A Data Warehouse Conceptual Data Model

Enrico Franconi and Anand Kamble[†]
Faculty of Computer Science, Free University of Bozen-Bolzano, Italy
franconi@inf.unibz.it, kamble@inf.unibz.it

Abstract

In this short paper we will briefly introduce a Data Warehouse Conceptual Data Model, which gracefully extends the standard Entity-Relationship (ER) conceptual data model with multi-dimensional aggregated entities. The conceptual data model has a clear model-theoretic semantics grounded on the extension of the standard ER semantics with the $\mathcal{GMD}$ logic-based multi-dimensional data model.

The goal of the work presented here is to extend the standard ER conceptual data model, as defined in the database textbooks, with constructs which allow the modelling of multi-dimensional aggregated entities together with their interrelationships with the other parts of the conceptual schema. An important aspect is that a formal model-theoretic semantics is given to the resulting conceptual data model, by combining the well known first order semantics of standard ER—as described, for example, by [2]—with the model-theoretic semantics of the $\mathcal{GMD}$ logical data model for multi-dimensional information [5, 6]. This work is also based on a similar preliminary work done by one author on the use of Description Logics as a mean to give precise semantics to a data warehouse conceptual data model and to study its computational properties [8]. In this short paper we do not present any formal aspect of the conceptual data model, but we only introduce it by means of an example.

The conceptual data model we have obtained is exemplified in this short paper in Figure 1. A basic multi-dimensional entity such as *Calls* is described in the diagram of the figure using a standard star schema—i.e., it is represented by means of a weak entity with respect to its dimensions. In this example, this basic multi-dimensional entity may be useful for analysing the nature of telephone calls by considering, among others, the dimensions related to the origin and the destination of the calls with respect to the type of the phone point (associated to consumer or business customers). So, the entity *Calls* represents a basic cube whose dimensions are *Date*, *Destination*, *Source*, which are restricted to the basic levels *Day*, *Point*, and again *Point*, respectively. This part of the diagram makes still use of standard constructs.

A first extension of the language can be seen with the *simple aggregated entities*—i.e., non-dimensional aggregations—*Weekday* and *Customertype*, which represent dimensional levels built from the basic dimensional entities *Day* and *Point*, respectively. A simple aggregations aggregate the collections of objects that are in the extension of the aggregated entities. So, in our example, since the entities *Mon*, . . ., *Sun* form a partition of the entity *Day*, the *Weekday* entity denotes exactly seven objects, one for all the Mondays, one for all the Tuesdays, etc. On the other hand, the aggregated entity *Customertype* denotes exactly two objects, one aggregating all consumer phone points, and the other aggregating all business phone points. In this way, by interleaving partitioning and simple aggregations, we are able to construct *level hierarchies* starting from some basic dimensional level. Obvious functional dependencies exist among the levels of a hierarchy, as analysed by [9].

A second extension of the language is the *multi-dimensional aggregated entity*, exemplified in Figure 1 by the entity *Calls-by-Weekday-and-Customertype*. This entity denotes all the cells of a cube whose coordinates are the weekdays of the date of the calls, and the customer types of the originators of the calls. All the necessary constraints that should hold for such an entity as a cube hold, enforced by the $\mathcal{GMD}$-based semantics given to this extension [5, 6]. A multi-dimensional aggregated entity is an entity itself in the ER diagram, and it can have attributes and can be part of further relationships or constraints.

The i.com tool [4, 7, 13] fully implements the extended conceptual data model briefly introduced here, and it is available online at the public web address http://www.inf.unibz.it/~franconi/icom/. i.com allows for the specifi-

This work has been partially supported by the EU projects Sewasie, KnowledgeWeb, and Interop.

[†] Also with Department of Computer Science, University of Manchester, UK.

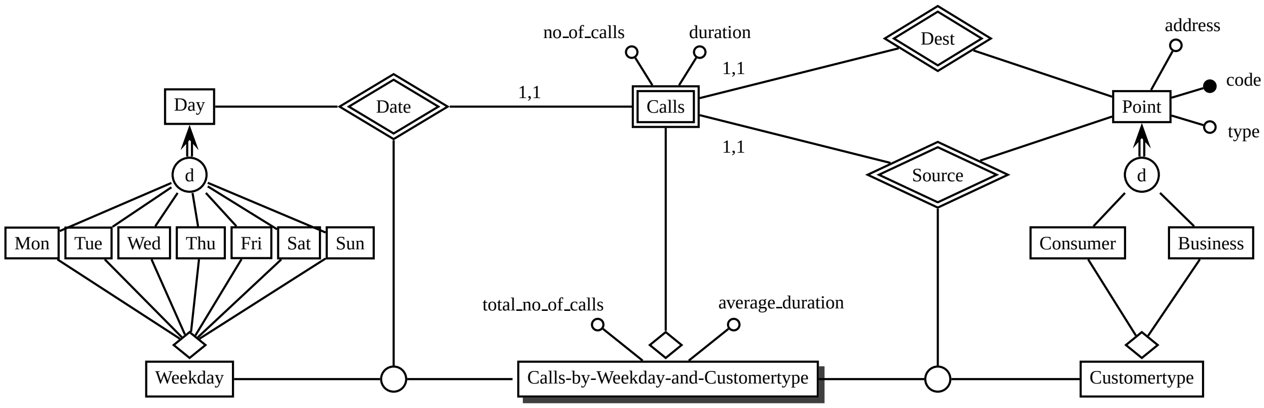

Figure 1. A Data Warehouse Conceptual Schema

cation of multiple extended EER (or UML) diagrams together with a rich set of inter- and intra-schema constraints. Complete logical reasoning is employed by the tool using an underlying description logic inference engine to verify the specification, infer implicit facts and stricter constraints, and manifest any inconsistencies during the conceptual modelling phase. A tutorial is available online.

References

[1] F. Baader, D. McGuinness, D. Nardi, and P. F. Patel-Schneider, editors. *Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, 2002.

[2] A. Borgida, M. Lenzerini, and R. Rosati. Description logics for databases. In Baader et al. [1].

[3] L. Cabibbo and R. Torlone. A logical approach to multidimensional databases. In *Proc. of EDBT-98*, 1998.

[4] E. Franconi, F. Baader, U. Sattler, and P. Vassiliadis. Multidimensional data models and aggregation. In Jarke et al. [12], chapter 5, pages 87–106.

[5] E. Franconi and A. S. Kamble. The $\mathcal{GMD}$ data model for multidimensional information. In *Proc. 5th International Conference on Data Warehousing and Knowledge Discovery*, pages 55–65, 2003.

[6] E. Franconi and A. S. Kamble. The $\mathcal{GMD}$ data model and algebra for multidimensional information. In *Proc. of the 16th International Conference on Advanced Information Systems Engineering (CAiSE-04)*, 2004.

[7] E. Franconi and G. Ng. The ICOM tool for intelligent conceptual modelling. In *Proc. of the 7th International Workshop on Knowledge Representation meets Databases (KRDB'2000)*, 2000.

[8] E. Franconi and U. Sattler. A data warehouse conceptual data model for multidimensional aggregation. In *Proc. of the Workshop on Design and Management of Data Warehouses (DMDW-99)*, 1999.

[9] M. Golfarelli, D. Maio, and S. Rizzi. The dimensional fact model: a conceptual model for data warehouses. *IJCIS*, 7(2-3):215–247, 1998.

[10] K. Hahn, C. Sapia, and M. Blaschka. Automatically generating olap shemata from conceptual graphical models. In *Proc. of the ACM DOLAP 2000 Workshop (DOLAP'2000)*, pages 9–16, 2000.

[11] B. Husemann, J. Lechtenborger, and G. Vossen. Conceptual data warehouse modeling. In *Proc. of the International Workshop on Design and Management of Data Warehouses (DMDW'2000), Stockholm, Sweden, June 5-6*, pages 6–1–6–11, 2000.

[12] M. Jarke, M. Lenzerini, Y. Vassilious, and P. Vassiliadis, editors. *Fundamentals of Data Warehousing*. Springer-Verlag, 1999.

[13] M. Jarke, V. Quix, D. Calvanese, M. Lenzerini, E. Franconi, S. Ligoudistiano, P. Vassiliadis, and Y. Vassiliou. Concept based design of data warehouses: The DWQ demonstrators. In *2000 ACM SIGMOD International Conference on Management of Data*, May 2000.

[14] T. Nguyen, A. M. Tjoa, and R. Wanger. Conceptual multidimensional data model based on metacube. In *Proc. First International Conference on Advances in Information Systems (ADVIS'2000), Izmir, Turkey, 25-27 October*, pages 24–33, 2000.

[15] J. Pei. A general model for online analytical processing of complex data design and evolution. In *Proc. of 22nd International Conference on Conceptual Modeling (ER2003)*, 2003.

[16] C. Phipps and K. C. Davis. Automating data warehouse conceptual schema design and evolution. In *Proc. of the International Workshop on Design and Management of Data Warehouses (DMDW'2002)*, pages 23–32, 2002.

[17] N. Tryfona, F. Busborg, and J. Christiansen. starER: A conceptual model for data warehouse design. In *Proc. of ACM Second International Workshop on Data Warehousing and OLAP (DOAP'99), Kansas City, Missouri, USA, November*, pages 3–8, 1999.

[18] A. Tsois, N. Karayiannidis, and T. Sellis. MAC: Conceptual data modelling for OLAP. In *Proc. of the International Workshop on Design and Management of Data Warehouses (DMDW-2001)*, pages 5–1–5–13, 2001.