Oriental Journal of Computer Science & Technology Vol. **1**(1), 55-58 (2008)

Data warehousing practices in business initiatives

G. VIJAY KUMAR and M. SREEDEVI

School of Computing, K.L.College of Engineering Vaddeswaram,
Guntur (D.T) Andhrapradesh (India)

(Received: February 12, 2008; Accepted: April 04, 2008)

ABSTRACT

The paper presents the data warehousing architecture and practices used at a major. Retailing company. Many considerations were assessed when deciding which data warehousing architecture to adopt. The paper discusses the two pre-dominant styles in data warehousing, namely the "Bill Inmon Style" or the top-down approach and the "Ralph Kimball Style" or the bottom-up approach. The com-pany chose the Inman style due to a unique combination of circumstances in their business and technical environments, which are being discussed in detail. Much of the information presented in this paper is based upon the direct experiences of the lead data architect assigned to the projects under which this retailing company's customer data warehouse evolved.

The architecture has evolved over time and currently has been accepted at the company as a best practice. It is interesting to mention that both the hardware platform (CPU and disk drives) and Relational Database Management System (RDBMS) soft-ware employed today at this company for data ware-housing is not the same as was selected for the first instantiation. The implication was that the best plan or practice was a flexible one. There were many challenges, like organizational, technical, data sourcing and data naming, needed to be solved during the pre-project, initial stages, and throughout the project and beyond. The initial data warehouse, implemented in 1996, was termed an overall success and approved for expansion. The current data warehouse data are being used by over six hundred registered users to fine-tune customer marketing and leverage and share data in an enterprise manner. The data warehouse has allowed the company to strengthen customer relation-ship management (CRM) core capabilities and business partnerships. Today, there are many departments benefiting from queries and requests for data warehouse data, many anticipated, some not.

Key words: Data warehouse, business intelligence, CRM**.**

INTRODUCTION

A diverse retailing company was experiencing the usual growing pains of the middle 1990's. The diversity of businesses supported by multiple business units and the company's Information Technology organization had resulted in "stove-pipes" of data, along with corresponding computer applications, which were built over several years. The data in these legacy systems were not easily accessed, causing difficulty in making information out of the data, discerning knowledge from the information, and implementing sound business decisions based upon this knowledge. Also, the legacy operational data were not integrated with other operational data, were organized along process or functional orientations, and were predominantly current-valued, containing little or no history. Because the data as such could yield very little business intelligence, the company decided in 1995 that data warehousing could be used to release their data from its "data jailhouse".

Data warehousing architecture

Many decisions must be made when implementing a data-warehousing environment. As if the technology decisions were not difficult enough in and of themselves, deciding which data warehousing architecture approach to use is sometimes even more difficult. There are two general styles from which to choose – one termed herein the "Bill Inmon Style" and the other the "Ralph Kimball Style"[8].

The Inmon style is considered application neutral, while the Kimball style has data prearranged by certain dimensions according to desired output[6,7,10,11,12]. If the Inmon style data warehouse has data covering most or all data subjects for the company, it can be termed an "enterprise" data ware-house. With the Kimball style, the sum of all individual multidimensional data structures is considered the "enterprise" data warehouse. Although highly debated in some data warehousing data architecture communities, the detail advantages and disadvantages, as well as the recommended analysis and selection process of each style, is beyond the scope of this paper. Companies usually pick one style over the other based upon a combination of employee expertise, assumed preference, consultant or vendor recommendation, budget, existing technologies, or perceived net advantages.

Business users had an overwhelming desire for detail, transaction-level data. Under the Kimball approach, data are typically summarized by higher-level dimensions[8]. In other words, it would be rare to employ "Transaction ID" as the lowest-level dimension, but rather "Product Type" or some other higher-level dimensional measure. Under the Inmon approach, data are typically kept at the lowest level of detail[13]. In other words, each transaction would be stored in its 3NF form and could be summarized by "Product Type" or other dimensional measure upon reporting to the business user.

Due to traditional business "stove-pipes" of data, potential cross-business use of data was unknown. Under the Kimball approach, data are arranged in an application- or data-view-specific manner[8]. Under the Inmon approach, data are arranged according to the rules of normalization and remain application-and data-view-independent. Sufficient expertise existed in the business community to support user self-sufficiency incorporating native SQL against an atomic-level data warehouse. There was also a general absence of Business Intelligence tools for accessing data warehouse data. Under the Inmon approach, while the SQL can get quite complex, it still will not be as complicated as that needed to access a multidimensional structure and perform drilling navigation[13].

The architecture adopted as the best practice, as shown in Figure 1, consists of four distinct, interacting components. As depicted, the legacy operational systems and ODS are used as

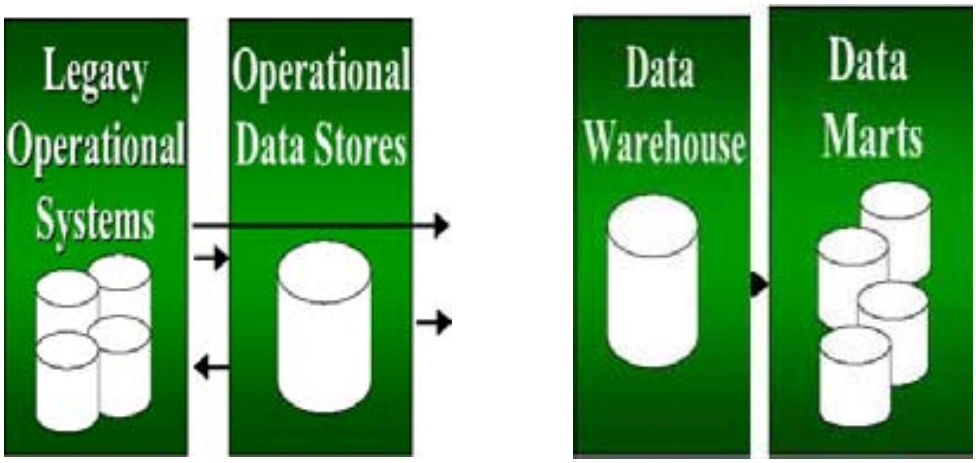

Fig. 1: Choosen datawarehousing architecture

sources for data ware-house data. Outside sources, such as household demographics or ZIP code geo demographics, may also be used as sources for data warehouse data. Any needed data marts are built with data from the data warehouse, thus being "dependent" data marts. "In-dependent" data marts should be discouraged, as the data used to build them would not be of the same assured quality as the data warehouse "Single source of truth".

Best practices

As a reminder, the thrust of this paper is not technological, so the various hardware and software selection decisions will not be covered. Those are best left up to the company technicians and consult-ants who are charged with selecting the best set of technological solutions to match the business problem. This analysis, research, testing, and selection process can be a project in itself, and was handled as such at the U. S. retailing company. What may be interesting is that both the hardware platform (CPU and disk drives) and Relational Database Management System (RDBMS) software employed today at champions, fully utilize or establish a data ownership stewardship function and process[2].

Data warehouse sponsorship

One of the basic best practices you can employ for data warehousing is to ensure that a high-level business champion exists, not just during building of the data warehouse, but ongoing continually after the data warehouse is built[1,2]. It is extremely important for the business champion to engage data warehouse business partners in an "enterprise" manner, not as individual vertical business units

Data warehouse growth

Most data warehousing initiatives have found that there is a continuous need for incremental additions to the data warehouse[2]. Treat the data ware-house as an ongoing system and spawn specific projects when appreciable expansion is needed. Keeping your data warehouse team intact after the initial build is very important in order to sustain the capability to react to this need. To paraphrase a popular saying, "Data warehousing is not a destination – it is a journey".

Data warehouse expertise

In addition to the high-level business champion, your organization should use data warehousing industry experts for both validation and expertise deficiencies[14]. Be sure to interview, hire, and contract with individuals and firms according to the data warehousing style, and perhaps even the technologies, you choose. Also, there are a number of Industry trade shows and conferences from which beginner to experienced practitioners can benefit greatly. Again, select and use these opportunities based upon the style of data.

Data warehouse scope

For the initial release of your data warehouse, limit the number of data subjects implemented and the extent of their content, perhaps employing an evolutionary prototype or proof-of-concept development methodology. This will minimize initial investment, help gain expertise with a smaller set of data (and, thus, a smaller set of technical challenges), and deliver business value sooner. This is an excel-lent way of demonstrating the informational and monetary benefits of data warehousing to the com-pany's top-level management, increasing their overall commitment and support of the concept.

Data warehouse data modeling

It is important, once a data warehousing architecture is chosen, to adhere to it from beginning to end . This may seem rhetorical, but there can be many opportunities and much pressure to short-cut the process necessary to create a quality data warehouse. Using a robust data modeling tool, follow a typical conceptual to logical to physical data model progression, maintaining all data models in as close to third-normal form (3NF) as possible .

Data warehouse attribution and keys

When defining attributes for the entities of the data warehouse data model, do not define intelligent, compound fields, especially when for attributes making up the key of the entity[3]. As a best practice, use native keys for primary keys; do not use token keys, which are made up "serial" type numbers with no meaning that represent a unique set of multiple native key values. With multidimensional data warehouse structures,

however, it is often recommended to use token keys because, with the multiple dimension enti-ties surrounding a central facts entity, the primary key list of the central facts entity would be the un-wieldy list of all primary keys of its dimension entities [8]. In practice, getting rid of the multiplicity of keys has more to do with minimizing SQL keying of power users than maximizing database performance. A potential compromise would be to carry both the native keys and the token keys, trading ease of use for more database space consumed.

Data warehouse loading

When populating the data warehouse from the legacy, external, and ODS files and databases, you should employ the use of utility Extract / Transformation / Load (ETL) purchased software. Similarly, build necessary dependent data marts from the data warehouse using an ETL tool .These tools are somewhat costly, but provide necessary structure and efficiency in ensuring data quality, transformation and standardization of data values, and in building and delivering the data stream necessary to load the data warehouse.

Data warehouse data marts

Independent ("end-run") data marts built directly from legacy, external, and/or ODS data files and databases should be avoided. It is best to first source the data into the data warehouse, thus becoming part of the "single source of truth", and then into a data mart, if necessary

ACKNOWLEDGEMENTS

We sincerely thankful to our chairman Sri K. Satyanarayana Garu, Principal Sri L.S.S.Reddy garu and staff of our department who support us to make out.

REFERENCES

1. Agosta, L., The Essential Guide to Data Ware-housing. Prentice Hall PTR, Upper Saddle River, NJ 07458 (2000).
2. Anahory, S and Murray, D., Data Warehousing in the Real World: A Practical Guide for Building De-cision Support Systems. Addison Wesley Longman Limited, England (1997).
3. Armstrong, R. White paper by NCR Corporation- The Fallacy of Data Mart Centric Strategies (Short Term Gain, Long Term Pain) (2002).
4. Chen, P., The Entity-Relationship Model – Towards a Unified View of Data, ACM Transactions on Database Systems,9-36 (1976).
5. Chowdhury, S., Lecture notes on Data Ware-house and Data Mining. The College of Business Ad-ministration, Roosevelt University, IL (2002).
6. Inmon, B., "The Problem with Dimensional Modeling" DM Review Magazine Archived Article (1999).
7. Ismail, W. and Chowdhury, S., Database Applications in Business. In Proceedings of the MBAA, Chicago, IL (2003).
8. Kimball, R., The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses, John Wiley (1996).
9. Kroenke, D., Database Processing – Funda-mentals, Design and Implementation (8th ed.). Pearson Education, Inc (2002).
10. Letowski, B. Parzatka, H.Woods, N., North-Wind Star Schema – a project work presented and submitted in Seminar on Data Warehouse and Data Mining at the College of Business Administration, Roosevelt University, IL. 2002 (2002).
11. McFadden, F. Hoffer, J. and Prescott, M Mod-ern Database Management. (5th ed.). Addison-Wesley Educational Publishers, Inc (1999).