Enhancing Business Intelligence using Data Warehousing: A Multi Case Analysis

Sonal Sharma[1]
IT- Department
Uttaranchal University
Dehradun - India

Rajni Jain[2]
NCAP
New Delhi - India

Abstract: In the recent years, data warehouses are designed for various industry in order to analysis their data and growth. It has been extensively used for storing the data and gaining knowledge. In this paper authors have focused on various benefits that could be brought to fruition by designing data warehouse. Further this work has been extended to analyse multiple cases of the organisations like GPCL, ICICI prudential life insurance, Singapore Land Transport etc. to understand the real benefits gained after successful implementation of data warehouse by them.

Keywords: Data Warehouse; OLTP; OLAP; ETL; Business intelligence.

I. INTRODUCTION

The emergence of data warehousing was initially a consequence of W. Inmon and E.F. Codd in early 90's that OLTP (On Line Transaction Processing) and OLAP (On Line Analysis Processing) can't exist efficiently in same database environment mostly due to their very different transaction characteristics [Inmon 96]. To analyze the development of an organization, measures such as the number of transactions per customer or the increase of sales during a promotion are used to recognize warning signs and to decide on future investments with regard to the strategic goals of the organization. In order to answer such questions the OLTP approach is not sufficient as OLTP deals with the daily transactional data stored in databases. Whereas Data warehouse have different requirements, they deal with OLAP. The data stored in data warehouses is cleaned, temporal (historic), summarized, and non-volatile. The requirement to have an upper edge in this competitive world leaded to the need of warehouse. The heterogeneously existing databases of any small or large organization are integrated to form a data warehouse. It is organized under a unified schema at a single site to facilitate management decision making. A Data warehouse (DW) is a collection of technologies aimed at enabling the decision maker to make better and faster decisions. Data warehouses differ from operational databases in that they are subject oriented, integrated, time variant, non-volatile, summarized, larger, not normalized and perform OLAP[6]. Currently, data warehouses are implemented in various sector viz- banking, telecommunication, insurance, retail industry, agriculture etc. Lot of research has been done for defining model for data warehouse designing. Since then data warehousing became an important strategy to integrate multiple heterogeneous data sources under a unified schema at a single site in order to facilitate management decision making. Barquin Devlin defines data warehouse as a single, complete and consistent store of data obtained from a variety of different source made available to end users in a way that they can understand and use in business context. The data warehouses are supposed to provide storage, functionality and responsiveness to queries beyond the capabilities of today's transaction-oriented databases. Also data warehouses are set to improve the data access performance of databases.

Since that era, data warehousing has emerged as a paranormal baton in an organization's hand. In this paper various aspects related to the need of data warehousing and how could it help any organization to have an upper edge in today's cut-throat

competition has been discussed. Since its advent, data warehouse has not only immensely supported the top management in improvising their business but has also provided them the true insight of their business, its pitfalls and strengths.

Section 2, discusses the advent of data warehouse where as the section 3 focuses upon the factors which enforces the urge of data warehouse in any unit. Section 4 reveals various benefits that could incur in any environment on the successful implementation of data warehousing. In Section 5, six case studies are referred where the aim is to understand the net benefit an organization has achieved.

II. ADVENT OF DATA WAREHOUSING

During its course of evolution the database technology has undergone through different phase. Right from earlier mechanism for collecting data and creating databases to the development of efficient, more reliable, secured mechanism for data storage, retrieval, query and transaction processing.[6].Earlier in 1960's primitive file processing was used for storage of collected data. In the era of 1970s & early 1980s the DBMS progressed from hierarchical and network based systems to relational database system. Tools for modeling data(ER model), techniques for indexing and organizing data (B+ tree, hashing) evolved in this time period[6][9].The development of 4GL, user interfaces forms, reports allowed users to interact, access and retrieve data from database more conveniently. Query processing, consistency control and the recovery of data at the time of crash played substantial role in the evolution of database technology [7][9]. Database system developed in mid 80's featured advance data models (extended- relation, object oriented, object-relational, deductive) and application oriented systems.

During this period data was distributed widely, heterogeneous database systems were emerging. In order to integrate the scattered data, analyze the data for decision making the concept of data warehousing was introduced [2][10].The need for separate enterprise-wide integrated information retrieval for decision making is the basis for data warehouse. Table 1, compares the database with data warehouse and justify the need of a separate data warehouse.

Database	Data warehousing
Holds current data	Holds historical data.
Data is dynamic.	Data is largely static
Data can update or change.	Data cannot update or change.
Not used for comparison, trend and forecasting.	Used for comparison, trend and forecasting.
Read/Write access.	Read only accesses.
Supports thousands of concurrent users in data base.	Supports few concurrent users relative to OLTP.
Optimized for validation of incoming data =during transactions; uses validation data tables	Loaded with consistent, valid data; requires no real time validation.
Doesn't reduce the cost much in relative to data warehouse.	Compared to database, data warehouse drastically reduces the 'cost-per-analysis.
Used by clerical staff for day-to-day operation.	Used by top managers for analysis.
Have different types of data.	Data warehouse has unique data.
Transaction driven.	Analysis driven.
Repetitive Processing.	Adhoc complex queries.

Table I: comparison b/w databases and data warehouse [6][10]

The generic data warehouse architecture [Fig 1] consists of three layers (data sources, DSA and primary data warehouse) (Inmon 2002; Vassiliadis, 2000). To build a DW we must run the ETL tool which has three tasks: (1) Extraction of data from different data sources, (2) propagation of data to the data staging area where it is transformed and cleansed, and then (3) loading the data to the data warehouse. Many researchers and practitioners share the understanding that data warehouse architecture can be formally understood as layers of materialized views on top of each other. Data warehouse architecture **[Fig 1]** exhibits various layers of data in which data from one layer are derived from data of the lower layer.

The Operational Data Store serves also as a buffer for data transformation and cleaning so that the data warehouse is populated with clean and homogeneous data. The next layer of views is local client, or warehouses, which contain highly

aggregated data, directly derived from the global data marts OLAP warehouse. There are various kinds of local warehouses, such as the databases, which may use relational database systems or specific multidimensional data structures.

Depending upon the available resource, technology, infrastructure, budget constraint and specific requirement a single data warehouse could be designed which would incorporate all the elements of all the units of organization (Enterprise Data Warehouse) or   a repository of data for a specific unit could be designed to answer a specific business query (Data Mart). The implementation schedule of EDW is comparatively lengthy as it is technology driven which affects various organizational units. Multiple departments of the organization are involved in the EDW so its consequences are very high on the entire organization. Whereas the amount of data stored is just few MB to a hundred of GB in a data mart. The focus is on the specific business matter so the development schedule is very short.
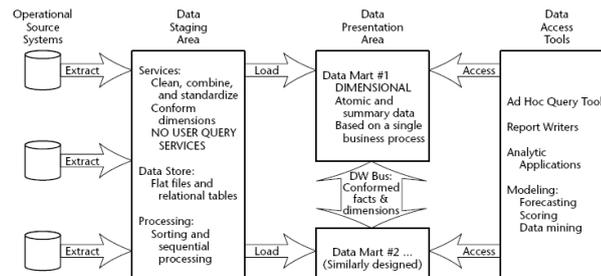

Figure 1: Components of Data warehouse [9]

III. NEED FOR DATA WAREHOUSING

A data warehouse is a repository of an organization's stored data. Their purpose is to facilitate analysis and reporting. It helps the organizations to analyze the trends depicted from the data stored with the organization over time. The main function of the data warehouse is to facilitate the organization in planning strategically on the basis of long term data. On the basis of analysis various forecasts, business models and prognosis can be made.

In order to take wise decisions and be ahead in competition, data warehouse has become a must to have tool. Data warehousing exclusively differentiate the data and information. As the processed data is the information which is generated through the data capture and properly stored in the data warehouse. That is why, data warehouses becomes the base of Business Intelligence (BI).

Information retrieval is possible if and only if the data is stored in organized, cleaned and in standard format then only business intelligence could act upon that data. The ultimate goal of the data warehouse is to integrate the data that is heterogeneously stored at various sites, then only that data could be used for decision support.

Beside to be useful as a tool for organizing data, data warehousing can also be used to act as an archival copy in case of system failure which may occur due to natural calamities, accidental deletion or any crash. Even if the company faces the trouble the industry around world keeps on progressing, so the loss of organizational data is not bearable at any cost. That is also one of the reasons that companies organize and maintain their very important information in the data warehouses.

Data warehouses, now a day, are implemented in various sectors viz banking, health care, academics, retail sector, telecommunications, biological surveillance etc[Hoffer et al., 2005;Inmon 2002]Even the government sector is adopting data warehouse technologies. The Government of India has accomplished a project of designing and deploying the Integrated National Agriculture Resource Information System (INARIS) data warehouse for the agriculture sector.[Sree Nilkanta et al 2008]. The data warehouses have been created by various countries in different domains.

As mentioned in section 1, the data warehouse design is very distinct as compared to traditional databases deign. As the data stored in warehouse is from the already existing information stored in the database. Data warehouse faces some constraints regarding the quality of data, amount of data to be stored, and the granularity of data that is to be considered for the design of

data warehouse. The very first requirement for designing the DW is of gathering the need of the major business process. The knowledge and access to the performance metrics is also needed so that the data can be feed to the data warehouse. The level of levels that can be stored in the warehouse depicts the granularity of the data warehouse. More the detail data less is the granularity and less the detailing more is the granularity. According to Bill Inmon, granularity is the most important issue in the design of data warehouse which affects the storage capacity and performance and in turn, the overall analysis done through the data warehouse.

IV. BENEFITS OF DATA WAREHOUSING

There are number of benefits of Data warehousing that are witnessed in the literature. Data warehouse help the organization immediately as well as it also enhances long term positive gains. In[6][7][8][10][, authors have explored the benefits of data warehouse on return on investment, enhancement in decision making, timely access to data, consistency in data and improved system performance.

- **Return on Investment (ROI):** They explained ROI as the amount increased or decreased on the money invested. Data warehouse implementation provides lots of saving for organization and has positive effect on the growth of the company. According to a 2002 International Data Corporation (IDC) study. "The Financial Impact of Business Analytics", analytics projects have been achieving a substantial impact on an organization' financial state.

- **Enhanced Business Decisions:** The decision of the organization depends on the encapsulated data of the organization which is evidently stored in the data warehouse. Because of the accurate analysis obtained from the DW, now the managers and executives need not to depend on their personal, learned knowledge, rather they can rely on the highly organized and accurate reports

- **Timely Access to Data:** as such the data of the organization is stored at various locations, so the reterival of the information is not an easy task but the creation of data warehouse enforces the integration of data at a unified place, so iit becomes quite fast for the organization to access their data in no time. The integration of data is done with the ETL tools on the regular basis. The management people now can access the entire information using one interface only. They need not to depend upon the compilation of heterogeneously stored data to answer each query. This will also facilitate the managers to directly use the query and analysis tools without the involvement of the technical professional. This would also reduce the waiting time.

- **Consistency of Data:** the inconsistency in the format of data storage will also vanish away. As the entire data of the organization would be stored in the standard format at a central location. It also allows all the functional unit of an organization to use same data source to respond to their queries. Thus every unit would get the same image of the organization growth and their decision would depend upon the actual position of the organization.

- **System Performance:** the main concern of data warehouse design is the speed of data retrieval.so the data in the data warehouse has to be stored in organized fashion so as to provide the most optimized query response. The data store for the daily processing is maintained with different perspective whereas the information lying in the data warehouse is stored to answer analytical queries. Due data warehouse the burden of large system is taken off from the operational environment, and it efficiently and effectively divides the load across entire infrastructure.

- **Increased ICT staff productivity.** Being able to proactively resolve incidents and quickly pinpoint the source of problems has enabled ICT to increase system uptime without adding additional staff. As part of the project, Imperial employed a third party to receive and respond to alerts outside working hours, so technical experts can be contacted outside working hours if a problem needs their immediate attention. This both reduces disruption to end users and provides faster incident resolution without additional ICT staff hours or overtime.

- **Increased customer satisfaction.** Before Fog light, IT depended on service tickets to know when an application had a usability problem. Today, users know that if an application becomes unavailable or suffers a performance problem the issue is likely to be identified within 15 minutes and subsequently resolved.
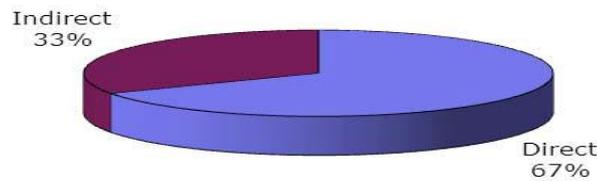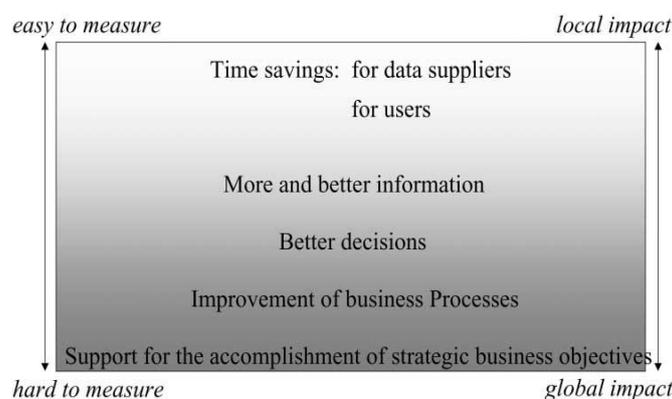

Fig 2: Benefits gained through implementation of data warehouse. Source [10]

Figure 2 depicts that the companies are directly benefited by the data warehouse implementation and indirectly they gain profit. Figure 3 elaborate the various impacts (both global and local) of data warehouse implementation.

- **A Data Warehouse Saves Time:** Since business users can quickly access critical data from a number of sources—all in one place—they can rapidly make informed decisions on key initiatives. They won't waste precious time retrieving data from multiple sources. Not only that but the business execs can query the data themselves with little or no support from IT—saving more time and more money. That means the business users won't have to wait until IT gets around to generating the reports, and those hardworking folks in IT can do what they do best—keep the business running.

- **A Data Warehouse Enhances Data Quality and Consistency:** A data warehouse implementation includes the conversion of data from numerous source systems into a common format. Since each data from the various departments is standardized, each department will produce results that are in line with all the other departments. As a result accurate data is available and accurate data is the basis for strong business decisions.

- **A Data Warehouse Provides Historical Intelligence:** A data warehouse stores large amounts of historical data so one can analyse different time periods and trends in order to make future predictions. Such data typically cannot be stored in a transactional database or used to generate reports fr1om a transactional system.

V. CASE STUDY

CASE I: Godrej Consumer Products Limited)

In 1995 GCPL (Godrej Consumer Products Limited) implemented MFG/PRO, an ERP package from QAD Inc. in its 120 companies and locations. Lots of ERP data was lying idle so in order to extract more out of the well-formatted ERP data GPCL implemented data warehousing. "We devised a way to extract the most out of this critical data. We decided to implement data warehousing applications and functions," said Mani B. Mulki, General Manager, Information Systems, Godrej Industries Limited.

Benefits achieved: The companies was able to get contribution analyses, profit and loss analyses, and sales breakup analyses from the data in its warehouse. This has given it competitive advantage and the ability to manage resources better. The applications have also paved way for future e-commerce initiatives. "You can view interesting trends in sales and costs. And if you like, you can integrate the findings from standard reports and the warehouse analyses to create strategic advances. The limit of the capabilities of a warehouse is bound by the creativity of the end-user," said Mani.

CASE II: Northland Group

Since its inception in 1982, the twin cities based Northland group is committed service to bank card and retail clients has made Northland Group a proven leader in those markets. In order to maintain and enhance its cutting-edge technology, the company implemented the data warehousing project designed by Lancet Software India, Pvt. Ltd

Benefits Achieved: "Today Northland Group has the tools in place to identify areas within our business where costs can be reduced," Trochlil explains. "Because we had such good coaching from Bruce, we also have recognized a genuine time reduction in retrieving massive amounts of data that needs to be analyzed. We now have the capability to export the data to a reporting format that can be readily shared among key decision-makers.""With the implementation of this project, there have been more departments identified that we are going to move into the warehouse. Overall, this process taught us that our ability to understand data is now greater than at any time in our history", according to Jennifer Trochlil, Senior Programmer Analyst at Northland.

Case III Supermarket Chain

Q3 designed a data warehouse for a client who has the largest supermarket chain in Italy. The first shop of supermarket was established in 1854. As of 2005, client operates nearly 1,300 shops. Company's requirement from a data warehouse was to provide strategic and tactical decision support to all levels of management. So, Q3 implemented the data warehouse in Oracle 10g R2. OWB (Oracle Warehouse Builder) was used as an Extract, Transfer & Load (ETL) tool for loading data into the data warehouse from different data sources: Data consolidated from various data sources such as ERP systems, OLTP databases. They build various cubes viz PROFIT_LOSS Cube - Keeps Profit & Loss, Cash Flow. BALANCE_SHEET Cube - Keeps the Balance Sheet.

SALES Cube - Keeps daily sales of all categories in all stores.

SALES_DAY Cube - Keeps sales amounts and counts at 45 minute intervals.

SALES_OPERATIONS Cube - Keeps daily sales summary for each store.

PRODUCT cube Contains all the categories like Garments (Jeans, women wear, men's wear and kids wear), cars, and other household stuff.

Q3 Technologies have used OLAP (Essbase) data mining algorithms to do a trend analysis which helps to all levels of management to make a decision for a specific brand, zone, and category in terms of sales and time graph. For reporting they used Oracle Business Intelligence Enterprise Edition (OBIEE) and Excel Add-in.

Benefits Achieved:

- Data Warehouse designed for analysis, pattern search and reporting has been created.

- Integrated budgeting & planning processes in a centralized Web-based application.

- Combines specific departments and business unit plans with organization objectives. Reduces budgeting & planning cycles cost.

- Avails reporting data for different perspective on the basis of time.

- It also minimizes both IT support and the technical expertise required by the management.

Case IV Financial Service Group (Infosys)

Infosys designed and implemented a data warehouse solution using iterative phased approach to extract information from the Mortgage Sales Application and administration systems of different brands and house them in a single data warehouse database for its Client who was one of the Europe's largest financial service groups.

Benefits Achieved:

- Transaction volume expected: 73 Million per year; annual growth rate of 110%

- Size expected: 180 GB at the end of Year 1; annual growth rate of 45%

Case V ICICI Prudential Life Insurance

The aim of Dipak Nair, VP-IT, ICICI Prudential Life insurance was to take the business analytics at higher levels by integrating data from business areas and correlate them accurately. These would help them to be ahead in the competitive world of insurance sector in India. "To enable business teams with greater control on business initiatives and increase returns from it, the right kind of data needs to be provided at the right time," says Dipak Nair, VP-IT, ICICI Prudential Life Insurance. "The platform not only needed to cater to the information needs of each business unit in parallel, but it also needed to meet performance objectives" says Nair. The new information architecture pushed ICICI Prudential to build an Active Enterprise Data Warehouse on which functional data marts were formed for targeted business units.

Benefits Achieved:

- Information and analytical capabilities were provided with the solution.

- Data across various dimensions could be integrated as a result complete performance of the company as a whole could be retrieved.

- Accurate and better decision making was possible.

- The solution also lowered down the processing and reporting time by 50%.

Case VI Singapore Land Transport (by Wipro)

In august 2008 higher authorities of Singapore Land transport decided to build a large data warehouse called PLANET[] from various business sources and support queries over 4.6 billions of transactions. The proofs of concept were promising, so in March 2009, it decided to invest $12 million in a full-scale implementation of Planet. The LTA worked with Wipro on the design and implementation, and the LTA chose data warehouse, data integration and data quality, and analytic technologies from Teradata, Informatica and SAP Business Objects respectively.

Benefits Achieved:

- Planet gives users new ways to view the data gathered.

- The system's storage capacity has increased to three years' worth of records for public transport transactions, compared to three months' worth with the old system.

- The time it takes for extraction, transformation and loading has fallen 67% while the time required for queries has dropped 99%.

- A query of 100 million records would have taken 18 hours to complete before Planet; now they take 15 minutes.

- Planet offers rich features, such as ad hoc queries and predictive modelling.

- Within the first three months of operation, users ran more than 4,000 new queries and created more than 70 new reports (in addition to 30 legacy reports) for policy reviews, analysis of new transport schemes and trend patterns to optimize resources.

VI. Conclusion

Since its advent the data warehouses has helped top management in understanding the insight of their organisation. It has predicted the future well in advance on the basis of historical data stored in the database and thus influenced the managers to take strategical and tactical decisions. The consistency of data could be achieved due to unified storage. The returns on investment were increased. It saves the time by quickly responding to the queries, in some cases up to 99% and thus increased the customer's satisfaction. The companies were able to get contribution analyses, profit and loss analyses, and sales breakup analyses from the data in its warehouse. This has given competitive advantage and the ability to manage resources better.

There are many more such case studies available in which the organization has been benefited extensively benefited by the implementation of data warehouse. To mention few, the NHTSA by WIPRO, DW for DP world Australia and Australia National University by Altis Consulting.

References

1. Arnott D., Pervan G.,"Eight key issues for the decision support systems discipline",Decision Support System,Elsevier,Pg No. 657-672,2008.
2. Arnett S.,"Data Warehousing: the Key for a successful Implementation",White paper
3. Bassiliadas N.,Elmagarmid A.K.,"Interbase-KB: Integrating a knowledge base system with a multidatabase system for data warehousing", Computer, IEEE, Vol 16, No. 5,Pg no. 1188-1205.
4. Beg Javed, Hussain Shadab,"Data quality- A problem and An Approach", White paper
5. Berndt D.J.,Fisher J.W.,Hevner A.R.,Studnicki J.,"Healthcare Data warehousing and quality assurance",Computer,IEEE,pg no 56-65,2001
6. Han J. and Kamber M.,"Data Mining: Concepts and Techniques", 2nd Ed, Elsevier, 2007
7. Inmon W.H.,"Building the Data Warehouse", John Wiley & Sons, Inc.1992
8. Jarke, M., Lenzerini, M., Vassiliou, Y., Vassiliadis, P.,"Fundamentals of Data Warehouses", second ed. Springer-Verlag.
9. Kimball R.,"The data warehouse toolkit: Practical technique for building dimensional data warehouses",John Wiley&sons,New York,1996
10. Lane P.," Oracle9i Data Warehousing Guide, Release 2 (9.2)", Oracle Corporation, Part No. A96520-01,1996, March 2002
11. Rai Anil," Data warehouse and its applications in agriculture", available at http://www.inaris.gen.in
12. Rizzi, S.," Conceptual Modeling and Evolution in DWs.", Perspectives Workshop: Data Warehousing at the Crossroads, Dagstuhl, August 1-8 (2004)
13. Samtani S., Mohania M., Kumar V., Kambayashi Y.," Recent Advances and Research Problems in Data Warehousing", Springer, LNCS 1552, p p . 81–92, 1999
14. Sharma S.D., Singh R.,Rai Anil,"Integrated National Agriculture Resources Information system" , http://www.inaris.gen.in
15. Sharma S, Kumar H,"Data warehouse design issues in forestry: - Eucalyptus tereticornis","IJCIR, Volume 7, Number 1 (2011), pp.109-114.
16. Sharma S., Soni P.," Data warehousing: Challenges and Opportunites", "in press".

Web Site

1. **www.dwinstitute.com**
2. **www.inaris.gen.in**
3. **www.wisegeek.com/what-is-data-warehousing.htm**
4. **www.exforsys.com/tutorials/data-warehousing.html**
5. **www.dwinfocenter.org**
6. **www.ibm.com/software/data/infosphere/warehouse**
7. **www.dwai.org**
8. **www.bi-dw.info**
9. **www.dwreview.com**
10. **www.altis.com.au**