DATA WAREHOUSING

September 24, 2001

ISM 6217

Prepared by:

Richard Gayle
Christie Knowles, III
Nick Saputo
Alan Whitney
Ali Yorkos

Contents

I. Executive Summary

Data warehousing, managing data from beyond the every day operational system structure, is a valuable, proven approach to providing business users at all levels of the organization with the information they need to make high-impact decisions. The key of data warehousing is that data is stored for business purposes or for analysis can be more efficiently used if stored outside the operational system.

There are some logical transformation procedures involved in moving data from an operational system to a data warehouse. Data in a data warehouse is de-normalized. The de-normalization of data reduces the need for joins in a SQL query. A data warehouse model incorporates a de-normalized structure mainly for performance and simplicity.

The type of architecture deemed most efficient in creating a data warehouse is considered the three-level architecture. The three-level architecture allows data stored in a data warehouse to be customized by using data mart technology that provides single subject data to a small group of people that need that specific data, thus providing customized decision support.

The dimension model must be created taking into account the business needs and detailed needs of the users. The star design, the best suited schema for a data warehouse technology, relates to business needs and supports simple queries. The star schema is a data modeling technique used to map multidimensional decision support data in a relational database that facilitates advanced data analysis requirements. The star schema has four components including facts, dimensions, attributes and attribute hierarchies.

Data warehouses were designed to solve several business and technical problems. Data warehouses shrink the length of time it takes between when business events occur and when executives become alert. Data warehouses also provide a complete picture by combining data from multiple sources. Data warehouses further contain years of data to support trend and seasonal analysis and give users tools for looking at the data differently. Lastly, data warehouses provide freedom from IS department resource limitations.

While there are many positives to adopting a data warehousing system, there are also drawbacks. These drawbacks include large costs from setting up the system and from maintenance, source data that is produced from the internal processes of the company and excess data creation. All of these drawbacks may not have a monetary value, but they certainly impact the bottom line of a firm in the long run.

According to industry analysts, the market for data warehousing products and services in government is exploding, fueled by a growing need by government and citizens to access and analyze data for a variety of purposes. The expansion is from government agencies, large amounts of data from legacy systems, e-commerce, growing business potential from decision support systems and business intelligence systems. Today data warehousing systems and business intelligence tools enable agencies to extract valuable information from their databases and deliver more useful services to citizens and other users.

II. Technology Description

Data warehousing is a valuable, proven approach to providing business users at all levels of an organization with the information they need to make high-impact decisions – only if the right technologies, architecture, and techniques are used (Simon, Alan). In essence, a data warehouse is managed data beyond the every day operational system structure. It consists of business related data obtained from various sources and then formatted to meet business decision-making. The data warehouse technology facilitates informed business decisions by providing a means for the collection, consolidation and organization of data for reports and analysis.

The key of data warehousing is that data stored for business purposes or for analysis can be more efficiently used if stored outside the operational system. In the early stages of the technology, systems archived data on magnetic tape, as it became old or inactive. Accessing 'old' or archived information involved a manual process of loading the tapes. However, factors such as relational database management systems, development of faster processors and greater storage capacities, as well as enhancements in end-user applications has made it possible for data warehousing to exist.

Logical Transformation of Operational Data

There are some logical transformation procedures involved in moving data from an operational system to a data warehouse. The model of the data warehouse model outlines the logical and physical structure of the data warehouse (Gupta, Vivek). The model needs to be independent of the relational data model that exists for the operational system. Data in the data warehouse is, functionally, de-normalized, and may have

redundant reference data. If data obtained from different systems, have varying information about the same product or process (according to the business), the data warehouse must consolidate all information provided for that product that is relevant to the business process. Figure 1 (Gupta, Vivek) shows this consolidation.

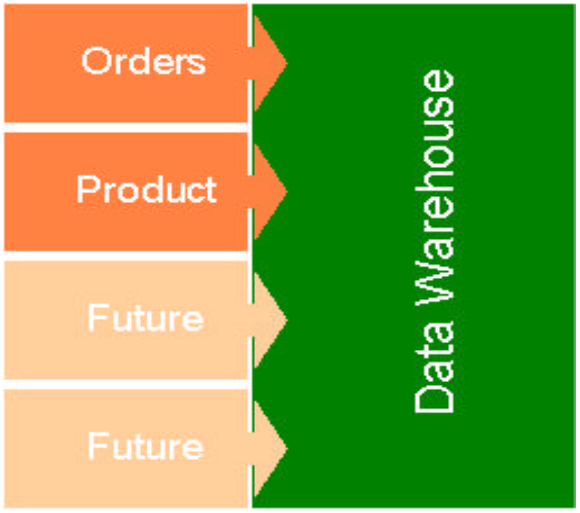


Figure 1

The de-normalization of data reduces the needs for joins in an SQL query. Normalization is used in relational modeling and involves breaking the database into various tables for flexibility purposes, however, this can make the data model very complex. A data warehouse model incorporates a de-normalized structure mainly for "performance and simplicity" (Gupta, Vivek); see Figure 2.

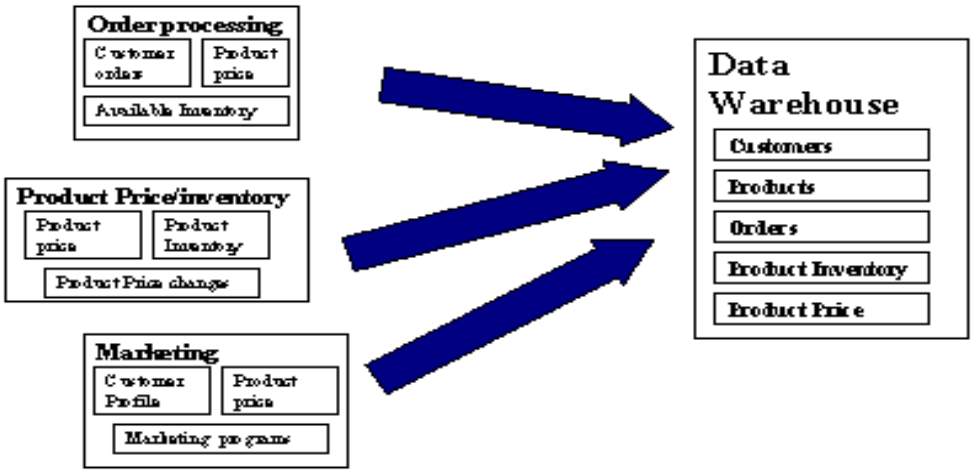


Figure 2

Data Architecture

The type of architecture deemed most efficient in creating a data warehouse is considered the three-level architecture. The conventional architecture, or two-level architecture, see Figure 3, shows all data being combined into one centralized unit, which can then be accessed by users. The three-level architecture allows data stored in a data warehouse to be customized using data mart technology that provides single-subject data to a small group of people that need that specific data. Decision support, in essence is customized for various groups, and only the specific data needed is given to a specific group – Figure 4.

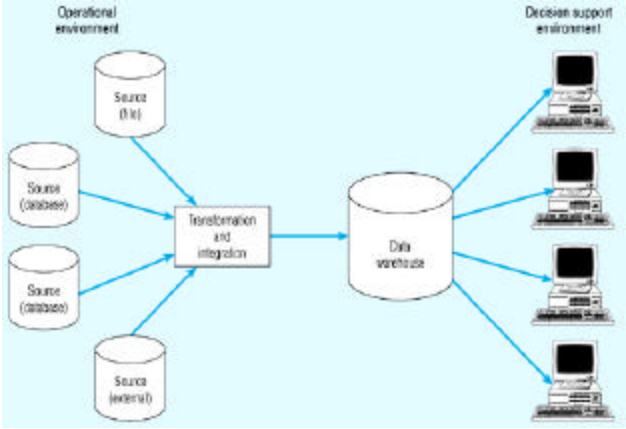


Figure 3

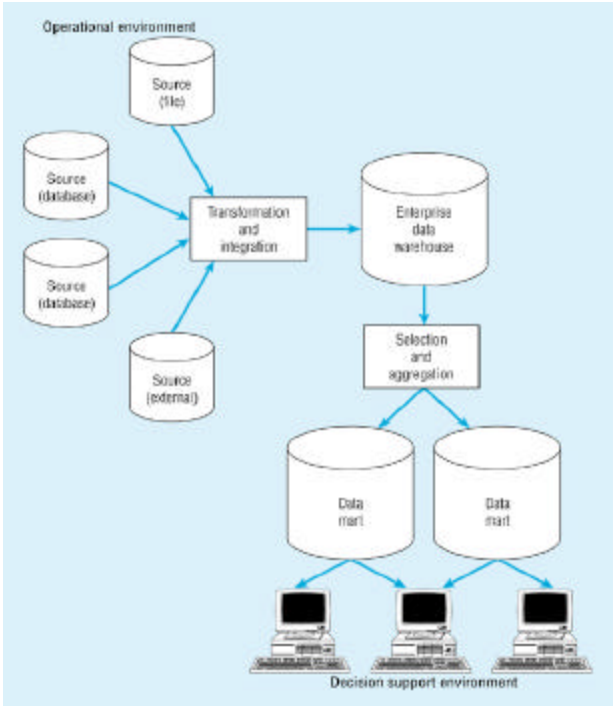


Figure 4

The Dimension Model – Star Schema

The dimension model must be created taking into account the business needs and detailed needs of the users. The model should provide ease of use and address the user requirements. The dimensional model uses a star design that relates to business needs and supports simple

queries. Other schemas exist such as the Snowflake schema, however the Star schema design seems to be well suited for a data warehouse technology.

The Star schema is a data modeling technique used to map multidimensional decision support data in a relational database. The Star schema was developed because the existing relational modeling techniques was not structured to facilitate advanced data analysis requirements. Star schemas allow an easily constructed model for multidimensional data analysis while still preserving the relational structures on which the operational database is built. The basic Star schema has four components including facts, dimensions, attributes, and attribute hierarchies. The reason it the technology is called a Star schema is because all of the dimension tables have a many to one relationship with the fact table. See Figure 5. The fact table contains facts that are linked through their dimensions. Some facts that are commonly used in business data analysis are units, costs, prices and revenues. Dimensions are qualifying characteristics that provide additional perspectives to a given fact. For example, a sale from a certain item may be compared by region, and by time period. Attributes are sometimes used to search, or for the classification of facts. Dimensions provide descriptions about facts by the attributes. The data warehouse must

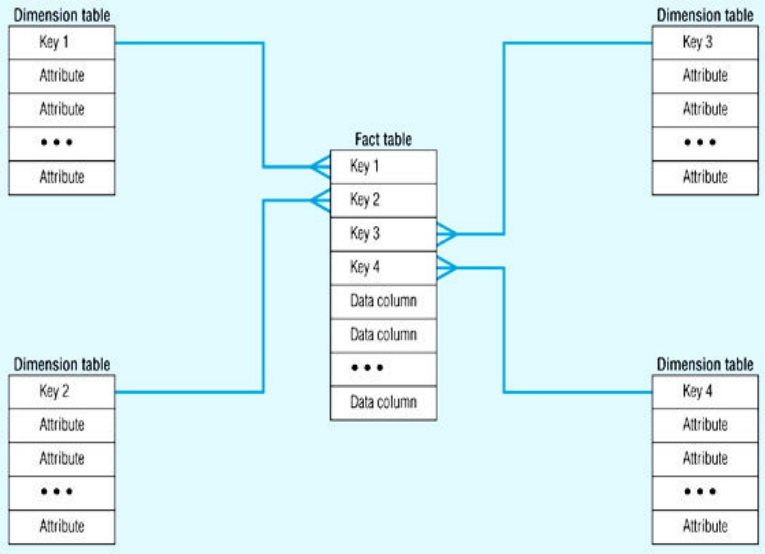


Figure 5

be designed to define common business attributes that will be used by in a narrow search by a data analyst. Finally, attribute hierarchies provide a top-down data organization that may be used for two main functions: aggregation and drill-down/roll-up data analysis as shown in Figure 6.

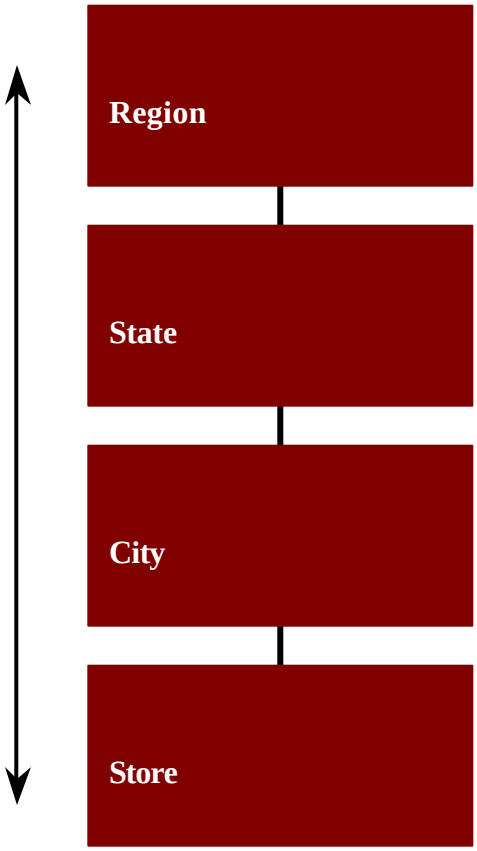


Figure 6

III. Business/Technical Problems the Technology is Designed to Solve

One important question that emerges in the business world is the question of "what are the business benefits of data warehousing." By highlighting some the more important and widespread uses of data warehouses, that answer can be attained. A data warehouse can answer many of the important questions facing today's dynamic businesses. Some of the more common answers and identifiable benefits are as follows…

Immediate information delivery:

Data warehouses shrink the length of time it takes between when business events occur and when executives become alert. For example, in many corporations sales reports are printed once a month, about a week after the end of each month. This means the June sales reports are delivered during the first week in July. Using a data warehouse, those same reports are available on a daily basis. Given this data delivery time compression, business decision makers can exploit opportunities that otherwise would have been overlooked. A good example of this can be observed through a statement from the Toyota Motor Corporation…

Toyota Motor Finland Organizes

"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."
Markku Pettersson, General Manager of Field Operations

Data integration from across, and even outside, the organization:

To provide a complete picture, data warehouses typically combine data from multiple sources such as a company's order entry and warranty systems. Thus, with a data warehouse, it may be possible to track all interactions a company has with each customer, from that customer's first inquiry, through the terms of their purchase all the way through any warranty or service interactions. This makes it possible for managers to have answers to questions like, "Is there a correlation between where a customer buys our product and the amount typically spent in supporting that customer?"

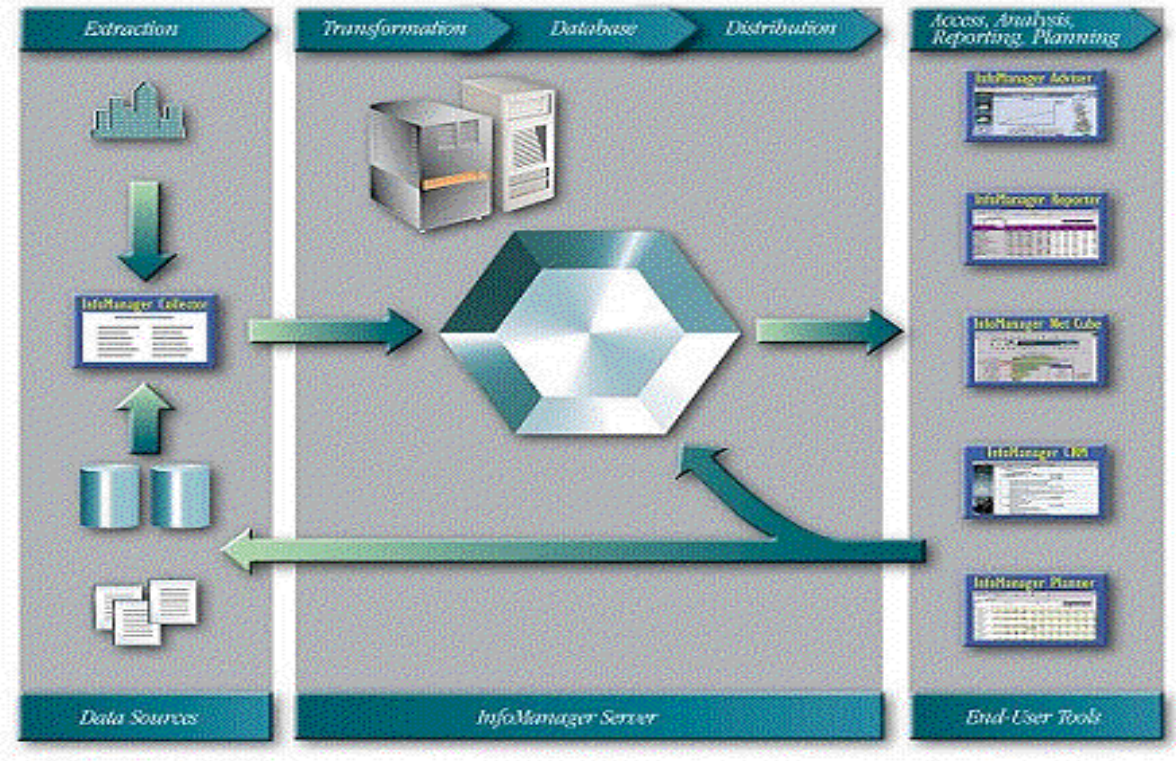


Figure 7

Future vision form historical trends:

Effective business analysis frequently includes trend and seasonality analysis. To support this, data warehouses typically contain years of data.

Tools for looking at the data in new ways:

Instead of paper reports, warehouses give users tools for looking at data differently. They also allow those users to manipulate their data. There are times when a color-coded map speaks volumes over a simple paper report. An interactive table that allows the user to drill down into detail data with the click of a mouse can answer questions that might take months to answer in a traditional system.

Freedom from IS department resources limitations:

One of the problems with computer systems is that they usually require computer experts to use them. When a report is needed, the requesting manager calls the IS department. IS then assigns a programmer to write a program to produce the report. The report can be created in a few days or, in extreme cases, in over a year. With a data warehouse, users create most of their reports themselves. Thus, if a manager needs a report for a meeting in half and hour, they, or their assistant, can create that report in a matter of minutes.

IV. Technology Limitations

While there are many positives to adopting a data warehousing system, there are also drawbacks. Even with these drawbacks, the negative aspect of data warehousing allows management to determine whether or not this system is right for their company. In the end, it allows for responsible decision-making.

The drawbacks can be quite daunting to a firm that has little patience for results. The costs of a data warehousing system are large. These 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 the system. While they may not directly manifest themselves in monetary terms, they will eventually impact a firm's bottom line if the company cannot effectively use a data warehousing system. These costs are the drawbacks of implementing a data warehouse system.

Data that is collected in a data warehouse is typically produced from the internal processes of the company using historical data. This information has limited value because the purpose of this information does not contribute to forward-looking decision-making. The information generated has such a limited scope that it may be overkill for a firm to invest in this system especially if the business process is relatively simple.

Creating excess data for the data warehouse is another drawback to adopting this system. This typically occurs when the data warehouse system has been established. Firms find that they can add unlimited data to a data warehouse. The problem arises due to the fact that data is added that doesn't have a clear relationship to its core business

strategy. By adding unrelated data to the data warehouse, the costs of retrieving data and transforming it into a useable form are increased.

Another data warehouse drawback is that the costs to collect and deliver data in a timely fashion may be too expensive for a firm. This is especially true when passing this information to an end user that needs to use the information. Cleaning the data is a vital part of this process. A data warehousing developer may not possess the necessary skills to effectively clean the data for the user. Compromises have to be made at times but this eventually leads to a compromising of the value of the data in the data warehouse.

Human capital is an expensive part of any database system. This is especially true with a data warehousing system. Many consultants and employees, who will use or maintain the system, don't have ample experience with the data warehousing life cycle and the associated processes. There aren't many people that have extensive experience with data warehousing because it is a relatively new data storage system. Those that do have experience, command a higher employment cost than others. These increased data warehousing human capital costs increase a firm's IT costs more so than those with regular database management and administration.

In conclusion, the costs associated with the data warehousing system are the major drawbacks to implementing this style of database. These drawbacks include limited value of the data, irrelevant data collected, data collection and retrieval costs, and costs of human capital. All of these drawbacks may not have a monetary value, but they certainly impact the bottom line of a firm in the long run.

V. Assessment of Technologies Prospects for Success in the Marketplace

According to industry analysts, the market for data warehousing products and services in government is exploding, fueled by a growing need by government and citizens to access and analyze data for a variety of purposes. Today data warehousing systems and business intelligence tools enable agencies to extract valuable information from their databases and deliver more useful services to citizens and other users.

The expanding market of data warehousing includes both federal agencies and state and local governments. 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, Va.-based IT Market Research Firm.

State and local government spending is rising even faster, growing from $550 million in 1999 to about $1.1 billion by 2004, according to Dataquest, a research arm of the Gartner Group in Stamford, Conn.

A recent study from Deloitte Research of New York backs up claims of the growing business potential of data warehousing and the related area of decision support. IT officials were asked to rank the most important technologies, data warehousing moved to No. 2 in 2000-01 from No. 5 in 1998-99, while decision support vaulted from eight to four.

The below charts in Figure 8 show the current demand for data warehousing:

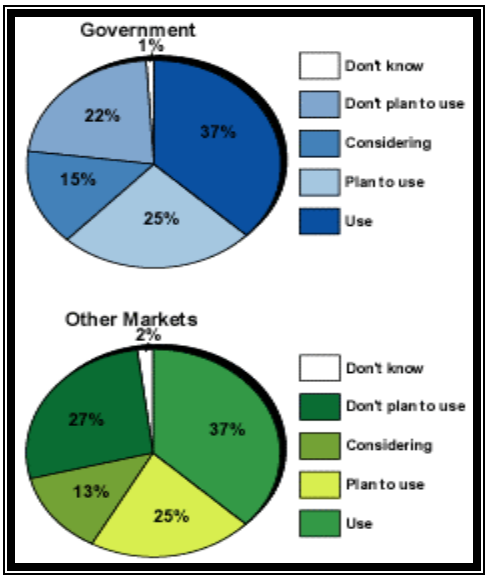


Figure 8

All the big names in the IT arena provide data warehousing solutions, including system integrators such as Andersen Consulting, IBM Corp. and Unisys Corp., major database vendors such as IBM, Informix Corp., Microsoft, Oracle Corp. and Sybase Inc., and other players like France's Bull and NCR Corp.

Successful results in government sector include:

- Decreasing costs and increasing income by exposing patterns of fraud and waste, such as in taxation, health care and welfare programs;
- Understanding the impact of programs and planning new programs, such as in health care, education and law enforcement;
- Reporting on how agencies use resources, such as personnel and finances.

Businesses will increase their annual spending by nearly 74% in the next three years on data warehousing and business intelligence (BI) solutions, according to newly released research from Survey.com, the leading eResearch services company. The study reveals that businesses of all sizes continue to make sizeable investments in data warehousing and BI projects.

The data warehousing and business intelligence market has experienced soaring growth in past years. The market is starting to show signs of maturity and its once spectacular growth rate has begun to moderate but continues to outperform many other technology markets.

Figure 9: Worldwide and North American Warehousing/BI market forecast

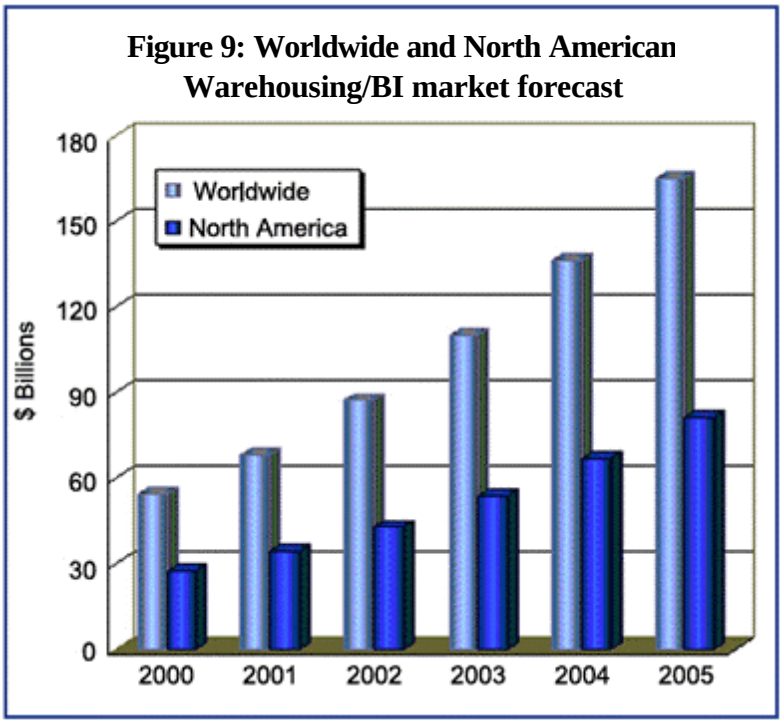
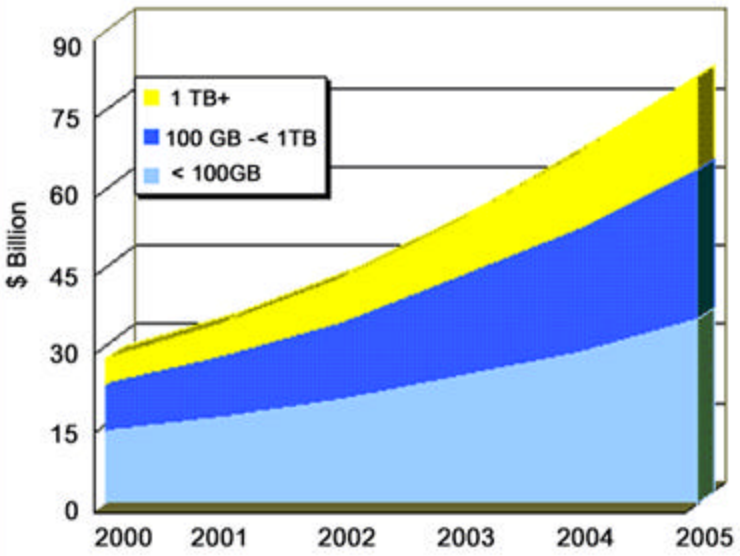


Figure 10: North American market forecast by Amount of useful data



The demand for data warehousing is emanating from several sources. First, years of work to remedy the Y2K problem finally have concluded. Second, many agencies have heaps of transaction data, often on legacy systems, with no easy means to access that data. Third, new government directives on reporting and oversight require accessing this data, if only in summary form. Fourth, government agencies are trying to satisfy their constituents by providing new services and access to information. And finally, using the World Wide Web for e-commerce or other transactions requires improved access to data. Data warehousing addresses all these issues, because it stages otherwise-inaccessible data in a structure that allows for simpler access and analysis.

Appendix A - References:

Websites:

The Data Warehouse Integration

http://www.dataspace.com/seminars_cvr.htm

Info Manager

http://www.dataspace.com/whatman.htm#whatexe

http://www.dwinfocenter.org/against.html

http://206.144.247.77/news/14_24/tech_features/1149-1.html

http://www.washingtontechnology.com/news/16_8/features/16846-1.html

http://www.ils.unc.edu/~mccuh/inls258/currentstate.htm

http://www.ebizchronicle.com/backgrounders01/jun/warehousing.htm

http://www.mdcomputing.com/ruffin.html

http://www.dwinfocenter.org

http://datawarehouse.ittoolbox.com

Print:

Designing a Data Warehouse: Supporting Customer Relationship Management
Author: Chris Todman, 1998, Second Edition.

Data Warehouse Project Management
Author: Sid Adelman, Larissa Terpeluk Moss

Data Warehouse Design Solutions
Author: Christopher Adamson, Michael Venerable. First Edition

The Essential Guide to Data Warehousing
Author: Lou Agosta, Prentice Hall, 2000

Data Warehousing and Data Mining, Implementing Strategic Knowledge Management
Author: Elliot King, Computer Technology Research Corp., First Edition, 2000

Database Processing
Author: David M. Kroenke, Prentice Hall, 2000

e-Data : turning data into information with data warehousing
Author: Jill Dychae, Addison-Wesley, 2000

Distributed data warehousing using Web technology : how to build a more cost-effective and flexible warehouse
Author: R.A. Moeller, AMACOM, 2001

Improving data warehouse and business information quality : methods for reducing costs and increasing profits
Author: Larry P. English, Wiley, 1999