The Study on Data warehouse Design and Usage

Seema T. Gopale[1]
MCA (IMCOST), University of Mumbai
ASM Institute of Management & Computer Studies
Thane, Mumbai – India

Snehal J. Patil[2]
MCA (IMCOST), University of Mumbai
ASM Institute of Management & Computer Studies
Thane, Mumbai – India

Abstract: This study provides a short survey of the field of Dataware house. Data ware housing is booming industry with many interesting research problem. Here we discussing about the data warehouse design and usage. It also presents the various approaches to the data warehouse design and usage process and steps involved. In this research paper we discussing about data ware house design process.

Keywords: Analysis, Data warehouse, Data warehouse design, process.

I. INTRODUCTION

Let's we first seen about what is Data warehouse? It is central repositories integrated data from one or more disparate sources. They store current and historical data and used for creating analytical reports for knowledge workers throughout the enterprises. "What is the need of data warehouse? What's going into data warehouse design? How are data warehouse used? How do data warehousing and OLAP relate to data mining? " In this research paper we discussing about business analyst framework for data warehouse design, and design process, data warehouse development, data warehouse usage for information processing.

The concept of Data warehousing is very much simple. Data is information process or stored by computer from the application that support business process. There it can be validated, reformatted, recognized, restructured, summarized and supplemented with data from other sources. The resulting data warehouse becomes the main source of information for analysis, presentation through ad hoc report, report generation, portals and dashboard. Building data warehouse is difficult. Many early adopters found that it to be costly, time consuming, and resource intensive and risky. This is especially true for those who have tried to build data warehouse themselves without the help of experts.

II. RESEARCH ELABORATION

[A] BUSINESS ANALYSIS FRAMEWORK FOR DATA WAREHOUSE DESIGN

"What is business analysis framework?" Business analysis framework is real and /or conceptual structure which includes the usage of an assembly of knowledge, practical techniques and establish concepts , in view of quickly discovering critically analyzing and accurately getting the business requirements.

"What can business analyst gain from having a data warehouse?" The Business analyst get the information from the data warehouses to measure the performance and make critical adjustments in order to win over the business holder in the market. Having data warehouse offers the following advantages:

- Since a data ware house can gather information quickly and efficiently, it can enhance business productivity.

- A data warehouse provide us a consistent of customers and items, hence it help us to manage customer relationship.

- A data ware house also help in bringing down the cost by tracking trends, patterns over a long period in a consistent and reliable manner.

If you wanted to do design effective data warehouse you must know the business need and construct business analysis framework. The construction of a large complex information system can be view as the construction of large and complex building for which the owner architect and builder have different view. This view is as follows:

- **The top-down view**- this views allows selection of the relevant information needed for a data warehouse.

- **The data source view**-this view presents the information being captured, stored and managed by the operational system. Data sources often modeled by traditional modeling techniques such as the E-R model.

- **The data warehouse view**-this view includes the fact table and dimension table. It represents the information store inside the data warehouse and the information regarding source, date and time of origin.

- **The business query view**- It is the view of the data from the viewpoint of the end user

So, building and using data ware house is a complex task because it requires technology skill, business skill, programming management skills. Building a data warehouse involves understanding how system store manage their data, how to build extractor for transfer data.
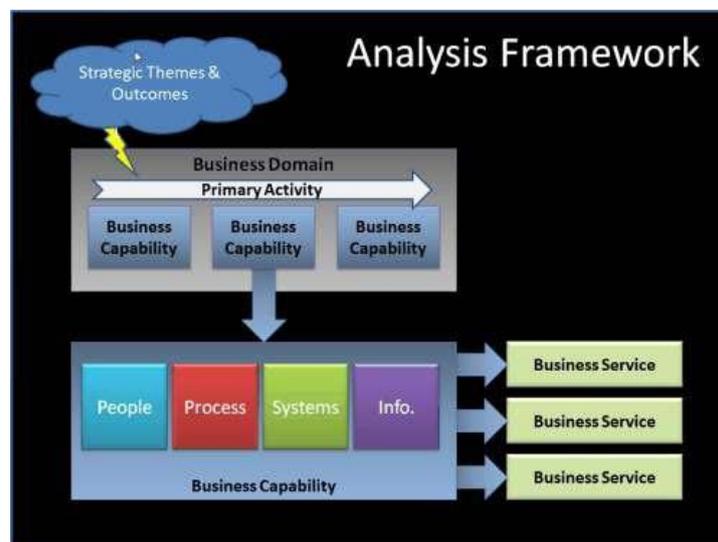

Fig. 1. Business Analysis framework

Business Analysis Framework (fig. 1.) provides the structure for business analyst, business process analysis, system analyst, risk analyst. It provides way to gather, communicate and develop the desired information required by business user, program management office, key stakeholders and technology developers.

Based on our experience, even for projects that are completed on time and on budget. There may be inefficiency in performing business analysis function. These inefficiencies includes following:

➢ Lost opportunities

➢ Rework

➢ No realization of benefits

By implementing framework you provide structure and standards that are intended to serve as support or provide guidance for your Business Analyst. Provide the ability to attract and retain experience and motivated BA's:

➢ Reduced waste

➢ Create solution

➢ Complete project on time

➢ Improve inefficiency

➢ Document the right things

A framework enables your organization to bring to market your competitive innovation more efficiently. This is increase in the delivery of successful projects.

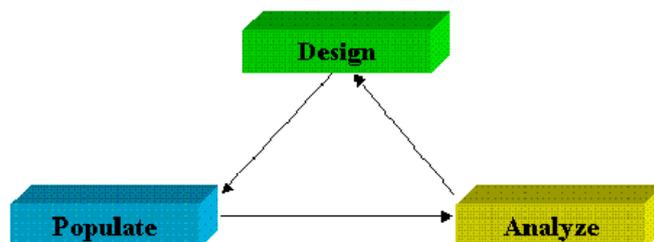*[B]. DATA WAREHOUSE DESIGN PROCESS*

Fig. 2. Data warehouse process

There are many ways to build data warehouses as there are companies to build them. Each data warehouse is unique because it must adapt to needs of business users in different functional areas. Here we discussed about various approaches to the data warehouse design processes. A data warehouse can built using top-down approaches, bottom-up approaches or combination of both.

The top-down approach starts with overall design and planning. The data warehouse holds the atomic or transaction data that is extracted from one or more source system and integrated within normalized, enterprise data model. It is useful in cases where the technology is mature and well known, and where the business problems that must be clear and solve and well understood. The bottom-up approach starts with experiments and prototypes. It consciously tries to minimize back-office operation, preferring to focus an organizations effort on developing dimensional designs that meet end user requirements. This useful to move forward at considerable less expenses and evaluate the technological advantage before making significant commitments. If we are thinking engineering point of view , the design and construction of data analysis , warehouse design, data integration and testing and finally deployment of the data warehouse.

Large software system developed by using one of the two technologies. The waterfall method and spiral method. The waterfall method is sequential design process used in software development process, in which progress is seen as flowing steadily downwards through the phases of conception, initiation, analysis, design, construction, testing and maintenance. The spiral method involves rapid generation of increasingly functional system, with shot interval between the successive releases. This is always good choice for data warehouse deployment, especially for data marts because the turn around time is short, modification can done quickly an new design for the technology can adopted in timely manner. So we here discuss about design process;

This includes various steps as follows:

Choose Business process model: It is first step in the model is to describe the business process which the model build on. It involves complex object collection; a data warehouse should be followed.

Choose the business process grain: the grain of the model **is** the exact description what model should be focusing on. You should pick the central process and describe in one sentence.

Choose the dimension: Dimension are the foundation of fact table and is where the data for the fact table collected.

Choose the measures: It is closely related to business user of the system, since this where the get access data stored in the data warehouse. Therefore most of the fact table rows are numerical, additive figures such as quantity or cost per unit etc.

The goals of a fundamental data warehouse implementation should be specific, achievable and measurable. So the data warehouse administration include data refreshment, data source synchronization, planning for disaster recovery, managing access control and security, management data growth, managing database performance and of course data warehouse enhancement and extension. Data warehouse development tools provide functions to define and edit metadata repository contents, answers queries, ship Meta data and output reports to and from relational database system catalogs.

[C]. DATA WAREHOUSE DEVELOPMENT:
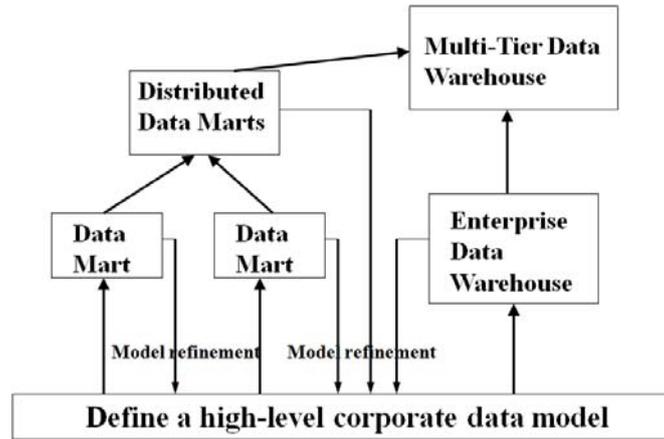
A RECOMMENDED APPROACH

Fig. 3. Data warehouse Development

A data model is graphical view of data created for analysis and design purposes. Data modeling includes designing data warehouses in detail. It include following fig.3:

Enterprise Data warehouse: It collects all the information about the subjects spanning the entire organization.

Data Mart: It is a subset of corporate-wide data that is of value to a specific group of users. Its scope id confined to specific, selected groups, such as marketing data mart.

Virtual warehouse: A set of views over operational databases. Only some of the possible summery views may be materialized.

Above diagram discussed various data model which build data warehouse incrementally. It starts with data models and then builds each data mart in the organization in parallel and then integrated the all dart marts.

[D]. DATA WAREHOUSE USAGE FOR INFORMATION PROCESSS
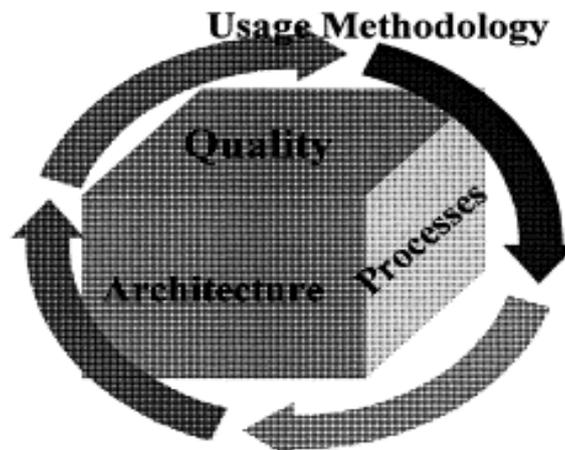

Fig.4. Information Process

The proposed Meta model (fig.4) 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:** A data warehouse allows processing the data store in it. It 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. 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.

[E].FROM ONLINE ANALYTICAL PROCESSING TO MULTI DIMENSIONAL DATA MINING

Among the many different paradigm and architecture of data mining systems multidimensional data mining is particularly important for the various reasons which are follows:

High quality of data in data warehouse: A data warehouse constructed by preprocessing steps while data warehousing constructed by such preprocessing serves as a valuable source of high- quality data for OLAP as well as data mining. Most data mining tool need to work on integrated, consistent and cleansing data transformation as a preprocessing steps.

Available information processing infrastructure surrounding data warehouse includes the accessing, integration, consolidation and transformation of multiple heterogeneous databases and OLAP analytical tools

OLAP based exploration of multidimensional data a user will often want to traverse through a database, select portion of relevant data, and analyze them at different granularities and present knowledge in different forms. Effective data mining needs exploratory data analysis.

Online selection of data mining function: BY integrating OLAP with various data mining functions, multidimensional data mining provides facilities of pivoting filtering, slicing and dicing on data cube and data mining result.

So, there are the various multi dimensional data mining resources for the data warehouse usage and designing.

III. RESEARCH RESULT

The paper is based on literature research. The intention is to provide an overview over the current state of the art and use that as base for presenting the data warehouse design and its usage. The idea of data warehousing is deceptively very simple. It is more important to prepare data warehouse by using proper methodology and process. We are discussing about business analysis framework for data warehouse design, data warehouse design process, data warehouse development, data warehouse usage for information process and it is from OLAP to multidimensional data mining. Data warehousing provides users with large amount of clean, organized and summarized data. Which helps in data mining? Suppose rather than storing details of each transaction, a data warehouse may stores summery of the transaction per item type for each branch or summarized to higher level of summarized data in data warehouse.

IV. CONCLUSION

Designing and managing data warehouse system is difficult. Many different tools are available to facilitate different aspect of the process which describe above. Development tools are used to design and edit schemas, scripts, views, rules, queries and reports.

The key to data warehousing is data design. The business users grasp what information they have and the way they require to use it. Target the users, verify what information is required, find sources for the info, and organize the info in a very dimensional model that represents the business desires. The remaining tasks flow naturally from a well-designed model— extracting, transforming, and loading the data into the data warehouse, making the OLAP and data mining analytical applications, developing or acquiring end-user tools, deploying the system, and tuning the system design as users gain experience.

ACKNOWLEDGEMENT

References

1. www.google.in
2. Inmon, W.H., Building the Data Warehouse. John Wiley, 1992.
3. http://www.olapcouncil.org
4. 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.
5. http://pwp.starnetinc.com/larryg/articles.html
6. Kimball, R. The Data Warehouse Toolkit. John Wiley, 1996.
7. Barclay, T., R. Barnes, J. Gray, P. Sundaresan, "Loading Databases using Dataflow Parallelism." SIGMOD Record, Vol. 23, No. 4, Dec.1994.
8. Blakeley, J.A., N. Coburn, P. Larson. "Updating Derived Relations: Detecting Irrelevant and Autonomously Computable Updates." ACM TODS, Vol.4, No. 3, 1989.
9. Gupta, A., I.S. Mumick, "Maintenance of Materialized Views: Problems, Techniques, and Applications." Data Eng. Bulletin, Vol. 18, No. 2, June 1995.
10. Zhuge, Y., H. Garcia-Molina, J. Hammer, J. Widom, "View Maintenance in a Warehousing Environment, Proc. Of SIGMOD Conf., 1995.
11. Roussopoulos, N., et al., "The Maryland ADMS Project: Views R Us." Data Eng. Bulletin, Vol. 18, No.2, June 1995.
12. O'Neil P., Quass D. "Improved Query Performance with Variant Indices", To appear in Proc. of SIGMOD Conf., 1997.
13. 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.