Proposed Local Data Mart Approach for Data Warehouse Architecture

Bhadresh Pandya\textsuperscript{1}, Dr. Sanjay Shah\textsuperscript{2}

\textsuperscript{1}Professor, Department of Computer Science, Kadi Sarva Vishwavidyalaya, Gandhinagar
\textsuperscript{2}Director, S. V Institute of Computer Studies, Kadi.

Abstract - Data warehouses are in use for more than a decade but companies are still finding it difficult to decide on the approach they should take for successful implementation. Data warehouse architecture plays a major role the successful construction of data warehouse. This paper proposes an approach of constructing local data marts at source and integrating them at enterprise level by creating physical data marts to give global view for the analysis.

Keywords - Data Warehouse Architecture, CIF, ETL, Bus Architecture, Data Mart

I. INTRODUCTION

A data warehouse is a data repository, containing historical data, prepared for analysis and supporting decision making process [3]. Data in the data warehouse needs to be refreshed at predefined regular intervals to give knowledge workers view of the data as per their needs.

Data is extracted from data sources, mainly various operational systems of the organization, cleansed, transformed, integrated and loaded into the data warehouse. This process is known as ETL (Extract, Transform, and Load) process. Data is then made available to the knowledge user for access using decision support tools.

There are two strategies mainly followed to build a data warehouse namely top-down approach suggested by Bill Inmon and bottom-up approach suggested by Ralph Kimball.

The Corporate Information Factory Architecture earlier known as The Enterprise Data Warehouse Architecture is advocated by Bill Inmon, known as “the father of data warehousing” [1].

Another prominent figure in data warehousing, Ralph Kimball suggests Data Warehouse Bus Architecture in which enterprise level data marts are created which are connected through conformed dimensions [1].

Data warehouse architecture complement enterprise architecture which set up a framework for decision making process. Data warehouse architecture must cater to following two requirements 1) flexibility – provide flexibility to adapt to the changing conditions 2) scalability – provision for expansion for systems. [4]

Quality of input data provided to the data warehouse is important for the success of the data warehouse and business intelligence. Finding, understanding, selecting and improving data sources are critical to the success of any data warehousing project. [4] Lack of understanding of data sources and poor data quality of data sources is a major cause for the failure of data warehousing project.

II. METHODS AND MATERIAL

The Corporate Information Factory Architecture

An extensive enterprise-level analysis of data requirement provides the basis for this architecture. It starts by designing an enterprise model for a data warehouse. Multi-tier architecture is deployed which is comprising of a persistent staging area, an enterprise data warehouse and set of dependent data marts. The data warehouse is enterprise-oriented with function specific data marts. The data warehouse has atomic level data typically maintained in a 3\textsuperscript{rd} normal form. Dependent data marts are developed for specific functional area or departments. It may have normalized, de-normalized, or summarized/atomic dimensional data structures based on user needs with summarized data marts. Data warehouse uses an enterprise based normalized data model. Data marts use subject specific dimensional model. Users can query data warehouse and data marts.
This approach supports other analytical structure in an architected environment, including data mining sets, and operational data stores. Once it is built successfully, it minimizes population of independent data marts. Normalized detailed data gives flexibility to meet new and unexpected needs.

Enterprise level modeling and deployment takes longer to deploy and cost more. Complexity of design increases with increasing level of detail. It requires building multiple data stores and platforms. Subject area specific data marts are built once the enterprise data warehouse is implemented. Data marts contain data in de-normalized form also called star schema. Data marts usually contain summarized data based on the end users analytical requirements. If data marts only store summarized data then drill through operation becomes difficult and take longer. In such cases you require to store detailed data in data marts.

The purpose to de-normalize the data in the data marts is to provide faster access and for the easy understanding to the end users for analysis. If normalized schema is used, the queries would end up in a complex multiple level joins which would also slow down the performance.

The top-down approach is good when there is complete clarity on entire data warehouse requirement and business is ready to invest considerable time and money.

Data Warehouse Bus Architecture

The bottom up approach suggested by Ralph Kimball is an incremental approach in building a data warehouse. The architecture is referred to as data warehouse bus architecture. The development of the data warehouse bus architecture starts by taking up business requirement of a single subject area or a specific business process and a data mart is created. Data marts use dimensional modeling which is de-normalized format. Data from all the sources related to the specific business process is integrated into a single data mart at enterprise level. Once one data mart is ready, end users can access it and start using it for analysis. Additional data marts are developed for different business processes. Different data marts are combined using conformed dimensions and conformed facts. Conformed dimensions and conformed facts are the one which are shared across different data marts. This results in logically integrated data marts which provide enterprise level view of the data. These integrated data marts combined altogether form an enterprise level data warehouse. The bottom up approach helps to build data warehouse incrementally by developing data marts as and when requirement is clear and integrating them. This approach allows building data warehouse in less time as we don’t have to wait for specifying over all requirements of the warehouse.

This approach is good when we have time and cost constraints. It is also useful when the complete requirement of the warehouse is not clear but requirement for one of the subject area or business process is clear.
The advantage of this approach is that implementation time is less and does not require high cost initially, which helps business using data marts much earlier for the analysis compared to the top down approach.

The disadvantage of bottom up approach is that if we do not store detailed data i.e. atomic level data, then flexibility to cater to detailed analysis and future requirements is lost.

III. RESULT AND DISCUSSION

Proposed Local Data Mart Approach

In this proposed approach local data marts are created at the source for all the business processes for which analysis at the enterprise level is required. These local data marts are then used to populate respective enterprise level data marts.

Figure 3: Proposed Architecture

Local data marts contain atomic data extracted from its source. Local data marts use dimensional modeling. Though this approach is similar to data warehouse bus architecture with the main difference being there is one to one correspondence between source and local data mart for each of the business process. Thus local data mart is populated from a single source only. Physical enterprise level data mart for each of the business process is populated by extracting and integrating data from all the local data marts related to the same business process. This architecture also follows bus architecture by sharing conformed dimensions across different data marts at local and enterprise level.

Benefits of this architecture are:

1) Easy to find and understand relevant data at local level
2) Easy to resolve data quality issues when populating local data marts
3) Easy to track-back data quality problems arising when populating enterprise level data marts
4) Gives flexibility to adapt to changes happening at any source

This is approach is useful where high level data quality issues from various source systems are anticipated. It is also useful when source systems spread across multiple geographical boundaries and heterogeneous platforms.

IV. CONCLUSION

Data warehouses have become integral part of decision making process for the businesses. To provide enterprise level view of data for analysis of different functions, data from all the operational systems need to be cleaned, transformed and integrated at enterprise level. There is no de facto standard for architecture and construction of data warehouses. However quality and completeness of data coming from different heterogeneous sources play a major role in success of data warehouse along with flexible and scalable architecture. Proposed architecture of creating local data marts at all the operational sources and integrating each data mart at enterprise level provides an option to deal with data quality issues efficiently where sources are widely spread across various platforms and locations having different local needs. This approach also gives a scope to manage local concept hierarchies and global concept hierarchies when designing a data warehouse.

REFERENCES


