# DFM as a Conceptual Model for Data Warehouse

**Matteo Golfarelli**
*University of Bologna, Italy*

## INTRODUCTION

Conceptual modeling is widely recognized to be the necessary foundation for building a database that is well-documented and fully satisfies the user requirements. In particular, from the designer point of view the availability of a conceptual model provides a higher level of abstraction in describing the warehousing process and its architecture in all its aspects.

Typically conceptual models rely on a graphical notation that facilitates writing, understanding, and managing conceptual schemata by both designers and users. The *Entity/Relationship* (E/R) model (Chen, 1976) is widespread in the enterprises as a conceptual formalism to provide standard documentation for relational information systems; nevertheless, as E/R is oriented to support queries that navigate associations between data rather than synthesize them, it is not well-suited for data warehousing (Kimball, 1998). Actually, the E/R model has enough expressivity to represent most concepts necessary for modeling a *Data Warehouse* (DW); on the other hand, in its basic form, it is not able to properly emphasize the key aspects of the multidimensional model, so that its usage for DWs is expensive from the point of view of the graphical notation and not intuitive (Rizzi, 2006).

Some designers claim that star schemata are expressive enough for conceptual modeling. Actually, a star schema is just a (denormalized) relational schema, so it merely defines a set of relations and integrity constraints. Using star schema for conceptual modeling is like starting to build a complex software by writing the code, without the support of any static, functional, or dynamic model, which typically leads to very poor results from the points of view of adherence to user requirements, maintenance, and reuse. For all these reasons, in the last few years the research literature has proposed several original approaches for modeling a DW, some based on extensions of known conceptual formalisms (e.g. E/R, *Unified Modeling Language* (UML)), some based on ad hoc ones. Remarkably, a comparison of the different models made by Abello (2006) pointed out that, abstracting from their graphical form, the core expressivity is similar, thus proving that the academic community reached an informal agreement on the required expressivity.

This paper discusses the expressivity of an ad hoc conceptual model, the Dimensional Fact Model (DFM), in order to let the user verify the usefulness of a conceptual modeling step in DW design. After a brief listing of the main conceptual model proposals, the basic and advanced features in DW conceptual modeling are introduced and described by examples. Finally, the current trends in DW conceptual modeling are reported and the conclusions are drawn.

## BACKGROUND

In the last few years multidimensional modeling attracted the attention of several researchers that defined different solutions each focusing on the set of information they considered strictly relevant. Some of these solutions have no (Agrawal, 1997; Pedersen, 1999) or limited (Cabibbo, 1998) graphical support, and are aimed at establishing a formal foundation for representing cubes and hierarchies and an algebra for querying them. On the other hand, we believe that a distinguishing feature of conceptual models is that of providing a graphical support to be easily understood by both designers and users when discussing and validating requirements. So we will classify "strict" conceptual models for DWs according to the graphical formalism they rely on that could be either E/R, object-oriented or ad hoc. Some claim that E/R extensions should be adopted since (1) E/R has been tested for years; (2) designers are familiar with E/R; (3) E/R has proved to be flexible and powerful enough to adapt to a variety of application domains; and (4) several important research results were obtained for the E/R model (Franconi, 2004; Sapia, 1999; Tryfona, 1999). On the other hand, advocates of object-oriented models argue that (1) they are more

*Table 1. Comparison between conceptual models (√: supported feature; −: not supported or not explained how to support; p partially supported; QL query language; A: algebra; C: calculus)*

| | Agrawal 1997 | Cabibbo 1998 | Pedersen 1999 | Franconi 2004 | Sapia 1999 | Tryfona 1999 | Luján 2006 | Abello 2006 | Tsois 2001 | Hüsemann 2000 | Rizzi 2006 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reference conceptual model | – | Maths | Maths | E/R | E/R | E/R | UML | UML | – | – | – |
| Formalism adopted to define operations over data | A | AC | A | – | – | – | – | A | – | – | QL |
| Shows how data can be aggregated | P | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ |
| Allows multiple aggregation paths along dimensions | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ |
| Describes hierarchies besides aggregation issues | – | √ | – | – | √ | √ | – | √ | √ | √ | √ |
| Allows multiple measure to be represented | P | – | √ | √ | √ | √ | √ | √ | √ | √ | √ |
| Models summarizzability semantics | – | – | – | – | – | P | √ | √ | – | √ | √ |
| Treats symmetrically measures and dimensions | √ | – | √ | – | – | P | – | P | – | – | – |
| Multiple subjects (fact) can be represented | – | √ | – | – | √ | – | √ | √ | – | √ | √ |
| Models many-to-many rel. within hierarchies | P | – | √ | √ | – | √ | √ | √ | √ | – | √ |

## Related Content

Data Mining for Lifetime Value Estimation
Silvia Figini (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 431-437).*
www.irma-international.org/chapter/data-mining-lifetime-value-estimation/10856

Subgraph Mining
Ingrid Fischer (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 1865-1870).*
www.irma-international.org/chapter/subgraph-mining/11073

Positive Unlabelled Learning for Document Classification
Xiao-Li Li (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 1552-1557).*
www.irma-international.org/chapter/positive-unlabelled-learning-document-classification/11026

Place-Based Learning and Participatory Literacies: Building Multimodal Narratives for Change
Sharon Peckand Tracy A. Cretelle (2020). *Participatory Literacy Practices for P-12 Classrooms in the Digital Age (pp. 74-94).*
www.irma-international.org/chapter/place-based-learning-and-participatory-literacies/237415

Association Bundle Identification
Wenxue Huang, Milorad Krneta, Limin Linand Jianhong Wu (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 66-70).*
www.irma-international.org/chapter/association-bundle-identification/10799