R

# Relationship Cardinality Constraints in Relational Database Design

**Dolores Cuadra**
*Universidad Carlos III de Madrid, Spain*

**Paloma Martinez**
*Universidad Carlos III de Madrid, Spain*

**Elena Castro**
*Universidad Carlos III de Madrid, Spain*

## INTRODUCTION

Database modelling is a complex task that involves conceiving, understanding, structuring and describing real universes of discourse (UD) through the definition of schemata using abstraction processes and data models. Traditionally, three phases are identified in database design: conceptual, logical and physical design. The conceptual modelling phase represents the most abstract level since it is independent of any database management system (DBMS) and, consequently, it is very close to the user and allows him/her to collect almost completely the semantics of the real world to be modelled.

A conceptual schema, independent of the data formalism used, plays two main roles in the conceptual design phase: a *semantic* role, in which user requirements are gathered together and entities and relationships in a UD are documented, and a *representational* role that provides a framework that allows a mapping to the logical design of database development. Three topics are involved in the database conceptual modelling process: data modelling formalism, methodological approach, and CASE tool support. One of the most extended data modelling formalisms, extended entity relationship (EER) model, has proved to be a precise and comprehensive tool for representing data requirements in information systems development, mainly due to an adequate degree of abstraction of constructs that it includes. Although the original ER model was proposed by Chen (1976), many extensions and variations as well as different diagrammatic styles have been defined (Hull & King, 1987; McAllister, 1998; Peckhan & Maryanski, 1988).

In database conceptual analysis, one of the most difficult concepts to be modelled is the relationship concept, especially higher order relationships, as well as its associated cardinalities. Some textbooks (Boman et al., 1997; Ullman & Widom, 2001) assume that any conceptual design can be addressed by considering only binary relationships since its aim is to create a computer-oriented model. We understand the advantages of this approach although we believe that it may produce certain loss of semantics (some biases are introduced in user requirements), and it forces to represent information in rather artificial and sometimes unnatural ways.

Concerning the logical design, the transformation process of conceptual schemata into relational schemata should be performed trying to preserve the semantics included in the conceptual schema completely; the final objective is to keep such semantics in the database itself and not in those applications that access to the database. Nevertheless, sometimes a certain loss of semantics is produced, for instance, *foreign key* and *not null* options in the relational model are not sufficient to control relationship cardinality constraints.

## CONCEPTUAL MODEL REVISED

Central concepts of the ER conceptual model are entities and relationships; these constructs were introduced by Chen (1976) and have been incorporated in other conceptual models although with different names: *class*, *type*, and so forth, for entities and *associations* for relationships. Nevertheless, those concepts do not have a precise semantics, and, consequently, it is necessary to fix their meaning.

Although the entity concept is widely used and accepted, there is no agreement on one definition; for instance, Thalheim (2000) collects 12 different entity denotations. Although experts are not able to give a unique definition, the underlying concept is coincidental in all of them and its usage as design element does not suppose great disadvantages. An entity definition is not given here, just to highlight, according to Thalheim (2000), an entity is a *representation* abstraction with modeling purposes. Date (2004) adds that the represented concept is a distinguishable object, but we do not consider this feature as essential because it depends on the designer's point of view.

The relationship concept is more confused; it is defined as an *association* among entities. This definition offers many interpretations; for instance, in several design methods there are some differences depending on the number of relations can participate in other relationships as in HERM (Thalheim, 2000) by means of association entities as in UML, OMG (2000), or by grouping as clusters a set of entities and relationships (Teorey, 1999). These differences occur because a relationship combines *association* features with *representation* features, and therefore, it might be considered a relationship (if association aspects are highlighted) or an entity (if representation aspects are emphasized).

Cardinality constraint is one of the most important restrictions that can be established in a relationship, and in general, in a conceptual schema. Its functionality is to limit the number of entity occurrences associated in a relationship. Even though it is a simple concept, the definition of this constraint admits several variants.

Two main approaches are discussed: the Chen's constraint that is an extension of the mapping constraint (a special case of cardinality constraint that considers only the maximum cardinality and that for binary relationships can be 1:1, 1:N or N:M) (Chen, 1976); the Chen's constraint has been adopted or extended in different data models and methodologies. On the other side, the MERISE approach (Tardieu, Rochfeld, & Coletti, 1983) incorporates the participation semantics. These two approaches meet each other when cardinality constraints for binary relationships are defined (excepting the natural differences in graphical notations). Both of them represent the same semantics in binary relationships although the way of expressing it is different. Table 1 summarises the semantics associated to cardinality constraint for the occurrences of *A* in the binary relationships *R*.

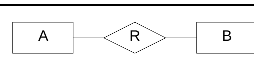
In n-ary relationships, the two approaches Chen and Merise, previously commented, do not represent the same semantics. Table 2 summarises the semantics associated to cardinality constraint for *A* in the n-ary relationship *R* depending of the approach.

## TRANSFORMATION OF CONCEPTUAL SCHEMATA INTO RELATIONAL SCHEMATA

The main difficulty when transforming a conceptual schema into a relational schema is information preservation. Generally, to achieve a complete mapping between both models and keeping their inherent and semantic restrictions from conceptual model to relational model is quite complicated. Usually, restrictions that cannot be applied in the relational model must be reflected in the application in some different way, that is, outside the DBMS. In this way, there are several extensions to the relational model proposed by Codd (1970), Codd, (1979), Date, (2004) and Teorey (1999) that provide a more semantic model.

The principal transformation rules are described in most database textbooks (Date, 2004; Elmasri & Navathe, 2003; Ramakrishnan & Gehrke, 2002), but these rules do not reflect the cardinality constraints transformation. In general, these rules can be classified for binary relationships transformation depending on the relations number that generated (Table 3). Option 1 makes a relation for each constructor, therefore the R relationship is transformed in a relation. Option 2 uses *key propagation* that is applied to a relationship which mapping constraint is 1:N or 1:1. Last, Option 3 only considers just one relation, to store the information of A, B and R constructors. This option

*Table 1. Cardinality constraint in binary relationship summary*




| Cardinality constraint for **A** | Minimum | Maximum |
|---|---|---|
| 0 | **Optional**: there are occurrences of entity A that not participates in the relationship R | Inapplicable |
| 1 | **Mandatory**: all occurrences of entity A participates in the relationship R | **Uniqueness**: there is at most one occurrence of the entity B related to an occurrence of the entity A |
| k>1 | **K-Mandatory**: each occurrence of the A participates at least K times in the relationship | **K-Limit**: There are at most K occurrences of the B related to each occurrence of the A |
| N | Inapplicable | Without limit of maximum participation |