GLOBAL EDITION

Processes, Systems, and Information

An Introduction to MIS

SECOND EDITION

Earl McKinney Jr. • David M. Kroenke

ALWAYS LEARNING

PEARSON

Processes, Systems, and Information

An Introduction to MIS

FIGURE 4-29
Report for League DB, Revision 1

example, Sark Justin changes his email, three rows will need to be changed. Hence, this new table is vulnerable to data integrity problems; it is not normalized. We need to fix it in the next revision.

League Database, Revision 2

Consider Figure 4-30, which shows an E-R model of the league database after the revision just described. The changed entity is shown in brown, and the new attribute, *Season,* is shown in blue. Neither the *Checkout* nor *Equipment* entities have been changed, so their attributes are omitted for simplicity.

ADDING THE COACH ENTITY Examining the model in Figure 4-30, we can see that *Team_Season* has two themes; one is about the team in a given season, and the second is about a coach and his or her email. Using our normalization criterion, we know that each entity should have a single theme. So, we decide to move the *Coach* attributes from *Team_Season* into a new entity called *Coach*, as shown in Figure 4-31.

OK so far, but what is the relationship between *Coach* and *Team*? If we look at the data in Figure 4-29, it appears that a given coach can coach many teams, in the same or different seasons (note Fred Dillingham, the coach who took off with the gear). It also appears that a team has at most one coach. Thus, the relationship from *Coach* to *Team_Season* seems to be 1:N.

However, it is dangerous to make such conclusions from sample data. We might just have an odd set of data. An experienced database design team knows to interview users (which could be you) to find out. In this case, let's assume that the 1:N relationship is correct.

The decisions yield the E-R diagram in Figure 4-32. Before we continue, notice we've added *Amount Due* to the *Coach* entity. The idea behind this addition is that at the end of the sport's season, an application program will compute the amount due based on the current

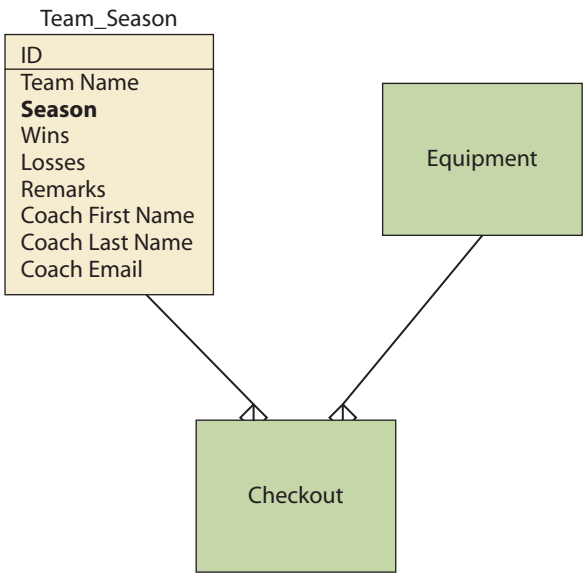
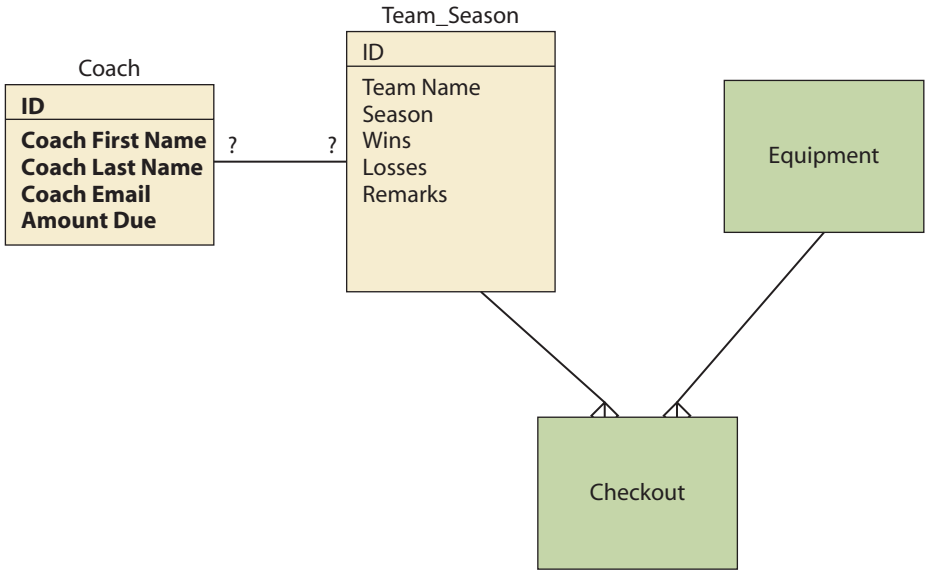


FIGURE 4-30
League E-R Diagram, Revision 1

FIGURE 4-31
League E-R Diagram with Coach Entity



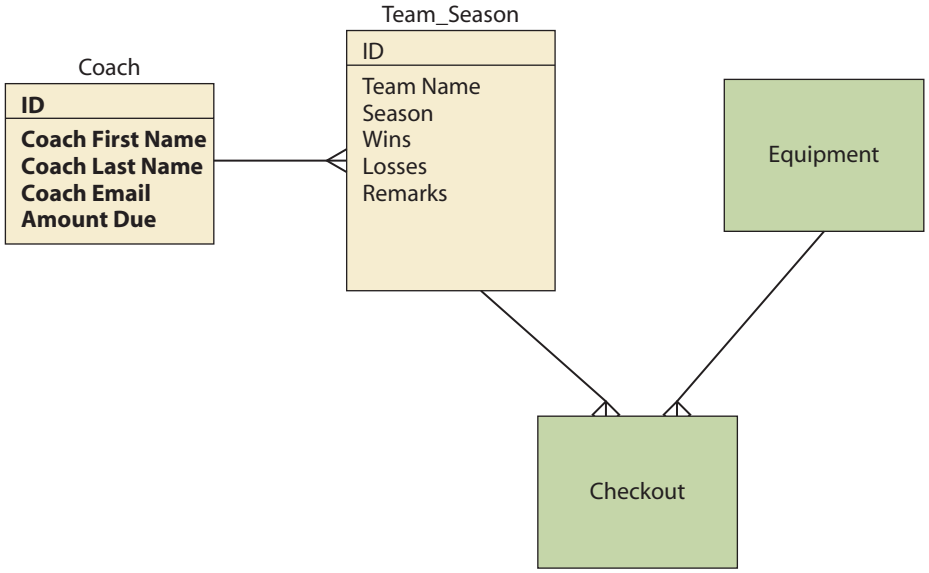
cost of equipment that has not been returned. As equipment is returned, this amount will be decremented appropriately.

REPRESENTING THE RELATIONSHIP IN THE DATABASE DESIGN As described in Q5, we represent a 1:N relationship by adding the key of the parent (the entity on the 1 side) to the child (the entity on the many side). Here we need to add the key of *Coach* to *Team_Season*. Figure 4-33 shows the result; *CoachID* in *Team_Season* is a foreign key that references *ID* in *Coach*.

With this design, every table has a single theme and is normalized. This design is therefore not subject to data integrity problems. Note, however, that it will be necessary to join rows in the table together to produce reports. DBMS products are programmed to do that efficiently, however. The report in Figure 4-34 shows equipment that has been checked out by coaches but has not yet been returned. This report was created by joining data in all four of the tables in Figure 4-33 together.

With these two changes, the intramural league can now allocate equipment checkouts to specific coaches. These changes, in and of themselves, will not solve the league's problem, but it will at least allow the league to know definitively who checked out what equipment. The complete solution to the problem requires a change in process as well.

FIGURE 4-32
League E-R Diagram, Revision 2



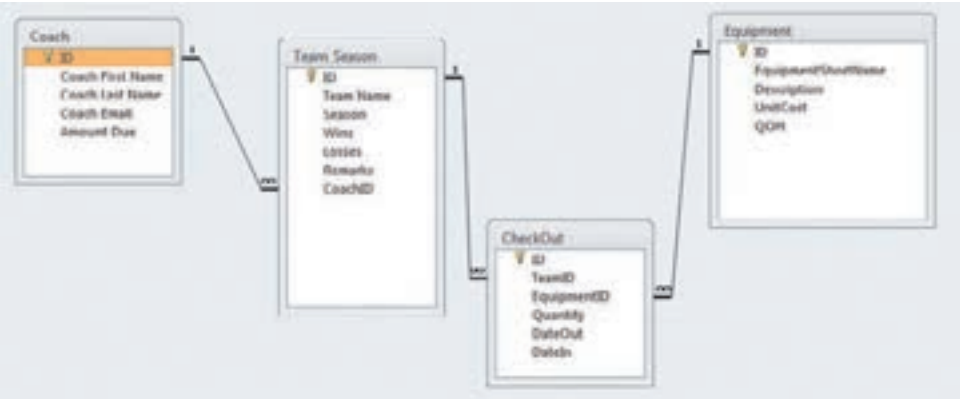


FIGURE 4-33

League Database Design, Revision 2

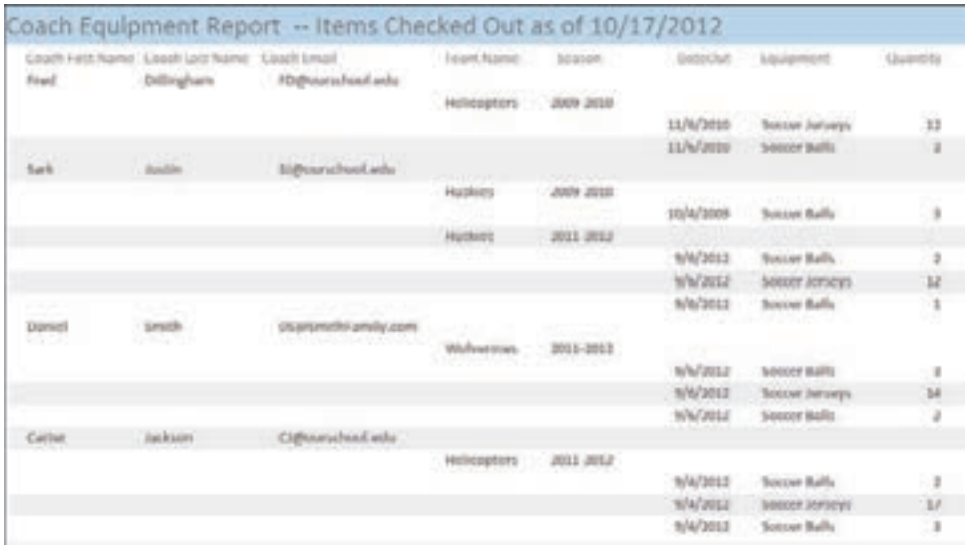


FIGURE 4-34

Report Showing Equipment Still Checked Out to Coaches

Ethics Guide

Querying Inequality?

MaryAnn Baker works as a data analyst in human relations at a large, multinational corporation. As part of its compensation program, her company defines job categories and assigns salary ranges to each category. For example, the category M1 is used for first-line managers and is assigned the salary range of $75,000 to $95,000. Every job description is assigned to one of these categories, depending on the knowledge and skills required to do that job. Thus, the job titles Manager of Customer Support, Manager of Technical Writing, and Manager of Product Quality Assurance are all judged to involve about the same level of expertise and all are assigned to category M1.

One of MaryAnn's tasks is to analyze company salary data and determine how well actual salaries conform to established ranges. When discrepancies are noted, human relations managers meet to determine whether the discrepancy indicates a need to:

- Adjust the category's salary range;
- Move the job title to a different category;
- Define a new category; or
- Train the manager of the employee with the discrepancy on the use of salary ranges in setting employee compensation.

MaryAnn is an expert in creating database queries. Initially she used Microsoft Access to produce reports, but much of the salary data she needs resides in the organization's Oracle database. At first she would ask the IS Department to extract certain data and move it into Access, but over time she learned that it was faster to ask IS to move all employee data from the operational Oracle database into another Oracle database created just for HR data analysis. Although Oracle provides a graphical query interface like that in Access, she found it easier to compose complex queries directly in SQL, so she learned it and within a few months was a SQL expert.

"I never thought I'd be doing this," she said. "But it turns out to be quite fun, like solving a puzzle, and apparently I'm good at it."

One day, after a break, MaryAnn signed into her computer and happened to glance at the results of a query that she'd left running while she was gone. "That's odd," she thought, "all the people with Hispanic surnames have lower salaries than the others." She wasn't looking for that pattern; it just happened to jump out at her as she glanced at the screen.

As she examined the data, she began to wonder if she was seeing a coincidence or if there was a discriminatory pattern within the organization. Unfortunately for MaryAnn's purposes, the organization did not track employee race in its database, so she had no easy way of identifying employees of Hispanic heritage other than reading through the list of surnames. But, as a skilled problem solver, that didn't stop MaryAnn. She realized that many employees having Hispanic origins were born in certain cities in Texas, New Mexico, Arizona, and California. Of course, this wasn't true for all employees; many non-Hispanic employees were born in those cities, too, and many Hispanic employees were born in other cities. This data was still useful, however, because MaryAnn's sample queries revealed that the proportion of employees with Hispanic surnames who were also born in those cities was very high. "OK," she thought, "I'll use those cities as a rough surrogate."

Using birth city as a query criterion, MaryAnn created queries that determined employees who were born in the selected cities earned, on average, 23 percent less than those who were not. "Well, that could be because they work in lower-pay-grade jobs." After giving it a bit of thought, MaryAnn realized that she needed to examine wages and salaries within job categories. "Where," she wondered, "do people born in those cities fall in the ranges of their job categories?" So, she constructed an SQL query to determine where within a job category the compensation for people born in the selected cities fell. "Wow!" she said to herself, "almost 80 percent of the employees born in those cities fall into the bottom half of their salary range."

MaryAnn scheduled an appointment with her manager for the next day.

DISCUSSION QUESTIONS

When answering the following questions, suppose that you are MaryAnn:

1. Given these query results, do you have an ethical responsibility to do something? Consider both the categorical imperative (page 48) and the utilitarian (page 68) perspectives.

2. Given these query results, do you have a personal or social responsibility to do something?
3. What is your response if your manager says, "You don't know anything; it could be that starting salaries are lower in those cities. Forget about it."
4. What is your response if your manager says: "Don't be a troublemaker; pushing this issue will hurt your career."
5. What is your response if your manager says: "Right. We already know that. Get back to the tasks that I've assigned you."
6. Suppose your manager gives you funding to follow up with a more accurate analysis and, indeed, there is a pattern of underpayment to people with Hispanic surnames. What should the organization do? For each choice below, indicate likely outcomes:
   a. Correct the imbalances immediately
   b. Gradually correct the imbalances at future pay raises
   c. Do nothing about the imbalances, but train managers not to discriminate in the future
   d. Do nothing
7. Suppose you hire a part-time person to help with the more accurate analysis, and that person is so outraged at the outcome that he quits and notifies newspapers in all the affected cities of the organization's discrimination.
   a. How should the organization respond?
   b. How should you respond?
8. Consider the adage, "Never ask a question for which you do not want the answer."
   a. Is following that adage ethical? Consider both the categorical imperative and utilitarian perspectives.
   b. Is following that adage socially responsible?
   c. How does that adage relate to you, as MaryAnn?
   d. How does that adage relate to you, as a future business professional?
   e. With regard to employee compensation, how does that adage relate to organizations?

Active Review

Use this Active Review to verify that you understand the material in the chapter. You can read the entire chapter and then perform the tasks in this review, or you can read the text material for just one question and perform the tasks in this review for that question before moving on to the next one.

Q1. What is the purpose of a database?

State the purpose of a database. Explain the circumstances in which a database is preferred to a spreadsheet. Describe the key difference between Figures 4-1 and 4-2.

Q2. What are the contents of a database?

Define the term *database*. Explain the hierarchy of data and name three elements of a database. Define *metadata*. Using the example of *Student* and *Office_Visit* tables, show how relationships among rows are represented in a database. Define the terms *key, foreign key*, and *relational database*.

Q3. What are the components of a database application system?

Explain why a database, by itself, is not very useful to business users. Name the components of a database application system and sketch their relationship. Explain the acronym DBMS and name its functions. List five popular DBMS products. Explain the difference between a DBMS and a database. Summarize the functions of a DBMS. Define *SQL*. Describe the major functions of database administration.

Name and describe the components of a database application. Explain the need for application programs. For multi-user processing, describe one way in which one user's work can interfere with another's. Explain why multi-user database processing involves more than just connecting another computer to the network. Define two broad categories of DBMS and explain their differences.

Q4. How do data models facilitate database design?

Explain why user involvement is critical during database development. Describe the function of a data model. Sketch the database development process. Define *E-R model, entity, relationship, attribute*, and *identifier*. Give an example, other than one in this text, of an E-R diagram. Define *maximum cardinality* and *minimum cardinality*. Give an example of three maximum cardinalities and two minimum cardinalities. Explain the notation in Figures 4-18 and 4-19.

Describe the users' role in database development. Explain why it is easier and cheaper to change a data model than to change an existing database. Use the examples of Figures 4-25(c) and 4-26(c) in your answer. Describe two criteria for judging a data model. Explain why it is important to devote time to understanding a data model.

Q5. How is a data model transformed into a database design?

Name the three components of a database design. Define *normalization* and explain why it is important. Define *data integrity problem* and describe its consequences. Give an example of a table from this chapter with data integrity problems and show how it can be normalized into two or more tables that do not have such problems. Describe two steps in transforming a data model into a database design. Using an example not in this chapter, show how 1:N and N:M relationships are represented in a relational database.

Q6. Why are NoSQL and Big Data important?

Explain the origins of NoSQL DBMS products and describe the applications for which they're used. Explain why it is unlikely that relational databases will be replaced by NoSQL databases. Name and explain the three key characteristics of Big Data and differentiate them from relational data. Describe the challenge that organizations face in attempting to glean information from Big Data stores. Explain what needs to happen for that challenge to be an opportunity for business professionals.

Q7. How can the intramural league improve its database?

What two factors caused the problem at the intramural league? Explain the first revision to the database. Explain what was lost in this revision. Explain why the revision caused a data integrity problem. Describe the need for the *Coach* table and justify the decision to model the relationship from *Coach* to *Team_Season* as 1:N. Explain how that relationship was represented in the database.