

The Data Warehouse: Roles, Responsibilities, and Functions

Chris Toppe, Ph.D.
Computer Sciences Corporation

Abstract

A data warehouse is a very complex operation, one that doesn't fit the traditional system life cycle model. Whether a warehouse is 200 megabytes or 200 gigabytes, in building and operating it there are many roles, responsibilities, and functions that must be covered. Roles include builders, providers, maintainers, miners, and analysts, to name a few. Responsibilities include concepts such as data cleansing, data integrity, metadata creation, and data transportation. Functions include warehouse administration, warehouse load/refresh, and information extraction. These topics are covered in this paper with the goal of helping you understand the design issues around a warehouse project, and how software helps these roles, responsibilities, and functions to be performed.

A Data Warehouse Defined

There are many ways to describe a data warehouse (DW), each with a slightly different focus. Among my favorites are:

- A DW is a stable and predictable source for data used in decision making.
- A DW is a system — software, hardware, and communications -- that has been designed and optimized for cross-record analysis.
- A DW is a system which allows many users to access key data independently of the owners of the data, independently of the information technology staff, and independently of their skill as traditional computer programming.

Your Future in Data Warehousing

It has been predicted that the DW market will exceed \$10 billion by 1998, with the average system costing around \$3 million. This figure includes, in approximately equal proportions, hardware, software, and consulting services. It has been found that the typical DW shows a return on investment of 400%, meaning that those who have ventured into the DW area have been able

to achieve significant business impact through the DW's improved access to corporate data. With these kind of numbers, it is likely that a DW project may loom large in your future.

Organizations that enter into data warehousing typically have large systems in place that have been designed and built to optimize transaction processing. By default, therefore, these systems have not been optimized to permit the extraction and manipulation of decision-making data. A system that has been designed to rapidly process one transaction at a time is not suited to analysis where many (millions?) transactions must be accessed for analytical purposes.

Further, organizations that build a DW can usually be characterized as having a centralized IT department in which access to the data is carefully controlled by a select few people who know how to access and manipulate the complex data structures associated with large transactional systems. While a data request from an analyst may not be difficult, it usually takes weeks -- as opposed to hours -- for a request to be fulfilled. It is this delay in responsiveness that a DW is designed to alleviate.

Practical Uses of a DW

Banks typically use their data warehouses for credit card rollouts. That is, they analyze the buying, credit history, and demographics of the potential client base to determine which people will get what offer. The objective, often reached, is to capture 70% of the market penetration on 50% of the marketing expenditures. That is, if a bank were to spend \$2 million rolling out a credit card offer with the anticipation of realizing \$4 million in first year charges, a DW-based marketing plan would have the bank cutting its marketing effort in half, to \$1 million with the anticipation of realizing \$2.8 million (70% or \$4 million) in first year charges. The return on investment is much improved by

using the warehouse. The bank then has the remaining \$1 million in marketing funds to repeat the process, using the warehouse to identify a separate group of people to which it can market a separate product. SAS is the default software system used in this type of analysis, being used by all the major credit card issuers for this type of market analysis.

A newer use of the DW in this industry is the analysis of buying patterns for re-sale to a bank's commercial customers. For example, a bank with a DW can offer a restaurant a mailing list of all people who spent at least \$100 in a restaurant over the last six months, limited to a particular geographical region. Again, it is the power of the DW to analyze millions of transactional records that makes this possible.

For an insurance company, the main uses of a DW have been in fraud identification. With fraudulent claims being estimated A HIGH as 3% of all claims, even a small reduction on fraud can have a substantial impact on a company's expenditures. A further use of a DW in this industry is on premium modeling. If the company has as a target achieving a claims-to-premium ratio of .79, it may reach that target for the company as a whole yet be substantially over-charging some client groups and under-charging others. A DW allows the company to examine its client base among many dimensions, providing the ability for it to fine-tune its premiums to particular sub-groups of customers. I know of an insurance company which has divided its customers into 250 distinct groups and is therefore able to track claim-premium data at a never before possible level of detail. This is the message we need to get to the market place: the value of a DW to the business. In the industry as a whole, SAS is the leading software used for these purposes.

In the services industry, including utilities and retail companies, the uses of a DW are very different. For utilities, the focus of the DW effort has been to identify customers at risk for defaulting on their utility bills. By adding credit data from a source such as Equifax with the company's own payment history data, utility companies have been able to establish payment profiles which allow them to set deposit requirements.

In the retail market, the use of a DW is currently moving into two new areas: shelf space management and electronic commerce. Both of these business process improvements are dependent on a DW for success.

In shelf space improvement, the DW helps a retail business determine how to best use its limited shelf space: how much "face space" to give competing brands, how many rows of space to allocate to each product. Further, the DW enables the analysis of concomitant purchases: what items are likely to be accompanied by the purchase of dissimilar products. These stores have the transaction information; what they need is the ability to analyze that data for making intelligent business decisions.

Electronic commerce (EC) is another area which has been enabled by the DW. Electronic Commerce is defined as a retailer entering into a partnership relationship with a limited number of vendors for the tracking and restocking of inventory. That is, a company may decide that it will let a small number of suppliers provide its stock of office products world wide. These suppliers are given access to the DW so that they can track usage of the particular items it is responsible for stocking. When the on-hand inventory falls to the re-order point, the supplier simply re-stocks the item using an open purchase order from the company. The inventory is delivered and the invoice generated (and maybe even paid) without intervention from company staff. This results in a substantial reduction in the paperwork required to track, order, stock, and pay for items. This is a blending of technologies, with the DW playing a key but minor role.

Given the above descriptions, it can be seen why data warehousing has become such a hot topic in corporate America. Market advantage will can no longer be achieved by improving transactional systems. After all, does it really matter if it takes ten or fifteen seconds to process your credit card approval? Market advantage will come from the ability to make intelligent business decisions in a rapidly changing business environment. The data warehouse allows this to occur.

Who's Mining the Store

I was in a meeting with a retail industry executive recently who described what he'd like to be able to do with his data warehouse: "I'd like to know what products are being purchased by married women from households with incomes over \$100,000 in each of 14 departments across my more than 1,000 stores, and how those buying patterns are influenced by their attitudes toward my store." Simple enough. After all, we have access to the cash register sales system, the customer profile system, the credit history system, and the customer attitudes survey system. By joining these systems in a data warehouse, this retailer's query not only would be easy to answer, it would be just one of a nearly unlimited number of items he and his team could investigate.

So let's say that this data warehouse is built and looks somewhat like any other data warehouse. It has a nice, user-friendly GUI interface that contains two main portions: a section of pre-designed reports, charts, and graphs, and a section for *ad hoc* investigation. Further, as with other data warehouses, the DW toolset is distributed to dozens of analysts with different responsibilities across many divisions and departments located in different buildings, if not cities.

But there's the rub: Exactly who are these analysts? And are they qualified to manipulate the data warehouse? More important, are they qualified to make corporate decisions?

The warehouse, by design, makes it easy for analysts to *quickly* probe the corporation. Like physicians, they determine the internal health of the corporate body. They ask questions and the patient dutifully answers. They poke and prod sensitive spots, looking for unexpected tenderness or unusual

lumps. They listen to the pulse, check circulation, and watch the vital signs. But unlike physicians, these corporate caretakers may not be trained to perform such duties.

Analysts who "work the warehouse" tend to be people from either of two sides of a company. On the one hand, there are data people. They understand data, how it is derived, how it can be used, and what it means. They understand programming. They know how to structure queries and how to probe data without violating its integrity. But they do not know business rules. Those are the bailiwick of the second set of DW users, business analysts.

Business analysts know business processes and rules. They know what questions to ask and how to interpret results. They know what's important and what to ignore. They know where to focus, and they understand the critical measures of the evolving business environment. But they don't understand data.

Consider the case of the amateur statistician who has sprung up armed with the large arsenal of statistical tools such as those in the data warehouse. Suddenly, you have an untrained, unskilled person performing complex statistical routines on your data and, without skipping a beat, telling you what it means. You have a business analyst conducting statistical analyses of your key corporate data—using techniques like linear or logistic regression—who couldn't begin to tell you how the two techniques differ in terms of input, output, or interpretation. Not only does he have access to the data warehouse and its toolset, he decides what is important and what is not.

Real problems start when an unskilled analyst decides that something isn't important, and his decision isn't reviewed by others.

When that happens, the information on which he passes judgment is not seen by people with different skills, objectives, and perspectives. And the conclusion that something does not warrant further action or investigation is final. After all, he has excellent decision-making data. But has he made a good decision?

If a data warehouse is to be successful, you don't want amateur data manipulators making important data decisions and you don't want amateur business analysts making critical business decisions. Yet most data warehouse plans and toolsets encourage exactly that, with the naive hope that good decision-making data will make people good decision makers. As corporate data warehouses increase in popularity and more people have access to them, it seems that fewer and fewer are qualified to use it. This is the data warehouse decision dilemma. And unless you carefully plan who

will do what with your data, you could be in big trouble.

Roles, Responsibilities, and Functions

As you can see, a data warehouse is not an easy undertaking. For those of you who have built a career as data analysts, the warehouse offers many challenges. You need to know that data analysis occurs only after the warehouse has been designed and populated, not trivial tasks. Source data has to be located, cleaned, summarized, and transported to the warehouse. Further, these steps are on-going as the warehouse has to be refreshed on a periodic basis, usually nightly. The warehouse has to be administered, including access control, security, and backup. User interfaces need to be built, and there will be a constant need to build interfaces to both new data sources and new output packages. The remainder of this paper presents selected slides from the presentation on these topics. This allows me to cover the most information in the smallest amount of space,

Chris Toppe, Ph.D., is a senior management consultant with CSC Consulting and Systems Integration. He advises clients on data warehouse and decision support system design. He is also strategic alliance manager to the SAS Institute, managing the strategic relationship between CSC and the Institute. He can be reached at (301) 921 - 3441 or with e-mail: ctoppe@cscmail.csc.com



SESUG

Goal: Understand Users



- ❑ The Data Warehouse Serves Three Types of Users:
 - Power Users
 - Business Analysts
 - Executives
- ❑ As CSC, we understand the different needs of each group, allowing us to target activity to each group in a phased approach.
- ❑ Our experience spans the range of users, from IT staff to the board room. We understand how information looks from many different perspectives.

© 1994 CSC. All rights reserved.



SESUG

Power Users



- ❑ Skills
 - "Programmers" or "Code Bangers"
 - Heavy data knowledge
 - Know how to get data
 - Understand systems
- ❑ Needs
 - Direct access to the source data, meta data, and the DW
 - Heavy duty tool(s) for:
 - Capturing, cleaning, enhancing, and mapping source data, and moving, summarizing, and loading data into the warehouse
 - Querying (SQL)
 - Analyzing (statistical)
 - Reporting
 - Charting and graphing
- ❑ Interface
 - Primarily keyboard

© 1994 CSC. All rights reserved.



SESUG

Business Analysts



- Skills
 - Occasional Programmers
 - Heavy content knowledge
 - Know what data they want
 - Understand business drivers
- Needs
 - Access to the meta data and DW
 - Easy-to-use tools for:
 - Querying (SQL)
 - Analyzing (statistical)
 - Reporting
 - Charting and graphing
- Interface
 - Primarily point and click

© 1996 CSC. All rights reserved.



SESUG

Executives



- Skills
 - Strategic and tactical decision making
 - Internal "focusing"
 - External "positioning"
- Needs
 - Access to information
 - Pre-defined and pre-designed
 - Reports
 - Charts and graphs
 - Early warning system
- Interface
 - Totally point and click

© 1996 CSC. All rights reserved.



SESUG

Goal: Flexible User Interfaces



- CSC understands that data warehouse users are varied in their skills and serve different functions.
- We understand that individual people will play more than one role, and that the tool set can't get in the way of a person changing roles.
- We are prepared to provide a data warehouse that will meet the needs of widely skilled users (power users, analysts, and executives) serving different roles:
 - Builders
 - Providers
 - Maintainers
 - Miners
 - Explorers
 - Planners
 - Analysts
 - Readers (Executives)

© 1996 CSC. All rights reserved.



SESUG

Builders



- Builders
- Providers
- Maintainers
- Miners
- Explorers
- Planners
- Analysts
- Readers (Executives)

- Duties
 - Build and maintain the data warehouse
 - Capturing, cleaning, enhancing, and mapping source data, and moving, summarizing, and loading data into the warehouse
 - Maintain metadata
 - Build interfaces and custom applications for other users
- Orientation
 - Present
- Level
 - Power users
- Development Effort
 - Low

© 1996 CSC. All rights reserved.



SESUG

Providers



•Builders
 •Providers
 •Maintainers
 •Miners
 •Explorers
 •Planners
 •Analysts
 •Readers (Executives)

- ☐ Duties
 - Acquire and deliver data to other users
 - Prepares and executes queries against many DW sources (including cross-divisional data marts)
- ☐ Orientation
 - Present
- ☐ Level
 - Power users
- ☐ Development Effort
 - Low

© 1996 CSC. All rights reserved.



SESUG

Maintainers



•Builders
 •Providers
 •Maintainers
 •Miners
 •Explorers
 •Planners
 •Analysts
 •Readers (Executives)

- ☐ Duties
 - Warehouse systems management
 - Access authorization
 - System security
 - Archive, backup and recovery
 - System monitoring and tuning
- ☐ Orientation
 - Present
- ☐ Level
 - Power users
- ☐ Development Effort
 - Low

© 1996 CSC. All rights reserved.



SESUG

Miners



- Builders
- Providers
- Maintainers
- Miners
- Explorers
- Planners
- Analysts
- Readers (Executives)

□ Duties

- Looks for business "keys" buried in data
- Tries to maximize:
 - profits
 - sales
 - revenue
 - life time value
- Work with know customer facts: they did something
- Predicts the future
- Heavy statistical analyses over small amounts of data

□ Orientation

- Future

□ Level

- Power users and business users

□ Development Effort

- Low

© 1995 CSC. All rights reserved.



SESUG

Explorers



- Builders
- Providers
- Maintainers
- Miners
- Explorers
- Planners
- Analysts
- Readers (Executives)

□ Duties

- Looks for new opportunities buried in data
- Looks for new customer/product groupings
- Tries to create new customer and business facts:
 - They might do something
 - They are alike in some ways
- Predicts the future
- Heavy statistical analyses over large amount of data

□ Orientation

- Future

□ Level

- Power users and business users

□ Development Effort

- Medium

© 1995 CSC. All rights reserved.



SESUG

Planners



•Builders
•Providers
•Maintainers
•Miners
•Explorers
•Planners
•Analysts
•Readers (Executives)

- ☐ Duties
 - Usual and customary business projections
 - Expenses, revenues, profits
 - Inventory
 - Product cycles
 - Works from known business facts
 - Quantifies goals
- ☐ Orientation
 - Future
- ☐ Level
 - Business users
- ☐ Development Effort
 - Medium

© 1996 CCC. All rights reserved.



SESUG

Analysts



•Builders
•Providers
•Maintainers
•Miners
•Explorers
•Planners
•Analysts
•Readers (Executives)

- ☐ Duties
 - Usual and customary business reports
 - Expenses, revenues, profits
 - Inventory
 - Product cycles
 - Works from known business facts
 - Reports on the state of the business
- ☐ Orientation
 - Past
- ☐ Level
 - Business users
- ☐ Development Effort
 - Medium

© 1996 CCC. All rights reserved.