ABSTRACT

Collaborative business making is emerging as a possible solution for the difficulties that Small and Medium Enterprises (SMEs) are having in recent difficult economic scenarios. In fact, collaboration, as opposed to competition, may provide a competitive advantage to companies and organizations that operate in a joint business structure. When dealing with multiple organizations, managers must have access to unified strategic information obtained from the information repositories of each individual organization; unfortunately, traditional Business Intelligence (BI) tools are not designed with the aim of collaboration so the task becomes difficult from a managerial, organizational, and technological point of view. To deal with this shortcoming, the authors provide an integration, mapping-based, methodology for heterogeneous Data Warehouses that aims at facilitating business stakeholders’ access to unified strategic information. A complete formalization, based on graph theory and the RELEVANT clustering approach, is provided. Furthermore, the authors perform an experimental evaluation of the proposed method by applying it over two DW instances.

1. INTRODUCTION

The Data Warehousing is the main Business Intelligence instrument for the analysis of large amounts of operational data. It permits the extraction of relevant information for decision-making processes usually inside one single organizations. In fact, Inmon defines it as a “subject-oriented, integrated, time-variant and non-volatile collection of data in support of management’s decision making process” (Inmon, 1992).
Traditionally, one single Data Mart (a building block of the DW) is focused on a particular aspect or subject area (thus, subject-oriented) and is confined to a single department, and the union of all the company’s Data Mart form the enterprise Data Warehouse.

DW integration is the process of combining strategic information from two or more heterogeneous Data Marts or Data Warehouses with the aim of providing users a unified view over the entire available information. The problem, although more and more frequent, has received little attention so far. There is in fact a series of scenarios where managers need to combine data and information from one or more Data Warehouses in order to obtain a unique overview of different areas of one single enterprise or a network of collaborating enterprises.

For example, in large organizations, different departments usually develop their separate, heterogeneous Data Mart without any knowledge about other departments. It is thus difficult for managers to have a coherent overview of the entire organization and to be able to take strategic decisions concerning all departments. Unfortunately, in such cases, the necessity of integration arises after the Data Warehouse has been built, and presents several difficulties due to the inherent heterogeneity of the data and to the different perspective with which different groups manage and represent the same data and information. The issue may be avoided when the integration goal is clearly defined a priori of the development phase. For example, Kimball proposes a design methodology, the Data Warehouse Bus Architecture for creating and maintaining common analysis dimensions for all DM of the group, eliminating thus schema and instance inconsistencies. The bus architecture creates “conformed dimensions”, as either identical or strict mathematical subsets of the most granular and detailed dimensions (Kimball & Ross, 2002). This way, facts can be analyzed through a simple union of the different Data Marts.

Unfortunately, most of the times developers and analysts lack the vision of creating a cross-organization Data Warehouse from the beginning, or create the individual Data Marts without any intent for integration.

This is also the case when companies collaborate, form alliances, or one company acquires another company. On top of the business structure, managers need to access a series of strategic indicators that describe all the companies involved in the collaborative effort.

One single company’s data is usually managed by a complex set of specialized tools, each with a specific task. Accessing all of it is a difficult objective, per se, because of to the distinct models that are used to represent each specific dataset. That is why accessing data from more than one company is not scalable, as the difficulty of managing the heterogeneity increases exponentially.

The goal of integrating data and structured information is rising also with the increasing adoption of specialized tools by Small and Medium Enterprises (SMEs), which initially used IT only as a support for the operational process. Subsequently, SMEs acknowledged the importance of IT in the strategic behavior of any firm seeking greater competitiveness (Blili & Raymond, 1993). SMEs provide an interesting setting as they are knowledge generators, but are poor at knowledge exploitation (Levy, Loebbecke & Powel, 2001). This means that they rarely capitalize on the data they produce and are not able to transform data and information into knowledge. Cooperation offers SMEs the possibility to access a larger array of resources and knowledge on which to capitalize for developing a deeper business vision and for obtaining competitive advantage.