

California State University, San Bernardino

CSUSB ScholarWorks

Theses Digitization Project

John M. Pfau Library

2002

Data warehouse applications in modern day business

Carla Mounir Issa

Follow this and additional works at: <https://scholarworks.lib.csusb.edu/etd-project>



Part of the [Business Intelligence Commons](#), and the [Databases and Information Systems Commons](#)

Recommended Citation

Issa, Carla Mounir, "Data warehouse applications in modern day business" (2002). *Theses Digitization Project*. 2148.

<https://scholarworks.lib.csusb.edu/etd-project/2148>

This Project is brought to you for free and open access by the John M. Pfau Library at CSUSB ScholarWorks. It has been accepted for inclusion in Theses Digitization Project by an authorized administrator of CSUSB ScholarWorks. For more information, please contact scholarworks@csusb.edu.

DATA WAREHOUSE APPLICATIONS IN MODERN DAY BUSINESS

A Project
Presented to the
Faculty of
California State University,
San Bernardino

In Partial Fulfillment
of the Requirements for the Degree
Master of Business Administration

by
Carla Mounir Issa


June 2002

DATA WAREHOUSE APPLICATIONS IN MODERN DAY BUSINESS

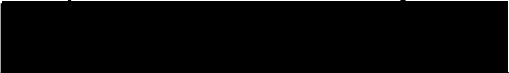
A Project
Presented to the
Faculty of
California State University,
San Bernardino

by
Carla Mounir Issa
June 2002

Approved by:


Dr. Robert D. Wilson, Chair, Information
and Decision Sciences


Dr. C. E. Tappe Rohm Jr.


Dr. Walter T. Stewart

June 10, 2002
Date

ABSTRACT

Data warehousing is not a new concept in the business world. However, the constant changing application of data warehousing is what is being innovated constantly. It is these applications that are enabling managers to make better decisions through strategic planning, and allowing organizations to achieve a competitive advantage. The technology of implementing the data warehouse will help with the decision making process, analysis design, and will be more cost effective in the future. Thus business owners should look at the cost benefit effect of implementing the data warehouse for the long run of the business.

ACKNOWLEDGMENTS

I would like to thank my mother and father, as well as Julia, Tania, and Dima for their continuous support. In addition I would like to thank my advisors Dr. Robert Wilson, and Dr. Tapie Rohm for their time. Last but not least I would like to thank Dr. Nabil Razzouk for everything that he has done.

DEDICATION

To Mom and Dad

TABLE OF CONTENTS

ABSTRACT	iii
ACKNOWLEDGMENTS	iv
LIST OF FIGURES	ix
CHAPTER ONE: DATA WAREHOUSE IN PRESEPTIVE	1
Purpose of the Project	3
Scope	3
What is a Data Warehouse?	4
Data Warehouse versus Database	4
Repository	5
Data Warehouse versus Data Marts	5
Data Warehouse versus Online Transaction Processing	8
Why a Data Warehouse?	8
Definitions of Terms	13
Transaction Processing Systems	14
Online Analytical Processing versus Data Warehousing	15
Data Warehouse and Management	18
Operational versus Informational Systems	20
Factors in Developing a Data Warehouse	21
Business Intelligence Tools	23
CHAPTER TWO: APPLICATION TO INDUSTRIES	24
Auditors	28
Price Waterhouse	29

Pharmaceutical and Healthcare	32
Insurance Companies	35
Retailers	37
Centralized Data Warehouse	37
Sears, Roebuck and Company	38
Data Warehousing is not Limited to Business	40
Governments	41
Security	42
Custodians	44
Geographic Information Systems	45
CHAPTER THREE: DESIGN AND TOOLS	
Implementing a Data Warehouse	48
Issues in Designing a Data Warehouse	50
Roles Within a Data Warehouse	52
Data Process	53
Data Warehouse Administrator	54
Online Activity Monitor	57
Activity Monitor	58
Monitoring the Data Warehouse	58
Transaction Processing Systems	59
Data Warehouse Iteration	60
Quality of Data	61
Testing Data	65
Measuring Capacity	66
Dimensionality Modeling	67

Open-Loop/Closed-Loop Architecture	71
Enhance Performance	72
Augment Security	78
Data Warehouse Tools	79
Online Analytical Processing	80
Data Mining	81
Information Management and Data Warehousing	81
The Management Decision Making Process	85
Data Warehouse on the Web	88
Data Warehousing and Marketing	90
CHAPTER FOUR: CUSTOMER RELATIONS MANAGEMENT	94
Market Segmentation	101
Graybar Electric Company	101
Cypress Semiconductor Corporation	102
Vendors	103
CHAPTER FIVE: COST-BENEFIT	
Evaluation of Cost-Benefit	108
Costs	109
Benefits	109
Cost of a Data Warehouse	110
Benefits of a Data Warehouse	114
Immediate Information Delivery	115
Data Integration	115
Trend and Seasonality Analysis	115

Freedom from Information Systems Department	115
Data Warehouse Tools	116
Limitations of a Data Warehouse	116
Poor Data Quality	116
Lack of Accessibility	119
Integrate Data into the Mainframe	119
Cost	119
CHAPTER SIX: SUMMARY	120
Advantages of Data Warehousing	122
Disadvantages of Data Warehousing:	123
Data Warehousing is not for Everyone	126
Conclusions	129
Recommendations	130
Charge Back	131
Ensuring Success	133
The Future	134
REFERENCES	137

LIST OF FIGURES

Figure 1. Data Warehouse versus Data Marts 7

Figure 2. Two-Dimensional Figure 16

Figure 3. Three-Dimensional Matrix 17

Figure 4. Structure of the Data Warehouse 52

Figure 5. Data Warehouse Iterations 61

Figure 6. Star Schema 69

Figure 7. Life Cycle of Information 83

CHAPTER ONE

DATA WAREHOUSE IN PRESEPTIVE

In this information age, optimizing data usage is crucial for more effective decision making in organizations. Technology is expected to assist users in utilizing the available tools and submitting the required information at the right place, time, and appropriate cost. Companies are constantly struggling to keep up with the changing business conditions and to stay ahead of competition. In order to succeed companies should be able to analyze, plan and react in a faster manner to changes in the environment.

Companies are realizing that information is the key to achieving a competitive advantage and for survival. The data warehouse is used for strategic decision making where management in the organization will utilize existing information to make decisions that have an impact on the future. Thus it is important to have information that is valid, accurate, and relevant on hand at all times. Enhancing data quality is what makes the data warehouse a necessity in today's business, data quality will be covered in more details in Chapter Three.

A data warehouse provides companies with tools to enable users to access and utilize a wide range of stored information that could either be in a summarized or detailed format. The tools will assist management to use the best available information to make decisions now while considering their impact on the future, and thus support organizations with the strategic management process. By effectively utilizing the information the organizations will achieve a competitive advantage in the market.

It is imperative for organizations to define what their business objectives are prior to implementing the data warehouse. Once the requirements are defined, the data warehouse is tailored to achieve the organizations goals. Management has to support the project of implementing a data warehouse by providing the necessary resources, and managerial support.

Within an organization there are typically multiple decision-support projects that operate independently, each serving different users but often requiring much of the same data. Thus, developing a system that will receive data from multiple operations sources within different departments is important. Users should be able to manipulate the information by running queries, producing reports, and performing analysis that will support the

organization's goals. Thus, "the challenge for organizations is to turn archives of data into a source of knowledge, so that a single integrated/ consolidated view of the organization's data is presented to the user" (Connolly & Begg, 2002, p. 1046).

Purpose of the Project

The purpose of this project is to explore the nature and purpose of data warehousing and its application as an instrumental tool in business intelligence for Decision-Support System (DSS). More specifically the project evaluates the role of the data-warehouse model in providing timely, accurate, and relevant information.

Scope

The scope of the project is limited to defining the business need for data warehousing in organizations. The role of the data warehouse model is providing timely, accurate, and relevant information. Thus organizations have to seek ways for integrating technology into business application to achieve their goals. In addition, management has to present reasons for utilizing a data warehouse in organizations that would lead to better decision making. This aspect will be covered in more details later in this Chapter.

It is vital to get business users to change their approach to decision making, analysis, and business intelligence, and implement data warehousing into their business and future plans. By doing so the business can to achieve a competitive advantage with the information they have, this will be covered in more detail in Chapter Three. How industries are using data warehousing in their business applications is covered in Chapter Two.

What is a Data Warehouse?

A data warehouse is often mistaken with a data mart and database. The following statement defines what constitutes a data warehouse, and differentiates data warehouse from other format of data storage.

Data Warehouse versus Database

A datum is a raw piece of information that is capable of being moved and stored. A database is a collection of different types of data organized to make the data easily attainable and manipulated through the use with programs such as SQL. A database is typically organized into records - one record per item, such as a customer order - that are themselves divided into several fields, with each field containing information about a specific aspect or attribute of the item. In addition to storing data, a

database management system handles security and access control.

A data warehouse is a type of organizational memory, where it is a snapshot of the organization at a particular point.

Repository

A repository can be another name for a database, but it can also refer to the "aggregation of data into some accessible storage location, without necessarily having the ability to manipulate or extract data" (Connolly & Begg, 2002, p. 1049).

Data Warehouse versus Data Marts

A data Warehouse is made up of a union of data marts. The data mart is a logical subset of the complete Data Warehouse; it is a "collection of data specifically designed for the use of a department" (Inmon, 2001a) that needs to be constantly updated for better results as presented in Figure 1.

The data mart is the queryable source of data in the enterprise and it must not be organized around an entity-relation (ER) modeling; because if it is there will be a loss of understandability of the information and poor performance. Examples of departments with data marts are sales, marketing, accounting and finance. Each department

looks at data in a different way according to their needs where the data can be summarized by the week, or month.

Data warehousing is the decision-support database that supports multiple corporate functions or business units. It is the process of managing the data warehouse and data marts. This process includes all the ongoing support needs of the refresh cycle, database maintenance, and continual refinements to the underlying data model. Whereas data mart, is a decision support database that is limited in scope to support a single business function or unit (Craig & Vivona & Bercovitch, 1999, p 105).

Data mart is a "smaller form of the data warehouse with a more specialized purpose" (McKeown, 2000, p. 134). Hence it increases access and usability of the data, and can run in size from megabytes to gigabytes. Data warehouses on the other hand usually run from gigabytes to terabytes. A data warehouse stores enterprise-level Online Analytical Processing (OLAP) data, while a data mart is smaller and typically covers a single function in an organization (MSDN, 2002). Often-selected data from a data mart is fed into a smaller database called a data cube for intensive processing.

According to a survey of more than 100 IT professionals worldwide conducted by Curt Hall and Data

Warehousing for Business Intelligence, "sixty-five percent of respondents indicated currently having a data mart or data warehouse in place, and eighty two percent indicated say they are currently building data warehouses and data marts". The same report indicated that approximately 65%-70% of companies worldwide are currently using and developing data warehouse applications to ensure they use data warehousing and business intelligence technology for maximum competitive benefit.

Figure 1 shows the relationship between data warehouse and the data mart. Whereby the data marts get their information from the data warehouse, which in turn gets its information from the source systems.

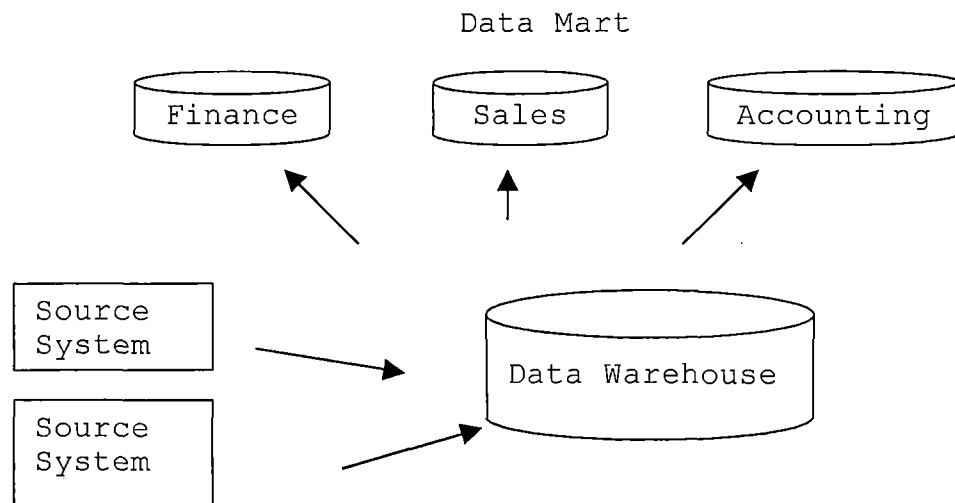


Figure 1. Data Warehouse versus Data Marts

Data Warehouse versus Online Transaction Processing

Online Transaction Processing (OLTP) databases give the user a quick response time, as well as consistent current data, making the information operational. An organization will have a number of OLTP that hold current data for business processes such as inventory control, customer invoicing and point of sale. However, an organization will have a single data warehouse, which holds all "data that is historical, detailed, and summarized to various levels" (Connolly & Begg, 2002, p. 1049). In addition OLTP is application and transaction driven whereas, data warehousing is subject oriented and is used for analysis. Furthermore, OLTP can not answer ad hoc queries like data warehousing systems.

Why a Data Warehouse?

From a business perspective data warehousing is a "reservoir for knowledge about corporate performance." It is a "centralized database of statistics that is highly tuned for decision making purposes" (Craig & Vivona & Bercovitch, 1999, p. 316). Data warehousing has a set of tools that enable users to query, extract, analyze, and present information in the format required.

Data warehousing is imperative for an organization that has a large collection of data available for users; where the information is scattered around the organization, with no summaries, making it difficult for users to retrieve and analyze the information on hand. In addition, there needs to be consistency for more effective decision making, especially when there is no commonality to the information, and each department views and use the data in a different way.

Managers and analysts must have instant access to the data in the warehouse from their computers. The connection must be immediate on demand to have better performance, especially when decisions need to be made rapidly to stay ahead of the competition due to market changes.

Management should be able to evaluate how their business is operating by measuring, managing and monitoring the business on an ongoing basis. Thus a system requirement is needed prior to the implementation of the data warehouse. By having a detailed list of the operation and flow of data, the users will be able to identify business weaknesses, and hence ratify and fine-tune these problems for better decision making.

Through data warehousing, organizations especially companies online, are able to develop a trend that will

indicate who the customers are; the purchasing trends of the customers, and what the customers are buying.

A data warehouse can be normalized (where the data is decomposed into natural structures) or denormalized. "It can be a relational database, multidimensional database, flat file, hierarchical database, object database, etc. Data warehouse data often gets changed. And data warehouses often focus on a specific activity or entity" (A definition of data warehousing, 1995).

When designing a Data Warehouse one should take into account the environment where the data will be used. This assists in the design stage "architecture" of the data warehouse, since it is important to know the flow of data in the organizations, and thus be able to decide on the data components. The data utilized has to be accurate and updated frequently to ensure validity and accuracy.

Organizations are undertaking numerous information technology (IT) initiatives to improve their competitiveness, such as implementing "enterprise resource planning (ERP) and developing client/ server and Internet technologies" (Craig & Vivona & Bercovitch, 1999, p. 18).

A data warehouse enables an organization to identify relationships within the data. Data Mining and Online Analytical Processing (OLAP) are both tools used to

extract and store data in the warehouse. The tools can be used by the organization to learn more about the customers and their product preferences

A data warehouse is a key factor for an organization's Strategic Management process that aids in the decision support system. Strategic management is using the best available information to make decisions now with an eye on the future. Data warehouse allows organizations to make strategic analysis by looking at external and internal aspects of the organization, through the data in the warehouse where the user can perform an extensive analysis on Strength, Weakness, Opportunities and Threats (SWOT).

The organization can make better decisions by identifying strategic alternatives, selecting the appropriate strategy and implementing a strategic plan that will enhance the future growth of the organization.

The external analysis involves an examination of relevant elements external to an organization that will help identify opportunities and threats that will have an effect on the organization. Information on the industry and economic conditions are stored in the data warehouse, and the user can make analysis of trends and movements in the market relevant to the organization based on the

stored information. The data warehouse will enable the user to evaluate the following external factors effecting the organization such as:

- Customer analysis. The data warehouse will enable the user to look at customers more closely and develop purchasing trends as well as help with segmenting the market.
- Competitor analysis. The warehouse allows the organization to look at their competitors' performance through sales history and strategies
- Market analysis. The user can identify weaknesses and strengths in the industry, as well as forecast product life cycles
- Environmental analysis. The data warehouse allows the user to look at economic changes taking place that have an impact on the organization such as inflation.

The data in the warehouse looks at factors affecting internal elements that will allow the user to identify strengths and weaknesses in the organization by looking at factors such as:

- Performance analysis, where the information stored in the warehouse allows the user to look at and analyze elements such as sales

profitability, changes in price, and performance of the salespeople.

- Determinants of strategic options such as past and current strategic problems, and may involve an evaluation of the financial condition of the organization.

By utilizing the data warehouse tools effectively (the tools are covered in detail in Chapter Three) organizations can forecast long-range options which will allow the organization to achieve a competitive advantage.

Definitions of Terms

Data Warehouse is a collection of integrated subject-oriented data bases designed to support the Decision System Support (DSS) function. Data warehouse integrates information from across the transactional system and external sources to help users analyze data, for more effective decision making.

Each unit of data is relevant to some moment in time similar to a snapshot of the organization at a particular time. "The data warehouse contains atomic data and lightly summarized data". The accepted definition of data warehouse as described by the father of data warehousing (W. H. Inmon) is " subject oriented, integrated, non

volatile, time variant collection of data designed to support management DSS needs" (Connolly & Begg, 2002, p. 1047).

The data warehouse is organized around ongoing issues of the organization rather than the major application areas. It is designed to provide an architectural model for the flow of data from operational systems to decision support environments. Thus "organizations are able to store decision-support data rather than application-oriented data" (Connolly & Begg, 2002, p. 1046).

Transaction Processing Systems

Creating a data warehouse involves using Transaction Processing Systems (TPS) where raw data is converted into usable electronic form. The data is cleaned to remove inconsistencies, redundancy and enhance accuracy; and then loaded into the data warehouse where it is then converted into a format that could be accessed by the users. TPS is an ongoing application that is used in the warehouse; data iteration is entering the data following TPS, will be covered in Chapter Three.

Transaction processing generates large amounts of data, which helps the organization identify its customers, and suppliers. According to Patrick McKeown in his book

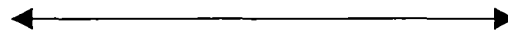
Information technology and the network (2001) states that the data in the warehouse is organized according to the subject rather than the application, and is hereafter constantly updated.

Online Analytical Processing versus Data Warehousing

Online Analytical Processing OLAP is designed to handle queries that will assist in discovering trends and critical factors to the user. OLAP queries typically require large amounts of data however it is not a Data Warehouse. OLAP data is organized into multidimensional cubes for better performance.

OLAP uses multidimensional data view for relational data such as it may represent information for a retailer on store location, product, time, promotion, and sales. The information may be presented in a two dimensional mode (Figure 2) or in a cube format (Figure 3) that will enable the user to trace information on total sales for a specific store location at a specific time. The information represents greater details on customer analysis.

Store Number	Time	Sales (\$000)
10009	Q1	1000
10008	Q1	2345
10110	Q1	3210
11001	Q1	4510
10009	Q2	1234


 Store Number

	10009	10008	10110	11001	
Q1	1000	2345	3210	4510	
Q2	1234				



 Time

Figure 2. Two-Dimensional Figure

Figure 2 represents information on sales of stores at a certain time. The relationship presented will enable the user to make a comparison of sales between a specific store (store number) over a specific time (quarters).

Store Number	Time	Product	Sales (\$000)
10009	Q1	0201TL3	1000
10008	Q1	0201TL3	2345
10110	Q1	0201TL3	3210
11001	Q1	0201TL3	4510
10009	Q2	0201TL3	1234

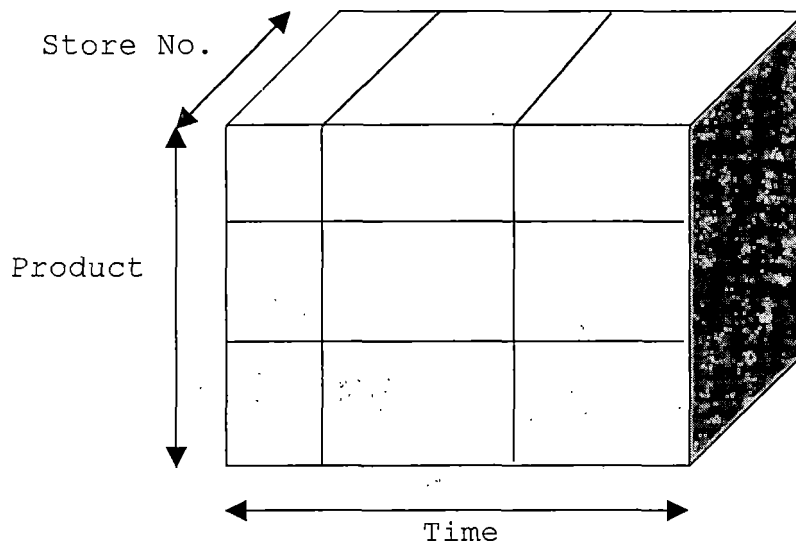


Figure 3. Three-Dimensional Matrix

The Three-Dimensional Matrix in Figure 3 allows users to look at the sales of a specific product (product ID) at a certain time (quarter) for a specific store location (store number). The OLAP has a slicing and dicing tool that enables the user to look at data from different views.

Microsoft Structured Query Language (SQL) was designed to "centrally store the data generated by the

daily transactions of large companies or government organizations" (MSDN, 2002). The databases have become efficient systems for daily operations, and transactions. Some of these transactions such as recording and placing orders are now referred to OLTP systems. OLTP systems may be good in recording the data required to support daily operations, however they are not organized to retrieve on the spot information to managers. Managers running OLAP queries need to reference data from several of OLTP systems. The lag in time for decision making is a disadvantage of this system.

Data Warehouse and Management

According to a report "Four in five companies interviewed by InformationWeek Research and Optimize Magazine Research have either business-intelligence strategies headed by business managers or have business executives on joint business-intelligence committees of IT and business managers" (Whiting, 2002). The same report stated that only two-thirds of companies say IT executives are members of a joint business-intelligence committee of IT and business executives or that their IT departments are sole managers of their business-intelligence campaigns. Therefore although it is important to have

management's input into the data warehouse requirements, management should delegate the responsibility to the data warehouse administrator and offer support.

To implement data warehousing successfully an organization has to have a "strong business management sponsor" (Kimball, Reeves & Ross & Thornthwaite, 1998, p. 47). Management should support the decision to implement data warehousing by committing financial resources and time. Prior to the design stage, management should allocate resources from the budget to support the project.

In addition, management should have a compelling business motivation to implement the system; for example the system will save time, money and resources. When designing and implementing the system there should be an understanding of the needs of the company prior to implementing. Thus the users and management should work closely with the information systems department or the external consultant to capture what the company needs. These needs are a work in process that is it is continuously updated; a prototype is required to show the users what they have asked for.

Feasibility analysis is important to assess if the Company is committed to invest in implementing a system.

The data warehouse has continuous costs that need to be considered when implementing the project. It is not a one time payment that needs to be made. There are ongoing costs that need to be considered and weighed with the benefits that a data warehouse can generate.

Management should assign consultants or administrators to collect information on implementing the warehouse. There are two basic techniques for gathering requirements and investigating what the organization requires:

- Interviews should be planned and conducted either with individuals or small groups by having open and closed ended questions.
- Observations should be objective to fully understand the needs. Thus allotting the time and resource for the interviews is important to have a more effective understanding of the organization.

Operational versus Informational Systems

There are two different types of information systems in organizations: operational systems and informational systems. The Operational systems are systems that help organizations run the daily operations such as inventory

and payroll transaction; and the information covers one area of the business.

Informational systems analyze data and aid in decision making such as forecasting future sales and prices, in addition to indicating how the organization is currently operating. The information covers several operating areas within an organization.

It is important to differentiate the two systems in discussions since each has different applications and functionalities.

Factors in Developing a Data Warehouse

There are several factors that organizations have to take into account when designing and building the data warehouse, these factors are covered in greater detail in Chapter Three. Organizations must have an idea about what could be accomplished with the data warehouse prior to implementing it. There are stages that the management has to follow to ensure a successful data warehouse.

Management has to perform extensive research on the advantages of having the data warehouse for meeting the goals of their particular organization. Management should justify the implementation by performing a cost-benefit analysis, which will be covered in Chapter Five. A budget

has to be developed with the costs involved in building and maintaining the data warehouse. Once the budget and the cost-benefit are preformed the organization has to decide whether it is prepared to commit to developing such a project; it has to commit by supporting the project and allotting both time and money.

Management has to gather information about whether it wants to hire a data warehouse consultant or just an administrator whose job is to design a suitable data warehouse system for the organization. The management then agrees to the choosing the hardware, the database management system, the data extraction tools, and the Business Intelligence tools. A data warehouse prototype is then built, tested, and evaluated. Once the management agrees to the final product, the data warehouse is deployed and put into operation, and users are trained on the applications. The management has to realize that the process of the data warehouse does not stop after implementation since it is crucial to maintain the data warehouse and have follow up updates. In addition it is crucial to preserve the integrity and validity of the information within the warehouse at all times. With the utilization of the system new possibilities are discovered, and modifications can be made.

Business Intelligence Tools

Business Intelligence (BI) Tools enable business users to use tools to access large amounts of complex data. There are three types of tools used for BI: Multidimensional Analysis, which can also be referred to as Online Analytical Processing (OLAP), this tool allows the user to look at the data from a number of different dimensions. Another tool is the query tool, which enables the user to ask questions about patterns or details in the data. And the third tool is the data-mining tool, which automatically searches for significant patterns or correlations in the data. These tools are covered in more detail in Chapter Three.

CHAPTER TWO

APPLICATION TO INDUSTRIES

Data warehousing has had an effect on the development of organizations by providing tools to enhance their strategic objective as well as help with the decision making process. The benefits that data warehousing provide vary between organizations and industries, in terms of quantified return on investment, to faster response time. It is difficult to compare data warehousing between organizations since the data warehouse objective, application and utilization differ between the various companies within industries.

In order to maintain its strategic market position, it is imperative for companies to maintain the integrity of the data in the data warehouse, provide training programs, and continue to innovate in finding better ways to extract, and format data for better usage within the organization.

To better illustrate the importance of implementing data warehouse this is a quote that sums up the importance of data warehousing from Louis V. Gerstner Jr., IBM Chairman and Chief Executive Officer "Today and in the future, companies that succeed will be those that know how

to manage knowledge faster than competitors. It isn't a question of getting new information. It's the ability to extract information from your existing business, to identify trends and insights faster than your competition."

Data warehousing is crucial in the development of strategic management within organizations in varying industries. The availability of information in the warehouse has helped in decision making, as well as helped in the organizations performance. The following Web Page (dataspace.com/whatman.htm) gave an analysis on how data warehousing has helped industries achieve their goals, by describing the scope of improvements that data warehouses have had on the organizations. The benefits are described in terms of:

- Sales Analysis. Data warehousing indicates vital pricing and distribution decisions, by monitoring historic data and analyzing past sales to determine whether changes in product/ packaging/ price / service decisions effected customer behavior. The data warehouse analysis also identified preferred customer profile based on revenue and marginal profits.

- Salespeople's Analysis. The data warehouse identifies the performance of sales personnel within an organization.

The data warehouse assists financial personal and accountants in developing budgets, based on past performance. In addition, financial analysis can be made where comparison of actual budgets versus forecast is performed on daily, monthly and yearly basis. The data warehouse provides such reports to managers upon request, which is a great advantage. In addition the warehouse generates financial ratios, and analyzes key expense generators.

Business Process Reengineering, and Enterprise Resource Planning helped organizations manage costs, and utilize the technology for their competitive advantage. Furthermore, data warehouse applications allowed companies to improve their processes for a single function e.g. finance, while also providing integrated transaction processing across multiple functions e.g., order management, billing, accounts receivable and other functions.

The airline and tourist industry benefited greatly from applying data warehousing tools to their business. For instance when a customer makes a reservation, the

system would be able to look at the customer information stored in the system via a unique identification number assigned to the customer. The system will generate alternative choices for the customer to select from depending on historical data gathered about the customer. For instance whether the customer is a vegetarian, whether the customer is a frequent flier and if so the preferred destination, and preferred seating arrangement which can be accessed and used. The companies that integrate their services together such as car rentals and hotels can share the information and give customers incentives such as mileage on air-tickets by using the service.

In addition, companies could also evaluate performances of the salespeople, as per sales per week, month or year. Companies can also gather information on the more popular routes that customers chose. Through proper utilization of warehousing tools companies can segment customers for example according to the demographics or purchasing trends.

The other change in business trend is the focus on customer centricity or Customer Relations Management (CRM), which will be covered in more detail in Chapter Four.

Auditors

Traditionally (and currently), financial reports have been produced on a periodic basis bi-monthly, quarterly, or yearly, mainly because obtaining information on a real-time basis is expensive and tedious in terms of hours spent in accumulating the information. In addition, by the time the reports are generated the information is already old, since it shows how the organization performed in the past over the audited period. The reports provide no accurate reflection on how the organization is doing "now."

With proper utilization of the data warehousing tools, organizations can produce standardized financial information on a realtime, online basis. This requires organizations to have their accounts updated on a daily basis to assure the quality of data. The electronic information is timelier, accessible, and can be easily stored, retrieved or summarized in the data warehouse. This does not mean that the auditors no longer comply with auditing regulations, on the contrary auditors will have to submit the same reports, but instead of doing it on paper they have to do it online.

"Continuous auditing enables auditors to significantly reduce and perhaps eliminate the time

between occurrence of the client's events and the auditor's assurance services thereon" (Rezaee, Sharbatoghlie, Elam, & McMickle, 2002). With real time audits, auditors can discover and report discrepancies and irregularities in the system in a timely manner. This will increase the assurance in the reliability and relevance of electronic documents, records, and data as well as help ratify problems faster.

Technology will enhance the compliance of organizations with the standard audit requirements. The monitoring agencies can detect errors in reporting; and shareholders will feel confident that organizations are operating according to the goals and objectives set forth by the management or the Board of Directors.

Price Waterhouse

Price Waterhouse began implementing data warehousing in 1990, currently they have a data warehousing practice of more than 1,600 consultants representing major countries in the world. An article titled "Data Warehousing: The Past 10 Years Have Been Quite a Ride" by Michael Schroeck (2001) in the Data Management Review (dmreview.com) talks about how it is crucial for organizations to have an integrated data warehouse as a foundation for better CRM. The data warehouse must include

all relevant information about a company's customers along with the "required analytics to perform customer segmentation, churn analysis, propensity to buy, one-to-one marketing and other mission-critical CRM programs" (Schroek, 2001).

Schroek is a partner and the global leader of PwC Consulting's iAnalytics solution, expresses in the above article, how building data warehouses increased the decision support through extracting and consolidating data from mainframe-based legacy systems, to analyzing data from a company's Web site. He states that for the past decade, data warehousing has gone from being a program typically owned and sponsored by IT to being a "strategic imperative tool for business executives across the enterprise to include areas such as finance, marketing, distribution/logistics, human resources, sales, customer service, manufacturing, etc" (Schroek, 2001).

In addition Schroek notes that "companies consider their information a strategic asset with the potential to provide real competitive advantage if the data can be aggregated and then fully leveraged using rich analytics." The author reflects on the influence that data warehouse had on organizations decision support, and achieve their goals. During the 1990s many firms acquired or merged with

other organizations. As these mergers took place the value of information began to extend beyond traditional financial analysis, and organizations began to recognize that their business intelligence was a strategic asset with inherent value. Hence, financial institutions looked at data warehouses as an asset and source for growth, where they can utilize information about their customers, and try to cross sell products. The data warehouse helped in placing information on various layers within the organization, and by such, it facilitated the usage of the information to the end user to improve their service to the customers.

Schroeck in summary concludes that there is more room for data warehousing applications to grow especially that more organizations are implementing data warehousing in their business.

With changes in economic trends especially with innovation in technology, utilization of ecommerce, and globalization data warehousing will seem to be the solution to many organizations. Thus organizations have to take into account these changes by incorporating data warehousing into the future plans in order to compete in the market.

Pharmaceutical and Healthcare

Organizations within Pharmaceutical and Healthcare Industries use data warehousing tools. Practices LLC, is a specialized company that deals with data warehouse benchmarking, (benchmarkingreports.com) developed a summarized report on companies that use data warehousing as a strategic management tool for decision making.

The report is titled Pharmaceutical and Healthcare Data Warehousing: Managing implementation to create greater ROI. The summarized report described how pharmaceutical companies used data warehouse tools to give them a competitive advantage and deliver targeted marketing messages to customers, as well as boost products' sales management.

The research conducted used a wide range of resources, including "Best Practices, LLCTMs" own proprietary databases, public and private databases, literature reviews, on-line data searches, conference, proceedings, professional journals and books, academic research, consulting assignments, and an in-depth analysis of data warehousing systems.

The overall effect of implementing data warehousing included "higher quality sales and marketing efforts, that translated into lower costs, more effective communication,

superior market share and enhanced economic returns”
(Pharmaceutical and Healthcare Data Warehousing: Managing implementation to create greater ROI). The report from Benchmark covered 13 companies in the Pharmaceutical industry:

- 3M Pharmaceuticals
- Bayer Pharmaceuticals
- Bausch & Lomb
- Eli Lilly
- GlaxoWellcome (now GlaxoSmithKline)
- Hoechst Marion Roussel (now Aventis)
- Johnson & Johnson
- Pharmacia & Upjohn (now Pharmacia)
- SmithKline Beecham (now GlaxoSmithKline)
- Bergen Brunswig
- Columbia HCA
- NDC
- Scott Levin

According to the above cited report, these companies enhanced their business value through the use of data warehousing tools. Some of the pharmaceutical manufacturers connected the data warehouse to their core corporate goals, which resulted in better customer

targeting, more consistent communication with major customers, and allowed the sales force to personalize the business. This enhanced the CRM of the companies and helped to develop trends, and utilize more effective strategies.

Other pharmaceutical companies focused data warehousing design and implementation on their sales force. "The sales force saw improvements within a year, including improved and increased pre-call analysis, post-call assessment, and integration of teamwork and selling efforts across business units and product teams" (PR Newswire, 2002).

In addition the report reflected on how proper planning of the data warehouse ensured that the organizations met their corporate needs, and gave them a return on their investment. Where in some cases the initial investment of the above companies was "as high as \$10 million, the companies noted a full return on investment, whether in the form of reduced operating costs, increased revenue, or both" (Pharmaceutical and Healthcare Data Warehousing: Managing implementation to create greater ROI).

Insurance Companies

In an article titled "Building a better warehouse" Lori Chordas (2001) evaluates changes that took place with insurance companies over the past 10 years. She uses a report by Conning & Co. to reflect the changes in attitude amongst insurers regarding data warehouse. In 1999 Conning & Co. asked a group of insurance companies what they thought about data warehousing. Twenty percent of the respondents said they found data-warehouse systems to be effective to their company's operation, and 68% viewed the techniques as ineffective or not innovative. In 2000 the same company conducted another study that showed insurers were beginning to "resolve some of the challenges and recognize more benefits in using data warehousing" (Chodras, 2001). However, the technology is still new and it is difficult to assess the long-term benefits of data warehousing. However, for now the success of using data warehousing is reflected with the insurers performances and the cost effectiveness from implementing the system

Although some insurance companies such as Axa Financial, and Chubb Corp are happy with the benefits that the data warehouse is yielding to their operations, other insurance companies in the Conning study cited poor data

quality and lack of understanding of the data as two of the chief reasons for the failure of the system.

In order for the data warehouse to be effective it is vital to maintain the integrity of the data warehouse by ensuring consistency, validity, accuracy and integrity. Data has to be cleaned prior to loading to the system, and regular maintenance has to be performed.

The report stated the benefits of implementing the data warehouse is that the system provided the insurers with the ability to better understand their market. By better understanding their customers not just from basic "demographics but also from their behavior" (Chodras, 2001) the organization can segment their customers and focus on targeting them with the appropriate marketing programs. In addition, the warehouse allowed them to build more data and focus on their CRM. This is achieved by assessing their customers and seeing how long they have been with the company, understanding the customers' needs, and identifying new opportunities. By developing records on customers, the companies were able to cross-sell customers other products.

Chodras reported on two studies that reflect on the growth in data warehouse spending by insurers, one of the studies was conducted by Meta Group and the other study by

Gartner Group. According to the study conducted by Meta Group, "the average user spent \$6 million on their data warehouse in 1998" (Chodras, 2001), for the same year insurance companies spent \$5.5 million. The study by Gartner Group, who surveyed 2,000 companies, estimated that U.S. companies spent \$7 billion in 1999 on the creation and operation of data, and companies are spending 35% annually since 1996. In the article Chodras stresses that cost is not the major determinant in the development of the data warehouse for an organization.

Retailers

Many retailers across the globe are utilizing the warehousing tools to develop forecasts based on data on hand. The data helps identify how the market will react to certain product if launched, at a certain price. In addition the stored data has information on sales of products, performance of stores and sales people.

Centralized Data Warehouse

For example Pepsi Co. and Zara (fashion retailer based in Spain) are using centrally located data warehouse for their retail stores across the globe. Salespeople can download and upload data into their online systems, such as the PDA's used by the salespeople and vice versa where

the information can be sent back to the central data warehouse. In Zara's case the retail stores across the globe send information to the data warehouse on high moving products; the warehouse the information is then accessed by production and manufacturing units who ship the products instantly to the destinations. The movement of the products are monitored at all times to identify where the shipments is (similar to the UPS system). The data allows the warehouse to observe the fast selling items daily and react swiftly to the demand.

Pepsi Co uses centralized data warehouse to establish the fastest selling product, and the location with the highest sales. This information is monitored to enable the management to make strategic decisions and maintain their competitive position in the market.

Sears, Roebuck and Company

Retailers such as Sears, Roebuck and Co. plan to deploy 95TB of new storage capacity, "tripling the amount it currently has installed and allowing the retailer to consolidate two key data warehouses and build a storage-area network" ((Mearian, L., 2002). Sears is trying to merge its inventory and sales data warehouse and another warehouse that holds its customer information on a single server. This will allow the users to combine the

information stored and utilize it effectively. The detailed data will give information about customers and their purchasing patterns, and help Sears focus on targeting customers with particular products.

Sears' current applications can trace sales to products where the company can identify the amount of a particular product sold on a particular day. Sears is focusing on using information on its customers more effectively however, the information does not reflect whether one customer made several purchases throughout the department store, nor does the information give details pertaining to the customer's future purchases.

Employees at Sears' are currently using the "two data warehouses" for analytical purposes, where they can obtain information on daily product sales. But the data there has no information on what individual customers bought and correlate that with previous purchases. In the future, with the integration of the new data warehouse, a correlation can be made between the purchases and the customers, which will assist management in quicker decision making, since management can react faster to changes in purchasing trends. The project should also speed up response time for end users because of an upgrade from the 450-MHz Intel Corp. processors used in the

retailer's current NCR servers to the 900-MHz Pentium III Xeon chips that are built into the new system.

According to the article by Mearian L, Sears triples its storage capacity (2002); the storage rollout will increase the EMC Corp. disk capacity that Sears has installed to 140TB. In addition Sears is purchasing World-Mark Unix server and a Teradata database from NCR Corp (Mearian, L., 2002). Currently, the data warehouses are located at separate facilities in different states, however with the new technology the warehouses will work parallel to each other.

Sears is also working with EMC to wire together a 25TB storage-area network (SAN) that will be used to link data residing on various Unix and Windows NT servers supporting applications in areas such as human resources and enterprise resource planning (ERP).

Data Warehousing is not Limited to Business

The advantage of accessing and processing information in data warehousing has been the basis of competitive information systems in many arenas including "telecommunications, retailing, banking/finance, insurance and pharmaceuticals" (Inmon & Terdeman, 2001). However, the usage of data warehousing is not limited to commercial

use only, but could also be used for governments and security agencies to protect and safeguard against threats from individuals, organizations, or countries. The major value of data warehouses for national security is that it gives the analyst the opportunity to look at information on a detailed basis.

Data warehousing can store large quantities of historical data over time, this is a great tool for governments and security agencies, since they can utilize this instrument to protect and defend the country against terrorist groups. By storing information from multiple sources and integrating the data, governments can monitor the activities of individuals coming in and leaving the country and establish a pattern that will highlight unusual behavior. In addition, government and security agencies can create profiles for individuals by gathering details and analyzing them over time. This safeguarding will enhance the national security of the country. Integration can be performed across a wide variety of systems and sources since the data warehouse allows for flexibility in usage.

Governments

The market for data warehousing products and services in government is booming. The data warehouse allows

governments to access valuable information, and provide better services to citizens. The expanding market of data warehousing includes both federal agencies and state and local governments.

According to a study by Gayle, Knowles, Saputo, Whitney, Yorkos on the web site:
www.geocities.com/datawarehouses/GayleKnowlesSaputoWhitneyYorkos_ResearchPaper.doc; the "Federal government spending on data warehouse-related projects is expected to rise from \$579 million in 1999 to \$911 million in 2004, a cumulative average growth rate of 9 percent" according to the Vienna, Virginia based IT Market Research Firm (Gayle, Knowles, Saputo, Whitney & Yorkos, 2001).

The report also indicates that State and Local government spending is rising for implementing data warehousing where it is estimated to grow from \$550 million in 1999 to about \$1.1 billion by 2004.

Security

Another benefit of data warehousing to the national security is that data can be easily and quickly accessed for analysis. This is especially important for people accessing information let us say at an airport, where they need to have information on hand prior to allowing the person entry to the country.

At a national security level data at the warehouse needs to be related by means of a "pass-through key, where the pass-through key has meaning only within the context of the data warehouse" (Inmon & Terdeman, 2001). This feature allows individuals and organizations to be called anything or multiple things without losing track of the true identity of the individual or the organization.

Data warehouse provides the support for more effective decision making and strategic policies through its tools. Inmon in an article "Data Warehousing and National Security," considers the advantages of implementing data warehouses for security purposes. Many advantages are mentioned in the article, however the major advantage pertaining to the security is the ability of the warehouse to store large amounts of historic information that allow the analyst the opportunity to look at things on a detailed basis. Some of the other advantages mentioned in his article are the ability of the warehouse to gather information from multiple sources. The data in the warehouse can be integrated into a physical structure that helps with the sharing of the information, without changing the format of the data. Data can be arranged for long-term viability of the physical security of the data warehouse.

All these features are important since the data warehouse can monitor and develop trends for unwanted activities. The great advantage is that a data warehouse can be shared across countries which can help government agencies track criminals, and assist in locating and extraditing them.

Custodians

The demand for faster and more detailed information has required custodians in Canada to make a large investment in technology by utilizing a data warehouse tools to collect, and store data for future analysis.

Prior to the data warehouse there was a need to minimize the risk of failed settlements, which resulted in the establishment of the Canadian Depository for Securities. The data warehouse created a global custody network that "support real time application, multi-currency, trade date, custody and accounting systems" (MacMillan, 2002). Thus with the application of the technology global custody companies are entering more markets and replacing local custody suppliers along the way. The technology will enable the custodians to compete in the future for the next 25 years.

Geographic Information Systems

Geographic Information Systems (GIS) were developed for government planners and mapmakers; however, they are now used increasingly in corporate settings. The spatial data warehouse and GIS allow user access to "vast stores of integrated, subject-oriented, geographic-related, and historical data" (Consulting - ESRI Professional Services, 2001, Spatial Data Warehousing SDE, Para. 1).

Ralph Kimball in his article: "Spatial Enabling Your Data Warehouse", correlates between linking data warehouse to Geographic Information Systems (GIS). Data warehouse, have geographic entities including addresses, point locations, sales districts, and higher-level political geographies. GIS data warehouse revolve around such unfamiliar terms as "vector data sets, cadastral databases, spatiotemporal information systems, and helical hyperspatial codes" (Kimball, 2001).

So far, organizations have not optimized the utilization of GIS tools to capacity by linking the data warehouse to GIS systems. Kimball explains the simplicity of the process. In addition, he discusses the benefits of implementing this synergy to organizations that need to monitor day-to-day operations. For example organizations may want to know whether customers come to their store

because of its location, or the proximity of their residence; in addition how can organizations reduce store lines and arrange faster flow to satisfy customer needs.

Data warehousing can be used in business development where a synergy can be built between GIS systems and a large data warehouse. Business people looking for existing opportunities for growth could use the warehouse. For example a retailer can use the information in the warehouse for building their new store location based on the data in the warehouse. The information would consist of the demographics of the area, the local and federal zoning regulations. This information would be beneficial for organizations targeting certain groups.

ESRI, a Redlands, California company, is the leader in developing GIS, and defines it as "taking the spatial components that already exist in the customers' databases such as store locations, telephone pole locations, and transportation routes and connecting these components to a physical location somewhere on the surface of the earth" (Consulting - ESRI Professional Services, 2001, Spatial Data Warehousing SDE, Para 1).

The Australian Geological Survey Organization (AGSO) provides the public with detailed information on the location, time and size of earthquakes, nuclear explosions

and volcanic eruptions as they occur within the Pacific region. It has the information in a data warehouse in the form of maps, as well as geoscientific information databases. The users of the website have faster response time to their queries, without going through to AGSCO and by just accessing the site.

AGSO provides its clients a faster way to access information through requests and queries. AGSO was able to place its databases on the Internet and no longer needed its information locked away in corporate databases. The users of AGSO's Website include government agencies, mineral and petroleum exploration companies and consultants, research agencies, universities and schools, private geologists, and the general public (Seismic safety increased as brio reports enables public access to geoscientific databases on the internet. Australian Geological Survey), in addition about 20 per cent of visitors to the AGSO website are international users.

CHAPTER THREE

DESIGN AND TOOLS

Organizations globally recognize the importance of data warehousing to their ongoing ability to innovate and compete in the world market. Problems occur in data warehousing when analysts and designers do not comprehend needs of the organization and hence are unable to design a fully operational system.

Thus, the data warehouse problem is not with the data warehouse technology, but with "people's failure to properly think through the business and organizational value that needs to be delivered" (Karakizis, 2002).

Implementing a Data Warehouse

Data warehousing technology itself is simple. Thus it is important to prove to business owners that implementing a proper data warehouse is important to compete in the information age. Technology helps with the decision making process, analysis design, and will be more cost effective in the future. Thus business owners should look at the cost benefit effect of implementing the technology for the long run of the business.

There are key elements that will support the success of the data warehouse when implemented:

- Identify key people in organizations who make decisions and have staff working for them to put together the proposals for new business initiatives.
- Long term planning where the data warehouse will be able to support any changes in the future.
- Collecting data that supports the decision making process where most businesses have business processes that provide guidance on future investments.
- Developing fact tables that directly measure the companies' performance over time.
- Developing various scenarios that will incorporate any future changes in the environment. This will be included in the forecasting.
- Gathering detailed transaction that will assist in future decision making.
- Using star schemas to store the most detailed information.
- Consider security conditions such alternative storing areas and backups.

Data Warehousing is experiencing a high failure rate-as high as 90 percent according to a 2000 report on

data warehousing by Conning Insurance Research and Publications (Voelker, 2001). The failure is not due to technology, but to lack of proper training of personnel, as well as poor management skills and inadequate business practice and planning, which accounts for foremost reason of failure.

Issues in Designing a Data Warehouse

A Data warehouse "is an architecture, not a technology" (Inmon, 2001a). Although the term technology is often used to describe a data warehouse, it is actually an architectural form, however the terms are used interchangeably. Data warehousing can exist in various forms, however the standard form is known as the quick online response time.

Prior to designing a data warehouse, the data warehouse administrator has to understand management's objective for building data warehouse, the administrator's responsibilities will be covered in more details in the next section. The administrator has to comprehend the various functions within the departments to design the proper information flow that will enable the user to utilize the information effectively.

Figure 4 shows the relationship within the data warehouse structure. Whereby the source in the data warehouse is data whose content can be materialized in the data warehouse. The figure illustrates the flow of information within the data warehouse where the data is extracted from the database and the source data and then transformed to the various applications.

The meta database is a repository for information about other components such as schema of source data. It is data about data. Metadata is information that describes the contents of a database. Metadata informs users when data was last updated. It enables the users to understand trends in the organizations, as well as assist in business transactions.

In a typical data warehouse there are various processes for the data to go through. There are those who are responsible for getting the data, cleaning it, and storing it, so that the end user can utilize the data for better managerial and analytical tools.

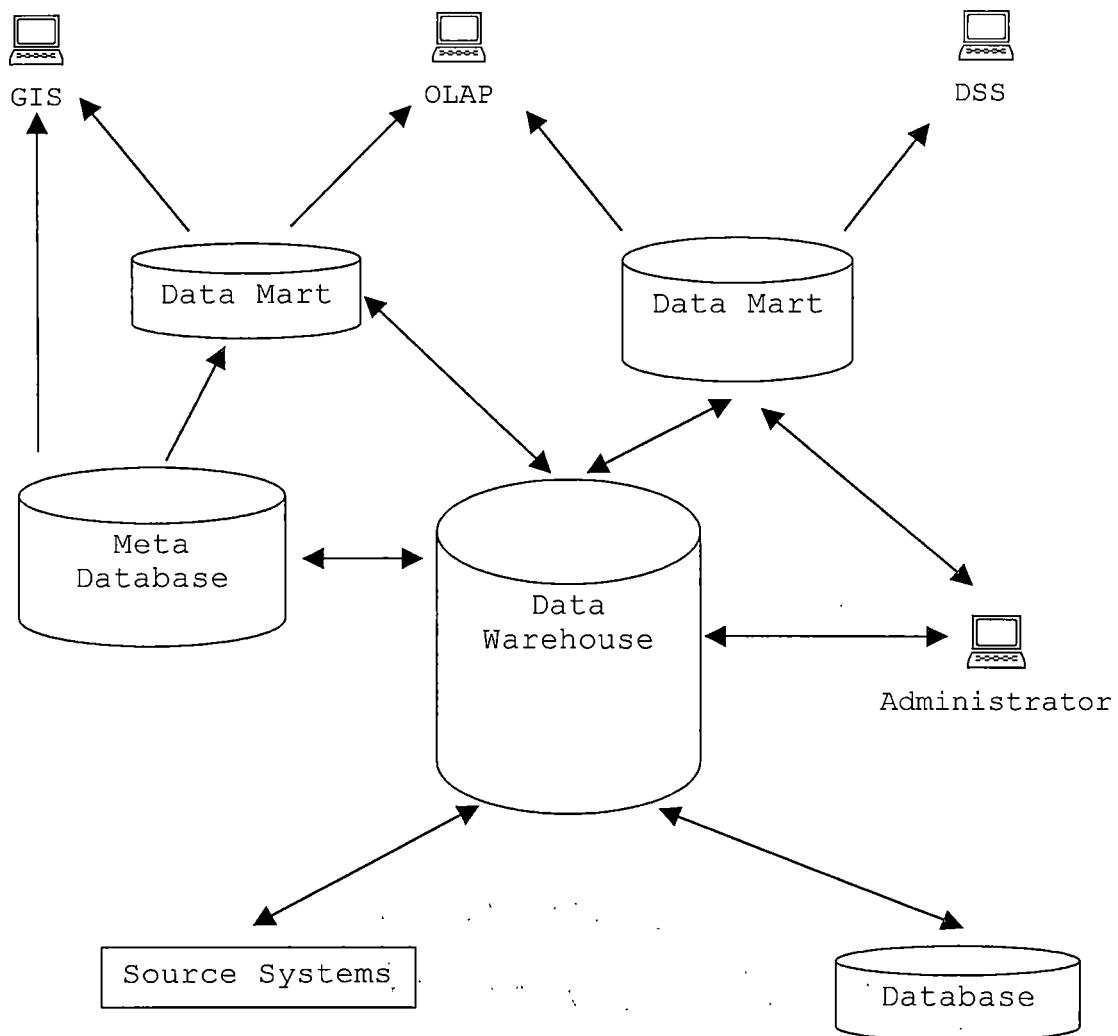


Figure 4. Structure of the Data Warehouse

Roles Within a Data Warehouse

The role and tasks performed within the data warehouse architecture include (Connolly & Begg, 2002, p. 1058).

- The Load Manager who performs all the operations associated with the extraction and loading of

data into the warehouse and is also referred to as a front-end component. Some of the transformation includes simple transformation of the data to prepare the data for entry into the warehouse.

- The Warehouse Manager performs all the operations associated with the management such as analyzing data to ensure data consistency, and merging the source data, in addition to creating indexes and views.
- The Query Manager, is also referred to as the backend component, performs all the operations associated with the management of user query. This includes directing the queries to the appropriate tables

There are activities associated with the processes of data flow within data warehouse architecture (Connolly & Begg, 2002, p. 1058) to visualize the process please refer to Figure 4. It is important to identify the processes to understand the stage in which data is at.

Data Process

- Inflow is the extraction, cleansing, and loading of the source data

- Up-flow is adding value to the data in the warehouse through summarizing, packaging, and distribution of data
- Down-flow is archiving and backing up the data in the warehouse
- Out-flow is making the data available to end-users
- Meta flow is process of managing the meta data with a description of the content of the data warehouse in terms of what is in it, and where it came from.

Data Warehouse Administrator

To effectively implement a data warehouse, a data administrator is required. The administrator has the responsibility of analyzing the requirements of the organization. Prior to implementing a data warehouse the organization has to identify the benefits and drawbacks of implementing the data warehouse.

The job of managing the warehouses falls to a group of people called the "Data Warehouse Administrators" (DWA). They ensure that the data model is complete with all the data defined and categorized, as well as the relationships recognized. The database administrator (DBA)

performs the project-design, monitoring, tuning, backup, recovery. The two administrators work together, however, this project will concentrate more on the functions of the DWA.

The administrator has to understand the entity-relationships in the organization. The latter concept is what enables the user to understand the dimensionality modeling which is a logical design technique aimed to present data in a standard, intuitive form that allows for high-performance access (Connolly & Begg, 2002, p. 1079).

These are some of the questions that the DWA should consider while developing the data warehouse (Inmon, 2002a).

- What should the database design look like?
- How should every record be time-stamped?
- How can the costs of the data warehouse be justified?
- Who should be leading the development effort?
- What tools should be selected?
- How big and how fast will the data warehouse grow?

According to Conning's, (Voelker, 2001) it is important to have scheduling for data warehousing as per

the 80/20 rule: 80 percent of the work involved in implementing a warehouse involves examination and preparation of source data, while only 20 percent is required to extract, transform, and load the data into the warehouse.

The data warehouse administrator forecasts the budget needed for implementing and managing the data warehouse. The budgetary expenditures include: hardware costs, software costs, consulting and service costs, and personnel costs as well as ongoing maintenance costs. It is important for the administrator when developing the budget to forecast future needs of the organizations and take expansion into account. "The administrator should build cost-effective, distributed systems, and add incrementally to the logical and physical design" (Kimball, 2002). The administrator has to start with a specific business objective and then decide how, or if, a data warehouse can support the achievement of that objective.

When designing a data warehouse the data administrator has to interview and research people in the organization by gathering information from both management as well as the users of the data to understand the needs and requirements of the organization. The key issue here

is communication since this what will make or break the project, thus involving the staff and getting their input is crucial.

Management has to fully support and be committed to the project by giving time, and money for it to be more effective. However when developing the data warehouse it is important to note that it is a work in progress, since the external environment is always changing thus when designing a data warehouse it has to be flexible to accommodate those changes. It is important to look for productivity gains, assess marketing opportunities, and evaluate the results of precision pricing.

It is critical for the administrator to test the data mart and data warehouse with sample data before loading the information; by building a small-scale data warehouse to vent the system and uncover problems.

"Newsletters, training sessions, and ongoing personal support of the end-user community should all be part and parcel of the first rollout of the data warehouse" (Kimball, 2002). It is important to have training sessions to users to be able to use the system more effectively.

Online Activity Monitor

It is important for the data administrator to monitor the activities in the data warehouse and perform analysis

based on the programs used to access the information such as using SQL for example. Monitoring determines the type of queries being used; establishes whether the queries are helpful in obtaining the necessary information, and verifies the performance of the data warehouse.

Activity Monitor

"The activity monitor helps the administrator in determining the dormant data in the system" (Inmon, 2000a). In addition, it helps the administrator with identifying the number of times the information is accessed, distinguishes who uses the information, analyze the types of queries being run, and establish the time of day the information is being accessed most.

Monitoring the Data Warehouse

The primary issue the data warehouse administrator looks for in the monitoring of the data warehouse environment is "that of the growth of data in the environment" (Inmon, 2000a). When organization growth is adequately managed, many other benefits follow, such as the control of Budget, Performance, and New and innovative uses of the warehouse.

Firms are required to understand the nature of their business needs usually through trial and error. The management should then modify the business practices to

successfully utilize the data warehouse to achieve higher efficiency.

Many warehousing projects break down after operation due to people's failure in recognizing the ongoing support costs (resources, time, and coordination) of refreshing the data.

Transaction Processing Systems

The data warehouse is not built overnight from scratch. It involves a collection of data over time. Transaction Processing Systems (TPS) are "computerized information systems developed to process large amounts of data for routine business transactions such as payroll or inventory" (Kendall & Kendall, 2002, p. 2). TPS allows business users and analysts to collect and create all the data they need, in addition it enables the user to access information from external sources, and delivers up to date information. However, it is the data warehouse tools that enable users to extract summarized information and present it in the required format for analysis.

The maintenance involved with TPS activity is burdensome since it involves continually collecting large quantities of data as business grows. "For example, marketing systems that previously tracked product sales

would now have to adjust to the expanding business and track service records, additional distribution channels, industry sales data and a deeper history" (Kendall & Kendall, 2002).

Data Warehouse Iteration

It is recommended that the data warehouse be built in iteration, and not at once; where the project should be broken into parts. The management will assign an objective for each part, and once the objective is met the management will complete the other parts. Figure 5 shows how the various iterations are built where the 1st iteration is built and is then populated by data. Then the 2nd iteration of the data warehouse is built and populated and so on. With the various iterations the administrator has to ensure that the system goes through the complete lifecycle of requirements, that include: proper design, testing, and implementation of data. However, throughout the building process the user can use the warehouse and provide the administrator with feedback. The data is extracted from the various data sources and database.

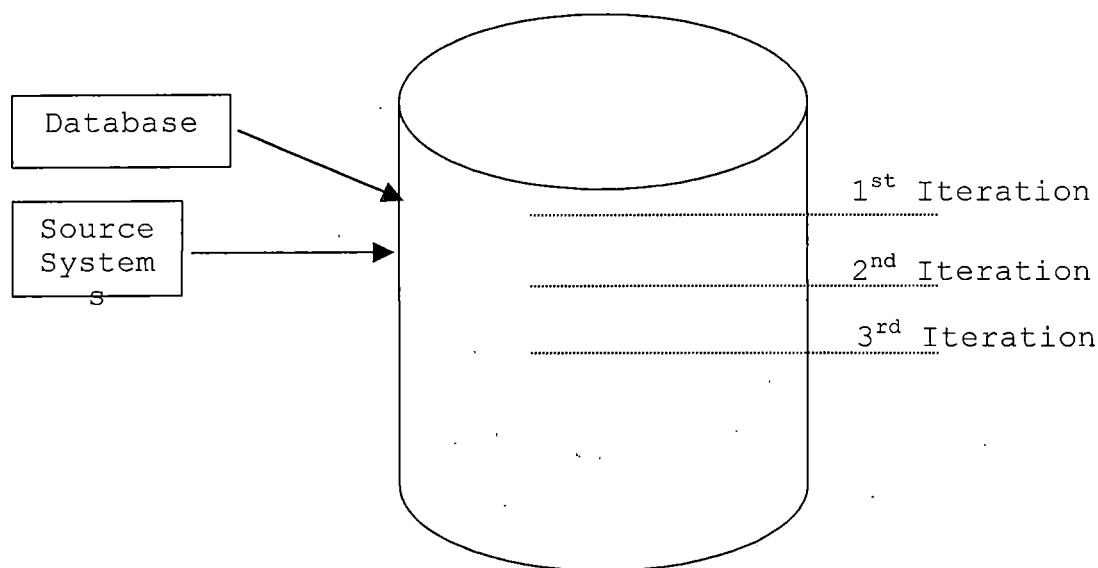


Figure 5. Data Warehouse Iterations

Quality of Data

Data Warehousing is only useful to an organization if it is used correctly, and maintained regularly to ensure high quality of data. Management has to encompass in its organizational policies and procedures a structure for responsibilities that implement data quality to enhance the application process of utilizing the warehouse. The quality control of the data warehouse is reflected in accessibility, usefulness, believability, and other validation factors.

The data warehouse should be accessible and available to everyone within the organization, and users should be able to access various tools within the data warehouse for

their transaction processing such as the source of data and the process application. However, there should be some privileges assigned to certain people to ensure that the security of the system is within the control procedure.

Users should be able to interpret the information provided to be able to use it for analysis, and better decision making purposes. The data should be complete, up to date, credible, consistent and accurate to ensure the validity of the information.

Many companies have redundant data, wrong data, missing data, and miscoded data in their systems, which "leads to miscommunication within the departments and with clients" (Nash, 2002). The DWA should therefore, monitor data closely to avoid miscommunication, especially on large projects such as installing enterprise resource planning (ERP), or supply chain management software (Nash, 2002). The miscommunication problems occur between different departments within an organization who run on various softwares and share information. The data becomes more difficult to organize and standardize for the different applications, hence the administrator has to cleanse, reconcile and integrate data from various departments within the company.

For example there are different sales and inventory recording methods at manufacturing factories, or at retailers at the point of sale. There are different measuring units, such as the European metric versus what the US uses; for example an item can be marked by different descriptions such as extra large (XL) in the US and Très Grand (TG) the French equivalent term in France or Canada. Both terms are used interchangeably and due to globalization retailers in different countries have the conversions at their stores. However the problem may occur when entering the product description at the different locations, and merging data. There should be a standard used for the different measurements to avoid errors.

Shell Exploration and Production wanted to combine data from their SAP AG financial applications with data from their "mishmash of volumetric systems, which process information on how much gas and oil the company finds and collects" (Nash, 2002). Shell realized the expense involved in the process, and they decided to use a tool from Kalido Ltd., a company based in the UK, that maps the data from various systems within the organization and combines it into one warehouse.

There is another cleansing tool from Trillium software, which searches for duplicate and incomplete

information. The software helps the user fill in missing information by matching information, such as Zip codes and compare it to the Zip code database in the system. For example if there was a number missing in the entry the system will alert the user; or if the Zip Code entered does not match the zip codes in the database, the user will be alerted. The data is then put in a warehouse to be analyzed and mined by sales agents. Having this tool would help the administrator greatly since the data is cleaned and does not have missing information prior to loading it to the warehouse.

Applying tools from vendors such as Kalido and Trillium, will save organization money and time, since it will increase the confidence of the user in the integrity of the data. The tools will also help the company in the long run where incomplete data, redundant data, and uncleaned data entries will be reduced. Having these tools will also assist in communication within the various departments since they will have common data to go by. Organizations have to benefit from these tools however they have to keep in mind that these tools are not a complete solution, and that continuous monitoring and control of data entry and application is important.

Testing Data

In the data warehouse, testing of data for decision making is problematic according to Bill Inmon in his article "The Data Warehouse and Testing", published in DM Review in February (2002a) unlike online transaction processing where everything is tested prior to implementation, and there is software and database that can be tested. It is important to have testing for data warehouse to ensure the quality of data is suitable for proper decision making (Inmon, 2002a).

Data in the warehouse is not regularly tested for many reasons. The main reason is the high cost involved in testing the data in the warehouse. The data warehouse is expensive as it is with the cost involved in the processors, and software needed to control and manage the terabytes of information.

Another reason is the fact that there are no programs to be tested, since when new data is added to the warehouse, the data is simply added. Data warehouses are built gradually, and any additions become part of the warehouse, thus if testing is to be performed it should be done on the new data not the whole warehouse infrastructure.

Data needs to be tested, modified, and verified within the data warehouse on a periodic basis, especially as business grows and changes occur in the business environment. This is important for achieving a competitive advantage in the business by maintaining effective and useful data, which will assist in better decision making, and place the company at a strategic position in the market.

Measuring Capacity

When measuring the capacity of Online Transaction Processing OLTP the unit of measure is transaction per second. This form of measurement is not applicable for data warehousing since the functionality of the data warehouse differs. There is inconsistency in the decision support workload in a data warehouse; in addition, the decisions that need to be made are more strategic. Whereby it is important for the data warehouse to manage data effectively for decision support rather than running transactions.

OLTP has a high degree in predictability once the pattern of access is established, in addition it does not vary in the transaction process. However, the data warehouse is more variable and changeable, in magnitude

and scope since it depends on the needs of the organization as well as the size of the data.

In data warehousing, the basic measurement of capacity and performance is "based on the effective number of gigabytes of storage that a hardware platform and/or a software platform can manage" (Inmon, 1999).

Although data warehousing is not a new tool, there has not been a measuring standard for capacity. However, there needs to be a way to manage and monitor capacity and performance to evaluate cost effectiveness of the organization.

Dimensionality Modeling

Within the dimensionality modeling of the data warehouse there are several model types, such as:

- The Star Schema is a logical structure that is better for flexible access of data, which is optimal for decision support systems (See Figure 6). The design contains factual tables and dimensions; dimension tables containing reference data, which can be denormalized, surround the data in the center. It provides multiple levels of summaries to monitor applications.

- The Snowflake Schema has dimension tables, however it does not contain denormalized data.
- The Starflake Schema is a hybrid of both the star and snowflake schema, and combines the functionality of both designs.

The Star Schema is a simple database design; it presents the multidimensional character of the data and allows for fast querying of the data in a data warehouse. In Snowflaking, some of the fields of the dimension tables are split into separate tables. This achieves a higher level of normalization, but makes the database design more complex and can reduce the performance and ease of use for decision making.

The Star Schema as presented in Figure 6 reflects the entity relationship between the various tables, and reflects how the data is being stored. The basic factual information is in the middle of the star. The points of the star represent various perspectives from which the factual information can be viewed. Each of the dimension tables has a single field primary key, with a one-to-many relationship with a foreign key in the fact table. The primary key is selected to uniquely identify each occurrence of an entity type. For example the primary key for Store is the Store_key, which is then used

(translocated) in the Sales fact table to relate the information. Thus when the user needs to analyze the overall sales generated by a certain store the Sales table can generate this information.

The Star Schema is an intentional simplification of the database design that would be achieved by following the standard rules of normalization.

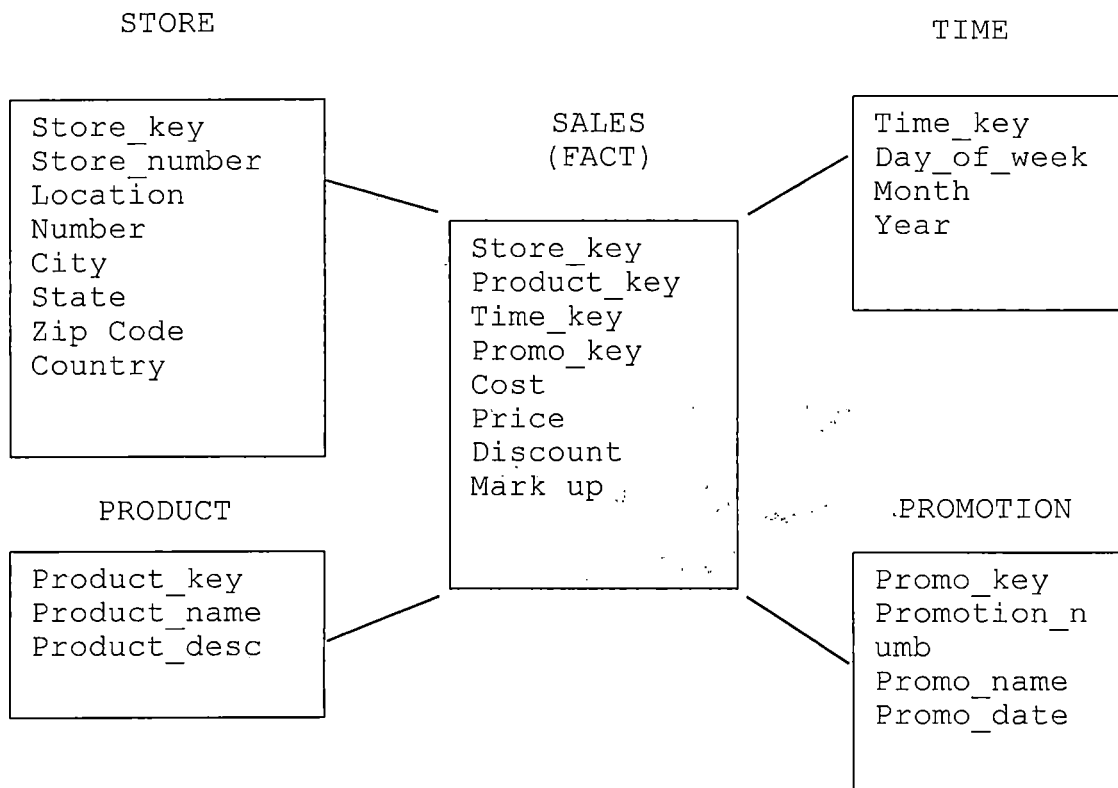


Figure 6. Star Schema

In dimensional modeling there are three forms of fact tables:

Transaction grain contains exactly one record for each individual "transaction in the source system from the beginning of recorded history (Kimball, 2002b).

Periodic snapshot grain transactions are viewed as transactions occur, for example a bank may use this for their large database of accounts, which need to be accessed on a monthly basis.

Accumulating snapshots of functions that are used for short-lived processes like orders and shipments; where the record is updated frequently as an activity occurs.

Whatever the design the objective is to have an efficient data warehouse where the user can access information using the various warehouse tools. When planning the warehouse design the administrator has to take into account the growth of the company, as well as changes in requirements that may occur as the company increases its operations. It is imperative for the design to provide fast response time, and provide decision support to the user.

Thus having realtime partition for the conventional static data warehouse is important, where the user receives updated information for decision making purposes. The realtime partition covers activities from the time the last update was made; it links to the grain and content of the static fact table information.

Open-Loop/Closed-Loop Architecture

The more traditional architectural form is the Open-Loop Architecture; which came as a result of the integration of business and computers. An open-loop architecture is one where the business decisions have an impact in the marketplace, and the "impact of the decision is measured indirectly by the corporation's computer systems" (Inmon, 2001b). This form of architecture is used with companies that do not know their customers and are the supplier of the product. In addition, it is used in organizations that conduct mass marketing.

With innovation offered by the Internet and data warehousing, a new form of architecture is developing. This is referred to as the Closed-Loop Architecture. Customers are approached directly, since they are given names and numbers. A Closed-Loop decision environment is one where the impact of decisions can be measured

precisely as they happen. In a Closed-Loop decision environment "management makes a decision, the computer system singles out exactly to whom the decision applies, and a direct action is taken as a result of the decision" (Inmon, 2001b). The results of the management decisions can be measured quickly and directly since the feedback and response time is quick.

There is a change in the architectural trend to meet changes in organizations requirements, where organizations are strategically positioning themselves for instant decision making. There is a movement towards Closed-Loop Architecture since the architecture offers more support to decision making within the organization. There is a cost involved in the development of Closed-Loop Architecture, which is the cost of the infrastructure required for the Web environment. However, the advantages of having the Closed-Loop outweigh the cost, since it will allow organizations to use the information to on hand to make faster decisions.

Enhance Performance

With time, the data warehouse's performance becomes a significant issue, especially as data compiles, thus it becomes necessary to tune the system, hence management has

to address this issue along with the administrators. It is important to tune a data warehouse system, especially following implementation.

Maintaining the performance of the data warehouse is imperative, since a more efficient data warehouse will help in better decision making, and utilization of the resources. The efficiency is related to updated information where data is accurate and relevant.

Tuning starts with monitoring the activities and data flows through the data warehouse and the decision support environment. When a "system is tuned, the system yields better performance without making major and wholesale changes to the system" (Inmon, 2001c).

In the initial stage of implementation the system is constantly being changed to fit the management needs through trial and error application. However, when the system is operational it is important to tune the system on a continuous basis for better performance.

Tuning the data warehouse for decision support purposes implies enhancing performance "that stop short of a wholesale rewrite and reconstruction of the data warehouse or a removal of the data warehouse off of one hardware platform and onto another platform" (Inmon, 2001c).

Within the data warehouse there are two functions: the operational transactions and the decision making process. The operational transactions are mission critical, and require two-to-three-seconds response time, and use a small amount of resources. However, the decision making system are not mission critical and use a longer response time. It contains one to two orders of magnitude of data more than the operational environment (Inmon, 2001c).

Thus, when it comes to tuning, the data warehouse administrator must fully understand the system. This involves monitoring and observing data flow and comparing it to the needs of the company. The observations have to be made prior to as well as after the tuning to ensure that all changes are made effectively. The administrator has to determine what data is or is not being used. Once the DWA determines what data is not being used, the data should be eliminated thus making the decision making environment more effective in processing and analyzing information. This will save resources in terms of time and space especially in the long run when the company grows and has more data on hand.

According to W. H. Inmon, in both his articles "Enhancing Data Warehouse Performance" and "Tuning a

Warehouse for Performance", mentions the following tools that enhance better performance:

Creating Indexes along the patterns of regularly accessed and used data in the warehouse, will reduce system resources required to locate data. The indexes will help users access the data needed much faster. The administrator has to monitor data closely to ensure that the data in the system is correct, and that the indexes used correspond correctly. However, the setback of indexes is that they require their own space for storing the information.

Creating Data Marts, allows end users to specify and extract a selective subset of data from the data warehouse. Data can be summarized, reduced and customized to meet with the management requirements. Furthermore the individual department can use the customized information. The cost of machines for the data mart is lower than the cost of machines for the data warehouse.

Creating Summary Data. This summary is made available to the end user in the data warehouse, and not only the user of the data mart. The data will be

faster to access, and it involves a process of summarization that describes the nature of the summary itself called "meta process". The DWA will know how to create summaries by monitoring the data warehouse activities.

Creating Star Joins or snowflake structures: these tools streamline the process with the aid of a fact table and a dimension table. Data that occurs frequently is stored with the related data. However the problem with these structures is that the administrator requires a full understanding on what the data will be used for, in addition different people have different uses for the data.

Developing Profile Records by aggregating together many detailed records. The user will only have to access a single profile. This is an efficient method of getting data. The administrator will develop this by analyzing the data activity.

Partitioning Data Finely, through a parallel system: this will enhance the processing data and hence increase performance. The administrator looks at the usage statistics for activities in the data warehouse.

Eliminating dormant data, this involves removing data that is not being accessed or used. By getting rid of the unwanted data, performance will be improved; since data takes up storage space and utilizes existing resources. By eliminating unneeded data users can make better decisions.

In addition to providing faster response time and improving performance, the administrator can opt for updating the existing hardware and software. Depending on the organization's budget, the administrator can purchase faster disk drives, bigger Central Processing Units (CPUs), as well as fast performing parallel processors such as Symmetric Multiprocessing Systems (SMP). The administrator can purchase specialized software such as multidimensional databases (MDDs), in and bitmap indexing databases, as well as utilize OLAP tools.

As the data warehouse grows the amount of information in the system also grows, and the administrator has to plan for the growth. Thus it is important to have information that is vital for decision making purposes, eliminate unwanted data, which will result in getting information faster. Therefore, the data warehouse administrators have to incorporate these performance suggestions into the data warehouse system, to ensure fast accessibility.

Augment Security

Security is an important aspect of any data warehouse and can not be ignored. However covering the various security issues is too broad of a topic. This project will just touch on certain issues.

The flow of data is the livelihood of any organization. Therefore, keeping data secure is becoming an imperative issue that management can no longer ignore. Organizations have to plan on making backups on a regular basis to ensure the security of the data at all times. This process protects organizations against unforeseen threats such as viruses or terrorist attacks (such as September 11) that impact to a complete loss of data. Management has to consider various alternatives such as remote backup areas to safe keep the data, as well as safe storage areas where the backup disks can be kept. The data warehouse administrator has to take the security issue as an important factor in planning and budgeting for the data warehouse.

It is important to control the flow and access of information. Although the data warehouse should provide information access to everyone in the organization, there are still data of certain security or confidentiality in the warehouse that should be monitored and have limited

access, this includes information such as financials and payroll as well as budgeting for research. Assigning passwords or encryption that will allow the user to retrieve the information can control the access. The administrator should understand the system flow to carefully allow the information to be used and viewed by the various departments. Management should also determine who has access to information and convey this information to the administrator at an early stage of building the data warehouse. The administrator has to look for firewall and other softwares that that will prevent hackers from accessing information.

Data Warehouse Tools

As mentioned in Chapter One there are BI tools that are used by the organizations, which will assist the end user needs to make strategic management decisions. The data warehouse must be able to "efficiently support ad hoc and routine analysis" (Connoly & Begg, 2002, p. 1115). The following are tools used by the end user to retrieve and analyze data in the warehouse:

Reporting and Querying tools include production reporting tools and report writers. The tools

used are designed to accept SQL statements and structure

Application Development tools include applications

associated with Online analytical processing tools OLAP tools and can access major database systems including Oracle, Sybase, and Informix

Executive Information System (EIS) tools are designed

to support high-level strategic decision making.

However it now includes all levels of management

Online Analytical Processing (OLAP) tools are based

on concept of multinational databases and allow a sophisticated user to analyze data using complex, multi dimensional views.

Data Mining tools discover meaningful new

correlations, patterns, and trends by mining large amounts of data using statistical, mathematical, and artificial intelligence (AI) techniques

The most popular tools in terms of usage in data warehousing are Online Analytical Processing (OLAP) and Data Mining.

Online Analytical Processing

OLAP describes the tools that can extract multidimensional data and present it from many different

points of view. It is designed for managers looking to make sense of their information, by providing trend analysis, detailed information, and summarization of data.

Data Mining

Data mining is the process of extracting valid, previously unknown, comprehensible, and actionable information from large database and using it to make crucial business decisions (Connoly & Begg, 2002, p. 1115). Data mining can identify patterns that a human is unable to detect, through a search of relationship within data, in addition it allows users to predict future trends and behaviors, allowing companies to make proactive decisions based on the data.

Information Management and Data Warehousing

Information management provides the opportunity to increase the individual and organizations' productivity through better decision making, and better time management of individuals and resources. To achieve better decisions one has to look at the flow of the information within the organization, as well as assess the life cycle of information.

Figure 7 shows information mapping and the life cycle of information matrix as presented by Dr. C.E. Tapie Rohm

at California State University, San Bernardino (Info 646) on the 9th of April 2001.

It is important for managers when building a data warehouse to understand the life cycle and stages of information. For example where the information is inputted from, where the information is stored, and the form in which it is stored, and how it is used to make decisions. The managers have to agree on how to output the information and distribute it so that is useful. Management hierarchy is flattening from the traditional pyramid format that existed to a flat model. The flattening of the structure is allowing for smoother information flow within the organization, whereby empowering employees at various levels within an organization to access information, and make decisions.

The information can be accessed more rapidly; however, it is still vital for managers to understand how the information is being used within the organization and at what stage it is at.

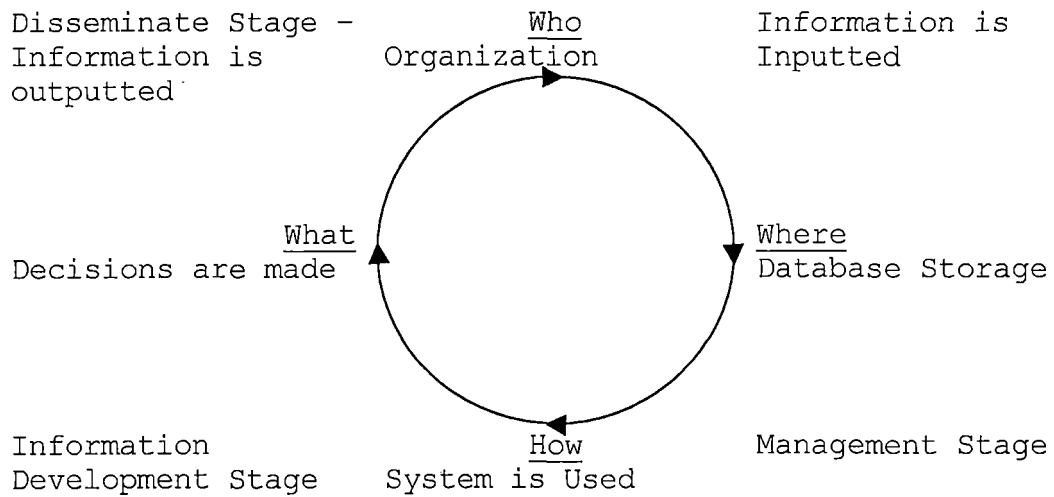


Figure 7. Life Cycle of Information

By transforming data into meaningful information, managers can perform more substantive accurate, and consistent analysis. Managers need summary information, which they can use to analyze trends that affect their organization.

Managers need to find the critical issues affecting the success of their organization, and how to adjust those elements to improve the success of their business. In addition, it will help them identify future trends in the environment and assist them in making proper forecasting decisions. Data Warehousing can be used in various forms:

Financial Data Warehouse focuses on capturing and analyzing financial and accounting data such as General Ledger, Accounts Receivable, budgeting and forecasting. There are a small number of

users in the financial data warehouse, and small volume of data. It does not require real time information, however it assists management in making forecasts as well as budgets

Customer Data Warehousing is used by banking, retail, health care, insurance, and telecom. It is used for customer relations' management CRM. The goal is to analyze customer behavior to improve market efficiency; hence management can react to changes in consumer purchasing trends. The data is large and dynamic since it is constantly updated

Transactional Data Warehousing monitors how various events effect customer loyalty. This supports management in making decisions about changes in their product or service.

Organizations have to define their needs prior to implementing an effective data warehouse that will assist in better decision making. Once the data warehouse is implemented it is more important that it used effectively to ensure the projects' success.

Prior to data warehousing, data was stored for transaction processing, not access and analysis, and data that has been stored for more then two years was

considered unnecessary data. There are certain things that can be examined with deep historical data. For example, fluctuations in interest rates. Some organizations may be interested to see how the fluctuations effected the business activities over time, this information can help management with forecasting sales and prices, using past information.

Organizations can perform more accurate analysis and forecasts when they have old data that can be traced to economic trends over certain periods of time. These analyses will help management; determine whether a recession is over, in the middle or just starting. There are many aspects of business that can be looked at over time, deep time. Thus, organizations have to utilize the historic data and not just store it, since it will help management in their decision making.

The Management Decision Making Process

The Management Decision making Process (MDMP) ensures all the conditions are set for starting the project and gives a "consistent reference point for judging the project's ongoing success" (Nolan, 2001); since it is important to remain focused on the goals set forth by the management.

With the availability of data warehousing, decision makers can gain access to information that was previously unavailable, and untapped. Once the information is analyzed, it assists the end user in developing trends, forecasting demand, and evaluating their business performance overall. This is a great tool that gives organizations a competitive advantage.

Decision makers can evaluate their business on a continuous basis, by assessing how their business is performing over time, and identifying warning signals (red flags) that will appear through the reports that are generated in a timely manner. The reports will contain up to date information that is accurate, valid, and relevant to the credibility of the decisions. The decision makers will be able to retrieve both detailed and summarized reports depending on their needs.

Having the reports available decision makers can contemplate alternatives and choose the appropriate option for better business returns. The data warehouse assists in the buildup of a corporate memory with initiatives and results. The results become a significant source of information for future business initiatives.

When presenting a data warehouse to the decision maker of the organization, the presentation should include

the reason for implementing the data warehouse. The presentation should illustrate the need for implementing the data warehouse. The presentation should cover the following areas: identify weaknesses in the information flow, indicate areas of information access and use that the organization currently lacks, the improvements the data warehouse will have on the organization, and the short and long term benefits versus the costs (cost-justification is presented in Chapter Five).

There is a need to provide information to support decision making needs throughout the organization, which will assist in evaluating management strategies. The Decision Support Systems (DSS) requirements "demand that information is presented in a way that is easy to understand and applies business rules to derive new information" (Laney, 1996). Data is presented in a descriptive format that is simplified and can be summarized, and is organized for more efficient access. This will assure the decision maker of the integrity, availability and comprehensibility of data.

What allows organizations to compete in the market and react to changes in the environment is instant access to the information enabling the user to make more accurate decisions that will reflect current market conditions.

Data Warehousing strategically positions the organization for better decision making. It enables management to make better assessment of where the organization is at a particular point in time, by providing the user with a snapshot of the organization.

Data Warehouse on the Web

Doing business on the World Wide Web (WWW) enables organizations to connect into the online world and improve all aspects of their business. In this innovative environment, business transactions are conducted in electronic form. Thus the technological advances have taken the form of "low-cost, high-speed digital data transmission by utilizing hardware that produces information quickly and easily, and using software that reduces and, in many cases, eliminates much time, space, and other constraints to information" (Rezaee, Sharbatoghlie, Elam, & McMickle, 2002).

It is essential for companies to develop a pattern on their customers' behavior online. Existing technologies allow organizations to "build more effective eCommerce efforts, and maximize the potential of the Web" according to an article in (Accenture, 2002) to gather data. Organizations can combine two tools to assist businesses

in achieving a competitive advantage. These two tools are the business intelligence tools such as data warehousing and the clickstream tools that collect information on users behavior online.

The business intelligence system is applicable for data warehousing and data mining. It captures and organizes data from corporate systems such as operational applications for example finance, and applies them to the traditional data warehouse. The Web "clickstream" tools gather information about online customers such as how many hits a site receives, how long customers stay on a site, and what portals are used. By combining these tools an organization will create a "Data Webhouse" (Accenture, 2002) whereby an interactive database is developed that captures information about customers and trace their behavioral pattern online. The information gathered will also provide opportunities for organizations to collect information on more profitable customers/ buyers; analyze products and services and their price elasticity; and develop a grasp on the websites that are most effective in terms of visitors' use and profit generation. In addition the Data "Webhouse" allows companies to react faster to changes in customer behavior, and thus adjust their strategies accordingly.

In an article titled "Changing the Five P's" (2001d) W. Inmon states that the term market share has changed with the new economy where market share originally referred to "how many products were sold versus how many products did the competition sell from the particular market". However in the new economy, market share refers to "what percentage of dollars consumers spend on products versus what percentage does a consumer spend on anything else" (Inmon, 2001d). Hence unrelated organizations are now competing with each other over customers and looking for ways to lead them into their business.

By going online customers allow organizations to trace the sites visited, and observe purchases made. The information is documented in the data warehouse, and is used by organizations to target customers each time they go online. Thus, the click stream data allows organizations to have more hands-on knowledge on their customers, so the next time the customer goes online suggestions are made according to what the system has.

The Internet, and data warehousing technology have given the consumer more options when selecting their products, by offering more information about the product, as well as a wider range of choices, and by creating a perfect economy where information is available for

everyone. Organizations can have more understanding on the price elasticity of the products by monitoring customers' behavior to changes in price. In addition, the supplier does not have the burden of reaching customers with their products, since the internet can be accessed globally. This allows the supplier to only think about shipping and delivering the product to the buyer. When promoting the product the supplier can measure the effectiveness of a promotion on the customer by analyzing purchases. The Internet allows for cheaper promotion, since it is reaching more people directly and a lower cost. The supplier can measure the attractiveness of packaging a product over the Internet, through the feedback of the customers.

By having information online, suppliers have to realize that their competition is also looking at what they are doing. Thus organizations have to remain innovative to maintain their market position. Moreover, organizations have to use an effective marketing mix to differentiate their products from the competition.

It is important when building a data warehouse to take the above factors into consideration to ensure that the warehouse would be of benefit to the organization. The planning stage of the warehouse is vital since the

administrator needs to know what the management need and the management will have a better understanding of information flow. Data quality is the most crucial element in building the warehouse, since it ensure that the information is reliable, valid and accurate.

Organizations are changing their focus and looking at customer satisfaction, through customer relations' management, which will be covered in the next chapter. CRM is an important element in strategic planning, which is what originations need to do now to be positioned well in the future.

CHAPTER FOUR

CUSTOMER RELATIONS MANAGEMENT

Customer Relations Management (CRM) is part of the organizations' strategic tools that utilizes data warehousing to create an integrated, historical foundation of customer information.

The data warehouse allows organization to perform strategic business planning in order to reach a maximum number of people anywhere and anytime; in addition it allows sharing information between various users which aides in achieving synergies within an organization. Strategic planning provides a formula that helps organizations compete, through a specific goal attainment process, which is crucial for decision making.

The process of strategic planning involves assessing and analyzing current and foreseeable business situations, in addition it helps in identifying strategies that will help the organization achieve a competitive advantage. Once the strategies are identified organizations can execute and monitor the strategies implemented

CRM is the competitive strategic tool that organizations are using in business to maintain old customers and attract new ones. CRM is a strategy for

optimizing the lifetime value of customers, where customers are the main focus of the organizations now, and achieving customer satisfaction is crucial to survive in the market. According to Channel Dynamics (www.CHDYNAMICS.COM) CRM is a two-stage concept. The first stage is building customer focus, by moving from product orientation to customer orientation and define market strategy from "outside-in and not from inside-out". The key elements in customer relations' management are:

Understanding the market. Organizations are examining the market carefully and performing various analyses to evaluate their competition have better understanding of market trends, as well as. Having historic data in the warehouse helps organizations to develop promotions, and sales forecasts, and anticipate customer reaction to external and internal trends within the industry

Segment the market to effectively target. Segmenting customers into groups enables the marketing and salespeople to focus more on customers needs and find ways to meet the demand. The data warehouse enables the user to filter the information needed according to the targeted group. The data

warehouse also helps with cross selling based on the information in the warehouse.

Present a fine product or service. Customer satisfaction is important and having the customer interest in mind increases the perception of quality and enhance loyalty amongst customers.

Follow-up. It is important to follow up with the customer through questionnaires and other marketing tools. The database in the warehouse will have on hand information for the user to follow up on new products. Follow-up is very important in understanding whether the product or service satisfied the customer need. Tools within the data warehouse will be especially beneficial for the launching of new products, and evaluating the reaction from customers.

The major challenge to organizations is maintaining old customers and attracting new ones. In addition to understanding and implementing changes to products and services according to customer needs from the follow up.

Data warehousing allows companies to look at their customers' information and analyze any purchasing trends or habits. The patterns reflect the customers' needs and

show changes in purchasing patterns. These indicators are useful tools for organizations that want to introduce new products, in that it will help them identify customer needs, as well pin point existing problems or flaws with the service/ product. Having this information allows companies to ratify problems in a timely manner. Call centers and interactive Internet sites give customers better access to the company (CRM a revolution in business practice, Channel Dynamics). These sites help the customers make faster purchasing decisions, convey problems and get faster responses to their queries. Companies have to concentrate on the unique value of their products and services, and try to find a market niche or a competitive advantage. Organizations have to realize that maintaining old customers is important since "the price of acquiring new customers is five times greater than the cost of keeping old ones" (CRM a revolution in business practice, Channel Dynamics).

"The quality, power, and usability of embedded analytics will soon be the yardstick by which software vendors and their customers measure the value of strategic business applications" (Ward, 2002).

Traditional data warehousing and data mining allowed companies to uncover hidden business opportunities, by

better understanding customer buying patterns across time, geography, and complementary product lines (Accenture, 2002). The fundamental tool for achieving value with real-time CRM analytics solutions is data mining. It is the mechanism that allows companies to make sense of their data by analyzing, segmenting and using predictive techniques that help support the customer interaction process through quicker and more informed decisions (Schwenk, 2002).

With a wealth of electronic information about what online customers buy and do, companies can use data warehouse tools to personalize the online experience in real time "to recommend related products, lead buyers to areas that reflect their interests, and continuously tailor the sales pitch to appeal to the individual customer" (Accenture, 2002).

Through data warehousing and data mining tools companies can retrieve details about their customers from "operational systems and external sources and predict likely customer behavior not on a segment basis but for each individual customer" (CRM a revolution in business practice, Channel Dynamics). CRM technologies are experiencing high growth rates especially that more organizations are implementing them. "The European market

for professional services that are looking into implementing these technologies is expected to grow from \$1 billion in 2003, achieving a compound annual growth rate of 44%" (CRM a revolution in business practice, Channel Dynamics).

There are two forms of CRM offered by vendors, the analytic such as E.piphany Inc., and the operational, which includes companies such as Siebel Systems Inc. and Clarify (Ward, 2002).

Operational application is similar to Enterprise Resource Planning (ERP), which handles business processes and data storage for significant collection of business units and business functions. ERP is growing and is being offered by vendors such as SAP who are expanding their product range.

ERP will continue to grow and be implemented especially by vendors such as SAP who are expanding their product range to include industry-specific solutions. To achieve the predicted growth, ERP will not only be offered in the back office (supply chain) according to Channel Dynamics.

"SAP, PeopleSoft, and Oracle, are the better-known application vendors. The products that generated higher revenues for them recently were CRM, SCM, and other

so-called point solutions" (CRM a revolution in business practice, Channel Dynamics). However, these products do not offer a standard function for data warehousing applications, although they began to integrate the applications to reach a higher market and achieve higher returns.

Companies with certain specializations such as Siebel in CRM, or Agile Software Inc. in product life-cycle management, or I2 Technologies Inc. in SCM, are constantly trying to come up with a software that will satisfy all the aspects of organizations. This is accomplished by developing a standard system for the various organization applications within the various departments. "Companies such as Applix Inc., Cognos Inc., and Business Objects SA are beginning to change their products to encompass the various applications and are trying to recast themselves with a CRM or SCM focus" (CRM a revolution in business practice, Channel Dynamics). The applications come with APIs or simple interfaces for the enterprise applications, such as SAP or Oracle. Companies offering these applications will only survive in this competitive environment if they tune into what their customer wants and integrate their applications.

Market Segmentation

When segmenting the market for CRM purposes, organizations should look into other ways for segmentation in lieu of the traditional segmentation, which is done, based on age, income, sex or education. Companies should look at other issues to differentiate their customers, and create customer focus. It is important to differentiate customers and at the same it is important to find a commonality to position customers in their market segment, for example, wine collectors are usually interested in gourmet food; thus magazines related to food should target these individuals from surveys conducted and from mailing lists obtained from sources such as The Food Channel, or wine magazines. Companies' say that targeted marketing, and customer segmentation have boosted returns.

Graybar Electric Company

Organizations are focusing their effort on CRM, and adopting applications and softwares that will help them survive with the changes in the market environment. For example St. Louis-based Graybar Electric Co. Inc., a \$4.8 billion electrical and data communications equipment distributor, is investing tens of millions of dollars to implement the mySAP.com e-business platform, which will include customer relations' management and supply chain

management components (Callahan, 2002). The company made this investment because customers asked for a more streamlined shipping and billing process.

Beatty D'Alessandro, is the vice president of IT strategy project at Graybar. St. supported the implementation of this project since he felt that by implementing the software, the organization would seize opportunity for growth. This will be achieved by streamlining the process of delivering the right product into a customer's hands at the right time. In addition D'Alessandro said this would allow Graybar's in-house staff to focus on value-added services for customers. For salespeople, the software will help automate customer relations' management and sell more products. The system will unify the data in the organization since the SAP software will help ensure that all parts of the organization from suppliers to local Graybar distributors to customers-will have access to the data they need.

Cypress Semiconductor Corporation

Cypress Semiconductor Corp based in San Jose, California. has made satisfaction surveys the cornerstone of their customer segmentation efforts and their entire customer relations management (CRM) program (Anthes, 2002). The company believes that what distinguishes one

customer from another is each customer's feelings toward the company. Cypress' real-time customer satisfaction monitoring system from Satmetrix Systems Inc. triggers the e-mails to management and will soon produce the data that determines the bonuses awarded to employees, based on responsiveness to customers.

Vendors

Vendors are providing organizations with strategic solutions to building data warehouse that enhance decision making. "Vendors are offering insurance-specific, pre-architect data models and automated data dictionaries" (Voelker, 2001). They (the vendors) understand that the key to a successful data warehouse is the architecture and not the technology, thus they are providing a whole solution. These packages differ from systems that are customized to meet the needs of a specific installation. Vendors are also developing Key Performance Indicators (KPI) that can be immediately accessed by users through role-based portals. Typical KPI's that organizations track are cash on hand, sales movement, orders satisfied.

Many top vendors that provide Customer Relations Management (CRM) and Enterprise Resource Planning (ERP), for example, offer suites comprised of their own

applications plus others they have acquired. "The products in the suites, were not built together and may not pass data back and forth smoothly" (Nash, 2002).

Oracle provides integrated analytics and data warehousing as part of its Oracle9i RDBMS. It "provides a mechanism for storing multiple dimensions and summary calculations on a table, for analytical functions" (Ward, 2002). It is powerful and certainly cost-effective because there is no additional purchase required beyond the database itself.

In an article by Sharon Ward entitled "Smart Move" (2002) she describes the various vendors and their applications. Siebel offers Siebel Analytic Applications as part of its CRM suite. In addition, it has a pre-built data warehouse and relevant applications that address elements within functional areas such as sales, marketing, service, and partner relationship management. Siebel allows users to access information stored in multiple applications to achieve timely, complete, and relevant answers.

PeopleSoft provides a simple linking ability to allow users easy access to information while making decisions. It uses a powerful analysis tool

completely integrated with business applications that may affect decisions, however this application is not stored in its own database (Ward, 2002).

SAP Portals present information from multiple sources, and gives users access to a role-based portal. In addition, it uses iViews, which integrate maps to common applications; and can also easily draw on multiple data sources for a single query.

SAS "provides business intelligence software and services that create true enterprise intelligence", (Knowledge Portal to be Acquired By SAS, 2002) by providing relationships with customers and suppliers to allow for more accurate and informed decisions, through streamline integration of the information.

Business Intelligence Companies and vendors are trying to come up with standard tools for organizations to assist in having better communication within the departments, streamline operations, as well as support complete decision making applications. The various vendors offer a variety of tools that are constantly being enhanced and updated to meet the customer needs. However,

the largest market for growth these days is the development of data warehouse total solution for businesses. Thus, data warehousing has become an "architectural world where the corporate information factory is at the center of the table and solutions, not technologies, are the driving force" (Inmon, 2001e).

For example, Aspect Medical is a prominent developer of brain-monitoring equipment. Aspect Medical was working with a variety of applications: Siebel Systems for CRM, Remedy for its call center and QAD for ERP. The company had a problem since these systems did not operate together, and there was no consistency in reporting methods within the various departments. Aspect Medical's success was in its ability to predict when hospitals will need to order more of its disposable sensors.

Aspect hired Creative Computing Inc., or CCI, which has expertise in business intelligence and data warehousing, to come up with a way to allow the various departments to communicate together for business purposes such as forecasting sales, and analysis of data. Thus the primary task was to provide 'Aspect Medical with one view of all its data, as well as provide tools for building accurate sales reports" (Darrow, 2002).

CCI consultants wrote a program that e-mails daily sales update to Aspect Medical executives' Blackberry devices, for daily, weekly, and yearly operations. CCI also used Cognos' DecisionStream-Extraction, Transformation and Load (ETL) software-that helped the solution provider unite data from disparate sources. The example was used to show that there is not a standard tool for companies to use.

CHAPTER FIVE

COST-BENEFIT

Evaluation of Cost-Benefit

When implementing a data warehouse it is important to evaluate the cost effectiveness of the data warehouse to the company, and the Return On Investment (ROI) from implementing the system. Management has to consider the cost-benefit of implementing the data warehouse, and evaluate whether it is worthwhile for the long run to implement it and consider other options. Once the project is initiated management has to assess the return on investment from having the data warehouse. For an effective evaluation of the cost-benefit and value the organization has to assess the hours and money saved. An organization has to look at cost in terms of out of pocket expense and compare it to the benefit in terms of percentage of time converted to hours saved if the system is implemented, and evaluate the value, which is the percentage of time converted to a monetary value. Since reducing time will increase productivity of the individual and hence save the organization money.

The cost of a data warehouses varies according to the company's requirements. It has variable determinants such

as the volume of data, the data warehouse architecture, the duration of the implementation, the number of users in various locations. There is cost justification in the implementation of the data warehouse. When evaluating the cost effectiveness of data warehouse, one has to look at the overall cost-benefit of the data warehouse and not just the reduction in the "back office" costs. If the data warehouse is effective, then organizations can maintain or increase their market share, and maximize profits while reducing expenses.

Costs

Within costs the following items are considered:

- Hardware, software, development personnel and consultant costs
- Operational costs like ongoing systems maintenance

Benefits

Benefits can be measured by looking at increased revenues and reduced costs. The increase in revenue can be evaluated by comparing the number of new customers with the old one and whether the old customers are coming back; the market condition and whether the company is ahead of its competitors; whether the company is keeping up with changes in the environment. In addition, it can also be

measured on the hours saved in retrieving material that can be quantified in monetary (Dollar) value. The decreased costs are evaluating the overall efficiency in the company after implementing the data warehouse.

Cost of a Data Warehouse

There are different approaches to the cost measurement and justification of the data warehouse. An effective data warehouse will reduce costs. More importantly an effective warehouse allows the organization to hold on to and increase its market share. By utilizing the data warehouse tools effectively to assess and analyze stored data, the organization can achieve a competitive advantage in the market. This leads to increasing revenue and generating higher profits in the long run. The data warehouse will enable management to identify opportunities in the market by analyzing existing data and tapping into new markets. Through the reduction in costs (reduction in time, resources), and evaluating the company's performance the organization can minimize expenses and maximize profitability. The data warehouse can identify existing weaknesses, for management to ratify.

The major validation to support the cost factors with data warehousing is that it reduces the cost of accessing

information to the end user. With no data warehousing it is costly and time consuming to acquire reports since the user has to locate the data needed for the reports, filter the information, convert data into the required format, and merge data from different sources to build the report. The process will take a lot of time, and utilize significant resources. With the aid of the data warehouse the user can generate multiple reports that can either be summarized or detailed in a fraction of time, which will save the company money and enable the user to make better decisions for the company, hence enhancing the effectiveness of operations. Management can monitor savings by looking at the before and after effects of implementing the data warehouse, and compare the expenses.

When building a data warehouse an organization has to consider whether it is more beneficial to have a data warehouse compared to a data mart. The concept of a data warehouse versus the data mart was covered in Chapter One. A data mart will only be beneficial on a small scale, for departmental use, or for organizations with a small database. The management has to justify the cost of building multiple data marts versus building a single data warehouse by weighting the advantages and disadvantages of building either one.

The normalization of the data in the data warehouse allows the data in the warehouse to serve all purposes. The data in the data mart, on the other hand, is summarized and is very denormalized. The data mart data serves the needs of a department, not the corporation, (Inmon, 2000b) and it contains limited information that can be used for a single department. Whereas the data warehouse has large volumes of detailed data. A data mart is part of the data warehouse, and organizations have to realize that they need to have many data marts to encompass the large amount of data on hand, and building numerous data marts is expensive. Thus for the long run it is more cost effective to have a data warehouse than multiple data marts.

An organization has to monitor the business with the implementation of the data warehouse. The organization can quantify whether the data warehouse generated revenue by comparing the before data warehouse to after the implementation. An example "using customer relations' management, which creates an integrated, historical foundation of customer information" (Inmon, 2000b).

The CRM makes it easier to measure the number of new customers that were added to the system through the data warehouse. The measurement can be also made via the

additional revenue generated by new customer accounts. In addition the data will enable the company to evaluate the response of the market to new product launching through sales to new and old customers. Management can make these analyses by comparing the new numbers to existing the number of customers, and thus make a better assessment on the new system's functionality.

A Data Warehousing Institute study of 1,600 companies who implemented data warehousing into their business, showed that from only 13% of the respondents track data warehousing ROI across the value chain; "37% said they plan to begin tracking ROI; and 27% said they are not tracking returns and have no plans to do so" (Lewis, 2001). Acxiom, Compaq, EMC, IBM, Microsoft, NCR, Oracle, Sagent Technology and Sybase sponsored the report. The institute also found 16 percent of respondents say their data warehousing implementation is better than expected; 42 percent said it met their expectations; and 41 percent said they are experiencing difficulties. However, companies had to modify their business processes, incentives, skills, and look at privacy issues; as well as develop their architecture to avoid collecting unnecessary data. It is vital that the implementation of the data warehouse is inline with organizations' policies, so that

is becomes inherent in the culture of the organization. These results agreed with another survey conducted by Globalinkage for data warehousing vendor Informatica, "the survey covered 134 unnamed Informatica customers" (Lewis, 2001).

If implemented successfully, both studies agreed that data warehousing increased overall productivity in the organization, and proved that there are many benefits of data warehousing such as reduction in costs and utilization of resources. However, neither study provided specific return on investment figures, and none of the companies in the two surveys used a huge variety of financial measures, making comparisons impractical.

Benefits of a Data Warehouse

The benefit of Data Warehousing can be summarized with a quote from Markku Pettersson, General Manager of Toyota, Finland. "The executives often need a single dealer's report on short notice. Collecting this information at the last minute overloaded the after sales services and sales departments. Nowadays, with the use of a data warehouse, the information is accessed with a single push of a button, and almost totally in real time."

There are many benefits from implementing the data warehouse:

Immediate Information Delivery

Having the information at the tip of ones hand allows the user to generate reports faster, analyze the information, and make better decisions. The reports are interactive, and thus assist the users in decision making process and allows the user to seize existing opportunities.

Data Integration

The data warehouse facilitates data integration within multiple departments in an organization, and across various businesses. The data integration can allow management to monitor data usage, and enhance CRM. In addition some of the information will be used for the future such as cross selling

Trend and Seasonality Analysis

A data warehouse contains historic data that can be used to perform market analysis on trends and customer reaction/ behavior over time. This will help management with marketing and sale decisions.

Freedom from Information Systems Department

The data warehouse is usually user friendly and does not require an expert in Information Management to

generate report. Thus, the ease of use encourages more people to access and use the data, and generate reports quickly.

Data Warehouse Tools

The data warehouse offers users tools to manipulate data to generate various scenarios, and reports that could be produced upon demand. Data warehouse has queries that allow the user to extract information as well as provide answers with a click of the mouse.

Limitations of a Data Warehouse

Data warehouse has certain limitations upon implementation such as:

Poor Data Quality

According to the Seattle-based "Data Warehousing Institute", poor quality data pertaining to customers costs US businesses around "\$611 billion a year in postage, printing and staff overhead" (Tremblay, 2002). The Data Warehousing Institute surveyed 647 U.S.-based information technology managers across a broad range of industries. Around 11% of respondents were from financial service industry, while 9.5% were in insurance companies.

The alarming figure reflected above comes from the fact that 50% of companies surveyed responded that they

had no plans to implement measures to improve data quality, where the definition of data quality is removing errors found in data. These errors were the result of transposing wrong data entries into the system, as well as not updating existing data to changes for example change in status, death, or address. The management of these organizations does not see the cost of having wrong data in their systems, and the effect this has on maintaining their customers. It is imperative that the data quality be enhanced at all times to avoid errors and cost of continuous cleaning. "Organizations can frustrate and alienate loyal customers by incorrectly addressing letters or failing to recognize [customers] when they call or visit a store or Web site" (Trembly, 2002). By mislabeling or misprinting envelopes organizations are losing customers' confidence in their ability to do business. According to Wayne Eckerson, director of education and research for the Data Warehousing Institute, the cost of "mislabeled, misprinted communications as "astounding"". He states that "even at an error rate of .001%, the claims data contains more than 754,000 errors per month and more than 9.04 million errors per year,"

In addition, there are different interpretations of data within different departments in an organization. The

misinterpretations can lead to discrepancies in the quality of data if it is not continuously monitored, controlled, and standardized. For example different departments have a different definition to what constitutes a sale, in the marketing department a sale can be the time that a customer agrees to the purchase whereas in accounting a sale could be the time that the money was collected for the sale of the product.

Eckerson suggests that organizations should be committed to keeping their data clean, and updating the information regularly to avoid these mistakes. By not doing so organizations will lose their loyal customers, suppliers. Organizations have to realize the importance of data as a resource that generates business, and use it as a strategic tool for better decision making purposes.

To maintain data quality, management should liaise with IT departments on the most effective way to maintain data integrity. Whenever possible the organization should hire, train or outsource experienced professionals to ensure that the data is correct, valid, and updated. Management could insist on having continuous auditing to provide continuous assurance about the quality and credibility of the information presented.

Lack of Accessibility

Information remains underused if only a limited number of users can access the information, thus the cost of installment is wasted.

Integrate Data into the Mainframe

This is a challenge to the administrator who has to integrate data, at the same time maintaining customer and organizations goals. The data entered has to be valid and accurate.

Cost

Costs include limited value of the data, irrelevant data collected, data collection and retrieval costs, and costs of human capital. These costs are all drawbacks to implementing the data warehouse

The cost of "extracting, cleaning and integrating data represents 60-80% of the total cost of a typical data warehousing project, or indeed any other decision support project" (Manning, 2000).

The data warehouse is useful in identifying where data problems exist; however corrections to problems must be made within the system by the administrator to capture "dirty data" or to ratify problems with the architectural design.

CHAPTER SIX

SUMMARY

Information is a decisive key to success, and it is significant in differentiating businesses these days. It is a fact that by effectively utilizing knowledge and information companies will be able to achieve and maintain a competitive advantage in the market.

What makes data warehousing important is its functionality of storing large quantities of data, and allowing the user to access the data on a timely basis. This enables users to manipulate the information and form educated analysis and forecasting which will enhance the organizations' performance. Companies are changing their focus and making customer satisfaction a center feature in the success of their business.

Data warehousing provides organizations with strategic tools to achieve the competitive advantage that organizations are constantly seeking. The use of tools such as data mining, indexing and summaries enables management to retrieve information and perform thorough analysis, planning and forecasting to meet with the changes in the market environment. In addition the data warehouse is providing security measures that if properly

implemented and planned are helping organizations ensure that their data quality and validity are intact.

Prior to implementing a data warehouse the management has to understand that there are systems that are similar and appear like a data warehouses but are not. Some of these systems are:

Data Mart the comparison to the data warehouse was made in Chapter One. The data warehouse is a collection of data marts

An Operational Data Store (ODS) although it appears like one; it is a "staging phase for data to be moved into the warehouse" (Inmon, 2001a). The ODS is a subset of data that is kept in order to satisfy the response time needs of the organization.

The Internet is claimed to be the largest data warehouse, however it is not since the in the data on the Internet "is not integrated, nor is it high quality cleansed data" (Inmon, 2001a).

CRM has become an important factor in organizations which are now focusing on enhancing their customer service as well as providing customers with products they need. Customer centricity is the core of business success thus having readily available information on customer

preferences and purchasing habits is important for organizations.

There are advantages and disadvantages for implementing a data warehouse in an organization. The advantages are listed on the web page of the Data Warehouse Information Center on the URL <http://www.dwinfocenter.org/casefor.html> and the same center lists the disadvantages, which can be accessed using <http://www.dwinfocenter.org/against.html>

Advantages of Data Warehousing

- A competitive advantage is obtained by allowing decision makers access to data that can reveal previously "unavailable unknown and untapped information on for example customer trends, and demands" (Connoly & Begg, 2002, p. 1048). Thus by transforming data into meaningful information, a data warehouse allows business managers to perform more substantive, accurate and consistent analysis
- The data warehouse tools perform tasks associated with querying and reporting. These transactions can be performed on a regular basis.

- Utilizing data models technologies such as a star schema to speed up querying and reporting.
- Small amount of knowledge is required to write and maintain queries and reports.
- The data warehouse provides an opportunity to clean up data without changing the transaction processing.
- Old data is cleansed from the transaction processing system which speeds up the response time.
- The data warehouse has a security aspect that prevents users who only need to query from having access to the transaction processing system.

Disadvantages of Data Warehousing:

- There is little value in some of the data stored, thus for small companies having the information in a data warehouse is a drain on resources.
- Data warehousing systems can complicate business processes significantly.
- If the firm has no interest in ad hoc query or reporting then data warehousing may not be useful, this is especially true for smaller

companies who will be affected greatly by the cost.

- The time aspect, where it may take too long to teach/ train users to use data warehousing, and some organizations may not have time or money. There is a limited number of people who work with the data warehousing system project from beginning to end of the data warehousing life cycle.
- Designing, building and maintaining a data warehouse can be costly. It is costly to capture data, clean it up and deliver it in the required format. The cost may accumulate with time, and even more so with changes in requirements

As mentioned above data warehousing has several advantages and disadvantages to organizations. The major element that has to be considered prior to implementing the data warehouse is a full-cost analysis of implementing, maintaining and adding on to the data warehouse. In addition organizations have to consider that without a data warehouse each request of information requires gathering of data, integration of data, merging and summarization of data before fulfilling the request. Thus one has to consider the cost of man-hours and

resources that need to be used without having the data warehouse. Data-warehouse benefits will not show up overnight, this is the reason behind the termination of the project (before completion) since managers are used to seeing immediate results.

Although, some companies are implying that data warehouse implementation has been a failure to their business, the reality is that one has to look closely prior to agreeing to these implications. In many cases the so-called-failure is attributed to the fact that the system was not designed to meet the needs of the organizations during implementation. This is attributed to lack of planning and understanding the system or organizations' requirements fully. Thus, organizations have to realize that the design stage of data warehousing is vital determinant in the implementation of the data warehouse.

In addition, maintenance and continuous monitoring of the system is crucial for the success of the data warehouse. Maintenance includes maintaining data integrity and validity, in addition to updating the hardware and software to function effectively. The data in the warehouse has to be updated, to meet with the different

requirements of the departments, where administrators have to reduce issues such as data redundancy and dormant data.

Data Warehousing is not for Everyone

Patrick Kassebaum in his article Do insurers really need a data warehouse? (2001). Argues that data warehousing is not for everyone and organizations should look at substitutes that will help them with the data. "Companies need to start exploring alternatives to data warehousing that can enable them to streamline operations, dramatically cut costs, and use their customer data to create competitive advantage" (Kassebaum, 2001).

The author in the above article argues that insurance companies are facing competitive pressure from other insurance companies, services provided on the Internet, financial services deregulation, globalization, and other environmental changes, and thus they can no longer afford expensive, static data storage systems that create operational inefficiencies and hinders the company from achieving its strategic objectives. Insurance companies thus, will not be able to react rapidly and effectively to the continuous changes in the market.

Kassebaum believes that data warehouses are static, since when changing the structure or any aspect of the data warehouse it requires changing and updating the whole

system, and thus is the foremost limitation of the system. Whereby when new data is added or modified into the warehouse, various departments within the organization should be able to work with the updated information, without having to wait until the entire "warehouse is remodeled" (Kassebaum, 2001). When updating the data in the warehouse specialized personnel have to be hired to implement the changes required at an associated cost. The author thus believes that the data in the warehouse has to be given a common structure (format) in order to be combined to create master records. He believes that data warehouses are inefficient since they require that data be transported to a single location, where it is modified, and transformed into a unified homogenous form prior to be used by various departments in the organization. Kassebaum believes that when data is stored it loses its value and richness, and becomes historic information that would be rarely accessed and used, and thus invaluable to the user.

Furthermore various departments such as marketing, customer service, and accounting need different information and by standardizing the information organizations will not be able to differentiate the customers and their needs if all the information is presented in the same format. This will lead the insurance

companies to lose their competitive advantage since they can not specialize the information in various departments for their customers.

The alternative solution to data warehouse that Kassebaum is suggesting is called "unified data view" software. The new system allows the insurance companies to:

- Obtain data from any source-without cleaning, standardizing, and display it on the screen according to what the various departments want.
- Permit users to work with data from unrelated sources
- Attach a new database or administration system to current data storage system without changing any old or new data-and to do so in a matter of hours.

According to the author the new system will be more cost effective to organizations, by decreasing maintenance cost charged by the DWA. In addition, the users can on a continuous basis update the system without accruing additional costs.

However, there are some discrepancies in the article that were covered in this project. The author should take into account that system implementation requires both time

and money to be applied effectively regardless of the system. The author states that data warehouse has limitations, however organizations have to realize that most systems have advantages and disadvantages, and organizations have to perform a complete cost and benefit analysis, and decide whether the benefits outweigh the costs prior to investing the money in the new system.

Conclusions

Information is key to reacting to changes in the environment, thus having accurate, valid reliable, and consistent information available will help make faster decisions. The development of organizations should reflect a long-term vision for organizations to gain a competitive advantage in the market. The project covered the advantages and disadvantages of data warehousing as a decision support tool.

Appropriate strategic planning can be based on information and analysis captured from historic data within the warehouse. Organizations have to utilize the best available information to make decisions now while considering their impact on the future. The data warehouse provides businesses with this tool. Therefore, management should grasp the opportunity of having information

instantly and using this tool to stay ahead of competition while maintaining an effective decision making process. For example an organization can look at an industry and decide whether there is growth or decline from observing signs that the data reflects. If the data shows stagnation within an industry, then the organization can reduce its spending and focus on new products.

Recommendations

Implementing data warehousing is pushing organizations to reconsider the benefits of reaching more customers, and reconsidering policies that concentrate on applying better CRM tools to achieve a competitive advantage. Thus, if companies have not already implemented data warehousing they are adopting data warehouses into their future plans for expansion, and forecasting which is reflected in their future budgets. Companies currently using data warehousing are modifying and updating their systems to stay ahead in the market

As organizations use data warehouses decision support tools for decision making, more frequently, the amount spent on data warehousing will increase significantly. Organizations have to grasp the opportunity and consider ways in which they can make a return on their investment

when implementing data such as using charging back as an option.

Charge Back

The concept is based on "requests, submitted and total rows of data returned" (Inmon, 2001f), the idea raises the awareness of the end-user as to the resources used, in addition to monitor the amount to be billed. In addition, organizations can use the information for better analysis and develop cost center for various departments. There are other factors that could also be used for measurement such as CPU time, line transmission time, response time, and so forth.

Organizations can develop data usage trackers that can be implemented at the end-user workstation. The tracker gathers statistics about query activities, as well as observes the access of information. In addition, there could be a monitor on the server as the query activity passes into and out of the server. Activities can be gathered and collected by department, in order to monitor department activities, and hence create a cost allocation to verify how much each department is using the data warehouse.

Time considerations could be used for the billing purposes, where a user can be charged more for using the

service at peak time, and will incur a lower cost on weekends. In addition, through the tracker the billing could be generated according to the amount of information accessed. Thresh-holding is another method for billing, where the user will accrue charges for one level of activity and another set of charges for another set of activity.

Organizations can promote the use of the data warehouse, by encouraging the end user to utilize the data warehouse where the first few queries are free; then as the user becomes more familiar and uses the data warehouse more often the organization can then charge them for the information. Organizations can develop ways to charge back; they can charge back using real money, a form of credit, or "funny money". Whatever the organization uses it has to take into account the effect this has on the user utilization of the data warehouse (Inmon, 2001f).

For an effective charge back organizations have to find ways that attract users to the information on hand. In addition, by having summarizations of data, users will be encouraged to access more information on a regular basis. This will be particularly useful, when other organizations follow suit and the way to differentiate one organization from its competitor is the way the

information is presented, the response time to queries, the level of summaries, and its ease of use, this will give the organizations a competitive advantage.

Ensuring Success

In order to have a successful data warehouse one has to look at various variables that are imperative to implementing a data warehouse. The project covered various issues that have to be taken into consideration when building a data warehouse, to ensure worthiness. In conclusion, the following points are key in ensuring success:

- It is important to know the needs of the decision makers within the organization, thus seeking the correct people for gathering information within the organization is necessary.
- Long term planning is vital to the data warehouse, as well as ensuring the warehouse will support various decisions and changes that will occur over time.
- Proper training and support will give the users confidence in using the data warehouse.

- Maintenance and update of information is a key success factor for the integrity of the data within the warehouse

The Future

Combining technology for the future is what will push organizations forward. Combining GIS technology and data warehousing will generate great remuneration with organizations since it will capture more up to date information that is not fully utilized. GIS is currently used more by government agencies and less by businesses. There should be a marriage between GIS and data warehousing, where GIS can help provide information to businesses on the demographics of certain areas, the topology and zoning information prior to businesses opening new stores. Thus this synergy will prove to be efficient if properly utilized.

The GIS service and the medical field can blend data warehouse information for the betterment of humans. Whereby, data stored in hospital's warehouses be accessed through the GIS for medical purposes to save money, and give the patient the proper medical attention needed. For example, if a patient with no medical records on them needs immediate medical attention, and the patients has a history of diabetes, high blood pressure. The patient's

physician needs to transmit the information to the closest hospital, which can be located through GIS. Having the information sent and shared with the other hospital can help the patient receive the attention required instantly.

In addition, properly combining both technologies GIS and data warehouse, whereby the patient carries a chip that can download the medical history from a remote data warehouse in which it is stored. The information is uploaded to the closest hospital and is transmitted through GIS technology (which itself locates the closest hospital). The patients' information is stored on a remote data warehouse, which is constantly updated with tests and exams that the patient had. Developing such a system will have positive results if the implementation is done correctly where hospitals can be more efficient in obtaining patient medical records (that could otherwise be lost), and the appropriate medical care can be given to patients. However, to implement this idea hospitals have to have proper and standard equipments on hand to retrieve information. This idea seems somewhat far-fetched however; it can be done, by combining the existing tools.

In summary data warehousing is not a new concept in the business world. However, the constant changing application of data warehousing is what is being innovated

constantly and thus it is important to implement data warehouse applications that will enable managers to make better decisions through strategic planning, and allowing organizations to achieve a competitive advantage.

REFERENCES

- A definition of data warehousing. (1995) Retrieved April 16, 2002, from <http://www.dwinfocenter.org/defined.html>
- Anthes, G. (2002). Picking winners & losers [Electronic version]. Computerworld, 36(8), page 34.
- Australian Geological Survey Organisation-Geoscience. (2002). Seismic safety increased as brio reports enables public access to geoscientific databases on the internet Australia. Retrieved April 17, 2002, from <http://www.agso.gov.au>
- Callahan, S. (2002). Distributors far from being dead [Electronic version]. B to B, 87(4), 16.
- Chordas, L. (2001). Building a better warehouse [Electronic version]. Best's Review, 101(11), 117-121
- CRM, a revolution in business practice- Channel dynamics. (2001). Retrieved April 20, 2002, from <http://www.chdynamics.com/files/Cap%20Gemini%20CRM.pdf>
- Connolly, T., & Begg, C. (2002). Database systems (3rd ed.). UK: Addison Wesley.
- Consulting - ESRI Professional Services. (2001). Retrieved April 17, 2002; from <http://www.esri.com/consulting/programming/progsdw.html>
- Craig, R., Vivona, J., & Bercovitch, D. (1999). Microsoft Data Warehousing. USA: Wiley, John & Sons.
- Darrow, B. (2002). Putting their heads together; Jericho; Apr 22, 2002. [Electronic version]. The Newsweekly for Builders of Technology Solutions, 992, 41-42.
- Gayle, R., Knowles, III, C., Saputo, N., Whitney, A., & Yorkos, A. (2001). Data warehousing report. Retrieved May 7, 2002, from http://www.geocities.com/datawarehouses/GayleKnowlesSaputoWhitneyYorkos_ResearchPaper.doc
- Inmon, W. (1999). Measuring capacity in the data warehouse. Retrieved April, 20, 2002 from http://www.billinmon.com/library/library_frame.html

- Inmon, W. (2000a). Data warehouse administration. Retrieved April 20, 2002, from http://www.billinmon.com/library/library_frame.html
- Inmon, W. (2000b). Data warehouse environment: Quantifying cost justification and return on Investment. Retrieved May 8, 2002, from <http://www.BillInmon.com>
- Inmon, W. (2001a). Information management: Charting the course: What a data warehouse is not. Retrieved April 16, 2002, from <http://www.dmReview.com/master.cfm?navID=55&EdID=3606>
- Inmon, W. (2001b). Closed-loop/open-loop architecture. Retrieved April, 17, 2002, from http://www.billinmon.com/library/library_frame.html
- Inmon, W. (2001c) Enhancing data warehouse performance. Retrieved April 20, 2002, from http://www.billinmon.com/library/library_frame.html
- Inmon, W. (2001d). Changing the five P's. Retrieved April 20, 2002, from http://www.billinmon.com/library/library_frame.html
- Inmon, W. (2001e). Information management: Charting the course: The data warehouse marketplace. Retrieved April 20, 2002, from http://www.billinmon.com/library/library_frame.html
- Inmon W. (2001f). Charge back in the data warehouse environment. Retrieved April 20, 2002, from <http://www.billinmon.com>
- Inmon, W. (2002a). The data warehouse and testing. Retrieved April 20, 2002, from http://www.dmreview.com/editorial/dmreview/print_action.cfm?EdID=4621
- Inmon, W. (2002b). Deep history what good is it? Retrieved April, 20,2002 from http://www.dmreview.com/editorial/dmreview/print_action.cfm?EdID=4811
- Inmon, W., & Terdeman B. (2001). Information management: Data warehousing and national security. Retrieved April 20, 2002, from http://www.dmreview.com/editorial/dmreview/print_action.cfm?EdID=4348

- Karakizis, S. (2002). How to win with your data warehouse: Advice from a data warehouse veteran. Accenture. Retrieved April 26, 2002, from http://www.accenture.com/xd/xd.asp?it=enWeb&xd=Services\se\se_warehouse.xml
- Kassebaum, P. (2001). Do insurers really need a data warehouse? [Electronic version]. National Underwriter, 105(20), 14-17
- Kendall, K., & Kendall J. (2002). Systems analysis and design (5th ed.). Newark, NJ: Prentice Hall,.
- Kimball, R. (2001). Spatial enabling your data warehouse. Retrieved April 17, 2002, from http://www.intelligententerprise.com/010101/webhouse1_1.shtml
- Kimball, R. (2002a). The anti-architect. Retrieved April 17, 2002, from http://www.intelligententerprise.com/020114/502warehouse1_1.shtml
- Kimball, R. (2002b). Realtime Partitions. Retrieved April 16, 2002, from http://www.intelligententerprise.com/ports/search_webhouse.shtml
- Kimball, R., Reeves, L., Ross, M., & Thornthwaite, W. (1998). The data warehouse lifecycle toolkit. USA: Wiley.
- Knowledge Portal to be Acquired By SAS. (2002). Business wire. New York. Business Editors/High Tech Writers.
- Laney, D. (1996). All roads lead to the data warehouse navigator on-line. International Data Warehousing Association. Retrieved May 29, 2002, from <http://www.idwa.org/spr96/roads.htm>
- Lewis, D. (2001). Data warehousing pays off. InternetWeek. Retrieved May 7, 2002, from <http://www.internetwk.com/story/INW20010313S0001>
- MacMillan, T. (2002). The next 25 years: Custody benefits Canada [Electronic version]. NewsWire, 26(2) 49-50.
- Manning, T. (2000). Data warehousing. Retrieved May 7, 2002, from <http://www.manning.demon.co.uk>

- McKeown, P. (2000). Information technology and the network economy. USA: Harcourt, Inc.
- Mearian, L. (2002). Sears triples its storage capacity [Electronic version]. Computerworld, 36(5), 1, 53
- MSDN. Data Warehousing and online analytical processing. Retrieved April 17, 2002 from http://msdn.microsoft.com/library/default.asp?URL=/library/psdk/sql/ar_dts_whs_7cbq.htm
- Nash, K. (2002). Merging data silos [Electronic version]. Computerworld, 36(16), 30-32.
- Nolan, P. (2001). Getting started and finishing well. Retrieved April, 19, 2002, from http://www.intelligententerprise.com/010507/webhouse1_1.shtml
- Pharmaceutical and Healthcare Data Warehousing: Managing implementation to create greater ROI. (2001). Retrieved May, 8, 2002, from <http://www.benchmarkingreports.com/r/r140.htm>
- PR Newswire. (2002). Pharmaceutical Data Warehousing: Leveraging Customer Knowledge to Drive Sales. New York.
- Rezaee, Z., Sharbatoghlie, A., Elam, R., & McMickle, P. (2002). Continuous auditing: Building automated auditing capability [Electronic version]. Auditing, 21(1), 147-163.
- Schroeck, M. (2001). Data warehousing: The past 10 years have been quite a ride. Retrieved April 17 2002, from <http://www.dmreview.com/master.cfm?NavID=55&EdID=3007>
- Schwenk, H. (2002). Real-time CRM analytics: The future of BI? [Electronic version]. Km World Camden, 11(2), 10-11.
- Transforming technologies. (2002). Accenture. Retrieved April 17, 2002, from http://www.accenture.com/xd/xd.asp?it=enWeb&xd=services/se/se_webhousing.xml

- Trembly, A. (2002). Poor data quality: A \$600 billion issue [Electronic version]. National Underwriter, Erlanger, 106(11), 48
- Voelker, M. (2001). Data Warehousing - What works? - Ten things you Oughta Be Doing technology decisions for insurance. Retrieved April 19, 2002, from http://www.technologydecisions.com/backissue/0701/7_9_01_17.asp
- Ward, S. (2002). Smart Move. Retrieved April 17, 2002, from http://www.intelligententerprise.com/020328/506feat1_1.shtml
- Whiting R. (2002). Business Intelligence: Feasible Or Feeble? InformationWeek. Retrieved April 19, 2002, from <http://www.informationweek.com/story/IWK20020314S0028>