

---

# An Enhanced design of Data warehouse using Business Intelligence to perform multi-case analysis

---

K.B.Jagadeesh Kumar

Dhruva Institute of Engineering & Technology, Hyd. Dept of CSE

**Abstract:** *Business intelligence includes tools and techniques for data gathering, analysis, and visualization for helping with executive decision making in any industry. Data mining includes statistical and machine-learning techniques to build decision-making models from raw data. 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 which uses Business Intelligence for designing a Data Warehouse which is used to perform Multi-case analysis.*

## 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 led 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. INCEPTION OF DATAWAREHOUSE

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.

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.

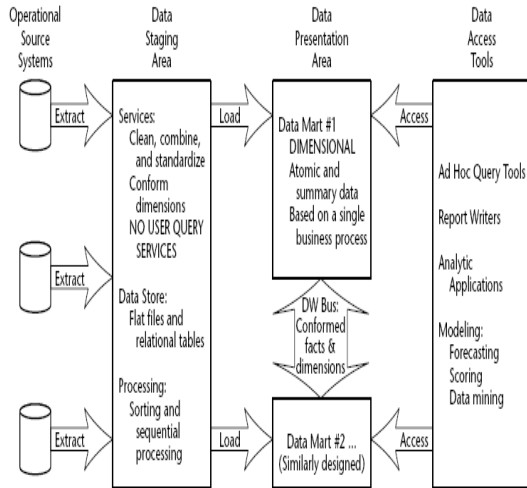


Figure 1: Components of Data warehouse [9]  
**III. NEED FOR DATA WAREHOUSING**

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 design. 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.

**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 analyze 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 from a transactional system.

## V. DW DEVELOPMENT APPROACHES

There are two fundamentally different approaches to developing DW: top down and bottom up. The top-down approach is to make a comprehensive DW that covers all the reporting needs of the enterprise. The bottom-up approach is to produce small data marts, for the reporting needs of different departments or functions, as needed. The smaller data marts will eventually align to deliver comprehensive EDW capabilities. The top-down approach provides consistency but takes time and resources.

## VI. BI FOR BETTER DECISIONS

The future is inherently uncertain. Risk is the result of a probabilistic world where there are no certainties and complexities abound. People use crystal balls, astrology, palmistry, ground hogs, and also mathematics and numbers to mitigate risk in decision-making. The goal is to make effective decisions, while reducing risk. Businesses calculate risks and make decisions based on a broad set of facts and insights. Reliable knowledge about the future can help managers make the right decisions with lower levels of risk. The speed of action has risen exponentially with the growth of the Internet. In a hypercompetitive world, the speed of a decision and the consequent action can be a key advantage. The Internet and mobile technologies allow decisions to be made anytime, anywhere. Ignoring fast-moving changes can threaten the organization's future. Research has

shown that an unfavorable comment about the company and its products on social media should not go unaddressed for long. Banks have had to pay huge penalties to Consumer Financial Protection Bureau (CFPB) in United States in 2013 for complaints made on CFPB's websites.

On

the other hand, a positive sentiment expressed on social media should also be utilized as a potential sales and promotion opportunity, while the opportunity lasts.

### **Decision Types**

There are two main kinds of decisions: strategic decisions and operational decisions. BI can help make both better. Strategic decisions are those that impact the direction of the company. The decision to reach out to a new customer set would be a strategic decision.

Operational decisions can be made more efficient using an analysis of past data. A classification system can be created and modeled using the data of past instances to develop a good model of the domain. This model can help improve operational decisions in the future. BI can help automate operations level decision-making and improve efficiency by making millions of micro-level operational decisions in a model-driven way. For example, a bank might want to make decisions about making financial loans in a more scientific way using data-based models. A decision-tree-based model could provide a consistently accurate loan decisions. Developing such decision tree models is one of the main applications of data mining techniques.

Effective BI has an evolutionary component, as business models evolve. When people and organizations act, new facts (data) are generated. Current business models can be tested against the new data, and it is possible that those models will not hold up well. In that case, decision models should be revised and new insights should be incorporated. An unending process of generating fresh new insights in real time can help make

better decisions, and thus can be a significant competitive advantage.

## **VII. BI Tools**

BI includes a variety of software tools and techniques to provide the managers with the information and insights needed to run the business. Information can be provided about the current state of affairs with the capability to drill down into details, and also insights about emerging patterns which lead to projections into the future. BI tools include data warehousing, online analytical processing, social media analytics, reporting, dashboards, querying, and data mining.

BI tools can range from very simple tools that could be considered

end-user tools, to very sophisticated tools that offer a very broad and complex set of functionality. Thus, Even executives can be their own BI experts, or they can rely on BI specialists to set up the BI mechanisms for them. Thus, large organizations invest in expensive sophisticated BI solutions that provide good information in real time. A spreadsheet tool, such as Microsoft Excel, can act as an easy but effective BI tool by itself.

## **VIII. CONCLUSION**

Since its advent the data warehouses has helped top management in understanding the insight of their organization. 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 strategically 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.

Business intelligence includes tools and techniques for data gathering, analysis, and visualization for helping with executive decision making in any industry. Data mining includes statistical and machine-learning techniques to build decision-making models from raw data.

## 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)