

Implementing BI concepts with Pentaho, an evaluation

Orhan Tuncer
Delft University of Technology
Delft, The Netherlands
Email: o.tuncer@tuncer.nl

Jan van den Berg
Delft University of Technology
Delft, The Netherlands
Email: j.vandenberg@tudelft.nl

Abstract—Since the late nineties of the previous century, many IT companies fell upon Business Intelligence (BI) and promised better business decision making by means of intelligent data analysis. Ever since those days new BI tools have been developed, usually provided as a closed software solution. As a consequence, the details of the underlying technology used are hidden to the user. Recently, an open source solution became available under the name of ‘Pentaho’. The related website announces the Pentaho suite to supply a complete BI enterprise solution that is comprehensive and to cover the full spectrum of BI functions. We decided to do the proof by applying the Pentaho software on a true business case. A secondary research goal was to test whether Pentaho is an appropriate training tool for use in university class rooms. This paper reports on our findings. Based on our experience, we think Pentaho can be a very useful tool in business, although, in order to compete with commercially available BI suites, extensions are necessary related to the presentation and dashboard layer, making it more user-friendly in the way how they are created. For educational purposes, where understanding of BI tools is essential, the Pentaho software is considered to be an excellent candidate simply since the environment forces users to truly comprehend what they are doing.

I. INTRODUCTION

Since the nineties of the previous century, Business Intelligence has become common practice in many (commercial) organizations as an umbrella term to improve business decision making by using fact-based support systems [1]. Ever since those days, new BI tools have been developed, at one time as add-on to existing software packages like enterprise resource planning or database management systems [2], [3], at another time as specialized business analytics software suites [4], [5]. In the cases mentioned, the usual approach has been a closed software solution where details about the technology used are hidden to the user. Therefore, (s)he may not fully understand how the underlying software executes certain tasks. This hampers an accurate assessment of the results that are obtained. In addition, the existing differences in implementation thwart integration of the various tools in order to get a complete set of BI functionalities. Finally, the differences in BI functionality of the various BI suites hinder a precise comparison of the various BI suites available, making the selection of the best suiting product a difficult choice. Based on these observations, we came to the conclusion that there is actually a high need for a comprehensive, fully integrated BI solution with complete BI functionality.

Relatively recently, an open source BI solution became available under the name of ‘Pentaho’ [6]. The corresponding website promises that Pentaho provides a complete and integrated platform with all crucial BI functionality, including data integration, data analysis, data mining, reporting, and dashboard visualization, put together in one easy-to-use BI platform. Having found this new BI platform, we asked ourselves whether this ‘world’s most widely deployed open source’ [6] BI platform is a truly good environment for use in business contexts. As a secondary goal, the question arose whether Pentaho can be applied successfully for educational purposes, namely, in a university classroom setting. To answer these questions we decided to do a test by applying the Pentaho software on a true business case. This paper presents our findings.

The rest of this paper is structured as follows. In section II, we introduce the set of criteria based on which we evaluate the Pentaho BI platform. Next, in section III, we introduce the software platform itself by describing the available functionalities. In section IV, we illuminate the preparation of the case study executed by sketching the context of the case, defining the KPIs selected and describing the BI components used. In section V, we continue by setting out the concrete execution steps of our case study, hereby providing all kind of execution details. We also include some (visualization) results obtained. Then, in section VI, we evaluate our findings against the criteria introduced in section II. Finally, in section VII, we conclude and provide an outlook. For more details than are given in this paper, we refer to the technical report [7] which also contains many software code details.

II. CRITERIA FOR EVALUATING BI SOLUTIONS

Using BI software involves many issues. Based on the insights that (i) the software should support BI-based activities, (ii) users need to be able to understand the fundamental underlying data processing activities, and (iii) users wish to have user-friendly interface, we identified the following set of criteria for evaluating the BI platform at stake. The first criterion is that it should possess all *BI functionality* needed in practice. In order to avoid a discussion on what this precisely means, we confine ourselves here to what are usually considered to be the major components of BI [8]. These major components concern software to:

- extract, transform and load (ETL) data from all types of databases into (enterprise) data warehouses;
- fill meta data repositories with relevant meta data about data warehouses;
- create domain-specific data marts from data warehouses;
- transform data from warehouses and marts into data that can be used by data analysis' software;
- analyze data with all kinds of BI techniques including on-line analytical processing (OLAP), data mining (DM) and prediction;
- report and visualize results as obtained in various steps.

The second criterion concerns the level of *transparency* of the software. It should be easy for users to identify, preferably in terms of BI concepts,

- what the function is of each software module;
- how (i.e., by means of what algorithm) each module executes data;
- how the different modules can be combined in order to generate the decision supporting information desired.

Finally, the BI software should be *user-friendly*, i.e., it should be

- easy-to-use (i.e., intuitive);
- flexible-to-use (i.e., easy-to-use under changing demands);
- consistent in its use; (i.e., same functions in different interfaces are offered in the same way)
- easy to integrate with other software modules.

Below, we will use the above-given criteria to evaluate the Pentaho BI software suite based on the insights and experience obtained.

III. THE PENTAHO BI SUITE

In this section, we analyze what the Pentaho BI software suite consists of. To do so, we have analyzed information available on the WWW. The Pentaho cooperation provides a lot of information on their products on the WWW (see [6] and related web pages). According to this, the software provides 'a full spectrum of business intelligence (BI) capabilities including query and reporting, interactive analysis, dashboards, ETL/data integration, data mining, and a BI platform'. Looking in more detail, Pentaho has generated a total BI solution by combining all BI concepts under one umbrella, managed by the Pentaho BI platform. All components can be used both 'out-of-the-box' and as an integrated component.

Comparing the information given in this section to the criteria introduced in section II, we may conclude that according to the Pentaho cooperation the Pentaho software suite promises to be quite complete with respect to the functionality provided since the above-mentioned criteria known from literature are said to be covered. In addition, the user-friendliness is suggested to be present. In order to check these promises in reality, we performed a case study, the set-up and results of which are presented below.

IV. PREPARING THE CASE STUDY: CONTAINER TRANSPORT

Global logistics and door-to-door transportation of goods by means of containers is a highly competitive market. This observation also holds for goods of liquid nature including liquid chemicals [7]. Therefore, companies active in this business field need to be highly flexible and innovative to survive in this market. Business Intelligence might help here by first giving insight in the state-of-affairs of the business and, later on, by performing more sophisticated studies based on scenario analysis and/or prediction methods.

We happen to have access to the real-world data of a company having a big container transport division for liquid transports. For confidentiality reasons, we are neither allowed to reveal the name of the company nor give insight into detailed data and the precise results of the BI analysis. This implies that, although the case study results shown with respect to the practical use of the Pentaho BI suite are genuine, the results shown with respect to the (interpretation of the) underlying data are just fictive. For reasons of convenience, we use an abbreviation of the big container company at stake and simply denote it as BCC.

In the rest of this section, we present the further preparations for the execution of our case study.

A. Fixing the KPIs

It is a common insight that the selection of the right Key Performance Indicators (KPIs) is crucial for successfully applying BI [8]. In order to determine KPIs that matter, an interview with the marketing department and the management of BCC was organized. When discussing possible KPIs, we kept the following requirements in mind:

- 1) each KPI must reflect a genuine success factor;
- 2) each KPI must be quantifiable (i.e., measurable).

Actually, a third requirement is thought to be important as well, namely, that all KPIs together must keep track of the progress against a business goal. In our case, we dropped this third requirement since for the purpose of our study (getting some first experience with the software), we selected just one KPI.

The above-mentioned interview revealed a lot of 'secrets of the container business'. Several success factors (KPIs) related to changing market conditions, based on which the marketing department may immediately need to react, were brought out during the interview. This means that these KPIs are important for practical decision support. One of the KPIs identified was the *utilization level* of the (container) tanks. We decided to choose that one in our case study.

B. Defining the KPI's

A necessary condition for obtaining unambiguous results is deciding upon a clear definition of the chosen KPIs. In order to come up with a clear definition of the KPI 'utilization level', we first give a few basic definitions related to the three possible states of a container tank:

- a tank is *on reposition* if it is in the state of being repositioned to another area or location to balance supply and demand;
- a tank is *on move* if it is currently being used, i.e. in use and on its way to its destination;
- a tank is *in depot* if it is currently not active, waiting for an assignment in a depot.

Now, let r represent the number of tanks being repositioned, let m be the number of tanks being on move, and let d be the number of tanks being in depot, then the utilization level u is defined as:

$$u = \frac{r + m}{r + m + d}. \quad (1)$$

It should be clear that for BCC the utilization level is an important quantitative KPI that clarifies, in relative terms, the (non-)utilization degree of container tanks. Especially, if its value traced during a longer period of time, management gets insight in what can be considered as average, above-average, below-average, etcetera utilization values.

It should further also be clear that the utilization level u is a variable that can be analyzed at all kinds of granularity levels, both in time and in space. Right from the beginning it was clear that management might like to have the possibility to perform detailed analysis by, among others, zooming in into data of small granularity. For the time dimension this means access to aggregation levels like year, quarter, month, week, day and, if possible, even smaller. These information needs are very important to know right from the beginning and the details how we dealt with them will be further illuminated below.

C. Selecting the BI tools to be used

In order to be able to calculate the utilization level as defined by equation (1), we analyzed the available data sets. Based on this analysis and the above-mentioned typical information needs, we gradually identified the basic set of BI tools needed. We have used the so-called ‘community edition’. The advantage of this choice (above the pre-configured ‘enterprise edition’) is that you are forced to configure different software modules yourself. Although it is somewhat more work, it should be clear that this way of working provides better insight in what the software is doing for you, how the components are linked together and managed through the BI Platform.

Eventually, we have used the following tools from the Pentaho suite:

- For Extract, Transform and Load (ETL), we used ‘Kettle’. The main advantage of this tool is that it provides an intuitive, graphical, drag-and-drop-based interface. This makes it easier and understandable how data is extracted, transformed and loaded into the main database system. This does not only let us define what will be transformed, but also how the transformation is being executed. Many objects are predefined and ready to be used during the transformation process.

- For Business Analytics, we used the ‘Mondrian’ OLAP Engine. Mondrian is the OLAP engine of Pentaho, enabling interactive analysis of an SQL [9] database and provides an XML schema based cube definition environment and is fully MDX compatible, which eases the creation of multidimensional complex analytical queries. (For history of multi-dimensional expressions (MDX) and the implementation used, we refer to [10].) The ‘schema editor’ is a tool to build the Mondrian cube and is capable of generating XML [11] and publishing the cube definition to the Pentaho Platform. Furthermore, it provides an opportunity to test the cube with MDX queries. ‘JPivot’ is used to visualize the defined, published cube definition within the Pentaho environment. Drilling down, slicing and dicing, swapping axis, showing graphs and rearranging dimensions are just a few options provided.
- The Pentaho dashboard model has been used for visualization purposes. This model is in fact not a separate tool, but an environment where dashboards can be generated by using the available (or third party) components. It supports both PHP and Java. The ‘Pentaho Designer’ is a tool provided to define the so called ‘action sequences’, on which Pentaho is basically built on. An action sequence follows certain rules (such as selecting data from a data source, executing some operations, and feeding the result as output to a Pentaho component) and can be triggered from any dashboards and views at any time. Together with other tools (such as ‘Weka’ for Data-mining and ‘JFreeReport’ for Reporting) a meta-data editor is also available.
- As the database environment, we have chosen MySQL [12] as our main database platform on which we have created our data warehouse and the data marts.

Having illuminated the preparatory steps of the case study, we are now able to describe the stages of executing it, i.e., of applying the software tools selected in order to provide management sufficient insight in the utilization level of their tank containers.

V. EXECUTING THE CASE STUDY

The case study has been performed using three main stages, namely,

- analysis of available data sources;
- preparation of the BI building blocks;
- creation of the OLAP views and dashboards.

A. Analysis of available data sources

An inspection of the available data sources with respect to the tank containers of our case study revealed that six databases of legacy type existed giving all kinds of details related to their possible statuses including ‘on reposition’, ‘on move’, ‘in depot’, ‘under maintenance’, ‘reserved’, etc. This information also included time stamps and locations. Unfortunately, as in many legacy systems, the data fields were often not standardized, sometimes unreliable or even

inconsistent, and also not having a unique key identifier. Such a situation, of course, yields a lot of extra work when the data need to be aggregated in a data warehouse (DWH). In practice, we had a lot of discussions with the users of these databases to precisely understand the meaning of all data fields. In the end, based on what we learned from the analysis, we took several decisions related to the (non-)adoption of certain data and fields [7].

B. Preparation of the BI building blocks

During this phase, the building blocks of our BI environment have been prepared and implemented. Note that this stage is the most important stage of our implementation. In what follows, we describe the 8 implementation steps executed. Note that the order of steps is described using a re-engineering perspective where, in general, the desired outcome of each sub-step steers the precise implementation strategy of that sub-step.

1) *Star Schema design*: The DWH design is based on the concept of dimensional modeling, called the ‘star schema’ [9]. Different than usual, where first the DWH is created and afterwards the DM is derived from the DWH. We used a reverse approach to create the star schema first, map it to the DM and afterwards populate the design to the DWH.

Star-based schema modeling contains a central controlling ‘fact table’ [9] where all the business facts are gathered. It contains the necessary attributes, operational metrics and all other metrics to analyze the organization’s performance. The fact table addresses ‘what’ the DWH (will) support for decision analysis while the related ‘dimension tables’ [9] address ‘how’ data will be analyzed. The star schema we developed is as follows. The *fact table*

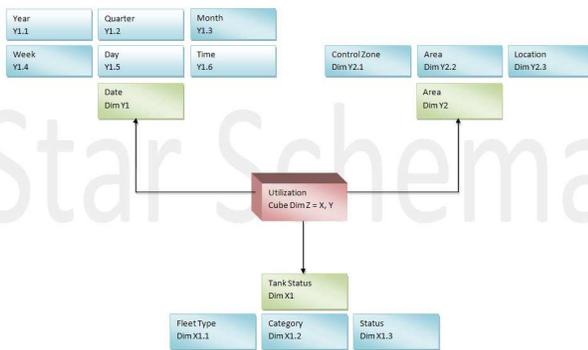


Fig. 1. Star schema of the PKI ‘utilization level’ with fact table *Utilization* and dimension tables *Date*, *Area* and *Tank Status*.

holds the following business facts:

- subtotal number of tanks for each date, area and tank status combination: this field is used to calculate the desired utilization levels;
- (foreign) tank key referring to which tank status the related subtotal belongs;
- (foreign) date key referring to the date the related measure applies;

- (surrogate, foreign) area key: since the area is not fixed, and can be changed over time, we introduced a so-called surrogate key to keep history information available; this covers the famous ‘slowly changing dimension problem’ [9].

The *date dimension table* holds the dates of fact table updates. Grouping analysis can be done on year, quarter, month, week, day and even the time stamp level. The *area dimension table* records are divided into a control zone (from where the related area is controlled and managed), location and its sub-region. This dimension controls the areas to which certain utilization levels refer to. The *tank status dimension table* holds the fleet type (which can be leased or owned tanks), the category it belongs to (whether it is ‘on reposition’, ‘on move’ or ‘in depot’), and the status of the related tank (such as its availability).

2) *Creation of OLAP schema*: We may analyze utilization levels within a certain area, in a certain period of time. We may discover during our analysis that a certain region contains a lower utilization than another region. Then we can ‘slice and dice’ [9] the cube according to a certain period and to certain sub-regions in order to investigate the underlying reasons while ‘drilling down’ [9] to a specific location. The software module Mondrian is the engine providing this OLAP functionality within Pentaho and is based on its own schema definition supporting MDX. We have created the Mondrian schema by using the star schema we previously created with ‘Schema Workbench’. After testing our Mondrian OLAP cube definition with a sample MDX statement, the schema has been published to Pentaho. For technical details, we once again refer to [7].

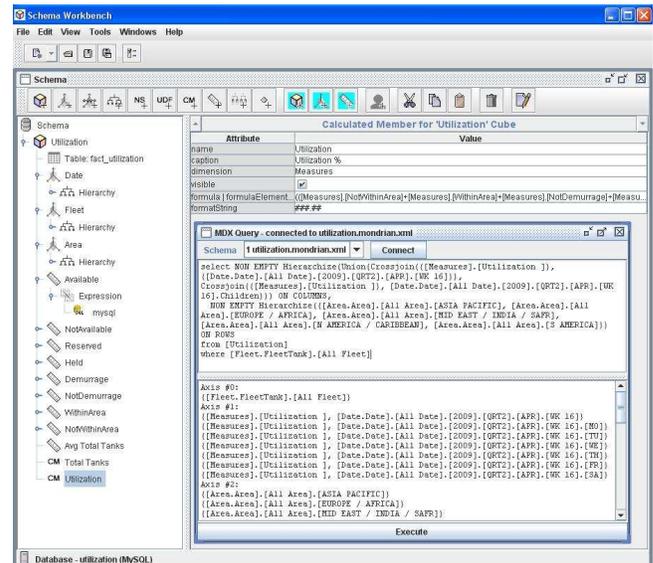


Fig. 2. Mondrian Cube Definition with Schema Workbench.

3) *DWH and DM design and implementation*: A DWH is a database optimized for analyzing the data in several ways, for different purposes. When the data are grouped into one

sub database for a specific purpose (such as mapping for a specific business process or KPI), it is called a data mart [8]. Several architectures exist. We have made a selection among the following types:

- 1) no DWH at all: all data are pulled and transformed directly from the systems;
- 2) centralized data ware house: all data are centralized into one database;
- 3) a DWH consisting of data marts.

Due to performance and modularity, we have chosen option 3. Additionally, we have set ourselves the following restrictions:

- 1) each datamart should reflect a single KPI (in our case: utilization);
- 2) each datamart should reflect a single business process (in our case: operations);
- 3) each datamart should be derivable from the DWH.

Further, the following items were considered, among others:

- all data must be consolidated first in the DWH and then de-normalized;
- indexes has to be chosen in such a way that maximum performance is gained while selecting the data needed to fill the datamart.

4) *ETL Process design, implementation and execution:* The ETL process is the first step in mapping the raw data into information. We have used the Kettle tool contained within the Pentaho suite to design our ETL process. Before the process is started, one stored procedure is called (through the ETL process) to prepare the tables for import. Our extraction, transformation and load process contains the following steps:

- 1) extract the data mapped to the information flow within each defined process: during this stage, the information is extracted by means of querying the legacy databases on the related tables; some details are shown in the screendump shown in figure 3:

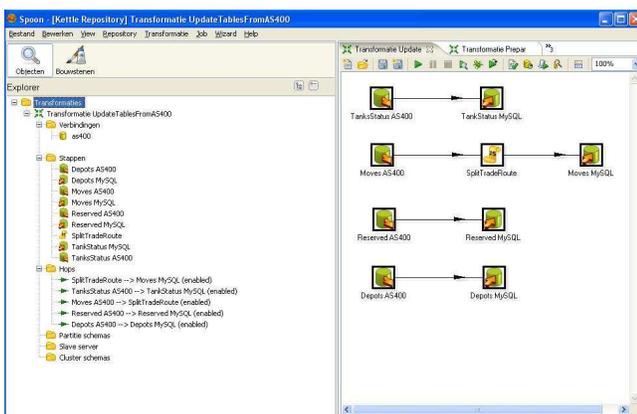


Fig. 3. Visualization of the etl processes performed where data are moved from the legacy databases towards the MySql database.

- 2) transform the data into information: during this stage, some fields are manipulated and re-joined to give the extracted data a meaning:

- first step was to split the reposition related fields into ‘within’ and ‘not within’ area to distinguish the repositions whether occurring within or outside the area; the output is passed to the next stage of the ETL process;
- as a next step, a procedure call is executed which prepares the tables for joining into the DWH;
- data cleansing operations are executed together with two very specific joins [7].

- 3) loading the transformed information into a data ware-house: the results of the two joins were inserted to the DWH;
- 4) extracting information from the DWH to feed the data-mart and the star schema: as a final step, a small procedure was added to the ETL process feeding the star schema with information:

- select the records from the DWH for the current time-stamp;
- sum Up all flagged fields per area and per fleet type, for all type of tanks;
- insert selectively to the dimension tables per fleet type and afterwards, linking them through the fact table;
- take the surrogate key on the area as the base link for the ‘fact table’.

5) *Verification of star schema data:* After the ETL process has been defined and executed according a test scheduling scheme, few days had to be passed while closely monitoring the whole process and system resources and while verifying that the DWH contains information we wanted to have. If performance issue would get important, redesigning the whole process would have been necessary. In our case, we only added and removed some keys within several temporary tables we created for the joining phase. The complete ETL process has been summarized in the following figure:

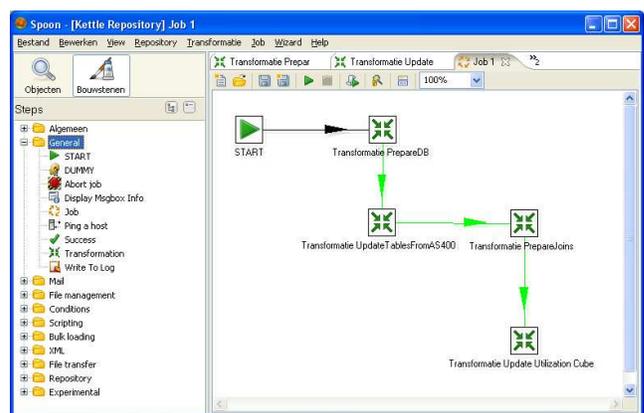


Fig. 4. Summary of the complete ETL-process.

- 6) *Re-scheduling of ETL:* The last and final step within the ETL process is to choose the schedules in such an away that it tactically supports the business. Since the BCC business

is global, we combined all opening and closing hours of all regions and selected the schedule in such a way that just before any business started in any region, the latest update was available and the star schema updated.

C. Creation of the OLAP views and dashboards

OLAP [9] is an approach to quickly provide answers to analytical queries which are multidimensional in nature. Generally, one or more dimensions contain aggregate data such as total or sum and in our case; the utilization formula. Pentaho provides per default an analysis view renderer called the 'JPivot' which uses the Mondrian OLAP engine. When the 'analysis view' on the main page is clicked, Pentaho allows user to select which Mondrian schema to use and shows a base view on the page for further analysis. In the example given in figure 5, we sliced the cube by filtering the data for the second quarter and swapped the axes for all fleet types to show utilization for each region based on the fleet type. The drill down option was enabled to show a drill down for available tanks within the selected slice.

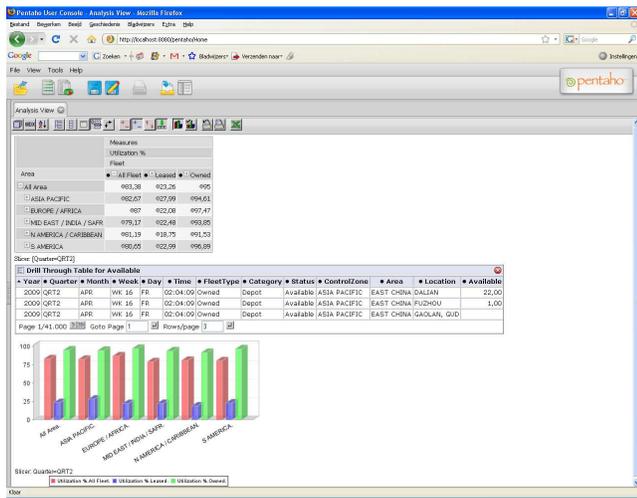


Fig. 5. Example visualization of the OLAP cube with JPivot.

D. Visualization with Google Maps

Generally, visualization contains the representation of the facts presented in a graphical format such that all the required information for decision making or about business performance can be viewed and interpreted at one glance to make a decision in a timely fashion.

During our implementation, a dashboard was implemented using Google maps' API [13] in order to show the utilization level per location on a map. This view allows to see the utilization of all locations by just looking to the colored pins on the map whether the related location falls below, in between or above the the defined thresholds. This enables management to react on any location (or region) in a timely fashion. The dashboard shown in figure 6 shows the utilization details managed by the South America region. At one glance, the



Fig. 6. Google map showing utilization details of business in South America.

following information is available for interpretation:

- utilization figures of a location are shown when clicking on the particular location 'pin' as info window, together with a dial chart; information shown is based on the selected utilization threshold values;
- tank status distribution is shown on a pie chart, as percentages;
- utilization details for the current week for the selected zone are shown as a bar chart, together with a line chart for utilization.

The 'pin' colors also change according to the selected threshold values. We added a small procedure call to the ETL process preparing the data for the dashboard [7] to speed up this process.

VI. EVALUATION

Having presented our case study were we used a set of tools of Pentaho, it is time to evaluate our experience against the criteria introduced in section II. As a first remark, it should be clear that using the Pentaho BI suite on a true business case is not a trivial exercise. For example, all kinds of technical and semantic details related to existing systems (databases) need to be understood and the KPI's need to be clearly identified before any step can be undertaken. In addition, unlike most of the commercial tools, use of this suite requires quite knowledge about technical BI concepts like DWHs, star schemes, and dashboards as well as about software 'languages' like SQL, PHP, Java and XML formats. Note that all kinds of specific details of the latter are not shown here (more details are available in [7]).

Looking at the criterion of *BI functionality* needed in practice, we first observe here that we only used a limited set of tools. However, based on our experience, we may conclude that the Pentaho suite has proven to support all aspects of business intelligence, starting from extraction, transforming and loading until building a DWH, creating and analyzing OLAP hypercubes till flexible visualization of the OLAP

cubes. In this case study, we did not use the ‘WEKA package’ for data mining and prediction (DM&P) but we happen to have quite some experience with that package in other cases. Again our experience (and that of many of our colleagues) is that, provided you understand the crucial concepts of data mining and prediction techniques (including the pitfalls), WEKA (and therefore Pentaho) is an excellent, easy-to-use tool with a lot of DM&P functionality.

Looking at the criterion of *transparency*, we conclude that the modular approach of Pentaho helps to keep an overview of the things you are doing and forces to think in genuine BI concepts. Since quite some technical and conceptual details related to the various steps need to be known and understood, one may conclude that the Pentaho approach supports transparency in the way we defined this notion above.

One may wonder why it would be more helpful if certain aspects need to be made invisible to the user by resolving problems automatically. For successful application of BI (what is supposed to end up into right information, to the right person, at the right time), we consider it advisable to limit this type of disclosing implementation details (as is usually done in closed software solutions). To quite some extent the user should be very aware about what (s)he is doing in order to avoid misinterpretations. Actually, the Pentaho suite seems to provide a nice balance between things that are hidden and things that are not hidden.

Looking at the criterion of *user-friendliness* it should be clear that the Pentaho suite cannot be said to be really user-friendly, simply since you need quite some software and technical skills to work successfully with it. If we compare Pentaho with commercial BI suits (such as Business Objects [14]), we think that more work should be done on the way how dashboards can be created. This should be done up to a certain extent, since too ‘user friendliness’ will restrict flexibility.

Although many wiki’s and separate documentations exist on the web, at the moment of writing this paper, there existed no proper single documentation explaining case-wise all the building blocks of Pentaho. All information had to be ‘gathered’ through different sources on the WWW.

In addition to this evaluation based on the section II criteria, we like to share the following general and more specific remarks based on the experience we obtained during our evaluation.

- BI is directly related to the ability of an organization to define clear KPI’s with a clear link to a business process. It enables organization to quantify its current performance and to set clear common goals, making performance measurable. In order to gain direct benefit from BI, organizations should be agile in redefining their business processes.
- BI requires commitment from higher management level, since many parties may have different conflicting interests and reasons not to cooperate.
- The Analysis phase is one of the most important initial steps to be executed before any table is created in the

DWH or DM. Understanding how data is recorded in the source application is imperative.

- ETL should not be seen as a simple export and import operation. ETL aims to load the warehouse with integrated and cleansed data from several sources after some operations are executed, preparing the data in such a way that it reflects the KPI or business process and contains reliable consistent information.
- Dumping the data to a DWH, by just joining different source tables as it is, makes no sense. It will generally not yield the information about the business management is interested in.
- Our way of working was based on thinking in terms of KPI and processes, mapping those aspects to a DM and designing the DWH bottom up and next implementing them top down. This also provided a modular approach which helped us in building further business-oriented cascaded KPI’s.
- For educational purposes (our secondary goal of research), for understanding the BI concepts, Pentaho would be a perfect candidate provided students are carefully guided through the BI step-by-step methodology.

VII. CONCLUSION

The need for having a comprehensive, fully integrated BI solution with complete BI functionality and the relatively recent availability of an open source BI solution termed Pentaho that promises to be such a solution, made us decide to investigate whether this platform is a truly good environment for use in business as well as educational contexts. To do so, we analyzed Pentaho’s software functionality and performed a case study.

Based on our experience we conclude that Pentaho is indeed a BI software suite having a full spectrum of BI functionality. Our business case has shown that, by using Pentaho, interesting management information can be extracted from a set of existing databases by using a step-by-step methodology where in each step different software modules are used. The integration of the software modules available is possible, although this requires quite some technical skills and knowledge related to software languages and implementation details. The presentation and dashboard capabilities of the Pentaho suite are, especially if you compare them to existing commercial, closed source BI solutions, rather flexible, but limited in the sense how dashboards can be created quickly.

For educational purposes, for understanding the BI concepts, Pentaho seems to be a perfect candidate. We therefore strongly advise to use this tool for those who want to get familiar with BI concepts, although in a classroom environment, quite some guidance of the students may be needed. In the near future, we will try to set up a lab environment based on the Pentaho BI software to be used for educational as well research purposes.

REFERENCES

- [1] D. Power, “A brief history of decision support systems,” *World Wide Web*, March 2007, <http://DSSResources.COM/history/dsshistory.html>.

- [2] Oracle, "Oracle Enterprise Performance Management and Business Intelligence," *World Wide Web*, visited May 13, 2009, available at http://www.oracle.com/solutions/business_intelligence/index.html.
- [3] SAP, "SAP NetWeaver Business Warehouse," *World Wide Web*, visited May 13, 2009, available at <http://www.sap.com/platform/netweaver/components/businesswarehouse/>.
- [4] Cognos, "Cognos Business Intelligence," *World Wide Web*, visited May 13, 2009, available at <http://www.cognos-bi.info/>.
- [5] SAP Business Objects, "Comprehensive BI and Performance Management for Midsize Companies," *World Wide Web*, visited May 13, 2009, available at <http://www.sap.com/solutions/sapbusinessobjects/sme/edgeseries/>.
- [6] Pentaho Cooperation, "About Pentaho," *World Wide Web*, 2008, available at <http://www.pentaho.com/about/>.
- [7] O. Tuncer, "Business Intelligence, Implementing BI Concepts with Pentaho," Delft University of Technology, Faculty of Technology, Policy and Management, Tech. Rep., 2009, MSc-course 'Fundamentals of Business Intelligence'.
- [8] E. Turban et al., *Decision Support and Business Intelligence Systems*, 8th ed. Pearson Prentice Hall, 2007.
- [9] T. Connolly and C. Begg, *Database Systems, A practical approach to Design, Implementation and Management*, 3rd ed. Addison Wesley, 2002.
- [10] Julian Hyde, "MDX Specification," *World Wide Web*, 2008, available at <http://mondrian.pentaho.org/documentation/mdx.php>.
- [11] W3C, "Extensible Markup Language (XML)," *World Wide Web*, 2009, available at <http://www.w3.org/XML/>.
- [12] Infobright Inc. and Sun Microsystems Inc., "A Guide to Open Source Data Warehousing for Communications Service Providers," *White Paper*, 2009, available at <http://www.mysql.com>.
- [13] Google, "Google Maps API," *World Wide Web*, 2009, <http://code.google.com/intl/nl/apis/maps/index.html>.
- [14] SAP, "SAP BusinessObjects Portfolio," *World Wide Web*, visited May 20, 2009, available at <http://www.sap.com/solutions/sapbusinessobjects/index.epx>.