ISSN 2250-3153

The Study on Data Warehouse Design and Usage

Mr. Dishek Mankad[1], Mr. Preyash Dholakia[2]

[1] M.C.A., B.R.Patel Institute of Computer Application [MCA Program]
[2]M.C.A. K.S.K.V. Kachchh University MCA College

Abstract- Data ware housing is a booming industry with many interesting research problem. The data warehouse is concentrated on only few aspects. Here we are discussing about the data warehouse design and usage. Let's look at various approaches to the data ware house design and usage process and the steps involved. Data warehouse can be built using a top-down approach, bottom – down approach or a combination of both. In this research paper we are discussing about the data warehouse design process.

Index Terms- Analysis, Data Warehousing, Data Warehouse Design, Process.

I. INTRODUCTION

Before we seen the design process, let's we seen about what is data warehouse? Think of a data warehouse as a central storage facility which collects information from many sources, manages it for efficient storage and retrieval, and delivers it to many audiences, usually to meet decision support and business intelligence requirements. "What is the need of data warehouse? What goes into a data warehouse design? How are data warehouse used? How do data warehousing and OLAP relate to data mining?" In this research paper we are discussing about business analysis framework for data warehouse design, data warehouse design process, data warehouse usage for information processing and from OLAP to multidimensional data mining. The concept of data warehousing is deceptively simple. Data is extracted periodically from the applications that support business processes and copied onto special dedicated computers. There it can be validated, reformatted, reorganized, summarized, restructured, and supplemented with data from other sources. The resulting data warehouse becomes the main source of information for report generation, analysis, and presentation through ad hoc reports, portals, and dashboards. Building data warehouses used to be difficult. Many early adopters found it to be costly, time consuming, and resource intensive. Over the years, it has earned a reputation for being risky. This is especially true for those who have tried to build data warehouses themselves without the help of real experts.

II. RESEARCH ELABORATIONS

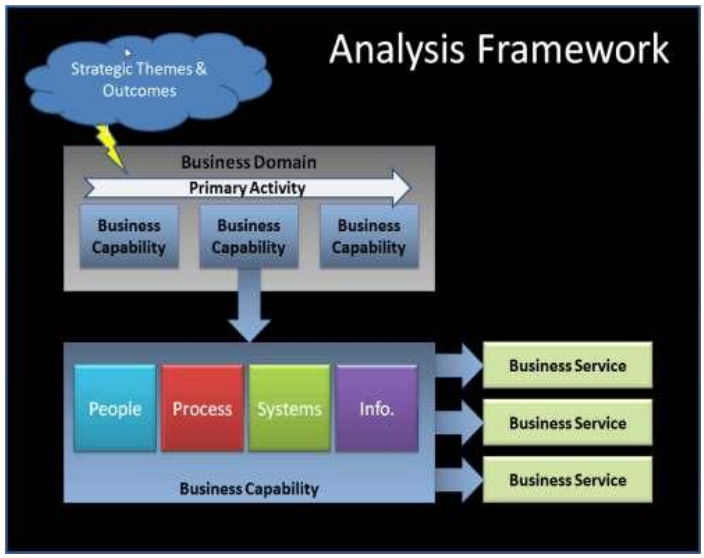
[A] A BUSINESS ANALYSIS FRAMEWORK FOR DATA WAREHOUSE DESIGN:

"What can business analyst gain from having a data warehouse?" **First,** we having a data warehouse may provide a competitive advantage by presenting relevant information from which to measure performance and make critical adjustments to help win over competitors. **Second,** a data warehouse can enhance business productivity because it is able to quickly and efficient gather information that accurately describes the organization. **Third,** a data warehouse facilitates customer relationship management because it provides a consistent view of customers and items across all lines of business, all departments, and all markets. Finally, a data warehouse may bring about cost reduction by tracking trends, patterns and exceptions over long periods. If you wanted to do design effective data warehouse you must know the business needs and construct a business analysis framework. The construction of a large and complex information system can be viewed as the construction of a large and complex building, for which the owner architect and builder have different views. This view are combined to form a complex framework that represents the top-down, business-driven, or owner's perspective, as well as the bottom-up, builder-driven, or implementor's view of the information system. **Four,** different views regarding a data warehouse design must be considered: the top-down view, the data source view, the data warehouse view, of the information system.

- ✓ **The top - Down view** allows the selection of the relevant information necessary for the data warehouse. This information matches current and future business needs.
- ✓ **The Data source view** exposes the information being captured, stored, and managed by operational system. This information may be documented at various levels of detail and accuracy, from individual data source tables to integrate at various levels of detail and accuracy, form individual data source tables to integrated data source tables. Data sources are often modeled by traditional data modeling techniques, such as the E-R model or DASE tools.
- ✓ **The Data warehouse view** includes fact tables and dimension tables. It represents the information that is stored inside the data ware house, including precalculated totals and counts, as well as information regarding the source, date and time of origin added to provide historical context.
- ✓ **The Business Query View** is the data perspective in the data warehouse form the end-user's view point

So, building and using a data warehouse is a complex task because it requires business skill technology skills, and program management skills. Regarding business skills, building a data warehouse involves understanding how systems store and manage their data, how to build extractors that transfer data from the operational system to the data ware house, and how to build

warehouse refresh software that keeps the data warehouse reasonably up-to-date with the operational system's data.



A frame work provides the structure for the upsurge of using analysts of all sorts: business analysts, business process analysts, risk analysts, system analysts, and provides a standardized way to gather, communicate and develop the desired information required by:

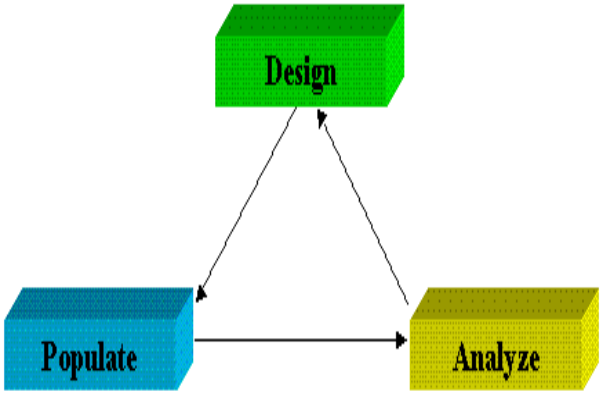
- The Program Management Office;
- Business users;
- key stake holders and
- Technology developers.

Based on our experience, even for projects that are completed on time and on budget, there may be significant inefficiencies in performing business analysis functions. These inefficiencies include:

- lost opportunities
- Rework
- No realization of benefits

By implementing a framework you provide structure and standards that are intended to serve as a support or provide guidance for your BA's. You can expect consistent quality outputs from your BA resources and provide the ability to attract and retain experienced and motivated BA's to: Reduce waste Create solutions Complete projects on time Improve efficiency Document the right requirements A Framework enables your organization to bring to market your competitive innovations more effectively and efficiently and to underpin an increase in the delivery of successful projects.

[B] DATA WAREHOUSE DESIGN PROCESS



Here we discussed about various approaches to the data warehouse design process and the steps involved. A data warehouse can be built using a top-down approach, a bottom-up approach or a combination of both. The top – down approach starts with overall design and planning. It is useful in cases where the technology is mature and well known, and where the business problems that must be solved are clear and well understood. The bottom -up approach starts with experiments and prototypes. This is useful in the early stage of business modeling and technology development. And it also allowed an organization ot move forward at considerable less expenses and evaluate the technological advantages before making significant commitments. In the combined approach, an organization can be exploit the planned and strategic nature of the top-down approach while retaining the rapid implementation and opportunistic application of the bottom – up approach. If we are thinking in from the software engineering point of view, the design and construction of a data analysis, warehouse design , data integration and testing, and finally deployment of the data warehouse. Large software systems can be developed by using one of the two technologies. The Waterfall method and The spiral method. So, here it is.

The Water Fall method performs a structured and systematic analysis at each step before proceeding to the next, which is like a water fall, falling form one step to the next. The Spiral Method involves the rapid generation of increasingly functional systems, with short intervals between successive releases.

This is always considered as a good choice for data warehouse development, especially for data marts, because the turnaround time is short, modifications can be done quickly, and new designs for the technologies and that can be adapted in a timely manner. So, here we are discussed about the warehouse design process. This includes various steps as follows:

Choose a Business Process to Model if the business process is organizational and involves multiple complex object collections, a data warehouse model should be followed. However, if the

process is departmental and focuses on the analysis of one kind of business process, a data mart model should be chosen.

Choose the business process gain, which is the fundamental, atomic level of data to be represented in the fact table for this process.

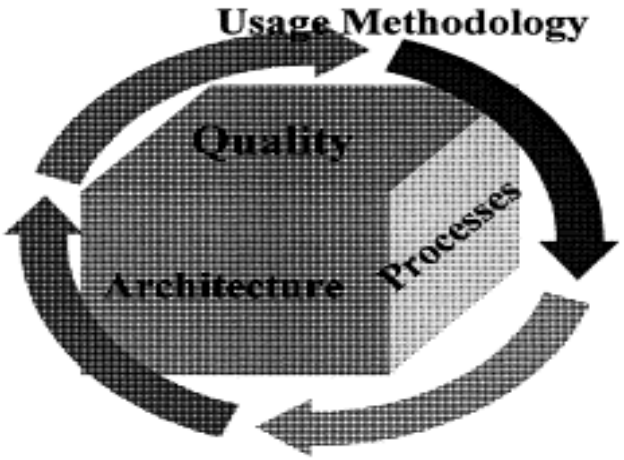
Choose the dimension that will apply to each and every fact table record. Typical dimensions are time, item, customer, supplier, warehouse, transactions type, and status.

Choose the measures that will populate each fact table record. Typical measures are numeric additive quantities like dollars_sold and units_sold.

Because the process of construction of data warehouse is a quite difficult and long-term task, its implementation scope should be clearly defined. The goals of an fundamental data warehouse implementation should be specific, achievable and measurable. This involves determining the time and budget allocations, the subset of the organization that is to be served. So, once a data warehouse is designed and constructed, the fundamental deployment of the warehouse includes the initial installations, roll – out planning, training, and orientations. And platform upgrades and maintenance must also be considered. So, the data warehouse administration includes data refreshment, data source synchronization, planning for disaster recovery, managing access control and security, managing data growth, managing data base performances and of course data warehouse enhancement and extension.

Data warehouse development tools provide functions to define and edit metadata repository contents (i.e. schemas, scripts, or rules), answer queries, output reports, and ship metadata to and from relational database system catalogs. Planning and analysis tools study the impact of schema changes and of refresh performance when changing refresh rates or time windows.

[C] DATA WAREHOUSE USAGE FOR INFORMATION PROCESSING



The proposed Meta model of data warehouse operational processes is capable of modeling complex activities, their interrelationships, and the relationship of activities with data sources and execution details. Moreover, the Meta model complements the existing architecture and quality models in a coherent fashion, resulting in a full framework for quality-oriented data warehouse management, capable of supporting the design, administration and especially evolution of a data warehouse. Data warehouse and data marts are used in a wide range of applications. Business executive use the data warehouses in data warehouses and data marts to perform data analysis and makes strategic decisions. In many firms, data warehouses are used as an integral part of a plan-execute-access" Closed-loop" feedback system for enterprise management. Data warehouses are used extensively in banking and financial services, consumer goods and retail distribution sectors, and controlled manufacturing such as demand-based production. Now, typically the longer a data warehouse has been in such a use, the more it will have evolved. This evolution should take place throughout a number of phases. Initially, the data warehouse is mainly used for generating reports and answering the predefined queries. Progressively, it is used to analyze, summarized and detailed data, where the results are presented in the form of reports and charts, later, the data warehouse is used for strategic purposes, performing multidimensional analysis and sophisticated slice-and-dice operations. So, at that stage we finally we reach the data warehouse may be employed for knowledge discovery and strategic decision making using data mining tools. In this context, the tools for data warehousing can be categorized into access and retrieval tools, database reporting tools, data analysis tools, and data mining tools. There are total three kinds of data warehousing applications: Information processing, Analytical processing, and data mining.

- ✓ **Information Processing** supports querying, basic statistical querying, basic statistical analysis, and reporting using cross tabs, tables, charts or graphs. A current trend in data warehouse information processing is to construct low-cost web-based accessing tools that are then integrated with web browsers.

- ✓ **Analytical Processing** supports basic OLAP operations, including slice-and-dice, drill-down, roll-up, and pivoting. It generally operates on historic data in both summarized and detailed forms. The major strength of online analytical processing over information processing is the multidimensional data analysis of data warehouse data.

- ✓ **Data Mining** supports knowledge discovery by finding hidden pattern and association constructing analytical models, performing classification and prediction, and presenting the mining results using visualizations tools.

  So these are the three various data warehouse applications which will help to design and use of data warehouse.

[D] FROM ONLINE ANALYTICAL PROCESSING TO MULITDIMENSIONAL DATA MINING

Multidimensional data mining integrates OLAP with data mining to uncover knowledge in multidimensional databases. Among the many different paradigms and architectures of data mining systems, multidimensional data mining is particularly important for the various reasons which are as follows:

- ✓ **High Quality of data in data warehouse:** Most data mining tools need to work on integrated, consistent, and cleansed data, which requires costly data cleaning, data integration and data transformation as preprocessing steps. A data warehouse constructed by such preprocessing steps. While a data warehousing constructed by such preprocessing serves as a valuable source of high-quality data for OLAP as well as for data mining. Now, we notice that data mining may serves as a valuable tool for data cleaning and data integration as well.

- ✓ **Available information processing infrastructure surrounding data warehouses:** Comprehensive information processing and data analysis infrastructures have been or will be systematically constructed surrounding data warehouses, which includes the accessing, integration, consolidation and transformation of multiple heterogeneous databases and OLAP analytical tools. It is prudent to make best use of the available infrastructure rather than constructing everything from scratch.

- ✓ **OLAP-Based exploration of multidimensional data:** Effective data mining needs exploratory data analysis. A user will often want to traverse through a database, select portions of relevant data, analyze them at different granularities, and present knowledge in different forms. Multidimensional data mining provides facilities of pivoting filtering, dicing, and slicing on a data cube and intermediate data mining results.

- ✓ **Online Selection of data mining functions:** Users may not always know the specific kinds of knowledge they want to mine. By integrating OLAP with various data mining functions, multidimensional data mining provides users with the flexibility to select desired data mining functions and swap data mining tasks dynamically.

So, these are the various multidimensional data mining resources for the data warehouse usage and designing. The data warehouse is concentrated on only few aspects. Here we are discussing about the data warehouse design and usage. Let's look at various approaches to the data ware house design and usage process and the steps involved. So, at the end of the research we are clearly said that data warehouse can be built using a top-down approach, bottom – down approach or a combination of both. In this research paper we are discussing about the data warehouse design process.

II RESEARCH RESULT

The paper is based on the literature research. The intention is to provide an overview over the current state of the art and use that as a base for presenting a data warehouse design and its usage and planning framework that emphasis the data warehouse specific needs. As we have seen, the introduction to data warehousing design and usage technology presented in this research paper is essential to our study of data warehousing. We are discussing about business analysis framework for data warehouse design, data warehouse design process, data warehouse usage for information processing and it is from OLAP's Multidimensional data mining. The idea of data warehousing is deceptively very simple. It is very much important to prepare data warehouse by using the proper design methodology and process. This is because data warehousing provides users with large amounts of clean, organized, and summarized data. Which greatly facilitates data mining. Suppose rather than storing the details of each sales transaction, a data warehouse may store a summary of the transactions per item type for each branch or summarized to higher level of summarized data in a data warehouse sets a solid foundation for successful data mining. Fundamentally, data is never deleted from data warehouses and updates are normally carried out when data warehouses are offline. This means that data warehouses can be essentially viewed as read-only databases. This satisfies the users' need for a short analysis query response time and has other important effects. First, it affects data warehouse–specific database management system (DBMS) technologies, because there is no need for advanced transaction management techniques required by operational applications. Second, data warehouses operate in read-only mode, so data warehouse–specific logical design solutions are completely different from those used for operational databases. For instance, the most obvious feature of data warehouse relational implementations is that table normalization can be given up to partially denormalize tables and improve performance. Other differences between operational databases and data warehouses are connected with query types. Operational queries execute transactions that generally read/write a small number of tuples from/too many tables connected by simple relations. For example, this applies if you search the data of a customer in order to insert a new costumer order. So, these kinds of queries are called an OLTP query. A data warehouse constructed by such preprocessing steps. While a data warehousing constructed by such preprocessing serves as a valuable source of high-quality data for OLAP as well as for data mining. So, the as per our research methodology data warehouse design and usage is very important but a little complex task.

III. CONCLUSION

Creating and managing a warehousing system is hard. Many different classes of tools are available to facilitate different aspects of the process described in Section 2. Development tools are used to design and edit schemas, views, scripts, rules, queries, and reports. Planning and analysis tools are used for what-if scenarios such as understanding the impact of schema changes or refresh rates, and for doing capacity.

ACKNOWLEDGMENT

The success of our research is never limited to the individual undertaking the paper. It is the cooperative effort of the people around an individual that spell success. For all efforts, behind this successful research, we are highly intended to all those personalities without whom this research would ever be completed. We find no words to express our gratitude towards those who were constantly involved with us throughout our work.

REFERENCES

[1] Inmon, W.H., *Building the Data Warehouse*. John Wiley, 1992.
[2] http://www.olapcouncil.org
[3] Codd, E.F., S.B. Codd, C.T. Salley, "Providing OLAP (On-Line Analytical Processing) to User Analyst: An IT Mandate."
Available from Arbor Software's web site http://www.arborsoft.com/OLAP.html.
[4] http://pwp.starnetinc.com/larryg/articles.html
[5] Kimball, R. *The Data Warehouse Toolkit*. John Wiley, 1996.
[6] Barclay, T., R. Barnes, J. Gray, P. Sundaresan, "Loading Databases using Dataflow Parallelism." *SIGMOD Record*, Vol. 23, No. 4, Dec.1994.
[7] Blakeley, J.A., N. Coburn, P. Larson. "Updating Derived Relations: Detecting Irrelevant and Autonomously Computable Updates." *ACM TODS*, Vol.4, No. 3, 1989.
[8]Gupta, A., I.S. Mumick, "Maintenance of Materialized Views: Problems, Techniques, and Applications." *Data Eng. Bulletin*, Vol. 18, No. 2, June 1995. 9 Zhuge, Y., H. Garcia-Molina, J. Hammer, J. Widom, "View Maintenance in a Warehousing Environment, *Proc. Of SIGMOD Conf.*, 1995.
[10] Roussopoulos, N., et al., "The Maryland ADMS Project: Views R Us*." Data Eng. Bulletin*, Vol. 18, No.2, June 1995.
[11] O'Neil P., Quass D. "Improved Query Performance with Variant Indices", To appear in *Proc. of SIGMOD Conf.*, 1997.
[12] O'Neil P., Graefe G. "Multi-Table Joins through Bitmapped Join Indices" *SIGMOD Record*, Sep 1995.
[13] Harinarayan V., Rajaraman A., Ullman J.D. " Implementing Data Cubes Efficiently" *Proc. of SIGMOD Conf.*, 1996.
[14] Chaudhuri S., Krishnamurthy R., Potamianos S., Shim K. "Optimizing Queries with Materialized Views" *Intl.Conference on Data Engineering*, 1995.
[15] Levy A., Mendelzon A., Sagiv Y. "Answering Queries Using Views" *Proc. of PODS*, 1995. 16 Yang H.Z., Larson P.A. "Query Transformations for PSJ Queries", *Proc. of VLDB*, 1987.
[17] Kim W. "On Optimizing a SQL-like Nested Query" *ACM TODS*, Sep 1982.
[18] Ganski,R., Wong H.K.T., "Optimization of Nested SQL Queries Revisited " *Proc. of SIGMOD Conf*., 1987.
[19] Dayal, U., "Of Nests and Trees: A Unified Approach to Processing Queries that Contain Nested Subqueries, Aggregates and Quantifiers" *Proc. VLDB Conf.*, 1987. 20 Murlaikrishna, "Improved Unnesting Algorithms for Join Aggregate SQL Queries" *Proc. VLDB Conf.*, 1992.
[21] Seshadri P., Pirahesh H., Leung T. "Complex Query Decorrelation" *Intl. Conference on Data Engineering*, 1996.
[22] Mumick I.S., Pirahesh H. "Implementation of Magic Sets in Starburst" *Proc.of SIGMOD Conf.*, 1994.Authors

First Author – Mr. Dishek Mankad, M.C.A., B.R.Patel Institute of Computer Application[MCA Progrm], Dahemi-Anand , dishek.mankad@gmail.com.

Second Author – Mr.Preyash Dholakia, M.C.A.,K.S.K.V Kachchh University,Bhuj and preyash_dholakia@yahoo.com.

Correspondence Author – Mr. Dishek Mankad, M.C.A., B.R.Patel Institute of Computer Application[MCA Progrm], Dahemi-Anand , dishek.mankad@gmail.com, 9712998513