

Toward Active XML Data Warehousing

Rashed SALEM, Jérôme DARMONT and Omar BOUSSAID

Université de Lyon (ERIC - Lyon 2)
5 av. P. Mendès-France, 69500 Bron, France
firstname.lastname@univ-lyon2.fr

Abstract. Warehousing data is not a trivial task, particularly when dealing with huge amounts of distributed and heterogeneous data. Moreover, traditional decision support systems do not feature intelligent capabilities for integrating such complex data. Therefore, we propose an approach for intelligent decision support based on active XML warehousing. We exploit XML as a pivot language in order to unify, model and store complex data. Furthermore, Web services tackle the distribution and interoperability problems of data sources. They are employed for complex data integration and react with active rules for realizing intelligent ETL. In this paper, we focus on the integration phase and propose an architecture for integrating complex data into a repository of Active XML documents, based on Web services and event-driven rules. We have finally developed a software prototype to validate this approach.

1 Introduction

Nowadays, huge volumes of heterogeneous data (e.g., operational data, Web data, text data, multimedia data, etc.) are available over networks. We term these data *complex data* (Darmont et al., 2005). Data may be qualified as complex if they are: diversely structured, represented in various formats, originating from several different sources, described through different perspectives and/or changing in terms of definition or value over time. Complex data must be warehoused into specific storage to be later analyzed for decision-support purposes. The complexity of data renders their structuring and exploitation difficult. However, the classical warehousing approach (Inmon, 1996; Kimball, 1996) is not very adequate when dealing with complex data, particularly in the data integration phase. It is technically difficult, crucial and ill-adapted for handling complex data. Neither classical ETL (Extracting, Transforming and Loading), nor dynamic ETL is alone sufficient for integrating complex data. Therefore, traditional decision support systems need to be extended with intelligent capabilities in order to: increase business competitiveness, automate decision-making, support near real-time decisions, improve consistency in decisions, improve management of data distributed throughout heterogeneous sources and support on-line analysis of complex data. Intelligent ETL is well-suited for achieving most of these promised issues. Intelligent ETL is constituted of autonomous services that are based on an event-driven environment for integrating complex data. It posses a set of services allowing to accomplish in autonomous way the different tasks of the ETL process.