Last Update: October 23, 2011 -- 7PM

Chapter 2: Data Modeling

Presenting the Chapter

This is the second chapter of material that must be covered before introducing the main concepts of database management. Just as systems analysis with its entity-relationship (E-R) diagrams naturally precedes database design in practice, so must a chapter on data modeling with E-R diagrams precede the database concepts material in a database textbook. We all appreciate that while database tables can be very elegant, there is always a need for being able to visualize the data and business environment using these diagrams. Furthermore, covering this material is essential since it will be needed when discussing logical database design a few chapters hence, which is based on converting E-R diagrams into relational database tables.

Depending on the curriculum in your school, your students may have been exposed to entity-relationship diagrams in a prior course on systems analysis and design, in which case this chapter will obviously be review. If this is their first exposure to the subject, then you can present this new material as something that is needed in both systems analysis and in database design, describing at a high level that these diagrams will eventually lead to relational database tables.

In any case, the material in this chapter is designed to introduce E-R diagrams as well as the essential database concepts of unary, binary, and ternary relationships, cardinalities, intersection data, associative entities, dependent entities, and so forth. The chapter takes a progressive approach, starting with the easiest cases of simple binary relationships and building up to the more complicated relationships. Clearly, it is important to lay a solid foundation in this material before proceeding on to database concepts.

Discussion Stimulation Points

The best piece of advice for presenting this chapter on data modeling is to engage the students in asking for examples of the various relationships and other concepts that are presented in the chapter. This can either be based on some readily understood business environment or on a student-oriented environment such as a university setting. These can then be discussed in parallel with the examples given in the chapter. Here is a set of examples for the university environment:

(Note that Exercise 1 in the chapter is based on a high school environment, which, if you assign it, will tie in nicely to discussing a university environment in class.)
One-to-one binary relationship: students and backpacks.
One-to-many binary relationship: departments and courses.

Many-to-many binary relationship: students and courses. Intersection data is date taken and grade received.
One-to-one unary relationship: students are paired with each other so that if one misses school one day the other gets her assignments *(similar to the example in the book.)*
One-to-many unary relationship: some professors report to other professors who are department chairpersons.
Many-to-many unary relationship: professors are chairpersons of committees on which other professors serve.
Ternary relationship: which professor taught which subject to which student. Intersection data is date taken and grade received.

Answers to End-of-Chapter Questions

1. *What is data modeling? Why is it important?*

   Data modeling is the diagramming and recording of a business environment's entities and their interrelationships.

2. *What is the Entity-Relationship model?*

   The Entity-Relationship model is a diagramming technique used for data modeling.

3. *What is a relationship?*

   A relationship is the way in which different entities are associated with each other.

4. *What is the difference between a unary relationship, a binary relationship, and a ternary relationship?*

   The difference is in the number of different entity types involved in the relationship. A unary relationship is between entities of the same entity type (i.e. there is one entity type involved.) A binary relationship is between entities of two different entity types. A ternary relationship is between entities of three different entity types.

5. *Explain and compare the cardinality of a relationship and the modality of a relationship.*

   The cardinality is the maximum number of entities that can be involved in a relationship. The modality is the minimum number of entities that can be involved in a relationship.

6. *Explain the difference between a one-to-one, a one-to-many, and a many-to-many binary relationship.*

   In a one-to-one binary relationship, an occurrence of one entity type is associated with exactly one occurrence of the other entity type and vice versa. In a one-to-many binary

relationship, an occurrence of one entity type is associated with many occurrences of the other entity type, while each occurrence of the second entity type is associated with exactly one occurrence of the first entity type. In a many-to-many binary relationship, an occurrence of one entity type is associated with many occurrences of the other entity type and vice versa.

7. *What is intersection data in a many-to-many binary relationship? What does the intersection data describe?*

Intersection data describes the relationship between two entities in a many-to-many relationship.

8. *Can a many-to-many binary relationship have no intersection data? Explain.*

Yes, it is possible that a many-to-many binary relationship has no intersection data. It can simply be a case of there being a many-to-many relationship between two entity types with no attributes that describe the relationship.

9. *Can intersection data be placed in the entity box of one of the two entities in the many-to-many relationship? Explain.*

No. This would imply that the intersection data describes one of the two entities. It does not. It describes the relationship between the two entities.

10. *What is an associative entity? How does intersection data relate to an associative entity?*

It is a new entity created from the many-to-many relationship between two existing entities. Intersection data comprises the attributes of an associative entity.

11. *Describe the three cases of unique identifiers for associative entities.*

In one case, the unique identifier is the combination of the unique identifiers of the two entities involved in the relationship. In another case, an additional field or fields, often of a time-dependent nature, must be added to the combination of the two unique identifiers. In the third case, the associative entity has its own unique identifying field.

12. *Describe the concept of the unary relationship.*

A unary relationship associates occurrences of an entity type with other occurrences of the *same* entity type.

13. *Explain how a unary relationship can be described as one-to-one, one-to-many, and many-to-many if there is only one entity type involved in the relationship.*

The issue is not how many entity types are involved. Instead, in the unary case, the issue is to focus on one entity occurrence and determine how many other entity occurrences (of the same entity type, since the relationship is unary) it is associated with. The same consideration must then be made "in the other direction."

14. *Describe the ternary relationship concept.*

A ternary relationship is one that involves three entity types.

15. *Can a ternary relationship have intersection data? Explain.*

Yes. A ternary relationship can, in effect, be considered an expanded form of a many-to-many relationship.

16. *What is a dependent entity? (See the description in the General Hardware example.)*

A dependent entity is an entity that only exists in the business environment as long as another entity (on which it is dependent) exists in the business environment.

Solutions to End-of-Chapter Exercises

1. *Draw an entity-relationship diagram that describes the following business environment.*

   The city of Chicago, IL, wants to maintain information about its extensive system of high schools, including its teachers and their university degrees, its students, administrators, and the subjects that it teaches.

   Each school has a unique name, plus an address, telephone number, year built, and size in square feet. Students have a student number, name, home address, home telephone number, current grade, and age. Regarding a student's school assignment, the school system is only interested in keeping track of which school a student currently attends. Each school has several administrators, such as the principal and assistant principals. Administrators are identified by an employee number and also have a name, telephone number, and office number.

   Teachers are also identified by an employee number and each has a name, age, subject specialty such as English (assume only one per teacher), and the year that they entered the school system. Teachers tend to periodically move from school to school and the school system wants to keep track of the history of which schools the teacher has taught in, including the current school. Included will be the year in which the teacher entered the school and the highest pay rate that the teacher attained at the school. The school system wants to keep track of the universities that each teacher attended, including the degrees earned and the years in which they were earned. The school system wants to record each university's name, address, year founded, and Internet URL (address.)

Some teachers, as department heads, supervise other teachers. The school system wants to keep track of these supervisory relationships but only for teachers' current supervisors.

The school system also wants to keep track of the subjects that it offers (e.g. French I, Algebra III, etc). Each subject has a unique subject number, a subject name, the grade level in which it is normally taught, and the year in which it was introduced in the school system. The school system wants to keep track of which teacher taught which student which subject, including the year this happened and the grade received.
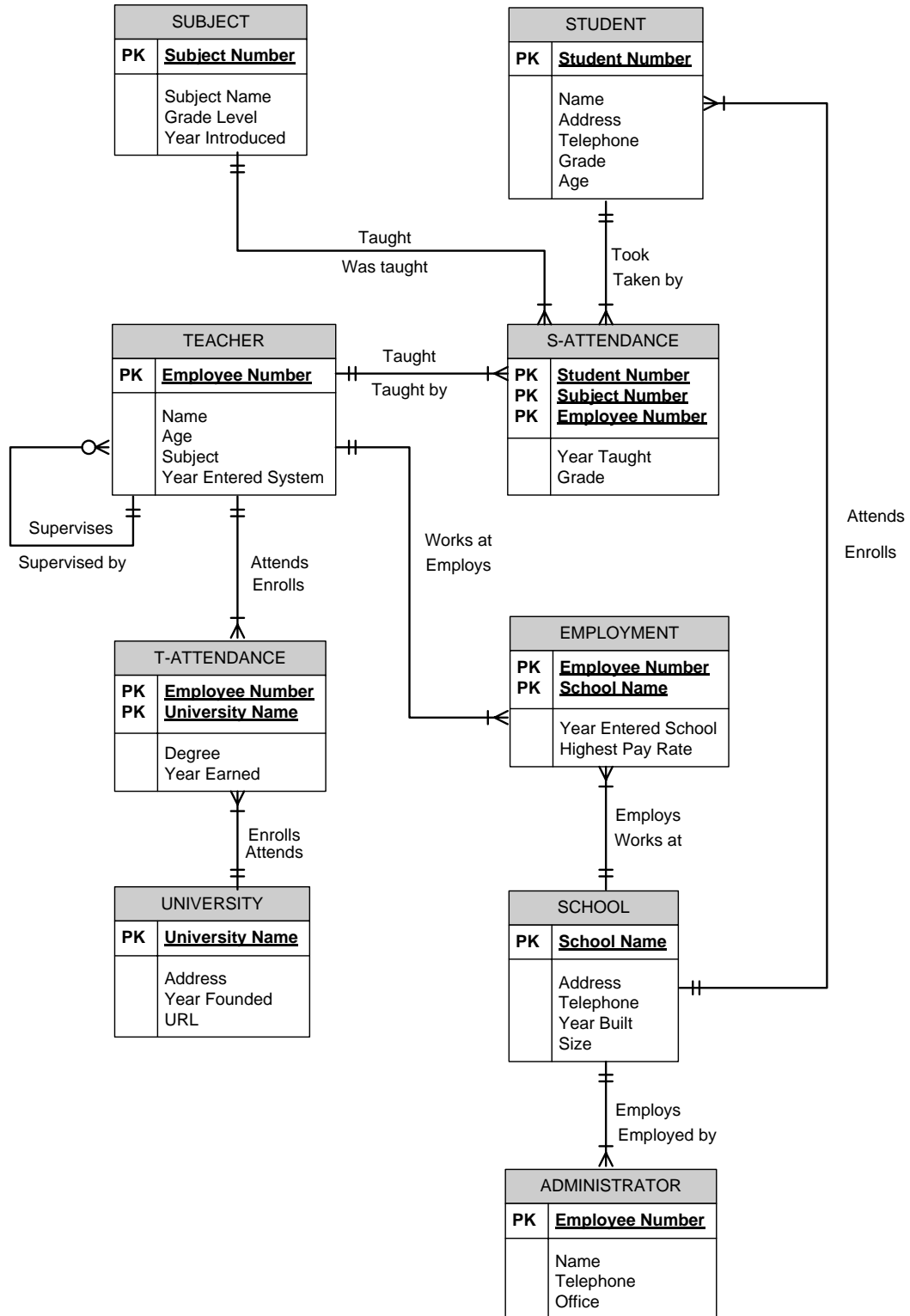
2. *The Entity-Relationship diagram [shown with this question in the text] describes the business environment of Video Centers of Europe, Ltd., which is a chain of videotape*

and DVD rental stores. Write a verbal description of how VCE conducts its business, based on this E-R diagram.

Begin with the movie entity type. Title is the unique identifier while length and year made are other attributes. According to the actor entity type, the actor name is the unique identifier; date of birth and nationality are additional attributes. Based on the cardinalities and modalities between actor and movie, we can infer that the company only wants to keep track of the star of each movie, since each movie is associated with only one actor. Also, an actor may have starred in many movies or may be in the database even if she has not starred in any movies.

According to the diagram, VCE must have at least one and may have many tapes or DVDs of a particular movie. Each tape or DVD has a unique serial number and a type. In turn, as we would expect, each store has many tapes and DVDs for rent. Obviously, a particular tape or DVD belongs to only one store. Store number is the unique identifier for stores; city, country, and telephone number data are also kept.

Customers are uniquely identified by a customer number. Other customer attriubutes include name, address, and telephone number. To be in the database as a customer, a person must have rented a tape or DVD from VCE at least once and possibly many times. Tapes and DVDs stay in the database even if they have never been rented. When a customer rents a tape or DVD, VCE records the date of the rental and the rental price.
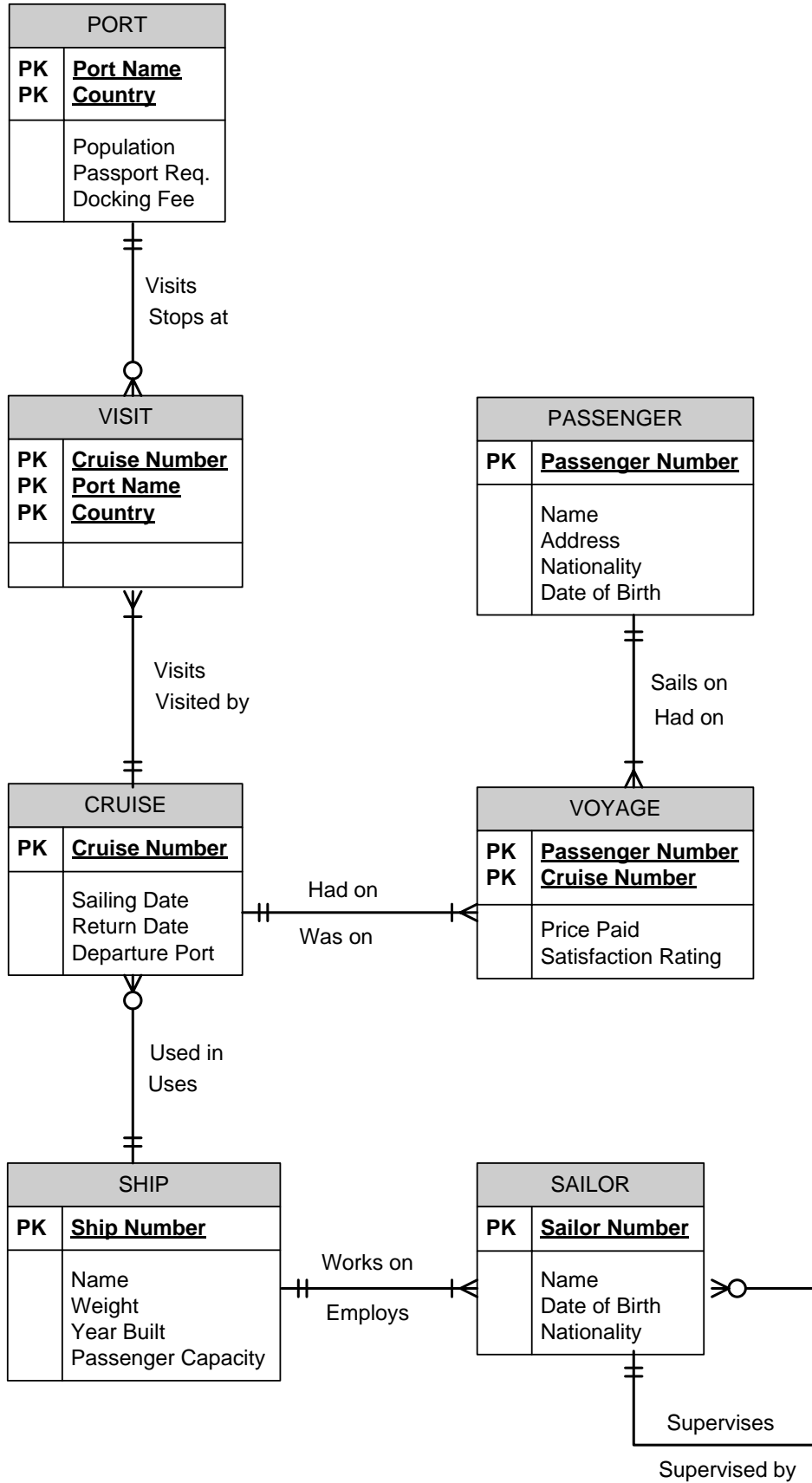
Solutions to End-of-Chapter Minicases

1. Draw an entity-relationship diagram that describes the following business environment.

Happy Cruise Lines has several ships and a variety of cruise itineraries, each involving several ports-of-call. The company wants to maintain information on the sailors who currently work on each of its ships. It also wants to keep track of both its past and future cruises and of the passengers who sailed on the former and are booked on the latter.

Each ship has at least one and, of course, normally many sailors on it. The unique identifier of each ship is its ship number. Other ship attributes include ship name, weight, year built, and passenger capacity. Each sailor has a unique sailor identification number, as well as a name, date-of-birth, and nationality. Some of the sailors are in supervisory positions, supervising several other sailors. Each sailor reports to just one supervisor. A cruise is identified by a unique cruise serial number. Other cruise descriptors include a sailing date, a return date, and a departure port (which is also the cruise's ending point.) Clearly, a cruise involves exactly one ship; over time a ship sails on many cruises, but there is a requirement to be able to list a new ship that has not as yet sailed on any cruises, at all. Each cruise stops at at least one and usually several ports-of-call, each of which is normally host to many cruises, over time. In addition, the company wants to maintain information about ports that it has not

as yet used in its cruises but may use in the future. A port is identified by its name and the country that it is in. Other information about a port includes its population, whether a passport is required for the passengers to disembark there, and its current docking fee, which is assumed to be the same for all ships. Passenger information includes a unique passenger number, name, home address, nationality, and date-of-birth. A cruise typically has many passengers on it (certainly at least one). Hoping for return business, the company assumes that each passenger may have sailed on several of its cruises (and/or may be booked for a future cruise.) For a person to be of interest to the company, he or she must have sailed on or be booked on at least one of the company's cruises. The company wants to keep track of how much money each passenger paid for (or will pay for) each of their cruises, as well as their satisfaction rating of the cruise, if it has been completed.

PORT	
PK **PK**	**Port Name** **Country**
	Population Passport Req. Docking Fee

Visits
Stops at

VISIT	
PK **PK** **PK**	**Cruise Number** **Port Name** **Country**

Visits
Visited by

PASSENGER	
PK	**Passenger Number**
	Name Address Nationality Date of Birth

Sails on
Had on

CRUISE	
PK	**Cruise Number**
	Sailing Date Return Date Departure Port

Had on
Was on

VOYAGE	
PK **PK**	**Passenger Number** **Cruise Number**
	Price Paid Satisfaction Rating

Used in
Uses

SHIP	
PK	**Ship Number**
	Name Weight Year Built Passenger Capacity

Works on
Employs

SAILOR	
PK	**Sailor Number**
	Name Date of Birth Nationality

Supervises
Supervised by

2. *Draw an entity-relationship diagram that describes the following business environment.*

   The Super Baseball League wants to maintain information about its teams, their coaches, players, and bats. The information about players is historic. For each team, the league wants to keep track of all of the players who have ever played on the team, including the current players. For each player, it wants to know about every team the player ever played for. On the other hand, coach affiliation and bat information is current, only.

   The league wants to keep track of each team's team number, which is unique, its name, the city in which it is based, and the name of its manager. Coaches have a name (which is only assumed to be unique within its team) and a telephone number. Coaches have units of work experience, which are described by the type of experience and the number of years of that type of experience. Bats are described by their serial numbers (which are only unique within a team) and their manufacturer's name. Players have a player number that is unique across the league, a name, and an age.

   A team has at least one and usually several coaches. A coach works for only one team. Each coach has several units of work experience or may have none. Each unit of work experience is associated with the coach to whom it belongs. Each team owns at least one and generally many bats. Currently and historically, each team has and has had many players. To be of interest to the league, a player must have played on at least one and possibly many teams during his career. Further, the league wants to keep track of the number of years that a player has played on a team and the batting average that he compiled on that team.

BAT	
PK	**Bat Number**
	Manufacturer

PLAYER	
PK	**Player Number**
	Name Age

Owns

Owned by

Plays

Played on

TEAM	
PK	**Team Number**
	Name City Manager

Employs

Employed by

EMPLOYMENT	
PK **PK**	**Team Number** **Playerf Number**
	Years Batting Average

Employs

Employed by

COACH	
PK **PK**	**Team Number** **Coach Name**
	Telephone

Has

Had by

WORK EXPERIENCE	
PK **PK** **PK**	**Team Number** **Coach Name** **Experience Type**
	Years