A Rule-Based Data Manipulation Language for OLAP Systems*

Mohand-Saïd Hacid[1], Patrick Marcel[2], and Christophe Rigotti[2]

[1] LuFg Theoretical Computer Science
RWTH Aachen, Ahornstraße 55, 52074 Aachen, Germany
hacid@cantor.informatik.rwth-aachen.de
[2] Laboratoire d'Ingénierie des Systèmes d'Information
INSA Lyon, Bâtiment 501, F-69621 Villeurbanne Cedex
Tél : +33 4 72 43 85 88 - Fax : +33 4 72 43 87 13
{*patrick,crig*}*@lisi.insa-lyon.fr*

On-Line Analytical Processing (OLAP) [4] has emerged to support multidimensional data analysis, by providing manipulations and aggregations of data according to multiple dimensions. Foundations of OLAP languages are now a growing field of interest in the database research community.

We introduce an extension of Datalog devoted to the manipulation of data organized in multidimensional cubes. In our data model, data are organized in *cells*. Our language is based on the point of view that a Datalog fact represents an entry (called *cell reference*) in a cube. Associations of cells contents with cells references are represented by *ground atoms* of the form:

$$N(N_1, N_2, \dots, N_p) : \langle N_{p+1}, \dots, N_{p+q}\rangle$$

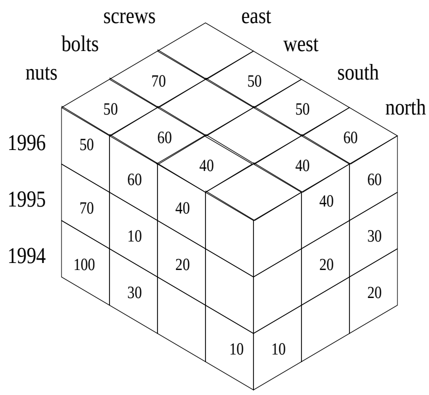
where $N(N_1, N_2, \dots, N_p)$ is the cell reference (i.e., the coordinate of the cell), and the tuple $\langle N_{p+1}, \dots, N_{p+q}\rangle$ denotes the cell contents. A *cube* is simply a set of ground atoms having a common cube name, in which the same reference does not appear more than once to ensure cell monovaluation, and a *multidimensional database* is a set of ground atoms in which the same reference does not appear more than once.

Consider the rule $p(X) \longleftarrow q(X, Y), r(Y)$. The standard (Datalog) informal meaning of this rule is *if q(X,Y) holds and r(Y) holds, then p(X) holds.* The basic intuition of our extension is to read such a rule in the following way: *if there are two cells of references q(X,Y) and r(Y), then there is a cell of reference p(X).* We also add the handling of cell contents, and then a typical rule will be: $p(X) : \langle W\rangle \longleftarrow q(X, Y) : \langle W\rangle, r(Y) : \langle X\rangle$. This rule will be informally read: *if there exists a cell of reference q(X, Y) containing W, and there exists a cell of reference r(Y) containing X, then there exists a cell of reference p(X) containing W.*

This language provides a declarative and concise way to specify the basic standard restructuring and summarizing operations used in OLAP systems. It can be used for example to specify the restructuring from one of the two representations below to the other.

* This work is partially supported by Esprit Basic Research Action no. 22469 - Foundations of Data Warehouse Quality.

3Dsales



2Dsales		1996	1995	1994
nuts	east	50	70	100
	west	60	10	30
	north			10
	south	40	20	
screws	east		10	10
	west	50	50	50
	north	60	30	20
	south	50	60	60
bolts	east	70	50	40
	west		10	40
	north	40	20	
	south		10	

This language is an extension of our previous works [2,5,7]. It has a model-theoretic semantics and an equivalent fixpoint semantics. Its semantics has been set up by combining techniques stemming from Hilog [3], F-logic [8] and Datalog with single-valued data functions [1]. Examples including summary and grouping specification illustrating its applicability for complex OLAP manipulations, as well as formal developments, can be found in the full paper [6].

References

1. S. Abiteboul and R. Hull. Data functions, datalog and negation. In *Proc. ACM SIGMOD*, pages 143–153, Chicago, IL, Jun. 1988.
2. J.-F. Boulicaut, M.-S. Hacid, P. Marcel, and C. Rigotti. Un langage de manipulation de données pour feuilles de calcul. Research report RR-97-01, LISI, INSA de Lyon, Jan. 1997. 24 pages, in french, submitted.
3. W. Chen, M. Kifer, and D.S. Warren. HiLog: a foundation for higher-order logic programming. *JLP*, 15(3):187–230, Feb. 1993.
4. E. F. Codd, S. B. Codd, and C. T. Salley. Providing olap (on-line analytical processing) to user-analysts: An IT mandate. White paper - http://www.arborsoft.com/essbase/wht_ppr/coddTOC.html, 1993.
5. M. S. Hacid, P. Marcel, and C. Rigotti. A rule based CQL for 2-dimensional tables. In *Proc. 2nd Int. Workshop on Constraint Database Systems (CDB'97)*, volume 1191 of *LNCS*, pages 92–104, Delphi, Greece, Jan. 1997.
6. M. S. Hacid, P. Marcel, and C. Rigotti. A rule based data manipulation language for OLAP systems. Research report RR-97-, LISI, INSA de Lyon, May. 1997. http://www.insa-lyon.fr/People/LISI/patrick/DMLOLAP.ps.
7. M. S. Hacid, P. Marcel, and C. Rigotti. A rule-based language for ordered multidimensional databases. In *Proc. 5th Int. Workshop on Deductive Database and Logic Programming (DDLP'97)*, volume 317 of *GMD-Studien*, pages 69–81, Leuven, Belgium, Jul. 1997.
8. M. Kifer, G. Lausen, and J. Wu. Logical foundations of object-oriented and frame-based languages. *JACM*, 42(4):741–843, Jul. 1995.