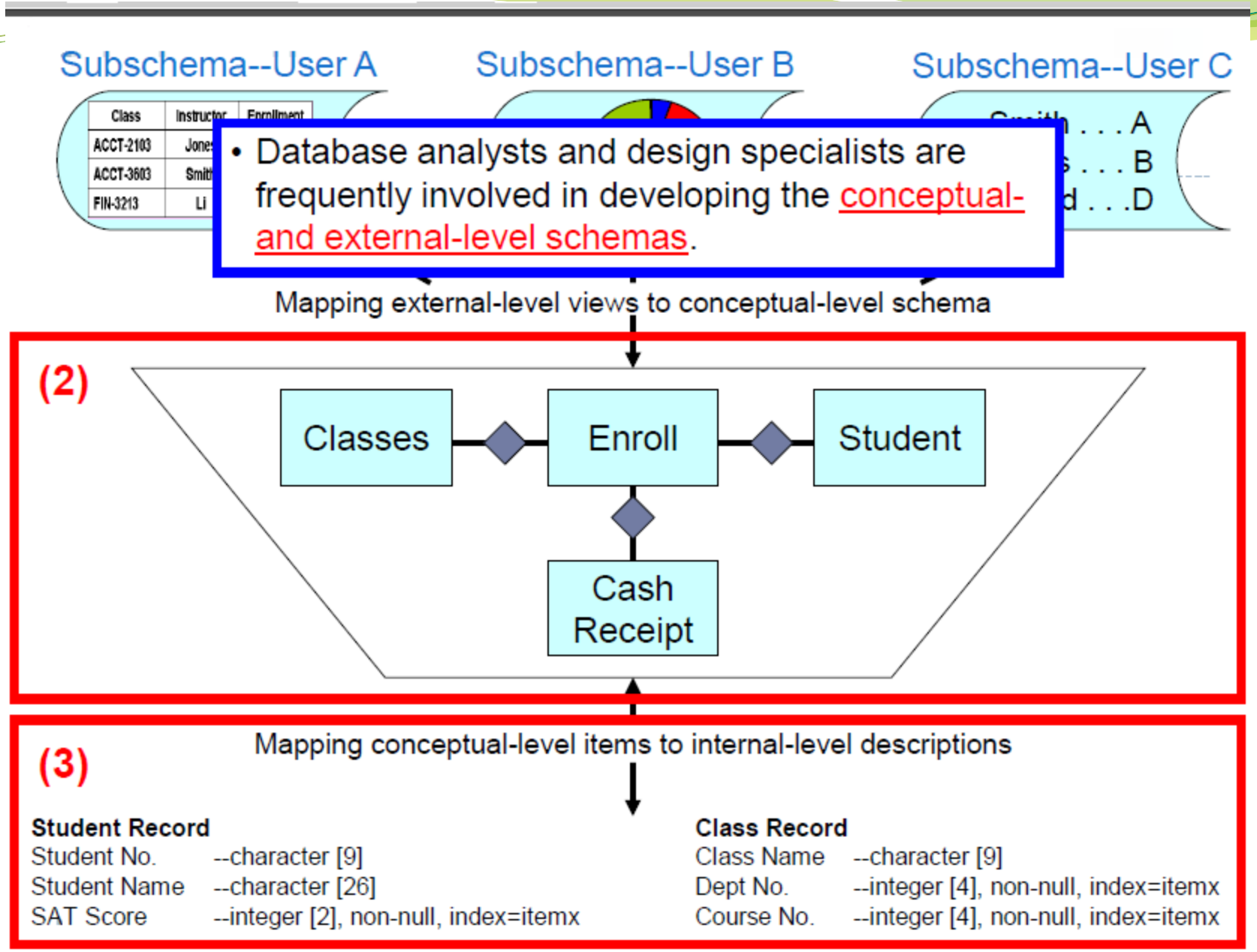


Providing a User View

- **Schema:**
  - Used to describe the entire database
  - Serves as the "blue print" to the design of the dbms and focuses on the relationships between entities.
  - Can be part of the database or a separate schema file
  - Three examples of schemas are shown on the next slide:
    - (1) External Level Subschema (end user view) is focused at the end user level. It serves as a map to the Conceptual View to the design of the DBMS.
    - (2) Conceptual Schema (design view) and the (3) Internal Level Schema (data dictionary view) is used by database analysts and design specialists to map the DBMS.


Subschema--User A
Subschema--User B
Subschema--User C
Class
Instructor
Enrollment
ACCT-2103
Jones
48
ACCT-3603
Smith
55
FIN-3213
Li
31
Smith . . . A
Jones . . . B
Arnold . . .D
Mapping external-level views to conceptual-level schema
Classes
Enroll
Student
Cash Receipt
Mapping conceptual-level items to internal-level descriptions
Student Record
Student No. --character [9]
Student Name --character [26]
SAT Score --integer [2], non-null, index=itemx
Class Record
Class Name --character [9]
Dept No. --integer [4], non-null, index=itemx
Course No. --integer [4], non-null, index=itemx

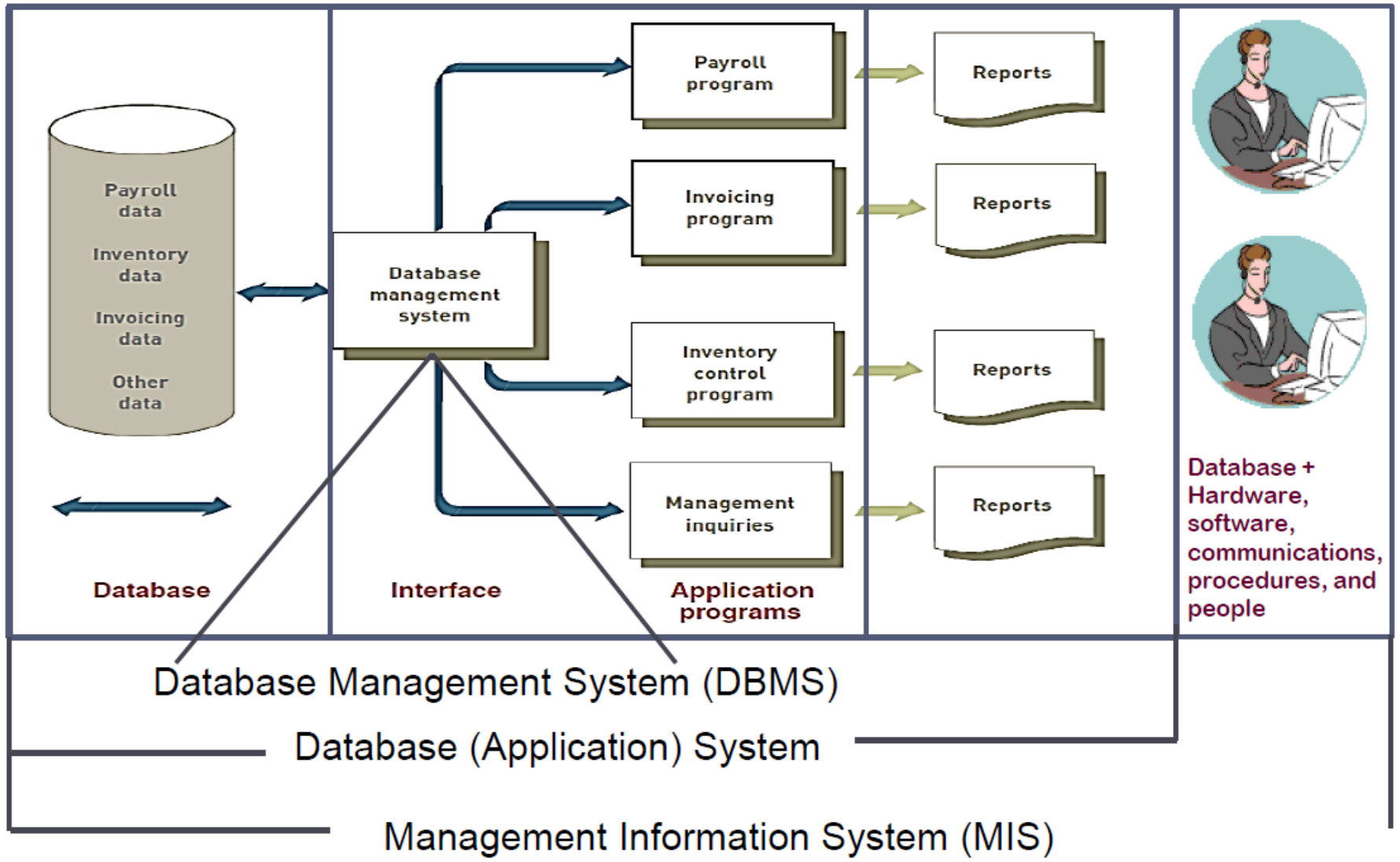

Subschema--User A
Class | Instructor | Enrollment
ACCT-2103 | Jones | 48
ACCT-3603 | Smith | 55
FIN-3213 | Li | 31
Subschema--User B
Subschema--User C
Smith . . . A
Jones . . . B
Arnold . . .D
(1)
Mapping external-level views to conceptual-level schema
• An employee's access to data should be limited to the subschema of data that is relevant to the performance of his/her job.
Cash Receipt
Mapping conceptual-level items to internal-level descriptions
Student Record
Student No. --character [9]
Student Name --character [26]
SAT Score --integer [2], non-null, index=itemx
Class Record
Class Name --character [9]
Dept No. --integer [4], non-null, index=itemx
Course No. --integer [4], non-null, index=itemx


Subschema--User A
Subschema--User B
Subschema--User C
Class
Instructor
ACCT-2103
ACCT-3603
FIN-3213
. . . A
. . . B
. . .D
• Database analysts and design specialists are frequently involved in developing the conceptual- and external-level schemas.
Mapping external-level views to conceptual-level schema
(2)
Classes
Enroll
Student
Cash Receipt
(3)
Mapping conceptual-level items to internal-level descriptions
Student Record
Student No. --character [9]
Student Name --character [26]
SAT Score --integer [2], non-null, index=itemx
Class Record
Class Name --character [9]
Dept No. --integer [4], non-null, index=itemx
Course No. --integer [4], non-null, index=itemx

DBMS | DB Application System | MIS

- The information system needs raw data which is stored on disk as a relational database.
- The relational database is managed by Database Management System (DBMS) software. The system calls the DBMS (behind the scenes) and the DBMS extracts data from the database.
- DBMS is a group of programs used to interface between a database and application programs, or a database and the user.
- 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.

DBMS | DB Application System | MIS



Using MS Access as a DBMS

- Manages the database (e.g. Microsoft Access)
- To use Microsoft Access you must have your design done first.
- When setting up the database as part of the DBMS, the database design schema (Table Structures, Primary Keys, Relationships) is defined before data is entered.
- When setting up the database system, the application interface screens are defined that utilize the DBMS to update the database tables.
- See Supplement Notes *"Using Access Screens"* for detail example

Using MS Access as a DBMS

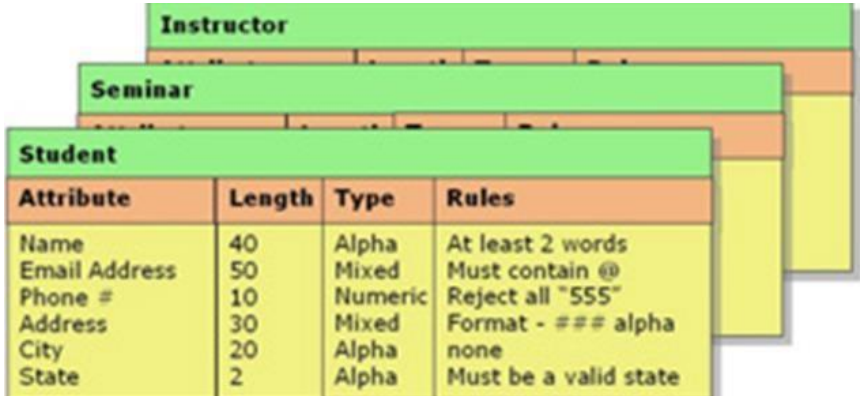
- MS Access can be used **as an information system developer tool**.
- We use it to build a customized Management Information System (MIS) for some specific purpose.
- When setting up the database as part of the DBMS, the database design schema (Table Structures, Primary Keys, Relationships) are defined before data is entered.
- The system will utilize application interface screens that serve as an interface to the DBMS and database tables.
- The system will have its own menus, input screens, output (query) screens, and reports.
- In the background, it will have a database with related tables, and programs that take the raw data from the database and convert it to the required queries and reports.
- It will also have controls, i.e. mechanisms to ensure that the output is correct and the at the data is safe from accidental or deliberate destruction.
- The information system will be set up on computers and a network and will provide and effective method for users to interface with the database without ever knowing they are utilizing MS Access.
- See Supplement Notes "Using Access Screens" for detail example.

Database Management System (DBMS)

- System software that sets up the database structure ('skeleton' on disk according to a certain model, fills the structure with data and retrieves the data to provide meaningful information)
- Parts of dbms:
  - **Data Dictionary**: defines each field and record, explains what each field means and who is authorized to update it
  - **Query Language**: used to extract data that satisfy certain criteria from a database. Used to develop query retrieval commands and reports – which are the two main forms of output from a DBMS

Data Dictionary

- Is a document that database designers prepare to help individuals enter data
- Provides several pieces of information about each attribute in the database including:
  - Name
  - Key (is it a key or part of a key?)
  - Data Type (date, alphanumeric, numeric, etc.)
  - Valid Value (the format or numbers allowed)

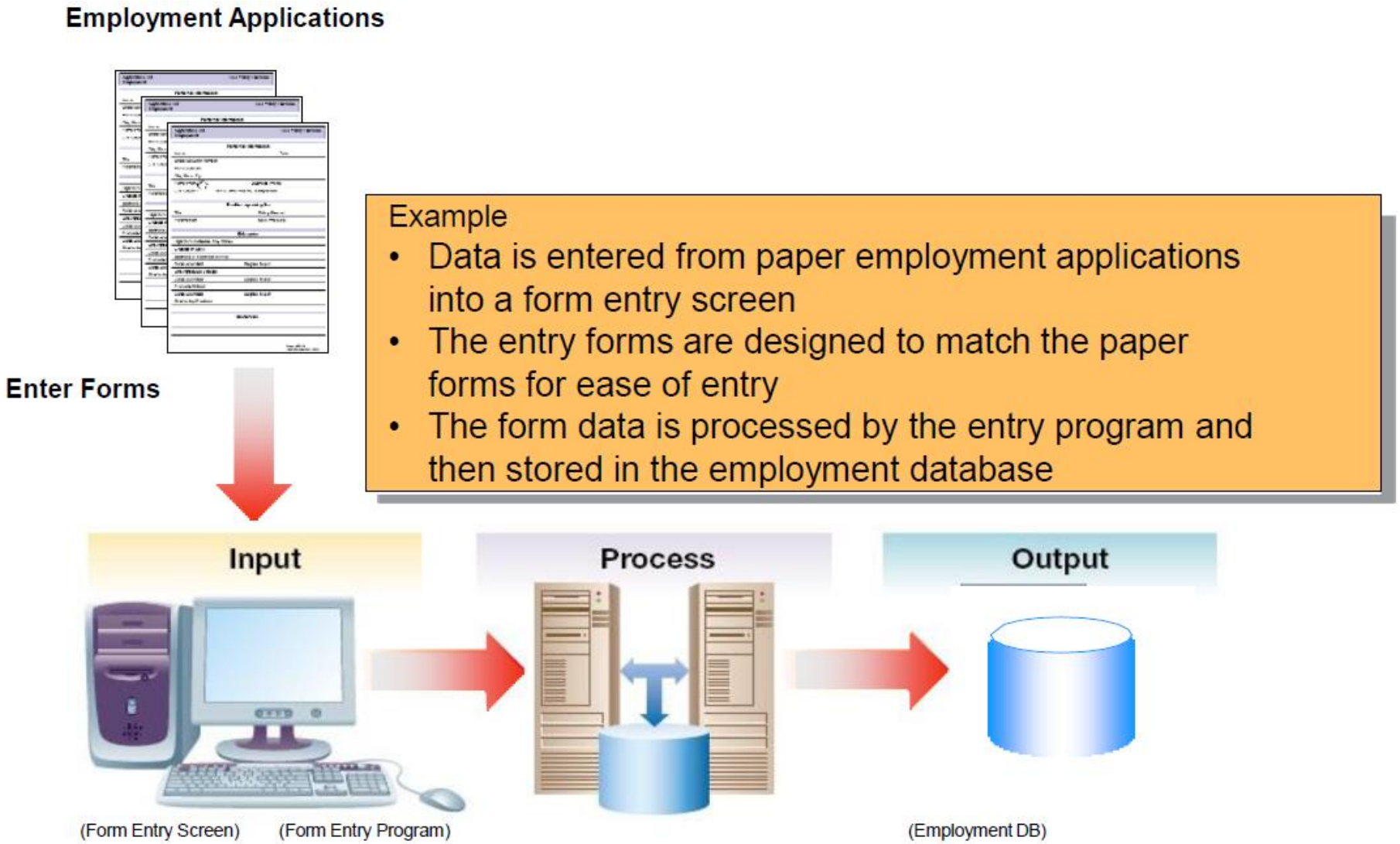


- Can be used to enforce Business Rules which are captured by the database designer to prevent illegal or illogical values from entering the database. (e.g. who has authority to enter certain kinds of data in specific files)

DBMS Functions

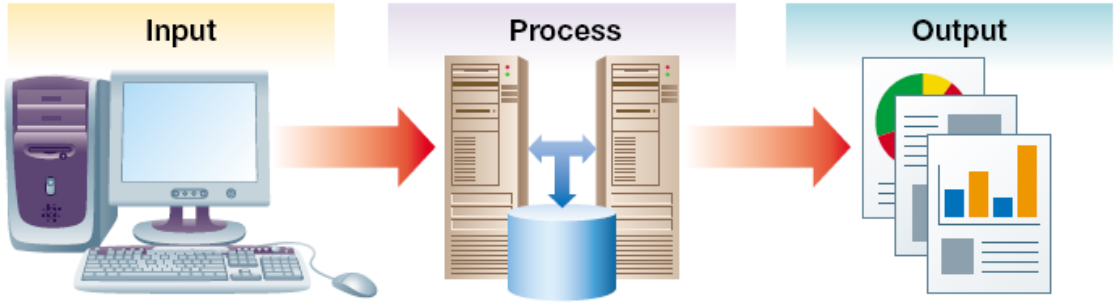
- A DBMS enables interactions with the database through activities such as:
  - Data Dictionary Management
  - Data Transformation and Presentation
  - Security Management
  - Backup and Recovery Management
  - Data Integrity Management
  - Database Access Languages and Application Programming Interfaces
  - Database Communication Interfaces
  - Transaction Management
  - Data entry, queries, etc.

DBMS – Data Entry



DBMS - Queries

- Query – A way to extract data from the database
  - Focuses on providing appropriate parameters to select the information required
- SQL (Structured Query Language)
  - A Language to select and extract data from a database
  - The industry standard language for relational databases
- QBE (Query by Example)
  - A technique that allows a user to design a query on a screen by dragging and placing the query field in the desired locations.



(Query Request) (Query Program) (Applicant Report)

The Database Management System – Query Language*

- SQL: Structured Query Language:
  - Popular language for making requests to a relational dbms

- **SELECT** LAST_NAME, FIRST_NAME, CITY
- **FROM** APPLICANT
- **WHERE** APPLICATION_DATE >= 'August 19, 2012'

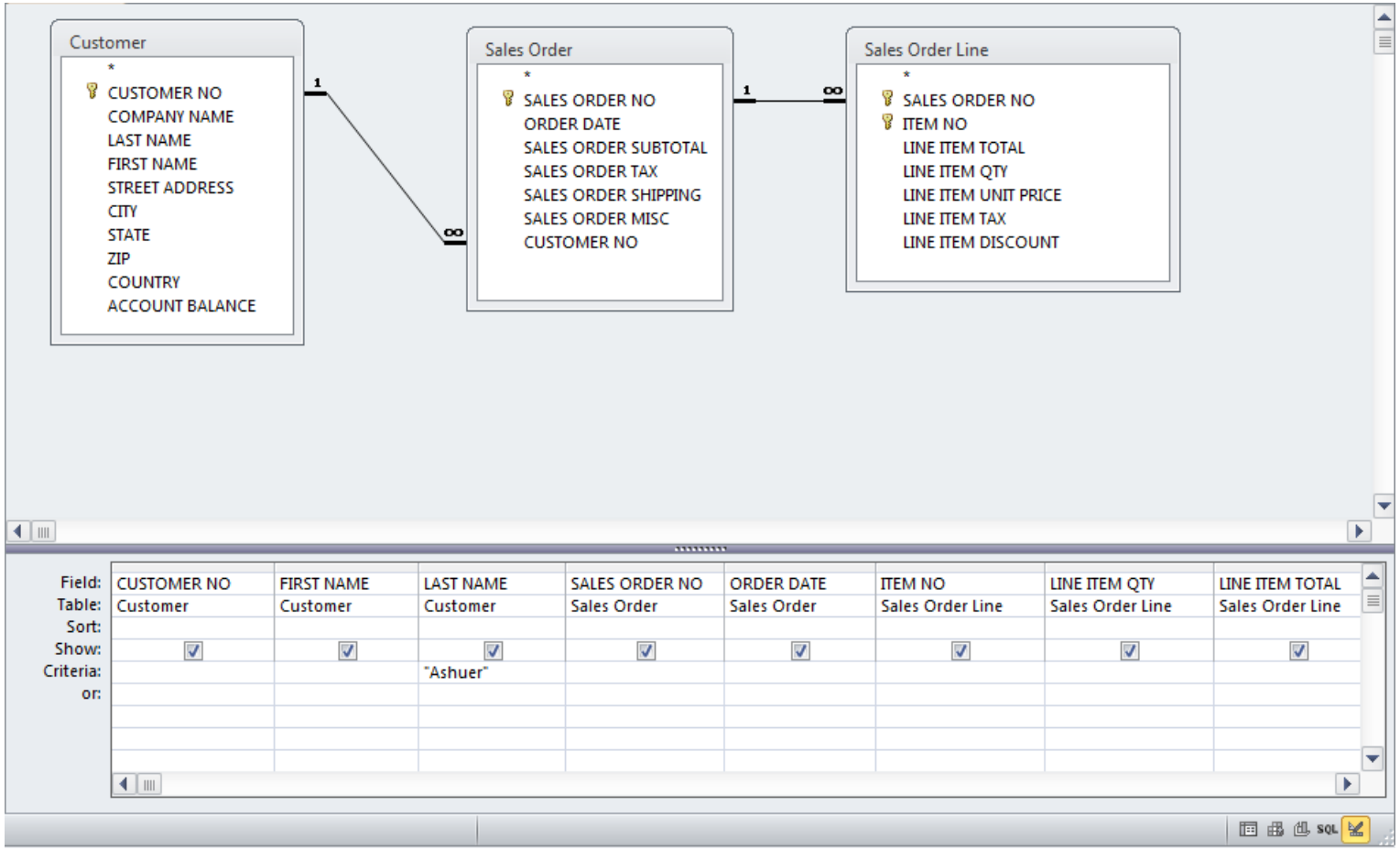
- **SELECT** LAST_NAME, FIRST_NAME, CITY SALARY
- **FROM** EMPLOYEE **WHERE** DEPARTMENT = '4530' AND SALARY > 25000

DBMS – Example of Query Results

WORKERS Query

wnum	wname	brith_date	rate	skill	certified
100	James Langdon	2/22/1953	$12.00	Painter	☑
200	Rekha Hindoch	1/8/1960	$30.00	Engineer	☑
246	Pierre Garceau	7/19/1947	$18.50	Electrician	☑
300	Mary Clutterha	12/30/1950	$12.50	Painter	☑
395	Donna Graham	9/27/1956	$21.00	Plumber	☑
452	Isabella Fong	5/12/1957	$15.00	Electrician	☐
453	Rosita Cordeirc	8/22/1955	$12.50	Painter	☐
565	Ernest Schneid	10/31/1957	$18.00	Electrician	☑
664	Maurice Favrea	5/17/1962	$12.00	Plumber	☐
743	Oleh Markiw	1/1/1959	$15.00	Carpenter	☑
754	Emil Juliano	9/9/1949	$15.00	Carpenter	☑
887	Salim Agarwal	11/20/1964	$17.00	Carpenter	☑
0			$0.00		☐

DBMS - Example of Query Results



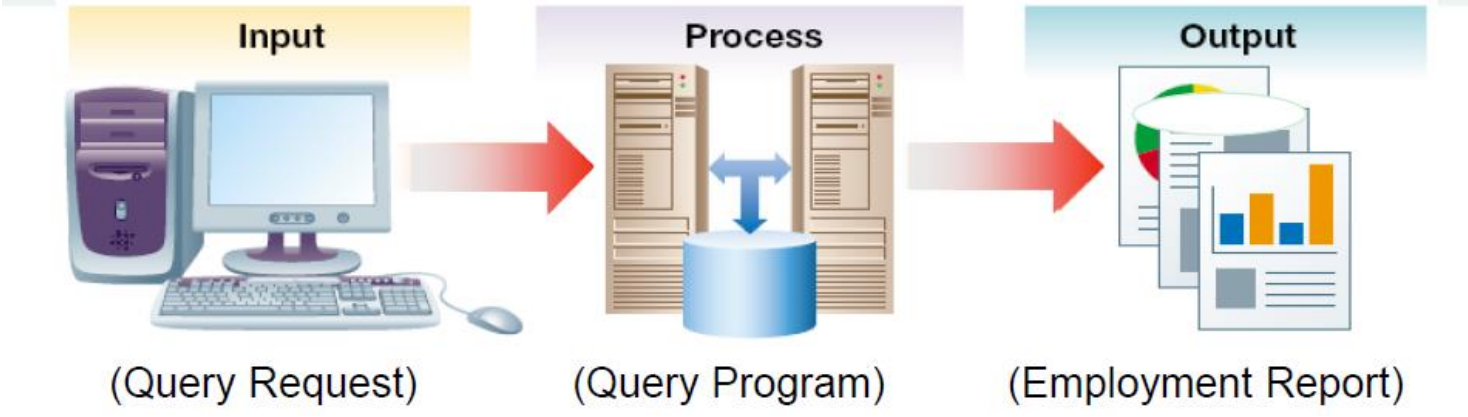
CUSTOMER NO	FIRST NAME	LAST NAME	SALES ORDER NO	ORDER DATE	ITEM NO	LINE ITEM QTY	LINE ITEM TOTAL
127127	Angela	Ashuer	5890	7/20/2010	9XB	1	$900.00
127127	Angela	Ashuer	5890	7/20/2010	ARQ1	10	$100.00
127127	Angela	Ashuer	5819	8/12/2010	WC2	1	$80.00

DBMS – Report

- Report – A database function that extracts and formats information from a database for printing and presentation

- Report Generator
  - A specialized program that uses SQL to retrieve and manipulate data (aggregate, transform, or group)
  - Report templates are designed using special report creation software (e.g. Crystal Reports, Jasper Reports, Cognos)

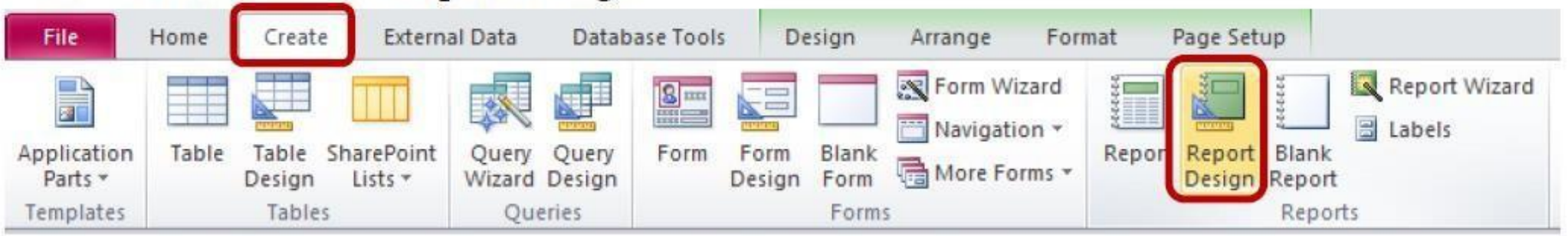
- Example – Report on applicants entered in the last 30 days
  - Report parameters are selected in the report request screen
  - The database program uses SQL to query and present the result



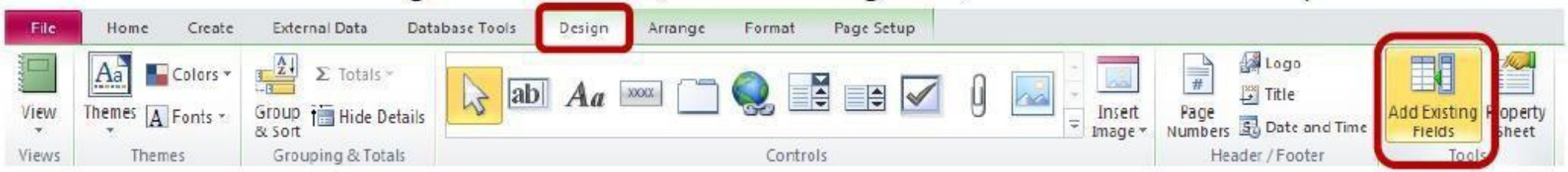
DBMS – Designing a Report

- To create reports, the developer must first select the table(s) or queries(s) upon which the report will be based.
- All fields will then become available to the developer to include on the report:
- *The developer formats and enters the report header, detail, and footer. The required fields are placed in the desired location on the report layout.*

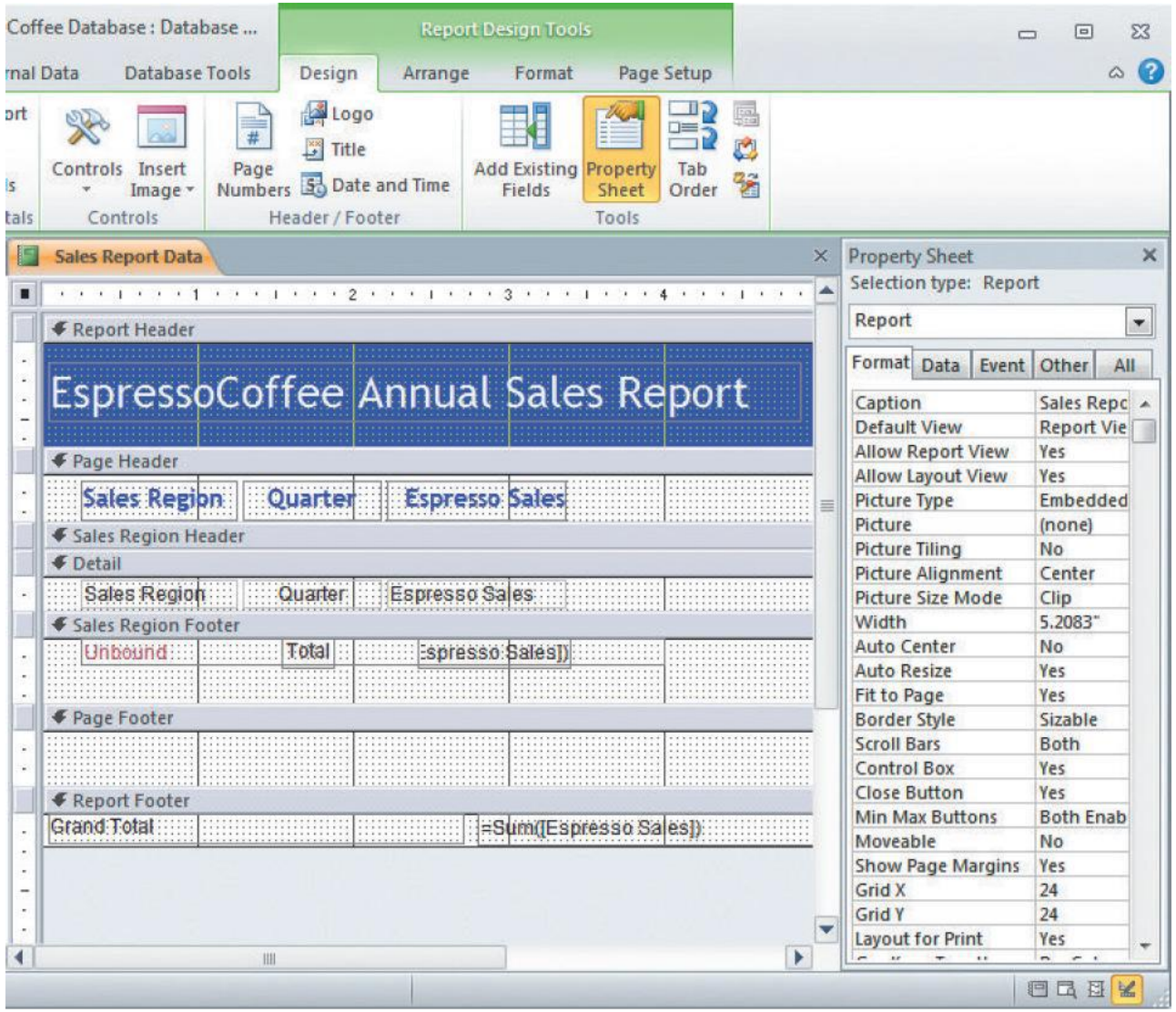
1. On the **Create** tab, click **Report Design**.



2. Click on the **Add Existing Fields** button, on the **Design** tab, to see a list of tables/fields.



DBMS – Designing a Structured Report



Database Reports

To build a database report, the developer uses a report design template as shown above. The developer formats and enters the report header, detail, and footer. The required fields are placed in the desired location on the report layout.

DBMS – Structured Report Results

WORKERS

WORKERS

wnum	wname	brith_date	rate	skill	certified
100	James Langdon	2/22/1953	$12.00	Painter	☑
200	Rekha Hindocha	1/8/1960	$30.00	Engineer	☑
246	Pierre Garceau	7/19/1947	$18.50	Electrician	☑
300	Mary Clutterham	12/30/1950	$12.50	Painter	☑
395	Donna Graham	9/27/1956	$21.00	Plumber	☑
452	Isabella Fong	5/12/1957	$15.00	Electrician	☐
453	Rosita Cordeiro	8/22/1955	$12.50	Painter	☐
565	Ernest Schneider	10/31/1957	$18.00	Electrician	☑
664	Maurice Favreau	5/17/1962	$12.00	Plumber	☐
743	Oleh Markiw	1/1/1959	$15.00	Carpenter	☑
754	Emil Juliano	9/9/1949	$15.00	Carpenter	☑
887	Salim Agarwal	11/20/1964	$17.00	Carpenter	☑

Tuesday, September 24, 2013 1

EspressoCoffee Annual Sales Report

Sales Region	Quarter	Espresso Sales
Austria	1	$610,911.00
	2	$901,574.00
	3	$465,460.00
	4	$671,190.00
	Total	$2,649,135.00
Canada	1	$635,144.00
	2	$777,186.00
	3	$338,432.00
	4	$226,018.00
	Total	$1,976,780.00
China	1	$61,241.00
	2	$643,284.00
	3	$834,940.00
	4	$497,871.00
	Total	$2,037,336.00
France	1	$969,279.00
	2	$61,797.00
	3	$353,502.00
	4	$779,811.00
	Total	$2,164,389.00

Implementing the Concepts

- After having reviewed all the basic database models in the last section, the following section will focus on a practical process to set up a database
- Databases refer to the way in which data is set up in the background (on disk) with the information system
- The purpose of this is that application programs are able to take such data and produce required results and reports

Steps to set up a Database

- First: Speak to people who will use the information system
  - go to every person that will need any kind of output from the computer when doing his / her job
  - Find out from them what data fields do they need for different reports that they will use
  - List those fields (data items), and write them all down
- The database is going to be set up as a relational database (most likely)
- Always think of the relationships between fields

Data Requirements Analysis

- Three different users tell you their information needs:
  - Person 1.
    - A report displaying Student number, Student Name, Address, and City for all students
    - A report showing Student Name, Student phone, Student Major, ordered by student name
  - Person 2
    - Student Name, Course name and grade that each student took last term
    - Major ,Student name and phone number for all students
  - Person 3
    - Course name, Course number, Student Name and Grade for each student
    - Course name, course number for each course offered by enrolment date over the past three years

Data Requirements Analysis

Example: School information System

- Three different users tell you their information needs:
  - Person 1.
  - A report displaying **Student number**, **Student Name**, **Address**, and **City** for all **students**
  - A report showing Student Name, **Student phone**, **Student Major**, ordered by Student Name
- Person 2
  - Student Name, **Course name** and **Grade** that each student took last term
  - Major ,Student name and phone number for all students
- Person 3
  - Course name, Course number, Student Name and Grade for each student
  - Course name, Course number for each course offered by **Enrolment Date** over the past three years

Data Requirements Analysis

Example: School information System

The director of XYZ school would like to assess the initial design of the system using the following student, course, and grade information:

- Student #: 1234; Student: A. Lam; Address: 491 Castle, Wpg; Phone: 204-258-6987; Major: Economics; Course: 1803, Computers, Grade: A; Course: 1901, Pascal, Grade: B; Course 2002, Finance, Grade A
- Student #: 9876; Student: R. Kelly; Address: 22 Miller Cres, Wpg; Phone: 204-258-7563; Major: Biology; Course: 2002, Finance, Grade: B; Course: 3421, Entomology, Grade: B; Course 4523, Zoology, Grade: C
- Student #: 4567; Student: J. Ng; Address: 399 High St.; Phone: 204-785-2145; Major: Business; Course: 1304, Accounting, Grade: C; Course: 2233, Marketing, Grade: A.

Example: School Database

- Data Items:
  - Student Number,
  - Student Name,
  - Student Street address,
  - Student City
  - Student Phone,
  - Student major area (only one),
  - for each course the student takes, the Course Number, Course Name, Enrolment Date, and Grade

Field Name	STNUMBER	STNAME	STREET	CITY	STPHONE	MAJOR	CNO	CNAME	ENROL	GRADE
Type	Text	Text	Text	Text	Number	Text	Text	Text	Numeric	Text
Length	4	25	25	20	10	15	4	20	8	2

Insert Data into your Table

- Insert all data records one by one:

ST Number	ST Name	Street	City	Major	Phone Number	Course Name	Course Number	Enrolment	Grade
1234	Lam A	491 Castle	Wpg	Economics	2042586967	Computers	1803		A

- Ensure all of the data attributes have been identified and populated in the table. If there are no data elements for a specific attribute, then enter data

ST Number	ST Name	Street	City	Major	Phone Number	Course Name	Course Number	Enrolment	Grade
1234	Lam A	491 Castle	Wpg	Economics	2042586967	Computers	1803	090912	A
9876	Kelly R	22 Miller Cr	Wpg	Biology	2042587563	Finance	2002	050112	B
4567	Ng J	399 High St	Wpg	Business	2047852145	Accounting	1304	050112	C

Storing Meaningful Information

- In order to store data on disk meaningfully, we notice we must set it up first, at least **in a file** so that we can get a variety of useful information out
- Storing just characters alone in a computer won't do; we need to group them
- Storing just fields alone in a computer won't do; we need to group them.
- Storing just one record alone in a computer won't do; we need a group of several records that follow the same layout and are somehow related (say students of the same class)

Inefficient Data Storage

STNumber	STName	Address	City	Major	StPhone	Cnumber	Cname	Enrolment	Grade
1234	Lam A	491 Castle	Wpg	Economics	2042586987	1803	Computers	090912	A
1234	Lam A	491 Castle	Wpg	Economics	2042586987	1901	Pascal	080911	B
1234	Lam A	491 Castle	Wpg	Economics	2042586987	2002	Finance	060111	A
9876	Kelly R	22 Miller Cr	Wpg	Biology	2042587563	2002	Finance	050112	B
9876	Kelly R	22 Miller Cr	Wpg	Biology	2042587563	3421	Entomology	080911	B
9876	Kelly R	22 Miller Cr	Wpg	Biology	2042587563	4523	Zoology	070910	C
4567	Ng J	399 High St	Wpg	Business	2047852145	1304	Accounting	050112	C
4567	Ng J	399 High St	Wpg	Business	2047852145	2233	Marketing	080911	A

Inefficient Data Storage

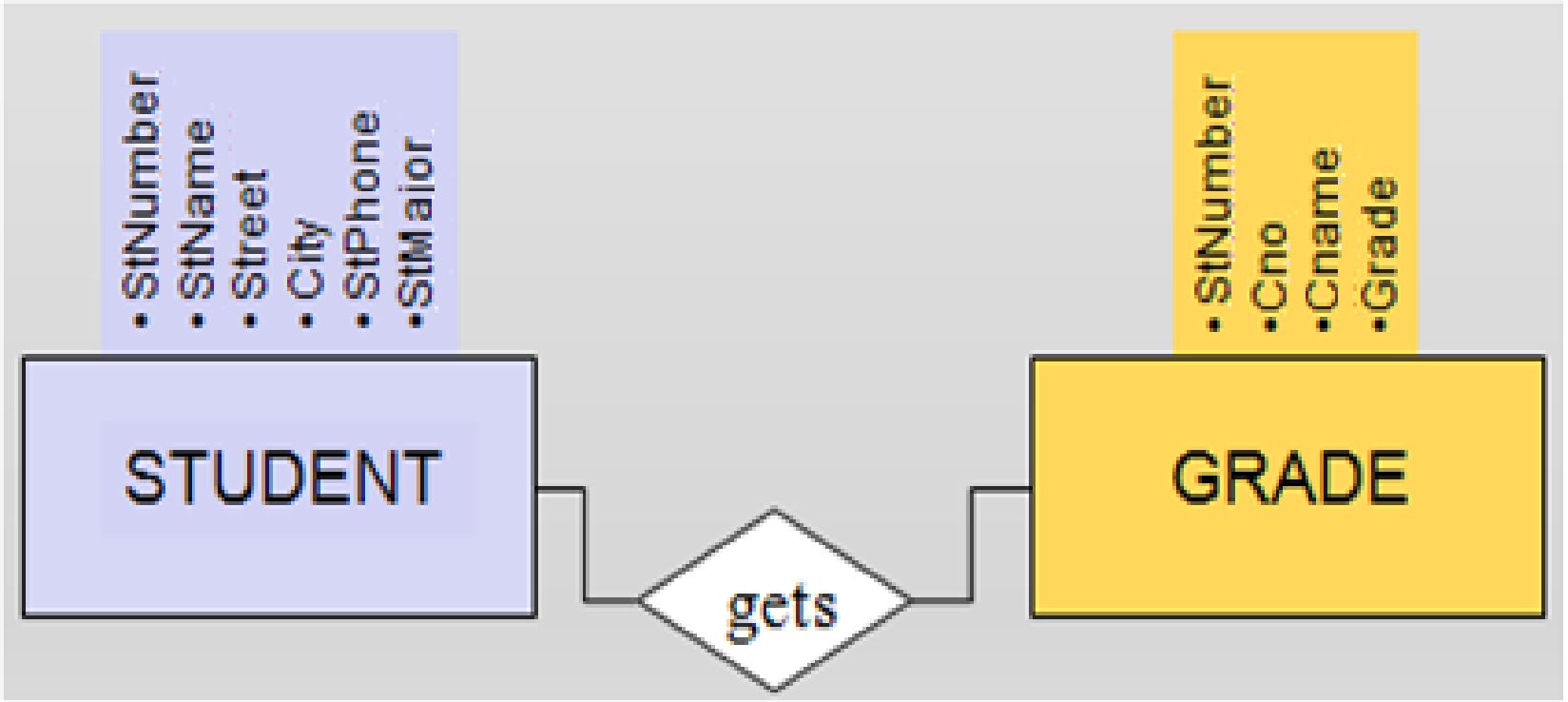
- Problems:
  - Data being Repeated
  - Inefficient Storage
    - Will take more space
    - Process for storing data will take longer
  - Inefficient Data retrieval
    - Data will take longer to be found
- Better to keep data in multiple tables (files)
  - Normalize the database!

School Database

- Determine the relationships when you have all the fields listed
- Group related fields into one table
  - Use logic, think about relationships
  - On this example, all fields with a one-to-one relationship go in one table, and all fields that have a one-to-many relationship in another table
- Determine what fields will be links between the tables
- Create you Data Model

Relationships

- When stored on Disk, *Student Number* is the common column linking data in the STUDENT table with data in the GRADES table
- An entity relationship (ER) diagram of this set up would be:



School Database

Student
StNumber
StName
Street
City
StPhone
StMajor

STNUMBER	STNAME	STREET	CITY	STPHONE	STMAJOR
1234	LAM A	491 Castle	Wpg	2042586987	Economics
9876	KELLY R	222 Miller Cr	Wpg	2042587563	Biology
4567	NG J	399 HIGH ST	Wpg	2047852145	Business

Grade
Student
Course Number
Course Name
Enrolment
Grade

STNUMBER	COURSENUM	COURSENAME	ENROLMENT	GRADE
1234	1803	Computers	090912	A
1234	1901	Pascal	080911	B
1234	2002	Finance	060111	A
4567	1304	Accounting	050112	C
4567	2233	Marketing	080911	A
9876	2002	Finance	050112	B
9876	3421	Entomology	080911	B
9876	4523	Zoology	070910	C

- This is a normalized database.

Keys

- Primary Key – Student Table
  - Uniquely identifies one record in the table
  - Student Number
- Foreign Key – Grade Table
  - Not unique in Grade Table but is unique in the related Student Table
  - Student Number
- Compound Primary Key – Grade Table
  - Consists of two or more simple keys
  - Student Number + Course Number + Enrolment
- Secondary Keys – Grade Table
  - Improves searches, may not be unique
  - Course Number, Enrolment

School Database

Student
StNumber
StName
Street
City
StPhone
StMajor

Grade
Student
Course Number
Course Name
Enrolment
Grade

Primary Key

STNUMBER	STNAME	STREET	CITY	STPHONE	STMAJOR
1234	LAM A	491 Castle	Wpg	2042586987	Economics
9876	KELLY R	222 Miller Cr	Wpg	2042587563	Biology
4567	NG J	399 HIGH ST	Wpg	2047852145	Business

Foreign Key

STNUMBER	COURSENUM	COURSENAME	ENROLMENT	GRADE
1234	1803	Computers	090912	A
1234	1901	Pascal	080911	B
1234	2002	Finance	060111	A
4567	1304	Accounting	050112	C
4567	2233	Marketing	080911	A
9876	2002	Finance	050112	B
9876	3421	Entomology	080911	B
9876	4523	Zoology	070910	C

- This is a normalized database.

School Database

Student
StNumber
StName
Street
City
StPhone
StMajor

STNUMBER	STNAME	STREET	CITY	STPHONE	STMAJOR
1234	LAM A	491 Castle	Wpg	2042586987	Economics
9876	KELLY R	222 Miller Cr	Wpg	2042587563	Biology
4567	NG J	399 HIGH ST	Wpg	2047852145	Business

Grade
Student
Course Number
Course Name
Enrolment
Grade

Compound Primary Key

STNUMBER	COURSENUM	COURSENAME	ENROLMENT	GRADE
1234	1803	Computers	090912	A
1234	1901	Pascal	080911	B
1234	2002	Finance	060111	A
4567	1304	Accounting	050112	C
4567	2233	Marketing	080911	A
9876	2002	Finance	050112	B
9876	3421	Entomology	080911	B
9876	4523	Zoology	070910	C

- This is a normalized database.

School Database

Student
StNumber
StName
Street
City
StPhone
StMajor

STNUMBER	STNAME	STREET	CITY	STPHONE	STMAJOR
1234	LAM A	491 Castle	Wpg	2042586987	Economics
9876	KELLY R	222 Miller Cr	Wpg	2042587563	Biology
4567	NG J	399 HIGH ST	Wpg	2047852145	Business

Grade
Student
Course Number
Course Name
Enrolment
Grade

Secondary Keys

STNUMBER	COURSENUM	COURSENAME	ENROLMENT	GRADE
1234	1803	Computers	090912	A
1234	1901	Pascal	080911	B
1234	2002	Finance	060111	A
4567	1304	Accounting	050112	C
4567	2233	Marketing	080911	A
9876	2002	Finance	050112	B
9876	3421	Entomology	080911	B
9876	4523	Zoology	070910	C

- This is a normalized database.

Data Reporting - Reports

- Example: School information System
- Three different users tell you their information needs:
  - Person 1.
    - A report displaying Student Number, Student Name, Address, and City for all students
    - A report showing Student Name, Student Phone, Student Major, ordered by Student Name
  - Person 2
    - Student Name, Course Name and Grade that each student took last term Student Major ,Student Name and Student Phone number for all students
  - Person 3
    - Course Name, Course number, Student Name and Grade for each student
    - Course Name, Course Number for each course offered by Enrolment Date over the past three years

Report – Courses by Enrolment

Courses by Enrolment Date for XYZ School

Enrol Date	Course No	Course Name
07-Sep-10	4523	Zoology
Total Number of Courses:		1
06-Jan-11	2002	Finance
Total Number of Courses:		1
08-Sep-11	1901	Pascal
	2233	Marketing
	3421	Entomology
Total Number of Courses:		3
05-Jan-12	1304	Accounting
Total Number of Courses:		1
09-Sep-12	1803	Computers
Total Number of Courses:		1

Data Reporting - Query

- Example: School information System
- Three different users tell you their information needs:
  - Person 1.
    - A report displaying Student number, Student Name, Address, and City for all students
    - A report showing Student Name, Student Phone, Student Major, ordered by Student Name
  - Person 2
    - Student Name, Course name and grade that each student took last term
    - Student Major, Student Name and Student Phone number for all students
  - Person 3
    - Course Name, Course Number, Student Name and Grade for each student
    - Course Name, Course Number for each course offered by Enrolment Date over the past three years

Query Reporting

- A report displaying Student number, Student Name, Address, and City for all students

St Number	St Name	Street	City
1234	Lam A	491 Castle	Wpg
9876	Kelly R	22 Miller Cr	Wpg
4567	Ng J	399 High St	Wpg

- Student Name, Course name and grade that each student took last term

STName	Cname	Enrolment	Grade
Lam A	Pascal	080911	B
Kelly R	Entomology	080911	B
Ng J	Marketing	080911	A

Big Data

- Extremely large and complex data collections
  - Traditional data management software, hardware, and analysis processes are incapable of dealing with them
- Three characteristics of big data
  - Volume
  - Velocity
  - Variety

Big Data Generators

Source	Magnitude of Data Generated
Large Hadron particle accelerator at CERN	40 terabytes of data per second
Commercial aircraft engines	More than 1 petabyte per day of sensor data
Cell phones	More than 5 billion people worldwide are making cell phone calls, exchanging text messages, and accessing Web sites
YouTube	48 hours of video uploaded per minute
Facebook	100 terabytes uploaded per day
Twitter	500 million tweets per day
RFID tags	1,000 times the volume of data generated by bar codes

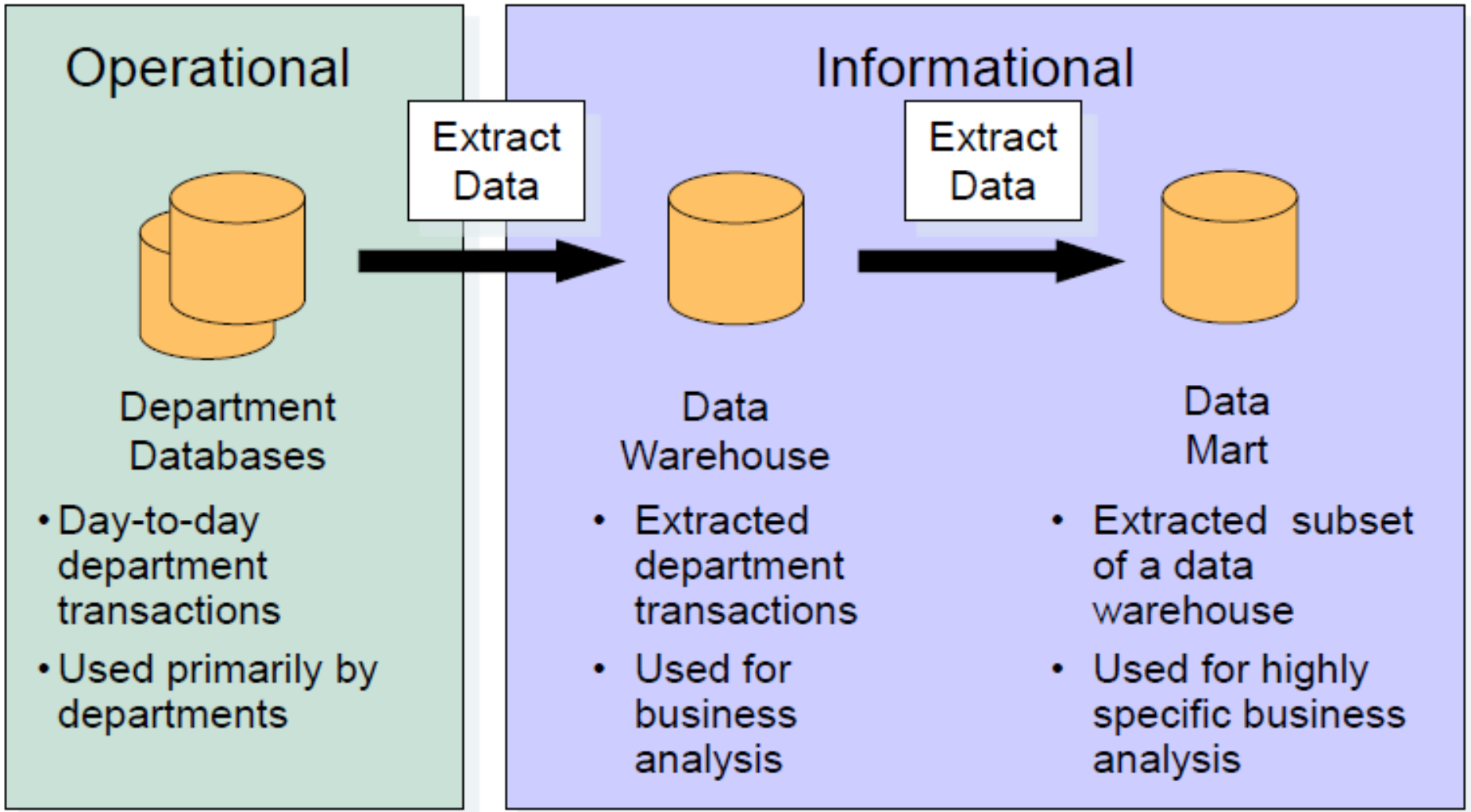
Challenges of Big Data

- How to choose what subset of the data to store
- Where and how to store the data
- How to find the nuggets of data that are relevant to the decision making at hand
- How to derive value from the relevant data

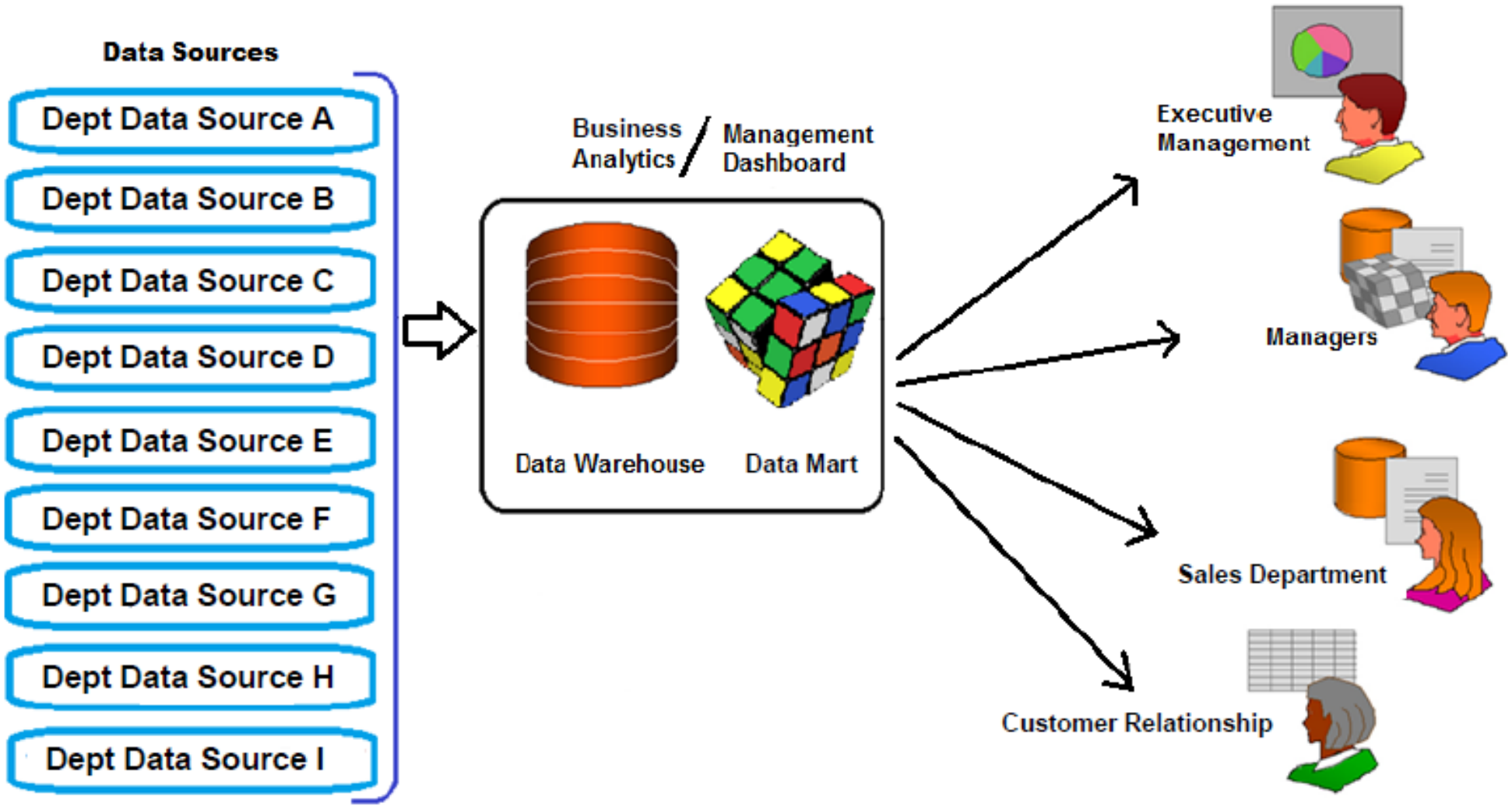
Data Warehouses, Data Marts, and Data Mining

- Data warehouse:
  - Database that holds business information from many sources in the enterprise
- Data mart:
  - Subset of a data warehouse that is used by small- and medium-sized businesses and departments within large companies to support decision making
  - A specific area in the data mart might contain greater detailed data than the data warehouse
- Data mining:
  - Information-analysis tool that involves the automated discovery of patterns and relationships in a data warehouse

Data Warehouses, Data Marts, and Data Mining



Data Warehouses, Data Marts, and Data Mining



Data Warehouses, Data Marts, and Data Mining

- Predictive analysis:
  - Form of data mining that combines historical data with assumptions about future conditions to predict outcomes of events
  - Used by retailers to upgrade occasional customers into frequent purchasers
  - Software can be used to analyze a company's customer list and a year's worth of sales data to find new market segments

Data Mining Applications

- Branding and positioning of products and services
  - Enable the strategist to visualize product behavior in different markets, while condensing the data in dimensions that are easily analyzed
- Customer Churn
  - Predict current customers who are likely to switch to a competitor
- Direct Marketing
  - Identify customer prospects most likely to respond to direct marketing practices
- Fraud detection
- Market Segmentation
- Trend analysis (sales, spending, promotions, etc.)

Data Management

End of Lecture 2-2