2000

Expanding the Data Warehouse Paradigm to Support the Virtual Company

Mary Ann Robert
Bentley College, mrobbert@bentley.edu

Peggy M. Beranek
Bentley Colleg, pberanek@bentley.edu

Expanding the Data Warehouse Paradigm to Support the Virtual Company

Dr. Mary Ann Robbert, CIS Department, Bentley College, mrobbert@bentley.edu
Dr. Peggy M. Beranek, CIS Department, Bentley College, pberanek@bentley.edu

Abstract

Virtual Organizations (VO) are on the rise fueled by a more global base for commerce, technology development to support VOs and a more global demand for products, and virtuality can fundamentally transform the ways in which organizations operate. In order to operate successfully a VO must develop a knowledge infrastructure which allows for seamless flow of information between geographically distributed people, processes and repositories. To support this a virtual data warehouse needs an architectural design with a logically common meta-data, common semantics and common business rules. While data warehouses have been designed to support the traditional organization the VO's high level of uncertainty and lack of centralized control make the construction of a data warehouse to support a VO more challenging. This paper presents two possible scenarios for accomplishing this along with advantages and disadvantages of both and measures of success.

Introduction

The organizational structure of a company, its business goals, its decision support requirements and the market characteristics play an essential role in building the data warehouse. When the company is virtual and the data exists in diverse locations, achieving a single source of information for immediate decision making is problematic. Not only must this data be in one spot, it must be logically integrated. A decision maker needs to know if the necessary information is included in the data warehouse and be able to assess the quality of a decision depending on the quality of the information in the data warehouse (Gatziu et.al.,1999).

In 1999 Doug Hackney defined a data warehouse as a common and unique repository for the information of an enterprise. (Hackney*)* The question then becomes what is an enterprise and if a virtual company is an enterprise, what is the set of information required for the virtual company? A corollary to this question is where should the unique repository exist?

Virtual companies' capabilities to deliver their products to the market are defined primarily by their ability to organize and maintain a network of business relationships, rather than their ability to manufacture a product or deliver a service. This virtual organization needs a metasystem, organizational point of view to support strategic decision making. A virtual data warehouse that has a logically common meta data repository and common semantics and business rules can support this. This is not the prototype virtual data warehouse environment described by Ken Orr but rather an architectural design that exists and component data warehouses or data marts are designed to fit into.

This paper views the data warehouse as a business driven model. The paper does not examine program integration which is technical, but rather the convergence of business modeling, business rules, existing application assets and a middleware foundation or businessware conceptually similar to that defined by Acly.(Desmond). The virtual organization and the traditional data warehouse are first defined then the informational structure necessary to support the virtual company is examined.

Virtual Organization

The impact of technology and the trend of increasingly global organizations have both contributed to the concept of a Virtual Organization (VO). Communication in these organizations takes place over geographic, time and organizational boundaries. A VO can be defined as a new form of a network organization (Miles and Snow, 1986) enabled by advances in communication technology (Davidow and Malone, 1992) (Jarvenpaa, and Ives, 1994). Virtuality can fundamentally transform the ways in which organizations operate, making communication, knowledge sharing and collaboration even more critical.

The emergence of the Virtual Organization has been driven primarily by two overriding factors, 1) the need for flexibility (Prahalad &Hamel, 1991), and 2) the need for efficiency by sharing resources with other partners. Organizations tend to concentrate on their individual core competencies, however, mass customization (individualized products for each customer) is becoming the norm requiring the combination of knowledge and cooperation of numerous companies. This requires a great amount of collaboration among organizations. By sharing resources and information companies can become surprisingly more flexible and efficient, and thereby create a Virtual Organization.

Zwass (1991) states: "electronic integration has led to dramatic shifts in the definition of a firm, with the emergence of virtual companies whose capabilities to deliver their products to the market are defined largely by

their ability to organize and maintain a network of business relationships, rather than by their ability to manufacture a product or deliver a service" (Zwass, 1991).

The number of virtual organizations are rapidly increasing, however the term virtual organization is used in many different ways. A working definition of a Virtual Organization comes from Travica ,"a Virtual Organization refers to a new organizational form characterized by a temporary or permanent collection of geographically dispersed individuals, groups or organization departments not belonging to the same organization - or entire organizations, that are dependent on electronic communication for carrying out their production process. (Travica, 1997)" However, more than just communication is required for success. The VO that can set goals and make decisions based on integrated information has the advantage.

According to Jagers (Jagers, Jansen & Steenbakkers, 1998) a VO is a special case of a network organization, a networked organization is a collaboration of people or organizations. The VO is an extreme and far reaching form of a networked organization characterized by a high level of uncertainty and a low level of control, at the other end of this spectrum lies a 'planet satellite' organization. A planet satellite organization would be characterized by a centralized organization, which closely subcontracts various services and/or manufacturing tasks to a number of other organizations. The high-level of control would be regulated through concise and unambiguous contracts. These contracts also serve to decrease the amount of uncertainty.

A VO attempts to manage the lack of control through the pooling of knowledge and information. Situations with high uncertainty require a flexible network form to support sharing of information. For successful operation a VO requires the development of a knowledge collaboration infrastructure, which is dependent on seamless knowledge flows between people, processes and information repositories, which may be geographically distributed (Skyrme, 1998).

Jagers (Jagers, et.al., 1998) further defined a VO to be comprised of a number of characteristics (Table 1).

Table 1 Virtual Organization Characteristics

Boundary Crossing
· Complementary core competencies
· Geographical dispersion
· Changing participants
· Participant equality
· Electronic communications

Data Warehouse

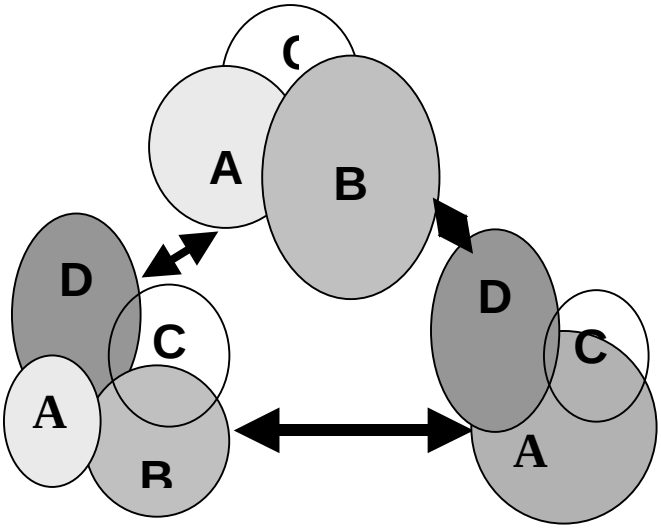
The goal of the data warehouse is to be an integral part of the organizational information system by efficiently, consistently and reliably storing large amounts of data in a central location. (Neary,1999) The data warehouse provides one single version of the truth, bringing together valuable corporate information for decision makers. The warehouse is scalable and can be configured to meet the growing corporate needs. The question is can it meet the information needs of a geographically disperse, boundary crossing organization with changing participants?

Whereas databases were designed as transaction processing systems and remained operational, data warehouses are part of the information system and are primarily a read only database used as an information source. In a database, data can be retrieved from separate sources which are not integrated and the data is dirty. The data warehouse must provide a clean, integrated data source. This integrated data set is combined and reconfigured as associations and partnerships are formed and is essential to provide the seamless knowledge flows. Individual organizational databases can feed a data warehouse but cannot be a data warehouse.

Data warehouses have been designed to integrate enterprise wide data into a repository from which decision making information can be retrieved. Though this works well for a traditional organization, the challenge comes in scaling it for a VO. Though the required decision support information is the same, the VO's high level of uncertainty and lack of centralized control do not lend itself to constructing a data warehouse using traditional methods.

Data warehouses must be built as complex adaptive systems with the business end user firmly in mind. Maxwell and Gutewicz contend that successful data warehouses invariably are individually small, heavily used by the target business unit organization, are constantly changing in terms of their content based on changing market needs, and are controlled by the business unit of an enterprise. (Maxwell & Gutewicz) However, Hackney would consider these legamarts (non-architected data marts) which would require an overall architecture enabling each of the business units data marts to have the capability of integrating into the enterprise data warehouse. (Hackney,1997) If we equate the business unit with the associate organization then each organizations repository or data warehouse must be integrateable into the virtual organization's repository.

Figure 1 Virtual Company's Information



Integrating the Information

Figure 1 shows the basic input, processing, output for the value chain with our virtual companies superimposed on top. A, B, C, D and E represent five different participating companies. A company can participate in more than one process and each company is not necessarily an equal participant at every phase. Companies can be added and subtracted as needed. In order to drive this organization as a whole, information must be available for viewing from multiple sites yet this must appear as that of a single enterprise. A single version of the truth must exist for the VO.

A single business plan exists for the virtual organization, not multiple plans. There also must exist a business process that ties together previously existing systems to provide end to end functionality. Accordingly, a single data source for decision making information must exist. Defined subject areas, common dimensions, common metrics, common semantics, common business rules, and common source systems of record are in this shared source. The source can be a logical integration of the required data or a physical repository. The important point is that there is a common point of entry to one fully integrated set of cleaned data and a single set of metadata.

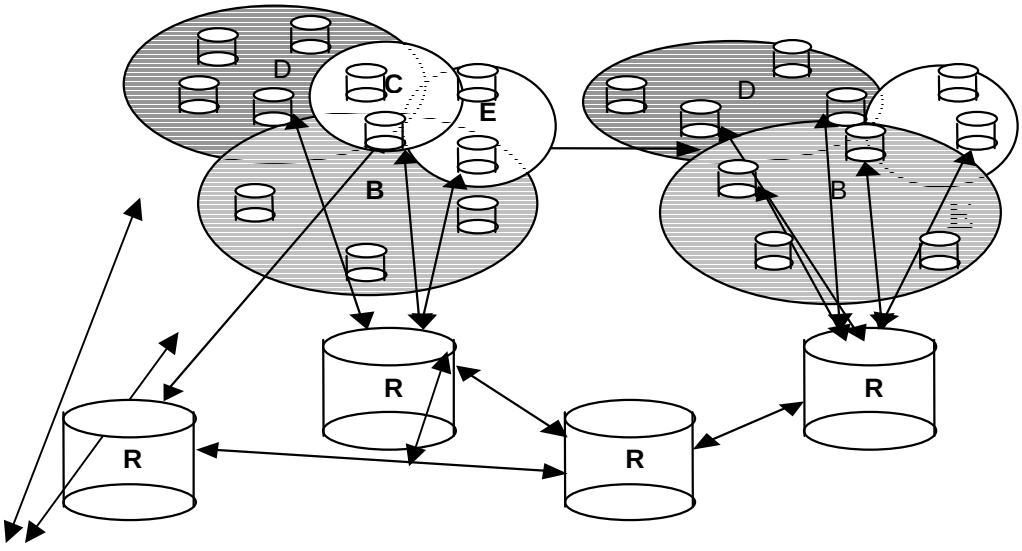
Metadata

The creation of an integrated environment centered on a common, extensible metadata repository, allows seamless metadata population and the sharing of information across all functions and processes. Separate metadata repositories rstricts reporting and advanced analysis on joint data. Metadata provides the understanding, rationale and audit trail. A logically common metadata repository that is available, accessible, current and complete, with one entry point is essential.

Metadata must be standardized across units and processes. The granularity of the metadata definitions will vary but it must be flexible and broad enough to accommodate a virtual organization of organizations. It must be scalable enough so that no repository restricts the global metadata. New participants must be able to easily plug into the metadata and convert their existing metadata into a compatible format.

The metadata model is an evolutionary development. The metadata definitions start out very generic and evolve into a more specific usable metadata. A hierarchical inheritance system can provide the basic structure for the metadata model. At the highest level, this model is a model of sub-metadata models used by various virtual

Figure 2 Virtual Company's Information



companies. A single model integrating ALL metadata from internal and external sources might be impossible to create in the required time period. However, sub-metadata models with feeds regulated by the transfer structure are usually constructable. The sub-models are independent units, which fit together to form the larger multi-organizational model. This construction allows for addition and subtraction of sub-models without rebuilding the entire structure.

Integration

Initially the virtual company appears as in Figure2 where the R's represent some type of shared data repository, data warehouses, or data marts and the cylinders within the companies represent local databases or data marts. However, multiple shared repositories or even warehouses do not assure a single information source capable of providing integrated support. The optimal solution would be a data warehouse designed specifically to meet the informational needs of the virtual organization. An average three year return on investment for a data warehouse is 401%. This is based on an implementation time from 6 months to 3 years. (Neary, 1999) Though the return is excellent the time is prohibitive.

The individual repositories (R) can be viewed as data marts and we can try to integrate them. This leads to the legamarts' problem, where individual data marts are built but then the time and effort to integrate them is equal to building a new data warehouse.(Hackney, 1997) Hackney's approach of building a series of incremental data marts requires an overall architecture. How and where can an architecture be designed to achieve the flexibility and scalability required by a virtual company? Equally important is, who is responsible for this architecture?

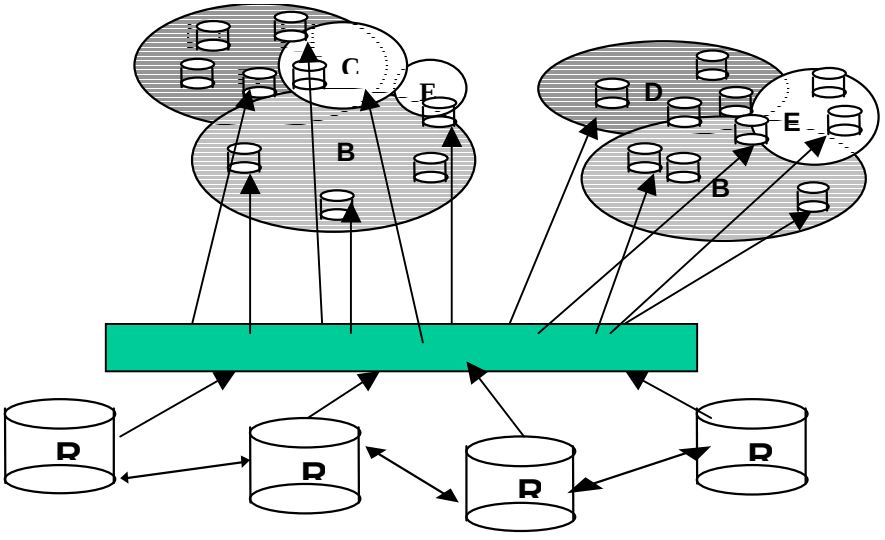
Two possible solutions are possible. One company can take the responsibility for the architecture and be the lead company with partners and associates required to adapt their repositories or warehouses to the lead architecture. Alternately is the existence of an independent company that maintains data warehouse schemas and implementation capabilities. The management of the informational resources for the VO can be outsourced to this organization.

In the first scenario, one company has all the business rules, metadata and access rules in a middle layer as shown in Figure 3. At a minimum this layer is a repository to store required specifications for the virtual organization and filter the data to provide a unified view. DTEAMM (Design, Transformation, Extract, Access, Monitor and Management) tools are available to assist with the technical integration. The business integration is the challenge. Assuming the virtual company is a partnership, the possession and the maintenance of the middle layer by one company can lead to domination by this company. The goals and directions of a planet satellite type organization can best be supported by this structure. The reliance upon one participant for the creation and maintenance of the middle layer denies the VO characteristic of participant equality.

The second scenario requires the existence of a center that has a plug and play operational data warehouse with a high level metadata structure which is modified to reflect the VO. The VO either outsources the data warehouse to this company or the information company becomes a partner in the VO. This information partners responsibility is the creation and maintenance of a single source of information and the information flow.

A data warehouse implies a lengthy, time-consuming evaluative process of potential new organizational

Figure 3: Information Stored in Middle Layer



linkages. It is also assumed that organizational relationships will be stable over a sufficient period of time to allow a careful examination of inter-organizational processes. Virtual organizations do not have the luxury of formalized, long-term evaluative periods. Electronic interactivity with other organizations, even organizations with which the firm has had no track record, are commonplace and demand information flow to and from an integrated repository. The information exchange is more than just EDI data flow, it's an integrated data warehouse used for tracking trends, analysis and decision making.

Where will the technology standards be set? Who will define operating standards? Who will bear the brunt of the technology costs necessary to maintain common interfaces regardless of underlying platforms and information systems? These questions point out the need for a new approach to virtual company design. Companies must introduce a *third party* who can link the dynamically bound virtual units. An integrated storage and retrieval center available to all elements participating in the current virtual company should exist. This "repository" of integrated data and metadata would lend itself to trusted third-party administration.

When information is required to support new VO activities or partnerships, a dynamic association can be established *on the fly*. While a standard in-house information repository contains all the information and links to support internal decision making activities, our "centralized" information repository requires the constant flow of data from outside sources. Therefore, this centralized repository is dedicated to the common goals, and participating companies' core processes and activities must be made amenable to providing data to that repository. This is a simpler task than building a data warehouse or unifying non-architected data marts or warehouses.

Measuring success

Since we have focused on the business perspective of the data warehouse, we cannot change course now and measure the technical aspects, such as gigabytes used or throughput. Rather we must focus on customer satisfaction and decision making information availability. Some measurements that must be included are ease of use, response time, validity of the data, quality of support. The metadata can be evaluated separately. The metadata is assessed on frequency of maintenance, currency, and again ease of use. When the data warehouse is used for a virtual company time to construct and inclusion of all data sources are additional metrics required.

Conclusion

Providing support information to a virtual company is not a technology problem but a business problem. The business drivers are at the center of a successful solution. The VO's need for efficiency in sharing resources points out the requirement for a centralized model and metadata. A data warehouse providing a single version of the truth can provide the analytical information required to support strategic decision making for the VO.

Information, if properly stored, can then be retrieved and analyzed. The outcomes can provide the virtual companies with a competitive edge in a world of saturated markets. The challenge is finding the optimal solution to provide a preset architecture that can be configured in a short period of time to address the informational needs.

References

Davidow, W.H. and Malone, W.S. . *The Virtual Corporation*, Edward Burlington books, Harper Business, Harper Collins Publishers, New York, 1992.

Desmond,J. & Acly,E., "Beating the Integration Blues", *Software Magazine*, Sept. 1999 http://www.softwaremag.com/archive/.

Hackney, D., *Understanding and Implementing Successful Data Marts*, Addison-Wesley, MA. 1997.

Hackney, Douglas, "The New World of Data Warehousing", http://www.egltd.com/production/homeindex2.html, (current March 14, 2000).

Gatziu,S., Jeusfeld,M., Staudt,M., Vassiliou, Y., "Design and Management of Data Warehouses: Report on the DMDW'99 Workshop", *SIGMOD Record*, (28:4) Dec. 1999.

Jagers, H.., Jansen, W. and Steenbakkers, W., "Characteristics of Virtual Organizations", *Proceedings of the VoNet Workshop*, April 27-28,1998, 65-76.

Jansen, W., Steenbakkers, W. and Jagers, H "Electronic Commerce and Virtual Organizations", *Electronic Journal of Organizational Virtualness*, (1:1), 1999, 54-68.

Jarvenpaa, S.L. and Ives, B. "The Global Network organization of the future", *Journal of Management Information Systems*, (10**:**4), 1994, 25-57.

Maxwell C.& Gutowitz, H., "Data Mining Solutions and the Establishment of Data Warehouses, Corporate Nirvana for the 21 Century"

Miles, R.E. & Snow, C.C.. Organizations: New concepts for new forms. *California Management Review*, (18:3), 1986, 62-73.

Neary, R., "Building a Data Warehouse and Data Mining for a Strategic Advantage", *Journal of Information Technology Theory & Application* , (1:1), Spring 1999.

Orr, Ken, “Data Warehousing Technology,” A White Paper, http://www.kenorrinst.com/dwpaper.html (current March 14, 2000).

Prahalad, C.K., Hamel, G., “The core competences of the corporation”, *Harvard Business Review*, 82, 1991, 79-91.

Skyrme, D.J., “The Realities of Virtuality”, *Proceedings of the VoNet Workshop*, April 27-28, 1998, 25-34.

Travica, B., "The Design of the Virtual Organization: A Research Model*", Proceedings of the Americas Conference on Information Systems*, August 15-17, Indianapolis, AIS,1997, pp. 417-419.

Zwass, V., “Electronic Commerce: Structures and Issues”, *International Journal of Electronic Commerce*, (1:1) 1996, pp. 3-23.