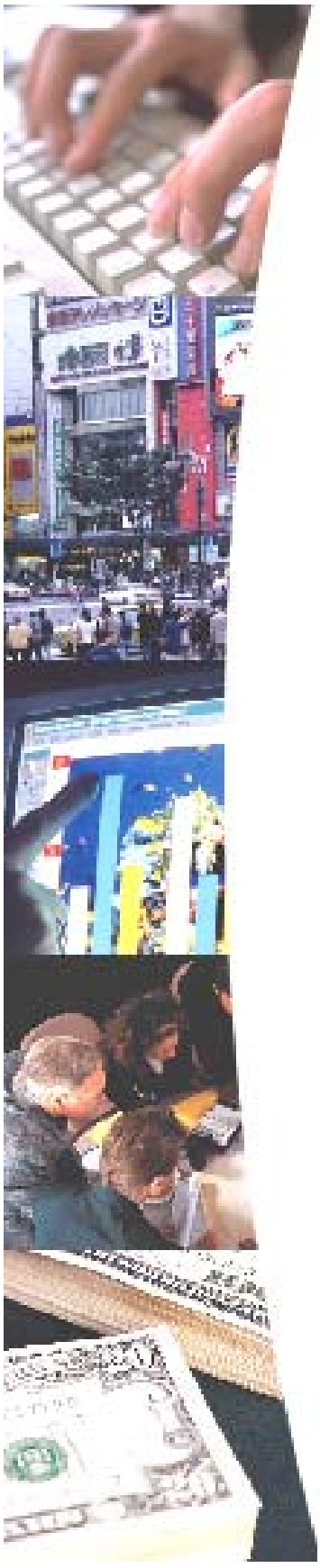


Current Practices in Active Data Warehousing

Richard Hackathorn
November 2002

BOLDER Technology, Inc.

1. Executive Summary

A series of interviews were conducted with people who have pioneered implementations of active data warehousing. This whitepaper summarizes these interviews and then synthesizes the common themes and lessons learned. We interviewed persons associated with seven companies (3M Corporation, Delta Technology, Ford Motor, Midwest Cards Services, PING Inc., Strategy.com, and Union Pacific), the details of which are in Section 3.

These interviews revealed the following common themes (in Section 4):

- Single Version of Truth, Business Value of Analytics, Business Value of Data Freshness, Actionable Granularity, Work Automation

…With the following factors contributing to their success (in Section 5):

- Securing executive support, Enabling business units, Adopting an open information policy, Simplifying architecture, and Stable vendor with scalable platform

…Facing the following business challenges:

- Balancing local diversity with global unity, and Discerning business opportunities

…Along with the following technology challenges:

- Maintaining terabyte warehouses, Managing unpredictable demands, and Moving from back-office to front-office

…Recommending the following best practices:

- Executive support from beginning, Use a normalized data model, Hands-on at project start-up, Educate users on the data, and Let alerting lead the way

…Realizing the following benefits:

- Improving asset management, Reducing customer support costs, Auditing billing practices, Terminating unprofitable product, "Being on top" of litigation cases, Reducing staff requirements, and Running the business

Companies are implementing active data warehousing successfully and reaping benefits.

The conclusion is that companies are implementing active data warehousing successfully and reaping benefits from these implementations. Active data warehousing is, however, in its early stages. The discernment of business opportunities appropriate to this approach is difficult but getting better as we learn from more successful examples. The barriers are often not technological but organizational because to realize the benefits of active data warehousing we may have to rethink the way we do business.

2. What is Active Data Warehousing?

At a Ford assembly plant in Los Angeles, a truck containing thousands of valuable parts waits to be unloaded. Another truck that just arrived is immediately unloaded. Why? For every truck throughout Ford's supply distribution system, the priorities of parts on those trucks are dynamically readjusted daily, depending on their importance to the current manufacturing process. The second truck contained parts critical to the assembly line today and needed to be unloaded first.

Situations like the above are occurring every day for Ford. Instead of following simple business rules (like first-in-first-out), these companies are analyzing the impacts of their decisions and choosing wisely. These are intelligent enterprises that are leveraging their business information to respond effectively to business demands.

Active data warehousing produces tangible impacts to the quality of business transactions.

At the heart of being an intelligent enterprise is Active Business Intelligence. It is using BI systems as an active (rather than passive) tool in performing the business of the corporation. Active data warehousing produces tangible impacts to the quality of day-to-day business transactions. Active data warehousing creates real differences in serving customers, delivering products, manufacturing goods, and securing supplies—across the entire value chain.

The shift from a passive role to an active role is the key differentiator with active data warehousing. From Webster's dictionary, the term 'active' implies 'to take action', rather than just thinking about or contemplating the situation. In contrast, traditional BI reports and analyzes the business dynamics within operational systems. The initiation of actions was a hope but not a design. For traditional BI, responsibility ends with the pixels on the screen.

We have used the framework of the 'intelligent enterprise' for understanding and explaining BI/DW trends, especially with active data warehousing. In particular, we emphasized the characteristics of: supporting tactical decision-making, leveraging actionable intelligence, and enabling the learning cycle. See a previous whitepaper for details.[1]

An intelligent enterprise is a company that pursues the goal of transforming insightful intelligence into effective action.

An intelligent enterprise is defined as a company that pursues the goal of transforming insightful intelligence into effective action. Although abused over the years, James Brian Quinn appropriately utilized the term as the central theme in his 1992 book.[2] He argued that companies should leverage their information resources through service offerings that enhance the usefulness of their product lifecycle.

Any enterprise (company, government agency, and the like) does not operate in isolation but in a matrix of relationships. Michael Porter[3] best expressed this matrix as a value chain—a flow of activities that add value to raw materials eventually terminating in the consumption of goods or fulfillment of services by the ultimate consumer. The focus is the incremental addition of value (or wealth) along the flow of the value chain.

The intelligent enterprise will exhibit certain characteristics to thrive within the value chain. In particular, the intelligent enterprise it will become more responsive to its business environment, pervasive in its business interactions, globalized in its business scope, and integrated across its business processes.

3. Case Studies

Each company described below has uniquely leveraged data warehousing technology in a variety of ways. The study interviewed seven persons to distill those insights. Although these persons were not randomly selected, the persons were knowledgeable about BI/DW technology with many years of experience, were familiar with the specific business situation, and were open and honest in sharing their experiences.

A 45-minute structured questionnaire was used and is listed in the Appendix. The emphasis was on their conception of and experiences with data warehousing, especially at pushing the boundaries on globalization, data freshness, data consolidation, and operational-level applications.

The companies are customers of Teradata, a division of NCR and the sponsor of this study. The companies were approached independently of Teradata to participate in an objective study of BI/DW trends. Although this is a bias sample and cannot generate statistical conclusions, the observations described accurately reflect the open and honest statements by experienced IT professionals who were closely involved with their data warehousing implementations and who are pioneering active data warehousing.

The following table lists the companies, interviewees, and themes observed.

Company	Interviewee	Themes
3M Corporation	Al Messerli formerly Director of Enterprise Info Mgt	▪ Enterprise-wide data consolidation with cross-functional leverage ▪ Single version of truth ▪ Globalization with local diversity and common views ▪ Atomic-level granularity to support actionable analyses ▪ Management of unstructured content ▪ Blending of ODS and data marts ▪ Data warehouse as data source for transactional systems ▪ Supporting external users
Delta Technology	Wayne Hyde formerly VP of Technology at Delta Air Lines	▪ Business need for data freshness ▪ Cost of real-time warehouse loading ▪ Operational analytics ▪ Information distribution using wireless devices ▪ Boundaries among staging areas, ODS, and warehouse ▪ Globalization and batch windows
Ford Motor Company	Jerry Hill President and CEO of Sagetree	▪ Supply chain management ▪ Data warehouse is a means to a business end ▪ Evolution of alerting tool to priority analysis ▪ Operational analytics ▪ Flowing data into versus out-of the warehouse ▪ Information speed and complexity for business requirements
Midwest Card Services	John Folkerts CTO and VP	▪ Process automation based on data warehouse ▪ Single version of reality ▪ Small IT support staff ▪ Stress business understanding of the data ▪ Tangible benefits with cost avoidance and staff reduction
PING Karsten	Kent Crossland Info. Services Director	▪ Operational applications ▪ Single database architecture supporting operational applications ▪ Small IT support staff
Strategy.com	Justin Langseth CTO and VP of Product Strategy	▪ Massive external user base ▪ Real-time alerting as a key feature ▪ Delivery via phone, pager, email, fax, and website ▪ Proactive information distribution
Union Pacific Railroad	Paul Evans Senior Mgr of Enterprise Data Warehouse	▪ Enterprise-wide data integration with cross-functional leverage ▪ Delivery consolidation for single information source ▪ Open information policy ▪ 3NF design that aids data integration

3.1 3M Corporation

Every industry in the world is impacted in some way by products from 3M, a $16 billion diversified manufacturer headquartered in St. Paul, Minnesota. With over one hundred years of history, the company has facilities in more than 60 countries and serves customers in nearly 200 countries, offering more than ½ million products, packaged in various sizes and in numerous languages.

3.1.1 Consolidation Initiative

In the 1990's, 3M was organized into 40 disparate business units with 60 international subsidiaries, all of whom tracked their customers, products, and sales without common standards. The business units had various independent data marts that supported little cross-unit sharing. Inconsistent summary-level statistics and inaccurate sales/customer data were the norm. Each 3M unit dealt with their customers and suppliers as if it were a separate company. Like many global corporations, top executives were unable to obtain a single view of the enterprise and, especially, a single view of their customers.

Beginning in 1996, 3M undertook an aggressive initiative to remedy this problem. With firm executive support, the goal was to construct a consistent global view of each customer, channel partner, and their buying habits. The corporation was realigned from 10 technology groups into six marketing groups. More importantly, all customer-related information was centralized into a global enterprise data warehouse (GEDW). The first version was put into production in 1997, so that employees, customers, distributors, and suppliers finally had access to one common, centralized source of information. Today, GEDW is operating a 32-node Teradata MPP platform housing 15 TB of storage.

GEDW was to be a major task for IT, considering the sheer volume of data, the hundreds of data sources, the multiple terabytes of storage, creation of thousands of reports, and information distribution worldwide to a massive user community.[4]

Drawing on the application-generic GEDW, new applications are spawning across many areas, such as demand planning, supply chain management, sales, pricing and profitability analysis, and market segmentation. The applications also span target customer information, order detail and service metrics, procurement, financial detail, and the digitalization of online product information for product commercialization, channel partners, and customers. Applications have increased the velocity of the primary business cycles, especially in order entry processes. Instead of daily loads, information is now being loaded into the GEDW near real-time, with batch loads every 2 hours. The GEDW is the information reference point for everyone, including many transaction systems.

3.1.2 The Benefits

3M sees opportunities to increase cross-selling, customer penetration, and proactive customer relationships. Customer behavior and sales activity can be understood and predicted, enabled by distributors sharing their sales data. Customers around the world can access extensive, detailed multimedia prod-

uct information. The loyalty and trust gained from customers and distributors will substantially increase corporate growth over the long term.

In the Realware Awards[5] in 1998, it was stated that 3M's consolidation of its disparate BI systems yielded a multimillion-dollar return on investment, by reducing IT infrastructure costs, eliminating redundant application development, increasing cross-sell opportunities, focusing marketing efforts, and enhancing the quality and timeliness of decisions. In addition, there were increases in productivity and customer/partner loyalty. On an investment of less than $50 million, 3M expected in 1998 to realize a net benefit in the first five years of operation to exceed $100 million.

In the Realware Awards[6] in 2002, it was stated that GEDW had reduced indirect cost by $350M.[7] Poised for expansion over the next few years, 3M executives have the goal that the GEDW will eventually reduce inventory costs by more than $1B over the project lifetime.

3M executives have the goal that the GEDW will eventually reduce inventory costs by more than $1B over the project lifetime.

Al Messerli, principal at Allen Messerli Enterprise Systems, LLC and former Director of Enterprise Information Management at 3M, was responsible for the justification, design, and implementation of the GEDW at 3M. Messerli recently stated[8], "3M is still the only and best practice [of global enterprise data warehousing] in the manufacturing industry. Not too many others have implemented a true GEDW. In the retail world, there are major retailers whose POS data is loaded into the warehouse by the time that the customer leaves the parking lot. I also think that some banks may have a GEDW, along with some firms in telecomm and transportation." Messerli continued:

> Most people are just thinking about doing enterprise data warehousing, which is targeted toward cross-functional data among multiple business units. The data from all the business units should be integrated so that units can cross-sell, leveraging their data about customers and sales across the enterprise.

Being global has special significance. All large companies must operate transaction systems in whatever countries they operate. Messerli explained:

> We just did it and made it work! 3M has independent transaction systems in 62 countries, all feeding into the GEDW. Each country needed access to its local content, such as customer name and address in its language and style. However, data was standardized for a single global view, designed to be used by people around the world. This task was very complex involving over 5000 tables that dealt with numerous factors, like currency conversions. You can do queries in the local currency, or you can do a global query in a common currency, all supported by automatic conversion routines designed by the finance people.

Messerli remarked that 3M "defines BI much broader, more than decision support, covering all of enterprise reporting, and more." The GEDW has become the place to go for any kind of information, such as customers, products, orders, or the like. "Although there are 100 order-processing systems at 3M, users can go to only one place for the information needed," remarked Messerli. "We have a single version of the truth."

The GEDW additionally acts as the operational data store and as multiple data marts. "We got rid of all our separate physical data marts and other data stores," stated Messerli. Data marts are now virtual, created with virtual views on the normalized relational data, avoiding multiple replications. "This approach was new and radical when we did it over six years ago."

In addition, many transaction systems go to the data warehouse for their source data. Messerli shared:

> 3M is using the GEDW to print product-shipping labels because it is available 24 hours per day and has the more accurate information. Also when the order entry processing was implemented for the website, the GEDW was tapped for reference information; however, the transactional programs do not execute on the Teradata box. The GEDW has become the self-service information access point for customers, suppliers, and channel partners. It goes beyond the corporate boundaries, supporting more than employees, even to the website. In addition, 3M has integrated all product literature, such as multi-media, catalogs, images, and documents. And all of that is in the data warehouse. This is general content management, all in the data warehouse.

When asked about the granularity of the data, Messerli discussed the need to have data at the atomic-level to enable people to take action based on that data:

The data warehouse is atomic level, which makes it actionable.

> The data warehouse is atomic level, which makes it actionable. We can drill down to get the credibility so that we can guarantee that contents match the source transaction systems. From that, you get quality, credibility, and actionability. [That is, you can take specific action based on the data.] That is why there is drill down to atomic detail. Most data warehouses have only aggregated or summary data. When they see a problem, they must delegate it to the department manager, who must go back to the source transaction system to see what the problem really is – where the transaction is out of line. Usually you cannot do the action [from data] in the data warehouse, but with the atomic detail there, you can close the gap in the data warehouse.

When asked about real-time warehousing, Messerli replied:

> 3M has gone from daily to 2-hour to sometimes real-time. The issue [with going real-time] is what is [happening in] the source system. For instance, the orders at 3M do have real-time, front-end processing; however, the actual allocation and release of the orders to the factories are done in batch every 2 hours. [So, real-time warehousing in this case does not make sense.]
>
> If the [transactional] system is truly real-time and you can extract the data in real-time, probably from the log files, then there is no problem to load the data into the data warehouse in real-time. Some of that is happening at 3M. But, I do not want to overstate the case there, because the actual [transactional] processing is often not done in real-time.

3.2 Delta Technology

Delta Technology, Inc. is the IT subsidiary of Delta Air Lines, the world's second largest airline, which offers over 5,000 flights daily to 410 destinations in 72 countries. Headquartered in Atlanta, Delta Technology employs 2,100 IT professionals whose mission is "to develop innovative technology solutions to improve the travel experience for Delta customers, enhance shareowner value, and increase employee satisfaction."[9]

Several years ago, Delta Air Lines started to replace its distributed and independent data marts with a centralized, enterprise-wide data warehouse.[10] It was difficult to perform cross-functional analyses, especially linking ticket revenue with customer data. No single group at Delta was responsible for administering all the data marts, resulting in the lack of common data standards. Delta concluded that a centralized data warehouse would be cheaper to administer and maintain because it involves fewer physical machines, interfaces, and administrators.

Complementing their enterprise data warehouse efforts, Delta recently announced the Delta Nervous System, a real-time digital network that shares

consistent and timely information with customers and employees. [11] It utilizes a variety of channels, including Gate Information Display Screens, the Flight Information Display Screens, various kiosks, and the company website.

In a recent interview, Wayne Hyde shared his experiences at Delta and with data warehousing generally.[12] He is principal at Reflection Technology and served previously as the Vice President of Technology at Delta Air Lines, where he was responsible for corporate-wide technology selection and implementation, including the construction of a global infrastructure for web, database, messaging, and data warehousing.

When asked about active data warehousing, Hyde focused on the business requirements for real-time information, which is only one aspect of active data warehousing:

> Understand it [active data warehousing] completely, and I am a huge fan. I have used the concept in my classes. The issue is whether the business demands it. If the business demands real-time information, then by definition, there is a need for active data warehousing for that business. There are some situations where active data warehousing is valuable, and quite a few more where active data warehousing is not.
>
> I spent time at a financial services company. Walking in the door, I expected that there was an opportunity for active data warehousing at this company. Until I talked with the business people, and there was no glimmer of hope for active data warehousing in their eyes. Their opinion was that yesterday's data is just fine; anything else would best be filled by an ODS off the transaction system.
>
> When I probe about the freshness of the data in the data warehouse, I answered the question by starting with yesterday's data and then discuss the possibility of more frequent updates. They [persons at the financial services company] were quick to point out that daily was enough freshness for their business. They did not have demand for information internally or even externally that pushed the envelope to real-time or near real-time. Their business model did not drive a ROI for anything more frequently than next day. Maybe their business model is obsolete and could stand some innovation to a more customer-centric perspective.

Do direct updates to the data warehouse mess up the enterprise architecture? Hyde replied:

> Enterprise architecture in general is not neat and clean in general. I often compromise EA with data marts on Teradata, if it is efficient. The box can be a little dirtier, as long it meets the requirements of the business. Tools, like the Teradata box, are usually not pretty if they are being used well. It may not be hard to manage. If it is good for the business, I do not care about the messiness.

When asked about tradeoffs of using the data warehouse for ODS processing, Hyde responded:

> That is not a bad idea because you can use it as a staging database. For instance, you could use messaging pipes to feed into staging tables, and then the ODS could batch update the data into the warehouse tables … all on the Teradata box. ODS can serve several functions. Positioning everything on the Teradata box has some appeal to it.

What are the implications of globalization a data warehouse?

> That is just a more complex environment, but it is not a huge concern. We were running a 20-hour window at Delta, and most railroads are running nonstop. What happens in global environments you have peaks during all operating

hours, instead not easy batch window during the night. Trickle feeds may be a more gentle way of loading the warehouse.

3.3 Ford Motor Company

Ford Motor pioneered the manufacturing revolution with its automobile assembly lines almost a hundred years ago. With sales over $160 million and employment over 350,000, Ford is the world's second largest automobile manufacturer behind General Motors, making brands such as Aston Martin, Ford, Jaguar, Lincoln, Mercury, and Volvo.

The focus of our interview was on the Ford Customer Service Division, which manages 2,000 suppliers, 5,900 dealers, 198,000 parts, $650 million in inventory, 8 distribution centers, 50 million vehicles, and 4,000 employees. We interviewed a key systems integrator who worked closely with Ford—Jerry Hill, President and CEO of Sagetree, whose vision is to revolutionize the global supply chain by transforming information into intelligence.[13]

The data warehouse is a means to an end.

Hill described the supporting role that the data warehouse plays with Supply Chain Management (SCM).

> I usually do not discuss the data warehouse with my colleagues. The data warehouse is a means to an end. To most business users, data warehouse is a foreign topic. The discussion starts with a business problem from a conception and ends with the output of the application.

His involvement with Ford was to develop an alerting tool but grew into a large and critical system.

> When we started with Ford, it was a simple alerting tool, such as a truck was late. As the relationship grew, the requirements were extended from alerting to analytical and then to a proactive operational capability. In other words, the system is measuring the net change in inventory on a nightly basis and then calculating, at the part number level, the days-on-hand inventory based on consumption and demand. Using the touch points in the supply chain, the flow of material is reprioritized for the next day. For example, imagine a truck on a five-day trip from Detroit to LA. The priority for the material on that truck was normal priority when it left Detroit; however, the demand spiked for a particular part, or the supply for that part dropped off. The system would sense that and recalculate the priority. When the truck arrived in LA, the work instructions would identify that truck as the Number One truck to unload, thus reducing the probability of a stock-out. They are simply readjusting the priorities for the work, rather than incurring additional cost (e.g., shipping via FedEx) or expending additional labor (e.g., overtime).
>
> This application is now called IMAS (Inventory Management and Alerting System) at Ford. They have become very dependent on IMAS, with 400 people who use it on a daily basis, [receiving] work instructions and analysis.

Hill continued by outlining the evolution of a DSS application into an operational (mission critical) one:

> [The IMAS system] was conceived as a decision support application, but it is actually used as an operational application. It is a unique situation because the application does not have the resources for being mission critical. Decision support has always been a second-tier application. This is no longer a second-tier application; it is an operational, 24x7, must-run (or the guys do not know what to do the next day) application with guaranteed uptime. The result set is analytics at the part-level number across all inventory touch points in the supply chain.

We need to think of it [operational application] differently [rather than reengineer it]. The Teradata system is very robust, with the best uptime stats. They went back and adjusted the priorities in the Teradata box so that IMAS have higher precedence in the job mix. It was competing with various queries. Now those queries are held until the IMAS processing is completed.

This situation did not create a problem. Very few decision support applications are used in such an operational manner. Normally, the users are analysts and management people in the office. Now, the users are union people on the docks. It is unique since few analytical applications push information that say 'unload this truck first'. By reconfiguring the Teradata system, the architecture did not need to be redesigned or reprogrammed.

Does active data warehousing mean anything to Ford?

Not at first. However, active data warehousing is now a building requirement.

You need immediate feedback while you have the part 'in your hand'.

Ford migrated from a 30-day refresh cycle of its inventory data to a one-day cycle, which was such a leap. There are aspects of the operation that they need to go real-time. I have not seen anyone [manager] who wants to analyze their business on a minute-to-minute basis. However on the operational side, the business does need a minute-to-minute guidance.

When material is being received, it is scanned. That transaction is entered into the system, and the system should immediately come back with what to do with it. …While you have it in your hand! We need to focus on those situations that generate a transaction and the people doing the transaction need an immediate response of what to do. Out of the gigabytes that we load each night, there will be a fraction that needs to be active and real-time.

Hill commented on the importance of delivering information along with its entire context.

…getting the data out of the data warehouse in real-time.

Most people focus on getting the data into the data warehouse in real-time. Instead, the focus should be on getting the data out of the data warehouse in real-time so that you can affect the quality of the transaction. We do need to feed the data warehouse with as current data as possible. But the trick is to get the information out of it in such a format that one can work with it. The person moving the pallet needs an immediate sense of what to do with it. The opportunity is there when he has it, not later when he has to go back. Likewise, the situation is similar for a teller at a bank. You have a customer in front of you now. In a few minutes, that customer is gone.

It all has to be detailed down to the part level number, and then rolled up to the bag, box, pallet, container, and finally the shipper. It does no good to tell the guy who is managing traffic to look for a part, when there are hundreds of trucks in the yard. There is a constant demand to keep the data at the level that the people are dealing with it. You are continuously dealing with bulk shipment of spark plugs, to individual washers in bags, to very large boxes (like fenders and chassis). The trick is to have sufficient details, but in a hierarchy that keeps it all straight for you.

The one thing that we learned was that the workers will revert to first-in-first-out if you do not give them sufficient information to get their job done. So, the reports must be tailored to the individual operation.

Hill concluded with this point:

I am not out to promote data warehouse as a solution in itself; it is the only way to get the job done! You need atomic detail in a relational manner on a high performance system to provide intelligence in a timely manner.

3.4 Midwest Card Services

Midwest Card Services (MCS), a subsidiary of Fishback Financial Corporation, has over fifteen years of credit card experience servicing prime, subprime, agent bank, affinity, and corporate card programs. In addition, MCS has expanded into telemarketing, debit card, merchant processing, and ATM servicing experience. [14]

John Folkerts, CTO and VP, was interviewed about MCS efforts in data warehousing.[15] He has the responsibility for all technology including the data warehouse, analytics, software development teams, networking teams, and systems interfaces – which covers "essentially everything" dealing with IT in the small company.

Folkerts explained the business of MCS as follows:

> MCS is basically a servicing company for clients having the financial products. For example, MCS does the application processing of credit cards, along with sending the cards out through a third-party, customer service, payment processing, collections, and all account maintenance. Our customer could be a bank, which has product managers. Those product managers decide what features that product should have, and then MCS services that product for the bank.
>
> Credit card processing is the big one. However, MCS does some merchant work, servicing the backend processing for their retail POS data. MCS also does short-term specialized loans, which is a very large business. MCS has also ATM networks and provided cash to those networks. We do collections for credit cards and loans for banks.

Allowed MCS to reduce 23% of their staff …by converting many manual activities to automated procedures driven by the data warehouse.

Folkerts shared that MCS currently has about 150 employees, from a peak of 400 employees. Their technology initiatives, especially process automation, have allowed MCS to reduce 23% of their staff with MCS by converting many manual activities to automated procedures driven by the data warehouse.

Folkerts calls their warehouse 'Diamond Well' or just DW. The telephony call center is called 'Gold Pond', and the accounting system is called 'Silver Lake'. A 2-node Teradata system is used for the data warehouse, along with processing all analytics. They started their efforts at the beginning of 2000 and installed the Teradata in late August of that year. Data loading and initial reports were done by late September or early October, resulting in a quick ramp-up time for the data warehouse.

Folkerts gave the following characteristics of their data warehouse:[16]

- Single version of the truth that is consistent
- Foundation for decision-making
- Accessibility of business information when business needs it
- A secure fortress that protects our biggest asset– information

Folkerts also described their user base and staff.

> We have primarily two groups: operational groups to achieve efficiencies, and business people to manage products. The data warehouse touches all 100 operational people, but they do not directly access the data warehouse. There are 5-6 daily users on the operational side and another half dozen people on the business side who manage the products, sales marketing, and product development.
>
> A third group is our application development team who create these process automation systems. There are 9 people. We have 4 people solely dedicated to

> developing applications and database ETL on the data warehouse. And, we have 2 persons on analytics, on the business and finance side. They are quite technical, but they are not programmers.
>
> The data warehouse supplies data to the development team for setting triggers to automate manual processes. Any time you have a repetitive manual process with strict business rules, you can automate it. So, that was one of the big wins that we had. By using the data warehouse to track history, monitor activity real-time, and then create automation systems.

Folkerts outlines the data sources for the data warehouse as follows:

> The data warehouse receives data from 6-8 source systems. The largest source is First Data Resources (FDR) who processes the credit cards. The daily feed is 30-40 MB.
>
> In addition, there is the real-time feed from our internal payment processing systems and telephony systems. Then, we get large external daily feeds from the short-term loans system, which are 10-20 MB each per day. The number of transactions is quite large, relative to their small size in MB. These are fairly compact. And, we have a few other sources inside the company.

Folkerts described the size of the data warehouse as follows:

> It is two four-way nodes with a 1 GB RAM per node. We have a total storage of 600 GB, with 350 GB of loaded data or about a third of a terabyte. We have grown steadily over the last 18 months by a couple of hundred GB every few months. A year from now, [we estimated that] the data warehouse will be at 600 GB.

When asked about the cycle time for data loading into the data warehouse, Folkerts explained:

> The data warehouse loads data twice per day, in the morning and in late afternoon. Most of the data is current as of yesterday.
>
> We do all our own custom ETL procedures that run automatically in scripts. We go out to a FTP site to get the external data. We have a few systems using TPump for near real-time loading directly into the data warehouse.

Folkerts emphasized that the automation processes use data refreshed every few hours:

> An example of an auto system is when a card is issued and not used for a 60-day period. We deactivate that account by doing an 'unwind' to reverse the annual fees, reverse the general ledger, close down the account, and remove the account from the system. Each step was a separate manual action. We have automated the whole process, interfacing with FDR.
>
> A second example of an auto system is one that detects fraudulent payments. It will go back and look for payments for that individual within a recent timeframe, comparing it to total payment history for that individual, giving a potential fraud indicator. If the payment is unusually high for that individual, we place a hold on that account until that payment clears. Again, this was done manually; now it is all automated and a little smarter because it looks at some past history.
>
> A third example is about an unusual life event, such as, the person won the lottery, his grandfather died, or he got a severance check. The call center should proactively call that person to determine the situation. If he did win the lottery, then we should offer him new products. If he did lose their job, we want to know about that because it affects his ability to perform on other products.

Folkerts stressed the concept of single version of truth and related his experience achieving it:

I agree with that concept 100%. For example, we had many people querying from a variety of operational systems. Sometimes they would build their query with subtle biases. They would query one system and compare it to another, thinking they had similar fields (but not really). We had multiple versions of truth!

We trained users that the data does not lie.

When we built our data warehouse, we trained users that the data does not lie. We spent extreme amounts of time insuring that the data was defined correctly and that they understood the data. It has helped our clients make some significant business decisions. The analysis paid for itself because we were able to do the analytics, predict the future, and know the real story.

For example, we never quite knew how a specific product or portfolio was performing, prior to implementing the data warehouse. There were always monthly adjustments; the data was a couple of days old; and so on. We did not think we were getting ahead on one product, but we could not put our finger on it. After the data warehouse, we had up-to-the-minute data that was accurate, and we made sure that everyone understood what the number meant.

In the case of one client's product, it was clear going forward based on forecasting, they were not going to make money on this product. We had very high confidence in the data and knew that it was the truth. The client then made the decision to stop the issuance of this product. That was about a $600,000 decision. Had we not made that decision for 3-4 months, there would have been a $600,000 impact. Luckily, the decision was made sooner rather than later.

Folkerts shared several suggestions based on their experiences:

Do not be afraid to ask for help. Know your internal strengths and weaknesses. Build on your strengths, and find resources (outside if need be) to counter your weaknesses.

The biggest thing is educating your users that the data warehouse is a tool for them to use. It is strictly a tool. It provides information to make decisions. We spend most of our time with those 10-12 power users, rather than the front-line people, educating them about the business value of information.

Get executive support at the very beginning. Win a few small battles along the way! Do not try to solve all the world's problems. Provide some immediate ROI.

Determine the long-term resources that are needed. One of the reasons that we went with Teradata is the lack of administrative overhead. I do not have the luxury of 2-3 DBAs around here; I am lucky to have one.

3.5 PING Karsten

PING®

In his garage in 1959, Karsten Solheim designed a revolutionary putter and founded PING, Inc., named for the distinctive sound make when striking the ball. Based in Phoenix, the company produces a range of golf clubs, accessories, and apparel. Interestingly, PING also supplies parts for the Patriot missile and to the helicopter industry.

Kent Crossland, Information Services Director, started the interview by referring to their data warehousing efforts as being 'from Mars':[17]

I forget to even use the term 'warehouse' because it is just a [single] database.

You just connected with Mars. We are aliens. We use the Teradata database because we built a system ten years ago using parallel processing for both transactional processing and decision support. We implemented the first phase of our order processing system around 1990 using SQL/DS under IBM VM. To upgrade that system, we needed a water-cooled IBM mainframe, which was too expensive for our company.

[Switching from SQL/DS] we found that, with the Teradata database technology with massively parallel architecture, we can incrementally add storage or proc-

> essing power. It was early client/server architecture before we knew to call it that. As an unconventional alternative, we evaluated Teradata's database computer [as the replacement for the IBM VM system]. We made it work! We were running counter to the industry wisdom then and continue even now.
>
> I forget to even use the term 'warehouse' because it is just a [single] database.

When asked whether PING's small size allowed them to pursue this unconventional single-database architecture, Crossland replied:

> We are fairly typical for a midrange company. We have 500-600 desktops, of which about 400-500 are connected to Teradata. However, there are only 1-2 analysts who are really exercising the system with large-scale queries. We really struggle with the small, simple transactions. But we get enough performance from Teradata. We fulfill about 1000 to 3000 orders per day.
>
> The data warehouse size is 100GB with 100 tables. The largest tables are: order item with 5M rows, and the Serial Number table with about 2M rows.
>
> I recently did a table-scan query of a 3 million-row table. I got the answer in 2 minutes, without impacting any in-house transactional application. It ran in low priority.

Crossland emphasized that their business required a customized system to meet their unique requirements:

> We are a highly customized configuration product, involving many variables that go into building a golf club. You order [a golf club] by specification. What model, clubs in set, shaft, length, color code, grip, and other parameters. We optimize material planning for better inventory. There is no standard application transactional software. We developed the applications in-house.
>
> Our applications are highly customer-centric, from customer order to product ship. Everything revolves around our relationship with our customers, recording who a customer is, taking his order, processing through assembly, recording the serial number and specs of the clubs, shipping it out, capturing the shipping charges, invoicing the customer, billing, accounts receivable, credit collections, and customer support.
>
> The whole order-processing cycle is on Teradata. Taking order, scheduling through production, assembling it, and billing the customer– that whole cycle is on the Teradata. We have recently added some CRM functions, and all that data will be on Teradata. Customers will be able to look up specifications on their clubs.

Crossland described their users as follows:

> Most of our users are operational types: customer service rep, sales people, production people, credit people, shipping clerks, and operators who are stamping serial numbers on the golf clubs.
>
> In marketing analysis, there is one analyst who generates reports that people can interact with. There are a couple of [power] users that can do their own queries. But, designing queries is a difficult thing for most users. We have about two dozen users on the Cognos analyst tool; however, most are using canned queries. It is a difficult sell [to get them to do more].
>
> The idea that you can be a causal user sounds nice, just point-and-click to analyze your data. Makes a great presentation. But, you have to know the data structure and be very familiar with it, to have confidence to pull the proper information from the database.

Leveraging SQL extensively, four developers implemented the system with APL under VM and with SQLWindows (Gupta tool) via ODBC. The Teradata database is both channel-attached to the IBM mainframe and network-

attached for ODBC access. The Cognos tool and Excel extract data using ODBC.

The concept of data freshness does not exist at PING. It is zero latency from transaction commit to analysis reporting.

Crossland related that they run their entire business on Teradata, but have only one DBA. And, he is doing it only part-time.

An interesting answer came when Crossland was asked about data freshness:

> The concept of data freshness does not exist at PING. It is zero latency from transaction commit to analysis reporting. We know each minute what the status is. We have a 48-hour turn-around on orders, so it is a highly dynamic process. We built the system so that the production people could manage that dynamic.

Crossland remarked that their data warehouse usage for event-driven analytics was probably in the middle ground:

> Instead of getting a report at the beginning of the day, most users sit in front of an application throughout the day, seeing the current information. Our users do not even know about batch processing and periodic reporting. Everything has been instantaneous. You go to the IT department, and an hour later you have a report.

Crossland emphasized the ease of management as the primary benefit from their single-database approach:

> The great benefit is ease of management. There are no issues of timing or versioning. There is no separate environment for querying and all the data extracts. There is no question as to what data is accurate. We have only one source for the data. This is all based on the ability to perform ad hoc queries in a timely basis without waiting for the data to update or for off-hours.

Crossland commented on developing operational applications for the data warehouse environment:

> I would recommend developing application software on the Teradata platform, since for us there is tremendous evidence that this concept works. What makes it works is the parallel environment of Teradata.
>
> Because its primary role [of Teradata as perceived by the industry] is as a data warehouse, you must engineer an application for the Teradata environment. It is not conventional thinking. When people build applications, they do not think about doing it all in one environment. We have proven that it works.

Crossland concluded by reflecting about their experiences as single-database pioneers:

> Years ago we thought that this single database would take off. We were early in that game but knew [expected] that one day it would be pretty standard stuff. … I am surprised that the single-database approach is not universal. … I felt that the industry would soon catch up. Well, more than 10 years later, we are still pioneers.

3.6 Strategy.com

…one big real-time data warehouse, designed to provide proactive alerting.

Ceasing operations in the fall of 2001, the company was called Strategy.com, a division of MicroStrategy Inc. Although no longer in business, Strategy.com is an important example of "one big real-time data warehouse, designed to provide proactive alerting and reporting services to consumers."[18] Strategy.com demonstrated the technical viability of large real-time data warehousing.

From financial markets to sports events to tornado forecasting, the system drove huge amounts of real-time alerts. A press release at its zenith provides a glimpse into the nature of the Strategy.com venture:[19]

> Strategy.com, a leading provider of one-to-one messaging through web, wireless and voice, announced today that it now has over 500,000 active subscribers and delivers over 3.5 million messages a week. Strategy.com services empower companies with online subscribers to deliver near real-time, personalized news and information services to their customers. Information on topics such as finance, news and weather are delivered through various electronic devices, including pager, email, fax, phone and web-enabled phones. … Strategy.com enables companies to offer personalized content to their customers via a full range of devices, providing the opportunity for tremendous brand awareness and customer loyalty.

Justin Langseth was the CTO and VP of Product Strategy and is currently a principal at Claraview. He mused about the capacity for the Strategy.com system: [20]

> Strategy.com got up to 350,000 to 500,000 users, supporting from worldwide stock feeds real-time stock alerts to users of Ameritrade and Wall Street Journal. A large Teradata system powered that service. Since the system was designed for 10 million users, we had much more capacity than the 500,000 users. We were ready for more users and pushing the boundaries on being more proactive and real-time.
>
> The article that appeared in Intelligent Enterprise Magazine[21] tried to point out some interesting things from our experiences at Strategy.com, as we used BI tools to do more proactive and real-time things. During those dot-com years, the consumer marketplace was the place to be. We were in the process to refocus on a more enterprise market, when our investors decided to pull the plug on the operation.

In a recent article, Langseth gave a good overview of the system architecture:

> All the data for these services was loaded into a large, six-node Teradata database system throughout the day. We used a combination of near real-time and true real-time data loading technologies, all custom built in C++ and Visual Basic. … At the peak of our service, we were providing real-time alerts to more than 350,000 users through email, wireless, and voice technologies, and we had capacity for more than one million users. While the business model of our dot-com didn't work out as planned (whose did?), the technology was rock solid. While some custom development may still be required to mitigate some of the challenges of real-time warehousing, advances in real-time ETL and other related technologies are now making things much easier.
>
> Unfortunately, most BI tools are limited to 5-minute cycles. For some business situations such as trading floors, we need to push that down to 30-second cycles. We are pushing the edge…

When asked about active data warehousing, Langseth defined it in terms of value to the enterprise:

Only 10% of the potential users get any value out of the data warehouse.

> The basic thought is that data warehouses are delivering value to the enterprise. However, only 10% of the potential users get any value out of the data warehouse because of the lack of training or ease-of-use. So the potential value can be much greater.

3.7 Union Pacific Railroad

BUILDING AMERICA℠

Union Pacific Railroad is one of the largest railroads in North America, generating $12 billion in revenues with 48,000 employees. With over 33 thousand miles of track, the system serves 23 western states, linking every major West Coast and Gulf Coast port. Union Pacific serves major gateways to the east (Chicago, St. Louis, Memphis and New Orleans), along with being the primary rail connection with Mexico and Canada. The Corporation's trucking operations include Overnite Transportation and Motor Cargo. The railroad has one of the most diversified commodity mixes in the industry, including chemicals, coal, food and food products, forest products, grain and grain products, intermodal, metals and minerals, and automobiles and parts.

The challenge is keeping track of 6,921 locomotives and 153,272 freight cars. We interviewed Paul Evans, Senior Manager of Enterprise Data Warehousing, to obtain an insight into this challenge.[22] He is responsible for all extract, transformation, and load processes that build their Enterprise Data Warehouse (EDW):

> I have a team of business analysts and data engineers who review business requirements for extending the data warehouse and deal with existing components for new requirements. I also have the technical DBA and the system administration staff. So, basically I have the complete environment. I really only have the support role with analytics side. The end-user consulting group does most of the analytics.

…the focus on getting all of our Decision Support data centrally managed.

Union Pacific has recently focused on data consolidation into the data warehouse and data delivery out of the data warehouse:

> The biggest things that we have changed is the focus on getting all of our DS [Decision Support] data centrally managed. Today we have some dependent data marts outside the data warehouse. First, we are rolling everything together into a common framework to simplify for us the management of the environment. Second, we are rolling out this enterprise-wide BI architecture to make it easier for our users to find information.
>
> We went production about 10 years ago. Almost all the subject areas across the business are now represented in the EDW and are integrated together. All the above applications have some components in our warehouse, and we add more as we go on. Recently most of our effort is in replacing our old systems. For example, the crew management system is being replaced and is being converted to relational web-based system. EDW is the pivot point for new systems.
>
> And… Everything comes into and goes out of the Teradata box.

Evans explained the complex feed of the data warehouse:

> From a technology standpoint, we are feeding data from mainframe (DB2, IMS, Focus, general files) along with our homegrown system called TCS or Transportation Control System. TCS runs and manages the railroad! In addition, there are a couple of hundred Oracle servers. Some servers are geographically dispersed, but most are consolidated in the data center.
>
> From a business application standpoint, most of DB2/IMS systems are the typical back office applications, such as payroll account receivables, etc. Most of the Oracle systems tend to be operational focused, such as train dispatch, crew assignments, and logistic systems.

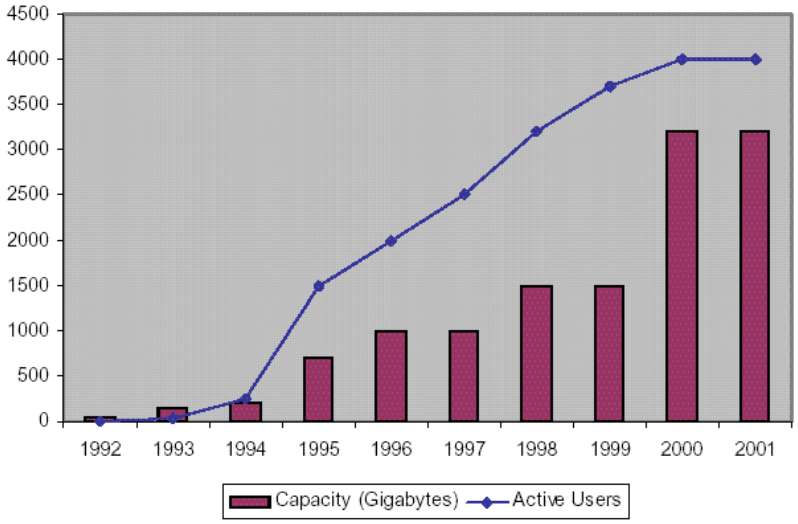
Evans described a fairly large user base from a cross-section of the company:

> We have a total of 4,000 persons who log into the EDW every month, cutting across all the different user departments. We have some field personnel logging

in. We have headquarters people who are doing actual line work, rather than management. For example, we have clerks who process cash receipts by identifying the actual freight bills and by applying the cash to the proper receivables. The FSD department develops financial applications that operate against the EDW.

The ratio of operational versus management users is roughly half-and-half. You may have 20 clerks using an EDW application and only one management person doing ad hoc analysis.

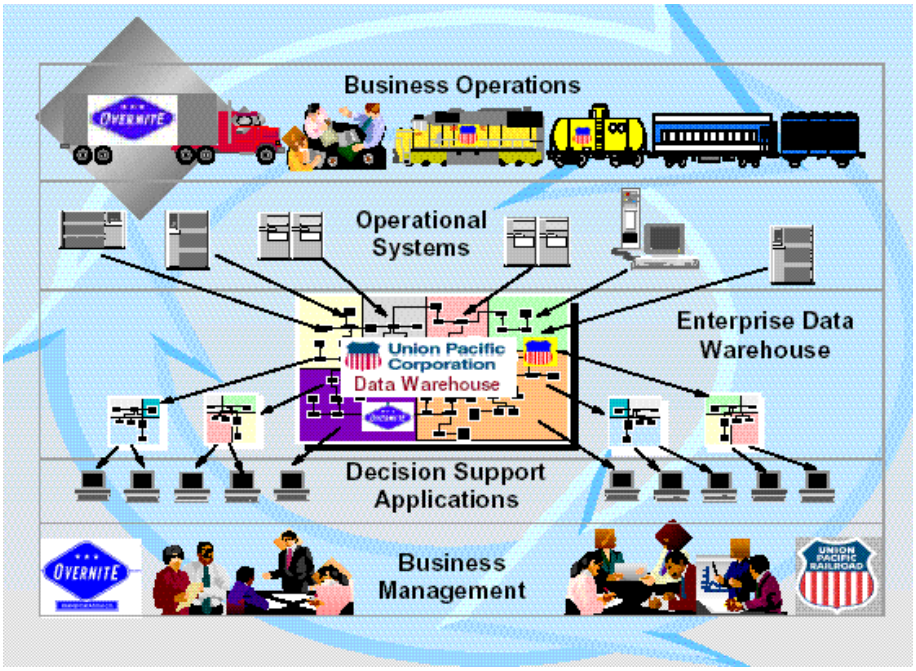
As shown in the chart, the number of users has plateaued. We have reached the saturation point of potential users. As we rollout the new BI platform, it will cause the number of users to grow. Total employees at UP is 48,000 so EDW users is roughly 10%. We have 10,000 to 12,000 desktops. I am expecting to see it plateau around 7,000 users in 3-4 years.



The data warehouse grew to 4 TB in 2001. In October of 2002, we will be at 7 TB. The jump is primarily caused by the philosophy shift of centralizing operational data on the EDW platform. We have roughly 2,000 physically tables with 4,000 to 5,000 views.

And… We have only four DBAs! If we went to an Oracle platform, we would not be able to manage with four DBAs.

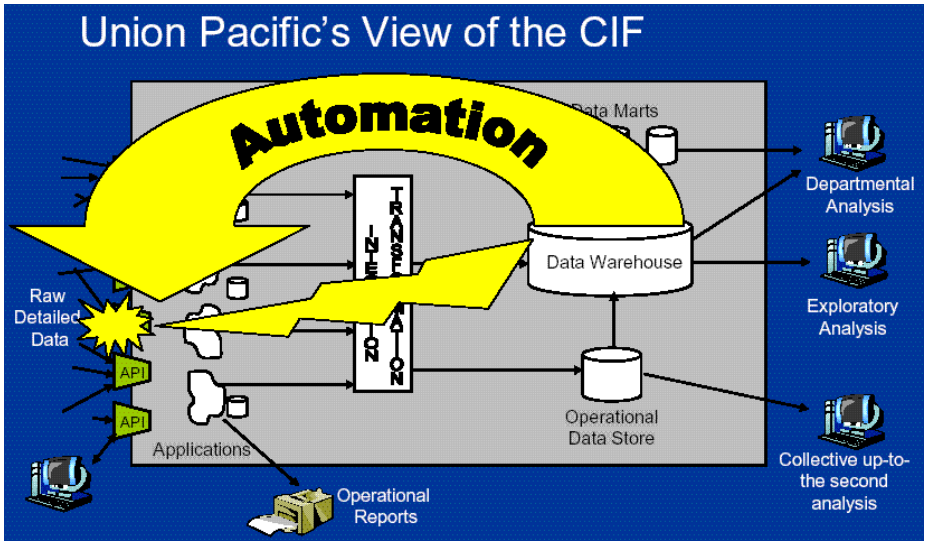
Evans illustrated their systems architecture with the following figure.[23]



Evans described the illustration as an inverted pyramid with a closed loop in the background:

> The key part is a closed loop system (in the background). In the Decision Support structure, those reports spawn actions. We must recognize that loop and automate it, so that we take the people out of the loop. We can prompt actions quicker and more effectively with higher repeatability, etc.

In another slide about Union Pacific's view of the Corporate Information Factory, he emphasized the closed loop as automation pointing back into the operational applications:



The operational side will definitely be growing more than the management side.

They are improving data latency, with hourly updates in the future:

> One of the new things that we are doing is improving the latency of our data, moving toward an active data warehouse. About 90% of the EDW data is refreshed daily, with the rest done monthly. Within the next year, a lot more data will be refreshed within an hour.

> As we put that [highly fresh] information out there, this will cause greater growth in operational users who are driving tactical decision-making. The operational side will definitely be growing more than the management side.

Evans pointed out several business areas that warranted hourly refresh cycles:

> Most of this data is TCS information with train movements, location of equipment, and the like. The crew management system is our first effort. They want to know where the trains are so they can place crews on those trains that are waiting for crews. The train is 'held for crews', implying that the train is sitting there idle. They need to know what trains are being held and the supply of crews. In addition, they should know what customers are waiting on those trains, what service agreements are in place for those trains, and what will it cost to defer this train versus another train. Here we are tapping into our financial, profitability, and customer contract data. This level of integration is difficult to achieve with just an OLTP system. We are pulling all that data into the EDW so that we can make a more informed decision about prioritizing and allocating the scarce resource of crews.
>
> The issue is whether having that [fresh] data has cross-application opportunity and integration potential? If it just a silo question that can be answered from the OLTP system, then build it as part of the OLTP without the overhead and expense of pulling the data to yet another location. So, we are really looking for those integration opportunities. [It is for] those types of data that we are improving the latency.

Evans also emphasized that data integration is the primary reason for the EDW:

Integration is now where we are getting the most return on investment.

> We had other reasons for starting our warehouse, but integration is now where we are getting the most return on investment. Being able to ask questions and get answers that we were never able to get before. Without the EDW as the integration point, we would be taking 6-12 months, for each question, to build interfaces among applications to answer those questions. In this case, the business deal that we were trying to close has long since gone to our competition. That level of integration with data staged and ready in the EDW gives us the power and flexibility to react to changes in the marketplace and to respond to requests from our customers. We have landed some multi-million contracts from this ability.

When asked about the benefits, Evans citied several:

> Litigation Research: We have had a surprising use of the EDW in our legal department. For example, one of my staff is flying to California to give a deposition in a litigation case based on data warehouse data, which is quite common. The key is that we have the information. Sometimes they are to our advantage, and sometimes they are not. As a company, we have a good comfort level of knowing the truth! If it was our fault, then we should settle this one up front. If it was not, then we have the data to support our case.
>
> Bill Auditing: We have been doing things with our auditing billing processes. We have discovered situations in our accounts payables where we have been paying for things that we should not have. In other situations, we did not bill for things that we should have. We have definitely recouped several millions of dollars per year by auditing our payables with EDW data.
>
> Customer Self-Service: We have several hundred customers who can query data in data warehouse via the web, such as getting pricing for services.

When asked about suggestions to colleagues facing similar issues, Evans suggested an important lesson from his experiences:

The one thing that works for us is that we have a very open philosophy with our data. We still must have security to sensitive data. However, our philosophy is that any user can ask any question of any data at any time. They do not have to ask IT or wait for a certain time of day. Just ask away! Again, they must have the proper security.

…any user can ask any question of any data at any time.

It is a two-edged sword. Takes IT out of the way, and the users can ask away. They are freed to answer their questions when they need the data. On the other side, we do not know the true value of our warehouse because they do not have to tell us. If we do not hear about it, then we do not know about it. We process 6 million queries per month, so there is too much volume to analyze. We must rely on word of month. It is a problem when we are trying to justify upgrades for our equipment.

We continue to stress the importance of metadata in the hands of users. Although it is industry wisdom, we need to continually do this. We publish on our internal website all the information (tables, views, columns, etc) in the EDW. When they want to answer a question, we have a search engine to find the right information. We built our own search engine to do this.

In passing, Evans offered an insight into their data warehouse design, which is based on Third Normal Form (3NF):

Although it may not be too popular in the industry, our EDW is actually a normalized model. We avoid dimensional or star-schema modeling, except when we have a performance problem. Even in those situations, we retain the normalized data. This has been advantageous to integrate and extend our data. The dimensional modeling is more evasive to bringing new data elements into the mix. Normalized modeling is easier to integrate new data.

When asked about data marts, he said that they exist but are managed within the central data warehouse:

We used them, but we do not have any independent data marts. We stored any of those data marts on the Teradata platform within the warehouse. They are simply an aggregation, summarization, or selected pieces of the normalized data, copied into separate physical tables. We call it 'managed redundancy'.

Finally, Evans offered two factors that he felt contributed to their success:

We have had a small group of individuals from the beginning (continuity) and a drive toward the goal of having everything integrated. We will have pressure to implement an application in a silo fashion. It is quick and easy. Just do it, so we can get on with the work. Achieving integration is not easy; it takes time. We have been fortunate to have management who recognize the benefits to the business of this integration.

Another key is that we were fortunate to have a technology vendor with a scalable platform that has supported us for an entire decade on the same platform. When you are trading horses, it takes time, and you are not delivering solutions. The vendor is Teradata. We were an early adopter of their technology, and they have treated us really well.

4. Common Themes

Now that we have read through the experiences of various companies with their data warehousing efforts, let's synthesize these experiences into common themes that distill the lessons learned.

4.1 Single Version of The Truth

The dominant theme through many interviews was the goal of a single version of the truth (SVT) within the data warehouse. The objective is to ensure that the data warehouse contains consistent and accurate information about business reality. By analyzing the warehouse data, the best understanding of our business, both its historical performance and current characteristics, can be obtained. A success indicator is the universal opinion that the warehouse is the only place to go for business information.

…consistent and accurate information about business reality.

3M and Union Pacific stood out as good examples of long-term, enterprise-wide efforts that are achieving SVT through source integration, data mart consolidation, and user training.

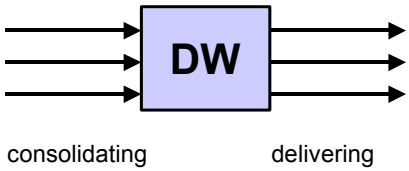
From the interviews, it was apparent that achieving SVT in a data warehouse is a continuing struggle involving many parties. A better phrase may be a 'unified version of business reality' because the word 'unify' conveys more meaning, having flavors of: integrated, united, amalgamated, fused, and cohesive.

Evans at Union Pacific is driving toward the goal of "having everything integrated," but cautioned, "Achieving integration is not easy; it takes time." He continued:

> We have been fortunate to have management who recognize the benefits to the business of this integration. [Without this integration] we would be taking 6-12 months, for each question, to build interfaces among the applications to answer those questions. The business deal that we were trying to close would have long since gone to our competition.

It is hard work, much more than selecting one version versus another. The step of determining which source is the 'system of record' for a certain subject is only the beginning of assuring SVT. There is always the temptation to implement new BI applications piecemeal, adding independent data stovepipes to the data warehouse architecture.

There are two perspectives to SVT, as shown in the figure. First, there is consolidating diverse data sources into the warehouse environment. Second, there is delivering information from the warehouse to diverse user groups in a variety of formats and channels. SVT requires excellence in both of these areas.



Hill described the situation for delivering information at Ford as follows:

> The main characteristic at Ford was the speed element (i.e., how fast information was flowing). It all has to be detailed down to the part level number, and then rolled up to the bag, box, pallet, container, and finally the shipper. It does no good to tell the guy who is managing traffic to look for a part, when there are hundreds of trucks in the yard. There is a constant demand to keep the data at the level that the people are dealing with it. You are continuously dealing with bulk shipment of spark plugs, to individual washers in bags, to very large boxes (like fenders and chassis). The trick is to have sufficient details, but in a hierarchy that keeps it all straight for you.

…implies a trust, in both directions, between business users and IT staff.

SVT also implies a trust, in both directions, between business users and IT staff. Business users must trust the IT staff to deliver accurate and consistent information about their business. Inaccuracies and inconsistencies must be few and resolved quickly. Likewise, IT staff must trust business users to utilize the data warehouse in a proper manner, achieving real business benefits. At Union Pacific, this trust is tested daily with their open information policy that, subject to security restrictions, allows any user to ask any question of any data at any time. Further, that policy continually tests the adequacy of their IT infrastructure.

4.2 Business Value of Analytics

The second theme was discerning and realizing business value from analytics. In other words, does putting 'smarts' into the hands of executives, knowledge workers, and operational personnel generate real benefits to the business? For years, we have debated the impacts on executives and knowledge workers. A decade of data warehousing history has demonstrated value here. Now with active data warehousing, operational personnel have been added to the mixture.

It is a complex issue, involving more than technology. Data warehousing technology has certainly improved, along with new modes for information delivery. More information is not always proven to be better for the business. The organizational culture and interpersonal dynamics are also important factors that influence the effective business value of information.

…more important is training in the business meaning of data.

From the interviews, it seems that training and motivation are critical factors. Training users in the proper usage of data warehousing technology is important. But even more important is training in the business meaning of data, which involves a lot of training about the business itself. After training is motivation. Compensation and supervision must be in synch with data warehousing innovations to solidify the role of analytics into the organizational matrix.

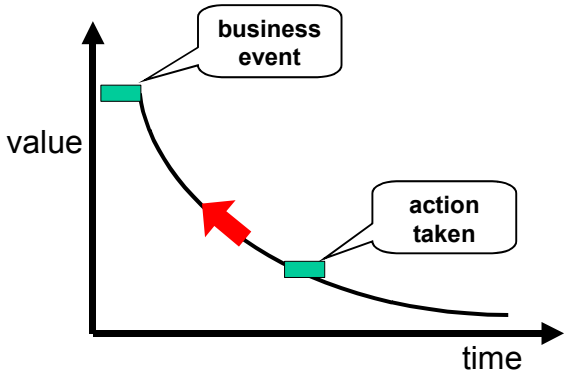
4.3 Business Value of Data Freshness

Whenever a question about active data warehousing was asked, the majority of the discussion was on real-time warehousing. In other words, active data warehousing is perceived to be similar to real-time (or near real-time) data warehousing. As we got into the discussion, it become apparent that this concept was complex and that everyone is searching for a common framework.

Some companies have boasted that, by the time the customer leaves the parking lot, their point-of-sales data is ready for analysis in the warehouse, thus resulting in a business benefit to the company. However, how has storing

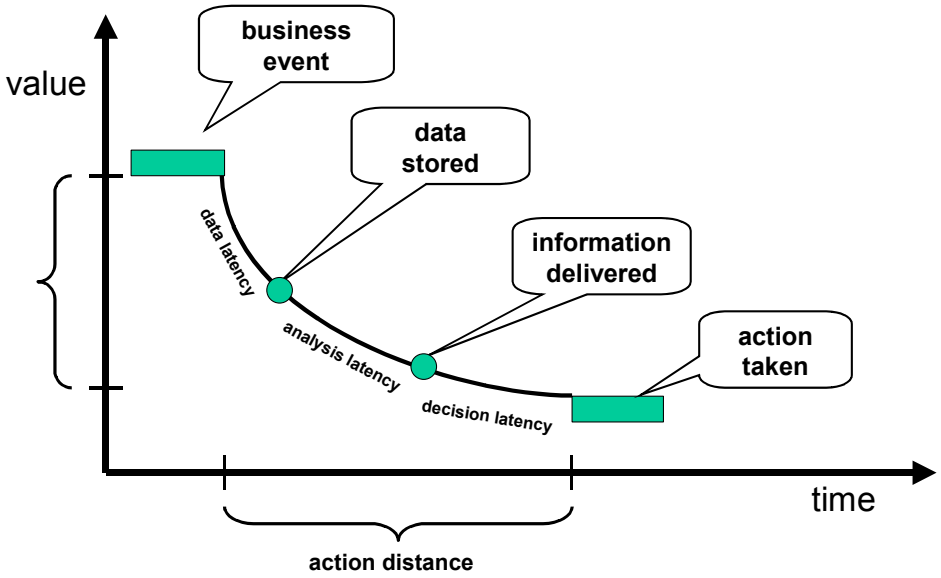
data faster into our warehouse resulted in a positive impact on the bottom line?

What is the relationship of data freshness to business value? Consider the following Value-Time Curve.



Most consider the relationship as a simple decay function. A business event happens; then an action is taken. If the business value of taking that action decays rapidly after the event happens, then the objective is to push up the value curve by minimizing the delay or latency.

The real situation may be more complex, as shown in the following figure.



A business transaction occurs over some time duration and then ends (commits/aborts). At a later time, the data about that transaction is stored within the warehouse environment. At a later time, the data is analyzed, packaged, and delivered to the proper person. At a later time, the person takes an action based on the analysis.

Action distance is the end-to-end time required to respond by taking action in response to the business transaction in an intelligent manner. There are three different factors involved with action distance. First, there is the data latency, the time between the business transaction and when the data is

ready for analysis in the warehouse. Second, there is analysis latency, the time of initiating the analysis, packaging its results, and delivering it to the appropriate person. Third, there is decision latency, the time required to absorb the information and respond in an appropriate manner.

Decision latency will be the limiting factor in the long term.

It is important to realize that only the last one—decision latency—really counts on the bottom line. The first two are overhead, simply infrastructure necessary for the third. Technology advances are greatly reducing data and analysis latencies. However, decision latency will increasingly become the limiting factor.

…when you have it in your hand.

In the supply chain management context at Ford, Hill illustrated this point by saying that you need the information when "you have it [the part] in your hand" – not before and not after.

> When material is being received, it is scanned. That transaction is entered into the system, and the system should immediately come back with what to do with it. …While you have it in your hand! We need to focus on those situations that generate a transaction and the people doing the transaction need an immediate response of what to do. Out of the gigabytes that we load each night, there will be a fraction that needs to be active and real-time.

In addition, Hill stressed the importance of delivering the entire context for the information.

> It all has to be detailed down to the part level number, and then rolled up to the bag, box, pallet, container, and finally the shipper. It does no good to tell the guy who is managing traffic to look for a part, when there are hundreds of trucks in the yard. There is a constant demand to keep the data at the level that the people are dealing with it. You are continuously dealing with bulk shipment of spark plugs, to individual washers in bags, to very large boxes (like fenders and chassis). The trick is to have sufficient details, but in a hierarchy that keeps it all straight for you.

This brings us to the concept of touch-point opportunity, which is taking advantage of a high value situation by directly impacting the quality of the business transaction within the scope of that transaction. A touch-point can be anywhere in the enterprise. If internal, it is like the handling of a part on the loading dock at Ford. If external, it involves a customer, or a supplier, or some business partner. In terms of impact on the value chain, the external touch-points are usually the high value ones.

…impacting the quality of the business transaction within the scope of that transaction.

For instance, a person walks into a bank and talks with a teller. If the teller can utilize insightful information about the customer during their discussion, the value can be quite high. If that data is ready for analysis after the customer has left the parking lot, we have lost the opportunity to assist that customer with his unique requirements during that specific visit.

4.4 Actionable Granularity

In the early days of data warehousing, the flow of source data into the data warehouse involved transformations that performed significant aggregations, so that the volume stored in the warehouse was considerably less than the volume extracted from source systems. The advantage was that the warehouse could be much smaller using less expensive technology. Besides, why would anyone want to retrieve detail data from the data warehouse when the transaction systems already contained that data?

That perspective is changing rapidly as the data warehouse becomes more actionable. Analyses to detect and understand a business problem are often general in scope, requiring aggregations across lots of data. Analysis to respond and correct a business problem requires precision and accuracy.

Actionable requires drill-down to the atomic level.

Messerli formerly of 3M stated, "The data warehouse is atomic level, which makes it actionable." He continued:

> We can drill down to get the credibility so that we can guarantee that contents match the source transaction systems. From that, you get quality, credibility, and actionable. … That is why there is drill-down to atomic detail. Most other warehouses have only aggregated or summary data. When they see a problem, they must delegate it to the department manager, who must go back to the source transaction system to see what the problem really is. But with the atomic detail, you can close that gap in the data warehouse.

4.5 Work Automation

The last common theme emerging from the interviews is work automation driven by analytics from the data warehouse. Automation strives toward mass production through functional specialization (i.e., doing a small task in a uniform manner). If we fully understand a business process and are able to control (or anticipate) the environment, then we can efficiently standardize products and services. Automation also implies that the human element is minimized or even eliminated, which maybe a cost saving.

Folkerts at Midwest Card Services described several examples, such as the 'unwind' of inactive accounts in cases where a credit card is issued and not used for 60 days. An automated procedure has replaced the manual tasks of reversing the annual fees, reversing the general ledger, closing down the account, and removing the account from the system, along with manually tracking the workflow.

Evans at Union Pacific described the crew management system as one of their first automation efforts. If a train is sitting idle waiting for a crew assignment, they need to tap a variety of information to make an informed decision:

This level of integration is difficult to achieve with just an OLTP system.

> They should know what customers are waiting on those trains, what service agreements are in place for those trains, and what it will cost to defer this train versus another train. Here we are tapping into our financial, profitability, and customer contract data. This level of integration is difficult to achieve with just an OLTP system. We are pulling all that data into the EDW so that we can make a more informed decision of prioritizing and allocating the scarce resource of crews.

5. Lessons Learned

This section summarizes those 'gold nuggets' from the interviews, organized as success factors, business challenges, technology challenges, best practices, and benefits realized. Pay close attention; the project that you save may be your own.

5.1 Success Factors

We asked about the factors that made the project successful. Answers varied from common industry wisdom to lucky coincidences to new insights.

5.1.1 Securing Executive Support

It is industry wisdom to have strong executive support for any IT project, and large data warehousing projects are no exception. 3M had a major fragmentation problem among their 50 business units and 60 international subsidiaries. Top executives at 3M recognized that the solution required both a steady multi-year effort and considerable technology risk. Crossland at PING attributed the selling and educating of management as invaluable to their success.

Attaining (and retaining) executive support is more than playing the right politics.

Attaining (and retaining) executive support is more than playing the right politics. It is an open, two-way partnership between corporate executives and IT management. Trust in the personalities, intentions, and competency between both groups is key. Trust builds slowly through incremental successes. The challenge is to effectively build trust within your unique organizational culture.

Folkerts at Midwest Card Services suggested, "Win a few small battles with immediate ROI, and avoid solving all the world's problems" as a way of advancing this support.

5.1.2 Enabling Business Units

It is the business unit, not IT, who understands and lives with the business issues. By pushing the analytical expertise out to the business units, greater efficiencies are achieved both in reacting to quick, short-term problems and in resolving tough, long-term problems. This implies that IT should give the proper technical resources, know-how, and motivation to the business unit for developing their own applications.

Folkerts emphasized that enabling business users in using the data warehouse was a key factors in their success.

> The biggest thing is educating your users that the data warehouse is a tool for them to use. It is strictly a tool. It provides information to make decisions. Educating them about the business value of information. We spend most of our time with those 10-12 power users, rather than the front-line people.

With data warehousing, there is a greater need to understanding the data, as opposed to operational systems where that understanding is buried within the applications.

5.1.3 Adopting an Open Information Policy

Another success factor was a policy of open information from the data warehouse, subject to security constraints. This policy gets IT out of the way as a

barrier to the corporate information. And, it places responsibility into the hands of the business users to determine proper usage and requirements on the warehouse.

Evans at Union Pacific stated, "Our philosophy is that any user can ask any question of any data at any time. They do not have to ask IT or wait to a certain time of day. Just ask away!" Of course, he quickly added that they must have the proper security.

5.1.4 Simplifying Architecture

Enterprise systems have a very complex architecture. Part of the reason is that these systems have evolved over a 20-30 year period, dragging old legacy systems along as burdensome barnacles on the next generation. With its short 10-12 year history, data warehousing has even succumbed to the legacy barnacles. The ETL infrastructure, in particular, was crafted for an earlier generation, ruled by daily batch windows and log tape (literally) extractions. Technology is better, allowing for many more architectural alternatives.

Simplifying the data warehousing architecture by consolidating data marts.

The recent trend toward simplifying architecture has been data mart consolidation. At 3M, Messerli remarked, "We got rid of all our independent data marts and other data stores, and we did it over six years ago." The function of those data marts are supported within the warehouse using virtual views on the normalized relational data, avoiding multiple replications from the same data sources. Union Pacific eliminated any independent data marts and consolidated everything into a common framework.

The extreme example of simplifying architecture is at PING. Crossland described their situation as: "From the traditional perspective in data warehousing, PING is 'from Mars' since we have, from the beginning over ten years, had a single database strategy for both transaction processing systems and for data warehousing. We are running counter to the industry wisdom then and even now." By using the massively parallel and scaleable technology from Teradata, they "made it work!" Crossland summarized it as, "I forget to even to use the term 'warehouse' because it is just a database."

5.1.5 Stable Vendor With Scalable Platform

With several of the above success factors, an extra burden is placed on the infrastructure for the data warehouse. With data integration, the data warehouse stores greater data volume and complexity. With a simpler architecture, the data warehouse supports more functions, such as ODS and various data marts. With an open information policy, data warehouse has greater demands from unpredictable workloads.

Evans at Union Pacific was appreciative of Teradata as "a stable vendor with a scalable platform that has supported us for an entire decade." Referring to the Ford implementation, Hill highlighted the importance of the underlying Teradata solution by stating, "You need atomic detail in a relational manner on a high performance system to provide intelligence in a timely manner."

5.2 Business Challenges

This section deals with the challenges faced from the business perspective.

5.2.1 Balancing Local Diversity With Global Unity

They must act locally, but they also must analyze globally.

Large global corporations operate their business in many locations, possibly across many time zones and with many languages, customs, and currencies. Yet, they must operate as a single enterprise, accountable to the stockholders of that corporation. Their problem is balancing local diversity with global unity. They must act locally, but they also must analyze globally. The enterprise data warehouse is at the heart of achieving this balance.

For example, 3M has 40 business units and 60 international subsidiaries. To assume that each is an independent profit center managed only by financial objectives is to ignore the realities of the global economy. In particular, the challenge is to realize the benefits from increased productivity, cross-sell opportunities, and customer loyalty, while eliminating redundant operations among units.

The integration effort at 3M was a major task for IT, considering "the sheer volume of data, the hundreds of data sources, the multiple terabytes of storage, creation of thousands of reports, and information distribution worldwide to a massive user community."

Beyond this technical effort is the larger challenge of business process improvements to capitalize on the data integration. In particular, the challenge is determining a common way of indicating business essentials globally without losing the richness of local information. This effort involves a continuing dialog among all parties, responding with unending flow of changes to the business environment.

5.2.2 Discerning Business Opportunities

The challenge of leveraging your data warehouse is discerning the business opportunities and then realizing the benefits. What are the real business opportunities from atomic-level warehouse data? From single version of truth? From near real-time data freshness? From wireless web access to warehouse reports?

Think business first and then technology.

When understanding how business requirements impact the technology, forget the atomic-level, single version, and real-time stuff. Instead, think of the union worker on the loading dock or reservation associate at the airport. Through improved management and even direct analysis reports, what would make a difference in: making the job easier, achieving higher productivity, enhancing customer satisfaction, and closing more sales? In other words, think business opportunities and then work forwards into the technology.

Visualize the key points in your value chain.

The second step is to visualize the key points in your value chain. Every successful enterprise has a clear vision of how they uniquely add value to creating and delivering their products or services. In the manufacturing environment of Ford, Hill of Sagetree sketched the challenge as pushing 'smarts' into the hands of the person who currently controls a vital point of the value chain. For example, whenever the forklift operator touches a pallet while unloading a truck, there is a business opportunity to maximize the value of that operation. It is a fleeting moment, never to reoccur. Simple rules/procedures (first-in-first-out, big parts to the left side, etc.) have traditionally dominated. Can we do better?

Break out of the doing-business-as-usual rut.

The third step is breaking out of the Doing-Business-As-Usual (DBAU) rut. Hyde, formerly VP of Technology at Delta Air Lines, related several conver-

sations with business people about data freshness. The basic issue was, "Is yesterday's data sufficient?" Invariably, the answer was, "Yes, just fine." From their perspective of doing business over the past many years, near real-time data freshness did not have any business value.

As IT professionals, we need to honor that answer. However as IT professionals, we need to challenge and probe that answer. Organizational inertia often blindly propels an enterprise forward in the DBAU mode. Stability can quickly turn to survival in changing markets. The challenge is achieving synergism between knowing the current business and innovating the future business.

5.3 Technology Challenges

In contrast to the business challenges, this section deals with the challenges faced from a technology perspective. In the early days of data warehousing, life was easy. Buy a server platform; choose a database; extract from a few mag. tapes; generate some Excel charts; and receive praise from the executives. Today, life for the data warehouse staff is much more difficult because of the following challenges.

5.3.1 Maintaining Terabyte Warehouses

A few years ago, a terabyte data warehouse was an impossible dream because the technology did not exist to support it at a reasonable cost. Now, a terabyte data warehouse is the norm for any large corporation. The amount of data has increased dramatically. However, if amount were the only characteristic that changed, our current challenge would be quite simple. Just buy more storage.

A terabyte data warehouse implies more than a terabyte of data, a couple of very large tables, and a few voluminous sources. The design typically involves hundreds and maybe thousands of tables, each with many columns and indices, along with hundreds of views and, by the way, add the physical configuration of storage allocations and caches, and so on. Because the enterprise is continually changing in significant ways, the data warehouse design must change in synch.

5.3.2 Managing Unpredictable Demands

Data warehouse users are varied and unpredictable, and the demands on the DW are likewise.

In the early days, one could enumerate data warehouse users on the fingers of one (maybe two) hands and pretty much know what they were up to. Today, data warehouse users are varied and unpredictable, and their demands on the data warehouse are likewise.

Evans explained that Union Pacific is now processing 6 million queries per month against their data warehouse. Our open philosophy is that "any user can ask any question of any data at any time." The users are free to ask their questions whenever they need the data. However, we did not know what the true value of our warehouse is because they do not have to tell us.

5.3.3 Moving From Back-Office To Front-Office

Over its brief 10-12 year history, data warehousing systems have gone through several stages. The early days were largely experimentation and innovation. Once one of those experiments became critical to the business, the

data warehousing system also became mission critical, along with a layer of new requirements for reliability, security, and the like.

We are now going through yet another stage in data warehousing history, which is moving from the back-office to the front-office. It is different than moving from experimental system to mission critical. The technical challenges in the front-office involve the support of a larger and diverse user base who are more focused on the day-to-day operations of the business.

At Union Pacific, the ratio of operational versus management users is currently half-and-half. Evans observed, "You may have 20 clerks using an EDW application and only one management person doing ad hoc analysis. As we put that information out there, this will cause greater growth in operational users who are driving tactical decision-making. The operational side will definitely be growing more than the management side."

5.4 Best Practices

Toward the end of the interview, the question was asked, "If a colleague asked for advice on a similar project, what would you suggest?" The answers were varied and insightful. Some interesting discussions were sparked by this question.

5.4.1 Use a Normalized Data Model

Several mentioned that they utilized a normalized (3NF) data model, rather than a star-schema model, as the design of their data warehouse. The reason was related to the objective of a Single Version of the Truth, as described previously. The normalized model allowed for easier integration and extension of the data warehouse.

Evans at Union Pacific stated the practice concisely as, "an advantage to integrate and extend our data. The dimensional modeling is more evasive to bring new data elements into the mix. Normalized modeling is easier to integrate new data."

5.4.2 Hands-On At Project Start-Up

Just like doing home improvements in an old house, you never know what you will find when you start a project. In the early stages of a data warehousing project, it is recommended that the development staff be hands-on. The emphasis is not on the side of data warehousing technology but more on the side of legacy systems.

5.4.3 Educate Users About The Data

Educate users on the data from the business perspective..

A best practice that was often mentioned was to educate users on the data from the business perspective. In other words, user need to understand what the data means to the business, along with how and why it was collected. Alert users who can interpret the data wisely, noting inconsistencies in data quality, are of tremendous help to the data warehouse.

Folkerts at Midwest Card Services recommended educating users about the business value of information. "The data warehouse is a tool for them to use. It is strictly a tool. It provides information to make decisions."

Evans at Union Pacific stressed the importance of metadata in the hands of users. They publish all the metadata (for tables, views, columns…) for their data warehouse on our internal website. He said, "When users want to answer a question, we have a search engine, which we built ourselves, to find the right information."

5.4.4 Let Alerting Lead The Way

Alerting applications often lead to an important breakthrough in DW automation.

Another practice that is not obvious is to explore new data warehousing applications by letting alerting to lead the way. In other words, identify a critical business area that needs some BI attention. What are the conditions that indicate a problem? What information needs be reported to describe the problem? Who would be most interested in knowing about the problem? Then, implement a simple alert that is fired on that condition, reports that information, and is sent to that person. It could be a short email message.

Now here is the important part. Observe carefully what happens when the person receives the alert. Have the person make a log of the alerts and responses. After a week or a month, schedule an assessment to analyze the business activity. This discussion should surface a clear idea of a new BI application that appropriately responds to abnormal business situations.

As Hill of Sagetree noted:

> When we started with Ford, it [Inventory Management and Alerting System] was a simple alerting tool, such as a truck was late. As the relationship grew, the requirements were extended from alerting to analytical and then to a proactive operational capability. In other words, the system is measuring the net change in inventory on a nightly basis and then calculating, at the part number level, the days-on-hand inventory based on consumption and demand. Using the touch points in the supply chain, the flow of material is reprioritized for the next day.

5.5 Benefits Realized

This section summarizes the benefits that have been realized by the data warehousing projects. This is where the 'rubber meets the road'—applications of data warehousing technology that generate actual business benefits.

5.5.1 Improving Asset Management

As the data warehouse accumulates data on the assets of the enterprise, the opportunities increase to analyze the utilization of those assets and to detect ways of improving that utilization.

3M has stated that they were able to reduce indirect costs by $350M and will reduce inventory costs by more than $1B over the next few years.

5.5.2 Reducing Customer Support Costs

The combination of the data warehouse and customer self-service via the Web is a natural step, assuming that the issues of security and performance are addressed. If done correctly, the customer perceives greater service quality and has greater satisfaction. And more importantly, the company has been able to reduce their costs for customer support.

Union Pacific has several hundred customers who can query data in the data warehouse via the web, getting pricing information.

5.5.3 Auditing Billing Practices

As a data warehouse matures toward a Single Version of the Truth, the opportunities increase to compare data across disparate systems, particularly for situations that are abnormal or improper. Union Pacific used their data warehouse to audit billing practices, both incoming and outgoing. Evans summarized:

We have discovered situations in our accounts payables where we have paid for things that we should not have. In other situations, we did not bill for things that we should have. We have definitely recouped several millions of dollars per year by auditing our payables with EDW data.

5.5.4 Terminating Unprofitable Products

The ability to forecast the profitability of products and services can have major implications to the financial health of the company. Only a data warehouse with a unified version of reality about the enterprise can bring all the factors together to make the proper analysis of profitability.

At Midwest Card Services, they offer varied financial services packages, which can be complex, requiring the historical analysis from the data warehouse. Folkerts related a specific case:

> We never quite knew how a specific product or portfolio was performing [profitable], prior to implementing the data warehouse. We did not think we were getting ahead on one product, but we could not put our finger on the problem. After the data warehouse, … It was clear going forward based on forecasting, we were not going to make money on this product. We had very high confidence in the data and knew that it was the truth. We then made the decision to stop the issuance of this product. That was about a $600,000 decision. If we had not made that decision for 3-4 months, there would have been the impact.

5.5.5 Leveraging Legal and Regulatory Actions

Being on top with litigation cases was a benefit mentioned by Evans at Union Pacific, as follows:

> The key is that we have the information [in the data warehouse]. Sometimes the information is to our advantage, and sometime it is not. As a company, we have a good comfort level of knowing the truth! If it was our fault, then we should settle this one up front. If it was not, then we have the data to support our case.

5.5.6 Reducing Staff Requirements

Although a data warehouse is seldom justified by staff reductions, the focus on streamlining and automating complex business processes has resulted in staff reductions. Folkerts at Midwest Card Services noted a 23% staff reduction through process automation from the data warehouse.

5.5.7 Running The Business

The final benefit was observed at PING. When Crossland was asked about the value of their data warehouse, he simply said, “It runs our business!” Because PING pursued a single-database architecture where transactional and informational data shared the same Teradata platform, a failure implied that PING could not accept new orders for golf clubs and ship orders with finished ones.

6. Conclusions

Companies are implementing active data warehousing successfully and reaping benefits from these implementations. Active data warehousing is, however, in its early stages. Successful examples (such as described in this paper) are helping to identify the best business opportunities for active data warehousing. The barriers are often not technological but organizational because, to realize the benefits of active data warehousing, we may have to change the way we do business.

Being an intelligent enterprise is not an option; it is a requirement.

Over time, active data warehousing will be accepted as common practice because the demands of our global economy require ever better ways of doing business. In our world, change may be the only constant. But in our business environment, continuous improvement may be the only viable goal. Being an intelligent enterprise is not an option; it is a requirement.

Achieving a single version of the truth within the data warehouse still requires a leap of faith for business people. It feels too much like technology for technology's sake. Amass all this data, and users will come. However, there are emerging many solid examples of tangible benefits from correlating data across functions, divisions, etc. CRM with its 360-view of the customer has been a huge contributor. Achieving a single version of the truth requires some new thinking, but the benefits are there. Build on incremental successes, educating slowly but steadily executives and workers alike.

The good news is work automation and actionable granularity. Work automation is the low-lying fruit for new applications based on analytics. Actionable granularity is evolving steadily as data warehouse usage expands and users asked more questions.

…on the verge of a new generation of enterprise systems.

I believe that we (as the BI profession) are on the verge of a new generation of enterprise systems, one in which BI takes a center stage rather than an extract-and-run appendage. There is a convergence of BI with transactional systems, with enterprise application integration (EAI), with eBusiness architectures, and the like. It may be that future enterprise systems will evolve with advances in analytics driving improvements in the business processes.

7. Appendix: Interview Questions

This appendix lists the questions used in the structured telephone interviews. Not all questions were asked of all interviewees, given the 45-minute time limit.

1. Company Background (prior to interview)
    - 1.1. Company Name:
    - 1.2. Company Profile:
    - 1.3. Industry:
    - 1.4. Geographic Location:
    - 1.5. Major Merger or Acquisitions:
    - 1.6. Product/Services:
    - 1.7. Organization Structure:
    - 1.8. Major Locations:
    - 1.9. Financial/Stock Performance:
    - 1.10. Trade Press Buzz:
    - 1.11. Relevant Press Releases:
    - 1.12. Competitors:
2. Interview Introduction
    - 2.1. Interviewee Name:
    - 2.2. Interviewee Title and Job Responsibilities:
    - 2.3. Interviewee Contact Info:
    - 2.4. >Start interview here. Explain study objective and methods.
3. System Characteristics
    - 3.1. How do you refer to your system or project?
    - 3.2. What are the key business tasks supported?
    - 3.3. Who are the key business users supported?
    - 3.4. Describe the general architecture
        - 3.4.1. Products utilized
        - 3.4.2. General sizing of database, flows, etc.
        - 3.4.3. Diagram publicly available?
4. Business Drivers
    - 4.1. What were your key business objectives?
    - 4.2. What were your expected benefits?
        - 4.2.1. In their terms:
        - 4.2.2. Responsive
        - 4.2.3. Pervasive
        - 4.2.4. Global
        - 4.2.5. Integrated
        - 4.2.6. Other…
    - 4.3. Who were the main champions/sponsors?
5. Implementation Experiences
    - 5.1. How long did it take to implement the first production version?
    - 5.2. Development organization
        - 5.2.1. How many developers worked on the project?
        - 5.2.2. Position titles and descriptions
        - 5.2.3. Skill sets available? Needed?
        - 5.2.4. How much training/education was required?
        - 5.2.5. Interface with business users
        - 5.2.6. Did you use outside services during implementation?
    - 5.3. Estimates of total resources required
    - 5.4. What difficulties did you experience during implementation?
6. Production Experiences
    - 6.1. When did the first version go into production?
    - 6.2. History of subsequent versions:
    - 6.3. What surprises (good or bad) occurred in the early months?
7. User Experiences
    - 7.1. What are the various groups that use the system?

- 7.2. For each major group (deal with only 1-3 groups)
  - 7.2.1. User profile: responsibilities, dept, mgt level, skills…
  - 7.2.2. Numbers: authorized, active, concurrent
  - 7.2.3. What is done with the data provided?
  - 7.2.4. What actions/decisions are taken with the data?
  - 7.2.5. Describe the decision sequence or workflow.
  - 7.2.6. What changes occurred in business process?
- 7.3. How did the skills and needs of users evolve?

8. ADW Characteristics
   - 8.1. Single version of reality
   - 8.2. Action distance
   - 8.3. Just-in-time data freshness
   - 8.4. Event-base triggers
   - 8.5. Workload mixtures
   - 8.6. Cross-application synergism
9. Business Benefits
   - 9.1. From your perspective, was the project a success?
   - 9.2. What was the biggest benefit to your company?
   - 9.3. If the system went down for three hours, what kind of 'pain' would the company feel?
   - 9.4. What are the major success factors?
   - 9.5. What are the major risk factors?
   - 9.6. Did you perform a ROI analysis on the project?
     - 9.6.1. What was the projected ROI during the first year?
     - 9.6.2. Has this projected ROI been realized?
   - 9.7. How did the system provide cost-reduction savings?
   - 9.8. How did the system provide value-added benefits?
     - 9.8.1. Specific revenue-generation benefits
   - 9.9. Is there a reasonable Total Cost of Ownership in…
     - 9.9.1. Hardware and special software required
     - 9.9.2. Skills and people required
     - 9.9.3. Implementation and maintenance effort
     - 9.9.4. Other…
10. Lessons Learned
    - 10.1. If a colleague asked for advice on a similar project, what would you suggest?
    - 10.2. If you did this project again, what would you do differently?
11. System Evolution
    - 11.1. How you enhancing the system?
    - 11.2. Are you expanding usage to other groups? Who? Why? How?
    - 11.3. Do you expect expanded functionality?
      - 11.3.1. Applications? Subject areas?
    - 11.4. Do you expect usage to increase over the coming year?
      - 11.4.1. Data volume?
      - 11.4.2. Query complexity?
12. Interview Conclusion
    - 12.1. >Thank interviewee; offer advice and assistance
    - 12.2. Are you willing to clarify your responses via email?

8. References

[1] Richard Hackathorn, Value Proposition for the Active Data Warehouse: Achieving the Intelligent Enterprise, February 2002.

[2] James Brian Quinn, Intelligent Enterprise: A Knowledge and Service Based Paradigm for Industry. Free Press, July 1992. ISBN 0-02-925615-1.

[3] Michael E. Porter, Competitive Advantage: Creating and Sustaining Superior Performance, Free Press, 1985. ISBN 0-02-925090-0.

[4] Realware Awards, 2002. Award Description for Best Business Performance Analysis and Management Application.
http://www.realwareawards.com/2002/2002winners.shtml

[5] Realware Awards, 1998. Award Description for Knowledge Management Application.
http://www.realwareawards.com/1998/sterling.shtml

[6] Realware Awards, 2002. Award Description for Best Business Performance Analysis and Management Application.
http://www.realwareawards.com/2002/2002winners.shtml

[7] It is not known exactly how the $100M amount in 1998 compares with the $350M amount in 2002. However, it appears that the GEDW significantly exceeded their expectations.

[8] Interview with Al Messerli on July 29, 2002.

[9] Website of Delta Technology at http://www.delta.com/.

[10] Kim Moutsos, The Trouble With Data Marts, Teradata Magazine, Spring 2000.
http://www.teradatamagazine.com/articles/spring_2000/industry.html

[11] Delta Airlines press release on April 25, 2002.

[12] Interview with Wayne Hyde on August 12, 2002.

[13] Interview with Jerry Hill on August 7, 2002.

[14] Website of Midwest Card Services at http://www.midwestcardservices.com/

[15] Interview with John Folkerts on August 9, 2002.

[16] Teradata Partners Conference, October 2001. Grow Your Business with Data Warehousing. Presentation by John Folkerts

[17] Interview with Kent Crossland on July 17, 2002.

[18] Justin Langseth, Agility Training, Intelligent Enterprise, May 28, 2002.

[19] MicroStrategy, "Strategy.com Delivers Timely, Relevant Information to More Than 500,000 Subscribers," press release, December 20, 2000. It describes Strategy.com at its zenith.

[20] Interview with Justin Langseth on July 30, 2002.

[21] Justin Langseth, Agility Training, Intelligent Enterprise, May 28, 2002.

[22] Interview with Paul Evans on August 15, 2002.

[23] Teradata Partners Conference, October 2001. Data Warehouse: Not Only a Destination, but a Way to Update the Source Data. Presentation by Paul Evans and Rick Royers.

Appreciation is given to Teradata, Division of NCR, for its support to pursue this investigation into Active Data Warehousing and to the many people that shared their experiences about these exciting developments.

Dr. Richard Hackathorn is president and founder of **Bolder Technology, Inc.** (BTI) in Boulder, Colorado. BTI is an eleven-year old consulting and education firm specializing in the Information Technology industry.

Richard has over thirty years of experience in the IT industry as a well-known technology innovator and international educator. He has pioneered many innovations in database management, decision support, client-server computing, database connectivity, data warehousing, and web farming. He founded MicroDecisionware Inc. (MDI), an early vendor of database connectivity products that was acquired by Sybase in 1994.

Richard has published numerous articles in trade and academic publications, presented regularly at leading trade conferences, and conducted professional seminars in eighteen countries. He writes for DM Review with a monthly column entitled The BI Watch that reaches over 75,000 IT professionals. He has written three professional texts, entitled Enterprise Database Connectivity, Using the Data Warehouse (with W.H. Inmon), and Web Farming for the Data Warehouse.

For over a decade, Richard was a professor at the Wharton School of the University of Pennsylvania and at the University of Colorado. He received his B.S. degree in Information Science from the California Institute of Technology and his M.S. and Ph.D. degrees in Information Systems from the University of California, Irvine.

He can be contacted at richardh@bolder.com.