# Communications of the IIMA

2005

# Optimizing Data Warehousing Startegies

Joseph O. Chan
*Roosevelt University*

Follow this and additional works at: http://scholarworks.lib.csusb.edu/ciima

Part of the Management Information Systems Commons

## Recommended Citation

# Optimizing Data Warehousing Strategies

### Joseph O. Chan

Roosevelt University, 1400 North Roosevelt Boulevard, Schaumburg, Illinois 60173
Phone: (847) 619-7306, Fax: (847) 619-4852, jchan@roosevelt.edu

## ABSTRACT

*Database technologies have evolved over the last two decades into different constructs to support the ever-growing information needs for organizations spanning the spectrum of operational and analytical processing. This paper examines the characteristics of transactional databases, operational data stores, data warehouses and virtual data warehouses. A framework is developed for an optimal data warehousing strategy based on organizational needs classified by the types of business processes defined by the requirements of supporting functional areas and the levels of decision structures. Enterprise architecture is described to provide an integrated and complementary view of various data warehousing constructs.*

Keywords: Data warehousing strategy, operational data stores, data warehouses, virtual data warehouses.

## INTRODUCTION

As data warehousing technologies matured over the past decade, organizations continue to build various data warehouses and data marts to meet the needs of integrating and consolidating information across the business enterprise. While data warehousing is used over a wide range of applications for analytical processing and reporting at various management decision levels, it is not a suitable structure to support operational processing and reporting. Day-to-day operations of a business are supported by transactional databases. Operational data stores are the means for data integration and consolidation across transactional systems to support enterprise-wide operational processing and reporting needs. The demand of real-time access to enterprise-wide data has caused the worlds of real-time, near-real time and batch analytics to come closer together than ever before. There is a resurgence of virtual data warehousing which allows queries to run against distributed operational data sources in real-time bypassing the physical constructs of data warehouses and operational data stores.

While any of these data constructs can theoretically serve as the model to satisfy most organizational information needs, each has its own characteristics that will perform optimally or sub-optimally under various situations. Therefore, in spite of marketing hypes from technology vendors, one size does not fit all. These technologies should be considered as complementary to each other and should be used according to the requirements that best utilize their capabilities. This paper provides a classification of the characteristics of different database constructs for Transactional Databases (TDB), Operational Data Stores (ODS), Data Warehouses (DW) and Virtual Databases (VDB). Virtual databases will categorically cover the constructs of Virtual Data Warehouses (VDW) and Virtual Operational Data Stores (VODS). Different types of analytics will be explored. They include real-time, near real-time and batch analytics supporting tactical and strategic decision making. The paper will further provide a description of the dimensions of organizational information needs based on the classifications of business processes according to the level of decision structures and the functional support across the business enterprise. Four types of business processes will be discussed: the uni-decisional functional processes, the uni-decisional cross-functional processes, the multi-decisional functional processes and the multi-decisional cross-functional processes. The selection of one or more of the four database constructs will be made according to the respective requirements of each type of business process. A conceptual architecture via an enterprise model is described to provide an integrated framework for transactional databases, operational data stores, data warehouses and virtual databases.

## AN OVERVIEW OF DATABASE CONSTRUCTS

*Transactional Databases*

Transaction processing systems (TPS) support the day-to-day operations of an enterprise such as order entry, machine control, accounts payable and accounts receivable. TPS are further classified by batch processing and online transactional processing (OLTP). Most transactional systems today are OLTP systems where users interact

with the computer in real-time. The database structures that support TPS are called transactional databases or operational databases. As described by Orr (2000), operational databases provide an efficient processing structure for a small number of well-defined business processes.

*Operational Data Stores*

While transactional databases focus on supporting the transactions of well-defined and targeted business processes, they are not equipped to support business processes and reporting requirements that span across multiple transactional systems. For example, the adjustment of credits for returned goods within a billing cycle spans across the order fulfillment and accounting systems. The systems may be reconciled through manual or semi-automated processes which can be time consuming and error-prone. Consequently, the adjustment may not be made until a few billing cycles later. Furthermore, if management wants to track returned goods by sales channels and manufacturers, the databases for sales, suppliers and order fulfillment need to be consolidated and synchronized to provide the necessary information. There is a need for various disparate transactional databases to be synchronized and coordinated as close to real-time as possible. The Operational Data Stores (ODS) provide the means of consolidating transactional databases, reconciling the differences in data structures, formats and semantics from different source systems. Sperley (1999) characterized ODS as containing only current or nearly current data typically used by operational personnel to respond to the daily needs of a business. Navas (2005) also pointed out that ODS are used for running day-to-day operations and are updated on the order of minutes or hours, and typically contain normalized data with very little summarization. As illustrated in Figure 1 page 3, besides transactional data, other data sources for the ODS may include legacy data and external data. For example, an ODS supporting a direct mail operation may require current sales information of active customers from transactional systems, inactive customer lists from legacy systems, and third party prospect lists from external sources.
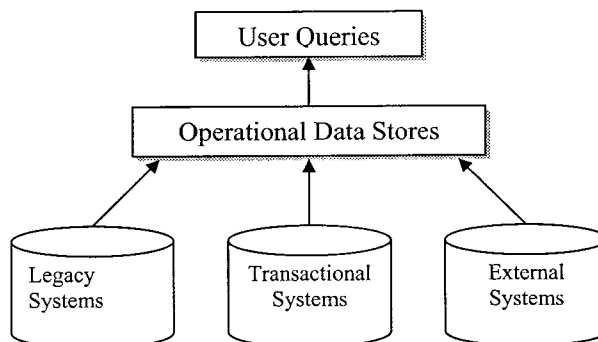


Figure 1: Operational Data Stores.

*Data Warehouses*

While operational data stores address the integration issue of disparate transactional databases, they are not suitable for analytic processing to support strategic decision-making. According to Singh (1998), tactical decisions are based largely on data in the ODS, whereas, long-term strategic decisions require the historical trend analysis that is only available from the data warehouse. It further stated that an ODS does not need to have the historical data that a data warehouse must store for use by strategic decision makers. Data warehouses have certain characteristics that differentiate them from operational databases. The classic definition by Inmon (1993) described these characteristics as subject oriented, nonvolatile, integrated and time-variant. In the following, we shall point out the differences between the two constructs. It should be noted that the properties of data warehouses discussed below are applicable to data marts, which are subsets of the enterprise-wide data warehouse at the departmental or functional levels.

Subject Orientation: Transactional systems are designed around specific business applications and are not subject oriented. Operational data stores, which consolidate multiple transactional databases, provide an integrated view across multiple business applications. Neither of these constructs is designed around subject areas. For example, operational databases that support sales transactions, marketing events and customer services may not provide the aggregated and summarized customer-centric views along the subject area of sales resulting from a customer's response to promotional programs. The construct of data warehouses is designed for subject-oriented views of data.

Non-Volatility: Data warehouses are characterized as non-volatile. Volatility is the degree by which the contents of a database can change with respect to time. Operational databases are volatile. Their contents are subject to change in real-time or in near real-time. The volatility of operational data may cause temporal results in queries, in that the same query run at different times may cause different results. While volatility is acceptable at the operational level where real-time and near real-time data is most relevant, it is not suitable for analytical process for strategic decision support.

Integration: Overtime, disparate transactional systems supporting various functional areas have become information silos, separated by system and organizational boundaries. Redundant and inconsistent data are embedded deeply in these functional silos prohibiting information sharing and optimization across the enterprise. While both operational data stores and data warehouses integrate data from different sources, they address different purposes utilizing the integrated database. In the case of the operational data stores, the integration of data is around business applications for operational purposes. For example, the data of returned goods is integrated with data in accounting in an ODS for the accurate and timely processing of credits. In the case of data warehouses, the integration of data is for analytical purposes. For example, a customer oriented data warehouse may integrate and aggregate data that pertain to a customer's interactions with a firm along various subject areas for the purpose of developing effective customer relationship management strategies.

Time-variance: As described by Todman (2001), the presence of time and the dependence upon it is one of the things that sets data warehouse applications apart from operational systems. While operational systems operate at the present environment where date and time are merely descriptive attributes, data warehouses contain historic data that are time specific. Marakas (2003) characterized the time horizon for data in the warehouse to be long (from 5 to 10 years) as compared to a much shorter time horizon in the operational environment (from 60 to 90 days). A characteristic for strategic decision support is the capability of performing analytic processing of historic data over a long time horizon. One of the reasons that operational databases are not designed for huge amount of historic data is system performance. Todman (2001) articulated the issue of system performance degradation by holding historic data in operational databases. Unlike the near real-time nature of ODS, data warehouses typically are batch-oriented with updates occurring on the daily basis versus hours.

De-Normalization: Transactional databases and operational data stores are optimized for operational data processing. They are normalized to avoid various kinds of processing and updating anomalies. Operational databases are designed to support day-to-day tactical decision making over a wide area of applications. On the other hand, data warehouses are optimized for analytical data processing to support strategic decision making. Data coming from different sources can be aggregated and summarized by subject areas and time dimensions in a data warehouse. The database design for data warehouses can be de-normalized to optimize system performance. De-normalization may also reduce the number of joint paths. Todman (2001) described the possibility of multiple join paths in operational databases for which different decisions on the paths of joining multiple tables may produce different results, whereas only one join path is allowed for decision support databases.

## VIRTUAL DATA WAREHOUSES, VIRTUAL OPERATIONAL DATA STORES

The concept of virtual data warehouses (VDW) gained popularity in the 1990's through the use of middleware technologies, where queries are run against distributed operational data sources in real-time bypassing the physical constructs of data warehouses and operational data stores. Eckerson (2003) indicated that the virtual data warehouse concept failed in that period due to operational system performance degradation issues and the belief that analytical processing systems require different architectures other than the operational systems. There is a resurgence of virtual data warehousing under the new name of Enterprise Information Integration (EII) which is positioned as a complement to data warehouses rather than replacing them (Eckerson 2003). As described by Russom (2003), EII is an integration platform that supports application access of virtual views from multiple data sources, and handles the connectivity with back-end databases and applications. Sperley (1999) described the virtual data warehouse as a decision support system that sends queries to the actual operation systems. Ankorion (2004) characterized "virtual data federation" as to provide access to "joined" views of data from multiple, heterogeneous sources, while leaving the actual data in place. Virtual data warehouses allow the real-time processing of queries to distributed data sources without the replication of data into another physical database. While the virtual data warehouse architecture emphasizes the perspectives of the virtual database and real-time queries to operational systems, its "warehousing" perspectives are different from the traditional data warehousing principles. Virtual data warehouses deal with real-time operational data and function more like virtual ODS (VODS), although the term has not been commonly used in the literature until recently (Navas 2005, MacVittie 2004, and Inmon 2004). While there may be technical

architectural differences between virtual database (VDB) products including virtual data warehouses (VDW), virtual data marts (VDM) and virtual operational data stores (VODS), the conceptual architecture that utilizes virtual views to support the application queries is common (Figure 2). Henceforth, the terms VDB, VDW, VDM and VODS will be used interchangeably.
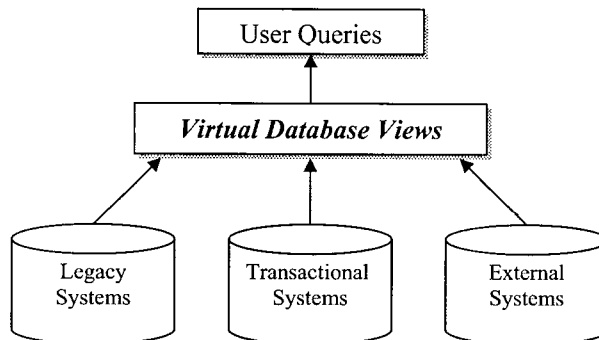


Figure 2: Virtual Data Warehouses and Virtual Operational Data Stores.

The obvious benefits of virtual data warehouses are savings in costs and efforts from not having to develop and maintain physical databases in ODS and data warehouses. Russom (2003) explained that the EII approach only moves data when an application requests it, avoiding the moving of excessive data needlessly. Thus, it helps to reduce loads on systems and networks. However, there are disadvantages. System performance of the operational systems can be degraded due to complex table joins. As described in Sperley (1999), the queries placed on the transactional systems may result in significant performance degradation in those systems. Due to different maintenance procedures in various operational systems, data may not be synchronized in real-time across various source systems. Sperley (1999) indicated that it is possible for a virtual warehouse query to produce different results at different times depending on the data archive schedule of the legacy systems. Virtual databases are not suitable choices for strategic analytic processing due to the issues regarding data volatility and the lack of subject oriented historical data in operational databases. Consider the analytical process of determining a cross-selling strategy that require years of customer sales data aggregated along different dimensions so that various analytic models and what-if analysis can be applied. To pull real-time data from distributed operational data sources with incompatible data structures to support such analysis would require very complex queries. The performance of the operational systems will be drastically degraded with these repeated queries.

Virtual databases can be used to support real-time operational processes and tactical analytics, particularly for unplanned and non-repetitive queries that require up-to-the-minute information from distributed data sources. For example, a salesperson upon making a customer-facing event may want to know the most recent interactions of the customer with the firm from systems supporting marketing, sales and customer services. Hotel operations may benefit in knowing the most recent interactions of a guest with the hotel, from systems supporting customer services, hotel reservations, membership events, and online surveys, so better services can be provided while the guest is still staying in the hotel. Operations can be greatly improved utilization up-to-the-minute information from VDB and historic information from data warehouses. As a complement to and in conjunction with data warehouses, virtual databases can be effective tools to support real-time tactical analytics.

## CHARACTERISTICS OF ANALYTICS

*Tactical versus Strategic Analytics*

According to the Merriam-Webster Online Dictionary (2005), analytics is defined as the method of logical analysis and operation as the performance of a practical work or of something involving the practical application of principles or processes. As pointed out by Chan (2004), there are clear distinctions between the data structures supporting operational processing and analytical processing. Chan further pointed out the symbiosis between these two computing environments in a feedback loop. Eckerson (2003) described the phenomenon of convergence of these two environments around real-time data warehousing and business performance management, where

Enterprise Information Integration (EII) tools allow firms to capture and integrate historical, transactional and external data. It should be noted with caution that while the two concepts are inter-related, their distinct characteristics warrant specific architectural design considerations in each of the respective environments. The world of analytics can be divided into different categories depending on the levels of decision-making they support. Analytics can support tactical decision making at the operational level or strategic decision making at the management level. For example, analytics at the tactical level can help to determine what products are in stock, based on inventory information and information in customer orders such as products, quantity, and delivery dates. Analytics at the strategic level can help to determine cross-selling strategies based on demographics, market segmentations and customers' buying behaviors over a long time-span of historical data. Analytical methods and models, and their data requirements are different for each of these decision support levels. We shall use the term "tactical analytics" for analytical methods supporting tactical decision making, and "strategic analytics" for analytical methods supporting strategic decision making. Tactical analytics deal with situations at the moment, whereas strategic analytics deal with situations in the future utilizing models in forecasting, projections and predictions. The data structures suitable for tactical analytics are transactional databases, operational data stores, and virtual databases. The data structures suitable for strategic analytics are data warehouses and data marts.

*Planned versus Unplanned Analytics*

Analytics can also be characterized by the types of queries according to whether they are planned or unplanned. As described by Bednarz (2004), data warehouses are best suited for planned analytic queries that are executed on a regular basis, whereas data integration tools (virtual databases) are suited for queries that are unplanned and not repeated.

*Real-time versus Batch Analytics*

Real-time processing is a relative concept. What constitutes real-time depends on the perception of the user and in how fast the user can absorb the information. Reimers (2003) pointed out that there may be no reason to use real-time analytics if a company can absorb its transaction information only on an hourly, daily or weekly basis. Reimers also described the characteristics of real-time analytics based on data collected within the past hour, and that of near real-time analytics conducted on data collected within the past 24 hours. Analytics based on data collected for over 24 hours are typically considered as batched. Thus, transactional databases and virtual databases support real-time analytics, whereas operational data stores support near real-time analytics and data warehouses support batch analytics. As pointed out by Reimers (2003), not every business process requires real-time data and that real-time data collected at different times may be inconsistent with each other and may not lend to better decision making. Therefore, it is important to understand under what situations that real-time analytics are required and feasible. Typically, tactical analytics requires real-time or near real-time data, whereas the data supporting strategic analytics can be batched.

## A SUMMARY OF DATABASE CHARACTERISTICS

As discussed in the previous sections, the suitability of transactional databases, operational data stores, data warehouses, and virtual databases depends on respective requirements of the queries supporting different types of analytics. Each of these data structures comes with its own set of characteristics that are optimal for certain types of applications and sub-optimal for others. Table 1, provides a summary of these database characteristics in ten categories that include subject orientation, data volatility, integration, time-variance, normalization, real-time versus batch analytics, historic data, summarized data, data sources and level of decision support.

|  | Transactional Databases (TDB) | Operational Data Stores (ODS) | Data Warehouses (DW) | Virtual Databases (VDB) |
|---|---|---|---|---|
| *Subject Orientation* | No | No | Yes | No |
| *Data Volatility* | Yes | Yes | No | Yes |
| *Integration* | Not Integrated across applications | Integrated | Integrated | Virtually Integrated |
| *Time-variance* | No | No | Yes | No |
| *Normalization* | Normalized | Normalized | De-Normalized | Typically Normalized Meta Models |
| *Real-time vs. Batch Analytics* | Real-time | Near Real-Time | Batch | Real-time, Near Real-time |
| *Historic Data* | Short Time Horizon (60-90 days) | Short Time Horizon (60-90 days) | Long Time Horizon (5-10 years) | No Physical Storage |
| *Summarized Data* | Very Little | Very Little | Yes | No physical storage |
| *Data Sources* | Application Specific | Cross-applications, distributed data sources | Cross-applications, distributed data sources | Cross-applications, distributed data sources |
| *Tactical vs. Strategic Decision Support* | Tactical | Tactical | Strategic | Tactical |

Table 1: Database Characteristics for Analytical Support.

## THE DIMENSIONS OF ORGANIZATION INFORMATION NEEDS

With many alternatives in how enterprise data can be represented, an organization needs to understand its business requirements in order to develop an optimal information strategy. In the following, we shall examine organizational information needs based on the levels of decision structures and the types of business processes.

*Levels of Decision Structures*

An organization can be viewed as a body of decision structures. Information needs can be identified by the levels of decision structures they support within and across functional areas. The levels of decision structures can be classified as operational, knowledge, middle management and executive management; based on the types of work performed (Laudon et al., 2004). Operational level decision structures support an organization's day-to-day activities. The knowledge level decision structures support the activities of knowledge workers. The middle management level decision structures support the management activities in planning and control. The executive management level decision structures support the activities of executives in the discovery of problems and opportunities, and making strategic decisions dealing with them. Simon (1960) characterized the types of decisions defined by the degree to which they are programmed. Thus, highly structured decision processes are programmed and highly unstructured decision processes are non-programmed. Semi-structured decision processes are in between these two ends of the spectrum. As described in Laudon et al. (2004), the operational level deals with structured decision processes, the middle management level deals with structured to semi-structured decision processes, the knowledge and executive levels deal with unstructured decision processes.

*Operational Level*

The operational level decision structure deals with the day-to-day activities of an organization supporting business transactions and events in all functional areas including finance and accounting, operations, marketing and sales, and human resources. The operational level decision structure is supported by Transactional Processing Systems (TPS). Data supporting the operational level are transactional and referential in nature. For example, data elements in a customer order such as the product and quantity are transactional data supporting the activities in order entry. A pricing reference table may be used to compute the total price for the order. Here, the data regarding the order are transactional and the pricing information is referential. For OLTP, transactional data are updated in real-time.

*Knowledge Level*

The knowledge level decision structure deals with the creation, capturing and codifying, and sharing of knowledge in an organization. The systems supporting these knowledge level activities are Knowledge Work Systems (KWS), Knowledge-based Systems (KBS) and Knowledge Management Systems (KMS).

Knowledge Work System: Laudon et al. (2004) described KWS as systems supporting highly specialized fields. Examples of KWS include Computer Aided Design (CAD) systems supporting engineers and investment workstations supporting financial specialists. Data supporting the knowledge work level consists of operational data and the rules in the knowledge base. For example, data supporting the CAD system consists of the bill of materials information and the rules of associations between different components. Knowledge work data can be a combination of real-time and historical data from internal and external sources. For example, data supporting investment workstations can integrate a wide range of data from both internal and external sources, including real-time and historical market data, and research reports (Laudon 2004).

Knowledge-based Systems: KBS are also known as Intelligent Systems. They deal with the decision structure that uses qualitative knowledge, rather than mathematical models, utilizing technologies in artificial intelligence (Turban et al. 2005). Systems in this category include expert systems, natural language processing, robotics, neural networks, genetic algorithms and various types of intelligent agents. Examples of KBS include medical diagnostic systems, credit analysis systems and production planning systems. Turban et al. (2005) described that the potential sources of knowledge include human experts, textbooks, multimedia documents, public and private databases, special reports and information from the Web. It can integrate a wide range of data from both internal and external sources, including real-time and historical data. Klebe (1998) described online credit fraud detection systems as requiring a detailed transactional history database coupled with a real-time artificial intelligence scoring system.

Knowledge Management Systems: Turban et al. (2005) described knowledge management as a process that helps organizations identify, select, organize, disseminate, and transfer important information and expertise that are part of the organization's memory and that typically reside within the organization in any unstructured manner. Turban et al. (2005) further described KMS as combination of technologies in communications, collaboration, storage and retrieval. Data supporting knowledge management in the discovery and codification of knowledge may include internal and external, structured and unstructured data from many difference sources and media.

*Middle Management Level*

Primary middle management functions include planning and control. Planning involves the development of programs and strategies for the future. Control involves the optimal acquisition and deployment of resources for the efficient and effective execution of the plans.

Management planning activities may include budgeting, demand forecasting, production scheduling, trend analysis, developing cross-selling and up-selling strategies, etc. Various analytical methods can be deployed utilizing statistical and other quantitative models, techniques in data mining and OLAP. Data supporting the analytic processes for management planning are optimized for the respective analytic methods. They may include historic, multi-dimensional and summarized data from internal and external sources. For example, data supporting the development of a cross-selling strategy may come from sales transaction histories by product line, by customer segment and by region as well as information from external sources regarding customer behaviors along market segments. Systems supporting the analytic aspects of management are called Decision Support Systems (DSS).

Management control activities may include the monitoring of actual versus planned outcomes, dealing with exceptions, acquisition and allocation of resources. Data supporting management control are real-time or near real-time summarized operational data. For example, retail managers may want to know sales data by store, by product

line and by time period in order to monitor actual versus planned revenue. These data are summarized from tens of thousands of sales transactions at many stores. Systems supporting the middle management level are called Management Information Systems (MIS).

*Executive Management Level*

Watson et al. (1993) described executive work activities as diverse, brief and fragmented. They are more unstructured, non-routine, and long-range in nature. Turban et al. (2005) described the executive decision roles as identification of problems and opportunities, and making decisions on what to do about them. They differ from other managerial decision roles that focus on the analysis of specific problems and opportunities. Systems supporting the executive level decision structure are called Executive Support Systems (ESS). Data supporting the executive level include aggregate historic operational data, results from analytical processes and external environmental data. For example, to determine what new products to develop, operational data from production, analysis from marketing regarding customer needs and competitive information from external sources are required. Furthermore, meaningful details are required to support the drill-down of any summarized information (Turban et al., 2005).

## TYPES OF BUSINESS PROCESSES

Information needs for an organization can be identified by their support for business processes. Business processes can be defined within or across functional groups (vertical organizations) such as sales, marketing, production, finance, accounting and human resources. For example, the quality assurance and line balancing processes are defined within the manufacturing function. The change order process that spans multiple functional areas in sales, manufacturing, shipping and accounting represents a cross-functional business process. Traditionally, companies organize themselves by functional groups. Information silos are created overtime confined by organizational boundaries. In order to optimize performance at the company level, a business needs to design effective processes across functional areas. As indicated by Harrington (1991), most processes do not flow vertically. Harrington further described the issues with horizontal work flows across vertical organizations that result in voids and overlaps and sub-optimization, negatively impacting the efficiency and effectiveness of the process.

Business processes can also span across the spectrum of decision structures. Processes defined within a decision structure will be called uni-decisional, whereas processes that span across multiple decision structures will be called multi-decisional. Examples of uni-decisional processes include point of sales (operational level), product design (knowledge level), inventory control (management control level), cross-selling strategy development (management planning level) and long-term revenue forecast (executive management level). A marketing campaign is a multi-decisional process that includes both analytical and operational decision structures. It requires the cross-selling strategy from management planning and the operations of conducting marketing events.

Business processes can be manual, or supported by one or more systems. Conversely, a system may support one or more business processes. For example, a change order process may be supported by systems in Sales, Order Processing, Production, Shipping and Invoicing. Conversely, these systems may support multiple vertical and horizontal processes. As indicated by Davenport (1993), strategic systems such as the Airline Reservation System are highly cross-functional, involving many aspects of operations. These systems come with the associated data files and databases tailored to the requirements of the respective applications. Furthermore, processes can be real-time, near real-time or batched. Optimal strategies in process design should consider the applicability of real-time versus other criteria such as the burden on computing resources, and absorption of data for real-time information for practical use. Having up-to-the-minute data may not be relevant for many business processes. The airline reservation process as noted above is real-time, based on data collected within minutes. The process of inventory replenishment based on demand forecasts, quantity on hand and usage can be near real-time, based on data collected within hours. The process of determining cross-selling strategies in a market segment utilizing data mining and statistical models over years of sales data along multiple product lines and demographics can be batched, based on data collected over 24 hours.

## ORGANIZATIONAL INFORMATION REQUIREMENTS BY BUSINESS PROCESSES

According to Harrington (1991), there are hundreds of business processes going on every day; over 80% of them are repetitive. Thus, the nature of information requirements for the majority of the business processes can be well defined. In the following, we describe the characteristics of the information requirements of an organization by the type of processes classified by the support of various business functions and decision structures.

*Uni-decisional Functional Processes*

Uni-decisional functional processes are defined by their level of decision structure in a functional area. For example, the accounts payable process in accounting is at the operational level. The engineering design process in manufacturing is at the knowledge level. The process of monitoring revenue by stores and by product lines in sales is at the management control level. The process of profitability forecasting in finance is at the management planning level. The process of developing a long-term product strategy in marketing is at the executive management level. The information requirements are determined by the requirements of the respective decision structure and functional area. At the operational level, the processes are supported by Online Transactional Processing (OLTP) systems that require real-time transactional databases. At the knowledge level, the processes are supported by Knowledge Work Systems (KWS), Knowledge-based Systems (KBS) or Knowledge Management Systems (KMS) that may require data from transactional databases, operational data stores, virtual databases, data marts, external data sources and the associated knowledge bases. At the management control level, the processes are supported by Management Information Systems (MIS) which produce summarized and exceptional reports from operational databases which may include transactional databases, operational data stores and virtual databases. At the management planning level, the processes are supported by Decision Support Systems (DSS) that require aggregate data in data marts from internal and external sources, with the associated model bases. At the executive level, the processes are supported by Executive Support Systems that require aggregate data in data marts from internal and external sources, with supporting detailed information for drill-down. The different levels of decision structures are characterized by the degree of real-time requirements. Notice that the source systems for VDB, ODS and data marts under this category of uni-decisional functional processes are within a functional area. Table 2 provides a summary.

| Uni-decisional Functional Process | System Support | Data Support | Real-time vs. Batch | Internal vs. External Data |
|---|---|---|---|---|
| *Operational* | TPS (OLTP) | Transactional databases | Real-time | Mostly Internal |
| *Knowledge* | KWS, KBS, KMS | Transactional databases VDB *(functional)* ODS *(functional)* Data Marts Knowledge Bases | Real-time Real-time Near Real-time Batch | Internal & External |
| *Management Control* | MIS | Transactional databases VDB *(functional)* ODS *(functional)* | Real-time Real-time Near Real-time | Mostly Internal |
| *Management Planning* | DSS | Data Marts Model Bases | Batch | Internal & External |
| *Executive* | ESS | Data Marts Drill-down | Batch | Internal & External |

Table 2: Data Requirements for Uni-Decisional Functional Processes.

*Uni-decisional Cross-functional Processes*

Uni-decisional cross-functional processes are defined by their level of decision structure across multiple functional areas. For example, the process of managing a change order spans across multiple functional areas at the operational decision structure level. Turban et al. (2005) described a knowledge management process supporting the call center operations that spans across multiple functional areas. Subject matter experts from various departments are involved in building the knowledge base that can provide the right answers to questions asked by millions of customers. Knowledge decision structures in sales and marketing, finance, and manufacturing may provide answers that require the knowledge of customer profiles, credit-worthiness, and product design. At the management control level, the process of monitoring order statuses may require the management control in different departments such as sales, order processing, production, shipping and billing. At the management planning level, the process of production

planning requires demand forecasts in sales and marketing, material resource planning in manufacturing and procurement planning in purchasing. At the executive level, the process of organization right sizing requires the executive support in finance for the projection of financial results, in sales and marketing for the projection of new markets and opportunities, in operations for the projection of store performances, and in human resources for the projection of skill requirements. The degrees of real-time criteria for data supporting these processes are the same as the uni-decisional functional processes decided by the level of decision structures. The major difference is in the sourcing of data from multiple functional areas for the cross-functional processes. Operational data stores and virtual databases are suitable data structures to provide the integration of real-time to near real-time data from distributed sources. Data warehouses are suitable data structures for supporting strategic decision making across multiple functional areas. Notice that the source systems for VDB, ODS and data warehouses under this category of uni-decisional cross-functional processes may span across multiple functional areas. Table 3 provides a summary.

| Uni-decisional Cross - Functional Process | System Support | Data Support | Real-time vs. Batch | Internal vs. External Data |
|---|---|---|---|---|
| *Operational* | Multiple TPS (OLTP) | Transactional databases VDB *(cross-functional)* ODS *(cross-functional)* | Real-time Real-time Near Real-time | Mostly Internal |
| *Knowledge* | KWS, KBS, KMS | Transactional databases VDB *(cross-functional)* ODS *(cross-functional)* Data Warehouses Knowledge Bases | Real-time Real-time Near Real-time Batch | Internal & External |
| *Management Control* | MIS | Transactional databases VDB *(cross-functional)* ODS *(cross-functional)* | Real-time Real-time Near Real-time | Mostly Internal |
| *Management Planning* | DSS | Data Warehouses Model Bases | Batch | Internal & External |
| *Executive* | ESS | Data Warehouses Drill-down | Batch | Internal & External |

Table 3: Data Requirements for Uni-Decisional Cross-Functional Processes.

*Multi-decisional Functional Processes*

A multi-decisional functional process is defined within a functional area but spans across multiple decision structures. For example, the process of production requires the management planning decision structure for scheduling and resource planning, the management control decision structure for inventory control and the operational decision structure for manufacturing. The types of systems and respective data support are combinations of the requirements in uni-decisional functional processes depending on the levels of decision structures required.

*Multi-decisional Cross-functional Processes*

Multi-decisional cross-functional processes span across multiple decision structures and multiple functional areas. For example, the process of a 1-1 marketing campaign requires the management planning decision structure for cross-selling strategies, the operational decision structure for direct mail and telemarketing, and the management

control decision structure to monitor customer responses. It spans across multiple functional areas in marketing, sales and customer services. The types of systems and respective data support are combinations of the requirements in the uni-decisional cross-functional processes depending on the levels of decision structures and the respective functional areas required.

## AN OPTIMAL STRATEGY

Theoretically, any of the data structures described above can be used in most if not all situations, albeit not optimally. Operational databases are not suitable for storing a large amount of summarized and historic data over a long time horizon. Data warehouses are not suitable for the processing of real-time data. Virtual databases can place heavy burden on operational systems with complex end user queries. The effectiveness of these data structures depends on the characteristics of the queries. For instance, checking the status of an order is real-time in nature and requires current operational data. It does not however require multi-year of historic data. On the other hand, to develop cross-selling strategies in a market segment requires analytical processing of multi-year of historic data where real-time operational data is not required. Some processes require both real-time operational data and analytic results from historic data. For example, up-to-the-minute customer service information about a customer regarding recent inquiries and complaints in conjunction with the knowledge of the customer's product preferences and buying habits can be of tremendous value to a salesperson before a face-to-face meeting.

The types of queries classified by their complexity, planned vs. unplanned, and the repetitiveness, also play a role in the selection of particular data structures. Other considerations include costs and performance. Navas (2005) characterized the relative costs of building EDW, ODS and VODS as very high, high and low, respectively. While the cost of VODS is relatively low in comparison, it may degrade the performance of operational systems with complex queries to extract and transform data from distributed sources in real-time.

The technologies and methodologies associated with transactional databases, operational data stores, data warehouses and virtual databases are complementary to each other. The requirements of an organization will determine how they are used in various combinations. The classification of organizational needs by business processes defined by the level of decision structures and functional requirements, in conjunction with the characteristics associated with the various data structures, provides a roadmap for an optimal data warehousing strategy for an organization.

## AN ENTERPRISE MODEL FOR DATA WAREHOUSING STRATEGIES

Chan (2004) described the construct of an enterprise model as a basis for designing data warehouses. The enterprise model consists of three levels: the external enterprise view, the conceptual enterprise view and the internal enterprise view. The external view consists of processes that support an organization's operational and analytical requirements. The conceptual view consists of enterprise data and function models independent of physical implementations. The internal view consists of the technical components in databases, software applications and tools, hardware and networks. The enterprise model provides an integrated framework for the constructs of operational data stores and data warehouses.

Depending on the requirements, a data warehouse may source data from ODS and/or directly from transactional systems. The optimal number of ODS depends on the number and size of data sources, and their relationships to the target data warehouses. In a sense, the virtual database is a special case of this construct, where data sourcing is directly from the distributed data sources in real-time. In this case, the virtual database consists of virtual views instead of physical data. Therefore, as a generalization of Chan's construct, one can consider the data warehouse, physical or virtual, sourcing data directly from source systems or via intermediate ODS. The enterprise model provides an integrated framework for the design of ODS, DW and VDB (VDW, VODS). Figure 3 illustrates the architecture. The business processes as defined in the previous section that span across various decision structures and functional areas of an enterprise are represented in the external enterprise view. The various source systems, the ODS, data warehouses and virtual data warehouses are represented in the internal enterprise view. The conceptual enterprise view provides the information model tying source systems to various data warehousing constructs.

The architecture illustrates the complementary nature of the various data structures using ODS, DW, and VDB. Various combinations of these structures can be optimal in supporting certain types of processes. For instance a data warehouse is suitable for strategic decision support that requires the analytic processing of multi-year of historic data. An operational data store is suitable for tactical decision support that requires near real-time data across

distributed data sources. VDB serve the same purpose as the ODS by creating virtual databases (views) directly from distributed data sources for tactical decision support. The trade-offs are in complexity of queries, burden on the operational systems, costs and performance. VDB and DW may work together in some situations. For instance, a process that requires the real-time combination of analytical results and real-time data may use a VDB sourcing data from a DW and other operational data sources.
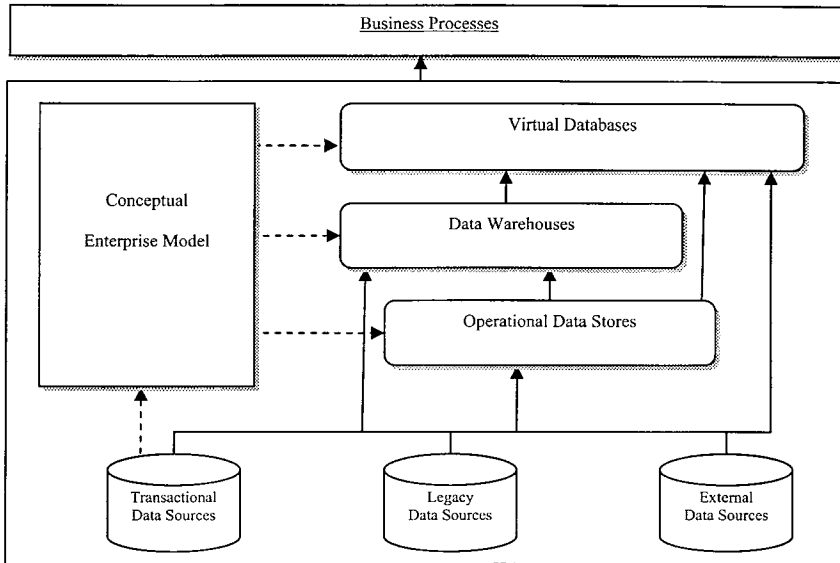


Figure 3: An Enterprise Model for Data Warehousing Strategies.

## CONCLUSION

The digital age empowered by the Internet has created the necessity for businesses to be able to access relevant information anytime, any place and for anyone who needs them. Information requirements are converging along many different dimensions that include operations and the various types of analytics supporting tactical and strategic decision making, utilizing real-time, near real-time and batch data. The advances in database and data warehousing technologies have provided many architectural choices for information management, which include transactional databases, operational data stores, data warehouses and virtual databases. This paper focuses on the needs of an organization based on the types of business processes to determine an optimal data warehousing strategy. A conceptual framework is proposed to integrate the various complementary approaches.

## REFERENCES

Ankorion, I. (2004, November). Using Virtual Data Federation in Business Intelligence. *What Works, Volume 18.* Retrieved March 1, 2005, from http://www.tdwi.org/Publications/WhatWorks/display.aspx?id=7305

Bednarz, A. (2004, April 19). Users Turn to Virtual Data Marts. *Network World.* Retrieved March 18, 2005, from http://www.nwfusion.com/news/2004/0419integration.html.

Chan, J.O. (2004). Building Data Warehouses Using The Enterprise Modeling Framework. *Journal of International Technology and Information Management, 13(2), 97-110.*

Davenport, T. H. (1993). *Process Innovation: Reengineering Work through Information Technology.* Boston, MA: Harvard Business School Press.

Eckerson, W.W. (2003, August 29). EII – The Return of the Virtual Data Warehouse? *ADTmag.com*. Retrieved June 13, 2005, from http://www.adtmag.com/print.asp?id=8152

Harrington, H. J. (1991). *Business Process Improvement: The Breakthrough Strategy for Total Quality, Productivity, and Competitiveness*. New York, NY: McGraw-Hill.

Inmon, W.H. (2004, July 8). ODS: For Flexibility and Speed. *Business Intelligence Network*[TM]. Retrieved June 13, 2005, from http://www.b-eye-network.com/view/201

Inmon, W.H. (1993). *Building the Data Warehouse*, New York, NY: John Wiley & Sons.

Klebe, S.W. (1998, June 15). Evaluating Online Credit Fraud With Artificial Intelligence. *Web Commerce Today, Issue 11*. Retrieved May 21, 2004, from http://www.wilsonweb.com/wct1/980615ai-screen.htm

Laudon, K.C., & Laudon J. P. (2004). *Management Information Systems: Managing the Digital Firm, 8th Edition*. Upper Saddle River, NJ: Prentice Hall.

MacVittie, L. (2004, November 18). CenterBoard Could Be Platform for Dive Into SOA. *Network Computing, 15(23)*, 24.

Marakas, G. M. (2003). *Decision Support Systems in the 21st Century, 2nd Edition*. Upper Saddle River, NJ: Prentice Hall.

Merriam-Webster Online Dictionary (2005). *Merriam-Webster Online*. Available: http://www.m-w.com

Navas, J. (2005, January). A New Approach to Operational Data Stores: The Virtual ODS. *DM Review*. Retrieved March 20, 2005, from http://www.dmreview.com/editorial/dmreview/print_action.cfm?articleId=1018474

Orr, K. (2000). Data Warehousing Technology. *The Ken Orr Institute*. Retrieved March 1, 2005, from http://www.kenorrinst.com/dwpaper.html

Reimers, B.D. (2003, April 14). Too Much Of a Good Thing? *Computerworld, 37(15), 38-39*.

Russom, P. (2003, April 5). A Virtual Point of View. *Intelligent Enterprise*. Retrieved March 18, 2005, from http://www.intelligententerprise.com//030405/606decision1_1.jhtml

Simon, H.A. (1960). *The New Science of Management Decision*. New York: Harper & Row.

Singh, H. (1998). *Data Warehousing: Concepts, Technologies, Implementations, and Management*. Upper Saddle River, NJ: Prentice Hall.

Sperley, E. (1999). *The Enterprise Data Warehouse: Planning, Building, and Implementation, Volume 1*. Upper Saddle River, NJ: Prentice Hall.

Todman, C. (2001). *Designing a Data Warehouse Supporting Customer Relationship Management*. Upper Saddle River, NJ: Prentice Hall.

Turban, E., Aronson, J.E., & Liang T.P. (2005). *Decision Support Systems and Intelligent Systems, 7th Edition*. Upper Saddle River, NJ: Prentice Hall.

Watson, H.J., & Frolick, M.N. (1993). Determining Information Requirements for an EIS. *MIS Quarterly, 17(3)*, 255-269.