



Enhancing Business Intelligence with Open Source Software Tools

Dr. Preet Kanwal*

Department of Computer Science, Sri Guru Gobind Singh College, Sector 26,
Chandigarh, India

Abstract– Automated data collection tools have led to storage of enormous amounts of data in data repositories. The organizations can benefit from this huge collection of data if and only if they are able to extract meaningful information for the benefit of their business. To survive in the modern competitive business world, the managers are in continuous need of right information at the right time in the right place to be able to make faster and better decisions. Business Intelligence (BI) plays an important role in enabling the business organizations to capture, extract, comprehend, utilize and harness multidimensional data to support decision making in order to improve business operation in complex and dynamic environment. A BI system includes capabilities for decision support, online analytical processing, prediction and data mining. The paper discusses the need for BI, the components of BI system, the benefits and challenges of adopting Open Source BI systems by organizations, and finally five Open Source BI tools are presented.

Keywords– Business Intelligence; Data Warehouse; ETL; OLAP; DSS; Open Source

I. INTRODUCTION

Automated data collection tools have led to storage of enormous amounts of data in data repositories. The organizations can benefit from this huge collection of data if and only if they are able to extract meaningful information for the benefit of their business. Meaningful Information and consequential knowledge form the backbone of the decision making process. To survive in the modern competitive business world, the managers are in continuous need of right information at the right time in the right place to be able to make quicker and better decisions. Business Intelligence (BI) plays an important role in enabling the business organizations to capture, extract, comprehend, utilize and harness multidimensional data to support decision making in order to improve business operations in the complex environment.

In this paper the importance of BI for organizations is explored. The remaining paper is divided into following sections: Section II – Background of BI, Section III - Need of Business Intelligence, Section IV - Components of BI, Section V - Benefits and Challenges in adopting Open Source BI systems, Section VI - the Open Source BI tools and finally Section VII - Conclusion.

II. BACKGROUND

Data is an asset for any organization as long as it can be used appropriately for the betterment of the organization. Historically, in ancient Egyptian times, trade and taxation were the big drivers for keeping manual detailed records though they were unable to provide on-demand reports as the manual system was tedious, time consuming and error-prone. In the digital age it has been possible for the businesses to organize their data in ways which can improve the efficiency of data storage as well as retrieval by way of using database management systems [1]. Database management systems have shifted from file based system to relational database systems. There is rise of alpha-numeric data including images and audio visual media in organizational data, thereby requiring a changeover from simple relational database comprising of tables to a system capable of storing graphical data also. To survive in such an increasingly tough business environment, it is crucial that cost-effective and rapid access to multidimensional business information is made available for business stake holders. Thus, a Business Intelligence system needs to be adopted which provides a set of technologies and products for generating the information needed to respond to business queries, and make tactical and strategic business decisions. Three generations of business information systems have been identified as: Host-Based Query and Reporting, Data Warehousing and Business Intelligence [2].

BI is an umbrella term including applications and technologies of capturing, accessing, and comprehending a large amount of data for the organization to make effective business decisions. The primary use of BI is to improve the timeliness and quality of information, and enable the managers to better understand and respond to changing trends by make better business decisions [3]. It combines the internal data of the organization with the external data derived from market analysis to gain intelligent insight into various business aspects like demand and suitability of the products or the services offered by the organization, new markets and impact of competition [4].

III. NEED OF BUSINESS INTELLIGENCE

The business organizations have already been using information technology for business transactions and its related operational decision making. The requirement of an enhancing from a simple operational decision making into intelligent decision making, the organizations need to have requisite BI system which can be integrated with their existing infrastructure and business processes. There are various driving forces which present a need of BI implementation in business organizations [2, 5]. They are listed as under:

- To gain understanding of all the forces which affect the business organization like customer behaviour, demographic and economic information, future trends as well as social, regulatory and political environments in which the organization operates.
- Managing huge amount of data collected by the organization through various channels and extracting meaningful information to generate intelligent knowledge for survival and growth of business organization. Data may come from various sources such as operational data, warehouse data, as well as related information managed by office systems and corporate web servers.
- To provide accurate and rapid analysis for decision making through sophisticated information discovery and analysis tools capable of handling and processing the complex business information.
- To attain the ultimate goal of the business i.e. to increase revenues, reduce costs, and compete more effectively in rapidly changing business environment.

IV. COMPONENTS OF BUSINESS INTELLIGENCE SYSTEM

BI systems take data from multiple sources and compile meaningful information from multiple formats into reports which are relevant and precise for the managers to act upon strategically and tactically. To achieve this goal, a BI system utilizes multiple components which act in cohesion. A BI system includes capabilities for decision support, online analytical processing, statistical analysis, forecasting, and data mining [5-9].

- **Data Sources:** Data Sources can be internal and external. Internal data sources include the organizations own operational data and historical data. External data include the data from other sources like market research companies, from the Internet or from public repositories and data warehouses. Different data sources may have different data structures, reside on different platforms and may be structured or unstructured.
- **Data Warehouse -** A Data Warehouse is a central decision support database created for making the data from all sources – internal as well as external, useful and accessible for decision-makers to extract the information they need for making critical business decisions. The data is systematically collected, organized and stored in a warehouse for integration, cleansing, aggregation, validation and query tasks.
- **Extract-Transform-Load (ETL) Tools –** ETL includes processes and tools used to extract data from external sources, transforming and pre-processing such data into a format compatible with the organization's Data Warehouse structures for loading.
- **Data Mart -** A Data Mart is a functional database that extracts specific information out of the overall data warehouse or directly from external regarding specific queries.
- **On-line Transaction Processing (OLTP) techniques -** Tools for reliable and efficient processing of a large number of transactions comprising of day-to-day operations and ensuring data consistency in traditional relational database.
- **On-line Analytical Processing (OLAP) technique -** These systems work with data warehouses to process queries required to discover trends, analyze critical factors and generates reports from large volumes of multidimensional data.
- **Dashboards -** Digital Dashboards allow organizations to capture and report specific data points from each department within the organization, thus providing a "snapshot" of performance [8].
- **Data Mining -** Tools specifically designed to identify patterns, relationships and rules within the data warehouse for discovery and insight into business knowledge and subsequent analytical prediction.
- **Decision Support Systems (DSS) –** These are software-based tools that first compile useful information from a combination of multiple sources such as raw data, documents, personal knowledge, or business models, and then use this compiled information to identify and solve problems and make better business decisions [9].

V. ADOPTING OPEN SOURCE IN BUSINESS INTELLIGENCE: BENEFITS AND CHALLENGES

Open Source Software (OSS) revolution has transformed the ways of use of information systems and services by individual users as well as business organizations. Nowadays OSS alternatives are available to cover the entire range of software for which commercial and proprietary alternatives exist. OSS systems for data warehousing as well as BI systems have also come up. Business organizations can benefit from the use of open source BI systems due to following reasons [10]:

- **Lower Cost -** Open Source BI systems are inexpensive from licensing point of view. Moreover, business organizations can use free versions of open source software for pilot projects thereby benefiting in terms of minimal acquisition cycles and associated entry costs.
- **Component Based design** allows them to fit into existing applications being used by the organizations.
- **Open Source BI systems** can provide better security features than their commercial equivalents as large community of volunteers addresses questions and provides solutions about installation instructions, updates, best practices, and feature requests in real time.

- Customization is possible with Open Source BI systems.
- Possibility of frequent updates if the adopted Open Source BI system is backed by a robust developer community which continue to provide support for feature improvement as well as customization in accordance with infrastructure of the organization.

However, there are certain challenges in using Open Source BI systems [10-11].

- Requirement of highly trained developer group – Since an Open Source BI system involves a community of developers who can provide customized system as per the specific needs of the organization, there is a requirement of developer group which has the skill set to develop the requisite BI system as per the requirements of the organization as well as who can integrate the business processes in the BI system.
- Since open source development is developer centric and not business centric, user interface may appear technical to the non technical employees.
- Lack of instant on-site support - As the open source development model depends more on community of developers who may be working in different time zones and scattered geographically, the organization may have lack of instant on-site customer support which may be available in case of commercial BI systems.
- There may be compatibility issues between hardware being used by the organization and the Open Source BI system.
- If the organization is using cloud computing implemented as Software-as-Service, then the access to source code may not be available even if the hosted software is open source.

VI. OPEN SOURCE BUSINESS INTELLIGENCE TOOLS

As discussed above, the Open Source BI systems have a range of benefits for organizations, especially small and mid-size organizations that can benefit from using them to reduce their costs. In this section, some available Open Source BI tools are discussed.

- **JasperReport Server-** JasperReports Server is a stand-alone and embeddable reporting server. It supports a Web based BI Platform that is used to generate, organize, secure, and deliver interactive reports and dashboards. It is available in two versions: i) as an open source community edition download and ii) a commercially licensed, enterprise-grade business intelligence solution. It provides reporting and analytics that can be embedded into: i) a web or mobile application ii) operate as a central information hub for the organization. It has the capability to deliver real-time information to the web browser, mobile device, printer, or email inbox in a variety of file formats. [12].
- **BIRT** - Business Intelligence and Reporting Tool is an open source software project developed by Eclipse Foundation. It can enhance the business intelligence of the organization through its components for designing and generating report and charts. The components included are : a visual report designer for creating BIRT Reports, a runtime for generating reports, a Chart Engine, Chart Designer, and Viewer. It integrates with any data source in any environment and is able to access information from multiple data sources easily and quickly in order to create reports and applications with stunning data visualizations [13].
- **Pentaho** - Pentaho Open Source BI Suite Community Edition (CE) suite includes ETL, OLAP analysis, data mining, reporting, dashboards. It provides a platform that allows creating complex solutions to business problems. By integration data with business analytics thereby enabling organization's stakeholders to easily access, visualize and explore the data that impacts business outcomes [14].
- **SpagoBI** - SpagoBI is a complete Open Source Business Intelligence suite. It covers all the analytical areas of Business Intelligence with innovative themes and engines. SpagoBI offers a wide range of analytical tools, as follows: Reporting, OLAP, Charts, KPIs, adhoc-reporting , location intelligence via maps, free enquiry, data mining, network analysis, ETL, office automation and manage external processes [15].
- **KNIME** - KNIME is written in Java and based on Eclipse. This is an open source multi-language software development environment comprising an integrated development environment (IDE) and an extensible plug-in system. It has its origin in the pharmaceutical research. It has a robust and modular design having highly scalable platform comprising of data loading, data transformation, analysis and visual exploration models [16].

VII. CONCLUSIONS

In times of globalization and seamless businesses processes, increasing standards, automation, and technologies have led to enormous amounts of data becoming available. There is a need for techniques and tools that can assist the organizations in converting this data into useful information and knowledge which can enable intelligent decision making for providing a competitive edge. Knowledge management and deploying of Business Intelligence by organizations for surviving in complex and dynamic business environment is of utmost importance in present times. Open Source BI systems provide an opportunity to organizations who want to integrate BI systems incrementally in their business processes to cut down costs while having the opportunity for making intelligent decisions.

REFERENCES

- [1] P. Lake and P. Crowther, "History of Databases," *Concise Guide to Databases*, Undergraduate Topics in Computer Science Part - I, Springer-Verlag London 2013, pp 21-40.

- [2] M. S. Almeida, et al, "Getting Started with Data Warehouse and Business Intelligence," IBM Corporation, International Technical Support Organization, 1999. pp. 1-8. Available: <https://www.redbooks.ibm.com/redbooks/pdfs/sg245415.pdf>
- [3] N. A. Rostami, "Integration of Business Intelligence and Knowledge Management – A literature review," *Journal of Intelligence Studies in Business*, vol 4, no 2, 2014, pp. 30-40. Available: <https://ojs.hh.se/>
- [4] "Business Intelligence". [Online]. Available: https://en.wikipedia.org/wiki/Business_intelligence
- [5] J. Ranjan, "Business Intelligence: Concepts, Components, Techniques and Benefits", *Journal of Theoretical and Applied Information Technology*, vol 9. no. 1, 2009, pp 60 – 70. Available: <http://www.jatit.org/volumes/research-papers/Vol9No1/9Vol9No1.pdf>
- [6] J. Lloyd, "Identifying Key Components of Business Intelligence Systems and Their Role in Managerial Decision making," Master's Thesis, University of Oregon, 2011. Available: <https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/11389/Lloyd-2011.pdf?sequence=1>
- [7] K. Withee, *Microsoft Business Intelligence for Dummies*, Wiley Publishing, Inc, 2010, pp. 9-21. Available: <http://citeseerx.ist.psu.edu/>
- [8] "Dashboards (business)". [Online] [https://en.wikipedia.org/wiki/Dashboard_\(business\)](https://en.wikipedia.org/wiki/Dashboard_(business))
- [9] "Decision Support Systems". [Online]. https://en.wikipedia.org/wiki/Decision_support_system.
- [10] J. Williams, "Open Source BI: Does it Fit into Your BI Environment?", *What works in Enterprise Business Intelligence*, vol. 24, pp 41. Available: <http://download.101com.com/tdwi/ww24/whatworks24digitaledition.pdf>
- [11] P. Rubens, "7 Reasons Not to Use Open Source Software," [Online]. Available: <http://www.cio.com/article/2378859/open-source-tools/7-reasons-not-to-use-open-source-software.html>.
- [12] "JasperReport Server". [Online]. Available: <https://sourceforge.net/projects/jasperserver/?source=directory>
- [13] "BIRT". [Online]. Available: <https://sourceforge.net/projects/opensourcebirtreportdesigner/?source=directory-featured>.
- [14] "Pentaho". [Online]. Available: <http://wiki.pentaho.com/display/COM/Community+Wiki+Home>
- [15] "SpagoBI". [Online]. Available: <https://www.spagobi.org/homepage/product/business-intelligence/>
- [16] "KNIME Open Source Story". [Online]. Available: <https://www.knime.org/knime-open-source-story>