

December 2003

Building Data Warehouses Using the Enterprise Modeling Framework

Joseph Chan
Roosevelt University

Follow this and additional works at: <http://aisel.aisnet.org/amcis2003>

Recommended Citation

Chan, Joseph, "Building Data Warehouses Using the Enterprise Modeling Framework" (2003). *AMCIS 2003 Proceedings*. 317.
<http://aisel.aisnet.org/amcis2003/317>

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2003 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

BUILDING DATA WAREHOUSES USING THE ENTERPRISE MODELING FRAMEWORK

Joseph O. Chan
Roosevelt University
jchan@roosevelt.edu

Abstract

This paper proposes an enterprise modeling framework for the deployment of data warehouses. The framework provides the information roadmap coordinating source data and different data warehouses across the business enterprise. The paper introduces a solution to address data warehousing issues at the enterprise level while avoiding the pitfalls of creating enterprise data warehouses and universal data marts. It further proposes a change of paradigm from point solutions focus to a methodology driven by enterprise requirements to meet the challenges of the new economy. The proposed framework emphasizes the separation of the conceptual construct from the physical and operational constructs of an enterprise. It points out the differences and dependencies of analytic and operational processes and how data warehouses and operational data stores respectively support their information requirements. This paper will demonstrate how the enterprise modeling framework for data warehousing can produce business benefits.

Keywords: Enterprise model, information roadmap, data warehouse, data mart, business intelligence, customer relationship management, supply chain management

Introduction

The 90's represent the decade of technology acquisition for many companies. Confronted with the Year 2000 problem, businesses unleashed vast amounts of capital spending in upgrading old systems or replacing them with new ERP systems. However, most companies are not reaping the anticipated benefits in information access and sharing. Technologies implemented in the past decades have created the legacies confined by the boundaries of systems and organizations. Information collected in these transactional systems got buried in their respective silos and hidden from the enterprise.

As we move towards a knowledge-based economy, empowered by the real-time infrastructure of the Internet, businesses are in a hurry to gather and analyze all sorts of data for competitive advantages. For example, businesses collect data about customers to enhance customer relationship management programs. The government collects data about crimes, court dispositions and correctional methods for crime prevention. Data collection seems to have evolved into a business on its own. The challenge however is the transformation of these data into useful information for the enterprise.

Data warehousing provides the data construct to aggregate data by subject areas for the purpose of decision support. Data warehouses have certain characteristics that differentiate them from transactional databases. The classic definition by Inmon (1993) describes these characteristics as *subject oriented, nonvolatile, integrated and time variant*. As data warehousing technologies are made available for data extraction, transformation and loading (ETL), some use data warehousing as a means of data consolidation at the transactional level for reporting and other purposes.

A disturbing phenomenon has developed over the years, which reminds us of the issues in file processing systems a few decades ago. During that period, a specific program was created with its associated data files when a specific need arose. If there was a need to solve another problem, another program was developed and another set of data files was created. Such practices caused the data redundancy and inconsistency problems for data management across an enterprise. The database management systems that came afterwards were trying to solve the problem by putting related data in one place, at least conceptually. The current issues of data warehousing are, to an extent, similar to those in file processing. If there is a need of information to solve a problem, a

data warehouse or data mart is created for the subject matter. If there is another need of information for another problem, another data warehouse or data mart is created for this other subject matter. Over time, it becomes a maintenance nightmare to coordinate and synchronize information in these data warehouses within an enterprise. Some companies resort to new strategies in consolidating old data warehouses into new data warehouses, and thus creating more legacy data warehouses that need coordination, synchronization and maintenance.

It is not uncommon in today's business environment that decision support requires information across the entire enterprise. Information needed to create business intelligence may require customer information, marketing and sales information, production information, supplier information, ...etc. It may involve many subject oriented data warehouses in the enterprise. Furthermore, some businesses are using data warehousing as a holistic means to address both transactional and analytical data requirements. Real-time information access with unanticipated requirements may put the usage of data warehouses in the most ineffective mode. Which, if any, of the data warehouses can answer such and such queries? And what if it requires multiple data warehouses to seek the truth?

This paper discusses a logical framework that supports the information requirements both at the transactional and analytical levels for the enterprise. It points out the different levels of data structures and how they are related. Furthermore, it provides the consistency of data for data warehouses and data marts throughout the enterprise.

Two Ends of the Spectrum

In the evolution of building data warehouses and data marts, two practices have emerged. One advocates a centralized enterprise data warehouse that serves as the repository of data from all data sources, from which various data marts can be created. The other advocates a purely bottom-up approach where data marts are built as point solutions to departmental or functional needs isolated from the rest of the enterprise. In addressing the data warehousing issues at the enterprise level, Simon (1998) refers to these two approaches as the "big bang" and the "loose confederation" approaches respectively. Simon further pointed out that the enterprise data warehouse "big bang" approach attains the maximum data extraction and the highest degree of source-neutral information abstraction. However, as pointed out by Simon, there are many pitfalls, which include cross-organization issues, technical complexity, data semantic issues and long delivery time. On the other hand, the bottom-up approach, while providing the point solutions with respect to departmental or functional needs, will result in information silos within an enterprise over time. Companies attempt to tie together these data marts into universal data marts. While the bottom-up approach avoids some of the pitfalls of the enterprise data warehouse, it creates a different set of issues. As summarized by Simon, these issues include redundancies in data extraction from data sources, unclear objectives by integrating data marts, technical challenges in platform integration and cross-functional issues.

The Enterprise Model Solution

Analogous to building a house, the enterprise data warehouse "big bang" approach gathers all the 2 by 4's, drywall, plumbing and electrical units in a storage area and then tries to figure out how they will be used in the construction of each room later. In the bottom-up approach, each room is built separately sourcing its own 2 by 4's, drywall, plumbing and electrical units. The task is then trying to fit these rooms together into a single house.

Each of these approaches focuses on the physical aspects of building a house. The enterprise data warehouse "big bang" approach puts all data in one place first, where in the bottom-up approach, each data mart sources its own data. In practice, various degrees of both of these techniques may be used depending on the requirements. For example, a garage may be prefabricated and fitted into a house easily. Drywall or floor tiles may be purchased for many rooms so that what is left over from one room can be used for another. Plumbing and electrical wiring may best follow an integrated plan so they all fit together as a unit. Therefore, a logical approach is to start with a blueprint of the house and determine a combination of these options, optimizing cost, time, and benefits. The enterprise model provides this framework in building data warehouses. It should be pointed out that there is a fundamental difference between an enterprise data warehouse and an enterprise model. The former is physical concept whereas the latter is a logical concept. The conceptual enterprise model provides the blueprint that can be used for various physical designs for building data warehouses or data marts for the enterprise.

The Enterprise Model framework is an extension of the "3-Schema" architecture originally proposed by ANSI/X3/SPARC (1975). The 3-Schema addresses the construct of data based on three levels of representation: the conceptual schema represents the logical view of data, the internal schema represents the physical data storage definitions, and the external schema represents the user

application views of data. The fundamental concept is the separation of the data definition (the “what”) from its physical representations and usage (the “how”, “where”, “who” and “when”).

While the 3-Schema provides the foundation of data definition in the development of databases and their applications, the fundamental concept can be extended to enterprise information management. This extended definition of the 3-Schema will be henceforth called the Enterprise Model. The following summarizes the basic concepts as introduced by Chan (2003).

The Conceptual Enterprise Model

Similar to the conceptual schema in the 3-Schema architecture of data, the Conceptual Enterprise Model (CEM) is a representation of the “what” in an enterprise. The CEM consists of the conceptual definition of data and business functions across business units and processes for an enterprise. Common methods in data modeling and function modeling such as the Entity-Relationship modeling introduced by Chen (1976) and functional hierarchy decomposition can be used to develop the CEM.

The Operational Enterprise Model

Similar to the external schema in the 3-Schema architecture of data, the Operational Enterprise Model (OEM) is a representation of the usage of data describing the business “how”, “where”, “who” and “when”. The OEM consists of the views defined by business processes, organizations and user applications. Common methods in process design and application design can be used for the development of the OEM.

The Technical Enterprise Model

Similar to the internal schema in the 3-Schema architecture of data, the Technical Enterprise Model (TEM) is a representation of the physical implementation describing the technical “how”, “where” and “when”. The TEM consists of the internal views of data structure and storage, software modules, hardware/networking platforms and communication protocols. Each one of these dimensions requires specialized knowledge in the respective disciplines for its design and implementation.

The Enterprise Information Roadmap

The key concept behind the Enterprise Model is the physical separation of the three layers of the enterprise (conceptual, operational, and technical), but yet preserving their logical connections. We will call this construct the Enterprise Information Roadmap (EIR). If we view the CEM, OEM and TEM as sets, then the EIR can be thought of as the set of relations between these sets. The virtuality of the model (the three layers being physically separated but logically connected) insulates enterprise information assets from process, organizational and technological changes.

Building Data Warehouses

Supporting analytic processing is the ultimate objective of building data warehouses. In the context of the Enterprise Model, the analytic processes are at the same level as the Operational Enterprise Model (OEM). The OEM concept is extended to an analytic component of the Enterprise Model. This will be called the Analytic Enterprise Model (AEM). The AEM consists of different types of analytic processes for an enterprise. They include various decision support processes, multi-dimensional analytics, data mining and other methods of data analysis that facilitate decision-making and predictions.

Figure 1 depicts the relationship between the AEM and OEM. Results of analytic processes affect operations, and conversely, data obtained in operations affect future analytic processes. For example, customer-centric analytic processes create customer insight, which affects the operations of a one-to-one marketing campaign. Data obtained from the one-to-one marketing campaign provides the feedback to the analytic processes, which in turn creates additional and new customer insight.

At the Conceptual Enterprise Model (CEM) level, the Enterprise Data Model (EDM) provides the framework for the conceptual design of the data warehouse data model. The Data Warehouse Data Model (DWDM) aggregates data along subject areas. For example, if the data warehouse or data mart under consideration is oriented around the subject area of customers, one can look into the CUSTOMER data entity in the Enterprise Data Model (EDM) and construct a model around it by identifying all relevant

data entities that have one-to-many relationships associated with it. Dimensional modeling techniques such as Star Schema and Snowflake Schema can be employed to develop the data warehouse data model as described in Todman (2001). The data warehouse data models in conjunction with the target data warehousing technologies provide the basis to the design of the respective data warehouses. Furthermore, the Enterprise Function Model (EFM) can be extended to include the analytic functional requirements in addition to operational functional requirements. Observe that these functional requirements can be implemented using various analytic processes at the AEM level.

The physical data warehouses and all source systems belong to the Technical Enterprise Model (TEM). The Enterprise Data Model (EDM) can be physically realized as one or more Operational Data Stores (ODS). The ODS provides data consolidation of source systems. Data are extracted, transformed and loaded into the ODS from different source systems which include the legacy systems, transactional systems and external systems. It should be observed that the total consolidation of data into one data store might not be practical for large enterprises.

ODS may be of great importance in situations where multiple source systems are needed for certain data access requirements. It provides a physical representation of consolidated and transformed data from source systems, and acts as the data feed to target data warehouses. Moreover, the ODS, not the data warehouse, serves the need for transactional reporting requirements.

The key point of the enterprise modeling approach is that the design allows each data warehouse or data mart to source its data from zero to one or more ODS. In the case of zero ODS, the data warehouse or data mart sources data directly from the source systems. This is similar to the case of the previously discussed bottom-up approach. The difference is that the sourcing of data under the enterprise modeling framework is guided by the enterprise information roadmap, which provides the conceptual integration of data from source systems to various target data warehouses. When all data warehouses and data marts source from one big ODS, the architecture is similar to the enterprise data warehouse big bang approach. The optimal design of the number of ODS for an enterprise will take into consideration the number, size and relationships of data sources and target data warehouses.

Figure 2 provides the Enterprise Model architecture combining the AEM, OEM, CEM (EDM, DWDM, EFM) and TEM components. Notice that the relationships of these models are defined through respective mappings. For example, the aggregation of data in DWDM can be mapped to the data entities in the EDM, which in turn are mapped to the physical construct at the TEM level.

Benefits of the Enterprise Modeling Approach

Separation of Conceptual, Physical, Operational and Analytical Models

While the monolithic enterprise data warehouse approach and the bottom-up universal data mart approach provide two extreme views, an optimal data sourcing strategy can be determined based on business and technical requirements. The enterprise model serves as the blueprint governing the relationships between source systems and target data warehouses in the enterprise.

From the operational side, the separation of business processes from the conceptual model is critical. For example, the processes to implement the business function of “managing customer billing inquiry” can be in the form of a call center operation or in the form of a Web based self-service process. Yet, the associated business functionality and data requirements for those processes remain the same. This is also true for the analytical side. Various processes in analytics may be deployed, which may include various decision support methods. However, the associated analytic functionality and data requirements for the various methods remain the same.

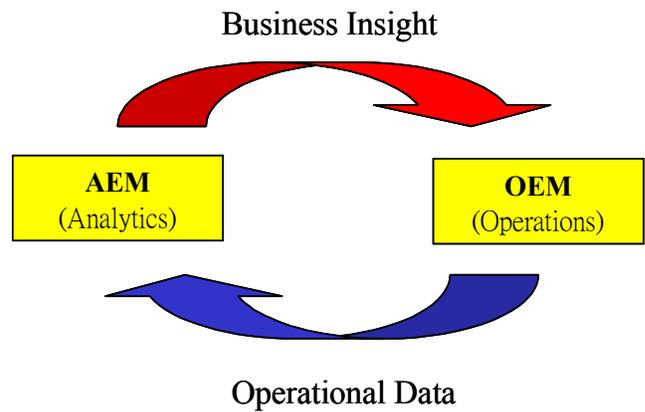


Figure 1. Analytics and Operations

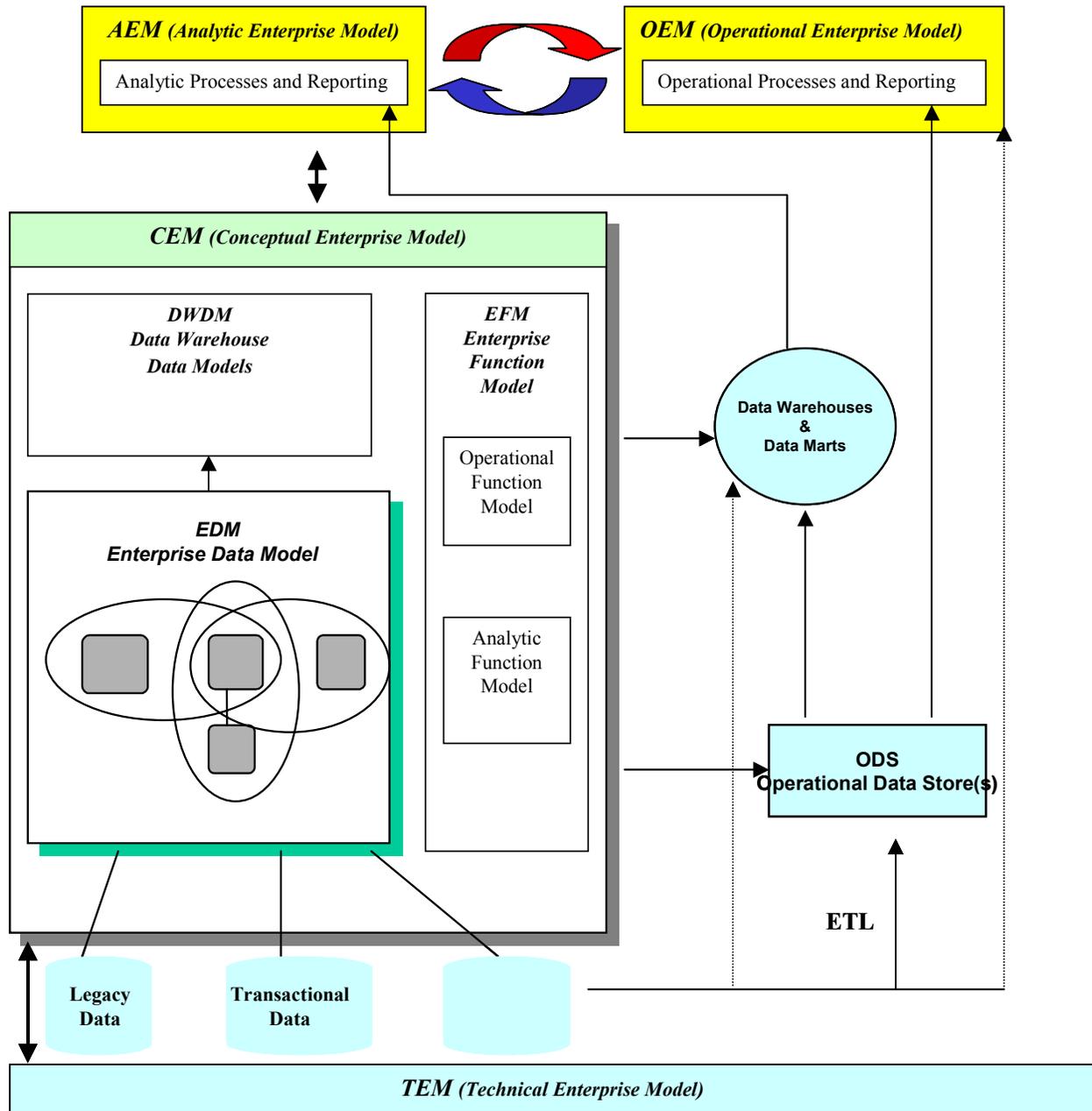


Figure 2. The Enterprise Model Architecture

Furthermore, the separation of the conceptual data from the technical construct is equally important. There should be a unified view of data, irrespective of their physical implementation methods. Therefore, a conceptual data element of a customer may be physically stored in different places, using different formats. They may include an ERP/Oracle/Unix system in New York, a CRM/DB2/Mainframe system in Chicago and a file system in Hong Kong. The integrity and consistency of data can be achieved through the mappings between the three levels of the Enterprise Model.

Clear Differentiation and Collaboration of Analytic and Operational Functions

It may be intuitively clear that data warehouses are not intended for transactional processing. However, for the lack of understanding or for convenience, many businesses are building data warehouses to satisfy immediate transactional reporting

needs. In fact, many IT departments are using data warehousing as an answer to all sorts of data access and reporting requirements. This puts data warehousing in its most ineffective mode of operation. Furthermore, chaos results when every data access and reporting requirement leads to the development of a new data warehouse. The proposed framework clearly separates the two domains. Data warehouses support analytic processes and reporting, whereas the operational data stores support operational processes and reporting. However, the collaboration between the two models can create better insight for the business.

From a Point Solution Focus to an Enterprise Model Driven Methodology

A point solution focuses on solving a narrow isolated problem. Business intelligence today seldom deals with a narrow isolated problem. More often than not, many aspects of the enterprise data are involved for decision support, making predictions and creating useful business insight. The enterprise model provides the framework for enterprise-wide information access.

Consistency of Data Across Data Warehouses

The vast number of disparate data warehouses and data marts built over the past years creates problems in data consistency across the enterprise. Many businesses are busy on consolidation projects. They are attempting to consolidate all these disparate data warehouses and data marts into a “common” data warehouse. The unfortunate reality is that as soon as the consolidation project is complete, the new data warehouse becomes another legacy data warehouse that needs to be maintained, synchronized and consolidated to some other data warehouses as the requirements grow. This is equivalent to putting all different files into one big file in file-processing systems. In fact, total physical consolidation of data warehouses is close to being impossible in any practical sense, especially for large organizations. It also defeats the purpose of subject-orientation for building data warehouses. The enterprise model can provide consistency of data via the mapping of its various components.

Enhancing the Success of New System Development Efforts

Building a new system in an existing enterprise is like building an extension for an existing building. It needs to leverage and integrate the existing infrastructure. For example, a company wants to implement a new CRM package. The CRM package comes with its data schema for “customer” and other data requirements. To implement the package, one needs to build interfaces to the existing ERP and all other legacy, transactional and external systems, each having its own data schema for customers. Each software vendor insists that its customer database is the one that should be used as the “master” database. Complex point-to-point interfaces are required and huge redundancies are created. That’s one of the reasons that most of these implementations are very expensive both in time and costs (typically in the millions of dollars for large companies). The failure rates for these implementations are high. The enterprise model provides the logical framework that leverages the existing information infrastructures. As a result, it reduces development costs and time, while enhancing the quality of the new systems. This applies to building new data warehouses as well as transactional systems.

Business Benefits

Two critical criteria dominate the information requirements of today’s businesses.

- a) Provide real-time access to accurate information across the enterprise to any authorized individual, at any time and any place.
- b) Create business intelligence and insight from the mountains of data collected at the transactional level.

The first criterion is transactional in nature and the second analytical. The enterprise model addresses both of these requirements in an integrated architecture. While the traditional business performance indicators evolve around internal efficiencies, the new economy requires businesses to extend the value system to their customers, suppliers and partners. Point solutions seldom provide the optimization required at the enterprise level. The following provides two examples that can benefit from the enterprise modeling approach.

Enhancing Customer Relationship Management

Two major areas of specific concerns regarding customer relationship management are being addressed in today’s businesses: creating customer insight and improving customer services. As pointed out by Todman (2001), the secret to customer relationship

management is “to know who our customers are and what it is that they need from us”. Companies are embarking on initiatives to collect and analyze various customer information so that they can better profile and classify customers, predict customer behavior, conduct target marketing, cross and up sell into existing customer base. Information about customers can span across all areas of an enterprise through many different business processes and involving many business units. Information collected through transactions in conjunction with external data can be used for various analytic processes to create customer insight enabling customer relationship processes. One of the issues many companies have to deal with is to make sense out of the islands of customer data marts to provide such intelligence. The enterprise model provides an integrated view of customers supporting various customer relationship management processes.

Enhancing Supply Chain Management

As we move from the production economy into the digital economy, the focus of manufacturing efficiency is shifted to effective supply chain management. Information sharing between various entities in the supply chain is a critical element of successful supply chain management. A notable example as pointed out in Turban et al. (2002) is Wal-Mart’s continuous replenishment. In this example, Wal-Mart the retailer shares inventory information of each Procter & Gamble (P&G) item in every store with P&G the supplier. P&G can then replenish automatically when a certain inventory threshold is reached. Furthermore, such information collected at the point of sales can be used in analytics in creating business insight for both retailers and suppliers in many areas including demand and production forecasts, product and facility planning. The enterprise model provides an integrated view for all the business entities in the supply chain.

Methodology

The following provides key activities of building data warehouses using the enterprise modeling approach. For this discussion, we assume that the Enterprise Data Model (EDM) is already developed. Here we focus on the key activities required for the development of data warehouses using the framework of the enterprise model. Some of the other required tasks such as cost-benefits analysis, product evaluation, migration planning, training ...etc. are not included.

Project Initiation and Scope Definition

The mission of the data warehouse project is confirmed with senior management, and executive sponsors are identified. Project scope is confirmed and project organization is defined. Project plans are developed and resources for the project are committed.

Identify Business Objectives and Key Performance Indicators

Business goals are clearly defined in measurable terms so that the return on investment can be measured. The key performance indicators can also provide the metrics for the design of the Analytic Enterprise Model (AEM).

Design the Data Warehouse Data Model and Analytic Functions

The Data Warehouse Data Model (DWDM) is developed based on the business requirements using the construct of the Enterprise Data Model (EDM). Techniques in dimensional modeling can be used in this phase. Analytic functional requirements are defined in the Enterprise Functional Model (EFM). The mappings between the EDM, DWDM and the EFM are constructed.

Design and Develop the Operational Data Stores

The number of Operational Data Stores (ODS) is determined based on requirements in the model taking into consideration the number, size and relationships of source systems and target data warehouses. They are then developed based on the construct of the Enterprise Data Model (EDM) and the target database technologies. This phase includes tasks in the design and development of the target databases. Routines for the extraction, transformation and loading of source data to the target databases are designed, developed, tested and implemented. The mappings between the EDM and the ODS are constructed.

Design and Develop the Data Warehouse

The physical data warehouse is designed and developed according to the Data Warehouse Data Model (DWDM) and the target data warehouse technologies. This phase includes the tasks in the design and development of the target data warehouse. The Operational Data Stores (ODS) provide the data feed into the data warehouse. Routines for the extraction, transformation and loading of ODS data to the target data warehouse are designed, developed, tested and implemented. The mappings between the ODS and the data warehouse are constructed. In the event that no ODS is used, the extraction, transformation and loading of data will be directly from the source systems to the target data warehouse. The mappings between source systems and the data warehouse are constructed.

Design and Develop the Analytic Processes

The analytic processes at the AEM level are designed and developed. This may include methods using mathematical models, decision support processes, multi-dimensional analytics, data mining and other methods of data analysis. Observe that various methods can be deployed to implement an analytic function. This phase includes the tasks of designing, coding, testing and implementation of respective software programs, which may be package or custom oriented. The mappings between the AEM and EFM are constructed.

Develop the Information Roadmap

The relations between the various components of the AEM, OEM, CEM (EDM, DWDM, EFM), and TEM in the enterprise model are developed. These mappings are created incrementally as each component is developed. They also serve to provide the cross check of the various models for accuracy and completeness.

Conclusion

The digital age has created a dilemma for businesses in determining how much data is enough data. It is no longer the case that there is not enough data out there. On the contrary, there's too much data. To find useful information in these mountains of data is like looking for a needle in a haystack. The challenges are both on the operational side as well as on the analytical side. On the operational side, information access across the entire enterprise based on real-time business requirements becomes a necessity. On the analytical side, one needs to sift through the mountains of data to find relevant information for decision support, making predictions, finding patterns, and creating business insight. Furthermore, technology vendors' one-size fits all approach contributes to the confusion. Short cuts and immediate gratification of point solutions also contribute to the overall information chaos. It is important for businesses to develop a sound information strategy before the next major data warehousing initiative. The enterprise model proposed in this paper provides the information roadmap for the deployment of data warehouses for an enterprise. It serves as the architectural blueprint to create adaptive enterprise-wide information technology architectures to enable business models in the information age.

References

- ANSI/X3/SPARC, Data Base Management Systems. Interim report. FDT (ACM SIGMOD bulletin) 7, No.2, 1975.
- Chan, J., The Enterprise Model – An Enterprise Information Management Roadmap, 2003 Proceedings of American Society of Business and Behavioral Sciences, Volume 10 Number 1, February 2003, pp. 197–201.
- Chen, P.P., The Entity-Relationship Model: Toward a Unified View of Data, ACM Transactions on Database Systems 1, no. 1, March 1976, pp. 9-37.
- Inmon, W. H., Building the Data Warehouse, John Wiley & Sons, New York, NY, 1993.
- Simon, A., 90 Days to the Data Mart, John Wiley & Sons, New York, NY, 1998.
- Todman, C., Designing a Data Warehouse Supporting Customer Relationship Management, Prentice Hall PTR, Upper Saddle River, NJ, 2001.
- Turban, E., King, D., Lee, J., Warkentin, M., and Chung, H.M., Electronic Commerce 2002, A Managerial Perspective, Prentice Hall, Upper Saddle River, NJ, 2002, pp. 643-644.