A Data Warehouse Design and Usage

A G P Kujur [1], Ajay Oraon [2]

[1] Assistant Professor, Department of Computer Sc and Engg, BIT Sindri, Jharkhand,India [1]
[2]Assistant Professor, Department of Chemical Engg, BIT Sindri, Jharkhand,India [2]

---***---

Abstract - *A Data Warehousing a subject oriented, integrated, time variant and nonvolatile data Collection organized in support of management decision making. Several factors distinguish data warehouse from operational database. The two system s provide quit different and require different   kind of data, it is necessary to maintain data warehouse separately from operational database. A data warehouse contain backend tool and utilities. These cover data extraction data cleaning, data transformation , Loading, refreshing and warehouse management. The paper considers What goes into a data warehouse design. How are data warehouse used .How do data warehouse and OLAP relate to data mining . how confident organizations have used them to increase control over data and decision making. This reveals that organization that can develop a strong system, data warehousing is value the cost. A physical repository where relational data are specially organized to provide enterprise, cleansed data in a standardized format*

Key Words: Data Warehouse, DBMS, Data Mining, Information System

1.INTRODUCTION

Data warehouse provides architectures and tools for business executive   to   schematically organize, understand and use their data to make strategic decision .Data warehouse generalize and consolidate data in multi dimensional space, The construction of data warehouse involves data cleaning, data integration and data transformation warehouses provide online analytical processing (OLAP) tools for the interactive analysis of multi dimensional data, which facilities effective data generalization and data mining Many other data mining function such as association ,classification , predict tion and clustering can be integrated with OLAP .Therefore, data warehousing and OLAP form an essential step in the knowledge discovery process. Major features of data warehouse are subject oriented , integrated, time variant and non volatile.

Subject oriented – A data warehouse is organized around major subjects such customer, sales, manufacturing etc. Rather than concentrating on the day-to-day operation and transaction processing of an organization .A data warehouse focuses on the modeling and analysis of data for decision maker hence data warehouse typically provide a simple and concise view of particular subject issue by excluding data that are not useful in the decision support process.

Integrated- A data warehouse is usually constructed by integrating multiple heterogeneous sources such as relational database system, flat file and online transaction records . Data cleaning and data integration techniques are applied to ensure consistency in naming conventions , encoding structure , attributes and so on.

Time variant – Data are stored to provide information from historic perpespective.Every key structure in the data warehouse contains either implicitly or explicitly, a element.

Non volatile – A data warehouse is always a physically separate store of data transformed from the application data found in the operational environment

III) Data Warehouse Models

From the Architecture point of view. There are three data warehouse model.

i) Enterprise Data Warehouse: An Enterprise data warehouse collects all the information about subject spanning the entire organization .It provides corporate wide data integration, usually from one or more operational system or external information providers and is cross functional in scope. It typically contains detailed data as well as summarized data and can range in size few gigabytes to hundred gigabytes or terabytes .An Enterprise data warehouse implemented on traditional mainframe, super computer and super server or parallel architecture platforms .It requires extensive business modeling and may take years to design and build.

ii) Data Mart: A data mart contains a subset of corporate wide data that is of value to a specific group

of users. The scope of confined to specific selected subject For example a marketing data mart may confine its subject to customer, item and sales. Data mart are usually implemented on low cost departmental server .The implementation cycle of data mart is more likely to be measured in weeks rather than months or year.

iii) Virtual warehouse: A virtual warehouse is a set of views over operational databases For efficient query processing only some of the possible summary views may be metalized. A virtual warehouse is easy to build but requires excess capacity on operational database server.

IV) Data warehouse :A Multi tiered Architecture

Data ware house adopt three tier architecture .Warehouse database that is almost always a relational database system like oracle, DB2, Sql Server etc.

Bottom Tier – Back end tools utilities are used to feed data into the bottom tier from operational databases or other external source. These tools and utilities perform data extraction , cleaning and transformation to merge similar data from different source into unified format as well as load and refresh function to update the data warehouse . The data extracted using application program interfaces known as gateways .A gateways is supported by the underlying database and allow client program to generate sql code to be executed at server.

Middle Tier – is an online transaction analysis server that is typically implementation using either (1),(2) or (3).

(1) a relational OLAO (ROLAP) Server – These are the intermediate servers that stand in between a relational back end and front end tools. They use of relational or external relational database to store and manage warehouse data. OLAP middle ware to support missing pieces. ROLAP server include optimization for each database back end .implementation of aggregation, navigation, logic and additional tools and services.

(2) a multi dimensional OLAP (MOLAP) Server – These server support multidimensional data views through array based multidimensional storage engines .They map multidimensional views directly to data cube array structure .The advantages of using data cube is that it allows fast indexing to pre computed summarized data.

(3) a Hybrid OLAP (HOLAP) Server – The Hybrid OLAP approach combines ROLAP and MOLAP technology, benefiting from the grater scalability of ROLAP and the faster computation of MOLAP.
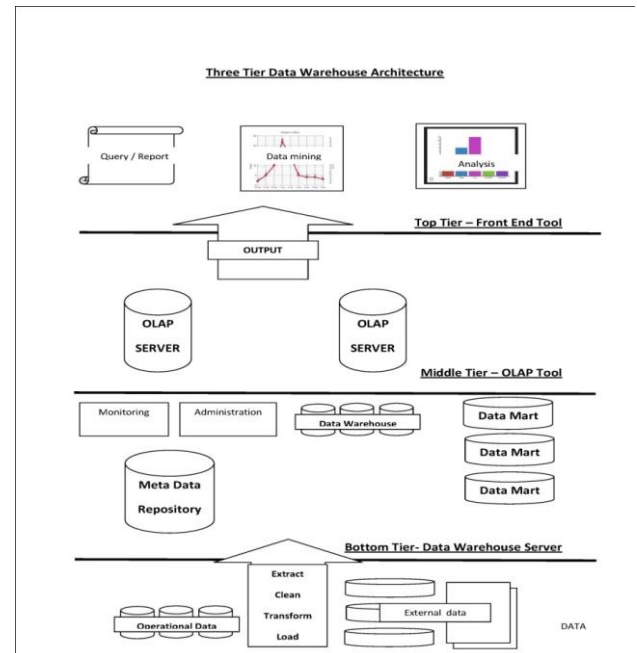
Top tier – is front end client layer which contain query and reporting tools , analysis tools, and data mining tools

V) Recommended method for the development of data warehouse system

A recommended method for the development of data warehouse system is to implement the warehouse in an incremental and evolutionary manner. As Shown in Fig. First of All A high level corporate data model is defined within a reasonably short period that provides a corporate wide, constant, integrated view of data among different subjects and potential usage .The high level model although it will need to be refined in the further development of enterprise data warehouse and departmental data mart will greatly reduce future integration problem. Independent data mart can be implemented in parallel with the enterprise ware house based on the same corporate data model .Distributed data mart can be constructed to integrate different data mart via hub server.
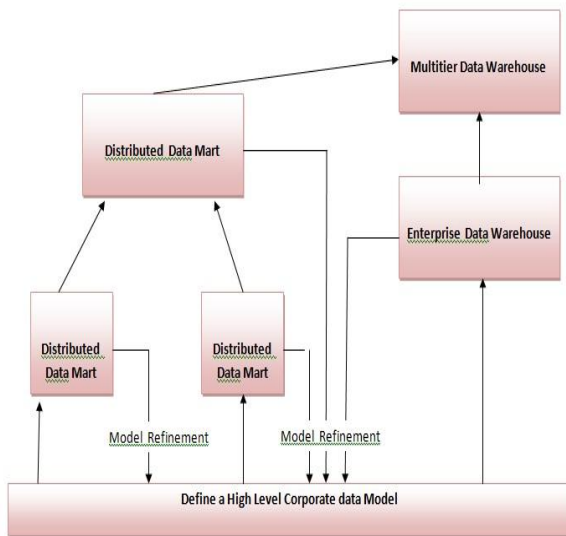
Fig : A Recommended method for the development of data warehouse system

VI) Conclusions

This paper introduced a model for building a data warehouse for a typical information system. The warehouse is an informational database whose data are extracted from an already existing operational database. The purpose of the proposed design method is to help decision makers and principles in performing data mining and data analysis over the data stored in the warehouse which eventually helps them in discovering critical patterns and trends. Data warehouse architecture provides a useful way of determining if the organization is moving toward a reasonable data warehousing framework.

VII) REFERENCES

[1] Pant, S., Hsu, C., "Strategic Information Systems Planning: A Review", Information Resources Management Association International Conference, May 21–24, 1995, Atlanta.

[2] Xingquan Zhu, Ian Davidson, "Knowledge Discovery and Data Mining: Challenges and Realities", Hershey, New York, 2007.

[3] Preeti S., Srikantha R., Suryakant P., "Optimization of Data Warehousing System: Simplification in Reporting and Analysis", International Journal of Computer Applications, vol. 9 no. 6, pp. 33–37, 2011.

[4] Kantardzic, Mehmed, "Data Mining: Concepts, Models, Methods, and Algorithms", John Wiley & Sons, 2003. [5] William Inmon, "Building the Operational Data Store", 2nd ed., John Wiley & Sons, 1999.

[6] Matteo Golfarelli & Stefano Rizzi, "Data Warehouse Design: Modern Principles and Methodologies", McGraw-Hill Osborne Media, 2009.

[7] Pang-Ning Tan, Michael Steinbach, Vipin Kumar, "Introduction to Data Mining", Addison Wesley, 2005.

[8]Microsoft Access Home Page, http://office.microsoft.com/en-us/access/default.aspx

[9] Jeffrey A. Hoffer, Mary Prescott, Heikki Topi, "Modern Database Management", 9th ed, Prentice Hall, 2008.

[10] Robert Laberge, "The Data Warehouse Mentor: Practical Data Warehouse and Business Intelligence Insights", McGraw-Hill Osborne Media, 2011.

[11] Anahory, S., D. Murray, "Data Warehousing in the Real World: A Practical Guide for Building Decision Support Systems", Addison Wesley Professional. 1997.

[12] Michael Goebel, Le Gruenwald, "A Survey of Data Mining and Knowledge Discovery Software Tools", Proceedings of the ACM SIGKDD, vol. 1, no. 1, pp. 20-33, 1999.