SimOLAP: A System for the semi-automatic implementation of Simulation Data Warehouses

Sandro Bimonte*, Nicolas Dumoulin*

*Irstea, 9 Avenue Blaise Pascal - CS 20085 - 63178 Aubière
sandro.bimonte@irstea.fr, nicolas.dumoulin@irstea.fr http://motive.cemagref.fr

Abstract. Data Warehouses and OLAP systems allow decision-makers exploring and analyzing huge volumes of data modeled according the multidimensional model, and extracted from heterogeneous data sources. Usually, DW design is a complex, and time and resources consuming task. Then, DW experts are necessary during design and implementation phases. In this paper, we present a new methodology and a tool allowing modelers (DW unskilled users) to design and implement DWs for analyzing simulation results data by themselves, without any intervention of DW experts.

1 Introduction and motivations

Nowadays in order to understand, explain and predict trends, dynamics modeling is extensively used to study complex phenomena and scenarios in different contexts: health, climate change, demography, etc. Nevertheless, to calibrate and validate these models, modelers need to make several replications of each simulation to get representative results, leading to huge volumes of results data sets. Although some efforts have been done to provide modelers with tools do design model experimentations (Reuillon et al., 2013), modelers cannot automatically store their simulation data results and explore them by means of interactive tools to validate their experimentations and discover unknown patterns. Then, modelers need to design and implement DWs by themselves in an incremental and iterative way allowing looking for the best way to analyze their data, without any intervention of DW experts. Thus, a hybrid DW design methodology where conceptual, logical and physical design phases are automatic is mandatory. Moreover, as stated in (Bimonte et al., 2013), DW unskilled users need real OLAP clients to validate generated prototypes according to an agile DW development methodology.

Several works propose to automatically generate multidimensional schema from data sources (Romero and Abelló, 2009) (i.e. data-driven approaches), but to best of our knowledge no work take into account data sources modeled as complex trees used to represent simulation data results. In requirement-driven approaches, the formalization of the requirements is usually expressed by DW experts using complex formalisms such as conceptual modelsor DW/DB languages such as SQL. They have a major drawback: they define a gap between users' intentions and their implementation, since DW experts have to translate them into a formalism that is not comprehensible by modelers. Thus, some works propose using the natural language to express needs and define DW schema, but too many ambiguities are issued for these methods to be useful in complex real projects.