DATA MODELING

Business rules, the foundation of data models, are derived from policies, procedures, events, functions, and other business objects, and they state constraints on the organization. Business rules represent the language and fundamental structure of an organization. Business rules are important in data modeling because they govern how data are handled and stored. Examples of basic business rules are data names and definitions.

After decades of use, the E-R model remains the mainstream approach for conceptual data modeling. Its popularity stems from factors such as relative ease of use, widespread computer-aided software engineering (CASE) tool support, and the belief that entities and relationships are natural modeling concepts in the real world. The E-R model is most used as a tool for communications between database designers and end users during the analysis phase of database development. The E-R model is used to construct a conceptual data model, which is a representation of the structure and constraints of a database that is independent of software.

Data modeling is the most important part of the systems development process for the following reasons:

1. The characteristics of data captured during data modeling are crucial in the design of databases, programs, and other system components.

2. Data rather than processes are the most complex aspect of many modern information systems and hence require a central role in structuring system requirements.

3. Data modeling facilitates interaction/communication between designer, application programmer, and end user, thus reducing misunderstandings and improving the thoroughness of resultant systems.

4. Data modeling can foster understanding of the organization (rules) for which the data model is being developed.

5. The value of data modeling can be demonstrated as an overall savings in maintenance or development costs by determining the right requirements before the more costly steps of software development and hardware acquisition.

6. Data modeling results in improved data quality because of consistent business data definitions (metadata).

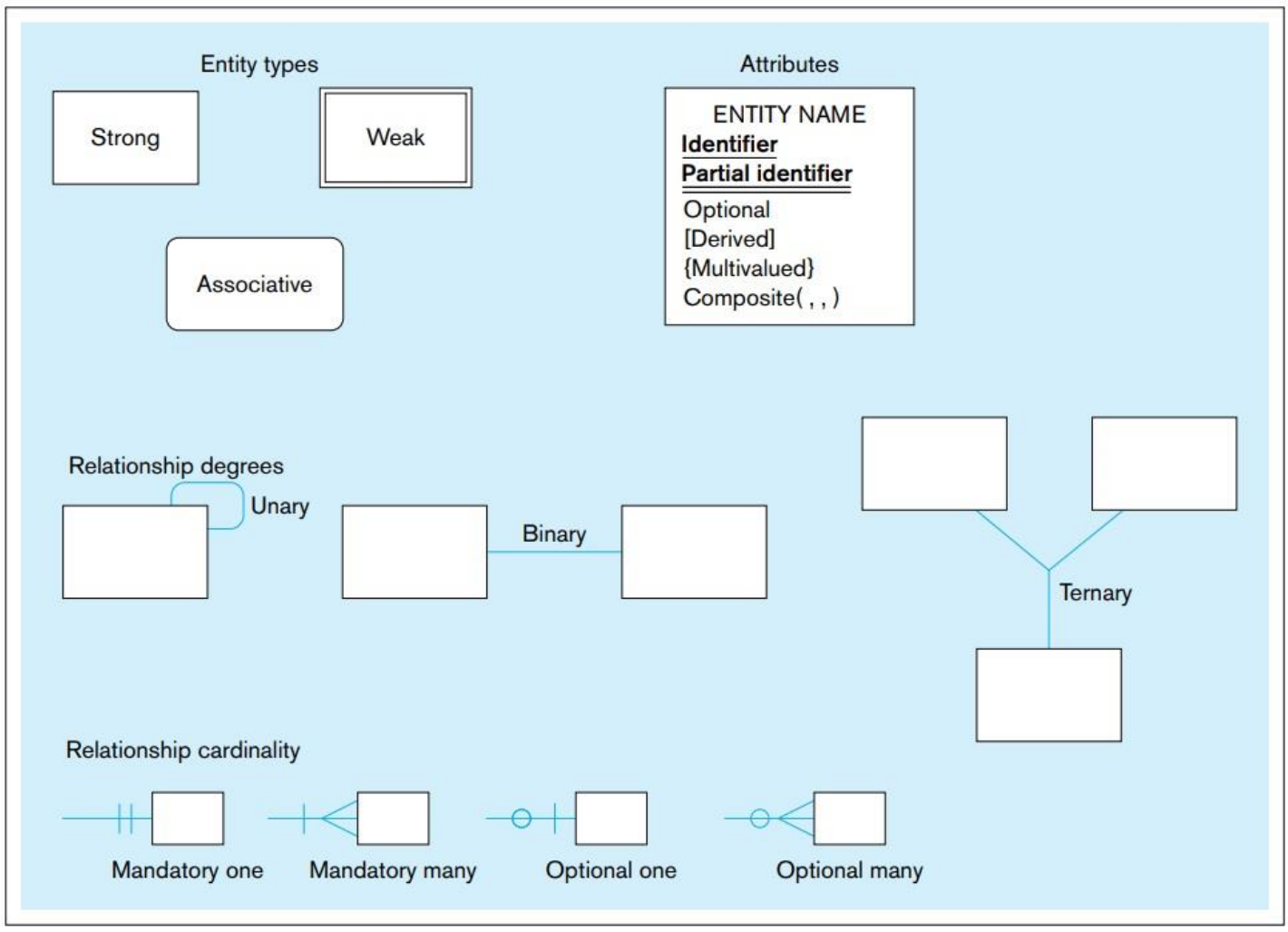
7. Data modeling reduces the significant costs of moving and translating data from one system to another.

THE E-R MODEL: AN OVERVIEW

An **entity-relationship model (E-R model)** is a detailed, logical representation of the data for an organization or for a business area. The E-R model is expressed in terms of entities in the business environment, the relationships (or associations) among those entities, and the attributes (or properties) of both the entities and their relationships. An E-R model is normally expressed as an **entity-relationship diagram (E-R diagram, or ERD),** which is a graphical representation of an E-R model.

The notation we use for E-R diagrams is shown in Figure 2-2.

FIGURE 2-2 Basic E-R notation



MODELING THE RULES OF THE ORGANIZATION

There are actually two business rules for each relationship, one for each direction from one entity to the other. Note that each of these business rules roughly follows a certain grammar:

<entity> <minimum cardinality> <relationship> <maximum cardinality> <entity>

For example, rule is <CUSTOMER> <may> <Submit> <any number> <ORDER>

Your job as a database analyst is to

• Identify and understand those rules *that govern data*

• Represent those rules so that they can be unambiguously understood by information systems developers and users

• Implement those rules in database technology

OVERVIEW OF BUSINESS RULES

A **business rule** is "a statement that defines or constrains some aspect of the business. It is intended to assert business structure or to control or influence the behaviour of the business, rules prevent, cause, or suggest things to happen" For example, the following two statements are common expressions of business rules that affect data processing and storage:

• "A student may register for a section of a course only if he or she has successfully completed the prerequisites for that course."

• "A preferred customer qualifies for a 10 percent discount, unless he has an overdue account balance."

Most organizations (and their employees) today are guided by thousands of combinations of such rules.

GATHERING BUSINESS RULES

Business rules appear (possibly implicitly) in descriptions of business functions, events, policies, units, stakeholders, and other objects. These descriptions can be found in interview notes from individual and group information systems requirements collection sessions, organizational documents (e.g., personnel manuals, policies, contracts, marketing brochures, and technical instructions), and other sources. Rules are identified by asking questions about the who, what, when, where, why, and how of the organization.

DATA NAMES

We will provide specific guidelines for naming entities, relationships, and attributes as we develop the entity-relationship data model, but there are some general guidelines about naming any data object. Data names should

• ***Relate to business, not technical (hardware or software), characteristics;*** so, Customer is a good name, but File10, Bit7, and Payroll Report Sort Key are not good names.

• ***Be meaningful,*** you should avoid using generic words such as *has*, *is*, *person*, or *it*.

• ***Be unique*** from the name used for every other distinct data object; words should be included in a data name if they distinguish the data object from other similar data objects (e.g., Home Address versus Campus Address).

• ***Be readable,*** so that the name is structured as the concept would most naturally be said (e.g., Grade Point Average is a good name, whereas Average Grade Relative To A, although possibly accurate, is an awkward name).

• ***Be repeatable,*** meaning that different people or the same person at different times should develop exactly or almost the same name; this often means that there is a standard hierarchy or pattern for names (e.g., the birth date of a student would be Student Birth Date and the birth date of an employee would be Employee Birth Date).

• ***Follow a standard syntax,*** meaning that the parts of the name should follow a standard arrangement adopted by the organization.