Chapter 3
NoSQL Data Modeling

ABSTRACT

The chapter discusses the necessity for data modeling in NoSQL world. The NoSQL data modeling is a huge challenge because one of the main features of NoSQL databases is that they are schema-free, that is they allow data manipulation without the need for the previous modeling or developing an entity-relationship (ER) or similar model. Although the absence of a schema can be an advantage in some situations, with the increase in the number of NoSQL database implementations, it appears that the absence of a conceptual model can be a source of substantial problems. In order to better understand the need for data modeling in NoSQL databases, first the basic structure of an ER model and an analysis of its limitations are summarized, especially regarding an application in NoSQL databases. The concept and Object modeling notation is presented as one of the possible solutions for data modeling in NoSQL databases.

INTRODUCTION

One of the more complex aspects of database design is the fact that designers, developers, and users typically see and use data in different ways. Unfortunately, if there is no common understanding of how the organization operates and what data is needed, the database design will not fully meet

DOI: 10.4018/978-1-5225-3385-6.ch003

the user's requirements. In order to have an accurate understanding of the nature of data and its use in the organization, it is important to develop a model for communication that, by its nature, would be nontechnical and free of ambiguity or vagueness. In this sense, a model is an abstraction that is focused on the essential, typical aspects of the organization and ignores its accidental characteristics. It needs to provide the basic concepts and notations that would allow database designers and users to communicate with each other unambiguously and accurately and to understand organizational data.

In the terminology of relational databases, the term logical model is used to mean a data model. A logical model represents a global view of the whole database, or as seen from the level of the entire organization. It is the basis for the identification and high-level description of the main data objects while avoiding any specifics of individual database models (Coronel, Morris & Rob, 2011). The logical level provides a view of data independent of the physical implementation. Namely, a logical model is independent of both software and hardware. Software independence means the model is independent of the database management software that is used for its implementation, whereas hardware independence means the model is independent of the hardware that is used for its implementation.

Relational databases are generally based on the Entity-Relationship (ER) logical model. However, as explained in Chapter 2, one of the main features of most NoSQL databases is that they are schema free, or they allow data manipulation without the need for the previous modeling or developing an ER or similar model. Although the absence of a schema can be an advantage in some situations (Chapter 2), with the increase in the number of NoSQL database implementations, it appears that the absence of a conceptual model can be a source of certain problems. That is why this chapter emphasizes the need for data modeling with NoSQL databases as well. In order to better understand the need for data modeling in NoSQL databases, first the basic structure of an ER model and an analysis its limitations are summarized, especially in terms of application in NoSQL databases. Hill's COMN (Concept and Object Modeling Notation) model is then presented. COMN is designed to make it possible to present the real world, or its objects and concepts—data about the real world—but also objects in a computer's memory, or to present everything from the naming of requirements to a functional database running by either a NoSQL or a SQL database management system.