Introduction to Databases

PWSP CLASS
DATABASE MANAGEMENT

INSTRUCTOR: SANJAY GOEL
FEBRUARY 26-27, 2008
Databases

Learning Objectives

• Learn need for relational databases
• Create Entity-Relationship Diagrams
• Learn the process of database normalization
• Learn the constructs of SQL
• Design and Implement relational databases using Microsoft Access
**Databases**

**Data and Information**

- Data is an individual fact or multiple facts, or a value, or a set of values, but is not significant to a business in and of itself.
  - What does this mean: 7/13/2002

- Giving data context, or meaning, turns it into information.
  - Date of change of bankruptcy law 7/13/2002
Databases

What is a database

• A database is simply a bunch of information (data) stored on a computer.
  – This could be a list of all your clients, a list of the products you sell, the results of a chess tournament or everyone in your family tree.

• If you run a matrimonial agency you probably have a spreadsheet with all your clients’ names and addresses on your computer
  – The top of the spreadsheet are typically column headings: Name, Address, Telephone Number, Email Address, etc.
  – The spreadsheet might have customer ID numbers and other data
  – This is your client table in the database

• In your filing cabinet, you might have a separate folder for every project you’ve worked on. This folder contains the project number, the name (or ID number) of the client, a description of the project, the current status, the budget and a completion date.
  – This is your projects table in the database.
Databases

Need

• A database is required to keep track of things
  – The climatic patterns over the last thousand years
  – Data from celestial microscopes of the night sky
  – Number of kids born with a mole on their finger
  – Number of fishes which spawn in Alaska
  – Number of people who also buy crackers along when they buy milk.

• As we will see, unlike a list or spreadsheet, a database can store complex information more efficiently than a simple list
Databases

Role

• Database is required to:
  – Organize data.
  – Retrieve information.

• Remember
  – you store data in a database
  – you retrieve information from the database.
Databases

Properties

• Persistence
  – Data can be stored as long as required (i.e. magnetic disks rather than computer memory)
  – Retrieve information.

• Sharing
  – Can be used by multiple users simultaneously
  – Unless two people are trying to change the same data at the same time they should be able to operate independently

• Interrelated
  – Link information about different elements to provide a complete picture
DATABASE COMPONENTS
Database Components

Database System

• The four components of a database system are:
  – Users
  – Database Application
  – Database Management System (DBMS)
  – Database
Database Components

User

• A user of a database system will
  – Use a database application to track things
  – Use forms to enter, read, delete and query data
  – Produce reports
Database Components

Database

• A database is a self-describing collection of related records
  – The database itself contains the definition of its structure
  – Metadata is data describing the structure of the database data

• Tables within a relational database are related to each other
Database Components

**Database Management System (DBMS)**

- A database management system (DBMS) serves as an intermediary between database applications and the database.
- The DBMS manages and controls database activities.
- The DBMS creates, processes and administers the databases it controls.
- Database management system (DBMS) has two goals.
  - Add, delete and update data in the database.
  - Provide various ways to view data in a database.
Database Components

**DBMS Functions**

- Create databases
- Create tables
- Create supporting structures
- Read database data
- Modify database data (insert, update, delete)
- Maintain database structures
- Enforce rules
- Control concurrency
- Provide security
- Perform backup and recovery
Database Components

Applications

• A database application is a set of one or more computer programs that serves as an intermediary between the user and the DBMS

• Functions of Database Applications include
  – Create and process forms
  – Process user queries
  – Create and process reports
  – Execute application logic
  – Control database applications
Database Components

Desktop Database Systems

- Desktop database systems typically support single users
  - Have one application
  - Have only a few tables
  - Have a few records

Database Components:
- Database
- Application
- Database Management System (DBMS)
- e.g. MS Access
- Database
Database Components
Organizational Database Systems

- Typically multi-user systems
  - Include more than one application and several databases
  - Involve multiple computers
  - Are complex in design (large # of tables and records)
RELATIONAL DATABASE
## Databases

### Example (Collection of Books)

<table>
<thead>
<tr>
<th>ISBN</th>
<th>Title</th>
<th>AuID</th>
<th>AuName</th>
<th>AuPhone</th>
<th>PubID</th>
<th>PubName</th>
<th>PubPhone</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-555-55555-9</td>
<td>Macbeth</td>
<td>5</td>
<td>Shakespeare</td>
<td>555-555-5555</td>
<td>2</td>
<td>Alpha Press</td>
<td>999-999-9999</td>
<td>$12.00</td>
</tr>
<tr>
<td>0-91-335678-7</td>
<td>Faerie Queene</td>
<td>7</td>
<td>Spencer</td>
<td>777-777-7777</td>
<td>1</td>
<td>Big House</td>
<td>123-456-7890</td>
<td>$15.00</td>
</tr>
<tr>
<td>0-91-045678-5</td>
<td>Hamlet</td>
<td>5</td>
<td>Shakespeare</td>
<td>555-555-5555</td>
<td>2</td>
<td>Alpha Press</td>
<td>999-999-9999</td>
<td>$20.00</td>
</tr>
<tr>
<td>0-99-999999-9</td>
<td>Emma</td>
<td>1</td>
<td>Austen</td>
<td>111-111-1111</td>
<td>1</td>
<td>Big House</td>
<td>123-456-7890</td>
<td>$20.00</td>
</tr>
<tr>
<td>0-55-123456-9</td>
<td>Main Street</td>
<td>9</td>
<td>Smith</td>
<td>123-222-2222</td>
<td>3</td>
<td>Small House</td>
<td>714-000-0000</td>
<td>$22.95</td>
</tr>
<tr>
<td>0-55-123456-9</td>
<td>Main Street</td>
<td>10</td>
<td>Jones</td>
<td>123-333-3333</td>
<td>3</td>
<td>Small House</td>
<td>714-000-0000</td>
<td>$22.95</td>
</tr>
<tr>
<td>0-103-45678-9</td>
<td>Iliad</td>
<td>3</td>
<td>Homer</td>
<td>333-333-3333</td>
<td>1</td>
<td>Big House</td>
<td>123-456-7890</td>
<td>$25.00</td>
</tr>
<tr>
<td>0-12-333433-3</td>
<td>On Liberty</td>
<td>8</td>
<td>Mill</td>
<td>888-888-8888</td>
<td>1</td>
<td>Big House</td>
<td>123-456-7890</td>
<td>$25.00</td>
</tr>
<tr>
<td>1-22-233700-0</td>
<td>Visual Basic</td>
<td>4</td>
<td>Roman</td>
<td>444-444-4444</td>
<td>1</td>
<td>Big House</td>
<td>123-456-7890</td>
<td>$25.00</td>
</tr>
<tr>
<td>1-1111-11111-1</td>
<td>C++</td>
<td>4</td>
<td>Roman</td>
<td>444-444-4444</td>
<td>1</td>
<td>Big House</td>
<td>123-456-7890</td>
<td>$29.95</td>
</tr>
<tr>
<td>0-123-45678-0</td>
<td>Ulysses</td>
<td>6</td>
<td>Joyce</td>
<td>666-666-6666</td>
<td>2</td>
<td>Alpha Press</td>
<td>999-999-9999</td>
<td>$34.00</td>
</tr>
<tr>
<td>0-321-32132-1</td>
<td>Balloon</td>
<td>12</td>
<td>Grumpy</td>
<td>321-321-0000</td>
<td>3</td>
<td>Small House</td>
<td>714-000-0000</td>
<td>$34.00</td>
</tr>
<tr>
<td>0-321-32132-1</td>
<td>Balloon</td>
<td>13</td>
<td>Sleepy</td>
<td>321-321-1111</td>
<td>3</td>
<td>Small House</td>
<td>714-000-0000</td>
<td>$34.00</td>
</tr>
<tr>
<td>0-321-32132-1</td>
<td>Balloon</td>
<td>11</td>
<td>Snoopy</td>
<td>321-321-2222</td>
<td>3</td>
<td>Small House</td>
<td>714-000-0000</td>
<td>$34.00</td>
</tr>
<tr>
<td>0-11-345678-9</td>
<td>Moby Dick</td>
<td>2</td>
<td>Melville</td>
<td>222-222-2222</td>
<td>3</td>
<td>Small House</td>
<td>714-000-0000</td>
<td>$49.00</td>
</tr>
<tr>
<td>0-12-345678-6</td>
<td>Jane Eyre</td>
<td>1</td>
<td>Austen</td>
<td>111-111-1111</td>
<td>3</td>
<td>Small House</td>
<td>714-000-0000</td>
<td>$49.00</td>
</tr>
<tr>
<td>0-99-777777-7</td>
<td>King Lear</td>
<td>5</td>
<td>Shakespeare</td>
<td>555-555-5555</td>
<td>2</td>
<td>Alpha Press</td>
<td>999-999-9999</td>
<td>$49.00</td>
</tr>
</tbody>
</table>
Databases

Why Use a Database?

• Most databases worth maintaining are quite complex.
  – Library of congress contains 16 million records
  – Social Security Database
  – Department of Motor Vehicles Database

• Why can’t we use a flat file like we had in word?
  – Redundancy

• Redundancy is unnecessary repetition of data
  – Wasted Storage
  – Database Anomalies
Databases

Redundancy – Book Database

- Multiple values in the column of a database
  - e.g. some books are authored by multiple authors.

- There are three choices
  - Accommodate multiple authors in multiple rows (one for each author)
    - Complete information about a book is repeated as many times as there are authors. (causing large redundancy)
  - Have multiple columns for the authors in each row.
    - You have to determine the max number of authors a priori
    - A lot of the fields will go unused
  - Add all the author names in one column
    - Searching and sorting become very hard.
Databases

Redundancy – Library of Congress

- Library of Congress Example
  - 10,000 publishers
  - 16 million records
  - Each address on average 50 characters long
  - Assuming each character takes 2 bytes, the difference in storage is:
    \[(16,000,000 - 1000) \times 50 \times 2 \text{ bytes}\]
    \[\sim 1.6 \text{ gbytes}\]
  - Duplication of address alone requires 1.6 gigabytes of storage
Databases

Redundancy – Project Database

• In a list, each row is intended to stand on its own. So, the same information may be entered several times
  – E.g. Consider a list of Projects that include the Project Manager’s Name, ID, and Phone. If the same person is managing 5 projects his information would be repeated 10 times

• In a list, each row may contain information on more than one theme. As a result, needed information may appear in the lists only if information on other themes is also present
  – E.g. A list of Projects may include Project Manager information (Name, ID, and Phone Extension) and Project information (Name, ID, StartDate, Budget) in the same row.
Databases

Anomalies

• A table anomaly is a structure for which a normal database operation cannot be executed without information loss or full search of the data table

• Three types of anomalies
  – Insertion
  – Deletion
  – Update
Databases

Insertion Anomalies

• Insertion anomaly occurs when extra data beyond the desired data must be added to the database

• If we need to add a new publisher to the database, but we do not have any book by that publisher.
  – we will need to add a new line and put NULL values in all but publisher related columns.
  – Also ISBN column which is supposed to be unique will have numerous null values.
Databases

Update Anomalies

• An update anomaly occurs when it is necessary to update multiple rows to modify a single fact.
  – If the phone number of a publisher changes then all the entries of the publisher need to be changed.
  – For instance changing the phone number of Big House publisher requires changing the phone number 6 times.
  – Why is this bad?
Databases

Deletion Anomalies

• A deletion anomaly occurs whenever deleting a row inadvertently causes other data to be deleted.
  – If we lose a book and delete a row containing the book, we lose the information of the publisher if it is the only book by that publisher.
  – If we remove books Macbeth, Hamlet, Ulysses, and King Lear we lose all information about Alpha Press and Shakespeare.
Databases
Relational Design

• To create a relational database
  – Break table into a collection of smaller tables.
  – Define relationships among the table

• Each smaller table has
  – a heading which contains the table definition
  – a body which contains the content

• The relationships are created by having common columns among tables
  – Matching values in the rows demonstrate relationships
  – These relationships are used to join tables while designing queries
Databases
Relational Design
Databases

Relational Design

• Increased complexity
  – Instead of simply sorting on columns in a table we need to gather information from multiple tables.

• Relational Integrity
  – Relational integrity should be maintained while changing data.
  – For instance if we delete publishers we can not let the books by that publisher reside in the books database as dangling references.

• Inadvertent data loss
  – During the design care must be taken to not lose any data
  – For instance without the books author table we will not know how to relate the books and authors.
Databases

Entities

• A database stores information about things encountered in real life i.e. person, places, things, or events.
• An entity is something of importance to a user that needs to be represented in a database
  – An entity represents one theme or topic
  – In an entity-relationship model thus entities are restricted to things that can be represented by a single table
• In the book database example Books, Authors & Publishers are the entities?
• All possible entities for a given entity type constitute the entity class.
• The subset of the entities from the entity class contained in a database is called an entity set
Databases

Entities Cont’d

• An entity class stays constant whereas an entity set can change.

• For the given book database
  – Book is an entity
  – Set of all possible books in the world is the entity class
  – The subset of the 14 books in the book table is the entity set
Databases

Relation

• A *relation* is a two-dimensional table that has specific characteristics
• The table dimensions, like a matrix, consist of rows and columns
• Characteristics of a Relation
  – Rows contain data about an entity
  – Columns contain data about attributes of the entity
  – Cells of the table hold a single value
  – All entries in a column are of the same kind
  – Each column has a unique name
  – The order of the columns is unimportant
  – The order of the rows is unimportant
  – No two rows may be identical
Properties of the entities that describe their behavior are called the attributes.

- Attribute values are the actual entries in each cell of a database table.

The attributes have three main purposes

- Represent the real data in the database.
  e.g. in the book table the title and price are the describing attributes

- Uniquely identify entities within an entity class.
  e.g. ISBN in the book table, PubID in the publisher table, and AuID in the author table provide unique identity to a book, publisher or an author.

- Define relationship of one entity with another entity
Databases

Attributes

• A set of multiple attributes can describe an entity uniquely so it is not absolutely essential to have an identifying attribute however it is useful to have these for the purpose of efficiency.
  – e.g. for all the U.S. residents Name, race, color, height are attributes that describe the data while Social Security Number is the identifying entity.

• What are the attributes that we need for our three entities?
  – Books: Title, Price, ISBN
  – Authors: AuID, AuPhone, AuName
  – Publishers: PubName, PubPhone, PubID
Databases

Attributes (Observations)

• From the books attributes there is no way to identify the publisher and author.
  – We need to add more attributes to describe the relationships.

• We need to distinguish between the unique attribute for an entity set vs. unique attribute for an entity class.
  – e.g. Book database: Title is a unique attribute for the current set of books, however, there are many books in the world with the same titles. ISBN on the other hand is a unique identifier for the book.
  – e.g. Adult males living in the U.S.: A lot of them have the same name, (probably the same age), however, SSN is a unique identifier.

• Even though Publisher can probably be uniquely identified by the Publisher Name & Phone Number, we have added PubID to make identification more efficient.

• The attributes and the unique identifier selection is context dependent and is the job of the database designer.
Databases

Keys & Superkeys

• A *key* is one (or more) columns of a relation that is (are) used to identify a row

• A set of attributes from the set of all the attributes for a given entity is called the super key for the entity class.
  – {ISBN} is the superkey for the Book entity
  – {PubID} or {PubName, PubPhone} are the superkeys for the Publishers entity class.
  – There can be multiple superkeys for a given entity.

• Superkeys should be evaluated on the basis of entity class not the entity set
  – Unique identifier for books in a database of 14 books may not work when additional books are added to the table
Databases

Candidate Keys and Primary Key

• A superkey is called a candidate key if no proper subset of the superkey is also a key.
  – i.e. a candidate key is a minimal superkey.
  – e.g. Both \{ISBN\} and \{ISBN, Title\} are superkeys for the books table since they both uniquely identify the Book.
  – However it is not necessary to include the Title in the superkey.

• A table may have multiple candidate keys

• A primary key is a candidate key chosen to be the main key for the relation

• If you know the value of the primary key, you will be able to uniquely identify a single row
Databases

Composite Keys

• A composite key is a key that contains two or more attributes

• For a key to be unique, often it must become a composite key

• To identify a family member, you need to know a FamilyID, a FirstName, and a Suffix (e.g., Jr.)

• The composite key is:

  (FamilyID, FirstName, Suffix)

• One needs to know the value of all three columns to uniquely identify an individual
Databases

Relationships

• A table may be related to other tables
• For example
  – An Employee works in a Department
  – A Manager controls a Project
• Relationships form associations between multiple entities
  – e.g. Book is written by an author
• Number of entities in a relationship is called the degree of a relationship
  – Binary relationship involves two entities
  – Ternary relationship involves three entities
Databases

Foreign Keys

• To preserve relationships, foreign keys are created
• A foreign key is a primary key from one table placed into another table
• The key is called a foreign key in the table that received the key
Databases

Binary Relationships

• There are three kinds of binary relationships
  – One-to-one (1:1): A single entity instance of one type is related to a single entity instance of another type
  – One-to-Many (1:N): A single entity instance of one type relates to many entity instances of another type
  – Many-to-Many (N:M): A single entity instance of one type relates to many entity instances of another type & vice versa
Databases

Binary Relationships

• One-to-one relationships are rare since they can be substituted by adding one or more extra attributes in one of the tables to model the attributes of the other. A strong justification is required for having such a relationship
  – e.g. Passwords are kept in a separate table for reasons of security.
  – e.g. If one of the fields contains a large data set it is maintained in a separate table for efficiency
Databases

Binary Relationships

• Cardinality specifies (maximum) number of instances of an entity that relate to one instance of another entity
  – e.g. Basketball team and starting players have cardinality of 5

• Ordinality describes the minimum number of instances of an entity that relate to one instance of another entity
  – if the minimum number is zero the relationship is optional
  – if the minimum number is greater than zero the relationship is mandatory
Databases

Weak Entities

• Weak entities are those that can not exist unless another entity also exists in the database
  – Entity that is not weak is a strong entity
    
    ![Employee](1:N) Dependent

  – The employee can exist without a dependent but not vice versa
    
    ![Building](1:N) Apartment

  – In this case the apartment address is a composite of building number and apartment number, so apartment can’t exist without building. (Such entities are also called id-dependent entities)

• The entity should not only depend physically but also logically to avoid ambiguities

  ![Advisor](1:N) Student

  – Even though a business rule says that each student should have an advisor student is still a strong entity
Databases

Weak Entities

- By business rule order would have a sales person associated with it but this is not a logical necessity this order is not a weak entity

- Prescription cannot logically exist without a patient thus it is a weak entity

- Thus a weak entity is the one with an ordinality (minimum cardinality) of 1 and a logical dependence on another entity
Databases

Examples

• University Database
  – Entities: Students, faculty, courses, offerings, enrollments
  – Relationships: faculty teach offerings, students enroll in offerings, offerings made of courses

• Water Utility Database
  – Entities: Customers, meters, bills, payments, meter readings
  – Relationships: bills sent to customers, customers make payments, customers read meters

• Hospital Database
  – Entities: Patients, providers, treatments, diagnoses, symptoms
  – Relationships: patients have symptoms, providers prescribe treatments, providers make diagnoses
E-R DIAGRAMS
**ER-Diagrams**

**Definition**

- They provide a way to pictorially depict the entities, attributes and relationships.
  - These are also called semantic networks.
- There are three elements of the ER-Diagram
  - Entities are represented by labeled rectangles. The label is the name of the entity.
  - Attributes are represented by oval boxes and contain the name of the entity.
  - Relationships are represented by a diamond connected to the two entities using solid lines
    (cardinality of many is represented by an infinity sign, cardinality of 1 is represented by a 1)
  - Weak entities are represented by a rectangle curved at the corners and the relationship triangle curved at the corners
ER-Diagrams

Book Database

- Among book authors there are people who are not primary authors but are contributors.
  - e.g. illustrators, indexers etc.
  - Each has a different level based on the contribution
- A separate entity can be used to represent contributors
  - Attributes: Level and Type.
- Let us now define the relationships.
  - A Book is written by authors
  - A Book is published by a publisher
  - A Contributor is an author
- Once this semantic model is created we need to create a relational database with this semantic model.
E-R Diagrams
Book Database

Books
- ISBN
- Title
- Price
- Published By
  - PubID
  - PubName
  - PubPhone

Authors
- AUID
- AuName
- AuPhone
- Written By
  - Books

Publishers
- PubID
- PubName
- PubPhone

Contributor
- ConID
- ConName
- ConLevel
- ConType

Is A
- Contributors

1
- Books

M
- Written By

N
- Authors

1
- Published By

1
- Contributors
E-R Diagrams

Interior Designer

An interior designer who specializes in home kitchen designs offers a variety of seminars at home shows, kitchen and appliance stores, and other public locations. The seminars are free; she offers them as a way of building her customer base. She earns revenue by selling books and videos and instructs people on kitchen design. She also offers custom-design consulting services. Her business is in selling products to the attendees at her seminars. She would like to develop a database to keep track of customers, the seminars that they have attended, and the purchases that they have made.

Please determine the entities, attributes and relationships that should exist in the database and draw an E-R diagram.

(Source: Database Concepts by Kroenke)
E-R Diagrams

Supplier

• An organization purchases items from a number of suppliers. It keeps track of the items purchased from each supplier, and it also keeps a record of suppliers' addresses. Items are identified by ITEM-TYPE and have a DESCRIPTION. There may be more than one such address for each supplier, and the price charged by each supplier for each item is stored. Suppliers are identified by SUPPLIER-ID.
E-R Diagrams

Supplier (Base)

Supplier
- SupID
- SupName
- SupPhone

Address
- SupCity
- SupStreet
- SupZip

Items
- ProdID
- ProdName
- ProdDesc
- ProdType

Supplies

Has

1

N

M
E-R Diagrams
Supplier (Expanded)
E-R Diagrams

Hospital

- A hospital stores data about patients, their admission and discharge from departments and their treatments. For each patient, we know the name, address, sex, social security number, and insurance code (if existing). For each department, we know the department's name, its location, the name of the doctor who heads it, the number of beds available, and the number of beds occupied. Each patient gets admitted at a given date and discharged at a given date. Each patient goes through multiple treatments during hospitalization; for each treatment, we store its name, duration, and the possible reactions to it that the patient may have.
E-R Diagrams

Hospital

- Patients
  - id
  - ssn
  - fname
  - lname
  - sex
  - Insurance
  - name
  - duration
  - Description

- Admitted To
  - PatientSSN
  - DepartmentId
  - id
  - location
  - name
  - head
  - has A

- Treatments
  - TreatmentID
  - PatientSSN
  - id

- Departments
  - id

- Beds
  - id
  - type

- Occupancy
  - Date
  - Time

- Get
  - Admission

- Occupy
  - Occupancy

- Admit To
  - M

- Has A
  - N

- Pat/Treatment
  - N

- PatientSSN
  - N

- TreatmentID
  - N
E-R Diagrams

Building

Each building in an organization has a different BUILDING-NAME and a BUILDING-ADDRESS. The meeting rooms in each building have their own ROOM-NO in the building, and each room has a specified SEATING-CAPACITY. Rooms are available for hire for meetings, and each hire period must start on the hour. The hour and LENGTH-OF-USE are recorded. Each hire is made by a group in the organization, and groups are identified by a GROUP-NO and have a CONTACT-PHONE. The equipment required for each hire period also are recorded. Each facility has an EQUIP-NO and a DESCRIPTION.
E-R Diagrams

Building
E-R Diagrams

Projects

A company has a number of employees. The attributes of EMPLOYEE include Employee_ID (identifier), Name, Address, and Birthdate. The company also has several projects. Attributes of PROJECT include Project_ID (identifier), Project_Name, and Start-date. An employee’s billing rate may vary by project and the company wishes to record the applicable billing rate (Billing_Rate) for each employee when assigned to a particular project.
E-R Diagrams

Projects

Employee

Project

Rents

Assignment

id

name

Address

DateOfBirth

ProjectId

EmployeeID

BillingRate

hours

StartDate

EndDate

id

Name

StartDate
E-R Diagrams

Chemists

A laboratory has several chemists who work on one or more projects. Chemists also may use certain kinds of equipment on each project. Attributes of CHEMIST include Employee_ID (identifier), Name, and Phone_No. Attributes of PROJECT include Project_ID (identifier) and Start_Date. Attributes of EQUIPMENT include Serial_No and Cost. The organization wishes to record Assign_Date, that is, the date when a given equipment item was assigned to a particular chemist working on a specified project. A chemist must be assigned to at least one project and one equipment item. A given equipment item need not be assigned, and a given project need not be assigned either a chemist or an equipment item.
E-R Diagrams

Video Rental Store

• You were hired by Blockywood Video to design a database for their video rental system. Luckily, you have had some experience renting videos (every Saturday night!) and you know that the video store needs to be able to information about customers, movies, and distributors.

• A movie title (e.g. Lord Voldemort of the Rings) has an ID number generated by the store. In addition, there is a title, release date, rating (G, PG-13, etc.), and price category (a newly released movie tends to be more expensive). Other movie related information saved is the director, producer, principal actor, and principal actress.

• Jane Doe is a customer and to rent movies, she is the owner of a Blockywood Video card which is scanned every time she makes a purchase. To obtain a card the owner needs to be above 18 (checked during application). The account associated with the card has a specific ID number (generated by the store), as well as the owner name (Jane Doe), her address, phone number, credit card information (optional), and amount credited/debited to the account. Additionally, she has both her husband (John Doe) and her son Jimmy Doe listed as other authorized users on the account and a rating for Jimmy of PG-13, which means that the video rental store is not allowed to let him have access to a higher rating video. Her husband also owns a Blockywood Video card in which Jane and Jimmy are listed as authorized users.

• Distributors have organizational names, an address, representative, and phone number. Videos are purchased from distributors on a specific date for a specific price. There can be multiple videos with the same movie title. Videos are also given a unique id number.

• The video rental system allows for a video rental history, which maintains the video rented, rental date, card that it was rented with, return date, and amount paid. This information helps the recommendation of new videos and to maintain accounts. The video rental system should also be able to specify the maximum number of videos a customer can rent and the amount of time a movie can be late until the account is frozen.
Assumptions:
1. Multiple distributors can sell same movie
2. Each member can be many accounts and the same account can have many members
3. Video rental store does not track the origin of a specific video from the distributor
4. Media type was added as additional attribute to the video (VHS or DVD)
E-R Diagrams

Real Estate Firm

• Prepare an E-R diagram for a real estate firm that lists property for sale. Please indicate primary key and foreign keys as well as relationships. The following describes this organization:
• The firm has a number of sales offices in several states. Attributes of sales office include Office_Number (identifier) and Location.
• Each sales office is assigned one or more employees. Attributes of employee include Employee_ID (identifier) and Employee_Name. An employee must be assigned to only one sales office.
• For each sales office, there is always one employee assigned to manage that office. An employee may manage only the sales office to which he or she is assigned.
• The firm lists property for sale. Attributes of property include Property_ID (identifier) and Location.
• Each unit of property must be listed with one (and only one) of the sales offices. A sales office may have any number of properties listed, or may have no properties listed.
• Each unit of property has one or more owners. Attributes of owners are Owner_ID (identifier) and Owner_Name. An owner may own one or more units of property. An attribute of the relationship between property and owner is Percent_Owned.
• **Note:** Components of Location include Address, City, State, and Zip_Code & Components of the name include f_name and l_name
E-R Diagrams

Concerts

• After completing a course in database management, you have been asked to develop an E-R diagram for a symphony orchestra. The symphony orchestra has the following characteristics

• A concert season schedules one or more concerts. A particular concert is scheduled for only one concert season.

• A concert includes the performance of one or more compositions. A composition may be performed at one or more concerts, or may not be performed.

• For each concert there is one conductor. A conductor may conduct any number of concerts, or may not conduct any concerts.

• Each composition may require one or more soloists, or may not require a soloist. A soloist may perform one or more compositions at a given concert, or may not perform any composition. The symphony orchestra wishes to record the date when a soloist last performed a given composition.

• Some other attributes of interest are the names and contact information of the soloist, composer and the location of the concert. The following entities are evident from the above discussion, i.e. ConcertSeason, Concert, Composition, Conductor, Soloist. Based on the above observations define what the appropriate relationships and attributes in the schema and construct an E-R diagram.
E-R Diagrams

Quick Oil

• Design the database and show an E-R Diagram for the case below. For many-to-many entities, show the additional tables that you would create.

• Quick Oil Company is in the business of changing engine oil for its customers. In addition to changing oil, the company also performs other services such as changing wipers, light bulbs etc. There are several technicians assigned to a car as it comes in for an oil change. They maintain the name, address, social security number, phone, and email address of the customer. In addition, the services that the customers purchase at different times are tracked. For each service, they store the name of the service, the cost of the service, description of the service, and number of hours of labor for the service. They maintain employee information including name, phone, social security number, address, date-of-birth, date-of-hire, and their hourly salary in the database. They also sell parts and the name, manufacturer, cost, weight and number of items of each part is stored. A list of parts is maintained in the database and customers that bought each part are tracked.
E-R Diagrams

Student Interviews

The School of Business would like to help its MBA students with placement in various companies and firms. To facilitate this, they need to design a database with the primary purpose of scheduling interviews and facilitating searches by students and companies that are looking for candidates. The database should have information about students, job openings, companies, interviewers, interviews, and conference rooms for interviews. Students have a lot of information that needs to be processed. This includes a first name, last name, student id number, social security number, concentration, and expected graduation date. Companies post job openings that students can apply for. The job posting date is recorded as well as the deadline of the posting. Also included is the title of the position, base salary, minimum requirements, and description. Students apply for open positions in companies and the date and time of each application is recorded. Students should also be able to see the status of their application (whether there will be a follow-up interview and whether they were turned down for the position). Interviews can be facilitated by the University or because of job openings and need to have a date, time, and conference room location. Multiple students can be interviewed at the same time and multiple interviewers can participate in a student interview. Companies send employees who interview students. An employee has an employee id, room #, telephone #, fax #, and position title. The company keeps track of which employees interviewed which students as well as the result of the interview (if it resulted in a job offer and/or hiring). A company has an identifier, name, main contact #, and main fax number. A company from the Business School reserves conference rooms and the company needs to specify the resources needed (e.g. computer, projector, etc.). There are multiple conference rooms and multiple instances of the same resource (e.g. there may be three projectors). The School should be able to track which company hired which students as well as how many students came from which concentration. In addition, such statistics as the average salary for the jobs per year should be available. Students should also be able to track their job offers.
E-R Diagrams

Handy Man

With a busy work life, home owners have difficulty in finding help for chores around the house and garden. Different jobs require different skills, such as, carpentry, plumbing, electrical wiring, etc. Several people have these skills and spare time to work on projects but are unable to find places for part time employment. This business entails connecting people who have jobs with workers willing to work on those jobs in their spare time. The workers should be able to list their contact information and skills. The employers should be able to list the jobs and the skills required for each job. The business model of the company is to get an enrollment fee for joining the network for both the home owners and the workers. Please design a database and create an E-R Diagram for the database.
E-R Diagrams
Handy Man

Job

- workerid
- memberIdid
- Grade
- semester

Provides

Job/Skill

Skill

Worker/Skill

Accepts

Worker

Acceptance

N

Takes

N

1

1

1

1

JobID

WorkerID

Salary

Acceptance

N

JobID

WorkerID
E-R Diagrams

Course Offerings

There are several courses on the books at a university which are offered from time to time. Students enroll for these courses. We assume there are no pre-requisites for the courses. Please create an E-R Diagram for the database. Assume reasonable attributes for the different entities and also write any assumptions that you make.
E-R Diagrams

Course Offerings
E-R Diagrams

Conference Center Bookings

Instructions: Draw an entity relationship diagram for the case, stating any assumptions you deem necessary.

Case: A conference centre takes bookings from clients who wish to hold courses or conferences at the centre. When clients make bookings they specify how many people are included in the booking, and of these, how many will be resident during the booking, and how many will require catered or non-catered accommodation at the centre. The centre contains a number of facilities which may be required by clients making bookings as follows:

A. There are 400 bedrooms for clients who will be resident during the Course or conference.
B. A maximum of 250 catered people can be handled at any one time.
C. Six main lecture theatres providing seating for 200 people.
D. Twenty seminar rooms each able to accommodate 25 people.
E. Video conference facilities. The video conference facilities consist of four separate video conference networks. Each video conference network has a large screen based in one of the main lecture theatres, along with 3 satellite screens each of which is based in one of the seminar rooms.
The London and Ireland Shipping Company PLC (LISC) was founded in 1852 and owns a fleet of cargo ships. The company had historically run passenger liners, but recent policy decisions involved the sale of all passenger-carrying vessels. The company currently has 14 vessels, including one oil tanker and one tugboat operating out of Liverpool. Most of the vessels are registered in Liberia for tax reasons.

Each ship has one or more holds divided into spaces. The holds are defined by steel bulkheads and the spaces are defined by shelf racks or other physical dividers. Sister ships, built by the same shipbuilders and to the same designs have similar names, such as *Pride of Ireland*, *Queen of Ireland*, *Song of Ireland* and *Warrior of Ireland*. Sister ships also have identical cargo storage facilities.

LISC issues contracts to agents for one or more manifests (lists of cargo items to be shipped). LISC's charges for cargo carried are based on the number of spaces the cargo requires for storage. The types of cargo typically carried by LISC include grain, coal and ores (carried only in ships equipped with bulk cargo holds). They also transport sacked grain, heavy cases, containers (which may be carried on deck), pallets and so on.

Cargo items may take up less than one space in a hold, or one or more spaces, depending on the size of the item. A space may therefore contain several small cargo items.

The ships owned by LISC are kept as busy and as full as possible, in order to maximise the profits that each vessel makes and minimise running & operating costs. LISC's ships ply most of the seas of the world, but tend to operate mainly in the Mediterranean, the North and Mid Atlantic and the Indian Ocean. Different ships require different crew complements.

LISC intends to create a computer based information system that will be able to perform the following tasks:

- Record the voyages of each ship with the start and end ports.
- Record the cargo held by a ship on each voyage
- Keep records of their employees and the ships they are assigned to
- Producing invoices for agents and customers
- Keep a record of customers' payments on invoices
- Analyse the efficiency of use of cargo space and of percentage wasted cargo space for ships voyages
Databases

Translating ERD to Database

- Each entity becomes a new table
- Each attribute becomes a column of the table
- Relationships
  - One to many relationship: Add the key from the many side of the relationship to the one side of the relationship
  e.g. add the publisher id to the book entity
  The pubID is called a foreign key because this is a key to a foreign entity
  - One to one relationship: Add the primary key of one entity to the other entity
  - Many to many relationship: We cannot add the foreign key of each to the other considering these as two one to many relationships. This leads to duplicated rows.
  - To implement a many-to-many relationship add an artificial entity to break the many-to-many relationship into two one-to-many relationships.
Databases

Integrity

• Referential Integrity: Each value of foreign key must have a matching value in the referenced key.
  – Otherwise we will have a dangling reference
e.g. If there was no publisher matching PubID then we have a problem.
• Violations can occur in two ways
  – We add a new entry in the books table with no corresponding publisher in the publisher table
  – We delete an entry in the referenced table without deleting the corresponding entry in the referencing table
• Two ways to ensure Integrity
  – Cascading updates: If a referenced key is changed then all matching entries in the foreign key are automatically updated
  – Cascading deletions: If a value of the referenced table is deleted by deleting a row then all rows in the referencing table that refer to the deleted key value will also be deleted.
Databases

Summary

• Modeling Databases
  – Abstract representation of the database (i.e., Entities, Attributes & Relationships)
  – Super Keys, Keys and Primary Keys
  – E-R diagrams (Semantic Network)