A Comparative Study: Business Intelligence Tools

Akshay Dinesh Badgujar[1*], Saurabh Shrikant Kadam[2], Manasi Mohan Zambare[3],
Shubham Raghavendra Kulkarni[4]

[1]*Student, Department of Mechanical Engineering, Imperial College of Engineering and Research, Pune, India*
[2,4]*Student, Department of Instrumentation Engineering, Padmabhooshan Vasantraoodada Patil Institute of Technology, Sangli, India*
[3]*Student, Department of Computer Engineering, Vishwakarma Institute of Information Technology, Pune, India*

Abstract: **Business Intelligence is a collection of technologies, processes, and architectures used to collect, store, and analyze the data produced by a company, BI helps in making a decision based in fact using historical data rather than gut feeling or assumptions. BI tools are used to make reports and do data analysis, summaries, dashboards, maps, graphs to provide users with detailed business information. Analysis tools affect the development and sustainable growth of a company, as costumer evaluation needs to develop in a competitive market is essential. With population development, processing large quantity of data has become increasingly difficult for companies. After certain period all the companies need a better and efficient way to manage data and improvise the decision making method. This paper compares two of the most used BI open source tools in the market: Jaspersoft and Pentaho. The purpose of the paper is comparative study of the most important BI tools in the market specifying the importance of the corresponding solutions.**

Keywords: **Business intelligence, Japersoft, Pentaho.**

1. Introduction

Business Intelligence (BI) has been compelling to IT executives for several years and this trend is expected to continue. While both the basic concept and also the term 'business intelligence' date back many decades. The emergence of the data warehouse as new infrastructure for reporting and analysis combined with OLAP and new of fact-based support decision support systems (DSS) leveraged interest in BI in the past decade, and what was initially considered another 'consulting fad', is now considered a potential source of competitive advantage.

Business Intelligence software transforms a company's customer profile database into a database that builds a database to create competitive and profitable market value. In addition, Business Intelligence is used to make a backup copy and to improve the business relevant data and to use it to analyze this data, to continuously improve the competitiveness of the organization.

Part of this analysis is to provide timely reports, for management's to make the decision based on factual information, so their decision making is based on concrete evidence. BI system gives enough data to use and evaluate the needs and desires of costumers and in addition it allows to: 1)

Design reports for departments or global areas in a company; 2) Build a database for costumers; 3) Create scenarios for decision-making; 4) Share information between areas or departments of a company; 5) Sandbox studies of multidimensional designs; 6) Extract, transform and process data; 7) Give a new approach to decision making; 8) Improve the quality of customer service. The benefits of systemization BI include the amalgamation of information from several sources. Creating user information management profiles, reducing dependence on program systems, reducing information retrieval time, enhances analysis, and improves access to specific business terms.
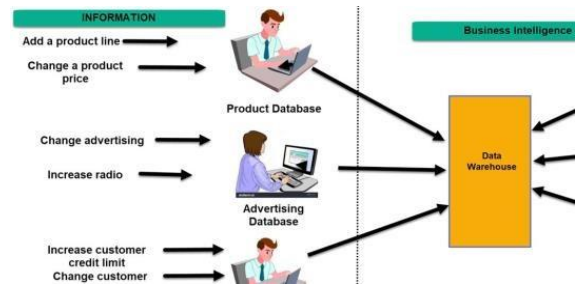

Fig. 1. Online Transaction Processing (OLTP) system

Big Data will aid to develop better procedures that allows (BI) tools to be used to gather information, such as: 1) Process and analyze volumes of information; 2) Increase the universe of data to consider when decision making and inherent historical data of the company, to incorporate data from external sources; 3) Provide an immediate response to the continued provision of real time data of the devices and the possibilities of interconnections between devices; 4) Working with structures of complex and heterogeneous data: logs, emails, conversations, locations, voice etc.; 5) To isolate from the physical constraints of storage and process by making use of scalable solutions and high availability at a competitive prices.

This paper presents an experimental analysis of the comparison of two of the best positioned open source BI systems in the market: Pentaho and Jaspersoft, processing big data and focusing on their Extract Transform and Load (ETL) and reporting processes. The aim of this paper is to analyze and

*Corresponding author: akshaydbadgujar@gmail.com

evaluate these tools and outline how they improve the quality of data, and inadvertently helps us understand the market conditions to make future predictions based on trends.

A. Pentaho

Pentaho is one of the most used Business Intelligence tool set (suite) across industries for management of the data. The suite is available in two editions- Community Edition (CE) and Enterprise Edition (EE). Analysts, data managers, software developers, and even students find the applicability of this tool. Companies like JP Morgan, Dell, TCS, Accenture, OLX, Bank of America to name a few have deployed Pentaho as an ETL tool.

Pentaho Business Intelligence tool suite is a set of tools that offers several benefits to businesses at an affordable cost and fast speed in terms of data management. Compared to other BI tools like SAS, SAP, BIA, Pentaho BI offers exceptional technical support to the costumer. It is highly scalable and offers large volume support to process data up to billion terabytes in size.

The scope of the Pentaho BI suite is vast supporting all kinds of data and data sources that furnish limitless visualization options. It supports an unlimited amount of data be it big data or existing data in the business IT. It can be used across different platforms that process hybrid data (text, graphics etc.) like cloud and mobile apps.

Pentaho is a Business Intelligence tool that offers many data solutions to its customers. The main feature of this tool are reporting, data integration, data mining, data analysis that account for the improvement of the business. A Pentaho suite enhances the overall performance of the business by generating informative reports in varied formats like text, XML, HTML, CSV, Excel, PDF etc. It is a multipurpose BI platform allowing enterprises to analyze, integrate and present data through comprehensive reports and dashboards.

With a self-exploratory designed interface, Pentaho has made valuable contributions in data analytics, providing business suggestions and presenting your data and information in interactive report formats.

The major reasons why enterprises are choosing Pentaho for their businesses are:
  a) *Controlled data delivery:* It merges trusted and timely data for powerful data analytics at scale for all users in all environments.
  b) *Easily embeddable:* Pentaho supports multi-tenant architecture, which allows embedding analytics into any workflow application like cloud, mobile and hybrid data models.
  c) *Power to integrate:* It accurately integrates and blends data from multiple sources, regardless of the deployment environments. Provides flexibility of analytics, turning big data into valuable insights.
  d) *Interactive and simple visual tools:* The visual drag and drop tools at Pentaho keep users away from the burdens of complex coding.

Table 1
The four layer's in Pentaho's architecture

Presentation Layer	It contains data available through analysis, reporting, process management etc.
Data layer	This is used to connect any database.
Server layer	It allows applications to run on top of it.
Client layer	Contains two clients.

Pentaho Features:
- Pentaho BI offers multiple features for the smooth workability of the business, such as:
- High end data analysis through well-defined ETL (Extract, Transform, Load) capabilities.
- Expertise in products across varied domains.
- Comprehensive report designer taking care of business needs.
- Additional reports or sub-reports along with the main detailed report.
- High scope for newer additions and updates.
- 24*7 technical support by the Pentaho Community.
- Unmatched reporting and query handling capabilities.
- Amplified functionality and efficient systems.
- Short integration TAT.
- Exceptional data source accommodability with high runtime metadata support.

Pentaho BI Suite:

The Pentaho BI suite is a three tier system that has different layers for exclusive functioning. It comprises of the following layers and components:

Tiers or Layers:
- Presentation Layer
- Business Intelligence Platform
- Data & Application Integration

Pentaho BI Components:

Pentaho BI suite includes the following components:

1) Reporting

The Pentaho BI reporting tool is used for generating reports on-demand and as per the fixed schedule set by the user. However, the reporting tool, however, works in association with the J free report project. The reports published by this tool are available in different formats like TXT, XLS, HTML, PDF etc.

2) Analysis

Another feature of this suite is an analysis of the extracted and transformed data which is now available in the form of reports. The analysis can be presented in multiple ways such as a Pivot table. The graphical user interface is well enhanced with projection tools like Flash, SVG etc. Other features include Workflow integration, portals and dashboard widgets that are integrated with the apps.

3) Dashboard

The dashboard serves as the front face of the suite that offers well reported content along with analysis and layout. The Pentaho suite also offers a self-service dashboard that has multiple layouts and templates to offer to its users. If the user willing to get some training, personalized dashboards can also be made.

4) Data Mining

Data Mining refers to extracting hidden patterns and future indicators from the available data that increases predictability of the future business and also accounts for forecasting. Data mining runs on the concept of machine learning which is backed by sophisticated algorithms that involve decision trees, networks, principal component analysis and clustering of data. This feature allows interaction with the data at the graphical and program level to enable future analysis.

5) Pentaho Data Integration

Pentaho data integration is a tool that allows and enables data integration across all levels. This tool possesses an abundance of resources in term of transformation libraries and mapping objects. This helps in data integration, big data analytics, data integration, and Hadoop data management.

B. Jaspersoft

Jaspersoft is an open source java based platform of BI developed by TIBCO. The product has been around for few years and is advanced in the field of embedded BI, visualization and data analysis. Jaspersoft product contains five different modules: Report Server, Studio, OLAP, Dashboard Designer and Analysis. Jaspersoft is used in more than 130,000 applications according to developers. This time, release 7 is the most current release. Jaspersoft is the company behind the famous and expanded Jaster reports. An open source reporting solution preferred by the most developers to be embedded in any Java application that requires a reporting system. Jaspersoft built its BI next to its reporting engine. This is done differently than Pentaho. Jasper put together its own projects and solved existing projects and put together however, it did not catch them. This strategy makes it "dependent" on Talend's solution regarding ETL and Mondrian – Pentaho with the OLAP engine. Jasper is able to access the Mondrian code which he can adapt to and continue his development with Mondrian. Jasper Report's main goal is to get the page redirected, ready to print documents in a sample and flexible way. The following flow chart shows the normal flow of activity during report.
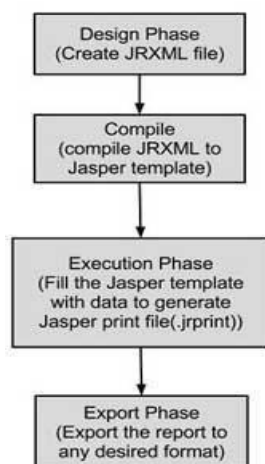

Fig. 2. Flow chart

As shown in the Fig. 2, the life cycle has following distinct phases,

- *Designing the report:* This step includes creating the JRXML file, which is an XML document that contains the definition of the report layout. We can use any text editor or iReportDesigner to manually create it. Real structure of the JRXML can be ignored if the iReportDesigner is used; the layout is designed in a visual way.
- *Compiling the report:* This step, includes compilation of JRXML is compiled in a binary object called a Jasper file (*.jasper). This compilation is done for performance reasons,
- *Executing the report:* Data from the application is entered in the compiled report. A Jasper print file is created, which can be used either to print or export the report.
- *Exporting the report:* Here we can create multiple representations of the data with the same input as Jasper provides various forms of exports.

2. Conclusion

This paper has tested two of the best positioned open source BI systems in the market: Pentaho and Jaspersoft. Both BI systems present notable features on their components. Pentaho on one side along with ETL component with great usability, maintainability and flexibility in making the transformations: Web Application with Java j2EE application 100% extensible, adaptable and configurable; the configuration management is integrated in most environments, that communicate with other applications via web services; it integrates all the information resources into a single operating platform; Reports with an intuitive tool that allows client to create reports easily; OLAP Mondrian with a consolidated engine widely used in environments of Java; Dashboard designer makes dashboards Ad-hoc, dashboards based on SQL queries or Metadata and a great freedom by offering a wide range of components and options. Jaspersoft on the other side has JasperETL (Talend) with Java / Perl native, Web Application with a Java j2EE application 100% extensible, adaptable and customizable; the management settings are very well resolved, it allows almost all through the same Web application; It integrates all information resources into a single operating platform; the editor Ad-hoc reports and Box Editor Ad-hoc command are best resolved; Reports are fast; Ad hoc and have a nice interface, with good flexibility and power, simple, However, this paper cannot tell all a business thinking of implementing BI needs to know but provides better understanding of some potential problem to be faced.

References

[1] L. T. Moss, S. Atre, Business Intelligence Roadmap – The Complete Project Lifecycle for Decision-Support Applications (Boston: Addison Wesley, 2003).
[2] N. H. Rasmussen, P. S. Goldy, P. O. Solli, Financial Business Intelligence – Trends, Technology, Software Selection and Implementation (NY: John Willey & Sons, Inc., 2002). www.intelligententerprise.com
[3] http://www.microsoft.com/presspass/press/2006/apr06/04-03ProClarityPR.mspx
[4] http://www.microsoft.com/mappoint/products/2006/default.mspx#1
[5] http://connect.microsoft.com/site/sitehome.aspx?SiteID=181
[6] http://www.oracle.com/pls/ebn/swf_viewer.load?p_shows_id=5801530
[7] http://www.oracle.com/appserver/business-intelligencece/index.html
[8] http://www.sox.com/News/detail.cf m?ArticleID=1035