Sofia University “St. Kliment Ohridski”

Faculty of Mathematics and Informatics

Business Intelligence

Pentaho BI Suite for eCommerce Analytical Dashboard

Knowledge Management Coursework

Stanislav Peshterliev, 61096

3 January 2011

Table of Content

1. Introduction

What is Business Intelligence (BI)?

Concepts and methods to improve business decision making by using fact-based support systems.

Howard Dresner 1993

Business Intelligence is the applications, technology and process for gathering, storing, analyzing, presenting, and providing access to data to help users make better decision.

Pentaho BI Suite Bootcamp 2010

BI applications gives users the tools to analyze vast amounts of complex data across all organization's levels. The main goal is to turn the mass of data, into knowledge that gives organization competitive advantage, improves performance and support better fact-based decision making. Such system have evolved from the decision support systems which began in the 1960s and developed throughout the mid-80s and came into focus in late 80s.

Different BI solution are intended for different users, most often for Business Analysts and Middle/Higher Management. Also it can be used as a tool for an agent to measure its individual performance compared to other agents and this way promote competitiveness which improves organization's performance.

BI systems analyze large volumes of historical business data like sales, revenue, marketing campaigns which is well structured explicit knowledge collected from different sources. They can dramatically improves organization's knowledge of its performance and how its action affects that performance. Some non-traditional systems are using unstructured and semi-structured data (email, video and etc.), which is a big part of the knowledge generated by organization, but they are not in widespread use.

The data to feed BI system is typically stored in centralized data warehouse optimized for Online Analytical Processing (OLAP). De facto standard for data warehouse design is Kimball's concept of Multidimensional modelling, where facts (e.g. sales, revenue) are explored in terms of dimensions (e.g. customer, time).

Major parts of Business Intelligence Applications are:

- Data Integration - Extract Transformation Load (ETL)
- Analysis - Data Mining, Predictions
- Reporting - Pixel Perfect, Dashboards, Interactive tools for drill, slice, dice, pivot

Main constrains for implementation and application of Business intelligence solutions are:

- Building a data warehouse and keeping it up to date may require large investment in hardware.
- Often data is so huge that a lot of caching is needed in order for application to be responsive which introduces additional implementation complexity.
- Most of the existing solution offered by big software vendors Oracle, Microsoft are quite expensive. Open Source alternatives exists as well.

2. Practical Application

The typical BI implementation goes through the following actions in its development. These are series of required steps that are consider best practice in any BI application.

1. Project Planning - high level analysis of business objectives, resources, skill-level staffing requirements, task assignment, and further steps that needs to be taken. The plan is dependent on business requirements. Depending on the requirements a team of 3-5 people should be need for period of 6-12 months.

2. Requirements Gathering - very important step that affects every almost every other decision. It is crucial to understand business requirements, needs and identify users. The most effective way is interviewing all stakeholders. Best practices are researching the organization, preparing questionnaires, engaging higher management, and recording interviews. This is the part where the knowledge from target users of the system is extracted and systematized.

3. Choosing the right BI platform - after the requirements are gathered and analyzed, BI platform should be chosen. There are several full featured BI platform with long history, offered by Oracle, Microsoft, SAP, and others. Open source alternatives, most famously Pentaho BI Suite, also exists. The decision whether to choose commercial or open source platform depends on various factors like money, level of customization, features usage, organization policy and others.

In my experience the open source Pentaho platform is not as mature and integrated as the tools offered by other commercial vendors but it is very flexible and customizable which can be ideal if you want to build a non-standard system where you won't need all the components offered by big vendors. My opinion is based on Pentaho course I have attended a month ago and my

recent work on Business Intelligent product based on the platform which is being built by the company I work for.

4. Data analysis - sources of data should be determined. They can be flat files, transactional databases, other data warehouses or third part systems. According to Kimbal[2] it is a good idea to do data profiling, this analysis will be able to describe the content, consistency and structure of the data.

5. Data integration - ETL is data integration process in data warehousing that involves:

- Extracting data from source database
- Transforming it to suit the target database
- Loading it into the data warehouse

ETL is necessary because it integrates data from heterogeneous sources, cleanses data removing transactional issues and introducing consistency, restructures data for optimal analytical processing. Before any ETL development is started target database should be designed.

6. Reporting - after data is integrated and stored in the data warehouse it have to be displayed to end users. Several options are in widespread use like print-perfect reports, dashboards, interactive tools for drill, slice, dice, pivot. Each can be used by different type of users, mangers, analysts, executives.

7. Deployment - after the above steps are executed application should be deployed on a client infrastructure. Hardware should be capable of processing required volumes of data in the given time span. Most of the tools process data daily or weekly, so ETL and caching should be executed for no more than 8-12 hours. Organization training should be given in order end users to be able efficiently to work with the solution.

3. Success criteria

According to Kimball these are three critical areas that you need to consider before starting BI project [2]:

1. Commitment and sponsorship of the project from senior management

This is very critical criteria for assessment because having strong management support will help overcome difficulties in the project. It is very important that the management personnel who participate in the project have a vision and an idea of the benefits and drawbacks of implementing a BI system. It is required that there is support from multiple members of the management so the project will not fail if one person leaves the steering group. All stakeholders in project should participate in this analysis in order for them to feel ownership of the project and to find common ground between them. If the Business sponsor is overly aggressive and demanding for new requirements, it is good idea to make sure that he is aware of his action because they will add extra months to the original plan.

2. Business-driven motivation

Implementation should be driven by clear business needs. Because of the close relationship with senior management, another critical thing that needs to be asesesed before the project is implemented is whether or not there actually is a business need and whether there is a clear business benefit by doing the implementation. The needs and benefits of the implementation are mostly driven by competition and the need to gain an advantage in the market. Another reason could be the acquisition of other organizations that enlarge the original organization and Business Intelligence solution can create more oversight.

3. The amount and quality of business data available.

This is may be the most important factor, since without good data – it does not really matter how good your management sponsorship or your business-driven motivation is. If you do not have the data, or the data does not have sufficient quality any BI implementation will fail. Before implementation it is a very good idea to do data profiling, this analysis will be able to describe the content, consistency and structure of the data. This should be done as early as possible in the process and if the analysis shows that your data is lacking;

In his thesis “Critical Success Factors of BI Implementation” [1] Naveen Vodapalli does research on different factors that can impact the final Business Intelligence product. He lists 7 crucial success factors for the implementation of a Business Intelligence project, they are as follows:

1. Business-driven methodology and project management
2. Clear vision and planning
3. Committed management support & sponsorship
4. Data management and quality
5. Mapping solutions to user requirements
6. Performance considerations of the BI system
7. Robust and expandable framework

4. Case study

Intelligent Trader

Currently I am working on a BI solution called eCommera Intelligent Trader[5]. The application is still work in progress, expected launch will be in April 2011. It is specially designed to be ecommerce analytics dashboard and decision support tools. The main target users of the application are Executives and Higher Management as one of the main goals is to create more oversight and support decision making. Other than Business Intelligence, eCommera’s[5] main business is in building large scale ecommerce websites for European retailers.

Intelligent Trader is build on the top of Pentaho platform. Data integration and OLAP components are used, while reporting applications are not in favor of custom Dashboard optimized for ecommerce analytics.

Key features of the implementation are[5]:

- Consolidate all your data from web analytics, merchandising systems and order management to provide a dashboard that makes sense of the vast amount of data.
- Identify the risks and opportunities through usable intuitive interface.
- Automatically identify and generate the key actions you need to take to optimise your performance

The company has experience offering analytical serveries which were performed manually by experts, using tools like Microsoft Excel. Now with Intelligent Trader we are working to extract expert's knowledge and integrate it the new product. Knowledge will be offered in forms of action the client should take in order to grow its business. There are predefined list of action and every action is linked to profit growth so you can see immediately which actions are worth taking, and which are not.

On the Dashboard homepage Key Performance Indicators (KPIs) are shown, the core of every trading action.

- Gross trading profit - gross profit less marketing, promotional and delivery costs.
- Lost profit - profit lost due to lack of product availability, declined or cancelled orders, or fraud.
- Return on inventory - measuring the effectiveness of driving profit from your stock investment.
- Customer satisfaction – A measure of overall customer experience including customer delivery to promise, customer service to promise.

User can choose to display additional Key Performance Indicators (KPIs) on the Dashboard homepage, but the above mentioned are not configurable and always shown. You can drill, slice, dice in every KPI to get more detailed information.

The implementation promise to be high quality and very well optimized for ecommerce business domain. I think that Intelligent Trader is a great solution for managing retail organization's explicit well structured knowledge.

5. References

Papers:

[1] Naveen K Vodapalli (2009-11-02). "Critical Success Factors of BI Implementation". IT University of Copenhagen.

Books:

[2] Ralph Kimball, Margy Ross, Warren Thornthwaite, The Data Warehouse Lifecycle Toolkit : Expert Methods for Designing, Developing, and Deploying Data Warehouses, Wiley, 1998

[3] Roland Bouman, Jos van Dongen, Pentaho Solutions: Business Intelligence and Data Warehousing with Pentaho and MySQL, Wiley, 2009

Web sites:

[4] http://www.pentaho.com accessed on 03.01.2011
[5] http://www.ecommera.com accessed on 03.01.2011