ISSN:0975-9646

Pushpinder Kaur / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (3) , 2016, 1184-1185

Data Warehouse

Pushpinder Kaur
Assistant Professor in Khalsa College Morinda

Abstract: Data warehouse are essential elements of decision support, which has increasingly become a focus of the database industry. 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.

Keywords: Data warehouse, Data warehouse design process

INTRODUCTION

Before we seen the design process, let's we seen about what is data warehouse? Data warehouse is a system used for data analysis and reporting. A data warehouse is constructed by integrating data from multiple heterogeneous sources. Data warehouses standardize the data across the organization so as to have a single view of information. Data warehouses can provide the information required by the decision makers. 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. However, we believe that a methodological framework for design is an essential requirement to ensure the success of complex projects. This can be easily confirmed by analyzing the statistic reports related to DWs project failures, which state that a major cause lies in the absence of a global view of the design process and, in other terms, in the absence of an organic methodology. In this direction, our research is aimed at defining the basic steps required for a correct DW design.

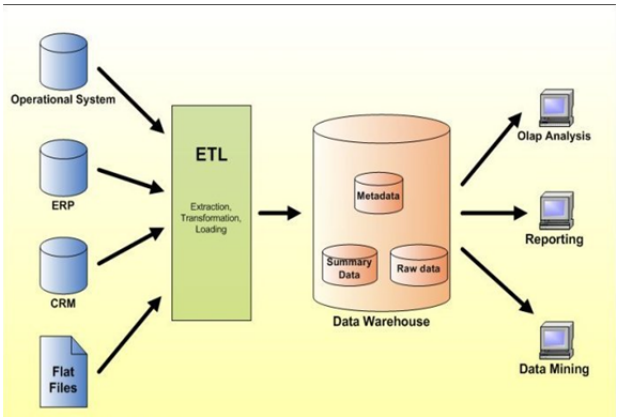

Figure 1.1: Data warehouse

Designing and rolling out a data warehouse is a complex process, consisting of the following activities:

- Define the architecture, do capacity planning, and select the storage servers, database and OLAP servers, and tools. Integrate the servers, storage, and client tools.
- Design the warehouse schema and views.
- Define the physical warehouse organization, data placement, partitioning, and access methods.
- Connect the sources using gateways, ODBC drivers, or other wrappers.
- Design and implement scripts for data extraction, cleaning, transformation, load, and refresh.
- Populate the repository with the schema and view definitions, scripts, and other metadata.
- Design and implement end-user applications.
- Roll out the warehouse and applications

1.1 Top down in Data warehouse: In the top-down design approach the, data warehouse is built first. The data marts are then created from the data warehouse. If you use a top-down approach, you will have to analyze global business needs, plan how to develop a data warehouse, design it, and implement it *as a whole*. This procedure is promising: it will achieve excellent results because it is based on a global picture of the goal to achieve, and in principle it ensures consistent, well integrated data warehouses.
Advantages of top-down design are:

- Provides consistent dimensional views of data across data marts, as all data marts are loaded from the data warehouse.
- This approach is robust against business changes. Creating a new data mart from the data warehouse is very easy.

1.2 Bottom up in Data warehouse: In the bottom-up design approach, the data marts are created first to provide reporting capability. A data mart addresses a single business area such as sales, Finance etc. These data marts are then integrated to build a complete data warehouse. The integration of data marts is implemented using data warehouse bus architecture. In the bus architecture, a dimension is shared between facts in two or more data marts. These dimensions are called conformed dimensions. These conformed dimensions are integrated from data marts and then data warehouse is built.
Advantages of bottom-up design are:

- This model contains consistent data marts and these data marts can be delivered quickly.
- As the data marts are created first, reports can be generated quickly.

- The data warehouse can be extended easily to accommodate new business units. It is just creating new data marts and then integrating with other data marts.

ANALYSIS OF THE INFORMATION SYSTEM

- Exploit the experience of the database manager in order to find out possible missing or abnormal data. In particular, correct handling of null values is an essential precondition for the quality of the query answers. A null value may mean that, at the time the measurement was taken, either the value did not exist or, conversely, that the measurement process failed to deliver the data [19]. This difference must be well understood by the designer when he plans the DW loading process.
- Select operational data sources by considering the data quality and the stability of their schemes. Other equivalent data sources should be determined as well, in order to allow the so-called *view* synchronization to take place.
- Determine which data can be usefully integrated in order to obtain a complete view of the database domain. Integration of heterogeneous views has been largely dealt with in the database literature.
- Deeply understand the data semantic in order to make cross-footing possible during the data staging process.

There are four major processes that contribute to a data warehouse:

- Extract and load the data: Data extraction takes data from the source systems. Data load takes the extracted data and loads it into the data warehouse.
- Cleaning and transforming the data: Once the data is extracted and loaded into the temporary data store, it is time to perform Cleaning and Transforming. Here is the list of steps involved in Cleaning and Transforming:
  - Clean and transform the loaded data into a structure
  - Partition the data
  - Aggregation

- Backup and archive the data: In order to recover the data in the event of data loss, software failure, or hardware failure, it is necessary to keep regular backups. Archiving involves removing the old data from the system in a format that allow it to be quickly restored whenever required.
- Managing queries and directing them to the appropriate data sources:

This process performs the following functions:

- Manages the queries.
- Helps speed up the execution time of queris.
- Directs the queries to their most effective data sources.
- Ensures that all the system sources are used in the most effective way.
- Monitors actual query profiles

CONCLUSION

In this paper we outlined a general methodological framework for DW design, discussing the different steps of top down and bottom up in data warehouse. The framework covers the conceptual, logical and physical steps. We have described the substantial technical challenges in developing and deploying decision support systems. While many commercial products and services exist, there are still several interesting avenues for research. We will only touch on a few of this here.While design and workload definitions have already been achieved on a set of sample applications producing satisfying results, we are currently studying effective algorithms for designing of data warehouse.

REFERENCES

[1] R. Agrawal, A. Gupta, and S. Sarawagi, "Modeling multidimensional databases", IBM Research Report, 1995.

[2] E. Baralis, S. Paraboschi, and E. Teniente, “Materialized view selection in multidimensional database", Proc. 23rd Very Large Database Conf. (VLDB97), Athens, Greece, 1997, pp. 156-165.

[3] J.W. Buzydlowski, I. Song and L. Hassel. "A Framework for Object-Oriented On-Line Analytic Processing", Proc. of the 1st Int. Workshop on Data Warehousing and OLAP (DOLAP’98), Washington D.C., November 1998.

[4] L. Cabibbo, and R. Torlone, "Querying Multidimensional databases", 6th Workshop on Database Programming Languages (DBPL’97), 1997.

[5] L. Cabibbo, and R. Torlone, "Un quadro metodologico per la costruzione e l' uso di un data warehouse", Proc. Sesto Convegno Nazionale sui Sistemi Evoluti per Basi di Dati, Ancona, Italy, 1998, pp. 123-140.

[6] Cabibbo, L., and Torlone, R., "A logical Approach to Multidimensional Databases". Lecture Notes in Computer Science, n. 1377, Proc. of the 6th Int. Conf. on Extending Database Technology, (EDBT’98), Valencia, March 1998, pp. 183-197.

[7] A.F. Cardenas, "Analysis and performance of inverted database structures", Comm. ACM, vol. 18, n. 5, pp. 253-263, 1975.

[8] C.D. French, " 'One Size Fits All' Database Architectures Do Not Work for DSS"P, roc. ACM SIGMOD Conf., 1995, pp. 449-450.

[9] M. Golfarelli, D. Maio, and S. Rizzi, "Conceptual design of data warehouses from E/R schemes", Proc. HICSS-31, VII, Kona, Hawaii, 1998, pp. 334-343.

[10] M. Golfarelli, D. Maio, and S. Rizzi, "The Dimensional Fact Model: a Conceptual Model for Data Warehouses" to appear on Int. Jour. Computer and Information Systems.