International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 05 | May-2016 www.irjet.net p-ISSN: 2395-0072

Review on Data Warehousing Architecture And Implementation Choices

Miss Poonam Wavare

Lecturer, Computer Engineering Department, V.P.M's Polytechnic, Thane, Maharashtra, India

---

Abstract - ***A data warehouse is an architectural construct of an information system that provides users with current and historical decision support information that is hard to access or present in traditional operational systems. It is also an integrated set of products that enable the extraction and transformation of operational data to be loaded into a database for end-user analysis and reporting. It usually contains historical data derived from transaction data, but it includes data from other sources. It separates analysis workload from transaction workload and enables an organization to consolidate data from several sources. This paper presents concept of Data Warehousing Architecture And Implementation Choices available for Data Warehousing.***

Three architecture choices are presented: the global warehouse, independent data marts, and interconnected data marts. There are several ways to implement these architecture choices: top down, bottom up, or stand alone. These three implementation choices offer flexibility in choosing an architecture and deploying the resources to create the data warehouse and/or data marts within the organization.

Key Words: **Data warehousing, Architecture Choices, Implementation Choices etc.**

1.INTRODUCTION

A Data Warehouse (DW) is defined as "a subject-oriented, integrated, time-variant, non-volatile collection of data in support of management's decision-making process". Data warehouses store huge amount of information from multiple data sources which is used for query and analysis.

It is not always the case, choosing an architecture should be done prior to beginning implementation. The architecture can be determined, or modified, after implementation begins. However, a longer delay typically means an increased volume of rework. And, everyone knows that it is more time consuming and difficult to do rework after the fact than to do it right, or very close to right, the first time. The architecture choice selected is a management decision that will be based on such factors as the current infrastructure, business environment, desired management and control structure, commitment to and scope of the implementation effort, capability of the technical environment the organization employs, and resources available. The implementation approach selected is also a management decision, and one that can have a dramatic impact on the success of a data warehousing project. The variables affected by that choice are time to completion, return-on-investment, speed of benefit realization, user satisfaction, potential implementation rework, resource requirements needed at any point-in-time, and the data warehouse architecture selected.

1.1 ARCHITECTURE CHOICES

Selection of an architecture will determine, or be determined by, where the data warehouses and/or data marts themselves will reside and where the control resides. For example, the data can reside in a central location that is managed centrally. Or, the data can reside in distributed local and/or remote locations that are either managed centrally or independently.

1.1.1 GLOBAL WAREHOUSE ARCHITECTURE

A global data warehouse is considered one that will support all, or a large part, of the corporation that has the requirement for a more fully integrated data warehouse with a high degree of data access and usage across departments or lines-of-business. That is, it is designed and constructed based on the needs of the enterprise as a whole. It could be considered to be a common repository for decision support data that is available across the entire organization, or a large subset thereof.

A common misconception is that a global data warehouse is centralized. The term global is used here to reflect the scope of data access and usage, not the physical structure. The global data warehouse can be physically centralized or physically distributed throughout the organization. A physically centralized global warehouse is to be used by the entire organization that resides in a single location and is managed by the Information Systems (IS) department.

1.1.2 INDEPENDENT DATA MART ARCHITECTURE

An independent data mart architecture implies stand-alone data marts that are controlled by a particular workgroup, department, or line of business and are built solely to meet

their needs. There may, in fact, not even be any connectivity with data marts in other workgroups, departments, or lines of business. For example, data for these data marts may be generated internally. The data may be extracted from operational systems but would then require the support of IS. IS would not control the implementation but would simply help manage the environment. Data could also be extracted from sources of data external to the organization. In this case IS could be involved unless the appropriate skills were available within the workgroup, department, or line of business.

The independent data mart architecture requires some technical skills to implement, but the resources and personnel could be owned and managed by the workgroup, department, or line of business. These types of implementation typically have minimal impact on IS resources and can result in a very fast implementation.

1.1.3 INTERCONNECTED DATA MART ARCHITECTURE

An interconnected data mart architecture is basically a distributed implementation. Although separate data marts are implemented in a particular workgroup, department, or line of business, they can be integrated, or interconnected, to provide a more enterprise wide or corporate wide view of the data. In fact, at the highest level of integration, they can become the global data warehouse. Therefore, end users in one department can access and use the data on a data mart in another department.

This architecture brings with it many other functions and capabilities that can be selected. Be aware, however, that these additional choices can bring with them additional integration requirements and complexity as compared to the independent data mart architecture. For example, you will now need to consider who controls and manages the environment.

Interconnected data marts can be independently controlled by a workgroup, department, or line of business. They decide what source data to load into the data mart, when to update it, who can access it, and where it resides. They may also elect to provide the tools and skills necessary to implement the data mart themselves. In this case, minimal resources would be required from IS.

2. IMPLEMENTATION CHOICES

The approaches to be discussed are top down, bottom up, or a combination of both. These implementation choices offer flexibility in determining the criteria that are important in any particular implementation.

The choice of an implementation approach is influenced by such factors as the current IS infrastructure, resources available, the architecture selected, scope of the implementation, the need for more global data access across the organization, return-on-investment requirements, and speed of implementation.

2.1 TOP DOWN IMPLEMENTATION

A top down implementation requires more planning and design work to be completed at the beginning of the project. This brings with it the need to involve people from each of the workgroups, departments, or lines of business that will be participating in the data warehouse implementation. Decisions concerning data sources to be used, security, data structure, data quality, data standards, and an overall data model will typically need to be completed before actual implementation begins. The top down implementation can also imply more of a need for an enterprise wide or corporate wide data warehouse with a higher degree of cross workgroup, department, or line of business access to the data

A top down implementation can result in more consistent data definitions and the enforcement of business rules across the organization, from the beginning. However, the cost of the initial planning and design can be significant. It is a time-consuming process and can delay actual implementation, benefits, and return-on-investment. For example, it is difficult and time consuming to determine, and get agreement on, the data definitions and business rules among all the different workgroups, departments, and lines of business participating. Developing a global data model is also a lengthy task. In many organizations, management is becoming less and less willing to accept these delays.

The top down implementation approach can work well when there is a good centralized IS organization that is responsible for all hardware and other computer resources. In many organizations, the workgroups, departments, or lines of business may not have the resources to implement their own data marts. Top down implementation will also be difficult to implement in organizations where the workgroup, department, or line of business has its own IS resources. They are typically unwilling to wait for a more global infrastructure to be put in place.

2.2 BOTTOM UP IMPLEMENTATION

A bottom up implementation involves the planning and designing of data marts without waiting for a more global infrastructure to be put in place. This does not mean that a more global infrastructure will not be developed; it will be built incrementally as initial data mart implementations expand. This approach is more widely accepted today than the top down approach because immediate results from the data marts can be realized and used as justification for expanding to a more global implementation.

The bottom up implementation approach has become the choice of many organizations, especially business management, because of the faster payback. It enables faster

results because data marts have a less complex design than a global data warehouse. In addition, the initial implementation is usually less expensive in terms of hardware and other resources than deploying the global data warehouse.

Along with the positive aspects of the bottom up approach are some considerations. For example, as more data marts are created, data redundancy and inconsistency between the data marts can occur. With careful planning, monitoring, and design guidelines, this can be minimized.

Multiple data marts may bring with them an increased load on operational systems because more data extract operations are required. Integration of the data marts into a more global environment, if that is the desire, can be difficult unless some degree of planning has been done. Some rework may also be required as the implementation grows and new issues are uncovered that force a change to the existing areas of the implementation. These are all considerations to be carefully understood before selecting the bottom up approach.

2.3 A COMBINED APPROACH

As we have seen, there are both positive and negative considerations when implementing with the top down or the bottom up approach. In many cases the best approach may be a combination of the two. This can be a difficult balancing act, but with a good project manager it can be done. One of the keys to this approach is to determine the degree of planning and design that is required for the global approach to support integration as the data marts are being built with the bottom up approach. Develop a base level infrastructure definition for the global data warehouse, being careful to stay, initially, at a business level. For example, as a first step simply identify the lines of business that will be participating. A high level view of the business processes and data areas of interest to them will provide the elements for a plan for implementation of the data marts.

3. CONCLUSION

Here we have analyzed regarding Data Warehouse Architecture wherein data marts are implemented, We develop a plan for how to handle the data elements that are needed by multiple data marts. This could be the start of a more global data warehouse structure or simply a common data store accessible by all the data marts. It some cases it may be appropriate to duplicate the data across multiple data marts. This is a trade-off decision between storage space, ease of access, and the impact of data redundancy along with the requirement to keep the data in the multiple data marts at the same level of consistency.

REFERENCES

[1] Chuck Ballard, Dirk Herreman, Don Schau, Rhonda Bell (Data Modeling techniques for Data Warehousing - redbooks.ibm.com)