

# THE INTERNATIONAL JOURNAL OF BUSINESS & MANAGEMENT

## Business Intelligence: An Open Source Approach

**Rafi Ahmad Khan**

The Business School, University of Kashmir, Srinagar, India

### **Abstract:**

*Business Intelligence (BI) systems permits managers and other business users to monitor their organizational information, analyze market trends, introduction of new products, retain the existing and attract new customers and make effective decisions in order to have competitive edge in a global market. Therefore, the significance of using BI tools is vital for the success of organizations. Business Intelligence (BI) systems utilize operational and historical data with analytical tools to present valuable information to business planners and decision makers for effective decision making which results in attaining competitive advantage. One of the biggest problems with traditional proprietary BI solutions is the huge investment needed before the organization achieves any value. This has become one of the main drivers for the growth and popularity of the open source BI solutions. Another reason which attracts many business organizations to open source BI software is their flexibility as it allows an organization's developers to embed BI tools into the application being developed. That flexibility to open source BI software allows not just IT workers or developers but a variety of users, to use its analysis and reporting tools. The best open source platforms use internet applications to increase the level of interactivity in the report creation process to be used by all types of business users. In this paper, an attempt has been made to sensitize the researchers, academicians and practitioners about the BI, its framework and various open source BI tools that are available.*

**Key words:** Business Intelligence, Data Warehouse, Data Mart, Open Source, OLAP, ETL, Scorecard, dashboard

### **1. Introduction**

The factors of globalization, mergers and acquisitions, competition and technological innovation are the primary factors that have forced organizations to reorganize their business strategies. With the result many organizations have resorted to Business Intelligence (BI) techniques to help them understand and control business processes to gain competitive advantage. BI is mainly used to increase the timeliness and quality of information so that support managers can better understand the position of their firm in comparison to their competitors. BI applications and technologies help organizations to analyze company capabilities, customer preferences, changes in customer behavior & their spending patterns, changing trends in market share and market conditions.

BI is an area of Decision Support System (DSS) that can be used to support complex decision making and solving complex, semi-structured, or ill-structured problems [1],[2],[3]. The first reference to BI was made [4], which has now replaced the other terms such as Executive Information Systems and Management Information Systems [5],[6],[7].

Stackowiak et al. view BI as the process of taking large amounts of data, analyzing that data, and presenting a high-level set of reports that condense the essence of that data into the basis of business actions, enabling management to make fundamental daily business decisions [8].

Zeng et al. contend that BI is "The process of collection, treatment and diffusion of information that has an objective, the reduction of uncertainty in the making of all strategic decisions" [9]. "BI is a business management term used to describe applications and technologies which are used to gather, provide access to analyzed data and information about an enterprise, in order to help them make better informed business decisions" [10]. Cui et al. opines that BI is the way and method of improving business performance by providing powerful assistance to executive decision maker which enables them to have actionable information at hand [11]. BI tools are viewed as technology that enhances the efficiency of business operation by providing an increased value to the enterprise information and hence the way this information is utilized. BI tools have evolved from being an Executive Information Systems (EIS) and Decision Support System (DSS) to provide much more flair in information delivery and ability to support techniques such as query, reporting, ad hoc analysis and multidimensional analysis which are also known as Online Analytical Processing (OLAP). This diversity in capability attracted organizations to start investing in building these types of intelligence systems as it offers enterprises "one version of truth", providing consistent and harmonized data to every department in an organization [12].

### **2. BI Framework**

BI technology has been continuously growing and improving to support managers in taking better decisions. The most widely applied BI technologies, that have emerged include data warehousing (DW), on-line analytical processing (OLAP), and data mining (DM).

The general framework of BI for understanding and guidance of practitioners, academicians and researchers is presented in fig.1. In organizations data is first entered or processed by a daily business process that is based on Online Transaction Processing (OLTP) systems and stored in operational database. This forms the internal source of data. Organizations also use data from external sources that include the data from suppliers, government agencies, customers, competitors, internet etc. Since the internal as well as external data is from heterogeneous sources, therefore, it needs to be processed and converted into homogeneous form. Extract, Transform and Load (ETL) is one of the software utility that performs this task as follows.

- **Extraction & Cleanse:** During data extraction data is acquired from multiple internal as well as external sources. Data cleansing cleans up and validates the extracted data to correct missing, inconsistent, or invalid values. This stage initiates error reports and applies corrective processes.
- **Transform:** It integrates data into standard formats and then applies business rules that map data to the warehouse schema. Aggregates (e.g., summary table data) and assigned characteristics are generated.
- **Load:** It loads the cleansed data into the data warehouse.

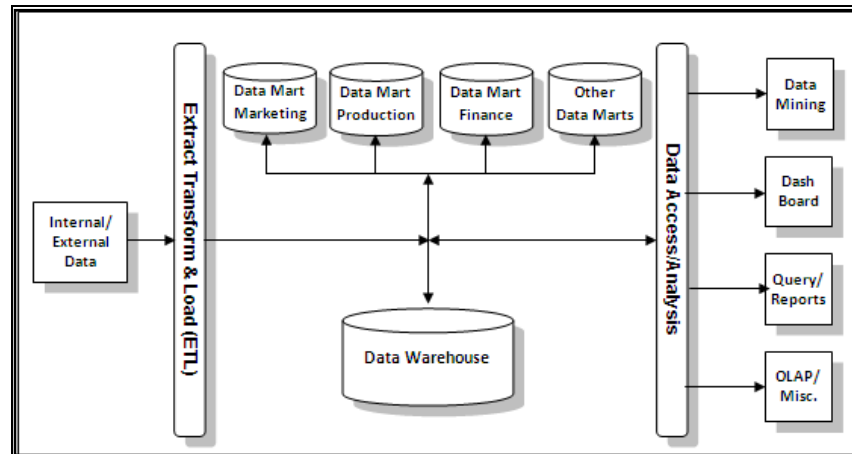


Figure 1: Framework of BI

After ETL, data is stored in data warehouse or data marts for future analysis.

Data warehouse is a subject oriented, integrated, time-variant, non-volatile collection of data that is used primarily in organizational decision making [13]. It stores copy of transaction data specifically structured for query and analysis and is informational, analytic and decision support oriented, not operational or transaction processing oriented [14]. Corey & Abbey, perceive data warehouse as a collection of corporate information derived directly from operational systems and some external data sources. Its specific purpose is to support business decisions, not business operations[15].

Data marts are small sized data warehouses or localized data warehouses, usually created by individual departments to provide their own decision support activities. Due to some constraints like risk of failure, huge investment, some organizations invest in data marts for a few functional areas like marketing or sales instead of a full-fledged data warehouse. While some organizations choose both data warehouse as well as specialized data marts as they considerably reduce the query complexity and thus results in significant reduction in the query response time[16].

To understand and locate data in the data warehouse/data marts users need information about the data warehousing system and its content. This information known as metadata which is data about data. It includes format, domain constraints, definitions of the data, and encoding/decoding algorithms.

BI technologies such Online Analytical Processing (OLAP), Data Mining, Machine Learning(ML) & other query reporting applications are used to extract hidden valuable information/knowledge from data marts or data warehouse. These tools and techniques provide a business user with direct, interactive, or batch access to data, while hiding the technical complexity of data retrieval. A variety of these tools and techniques are typically used in an integrated manner to serve the needs of different groups of users.

- **OLAP:** The best known knowledge discovery techniques are Online Analytical Processing (OLAP) and data mining (DM) techniques [17]. OLAP provide users with the means to explore and analyze large amounts of data, involving complex computations, their relationships, and visually present results in different perspectives. OLAP tools are a combination of analytical processing procedures and graphical user interface. The key features of an OLAP application are: multidimensional views of data, calculation intensive capabilities and time intelligence [18]. OLAP applications provides quick and flexible access to data and information through multidimensional view of data. Typical applications performed on multidimensional data views are: roll-up (data is summarized with increasing generalization), drill-down (increasing levels of detail are revealed), slice and dice (performing projection operations on the dimensions), and pivoting (cross tabulation is performed) [19].
- **Data Mining:** Data mining, as it is also known, is the nontrivial extraction of implicit, previously unknown, and potentially useful information from data. This encompasses a number of different technical approaches, such as clustering, data summarization, learning classification rules, finding dependency networks, analyzing changes, and detecting

anomalies. (William et. at, 1994). Data mining is the search for relationships and global patterns that exist in large databases but are hidden among the vast amount of data, such as a relationship between patient data and their medical diagnosis [20].

Data mining is generally recognized to be a single phase in a larger life cycle known as Knowledge Discovery in Databases (KDD)[21]. The term KDD refers to the broad process of finding knowledge in data stores. The field of KDD is particularly focused on the activities leading up to the actual data analysis and including the evaluation and deployment of results[22].

- **Dashboards:** Dashboards appear into a user interface which puts together and shows information in an easy and intuitive manner. Practically the desktop is organized so that it can offer necessary information, in the most proper way and to assure interactivity with the manager user[16].
- **Balanced Scorecards:** Balanced Scorecards help the users to put the strategy into practice. They are used to measure performances, derived from the objectives and strategy of organization, which reflects the most important aspects of the business[23].

The key reason that BI has become so successful is because of the development of scorecards and dashboards. These tools take huge volumes of data and deliver the information in a usual format which makes it easy to respond to time sensitive and critical events. Other tools include online analysis tools, reporting tools, , Web content management, Document management, groupware, knowledge portal, e-learning and work flow management that support dissemination and sharing of information and knowledge in a collaborative environment.

- **Machine Learning (ML):** ML is part of an emerging Artificial Intelligence (AI) technology that over the past two decades has been employed by an increasing number of disciplines to automate complex decision making and problem solving tasks. Among the various available methods of ML, Artificial Neural Network (ANN) is the most popular which has been inspired by the biological neural networks of the human brain and started as an attempt to model the learning capabilities of humans. Other techniques include genetic algorithms, NLP, inductive learning, case-based reasoning, etc.

### 3. Open Source BI Softwares

Broadly, open source refers to a software in which the source code is available free of cost to the general public for use as well as modification from its original design. It is usually created as a collaborative work in which programmers create, redistribute and modify the source code, driving a convenient evolution of the software product. Open source developed as a response from technological community to branded BI software offered by companies. The idea of open source relies on peer review to find and eradicate bugs in the software code, a method which commercially developed software do not use. Since open source software is community driven as the source code is shared throughout the community, the process of elimination of bugs and improving the software occurs at a much faster speed than through the traditional development of commercial BI software.

Open Source BI solutions are enterprise-wide, cross-technology based, business-centric, cross-customer viable and cross-functional. Open source BI software tools provide facilities ranging from data modeling to reporting to OLAP to ETL. Various BI open source softwares available are as follows:

- **Jasper Intelligence:** The Jasper Intelligence product line will include a server for a component for doing analysis, ETL and generating. JasperReports is a powerful open source Java reporting tool that has the facility to provide rich content onto the screen, to the printer or into HTML, PDF, CSV, XLS and XML files. It provides reporting, dashboards, data analysis, and data integration facilities[24].
- **Eclipse BIRT:** Eclipse BIRT is the most widely used open source BI suite. It runs on the Eclipse IDE platform and offers several reporting and data visualization tools. This provides the foundation for report design and viewing. The BIRT engine is at the heart of this offering, and this is a collection of Java classes and APIs which execute BIRT reports and generate them in appropriate formats[25]. Actuate founded and leads the BIRT (Business Intelligence and Reporting Tools) open source project within the Eclipse Foundation, and BIRT is at the center of its applications[26].
- **Pentaho:** The Pentaho BI Project is an ongoing initiative by the Open Source community to provide organizations with best-in-class solutions for their enterprise Business Intelligence (BI) needs[24]. Pentaho was able to grow considerably faster than commercial vendors by means of Open Source technologies and the assistance of the Open Source development community. Therefore, it has been able to outshine all commercial offerings in terms of functions, features, and benefits. Talend Open Studio: Provided as a packaged, out-of-the-box, ready-to-install platform, Talend Open Studio is one of the most promising open source ETL tool[24].
- **DeepGreen:** It is the first open source, complete database server dedicated entirely on supporting BI applications. It offers a feature-rich with robust platform is freely available for download. . DeepGreen is ideal for entry-level workloads such as data marts and reporting applications in the 10-300 GB range. The product offers Easy-To-Install Packaging, Cross Platform Support, Bulk Data Loader, and Open Connectivity to other Business Intelligence Tools[27].
- **BizGres:** Bizgres is the open source, production ready database server focused exclusively on supporting Business Intelligence applications. Bizgres targets entry-level and departmental workloads such as data marts and reporting applications in the 10-300 Gigabyte range.
- **SpagoBI:** SpagoBI is an open source Business Intelligence platform which uses many tools such as analytical engines, OLAP, Data Mining, Reporting, and Dashboard. SpagoBI is a complete suite for the development of Business Intelligence projects in an integrated environment[24].

- **Palo:** Palo is an open source business intelligence suite focused around OLAP and Excel and web interfaces. The Palo OLAP Server is at the heart of the offering and provides multi-user, high performance access to data. It supports real-time aggregation and is an ideal platform for BI collaboration. Data is loaded into the OLAP server using the Palo ETL Serve[25].

#### 4. Advantages of Open Source BI Tools

- Easy to initiate: With open source, BI project can be initiated easily and inexpensively with free download of the BI software.
- Lower cost: With small initial investment and lower ongoing maintenance/support cost, the cost for open source BI software is very little or negligible as compared to traditional BI software.
- Easy to customize: since users can access and modify the source code directly, it is possible for developers customize the open source BI tool and add their own features. While, it is very difficult to do this with traditional BI software because there is no way to access the source code.

#### 5. Disadvantages of Open Source BI Tools

- Features are not as robust: BI software vendors put in a lot of resources and money into R&D, as a result the product has a rich feature set. While, open source BI tools, bank on community support, and hence lack a strong feature set.
- Support not as readily available: Most of the commercial BI software - Business Objects, MicroStrategy, Cognos, Oracle etc., have been around for a long time. As a result, there are a lot of individuals with skill with those tools, therefore, it is easy to get consultation help to implement these solutions. Since open source BI tools are a recent development, therefore, relatively few people have their implementation experience, hence finding consultation is difficult.

#### 6. Conclusion

In today's highly competitive world, no business organization can deny the benefit of BI. Latest industry analysis reports indicate that in the future millions of individuals will use BI tools and analytics daily. This rapidly changing business environment will increase the need for BI. However, the huge cost of investment on proprietary BI software owned by corporations are out of the reach of most organizations. In order to accelerate business benefits by adopting BI without huge investment, organizations take advantage of freely available open source software, which allows organizations to rapidly deliver the business solutions. The open source BI allows the organizations to solve discrete business problems one after another, build knowledge and deliver value along the way by quick teams that too with little or no investment.

#### 7. References

1. Azevedo and M. F. Santos, "Business Intelligence: State of the Art, Trends, and Open Issues," in First International Conference on Knowledge Management and Information Sharing-KMIS 2009, 2009.
2. H. R. Nemati, D. M. Steiger, L. S. Iyer and R. T. Herschel, "Knowledge Warehouse: An Architectural Integration of Knowledge Management," Decision Support, Artificial Intelligence and Data Warehousing. Decision Support Systems, vol. 33, no. 1, p. 143–16, 2002.
3. J. P. Shim, M. Warkentin, J. F. Courtney, D. J. Power, R. Sharda and C. Carlsson, "Past, Present, and Future of Decision Support Technology," Decision Support Systems, vol. 32, no. 1, p. 111–126, 2002.
4. H. P. Luhn, "A Business Intelligence System," IBM Journal of Research and Development, vol. 2, no. 4, p. 314–319, 1958.
5. S. Negash, "Business Intelligence," Communications of the Association for Information Systems, vol. 13, no. 1, p. 177–195, 2004.
6. E. Turban, R. Sharda, J. E. Arosen and D. King, Business Intelligence: A Managerial Approach, Upper Sadle River, New Jersey: Pearson Prentice Hall, 2008.
7. E. Thomsen, "BI's Promised Land," Intelligent Enterprise, vol. 6, no. 4, pp. 21-25, 2003.
8. R. Stackowiak, J. Rayman and R. Greenwald, Oracle Data Warehousing and Business Intelligence Solutions, Indianapolis: Wiley Publishing, 2007.
9. L. Zeng, L. Xu, Z. Shi, M. Wang and W. Wu, "Techniques, process, and enterprise solutions of business intelligence," in IEEE Conference on Systems, Man, and cybernetics, Taipei, Taiwan, 2006.
10. L. Wu, G. Barash and C. Bartolini, "A Service-oriented Architecture for Business Intelligence," in Service-Oriented Computing and Applications SOCA '07, IEEE International Conference, 279-285, 2007.
11. Z. Cui, E. Damiani and M. Leida, "Benefits of Ontologies in Real Time Data Access," in Digital Ecosystems and Technologies Conference, DEST '07. pp. 392-397, 2007.
12. P. Bochner and J. Vaughan, "BI today: One version of the truth," Application Development Trends, vol. 11, no. 9, pp. 18-24, 2004.
13. W. H. Inmon, Building the Data Warehouse, New York: John Wiley & Sons, 1996.
14. R. Kimball, The Data Warehouse Toolkit : Practical Techniques for Building Dimensional Data Warehouses, John Willy & Sons, 1996.
15. M. J. Corey and M. Abbey, Oracle Data Warehousing, Osborne McGraw-Hill, 1997.

16. R. A. Khan and S. M. Quadri, "Dovetailing of Business Intelligence and Knowledge Management: An Integrative Framework," *Information and Knowledge Management*, vol. 2, no. 4, p. 2, 2012.
17. S. Forsman, "OLAP Council White Paper," 1997. [Online]. Available: [www.olapcouncil.org/research/whtpapply.htm](http://www.olapcouncil.org/research/whtpapply.htm).
18. M. Jarke, M. Lenzerini, Y. Vassiliou and P. Vassiliadis, *Fundamentals of Data Warehouses*, Springer Verlag, 2000.
19. M. Holsheimer and A. P. J. M. Siebes, *Data Mining: The Search for Knowledge in Databases*, The Netherlands ., Amsterdam: CWI (Centre for Mathematics and Computer Science), 1994.
20. E. Turban, *Decision Support System and Business Intelligence*, Pearson Education, 2012.
21. K. Collier, B. Carey, E. Grusy, C. Marjaniemi and D. Sautter, *A Perspective on Data Mining*, Northern Arizona University: The Center for Data Insight, 1998.
22. R. S. Kaplan and D. P. Norton, "The Balanced Scorecard: Measures that Drive Performance,," *Harvard Business Review*, pp. 71-79, 1992.
23. "Leading Open Source Business Intelligence software," 6 March 2014. [Online]. Available: [http://www.opensource-it.com/open\\_source\\_business\\_intelligence\\_software](http://www.opensource-it.com/open_source_business_intelligence_software).
24. "5 Free Open Source BI Suites," 6 march 2014. [Online]. Available: <http://butleranalytics.com/5-free-open-source-bi/>.
25. "Open Source Business Intelligence Software On the Rise," 4 March 2014. [Online]. Available: <http://ostatic.com/blog/open-source-business-intelligence-software-on-the-rise>.
26. 3 March 2014. [Online]. Available: <http://www.opensource-it.com/software/greenplum>.
27. J. F. William, G. P and C. J. Matheus, "Knowledge Discovery in Databases: An overview," *AI Magazine*, vol. 13, no. 3, 1992.
28. "Data Warehousing - Open Source Business," 15 March 2014. [Online]. Available: <http://www.1keydata.com/datawarehousing/open-source-business-intelligence.html>.
29. "open source," [Online]. Available: [http://www.webopedia.com/TERM/O/open\\_source.html](http://www.webopedia.com/TERM/O/open_source.html).
30. "Open Source Business Intelligence," Jaspersoft, 2012.