Data Warehouses and Business Intelligence in Croatia:
Do Managers Know How to Use Them?

Kornelije Rabuzin, Faculty of Organization and Informatics, University of Zagreb, Zagreb, Croatia
Darko Škvorc, Faculty of Organization and Informatics, University of Zagreb, Zagreb, Croatia

ABSTRACT

Authors of this paper were directly involved in several data warehousing and business intelligence projects that were carried out in companies in Croatia. Based on the experience from those projects, this paper presents how those solutions were implemented — and more importantly, how they were used by managers at different levels of management. Three different companies are described, including three different implemented solutions. One company was small, one was mid-sized and one was large. All had differences in ownership and education level of managers. Implementation (technical) details are omitted, but this paper shows that managers do not always see what business intelligence can do. Furthermore, it is shown that such projects do not always have to be expensive.

KEYWORDS

Business Intelligence, Data Warehouses, Databases, Managers

INTRODUCTION

In the past decade, many Data Warehouses (DW) have been implemented. The reason is quite simple, i.e., the companies were not able to make decisions without unified structures that are called data warehouses. Many sources that contain data have to be cleaned and integrated, and this is not so easy to do. For our purposes, the data warehouse is a database that is organized according to other design principles than those that are used for good database design. However, once a data warehouse is implemented, reports can be built within seconds and the developed data warehouse can become a real gold mine. Furthermore, it could help in making decisions and improving the business. 60 percent of executive managers state that the use of a performance management tool has a positive impact on shareholder value (https://www.atkearney.com/paper/-/asset_publisher/dVxv4Hz2h8bS/
Building a data warehouse is not as easy as it may seem. Many problems have to be solved along the way and many situations occur that require one’s attention. But generally, three important steps can be distinguished: one has to extract data from different sources, one has to clean and integrate the data and one has to load the data into the data warehouse (Extract Transform Load or ETL procedure).

ETL

Regarding sources of data, many different database management systems (DBMS) can be used; and it is usually best when one such system can be found. The reason is quite simple, i.e., one can use Structured Query Language (SQL) to extract data from such data sources (Rabuzin, 2012; Rabuzin, 2014). More often, flat or Excel files, or even old, not well-documented applications exist as data sources. Other types of data sources may require specialized knowledge and skills in order to extract and to use data. As Liu and Shi (2014) recognized, business data could cause severe problems:

- Business data often encounters quality issues and needs substantial cleaning efforts;
- Business data is large in overall size but cannot be fully shared due to the concern of data security;
- Business data often needs to be cross-referenced with public databases to reveal more information and knowledge.

In recent years, many NoSQL systems became available that contain relevant data. NoSQL systems are usually characterized by large volumes of data that do not share the same structure and that do change quite often. Volume, variety and velocity represent big data challenges, which cause difficulties in capture, storage, search, sharing, analysis and visualization (Abdelhafez, 2014). So there is a problem when one needs to extract data from all those sources, because they are not standardized; each system uses a different query language. For example, graph databases are interesting as they can represent nodes and their relationships. The query language that is used could be Cypher Query Language or Gremlin. For column-oriented databases (e.g., HBase), another specialized language is used. Key-Value databases are not an exception; neither are document-oriented databases. So one needs to be familiar with these systems in order to extract data. For more information on NoSQL databases, one could look at (Redmond & Wilson, 2012) or (Robinson, Webber & Eifrem, 2013).

Furthermore, in recent years Facebook and Twitter are becoming of increasing interest, because people write and publish large amounts of text on them; text analysis (mining) is becoming more and more important. Measuring a user’s opinion and attitudes could be very challenging, but at the same time it is known that people do publish things that could be disturbing or even dangerous. There are an increasing number of cases where people wrote on Facebook that they were going to do certain things that were criminal or dangerous, and then really did those things.

But there are also situations where entities have to extract data from very old data sources that are not documented. In the past, people used hierarchical and network databases. Although these two technologies belong in the past century (they were used several decades ago), our team had a conversation with a friend who confirmed that one such old database system was still used. This person had a project requiring that data be moved to a relational database management system.

In one of the projects that some of this paper’s authors worked on, data had to be extracted from a very old DOS (Disk Operating System) application that was built approx. 25 years ago. Since the data format was antiquated and it was not possible to extract data from the application, a team had