INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH & DEVELOPMENT

ISSN 2278 – 0211 (Online)

Big Data-Compelling Organizations beyond Data Warehousing

Mrunal Joshi
Assistant Professor, IES Management College and Research Centre (IESMCRC), Mumbai, India
Pradip K. Biswas
Associate Professor, National Institute of Industrial Engineering (NITIE), Mumbai, India

Abstract*:*
Data refers to the raw facts scattered across the organization. It serves the basis for generating information which leads to organizational decisions. The data is consolidated and stored in a data warehouse across the organizations over a period of time for trend analysis, comparisons and forecasting, allowing the enterprise data warehouse, enjoy the benefit of a strategic system. However, in past few years internet clicks are generating the volume of data which is precious for innovation. With comprehensive performance data available, companies can identify and focus on the high return improvement opportunities in business leading the improved performance in revenue. Storage and analysis of this data is imposing a challenge in traditional data warehousing technique. This paper is an attempt to examine the importance of data warehousing as well Big Data for an organization. The paper also attempts to explore what can be a way beyond data warehousing for effective management of Big Data.

Keywords*: Data Warehousing (DW), Big Data, Logical Data Warehousing (LDW), Organizational Performance, Emerging Trends in DW*

1. Introduction

Data refers to the fact figures. Organizations are getting larger and amassing with ever increasing amounts of data. Historic data encodes useful information about working of an organization. However, data is scattered across multiple sources, in multiple formats. The process of consolidating data in a centralized location is possible through data warehousing. Google's chief executive, Eric Schmidt, believes the world creates 5 Exabytes of data every two days. That's roughly the same amount created between the dawn of civilization and 2003. Today the volume of data generated is increasing exponentially whereas the cost of data storage is falling.

2. Data Warehousing

The data warehouse is the central repository where data from various systems and sources is collected. It is time variant since the data is collected over many time periods [7] for use in comparisons, trend analysis and forecasting [4]. The development methodology of data warehouse in different cases has been reported in various literatures [9],[11],[13]. When the organizational structure is high with a huge database, it is advisable to create appropriate number of data marts for independent functional divisions and then integrate all of them to get the total data warehouse. According to Craig (1997), data marts store subsets of information about product sales and other topics and may speed up the process of getting critical decision-support data to end users. But the proliferation of data marts can be a hassle for the warehousing staff that has to keep everything in synchronization. Seeking a way out of that trap, some companies are building virtual data marts that share a database but are presented to users as separate entities. Virtual data warehouses are best suited for applications that are of limited scope or duration [10].
Information from a customer database can be used to identify needs of different groups of customers. This knowledge can help shopping centers to improve marketing communications and customer satisfaction [3].
Data Warehousing along with the concepts of Knowledge Management and Data Mining helps in strategizing customer relationship management (CRM) in order to support the organization's decision-making process to retain long term and profitable relationships with its customers [2]. The design of the CRM data warehouse model directly impacts an organization's ability to readily perform analyses that are specific to CRM. Subsequently, the design of the CRM data warehouse model contributes to the success or failure of CRM. In fact, statistics indicate that between 50% and 80% of CRM initiatives fail due to inappropriate or incomplete CRM processes and poor selection of technologies [14],[16].

With increasing competition within the industry, information required by customer centric organizations need to be much more accurate and time efficient so that it can support the strategic decision-making process and improve CRM. Kincazid (2003) has viewed CRM as the strategic use of information, process, technology, and people to manage the customer's relationship with the company (e.g., Marketing, Sales, Services, and Support) across the whole customer life cycle.

With the increasing volume of data, traditional data processing methods will no longer provide sufficient support. The CRM database can be looked upon as the data warehouse (DW). It is a new concept as a result of technology improvement which collects data from several dispersed sources and builds a central control. Users can then use appropriate data analyzing tools to store and analyze needed data (Hwang, Ku, Yen & Cheng, 2004). Cunningham, Song and Chen (2006) have indicated that "by utilizing a DW, companies can make decisions about customer-specific strategies such as customer profiling, customer segmentation, and cross-selling analysis."

Across a wide range of business sectors, companies see the freshness of their business data as a critical factor for generating revenue and maintaining competitiveness. Enterprise data from core transactional processing systems should be continuously available wherever it is needed. However, making real-time information available across the enterprise has been perceived as costly and technically difficult. The study by Hackathorn challenges the belief that low-latency data is too expensive to support analytics and reporting for the business. Hackathorn explains, "Knowing about your business sooner, rather than later", provides a competitive edge no matter what industry you pick [8].

3. Organizational Importance of Data Warehousing

Data warehouse integrates data across time and across operational areas so users can easily obtain facts about the organization and business. Decisions involving firm's assets need to be based on an analysis of the fact-figures. Decisions based on good analysis are more valuable to a dynamic business environment. The data warehouse should be built for users not be left solely to the Information System (IS) community. The user interface and reporting tools must be aligned with the business strategies to get correct answers to questions. A good data warehouse can summarize all transactions it contains into time series data for monitoring and analyzing performance. It can help a firm to understand its customers better than those customers understand themselves. The flexibility in accessing the information can empower mangers with ad-hoc reporting capabilities to summarize data without the intervention of the IS people. Implementation of high end Enterprise Resource Planning (ERP), Supply Chain Management (SCM), Customer Relationship Management (CRM) systems demand normalization and standardization of data which is coming from different systems. If good corporate data warehouses have been built, the implementation of these detailed systems is faster and smoother. To reap the maximum benefits, corporate should treat warehouse as a part of value-based information infrastructure. Value based information is the data that support the High Impact Processes (HIPs) [6]. It supports the data warehouse's technical architecture, which must be flexible, yet stable enough to support the ever-changing customer demands and other business requirements. Flexibility in technical architecture leads to warehouse's longevity. The technical architecture should also be cost-effective, adaptable, and easily implemented. Since the early 1990s, the data warehouse has become the foundation of advanced decision support applications (Shim et al., 2002). Using sophisticated online analytical processing (OLAP) and data mining tools, some corporations are able to exploit insights gained from their data warehouse to significantly increase sales (Cooper et al., 2000; Heun, 2000; Whiting, 1999), reduce costs (Watson and Haley, 1998; Whiting, 1999), and offer new and better products or services (Cooper et al., 2000; Levinson, 2000; Watson and Haley, 1998). The payoff from a well-managed data warehouse can be huge. For instance, a study conducted by IDC, a leading research firm, found the average return on investments in data warehousing projects to be about 400 percent (Desai, 1999). However, the implementation of a data warehouse is both very expensive and highly risky. Building and marinating a data warehouse routinely cost a corporation millions of dollars (Gagnon, 1999; Jukic, 2006) whereas success seems to be the exception rather than the rule (Dagan, 2007). Kelly, 1997 reported that one-half to two-thirds of initial data warehousing efforts fail. Voelker, 2001, placed the failure rate at 60 to 90 percent. Whereas according to Beal, 2005; Madsen, 2005; Watson, 2005, it is not unusual to hear the failure rate of 50 to 75 percent. Data warehouses are growing in size, scope, and complexity as they serve increasing numbers of demanding users. They require a sizeable commitment of organizational resources.

4. What is Big Data?

Big data is generated from an increasing plurality of sources, including Internet clicks, mobile transactions, user-generated content, and social media as well as purposefully generated content through sensor networks or business transactions such as sales queries and purchase transactions. In addition, genomics, health care, engineering, operations management, the industrial Internet, and finance all add to big data pervasiveness. For understanding the trends and patterns within these large socioeconomic datasets we need powerful computational techniques. New insights gathered from such data-value extraction can meaningfully complement official statistics, surveys, and archival data sources which are mainly static in nature, adding depth and insight from collective experiences—and doing so in real time, thereby narrowing both information and time gaps (George, Haas, & Pentland, 2014).

The "bigness" of big data attracts researchers' attention to the size of the dataset, but there is an emergent discussion that "big" is no longer the defining parameter, but, rather, how "smart" insights the data can provide is important. One could analyze the social networks and social engagement behaviors of individuals by mapping mobility patterns onto physical layouts of workspaces using sensors, or the frequency of meeting room usage using remote sensors that track entry and exit patterns. This could provide information on communication and coordination needs based on project complexity and approaching deadlines. These micro data provide a richness of individual behaviors and actions that have not yet been fully tapped in management research. Whether it is "big" or "smart" data, the use of large-scale data to predict human behavior is gaining currency in business and government policy practice, as well as in scientific domains where the physical and social sciences converge (recently referred to as “social physics”) (Pentland, 2014). Purposefully assembled and organized, Big Data is a source of enormous organizational power. Nunan & Domenico, 2013 have explained the concept of big data by three perspectives. The first is a response to the technology problems associated with storing, securing and analyzing the ever-increasing volumes of data being gathered by organizations. This includes a range of technical innovations, such as new types of database and ‘cloud’ storage that enable forms of analysis that would not have been cost effective previously. The second perspective focuses on the commercial value that can be added to organizations through generating more effective insights from this data. This has emerged through a combination of better technology and greater willingness by consumers to share personal information through web services. The third perspective considers the wider societal impacts of big data, particularly the implications for individual privacy, and the effect on regulation and guidelines for ethical commercial use of this data [15].

5. Importance of Big Data

Big Data takes data from a range of systems and combine it into meaningful, actionable insight. Retail stores can use lagging indicators with current and predictive data to give a more pertinent view of how the product is performing and the contribution it is making to revenues. With an ability to ingest data from structured and unstructured sources Big Data solutions can combine standard performance indicators with activity and operational data such as customer satisfaction measures, real-time sales data, competitor advertising, blogs, website comments, campaign management reports, customer surveys and a range of other product and brand data to deliver a clear message on performance of each product in each market over time. With comprehensive performance data available, companies can identify and focus on the high return improvement opportunities in product development to improve performance in revenue.

For a customer focused business like banking Big Data solution can integrate structured customer data with unstructured reports and social media and deliver a real time, adaptable campaign management solution. It can also be useful in developing the strategies to maximize share pricing by identification and prioritization of global opportunities and threats, diverse stakeholders, and counter-balancing goals. The need of predictive analytics, and collaborative decision making around opportunities like countries, industries, and major clients, competitors and shifting macro-economic conditions can be addressed by proper usage of Big Data.

It can also be useful for senior decision makers and their supporting strategic management community in government agencies to improve their ability to prioritize what they do to deliver on their required outputs and mitigate enterprise risks.

Thorsten Engel, Wellington-based consulting partner and leader of Deloitte New Zealand’s enterprise information management team believes that Big Data challenges managers to be more fact based and savvy about the decisions they make. It can be a critical tool for realizing the improvements in yield, particularly in any manufacturing environment in which process complexity, process variability, and capacity restraints are present. Companies that successfully build up their capabilities in conducting quantitative assessments can set themselves far apart from competitors [1].

6. Moving towards Logical Data Warehousing

A traditional Data Warehouse (DW) is a physical implementation of a data consolidation process. Gartner reports that 90 percent of all DW implementations deployed in the world is to build some form of a repository. It always had a primary goal of being data consolidation engine or service. For many years the storage access rates, processing capacity and software used to deploy the warehouse, relied upon a physical deployment to meet many of the performance and consistency requirements in the data warehouse.

Today with the availability of new hardware and processing solutions Data Warehouse designers and architects are required to decide the changes required in the traditional approach. When pervasive use of any representations of consolidated information is expected above and over, any consideration, a repository is must. Latency tolerance allows for data to experience any delay in arrival, and with proper management of that latency interval the information still meets the use case requirements. In this case the information quality standards are agreed and rationalization of data requires data quality and enrichment to be infused into latent, persisted, pervasive data use case.

Data architects and designers consider the data virtualization when representation of consolidated information is not agreed and ongoing experimentation and discovery of how information assets relate to each other is under way. Data virtualization is suitable when transient consolidations that experience a temporary end-use-need are required and information quality and determination of anomalies can be delegated to the sources or can simply be resolved at runtime. Whereas when data collection does not need to follow a strict governance or data quality expectations but requires pattern analysis to infer the relationships, distributed processing option is considered by the data architects and designers.

This also infers that the traditional data warehouse is an enterprise repository implementation. However its role will be determined by the architectural preferences of each organization and should include evaluation of the use cases. A single view of data without moving it to some other place has given birth to a concept of Virtual or Logical Data Warehouse (LDW). The enterprise DW can be accessed beneath a semantic manager that directs processes to the repository or call out to distributed processes or even run the data virtualization task required. It can jointly control semantic management with a semantic layer, sharing auditing metrics and coordinating which tasks will run in each portion of LDW. The DW can become a processing platform to assist in tasks to support distributed processes and data virtualization tasks-especially in the case of scoring data records for use in analytics. It can become a place where patterns discovered using distributed data processing are stored and used along with other traditionally-collected data elements [13].

7. Conclusion

Over the years Data Warehousing has become a part of organization's long term strategic investment. It has emerged as an integral part of the corporate infrastructure. This technology was developed in response to management's need to make effective use of data resources. Today the term "Big data" which refers to data being generated across multiple sources, including internet clicks, mobile transactions, user-generated content, and social media as well as purposefully generated content through sensor networks or business transactions such as sales queries and purchase transactions is imposing a challenge before organizations to build the solution which will provide agility and support the storage and analysis of big data. For addressing the issues the traditional data warehousing solution needs to incorporate few changes like virtualization and distributed processing. Building a logical data warehouse can serve the purpose of maintaining the essence of data warehouse and serve the purpose of Big Data requirements.

8. References

1. Auschitzky, E., Hammer, M., & Rajagopaul, A. (2014, July). How big data can improve manufacturing. McKinsey & Company .
2. Cunningham, C., Song, I.-Y., & Chen, P. P. (2004). Data Warehouse Design to Support Customer Relationship Management Analyses. ACM .
3. Dennis, C., Marsland, D., & Cockett, T. (2001). Data Mining for Shopping Centres – Customer Knowledge-Management Framework. Journal of Knowledge Management , 368-374.
4. Dukart, J. R. (1998, June 01). Data warehousing. Inform , 12 (6), p. 48.
5. George, G., Haas, M., & Pentland, A. (2014). BIG DATA AND MANAGEMENT. Academy of Management Journal , 57 (2), 321–326.
6. Given, B. (1997). Master the art of storage and retrieval in a data warehouse. Transportation & Distribution;ABI/INFORM Complete , 68-70.
7. Gray, & Watson, J. (1998). Present and Future Directions in Data Warehousing. . Database for Advances in Information Systems , 83-90.
8. Hackathorn, R. (2007, Mar 29). http://search.proquest.com/docview/445070930?accountid=49672. Retrieved Dec 18, 2012
9. Inmon, W. H., Imhoff, C., & Sousa, R. (1998). Corporate information factory. New York: John Wiley & Sons.
10. Jigeesh, N., & Kondabolu, S. S. (2011). Creating a Virtual Data Warehouse for Manufacturing Industry. IUP .
11. Kimball, R. (1996). The data warehouse toolkit. New York: John Wiley & Sons.
12. (2013). Logical Data Warehousing for Big Data. Australia: Thoughtweb.
13. Mohanty, S. (2007). Data Warehousing: Design, Development and Best Practices. South Asian Journal of Management , 144-146.
14. Myron, D., & Ganeshram, R. (2002, July). The Truth About CRM Success & Failure. CRM Magazine .
15. Nunan, D., & Domenico, M. D. (2013). Market research and the ethics of big data. International Journal of Market Research , 55 (4).
16. Panker, J. (2002, June 20). Are Reports of CRM Failure Greatly Exaggerated? SearchCRM.com