THE STUDY ON DATA WAREHOUSING DIFFERENT CONCEPTS

[1]Praveen Kumar, [2]Dr. Kavita

[1]Research Scholar, [2]Research Guide, Department of Computer Science
Jayoti Vidyapeeth Women's University, Jaipur, Rajasthan, India

ABSTRACT - A data warehouse is a repository for all the data which is composed by an organization in various functioning systems; it can be either physical or logical. It is an integrated nonvolatile and time variant collection of data which helps in making decisions for organization success. The data warehouse is focused on very few aspects.

In this paper we are discussing data warehouse design and its usage. Let's look at different approaches to the data warehouse design and its usage process and all the steps involved. Data warehouse can be built using a bottom up approach, top down approach or a combination of both.

INTRODUCTION

Information assets are greatly valuable to any enterprise and because of this these assets must be properly stored and readily accessible when they are needed.Think of a data warehouse as a central storage system which collects data from many different and delivers it to many audiences usually to meet decision support and company intelligence requirements.

Some Areas of Application for Which Data Warehouse Technologies Are UsedSuccessfully:
- Risk analysis, Financial services and credit cards fraud detection.
- Vehicle management under Transport industry.
- Production cost controlsupplier and order support.
- Various health care service Patient discharge and admission analysis and bookkeeping in accounts departments.

What is the need of data warehouse? How are data warehouse used?What goes into a data warehouse design? In this paper we are focusing about company analysis framework for data warehouse design process, data warehouse usage for information processing and from OLAP to multidimensional data mining. The concept of data warehousing is simple, data is extracted at times from the applications that support company processes and copied on dedicated systems. There it can be reorganized, validated, supplemented and restructured with data from other sources. Finally, the data warehouse becomes the main source of information for

further analysis and presentation through multiple reports portals and dashboards. Building data warehouses used to be difficult andmany early users found it to be very costly and time consuming and resource intensive. From several years it has earned a reputation for being risky. This is true only for those who have built data warehouses themselves without the help of real experts.

A Company Analysis Framework for Data Warehouse Design:

- First, having a data warehouse can provide an advantage by presenting accurate information from whichanalyst can measure performance and make critical decisions to help win over competitors.
- Second, a data warehouse can enhance company productivity because it is able to quickly and efficient gatherinformation that accurately describes the organization.
- Third, a data warehouse strengths customer relationship as it provides a consistent view of items and customersacross all departments and markets. Finally, a data warehouse can bring costreduction by tracking different trends patterns and exceptions over long periods.

If you want to do design effective data warehouse you must know the company needs and construct a companyanalysis framework. The development of large and complex information system can be imagined as the construction ofa large and complex building for which the owner architect and builder have different views. This view will showcase a complex framework that represents the top down approach as well as thebottom up. Different views regarding a datawarehouse design must be considered like top down view, the data source view, the data warehouse view of theinformation system.

The top down view permits the selection of the related information necessary for the data warehouse. Thisinformation matches current and future company needs.The Data source view publish the information being managed and stored by operational system. Thisinformation can be documented at various places of detail and accuracy from individual data source tables tointegrate at the various levels.

Data sources are modeled by traditional data modeling such as the ER model.The Data warehouse view includes dimension and fact tables.

The Company Query View is the data perspective in the data warehouse form the enduser's viewpoint.So, building and using a data warehouse is a complex task because it requires company skill technology skills and programmanagement skills. Regarding company skills building a data warehouse involves understanding how systems store andmanage their data

how to build extractors that transfer data from the operational system to the data ware house and how tobuild warehouse refresh software that keeps the data warehouse reasonably up to date with the operational system's data.

Data Warehouse Design Process:
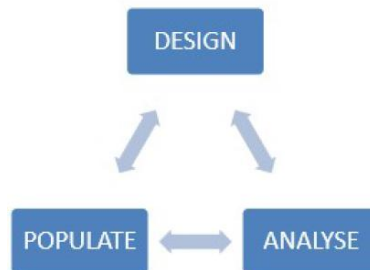
Here we discussed about several approaches of the data warehouse design process. A datawarehouse can be constructed by using top down approach or bottom up approach or a combination of both. The top downapproach begins with design and planning. It is useful in cases where the technology is mature and well known andwhere the company problems that must be solved are clear and well understood. The bottom up approach begins withexperiments and with different prototypes. This is very much useful in the beginning stage of company modeling and technology development. And italso allowed an organization to move forward at considerable fewer expenses and evaluate the technological advantagesbefore making significant commitments. In the combined approach organization can be exploit the strategic and planned nature of the top down approach and retain the opportunistic application of thebottom up approach.

If we think from the software engineering point of view the construction and design of data warehouse, data integrationand finally deployment of the data warehouse, softwaresystems can be developed by using any one of the two technologies,spiral method or waterfall method.The waterfall method performs systematic analysis at every step before moving to the next stage which islike water falling form one step to the next. The spiral method consists of rapid generation of systems with short intervals in between different successive releases. This is always been considered a good choice for datawarehouse development mainly for data marts because the turnaround time is short modifications can be done quicklyand new designs for the technologies and that can be adapted in a timely manner.

Data Warehouse Usage for Information Processing:

This model of data warehouse methodscanshow complex activities with their relationship of activities with data sources and execution details. Moreover, this model matcheswith the existing architecture and quality models in a coherent fashion which results in a full framework. Data marts and warehouses are used in a large range of applications. Company executive use the data warehouses in data warehouses and data marts to perform data analysis and makes strategic decisions. In many organizations, data warehouses are used as an important part of a system for enterprise managements. Data warehouses are used mainly in financial services like banking and retail distribution sectors and manufacturing units such as demand based production.

Now generally, the longer a datawarehouse is being used in such scenarios the more it will have evolved. This evolution should continue throughout several phases. Initially data warehouse is mainly used for providing customized reports and answering the user queries.

Progressively it was used to analyze detailed &summarized data where results are presented in the form of charts or reports, later the data warehouse is used for strategic purposes performing severaldata analysis and multiple slice and dice operations. So, at this stage data warehouse may be employed for strategic decision making and knowledge discovery and using data mining tools. In this context, data warehousing can be categorized into access and retrieval tools like reporting tools, different data analysis and data mining tools. There are total three kinds of data warehousing applications data mining, Information processing and Analytical processing.

Information Processing - supports finding of statistical dataon the basis of statistical analysis and reporting using charts or graphs. Recent trend in

data warehouse information processing is to produceweb based tools with low cost which can be integrated with web browsers.

Analytical Processing - supports OLAP operations like slice and dice, rollup and pivoting. It mostly operates on historical data in both detailed and summarized forms. The major usage of online analytical processing as compare to information processing is the multidimensional analysis of data.

Data Mining - supports knowledge discovery by querying hidden pattern and relation between the data and constructing analytical models and prediction and presenting the results using various visualizations tools.

RESEARCH RESULT

The paper is based on the literature research. The idea is to give an overview over the current state anduse that model as base for presenting a data warehouse concepts. As we have seen in the introduction part, data warehousing concepts presentedin this research paper is essential to any study of data warehousing. We havediscussed about company analysis frameworkfor data warehouse design process and it is fromOLAP's Multidimensional data mining. The idea of data warehousing is unreliably very simple. It is important to construct data warehouse by using the proper design principals and processes. This is because datawarehousing provides huge amounts of organized and summarized clean data to the users. Suppose rather than savingonly summary of each transaction, a data warehouse is storingevery detail of transaction, then it will built a solid foundation for data mining. Fundamentally data can never be deleted from data warehouses and updates arenormally performed when data warehouses systems are offline. This means that data warehouses can be viewed as readonly basis. This satisfies the overall users' need which he requires from a data warehouse system.

First, it affects data warehouse specific DBMS technologies because there is no need forany more advanced transaction management techniques which are required by operational applications.

Second, data warehouses mostly operate inread only mode so logical design solutions are completely different from those used foroperational databases. For instance, the most common feature of data warehouse relational database implementations which is tablenormalization can be dropped to partially renormalize tables which will help in improving performance.

Other differences between data warehouses and transactional databases are related with query types. Operational queries generally execute transactions that read/write few tuples from/too many tables which

are connected by simple relations. For example, if user search the data of a customer or insert a new costumer, these queries are called OLTP queries. A data warehouse is developed by using OLAP queries, this uses dimensions and cubes to store and query the data. So as per ourresearch methodology data warehouse design and usage is very important but a little complex task.

CONCLUSION

Creation and managementof a data warehousing system is quite complex. There are several different classes of tools available to simplifyvarious aspects of the process described in this paper. Various development tools can be used to design and edit views, schemas, scripts rules, queries and different reports. Planning and analysis tools are also used to design any new data warehouse projects. Before designing of a new data warehouse, we should have proper understanding of all its concept, usage, guidelines and should be aware about different design principals.

REFERENCES

- Inmon W.H. Building the Data Warehouse. John Wiley 1992.
- Kimball R. The Data Warehouse Toolkit. John Wiley 1996.
- Barclay T. R. Barnes J. Gray P. Sundaresan "Loading Databases using Dataflow Parallelism." SIGMOD Record Vol.23 No. 4 Dec.1994.
- Blakeley J.A. N. Coburn P. Larson. "Updating Derived Relations Detecting Irrelevant and Autonomously ComputableUpdates." ACM TODS Vol.4 No. 3 1989.
- Gupta A. I.S. Mumick "Maintenance of Materialized Views Problems Techniques and Applications." Data Eng.Bulletin Vol. 18 No. 2 June 1995.
- Zhuge Y. H. GarciaMolina J. Hammer J. Widom "View Maintenance in a Warehousing Environment Proc. OfSIGMOD Conf. 1995.
- Roussopoulos N. et al. "The Maryland ADMS Project Views R Us." Data Eng. Bulletin Vol. 18 No.2 June 1995.
- O'Neil P. Quass D. "Improved Query Performance with Variant Indices" To appear in Proc. of SIGMOD Conf.1997.
- O'Neil P. Graefe G. "MultiTable Joins through Bitmapped Join Indices" SIGMOD Record Sep 1995.
- Harinarayan V. Rajaraman A. Ullman J.D. " Implementing Data Cubes Efficiently" Proc. of SIGMOD Conf. 1996.
- Chaudhuri S. Krishnamurthy R. Potamianos S. Shim K. "Optimizing Queries with Materialized Views"Intl.Conference on Data Engineering 1995.
- Levy A. Mendelzon A. Sagiv Y. "Answering Queries Using Views" Proc. of PODS 1995.
- Yang H.Z. Larson P.A. "Query Transformations for PSJ Queries" Proc. of VLDB 1987.
- Kim W. "On Optimizing a SQLlike Nested Query" ACM TODS Sep 1982.
- GanskiR. Wong H.K.T. "Optimization of Nested SQL Queries Revisited " Proc. of SIGMOD Conf. 1987.
- Dayal U. "Of Nests and Trees A Unified Approach to Processing Queries that Contain Nested Subqueries Aggregatesand Quantifiers" Proc. VLDB Conf. 1987.

- Murlaikrishna "Improved Unnesting Algorithms for Join Aggregate SQL Queries" Proc. VLDB Conf. 1992.
- Seshadri P. Pirahesh H. Leung T. "Complex Query Decorrelation" Intl. Conference on Data Engineering 1996.