Fusion Cubes: Towards Self-Service Business Intelligence

Alberto Abelló, School of Informatics, Universitat Politècnica de Catalunya, Barcelona, Spain
Jérôme Darmont, Université de Lyon (Laboratoire ERIC), Lyon, France
Lorena Etcheverry, Computer Science Institute, Universidad de la Republica, Montevideo, Uruguay
Matteo Golfarelli, DISI - Department of Computer Science and Engineering, University of Bologna, Bologna, Italy
Jose-Norberto Mazón, Department of Software and Computing Systems, Universitat d’Alicante, Alicante, Spain
Felix Naumann, Department of Information Systems, Hasso Plattner Institute, Potsdam, Germany
Torben Bach Pedersen, Department of Computer Science, Aalborg University, Aalborg, Denmark
Stefano Rizzi, Department of Computer Science and Engineering, Università di Bologna, Bologna, Italy
Juan Trujillo, Department of Information Systems and Languages, Universitat d’Alicante, Alicante, Spain
Panos Vassiliadis, Department of Computer Science, University of Ioannina, Ioannina, Greece
Gottfried Vossen, Department of Information Systems, Universität Münster, Münster, Germany

ABSTRACT

Self-service business intelligence is about enabling non-expert users to make well-informed decisions by enriching the decision process with situational data, i.e., data that have a narrow focus on a specific business problem and, typically, a short lifespan for a small group of users. Often, these data are not owned and controlled by the decision maker; their search, extraction, integration, and storage for reuse or sharing should be accomplished by decision makers without any intervention by designers or programmers. The goal of this paper is to present the framework we envision to support self-service business intelligence and the related research challenges; the underlying core idea is the notion of fusion cubes, i.e., multidimensional cubes that can be dynamically extended both in their schema and their instances, and in which situational data and metadata are associated with quality and provenance annotations.

Keywords: Business Intelligence, Data Cube, Data Fusion, Data Integration, Data Warehouses, ETL, Metadata Discovery, Metadata Quality, Open Data, Schema Discovery

DOI: 10.4018/jdwm.2013040104
INTRODUCTION

Today’s business and social environments are complex, hyper-competitive, and highly dynamic. When decisions have to be made quickly and under uncertainty in such a context, the selection of an action plan must be based on reliable data, accurate predictions, and evaluations of the potential consequences. Business intelligence (BI) tools provide fundamental support in this direction. For instance, in medium and large companies, BI tools lean on an integrated, consistent, and certified repository of information called a data warehouse (DW), which is periodically fed with operational data. Information is stored in the DW in the form of multidimensional cubes that are interactively queried by decision makers according to the OLAP paradigm (Golfarelli & Rizzi, 2009).

In this work, we call stationary the data that are owned by the decision maker and can be directly incorporated into the decisional process. Stationary data may take either operational or multidimensional form; in both cases, their quality and reliability is under the decision maker’s control. In a corporate scenario, the data stored in the company DW and information system are stationary.

However, well-informed and effective decisions often require a tight relationship to be established between stationary data and other data that fall outside the decision maker’s control (Pérez et al., 2008; Trujillo & Mate, 2011; Golfarelli, Rizzi, & Cella, 2004; Darmont et al., 2005). These valuable data may be related, for instance, to the market, to competitors, or to potential customers, and are sometimes called situational data (Löser, Hueske, & Markl, 2008):

We call situational those data that are needed for the decisional process but are not part of stationary data. Situational data have a narrow focus on a specific domain problem and, often, a short lifespan for a small group of decision makers with a unique set of needs.

In some cases, situational data can be retrieved (for free or for a fee) in a semi-structured form by accessing established data providers, such as DBpedia (Auer et al., 2007, cross-domain), ProductDB (productdb.org, commerce domain), Geonames (sws.geonames.org, geography), or DATA.GOV (www.data.gov, public institutions); for instance, in DBpedia the structured content extracted from Wikipedia is mapped onto a cross-domain ontology and can be queried using SPARQL. In other cases, situational data are chaotically scattered across heterogeneous and unstructured sources available on the Web (e.g., opinions expressed by users on social networks, ratings of products on portals, etc.). In general, situational data tend to be highly dynamic in contrast to stationary data, which are used to address a large set of decisional problems and impose a slow and careful management. A quick comparison of the main features of situational and stationary data is reported in Table 1.

The capability of incorporating situational data into the decision process gives rise to a new class of applications which, in the context of BI 2.0, are often labeled as situational BI, on-demand BI, or even collaborative BI. In this paper we use the term self-service BI to emphasize that the search, extraction, and integration of situational data should be accomplished by users through a continuous interaction with the application, without any mediation or intervention by analysts, designers, or programmers.

As also emphasized by Gartner Inc., self-service BI appears to be the big wave in BI since 2010 (Horwitt, 2010). The key idea is to enable non-expert users to make well-informed decisions when required, by letting them navigate information in a “surf and save” mode, meaning that data can be stored for reuse or sharing.

Among the main applications for self-service BI in a corporate scenario we mention brand positioning, pricing, competitor monitoring, policy planning, and risk management; the domains involved range from retail, telecommunications, and entertainment to finance and public services, such as health and transportation.

This trend also finds an echo at smaller-scale levels, i.e., in non-governmental organizations, watchdog groups (non-profit groups
A Single Pass Algorithm for Discovering Significant Intervals in Time-Series Data
[www.igi-global.com/article/single-pass-algorithm-discovering-significant/1788?camid=4v1a](www.igi-global.com/article/single-pass-algorithm-discovering-significant/1788?camid=4v1a)

Vertical Fragmentation in Databases Using Data-Mining Technique
[www.igi-global.com/article/vertical-fragmentation-databases-using-data/1812?camid=4v1a](www.igi-global.com/article/vertical-fragmentation-databases-using-data/1812?camid=4v1a)