DATA WAREHOUSE – A KEY TO MANAGEMENT DECISIONS

Meenakshi Nair

Asst Prof, St Xavier's College, Jaipur

Abstract

Internet technology with its world-wide presence has made day-to-day communication and interaction very easy. Extended availability, affordability and wide outreach of the Internet has enabled small, medium and large businesses to use it as one of the key sources for marketing their products and business. Internet has facilitated them to reach around the world with nominal cost. This has given rise to an all-time high competition.

Internet has helped to capture and store large amount of raw data. Now, there is the requirement for organizing the raw data in a meaningful format so that it can be used as a guide for business. The technique and process used in ordering raw data into more meaningful information is termed Business Intelligence (BI). BI requires large amount of raw data to be stored properly and efficiently. This has been made possible through data warehouse.

Introduction to Business Data

Business operations are the key sources from where large amount of data is collected and this huge data require better tracking and managing. Companies these days use a wide range of software programs such as Excel, Oracle, DB2 and other database applications in various departments throughout their organization. When multiple software programs are used in one business organization, it becomes very difficult to maintain and manage them. Timely retrieval of data and information also becomes very difficult.

With the advancement of technology and changing trends in the business sector, latest tools and techniques are being used. These tools are termed as Business Intelligence (BI). They enable companies to gather, store, access and analyze business data and subsequently help in decision making. Business executives who take strategic and tactical decisions cannot take decision merely based on financial statements available to them. They need answers to queries such as: "Which product is still in competition?" "What new venture could be made?" "What is the current market swing?" "What long-term effect could follow a particular decision?" etc.

These are some critical questions which need accurate answer. Here, BI system plays a prominent role in providing answers to these key questions. BI collects massive data from operations system and stores them in a format which is latest and understandable. The management of these massive data of BI system is handled through data warehouse.

What Is Data Warehouse?

The process of transforming data into information and making it available to the end user in a timely manner is known as data warehousing. The concept of data warehousing dates back to the late 1980s. A data warehouse is a repository of an organization's electronically stored data. It is designed to facilitate reporting and analysis (Inmon).

Different people have different definitions for a data warehouse. The most popular one came from Bill Inmon, who provided the following definition: "A data warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process" ("Data Warehouse").

Ralph Kimball provided a more concise definition of a data warehouse: "A data warehouse is a copy of transaction data specifically structured for query and analysis" ("Data Warehouse").

Data, Information and Knowledge

In order to discuss the point, first we need to understand the difference between data, information and knowledge. These three keywords are the most commonly used terms in several IT-related (and other) communities. But it is very difficult to differentiate the three on representational basis, because representation is always based on signs (e.g. ASCII – characters, bits, or handwriting). But let us try to differentiate them.

To do this, let us first note that unless and until these signs do not include additional information, we cannot interpret them. These signs can be transformed to data, information and knowledge when we associate them with respective thought or view. For example, when we say "a motorcycle," this represents any two wheeler fitted with an engine. It can be treated as data but without a concrete thought or view. But when we say a Bajaj motor cycle, we have branded it, adding specificity and concreteness. Here we are referring to a real object (motorcycle) of Bajaj Company. But even here there is lack of specificity, for the customers are unable to decide which model of motorcycle from Bajaj Company. But if we say Pulsar Motor Cycle of Bajaj Company or Splendor of Hero Honda, we are more specific. This adds a valid meaning and it becomes knowledge.

Now, signs can be represented in a system which could be referred to as data-based system, information system and knowledge-based system or any other system. But all of them are similar within the computer system, i.e. all of them are sign/data. When these signs/data are associated with their respective context, users or customers, they become information and knowledge. Thus we can say that data warehouse is nothing but a data storage system which contains sings/data.

How Business Growth Can Be Increased by Data Warehousing?

The data present in data warehouse represents a major part of a company's knowledge, e.g. the company's clients and their demographic attributes. This data store will turn to be a good knowledge base if it can be transformed into information and made available to the end user in a timely manner. The objective of data warehouse is to facilitate reporting and analysis. Technically, it is referred to as Online Analytical Processing (OLAP) or Knowledge Discovery in Database (KDD).

Data analysis has resulted in the formation of components for data warehouse. These components have the ability to retrieve and analyze data, and to extract, transform and load (ETL) the data. Discussion of data warehouse always refers to these broader contexts.

There are many ways to integrate data and applications to provide important and timely information to the end user. Each business challenge and/or process should be analyzed to understand whether a data warehouse or any other type of information access tool presents the best solution. The following are the key reasons that determine whether the data warehouse is the best solution as compared to other business solutions (Guerra).

1. The reports and information written in any BI application software needs modification and upgradation based on their new releases and updates. Field names are often hard to interpret and some are just meaningless strings of characters. Format of data in any BI application is not well represented, especially dates and numbers. Retrieval and analysis of data in tables are usually very difficult because they are designed to facilitate data entry and field validations which are accordingly implemented. Another application of BI tool is that it does not provide any well-defined process to incorporate valid data from other sources. There is no clear, distinct way to maintain metadata for the data being stored. Transaction processing has the lead over reporting and analysis function. This results in poor performance in doing data analysis. Data is not secure in BI applications; they could get modified or mishandled by the users. This results in wrong data analysis.

2. Data warehouse has provided different ways for doing data analysis. OLAP and KDD provide facility to analysis data in the format that is best suited and required by the management. OLAP is one of the technologies that is being widely used by data analysts for generating reports. OLAP provides different reporting and visualization features, by the use of which data can be aggregated and presented from different angles. Besides this, graphical pictures could also be used to represent the report. A data analyst using OLAP can create wonders through data warehouse by using formulated queries and the output of queries to present the data to management. And this report could become a valid and authenticated source for further decision making and action. Here if we recall the definitions of data, information and knowledge, we identify a similar scheme. Data is stored in the data warehouse. The data analyst enquires parts of the data, which are then interpreted and represented in the format that is understandable to human users. The methods used to enquire data needs some knowledge, and if the yielded information helps the management in taking some useful decisions or actions, that, in turn benefits business or environment and then this information becomes knowledge.

The examples below will give a clear picture of how a business benefits from data warehouse:

First, let's consider the successful launch of Tata Nano by Mr. Ratan Tata. Popularly, Tata Nano is called a common man's car. The thought process that went through the mind of the great industrialist Mr. Ratan Tata when he asked his team to get a car on the road for just Rs 1 lakh, and the way his team responded to his vision tell an enlightening story. According to a featured story available on the Internet, "Suggestions came in like, a door-less car with a bar as a safety measure, having soft doors in vinyl with plastic windows, a cloth roof, two big doors (instead of four). But all these were turned down by Mr Tata; he was very clear that it had to be a complete car" (Agarwal).

A four-year-long research was carried out with full enthusiasm and zeal. The result was the big wonder, i.e. Tata Nano. During the development phase, it was decided that the major part of the work needs to be outsourced. So the company formed a vendor group comprising many Tata group companies "including TACO IPD, Tata Toyo Radiators, Tata Johnson Controls, Tata Visteon, Tata Yazaki, Tata Ficosa and Tata AutoComp GY Batteries. There's also Tata Ryerson for the steel service centre and roll form sections, Tata Bearings for bearings and Tata Steel Tubes for the engine cradle."

The designing and production of Nano is a persuasive example of efficient use of the latest technology. Nano would have remained a dream if the production team would not have had efficient and effective reporting mechanisms. The success of Tata Nano is actually the product of intelligent, effective and efficient usage of data, information and knowledge.

The second example is the way Vodafone wins customers by using BI approach. Vodafone India has a large customer base. It is the India's second largest telecom service provider with more than 14 crore subscribers across its voice and data services. The heterogeneous system, decentralized business process and high data volumes had made information analysis a highly complex task at Vodafone India. Prior to BI, the company faced data latency issues in its data warehouse, affecting timely information reporting. The volume of data amounted to more than 11 lakh files per day. Gaining valuable insights from such a large data pool was a formidable challenge.

To meet the challenge, Vodafone implemented a knowledge-based system from IBM. It includes the following:

- A multi-layered BI architecture InfoVision implemented by IBM
- Teradata v12 as Enterprise Data Warehouse
- IBM Data Stage as ETL
- IBM Tivoli Workload Scheduler as Monitoring Control
- IBM Cognos as OLAP and Reporting
- Oracle 11gR2 Database as Federated Data Marts ("Vodafon Wins").

This system brought many benefits. The benefits delivered by BI were:

- With InfoVision, Vodafone India has created a scalable architecture for BI which can integrate new functions and services such as 3G and ICR.
- IT governance, policy, processes and guidelines implemented in the data integration layer provides audit transparency into current business reporting and analytical channels related to CDR (call detail record) information needs.
- With BI, Vodafone has eliminated prepaid CDR information latency issues in data warehouse to enable timely reporting for business users.
- With the implementation of data mart and ETL along with BI, Vodafone India estimates to have already saved Rs 9.1 crore and expects an additional Rs 18.2 crore saving over the next two years ("Vodafon Wins").

Challenges Facing Data Warehouse

Now that data warehouse is being extensively used in decision making, it has turned out to be the silver bullet out of operational data abyss. When data warehouse is approached with some information and (mis)management principles, it could result in producing disintegrated and untrustworthy data, and this could result in the failure of data warehouse.

Analyzing the reasons for the failure of data warehousing projects, Larry P. English writes:

Data warehousing projects fail for many reasons, all of which can be traced to a single cause: non-quality. Poor data architecture, inconsistently defined departmental data, inability to relate data from different data sources, missing and inaccurate data values, inconsistent use of data fields, unacceptable query performance (timeliness of information), lack of business sponsor (no data warehouse customer), and so forth, are all components of non-quality.

Data or information of poor quality paralyzes virtually every area of a business, from mailroom to executive office. The time invested by the business in digging missing data, correcting inaccurate data, working around data problems, scrambling to assemble information across disintegrated databases, resolving data-related customer complaints, etc. adds to the cost, which is inflated and passed on to the customer. That time is not available for value-adding work. Senior executives at one large mail-order company personally spend the equivalent of one full-time senior executive in reconciling conflicting departmental reports before submitting them to the chief executive officer. (This means there is an equivalent of one senior executive redundantly required because of redundant and inconsistent (non-quality) data! ("Information Management").

With all the emphasis on data warehousing technologies, it is now important to understand that:

- The product of the data warehouse is information.

- The customers of the data warehouse are the knowledge workers who must make increasingly important decisions faster than ever before.

- If the data warehouse does not deliver reliable information that supports the customer's decisions and strategic processes, there will be a greater chance of failures in decision making which results in heavy business loss ("Information Management").

Solution and Further Enhancement

Non-quality data can be minimized by keeping an integrated data environment. This requires that data warehouse professionals need to perform the following steps:

- Map undefined and un-integrated data from many disparate and redundant database and files

- Standardize the schema and data representations.

- Remove redundant occurrences of data both within single files and across redundant files

- Integrate and consolidate data and format it into an integrated data warehouse data architecture.

- Identify and manage different data integration issues.

- Should not include the cost of customer complaints, product liability lawsuits, redoing defective work, etc.

- Purpose of information quality should be to improve customer and stakeholder satisfaction by increasing the efficiency and effectiveness of business process ("Information Management").

This could be further achieved by developing more powerful tools that are data warehouse aware-tools, designed for and are capable of dealing with very large databases. Some of the features that will appear in tools in the near future are:

- Optimizer-aware tools.

- Accurate query time prediction.

- Accurate resource usage prediction.

- Improved summary management.

- Improved integration of data generation, visualization and browsing (Anahory).

Conclusion

With the vast enhancement in business sectors and high competitive environment, there arose the need for some method that helps in taking proper decisions. Through this paper, a few ideas have been put together to show how data warehouse will help in taking complex business decisions. As compared to the invention age, today, the business world is known to be more dependent on knowledge base.

Quite often it has been quoted in magazines and discussions that nowadays information in the hand of the right people is more valuable than gold. All decision makers are dependent on how much information they have. A simple piece of information can make a big difference in the life of a company.

The great advancement in technology today has given a variety of tools to analyze data and information. It has also opened up a number of BI vendors like SAP, Oracle Apps, PeoplesSoft, Cognos, Micorsoft, Cognos, etc. to name a few. Business individuals are an expert in their field. Today, to succeed they will need to master business intelligence, or hire expert BI consulters from the above-mentioned vendors.

Business needs to choose the right IT solution to strengthen itself. It will not only create efficiency but also provide a number of opportunities to produce competitive advantages, revenue opportunities and new avenues for its customers and partners to communicate.

This paper can be enhanced to include advanced and proficient algorithms that could be used along with data warehouse in processing data and information more efficiently and quickly. More and more data is being collected and stored in data warehouse. Now the challenge before the data analyst and data manager is how the proficiency of data retrieval could be enhanced and how the queries could be improved so that the changing demands could be easily addressed.

Works Cited

Agrawal, Sujata and Jai Wadia. "Tata Nano: The Big Wonder," 2 Jan. 2013. <http://www.Tata.com/company/articles/inside.aspx?artid=pjl3huSRc+0=>

Anahory, Sam and Dennis Murray. *Data Warehouse in Real World.* 4th impression. Delhi: Pearson Education, 2009.

"Data Warehouse Definition." 1KeyData.com. Web. 12 Jan. 2013 <http://www.1keydata.com/data-warehouse-definition.html>

English, Larry P. "Improving Data Warehouse and Business Information Quality." Wiley Computer Publishing, 1999. Web. 29 Dec. 2012 <http://www.infoimpact.com /IQBook/iq.pdf>

Guerra, Joseph and David Andrews. "Why You Need a Data Warehouse." Andrews Consulting Group, 2011: 4. Web. 29 Dec. 2012 <http://www.andrewscg.com /pdfs/why-you-need-a-data-warehouse.pdf>

"Information Management." Web. 2 Jan. 2013 <http://www.information-management.com/issues /19980101/771-1.html>

"Vodafone wins customers with a landscape approach to BI." SearchBusinessIntelligence.in. TechTarget. Web. 12 Dec. 2012. <http://searchbusinessintelligence.techtarget.in /survey /Vodafone-wins-customers-with-a-landscape-approach-to-BI>