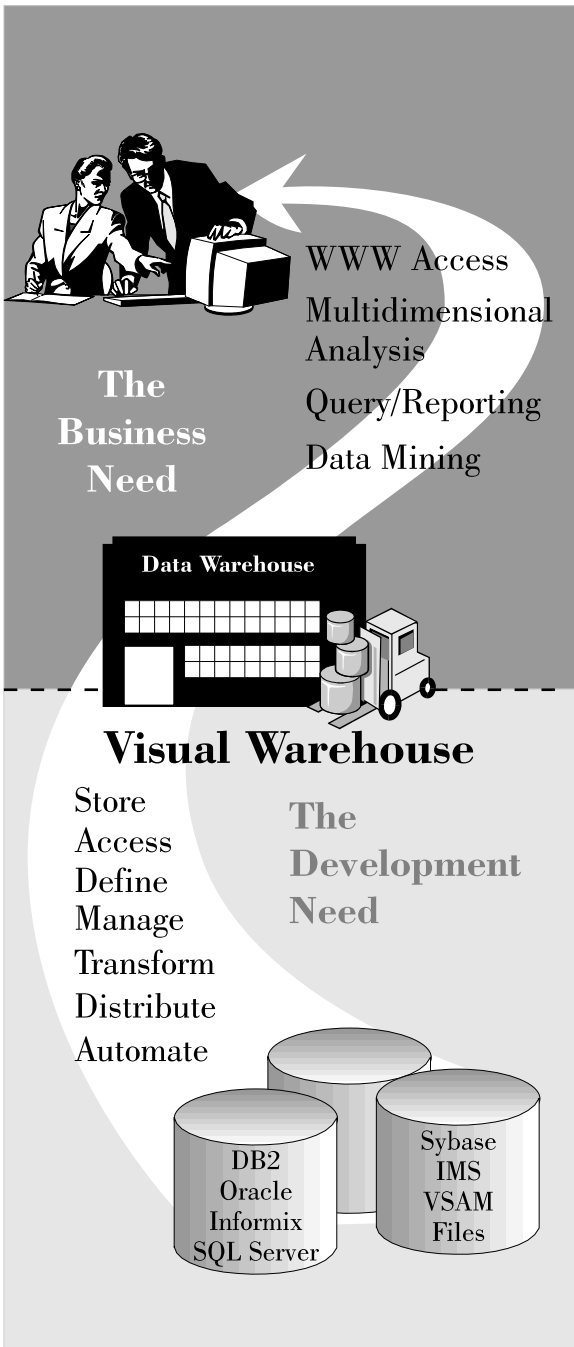


A Practical Guide to Getting Started with Data Warehousing

ATG's Data Warehousing
Technology Guide Series

This guide has been sponsored by





Visual Warehouse for Windows NT

The IBM Visual Warehouse is a comprehensive product that has integrated many of the functions and processes required to define, build, manage, monitor and maintain a data warehouse. This enables the rapid development of data warehousing to meet the informational needs for decision making in a cost effective manner.

- **Low Cost**

- **Self Contained**

- **Expandable**

- **Quick Start**

Visit ATG's Web Site
to read, download, and print
all the Technology Guides
in this series.

www.techguide.com

“The significant problems we face cannot be solved
by the same level of thinking that created them.”

Albert Einstein

Table of Contents

Considerations for a Successful Warehousing Project	4
Overview	4
Aligning Technology with Business Objectives....	6
Business Criteria	6
Process Criteria	14
Technology Criteria	21
Implementation Tactics	28
Planning the Project	28
Architecting the Solution	32
Designing the System.....	32
Developing the System.....	33
Piloting the Implementation	33
Deploying the System	34
Training Users.....	34
Managing Warehouse Operations	35
Summary.....	36
APPENDIX A.....	36
Elements of a Warehouse	36
Case Study	41
Glossary of Terms.....	49

About the Editors...

Thomas Flanagan and Elias Safdie are co-founders of The Cordis Group, Inc. a consulting firm specializing in helping management use information technology to meet their business objectives. Both Tom Flanagan and Elias Safdie have nearly fifty years of combined experience advising both vendors and system users in the strategic and tactical uses of information technology. You may contact the authors at elias@pacificnet.com or tomf@pacificnet.com (or telephone at 508-256-4974 or 508-649-3388 for fax).

This book is the property of The Applied Technologies Group and is made available upon these terms and conditions. The Applied Technologies Group reserves all rights herein. Reproduction in whole or in part of this book is only permitted with the written consent of The Applied Technologies Group. This report shall be treated at all times as a proprietary document for internal use only. This book may not be duplicated in any way, except in the form of brief excerpts or quotations for the purpose of review. In addition, the information contained herein may not be duplicated in other books, databases or any other medium. Making copies of this book, or any portion for any purpose other than your own, is a violation of United States Copyright Laws. The information contained in this report is believed to be reliable but cannot be guaranteed to be complete or correct.

Copyright © 1997 by The Applied Technologies Group, One Apple Hill, Suite 216, Natick, MA 01760, Tel: (508) 651-1155, Fax: (508) 651-1171
E-mail: info@techguide.com Web Site: <http://www.techguide.com>

NOTE TO THE READER: This guide is oriented to both business managers exploring the potential of data warehousing for their organizations, and technical implementors who are seeking practical advice on implementing data warehouse systems. To those unfamiliar with data warehousing terms we would suggest beginning with the Appendix which reviews the major components and nomenclature. Business managers may wish to focus on the Criteria section (specifically the Business and Process Criteria) and the first few sections of Implementation Tactics. Readers looking for technical information may wish to focus on Technology Criteria and the Implementation section. All readers may wish to look at the Case Study which highlights how three different companies use IBM's Visual Warehouse to address Warehouse requirements.

Considerations for a Successful Warehousing Project

Overview

Enterprises today, both nationally and globally, are perpetually seeking competitive advantages. It has become an incontrovertible axiom that information is the key to determining how to gain such a competitive advantage. The problem today is how to deal with the mountains of raw data which our ever more efficient information systems are collecting, massaging, processing, deriving, and disseminating. We literally are in the midst of a volcanic eruption of data. Somewhere in this explosion of data are the clues management needs to define their strategic positioning in the market so as to maximize their competitive stance.

Into this picture, technology has inserted the concept of Data Warehousing as one alternative to

copied with the well-known "information overload" described above. Data Warehouses exist to help management and decision makers transform raw data into information: to help management identify key trends; to help the enterprise foresee predictable events and act in anticipation of those events; to help management understand the entire picture of what happened and develop a good systemic understanding of the events, thus allowing focused reactions to those events, such as redefining and reengineering business processes to take advantage of that understanding.

A prerequisite to enabling management to accomplish all of the above is clearly the fact that data management access has to be relevant, it has to be current, and it has to be available in a format which is usable by the requester. To date, too much emphasis has been placed on the raw technology which embodies the concept of Data Warehousing, and not enough on the underlying and concomitant strategy, planning, processes, and services which develop, maintain, and use the Data Warehouse technology. Experience has shown us that in projects where the technology has been perceived as a failure, the problem does not usually lie with the technology itself, but rather with the way in which the technology was applied: it is applied to the wrong problem, at the wrong scale, with insufficient training, with insufficient planning, with little or no thought to how users need to access the data, etc.

Data Warehousing technology can benefit enterprises at a number of different levels, from departmental systems running on commodity platforms with operating systems such as Microsoft NT™ and database servers such as DB2™ for Windows NT, up to the enterprise level systems running on scaleable MPP architectures such as IBM's SP/2™ with DB2 Universal Database. Therefore, thought needs to be given to what the initial strategy needs to be to ensure that an organization truly benefits from the technology. Moreover, we all know we are living in a very dynamic

and fast changing business environment, where the only certainty is uncertainty and change. Therefore, the strategy must deploy Warehouses such that they can grow and adapt to the changes we know are coming, even though we don't know today what those changes are. This calls for prudent planning in developing the strategy to select architectures which can react flexibly to changes in market dynamics, organizational restructurings, economic fluctuations, etc.

Aligning Technology with Business Objectives

Data Warehousing is a nascent technology, and therefore by definition is maturing as we speak. There is no Rosetta Stone which will tell any one organization what works and what doesn't. Consequently, experience and evolution are the best overall planning principles which can be deployed by management today. There are a number of criteria which should be kept in mind as strategies are developed. The criteria can be grouped into three categories: Business Criteria, Process Criteria, and Technology Criteria.

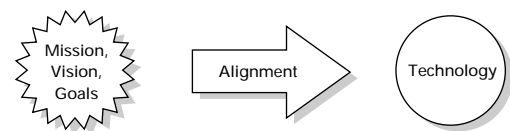
Business Criteria

The first set of criteria have to do with the business problem at hand and what demands are made on the project by the business dynamics encountered:

Critical Success Factors: Why are you doing this?

No technology project will ever succeed if it is not properly aligned with the business Mission, Vision, and Goals. Therefore it is vitally important to understand such elements of the strategy as:

- What is the problem at hand? Is it a problem related to cycle time, customer satisfaction, more cost-effective decision making, better business intelligence, or a general lack of information with which to make decisions on any of the above?
- Which of the departmental or enterprise goals and responsibilities are directly related to the problem at hand?
- What are the Critical Success Factors—those things which must be done well to solve the problem?
- Which of the organizational components of the enterprise is best positioned to solve the problem?
- Which audience within this organization will best use this technology and how: executives, financial analysts, scientists, engineers, clerical, and administrative users, line managers, others and why do they need this (who will it benefit)?
- How can this technology be used to solve the problem? This is where the alignment of the technology to the problem will occur.

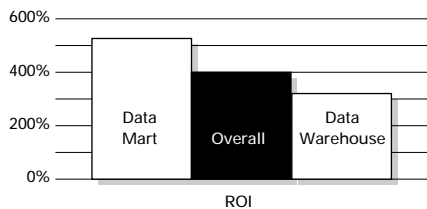


Quantifying Benefits

Once the benefits of the technology have been aligned to the business objectives, they should then be quantified. The reason for this is that management must be able to answer the question: “How will we know if this project is successful?” The answer must fit the form: “This project will be successful if it allows us to achieve the following goals...”.

In many organizations, quantifying benefits takes the form of a financial analysis, specifically a Return on Investment (ROI) analysis. A recent study by International Data Corporation (IDC), co-sponsored by IBM, showed that Data Warehousing can provide significant and impressive ROI numbers. The study, which included 62 participants, demonstrated that the overall ROI on warehouse projects was 401% with payback periods of two to three years.

What was interesting about the study, however was that the smaller, more focused departmental implementations, sometimes known as Data Marts, had a whopping 533% ROI, while the larger, enterprise efforts showed a smaller, although still impressive 322% ROI. The implications of this study are clear and were summarized in a June 1996 *Computerworld* article by Mr. John Margonis, manager of underwriting systems for Blue Cross/Blue Shield of Oregon: “By implementing data marts, you get a quicker return on investment and you have a system up and running in two or three months”.



The IDC study identified three kinds of benefits in the use of Data Warehouses:

- Cost avoidance benefits. These benefits were the ability *not* to spend money which is presently spent on generating endless reports to end users. This included the resources expended by IT to generate answers on ad hoc queries. In many ways, Data Warehousing represents a liberation for IT by

providing users with the tools they have needed over the years to generate their own reports. By allowing the users to do so, one eliminates the often endless loops of the user requesting a report from IT, IT delivering the report, the user either not approving the report due to some miscommunication or changing their minds after seeing what they asked for, etc.

- Efficiency gains from increased productivity among end-user professionals who gather and analyze data. The analyst who must stop an analysis to get information and who has to ask someone else to get the information loses efficiency in two ways. The first is in the “loop” described above, where there might be several iterations of request and response between the analyst and IT before the analyst is satisfied with the results of the request. The second is in the interruption of the analysis, and the inefficiencies associated with recovering the thought processes which were underway when the analysis was suspended.
- Warehouse dependent savings due to decisions based on analysis which could only come from data in a warehouse. This is a “quality of decision” issue which comes from the fact that certain data associations may not exist in any one operational system and therefore are not available to the analyst. By building those requisite associations and relations in the Warehouse, a situation where the whole is greater than the sum of its parts occurs and those relationships allow the analyst to do their work better, and become more effective. This is not a case of people making bad decisions prior to the advent of the system, this is about giving people better tools to empower them to do better work.

Not all benefit quantifiers are in terms of ROI. A recent article described how firms often eschew formal ROI analysis because they consider the data warehouse a “strategic investment”. In this case, these organizations were convinced prima facia that the benefits would be worth the costs. A cautionary word: Make sure that there is an understanding of the projected costs before starting the project if there is no formal financial measurement technique such as ROI.

Regardless of the culture of the organization, whether it accepts “soft” benefits in its approval process or not, be sure to quantify the anticipated benefits in business terms. Without this, the organization will have no metrics to determine whether or not the project is successful.

Perspectives on data marts and warehouses

The IDC finding would lead one to believe that since the data marts delivered the highest ROI, a viable strategy could be to allow each department to implement their own data marts independently. The problem with this thinking is that it ignores the issue of cross functional analysis. What happens when Finance and Marketing each develops their own data marts independently of each other. Let us further assume that each keeps track of sales, but defines them differently. For example, marketing defines a sale as a booking while finance defines it as a payment. What are some of the potential ramifications of this situation? A primary consequence is that analyses of the same business dynamic—sales—may result in two departments having significantly different views and analyses. More importantly, what if finance needs something from the marketing Data Mart, or vice versa. Without a unifying design or data standards, the analysis of data which spans two or more organizational Data Marts will not be possible using this strategy. It is therefore important in developing a Warehousing strategy to understand the implications of an independent Data Mart strategy

and factor these risks into its formulation. A more detailed discussion of independent vs. dependent Data Marts is presented in the Process Criteria.

“The heat is on...to get warehouses up and running fast” [but] “The biggest benefit comes down the road, when you can support 20 different decision support applications with the same architecture”—Gartner Group. This quote from the Gartner Group highlights the need to develop an overall strategy and architecture for the organization prior to or at least in parallel with the development of the first Warehouse project. At the beginning of these projects, everyone is in a hurry to reap the immediate benefits of the systems and becomes impatient with the planning process which is sometimes seen to “slow things down”.

However, experience tells us the payoffs will be that much greater, and subsequent improvements will come that much sooner if a sufficient amount of time is invested at the front end of the project in defining an overall enterprise architecture into which the Data Marts can connect.

Understanding product integration issues

The relative youth of the Data Warehousing industry leads to a proliferation of disconnected offerings in the marketplace. Many small companies have joined the fray with products targeted at one or two elements of the Data Warehousing architecture. As a result, there are very few offerings in the market which answer all of the needs of a potential end user. This leads to a risk regarding how well different packages will integrate, what common platforms are supported, and so on. IBM has addressed these product integration issues with its Visual Warehouse solution.

Dealing with cultural issues

Data Warehousing is about pooling resources (data), which implies sharing, which in turn can imply a loss of control, a concept sometimes inimical to

many data owners. This kind of organizational provincialism can sometimes throw up impediments to a Warehousing project and must be dealt with in the early phases of the project.

Process considerations

Technology is only useful to the extent that it supports our ability to carry out our assignments and achieve our corporate goals. Therefore, the introduction of any new technology must be aligned with the business processes which it is intended to support. Too often, technology is developed independently of any process considerations, many times with catastrophic results. A successful Warehousing strategy will harmoniously integrate the Warehouse with redesigned business processes which truly can take advantage of the benefits the technology has to offer.

Choosing services

Very few, if any, IT organizations have the requisite combination of skills and resources required to perform all of the technology, planning, and implementation tasks required for successful Data Warehouse projects. This is not intended as an affront to IT organizations, but rather a simple observation that it will take a broad spectrum of talent which crosses many disciplines to make this work. The chances that a single IT organization will have all of these talents available for this project at the same time are minimal. Therefore, at some point in time, many organizations will need to locate a partner to consult in the planning for the Warehouse and then eventually assist in the implementation of the system. The partner selected must be able to operate within the constraints enumerated in the organization's Warehousing strategy, including the methodology chosen, implementation style chosen (see Process Criteria—Methods), and so on. Choosing the wrong partner, for example one who has Data Ware-

housing experience but not Data Mart experience, or one who does not have experience across the entire spectrum of products and services, can increase the risks associated with these systems. IBM offers specific services in conjunction with its Visual Warehouse solution as well as custom services for any scope of warehouse implementation.

The state of standards.

One of the inherent risks of a new technology is the lack of standards, and Data Warehousing is no exception. There are countless examples of competing technologies which resolved themselves into one standard, and most of the time the resolution creates winners and losers. Eight track tape owners were losers in the technology battle with cassettes, Betamax owners lost against VHS in video recording, CPM lost to MS/DOS, the list goes on. Obviously one risk mitigation strategy in this arena is to align the project with a “big player” in the industry—one which will have an influence in determining the winning standards. Another innovative mechanism which was used in a major bank IT shop was to include in the cost/benefit analysis an estimated cost to “bury” the existing system and replace it with a new one in case the wrong decision was made. If the project still made sense after including this cost, then the bank would go ahead with it in spite of the standards risk.

Lack of attention to training

For many years, exposure to the world of the database was limited to the inhabitants of the glass house—the IT department. Consequently, many end users are not familiar with the concepts behind navigating a data schema or unraveling the mysteries of joining tables via keys to get queries answered. A liberation of sorts will ensue from allowing users access to their own data. However, the users must be prepared for life in this liberation and must be ready to accept

the responsibilities of runaway queries, etc. Therefore it is incumbent on management to make sure that adequate training is provided to users to allow them to use the system without (a) getting so frustrated that they give up or worse, poison the project by maligning it to others, (b) bring the system to its knees by constructing queries which run forever, or (c) base an analysis on faulty data because the user was not familiar enough with the system to understand what he or she was asking for.

Process Criteria

A number of elements of the Warehousing strategy have to do with processes. Processes by which the strategy is implemented, and processes which are supported by the overall strategy.

Scope of Effort: How big should it be?

Many large warehouse projects have failed because of an inability of the organization to handle the size and scope of the project. It is tempting to think of a single repository where all of the enterprise's data problems can be solved in one fell swoop. And if the organization can indeed come up with an integrated data model and solve all of the issues associated with such an architecture, the benefits are indeed significant. However, this is sometimes not realistic and not necessary for the problem at hand. Industry studies have estimated the average size of a large warehouse project to be nearly three million dollars with a time to completion measured in years. No doubt, some business problems require the integration of data from many systems and will require a global strategy. Many others will not, and a simpler strategy, starting with data marts will likely be less risky.

This question of project scope should be readily answered from the exercise described earlier on aligning the technology to the corporate Vision and

Mission. Once the questions relating to who needs the technology and what problems are being addressed are answered, the scope should be relatively straightforward to determine, which should allow management to allocate appropriate resources to manage it.

A strategy which entails an enterprise wide scope has certain implications associated with it which should be understood. First and foremost it requires integration and cooperation among multiple organizational elements. Many issues will arise regarding different definitions of similar or identical terms, competing objectives and agendas, data parochialism, and an unwillingness to give up control. This is a situation which can be difficult to manage and successfully navigate. Oftentimes, change management is necessarily intertwined with this exercise, since the organization will have to wrestle with interdepartmental issues as described above. Management must assess whether or not the organization is ready to deal with these kinds of issues, or whether they might not better wait until a more appropriate time.

A Departmental, or Data Mart approach is by definition smaller in scope, more focused in its outcome, quicker to achieve, less costly. However, there is a risk to developing Data Marts in a vacuum, as described in more detail in the "method" section below. Ideally there is a need to think globally about future integration with other departmental applications and Data Marts to avoid developing "Data Islands".

Data Marts have their place and present a strong business case for starting with such an implementation, but if an organization determines that an enterprise warehouse is the appropriate strategy than there are many successful models to emulate.

Implementation Approach Options

Deciding whether a Data Warehouse or Data Mart is right for the enterprise is an appropriate beginning. It

must, however, be followed by a decision on whether to buy an integrated package from a single vendor, engage a systems integrator to bring together a collection of “best of breed” products, or have the enterprise’s IT department create a “best of breed” solution.

The obvious advantage to dealing with an organization which can offer a complete solution is, of course, faster realization of business goals, usually at a reduced cost, and with a good degree of certainty that the ultimate solution will work (lower technological risk). However, the enterprise may have to adapt business processes to fit the specific characteristics of the “one-stop-shop” technology.

The “best of breed” concept is certainly not new and has as its foundation the tenet that the enterprise will be better off if it can somehow bring together the “best” extraction tools, databases, SMP/MPP hardware, disk drives, analysis tools, network, etc. and get them to function as a unified system. Aside from the difficulty and religious wars which accompany the attempts to define “best”, the price paid for the anticipated exceptional performance is primarily the pain associated with integrating the disparate components. Different vendors value different architectures and functionality characteristics, and the “best” extraction tools, for example, may not integrate well with the “best” metadata repository, and so on. Many implementations of components “should” integrate, are advertised as “compatible”, but require extensive work to get all details to work for an organization’s application. This kind of scenario is especially significant in technologies which are young, and in which standards are not yet entrenched. Most integrators find that all projects require compromises to achieve integration, although some level of custom fit with the enterprise is generally achieved.

There is a second, more subtle price associated with this custom fit, and that is what to do in the future about upgrading to later releases. Assume, for example, that an integrator combines extraction tool A with metadata repository B, database C and analysis tools D. Assume further that all of the products, (A,B,C,D) are at release 1.0. The integrator finishes the job, and the customer is satisfied, and some time later extraction tool A moves to release 2.0, three months from then the analysis tool moves to release 2.0, and so on, each release bringing with it features and functions the organization would like to incorporate. The challenge is: how to do that? Are these independent release updates compatible with each other? Will the resultant system be backwards compatible with the original system? If the organization continues to rely on the integrator to maintain a system at the latest “best of breed” status, this is tantamount to a full employment act for the integrator, and certainly of dubious value to the customer. On the other hand, a customer who has purchased an integrated package from a single vendor can put the onus of compatibility and upgrading on the shoulders of the vendor, and upgrade the warehouse to the next release by buying a single upgrade.

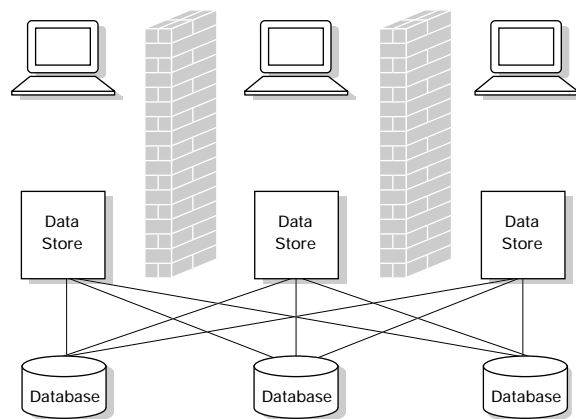
The third option, that of using the enterprise’s own IT shop to integrate the package is probably not a viable option except for very large and sophisticated organizations. Few IT shops have the expertise to integrate “best of breed” components, and fewer still have the available resources. Remember that one of the reasons to build a Warehouse is because the IT department is buried with requests from users who are already frustrated with the inability to get answers to their mission-critical planning questions. The idea of taking an entire team of IT specialists out of the front lines for months (or years for a global warehouse), and simultaneously maintaining or reducing the backlog could be a formula for disaster.

Top-Down and Bottom-Up Approaches

How should an organization approach the design and construction of the Warehouse? Should it go for a top-down technique, setting up an enterprise-wide architecture and then constructing Warehouses or Data Marts which conform to that architecture? Or perhaps take a “bottom-up” philosophy, starting right in with highly focused and targeted Data Mart projects aimed at specific critical areas of the business? The issues here have to do with deciding between addressing short term tactical requirements to help individual departments and long term strategic planning issues regarding data architectures which have to cut across political and organizational boundaries.

The Top-Down approach can yield the best long term results, but it also invokes the most angst within an organization. It is very difficult, expensive, and time consuming to achieve consensus on a single, consistent, accepted, and valid view of the business, the data it needs, etc. In the prior example of the Marketing and Finance departments, two different operational definitions for “sales” were developed. Which of the definitions will be used in the Warehouse, or will there be two terms which must now be defined? If so, they can’t both be called “sales”, and the metadata had better be clear as to what the final determination became. It is clear that many more change management issues will have to be dealt with in this approach, as organizations grapple with the sins of the past in not having promulgated an accepted data standardization program, etc.

The Bottom-Up approach favors the use of smaller, more focused applications of Warehouses which can avoid the pitfalls of the Top-Down by simply limiting the extent of the implementation. This approach also exhibits simpler data archaeology problems: there are usually limited data sets, limited user views, a good understanding of the data needs and how they relate to the business problem. In its purest form, this

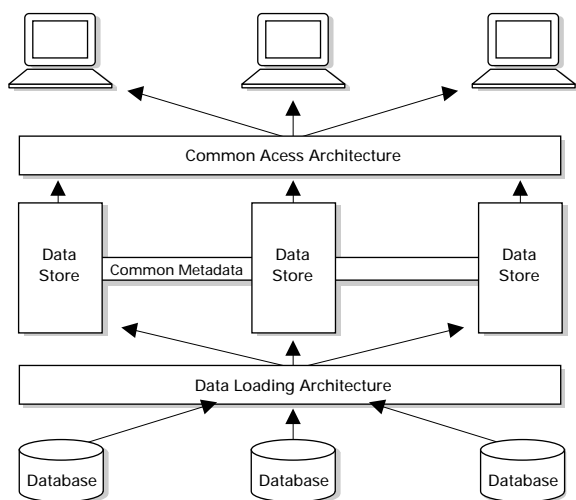


approach trades the near term pain of dealing with data standardization issues for longer term pain of dealing with cross-organization operational issues. In this approach, each department is responsible for extracting whatever data they need, defining their own metadata, and using their own private Warehouses for decision support at the departmental level. Two obvious problems arise: (1) this architecture is difficult to evolve to an enterprise view; and (2) the lack of data standardization prevents analysts in one organization from accessing information from another organization’s Warehouse which might be of use to them in their analysis. This is demonstrated by the brick walls inserted between departmental systems in the figure above, preventing interactions. Each mart can create confusing, overlapping and contradicting views of the business, like the proverbial six blind men trying to describe an elephant, each only being able to relate to the elephant according to the portion of the animal they were feeling. What is a customer? a product? a sale?

This approach works if the organization has a business problem with a single focus and the data to solve that problem exists in only a few places, with no political ownership issues.

Some organizations use a hybrid approach to gain the speed and cost advantages of the highly focused departmental approach, yet at the same time making sure that the implementation is consistent with the overall goals of the organization.

This approach uses the principles developed in Rapid Application Development (RAD) methodologies and intentionally delivers iterations of the departmental Warehouse, attempting at each iteration to come closer to an overall enterprise data model and data architecture.



This allows the organization to take advantage of the speed and cost savings of the smaller approach but at the same time mitigate, and eventually overcome the lack of integration and “islands of automation” problems inherent in this approach. Other organizations are using the smaller implementation as a proof of concept or prototype/pilot installation, proving the benefits of the technology on a smaller scale, with smaller risk, and then scaling the solution into a more global implementation, either by building more inte-

grated departmental Warehouses (see figure, left) or by moving to a full global enterprise Data Warehouse implementation. Any of these approaches can work for your organization. What is important is that you are aware of the issues associated of each and actively mitigate the cons.

Technology Criteria

The technology dimension will of course play a major role in the enterprise strategy. Different strategies will require different technological characteristics and features. Just as the technology must align itself with the business mission, the strategy must also consider the technology.

Scalability

Scalability refers to the ability of a system to increase in capacity as users demand more, as data stores grow, as more users are added to the system, and as more applications are developed against the Warehouse.

One of the challenges associated with introducing new technologies and new applications is that of determining the actual system load after implementation. JAD (Joint Application Development) sessions attempt to mitigate the risk by attempting to produce a picture of user requirements, but the truth is that users who have never had the opportunity to work with a new technology don't really know what to ask for. Further, they don't know ahead of time what kinds of demands they will make on the system until they actually sit down and start working with it. Therefore, as the users become familiar with the query capabilities and the navigational issues, their own success will cause them to demand more from the system as word spreads and more users exploit the features of the system. In time, users will become more sophisticated and begin exploring with ideas of data mining and visualization

as their analyses become more complex. All of these are factors which demand scalability in a system.

Manageability

Data Warehouses require the development and implementation of new processes, tools, and work systems to manage the extraction/transformation of operational data, the administration of users (adding, deleting users, changing access rights) and so forth. In the course of deploying a Warehouse, a number of operational management decisions must be made:

- How frequently should data loads and transformations be made? Should they be daily, weekly, monthly, quarterly?
- How much data needs to come in on each load, and how long will it take for the system to recompute all the indexes, metadata updates, and other administrative details which must be undertaken?
- How will the system deal with backups/restores and what should the process be to administer a disaster recovery plan?
- If the system is to be paid for by usage chargebacks, is there a mechanism for the systems administrator to keep track of usage by account number or password, and are there reports available to facilitate this feature?

Performance

How well the system performs will be the ultimate arbiter in the success or failure of the project. The intent of the Warehouse is to help people do their jobs more effectively and efficiently. If the response time is not adequate, users will not use the system, the enterprise will not derive any benefit from the expenditure, and the project may wind up with a negative ROI. Therefore, the technology dimensions of performance must be thoroughly considered in developing a strategy.

Over the years, there have been a number of studies to determine the limits of human patience in dealing with computer response times. In general, users want to see something back in a timeframe which does not interrupt their thought processes. Some have said that two to five seconds is a good target response time, but when we are dealing with such gargantuan database sizes, it is difficult to conceive of doing a table scan on a multi-million row table in that timeframe, which leads us to the second human factors point. That is the fact that the amount of time a user will wait is proportional to the perceived difficulty of the procedure requested. Therefore, if a user knows they have entered a particularly nasty query, they will be more tolerant of delay. The best advice on this subject is to work with the user community in establishing meaningful metrics and working cooperatively to set and meet expectations on both sides.

There are several areas which directly impact the performance of a Warehouse system: the hardware architecture and the database architecture. The hardware dimension may involve the use of Symmetric Multi Processing (SMP) and Massively Parallel Processing (MPP) architectures. The top end of the MPP capacity does outstrip the top end capabilities of SMP. However, the applications where this kind of performance is necessary are few and far between. SMP architectures have been shown to scale nicely to as many as 64 processors, and most small departmental applications are well served with configurations as small as four processors. For this reason, many industry analysts are predicting that SMP will be the hands down winner in the Warehousing market. This does not mean that every warehouse implementation should be on a parallel technology of one sort or another, many applications can perform satisfactorily on non-parallel systems.

Many database vendors have followed the hardware architecture by introducing parallelism into the data-

base functions. Parallel loads, joins, queries, scans, etc. are available and are of tremendous value. IBM offers parallel database capabilities in DB2.

Flexibility

If there is one thing which distinguishes market conditions today, it is the pace of change. In fact, as pointed out in the introduction, one of the drivers leading enterprises to consider Warehousing technology is the need to keep up with that change. Deployment of a Warehouse will not alter the pace of change or the need to keep up with the change, which means the Warehouse technology itself must be flexible, allowing for rapid responses to changing conditions. The Warehouse must be adaptable to changes in the enterprise, such as reorganizations, mergers, and acquisitions. It may be necessary to implement new queries based on responding to a competitor's product introduction—queries not envisioned in the original design. If the database needs to be redesigned, reloaded, or indexed, this could mean a significant delay in responding to the competition.

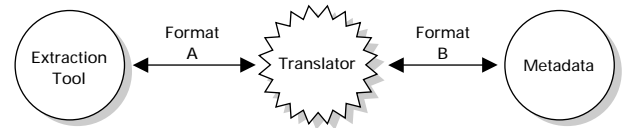
Ease/Speed of Implementation

The development and maintenance tools available with the Warehouse will be key to the success of the project. Are the tool sets tightly integrated, as is usually the case in a “one-stop-shop” solution, or will the development team have to wrestle with the tools as well as with the disparate products to make them work together? Are the user interfaces graphical or the old command line metaphor? Do the development tools automatically generate the metadata content, or will there be a second step to build the metadata and then a third to reconcile the mistakes made in doing this manually? These considerations apply not only to the initial deployment of the Warehouse, but also speak to the flexibility, since changes to the Warehouse will likely at some point involve the use of these tools again. A

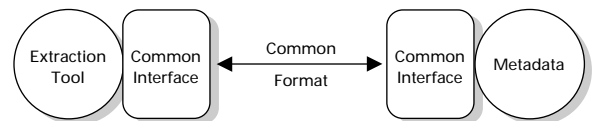
recent trade article in Datamation put it best: “Having tools that are well and deeply integrated with each other is a necessity, not a luxury.”

Integration

Integration deals with how well and how smoothly the different architectural components interact. Obviously these are subjective terms, and therefore there are degrees of integration in many different areas such as administrative interfaces, packaging and installation, problem determination, metadata, etc. Tight meta data integration would yield benefits from needing only to input meta data once and once operational preventing meta from becoming out of synchronization. For example, an extraction tool and a metadata repository could interact according to several different models as follows:

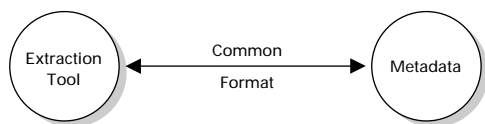


- In the first model, each component has its own structures and syntax for the metadata and a third component must be interposed between them to affect information transfer between them. In the figure above, a user-written translation module changes the formats of the information from one component to the other. Therefore in this model there is no integration between the components.



- A second option is to have each module maintain different internal structures and syntax, but to have a common defined interface, such as a meta-

data standard. In this case, the accepted transfer mechanism and translation is built in to each module by the appropriate vendor. Compromises must still be made because the internal structural differences may yet interfere with some functions, but overall the interaction between the components is facilitated by the acceptance of a common mechanism for information interchange. This situation is considered to be moderately well integrated, and in fact to have an accepted architecture which each component follows to affect this integration.



- Finally, on the other end of the spectrum are tools which use the same structures and syntax internally as well as present a common interface to each other. These components are completely integrated. The more vendors involved in the solution the more challenges exist integrating them—some single vendor solutions make this option relatively achievable.

The degree of integration of the Warehouse is an important consideration in developing an architecture and an implementation strategy. First of all, the Warehouse must integrate with the existing architecture and infrastructure of the organization. Reverting to the three models above, if the Warehouse can only integrate with the existing architecture by means of extensive interposition of custom code, the project will be very expensive, lengthy, and complex. Questions must be asked about the degree to which the proposed Warehouse must conform to existing supported operating systems, networks, data standards, existing data-

bases, existing application development environments, and more.

If the Warehouse supports common, open architectures, such as in the second model, the likelihood of being able to easily add new analysis functionality later, such as data mining for example, will be higher. Open architectures also make customization of products easier due to the more regular and predictable nature of the interaction among architectural components.

With regard to the existing operational applications from which the Warehouse will extract its information, the tighter the integration with all data sources, the better. Remember that some of the systems in the enterprise will have flat files or older hierarchical or network database architectures which can present a challenge.

Data can be extracted from the databases in discrete batches or continuously, on a transactional basis. Depending on the Warehouse application, either of these may be preferable or both may be required. Do extraction tools exist for both cases, or will there need to be some special code written? Finally, can the existing Warehouse automatically extract portions of the metadata from the existing operational systems, or will there need to be special software or processes employed to do so?

How easy will it be to integrate the Warehouse with existing operations? Data extracts, cleansing, transformation, and loading should not be done manually if it can be avoided. Not only does this introduce more cost in labor, but can also it interject too many vulnerabilities for error. Automating this process is a must.

The Warehouse should also integrate with enterprise standards, both de jure standards, those promulgated by officially sanctioned bodies such as ANSI, and de facto standards, which are practices and products generally accepted in the industry.

Completeness

The completeness of a solution talks to the existence of all the architectural components. This means that all of the parts an organization needs to work are in place and functional, from extraction and transformation to storage and metadata, to the analytical, management, and change processes required.

Implementation Tactics

Following is a high level guideline for implementing successful Data Warehousing projects.

Planning the Project

Many organizations are in such a hurry to install a system that they tend to gloss over this vital step. To paraphrase what Lewis Carroll's Cheshire Cat told Alice when she told him she didn't know where she was going, "if you don't know where you're going any road will get you there". The corollary, of course, is that if one does have a destination certain, then the choice of roads is important, and one had best spend some time setting it up.

Gaining commitment at top

Senior Management Sponsorship is crucial in these projects, and the level of support is commensurate with the scope of the project. If the project is Departmental in scope, then the senior Departmental leadership must be on board. Obviously, support even higher will only help, but be careful of bypassing chain-of-command positions which could turn a potential ally into a snubbed detractor.

Determining metrics for success of the project

How will the organization know that the project was successful? If there is an intention of using the original project as a proving grounds for an enterprise-wide rollout, how can there be a rollout if there is no yardstick for measuring the results of the pilot? As part of the planning process, it is critical that metrics be defined which clearly demonstrate that the results of the project meet and are aligned with the original business goals which drove the project to begin with. It is also imperative that the metrics be objective rather than subjective.

Writing a project plan

This step is anathema to many technologists, but a project plan detailing the objectives, approach, strategy, ownership, timeframe, and resources' responsibilities is a must if the project is to be managed with any degree of professionalism and efficiency.

Identifying the areas of expertise required

One of the outputs of a good plan is identification of the resources required and an analysis of whether they exist in-house and whether or not they are available. Make sure they are represented on the team. Having senior management support is a good prerequisite for being able to get the people you need, and if they are not available, for being able to get permission ahead of time to go outside (hire service providers) if necessary.

Developing a communications plan

A Warehouse is only useful to the organization if users exploit its abilities. It is foolhardy to spend effort on developing such a project only to spring a surprise onto a department or enterprise which is not ready to take advantage of it because they didn't know it was coming, or when it would be ready, etc. A communica-

tions plan is essential to disseminate information about the project.

Using benchmarking techniques

If possible, investigate other organizations who have successfully completed projects similar to the one at hand. This can take the form of literary research or actual trips to view operational systems, interview users, developers, and management as to the best practices they have encountered in their project. Remember that not all practices which work in one environment are universally portable to all environments and so make sure that the context in which a particular practice is said to work is extensible to the target environment.

Getting user/IT involvement

Given the relative newness of Data Warehousing technology, it is understandable that users don't know nearly as much about their data requirements as OLTP users. Therefore, it is critical to form a team where both technical and end-user personnel can work together to develop a mutually acceptable and technologically achievable set of specifications and requirements.

This is a good area to use RAD (Rapid Application Development) techniques to allow technicians to demonstrate to end users the potential capabilities of the system and to allow users a tangible feedback mechanism.

One implementation of Data Warehousing technology advocated the use of business managers as well as end users. The business managers can not only talk about what's done today in the existing processes, but are in a better position to articulate the business vision to both the technicians and the end users.

Selecting a methodology

A methodology that is known and accepted by the organization will go a long way to smoothing out many of the rough spots which any project will hit. A meth-

odology provides three components to a project:

- A logical series of activities to achieve a desired end. This structure will tell the organization what has to be done and when in order to produce a quality product.
- A definition of deliverables associated with each activity. This tells the organization what the output is of each step.
- Roles and responsibilities for participants in the activities.

The methodology will also help define a project structure and what has to be done to manage the project: when reviews are to be held and what each review will cover. One trap to avoid here is the one of putting someone in charge who is too narrow in their scope. For example, a technologist who has no appreciation for or understanding of the business problem, or a functional specialist who has no patience for technological issues. There has to be a blending of business and technology in order for these projects to succeed, and therefore, the lead people have to be sensitive to the various cultural differences of the team members. As the Data Warehousing Institute said in a recent publication: "Data Warehousing is a service business, not a storage business."

Using service providers

Any organization will have areas where there is either a lack of expertise or where there are insufficient resources available who have that expertise for this project. The project manager should look for areas of expertise which are not represented on the team. In many IT shops, the most significant missing element where consulting services might be required are the planning elements.

Architecting the Solution

Architecture is defined as the definition of the components of a solution and their interaction. This guide has delved into several areas where the benefits of a well defined and integrated architecture have been demonstrated. Defining the components of the solution and how they interact is a critical step in implementing a successful project. The solution must allow for all the characteristics described earlier, including scalability, flexibility, and manageability. The metadata component must be designed so that everyone understands the data in the Warehouse. Robert Typanski, Data Manager at Bayer was recently quoted in *Data-mation*. “unless people can identify the data that’s in the warehouse, they’re not going to be able to access it any better than if it were buried in some legacy operational system.”

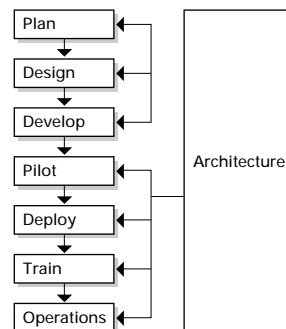
Defining a directory such as IBM’s DataGuide into the architecture will allow the system to draw business metadata from 25 other products, including DB2, Oracle, Sybase, Arbor Essbase, CASE tools, and more. Tools such as DataGuide can help provide users the keystone component of consistent, synchronized metadata that helps developers, maintainers and users alike.

Designing the System

Designs should be based on open, well understood architectures with well defined interfaces between the components. Use of standards will help the scalability and flexibility of the system over time. Navigational tools for users and user interfaces in general should be designed with the business problem in mind in a cooperative process with end users and managers.

Departmental systems should be designed within the context of an enterprise framework. That is, where possible and practical identify data elements and concepts which span multiple departments and try to

define them in the broadest possible terms so as to be able to include other departments later on.



Developing the System

The development environment should be capable of using RAD techniques to help end users who might not be able to grasp Warehousing and analysis concepts immediately. Use iterative prototyping with time boxing and multiple releases where necessary to gain consensus from the user community.

Use integrated development tools which automatically develop the metadata repository at the same time to avoid the expense of manually building the repository.

Piloting the Implementation

Pilot systems, also known as prototype systems or proof of concept systems are among the most misunderstood concepts in the industry. There are three possible reasons for an organization to embark upon a pilot program:

- The technology is foreign to the organization, and there is a need to understand the benefits of the technology to see how it might help in the business problems at hand.

- The technology is understood, but finding an appropriate application is not. The organization wants to know if a particular application of this technology to a business problem is appropriate, which is to say, will this technology solve the problem?
- The technology is understood as well as the application, but the cost/benefit equation is not understood. The organization wants to know if the cost of the solution is worth it.

Each of these motivations will result in different pilots, in different places in the organization and with different associated metrics.

Regardless of the motivation, however, a primary question which must be asked is: “How will we know if the pilot answered our questions?” The only way is to develop quantitative as well as qualitative metrics and in addition develop tools and techniques for capturing and analyzing the results at the end of the pilot program.

Deploying the System

If the strategy calls for replicating successful projects in other departments, then a roll out plan must be developed to identify the order of the rollout as well as any integration efforts which must be dealt with. These might include process reengineering, especially if multiple departments are today involved in a process which is not automated, and one of these departments will receive the automation prior to the others.

Training Users

One sure way to snatch defeat from the jaws of victory is to develop a technically outstanding Warehouse and then let users loose on it with no training. That which is intuitive to a technologist steeped in

Warehousing concepts is gibberish to an end user whose focus is running a business. Training programs must be developed which target not the technological niceties behind the screens, but rather that address the “how’s” of using the system from a business viewpoint. Focus on training the user on understanding the meta-data and the navigation capabilities as well as how to use those in analyzing a business problem.

Managing Warehouse Operations

Developing a Warehouse is one thing, keeping it operational is another. Daily operations create meta-data changes, such as the addition of a user, or the incorporation of an extract from a database, which means the metadata repository must be managed. In addition, the organization will need feedback reports on the Warehouse operations. Managers will want to know that the quarterly data extraction actually was started on schedule and completed on time with all appropriate indexes regenerated.

Some users, in spite of all best training efforts will construct queries that are capable of bringing a system down to its knees. In these situations, organizations will need database management tools that will report on how much system resources are being used by a query and that will allow the interruption and subsequent termination of the query.

Some reporting tools are also able to identify which areas of the database are being hit hardest, or most frequently, or even identify what times of the month or quarter or year those areas are most likely to be accessed. This will allow managers to govern the extraction and indexing processes accordingly to change table structures, drop columns, define different aggregations, and so on to tune the system according to the anticipated uses.

Summary

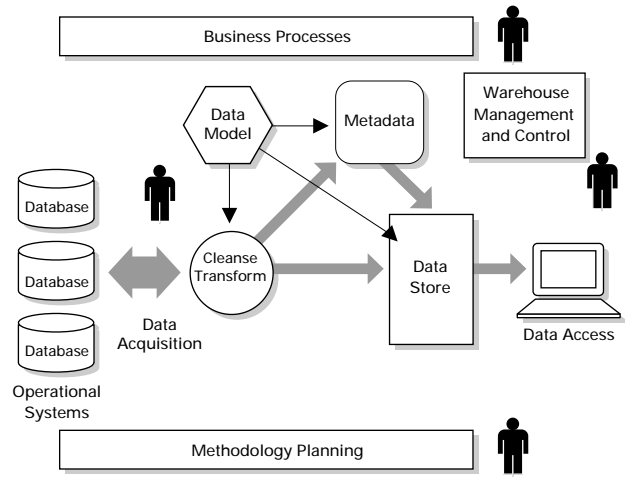
Data Warehousing can provide significant new benefits to an organization. The technology is moving from an early adopter stage to a maturity phase. This has removed much of the technology risk associated with early adopters stage but risk associated with implementation and strategy are still very much factors. According to some analyst statistics' over 50% of warehousing implementations fail to meet their goals. By considering the issues described in this guide you should be able to reduce risk associated with implementing data warehouse solutions and deliver higher value solutions in a reduced timeframe.

APPENDIX A

Elements of a Warehouse

The figure at the top of the next page represents the basic elements of a Data Warehouse architecture which responds to the needs of enterprises today. This model is slightly different from most of the ones being promulgated throughout the industry in that it combines the technological elements with the process and planning elements mentioned earlier. The human figures in the diagram represent those elements which are primarily of a service nature, such as planning, analysis, processes and management.

In this section we will briefly touch on the individual elements and their contribution to the overall architecture.



Planning/Analysis

Arguably the most important element of the Data Warehouse architecture, the planning and analysis constituent, is the one in which the enterprise determines which problems it needs to solve, what the characteristics of the problem are, and how best to solve them. Once the problem is clearly framed, then the team can move on to identifying and resolving other dependent questions such as:

- What existing business processes are involved in the defined problem,
- How large a warehouse do we need to satisfy our business problem,
- What operational data systems (existing operational data sources) do we need to access to get the data we need,
- What transformations will need to be built to convert the operational data into usable warehoused data, and
- What are the characteristics of the Data Model needed to build to answer the anticipated queries.

Data Acquisition

Data acquisition includes extraction of data from operational systems, cleansing the data (restructuring records, translating field values to a common data dictionary, checking data integrity and consistency), transformation (adding time fields, summarizing data, and deriving new fields), loading of the “clean” data into the Warehouse database, updating the Warehouse indexes, etc. It is the component of the architecture which provides the raw data into the warehouse repository which will then be used for analysis. Clearly, it will be important to understand how the data will be used so that transformations, cleansings, etc. render a suitable format with which to accomplish our query goals.

Data Storage

The Data Storage element represents the database and accompanying structures which are used in the analytical process. Relational and multidimensional databases are primarily used in architectures today, and while there is a fair amount of debate in the industry as to which provides the “best” results, relational databases offer more flexibility, smaller overall size, and easier access to atomic data for drill down operations. A trend emerging within the industry is for database vendors to increasingly add multi-dimensional functions into relational products.

Metadata

Just as any other kind of warehouse needs to keep an inventory of its holdings, a Data Warehouse needs to keep track of what data it is currently holding along with the “pedigree” of that data. This is the role of the Metadata repository, to give users and technicians information about the data such as where the data came from, what rules were used in creating the data, what the data elements mean, etc. Many systems divide the Business or End-User Directory (such as IBM’s

DataGuide) from the Technical or Development Directory (such as IBM’s Visual Warehouse), so that technical information about the data, which would be of limited use to an end-user, and could arguably make the end users’ task more difficult, are kept in a separate directory from that which is required by the user to understand the data from a business perspective.

Data Access

Here, the rubber meets the road in that the tools presented here to the end user are the ones which will ultimately drive their perception of the utility of the system. Access can be categorized using one of four common means: Standard Query, Data Interpretation, Multidimensional Analysis, and Data Mining.

Standard query tools allow users to develop a hypothesis and create questions (queries) to test its validity. This is sometimes called a verification-driven approach. Data interpretation tools support advanced analysis of data. Examples include business statistics and optimization (linear programming). Multi-dimensional analysis tools facilitate flexible investigation of the data along various dimensions, applying operations such as time series analysis, and enabling interactive “drill-down” capabilities. Data mining tools use a discovery-driven approach where sophisticated data mining “tools” are automatically applied to detect trends, patterns, and correlations hidden in the data.

Data Delivery

This element refers to how the resultant data is presented to the end user, and has two dimensions, the physical and logical conduit by which it reaches the end user, and the mechanisms by which the data may be visualized. Geographic dispersion, security, and data volumes will dictate the use of various conduits such as Local Area Networks, Internet, Wide Area Networks (public and private), etc. Query complexity and output

volume will dictate the rendering of the data. Options range from simple tables such as spreadsheets to simple two dimensional graphics such as bar graphs and pie charts to very sophisticated visualization technologies which utilize three dimensional landscapes to portray the results.

Management

Management of the Warehouse represents an indispensable architectural element. Once the Warehouse is set up, refreshes of the data must be scheduled and efficiently and correctly executed. Backups of the Warehouse must be made at reasonable and predetermined frequencies. Users must be added and deleted. Security must be administered. The list goes on. The larger the warehouse, the more difficult the task becomes, and therefore the more important that this be done well. A recent article in Computerworld stressed the fact that some Warehouses have reached a size such that data refreshes cannot be done without taking the system down during operational hours due to 18 and 20 hour upload processes.

Methodology

Methodologies are key ingredients for any complex project. A good methodology will define:

- A logical stepwise approach to solving a problem where each step builds on the previous ones,
- A complete set of deliverables so that nothing is “forgotten” in the project, and
- A project structure including roles and responsibilities of each participant.

Use of a good methodology will greatly increase the chances of success of any project, but in Data Warehousing it is essential due to the integrative (and therefore complex) nature of the problem.

Project Management and Administration

Although this technically falls under the purview of a good methodology, it is important to emphasize the need for solid project management in developing and setting up the Warehouse. Assembling the team, setting and managing expectations, and working an effective communications plan are all a part of a good project management team.

Case Studies

With Visual Warehouse you can effectively, easily, and inexpensively build and manage data marts using information stored in a variety of different source systems, including legacy, relational, IBM, and non-IBM databases. This information can be used for business intelligence and decision support purposes in a wide range of applications, including marketing, finance, manufacturing, real estate, and transportation.

Meeting Critical Data Requirements

Ryder System, Volt Information Sciences, and Intentia—three major corporations representing the entire spectrum of data warehouse customers, from end user through systems integrator and applications vendor—have all reached the same conclusion: IBM’s Visual Warehouse 2.1 for Windows NT is a powerful, economical, simple to implement and install data warehouse solution that provides consistent and coherent data for business intelligence and decision support.

All three businesses have been extensively involved in the testing and implementation of beta versions of Visual Warehouse Version 2.1 for applications ranging from transportation and real estate to banking and manufacturing. So impressed are they by the flexibility

and simplicity of IBM's new data warehousing solution for Microsoft Windows NT environments that they are incorporating Visual Warehouse for Windows NT as a core business service.

Ryder System Gets 10:1 ROI

Ryder System Inc., well-known for their transportation operations, is a company with more than 200,000 vehicles, 45,000 employees, and revenues of \$5.17 billion. While Ryder's largest product line is their \$2-billion full service truck leasing, their fastest growing product line is integrated logistics, with revenues now approaching \$1 billion.

Ryder supplies integrated logistic solutions to 500 customers in 30 different industries. These customers rely on Ryder's information technology to provide the information necessary to make accurate and profitable decisions. The data warehouse plays a major role in obtaining and processing data into decision support information that Ryder and their customers can understand and act on.

Three years ago, when the data warehousing explosion began, Ryder assembled a 20-person team from their Information Technology (IT) group and the user community to select a business partner that could help design and implement an enterprise-level data warehouse that could provide strong decision support for a very large number of projects. They chose IBM as their partner and Visual Warehouse as their solution. Ryder now uses Visual Warehouse to load millions of transactions from LANs, DB2 mainframe databases, VSAM flat files, and other relational databases, onto DB2 for AIX and DB2 for Windows NT.

Jim Sutter, senior data administrator of the IT group, said, "IBM built something that would allow us to do our jobs. We needed to start small and then rapidly scale up to process high volumes of operational data. IBM listened to our data warehouse requirements

and provided a powerful solution."

According to Sutter, Visual Warehouse's flexibility was a key factor in their decision. "You can't take a year to give somebody five minutes worth of data. With Visual Warehouse, we can build an application faster than we can talk about it. We can architect the data for easy access so that we can answer even the unpredictable questions. Visual Warehouse gives us visibility of our metadata, which helps provide the flexibility we need," he said.

Stephen Hendrick, a consultant with International Data Corporation (IDC), agrees with Sutter that the visibility of metadata is a key feature of Visual Warehouse. At a recent industry conference he said, "Visual Warehouse is one of the first products to emphasize the importance of driving the warehouse implementation process logically by means of the extensive use of metadata. Visual Warehouse's wide ranging functionality, combined with a usage metaphor which stresses a logical and metadata-driven development process, defines a product that is ideally positioned to address market needs."

While flexibility was a key concern, Sutter noted that cost was also very important. "Visual Warehouse has the architecture to do the job at a reasonable cost. In addition, because Visual Warehouse gives us easy access to crucial data, we can quickly provide information to our end users. This information is used to make accurate decisions that are profitable for both Ryder and its customers. Visual Warehouse has enabled us to improve the level of service we offer, while decreasing our costs and increasing our customers' satisfaction. Now, for every dollar we spend, we get back ten," said Sutter.

For Volt, Visual Warehouse for Windows NT is Key to Market Response

Finding a data warehouse product with a good return on investment was also extremely important to Volt Information Sciences, a billion-dollar services conglomerate based in New York City and Los Angeles. The company specializes in delivering complete, turn-key solutions-including everything from providing the technical and support personnel, to installing every aspect of the IS system-and now has more than 6,000 contract programmers and technicians at various customer sites.

As the key implementor of IBM contracts for data management systems in 11 Western states, the Volt Consulting Services (VCS) Division of Volt Information Sciences was also determined to uphold IBM's mission of providing the highest quality products with the highest quality service. The division immediately recognized the benefits of Visual Warehouse after installing it in their own internal enterprise.

VCS's internal enterprise set up a test-bench of the Visual Warehouse for Windows NT solution to access MVS DB2 databases; MVS VSAM files; Oracle, DB2 for Windows NT and DB2 for OS/2 databases; Microsoft SQL Server; and Btrieve database files. Volt's Professional Development Center in Orange, California pulls data across its T-1 data network from IBM MVS mainframes in Bluebell, Pennsylvania and also accesses data from client/server databases on its enterprise network. This setup enables the division to pre-stage customer requirements and test data from the customer, duplicating the customer's environment. VCS can then make any adaptations and create the appropriate business views of the data. Customers can dial in from their sites via a remote access server to fine tune the business views. This enables VCS to provide a prototype to the customer before the actual on-site installation.

"We think Visual Warehouse is a real winner because of its superior price/performance ratio," said Daniel S. Rimkus, director of Volt Consulting Services. "It allows us to learn our customer's business and environment. It also leads us to more sophisticated data warehousing and business applications such as data mining, software agents, and IBM's complete line of data interfaces. For example, Visual Warehouse naturally evolves into use with IBM's DataJoiner and Data-Hub to support sophisticated data mining applications."

VCS has already begun multiple Visual Warehouse applications, including a project for the real estate management industry that each day processes around 100 MB of data, which is utilized for sales lead management and demographic studies. Projects are also in the works for the financial industry and for various marketing information systems.

VCS found that many of their customers wanted to reap the benefits of the new, powerful, and easy-to-use data warehousing software, but were not willing to make a major investment to re-engineer information systems.

"IBM's Visual Warehouse can utilize the customer's existing databases. It can be used with whatever legacy database the customer has running. It is an extremely scalable, multiprocessing solution that enables our customers to build and manage department-size data marts and to share this information in a larger enterprise environment. With all of these benefits, IBM's Visual Warehouse gives us a 10 to 1 price implementation advantage," said Rimkus.

With Visual Warehouse, Volt can build a data mart on DB2 for OS/2, AIX, OS/400, Windows NT, HP-UX, and Solaris. Data can be retrieved from a variety of sources, including the DB2 family, VSAM, IMS, VM flat files, Oracle, Sybase, and binary large objects (BLOBs). In addition, Visual Warehouse's flexible configuration enables it to reside on the same

server as the data warehouse or on a remote server, and it can manage several different data warehouses.

Gerald C. Ostrowsky, National Director of Strategic Services for Volt Information Sciences, adds, “As a full service systems integrator, we see many advantages to IBM’s Visual Warehouse. Its primary value is its ability to interface to all the major databases and operating system platforms, making it the perfect tool for today’s enterprise data mart solutions. Visual Warehouse installs easily and is priced very competitively. Plus, tailoring the product to our clients’ needs gives us a better understanding of their business environment. This increases the overall value of our relationship to our customer, and creates more opportunities to enlarge this relationship.”

International Product Support Crucial to Intenia

Intenia, with annual sales of \$170 million, is the developer of MOVEX[®], a leading international business application software program. The Movex Business Application System is used to support business processes in distribution and manufacturing. To maximize the benefit of its integrated software systems, Intenia has put data warehousing and decision support systems at the top of its agenda, seeking the best tools on the market to implement these systems. As one of the world’s ten largest integrators of business application systems, and among the five most preferred Business Partners for IBM’s AS/400 family, the Stockholm Sweden-based company employs 1,200 people and is experiencing an annual growth rate of more than 50 percent.

Intenia’s data warehousing solutions, developed by the company’s research and development division for their network business partners and sales consultants, are being implemented by customers in close to 40 countries around the world.

“Our customers expect us to deliver turn-key manufacturing systems that are transparent to the manufacturing operations,” said Thomas Ahlerup, senior development manager of Intenia’s Business Intelligence Systems. “They buy the entire value chain, and they expect support.”

Intenia provides this support by maintaining a high level of involvement with their customers, conducting multiple sets of pre-studies and analyses before implementing the installation. While this preparation is extremely necessary, it is often performed under very tight deadlines.

Said Ahlerup, “Simplicity is the key benefit of Visual Warehouse. We have to have our customers up and running quickly. Another reason Visual Warehouse is so promising is that it is completely transparent to all of our existing installations. Regardless of what release of our software the customer is running, we can easily implement a warehouse strategy and do it using less of our consulting resources, which saves everyone money.”

Intenia uses Visual Warehouse to access transactional and operational data in Movex to build enterprise-wide data warehouses as well as advanced departmental data marts. The company was among the first in the world to ship a proprietary application embedding Visual Warehouse and plans to offer this data warehousing solution to its entire existing customer base of approximately 1,700 sites worldwide.

“With our global network of customers, IBM’s worldwide support for Visual Warehouse is extremely important. They give us true international representation,” said Ahlerup.

A Solution Whose Time is Ripe

Ahlerup is not alone in noticing that, from a business perspective, the importance of the functionality of the data warehouse is growing. IDC analyst Stephen Hendrick said, “With the data warehousing market

approaching adolescence, there has certainly been no lack of tactical products and tools to address various warehouse implementation and generation tasks. However, virtually all of these tools are very physical in their approach to addressing data warehouse needs and limited in their ability to address the wide-ranging requirements associated with designing, implementing, administering, and maintaining a data warehouse. IBM's Visual Warehouse is the first affordable production-quality tool to address all of the major warehouse implementation, generation, and administration issues in an integrated and unified product."

Today's businesses are becoming more aware that they need to reengineer their transactional and operational data so that it can be easily used to analyze and support management processes. On the other hand, IS departments, systems integrators and application vendors are also discovering that the business world is moving too fast and running too lean to implement big-budget, multi-year data warehouse projects. For these reasons and more, Visual Warehouse is becoming the warehousing solution of choice.

For more information about Visual Warehouse, visit the IBM World Wide Web page at:
<http://www.software.ibm.com/data/warehouse/vw>

These case studies illustrate how three customers use IBM products. Many factors have contributed to the results and benefits described. IBM does not guarantee comparable results. All information contained herein was provided by the featured customers and IBM Business Partners. IBM does not attest to its accuracy.

Glossary

Ad Hoc Query—Any query that cannot be determined prior to the moment the query is issued.

Business Directory—The portion of the Metadata Repository which deals with information about data pertinent to the business users. For example, metadata in the Business Directory will tell users what factors are included in the calculations of gross vs. net sales.

Catalog—A component of a data dictionary that contains a directory of its DBMS objects as well as attributes of each object.

Centralized Data Warehouse—A Data Warehouse implementation in which a single warehouse serves the need of several business units simultaneously with a single data model which spans the needs of the multiple business divisions.

Change Data Capture—The process of capturing changes made to a production data source. Change data capture is typically performed by reading the source DBMS log. It consolidates units of work, ensures data is synchronized with the original source, and reduces data volume in a data warehousing environment.

Copy Management—The process of taking a snapshot of data from a source and copying it to a target environment.

Crosstab—A process or function that combines and/or summarizes data from one or more sources into a concise format for analysis or reporting.

Data Administration (DA)—The processes and procedures by which the integrity and currency of the data in the warehouse are maintained.

Data Aggregation—A type of data derivation where a data value is derived from the aggregation of different data occurrences of the same subject data. For

example, yearly sales data can be aggregated from monthly sales data.

Data Cleansing—The process of removing errors and resolving inconsistencies in source data before loading the data into a target environment.

Data Dictionary—A collection of definitions and specifications for data categories and their relationships. A database of data about data (metadata).

Data Extract—The process of copying a subset of data from a source to a target environment.

Data Loading—The process of populating the data warehouse. Data loading is usually done using vendor-supplied utilities or user-written programs.

Data Mart—A type of data warehouse designed to meet the needs of a specific group of users such as a single department or part of an organization. Typically a data mart focuses on a single subject area such as sales data. Data marts may or may not be designed to fit into a broader enterprise data warehouse design.

Data Mining—A process of analyzing large amounts of data to identify hidden relationships, patterns, and associations. This is often called “discovery-driven” data analysis.

Data Model—A logical map that represents the inherent properties of the data independent of software, hardware or machine performance considerations. The model shows data elements grouped into records, as well as the association around those records.

Data Propagation/Replication—A process for distributing data from a source database to target databases while usually keeping the databases synchronized.

Data Scrubbing/Transformation—The process of filtering, merging, decoding, and translating source data to create validated data for the data warehouse.

For example, a numeric regional code might be replaced with the name of the region.

Data Warehouse—A subject oriented, integrated, time-variant, non-volatile collection of data in support of management’s decision making process. A repository of consistent historical data can that can be easily accessed and manipulated for decision support.

Database—A collection of data which are logically related.

Database Management Systems (DBMS)—A software system for creating, maintaining and protecting databases.

Decision Support System (DSS)—Systems which allow decision makers in organizations to access data relevant to the decisions they are required to make.

Derived Data—Warehouse data that results from calculations or processing applied to the source data before it is stored in the Warehouse environment. For example, source data might be used to calculate ROI (Return On Investment) which is stored as derived data in the Warehouse.

Drill Down—A method of exploring detailed data that was used in creating a summary level of data. Drill down levels depend on the granularity of the data in the data warehouse.

Enhanced Data—Warehouse data that has been cleansed, scrubbed, transformed, derived, summarized, or aggregated.

Enterprise Data Model—A blueprint for all of the data used by all departments in the enterprise. An Enterprise Data Model has resolved all of the potential inconsistencies and parochial interpretations of the data used and presents a consistent and commonly understood and accepted view and definition of the enterprise data.

Enterprise Data Warehouse—An Enterprise Data Warehouse is a Centralized Warehouse which services the entire enterprise.

Executive Information System (EIS)—Tools programmed to provide canned reports or briefing books to top-level executives. They offer strong reporting and drill-down capabilities. Today these tools allow ad-hoc querying against a multi-dimensional view of data, and most offer analytical applications along functional lines such as sales or financial analysis.

Extract—The process of copying a subset of data from a source to a target environment.

Information Directory—A system for browsing Data Warehouse metadata. Sub-components usually include a business directory and a technical directory. See the entries for Business Directory and Technical Directory.

Information Warehouse™—IBM's approach to data warehousing that supports the implementation of either functional, central or decentralized warehouses.

Informational Applications—Applications which are written to analyze data from a Data Warehouse for Decision Support purposes.

Informational Data—Data which has been extracted, summarized and stored in a Data Warehouse for purposes of supporting Informational Applications.

Metadata—Data about data. For example, information about where the data is stored, who is responsible for maintaining the data, how often the data is refreshed, etc.

Middleware—A communications layer that allows applications to interact across hardware and network environments.

Multi-Dimensional Analysis (MDA)—Informational Analysis on data which takes into account many different relationships, each of which represents a dimension. For example, a retail analysis may want to understand the relationships among sales by region, by quarter, by demographic distribution (income, education level, gender), by product. Multi-dimensional analysis will yield results for these complex relationships.

Multi-Dimensional Database (MDB)—A database which has been constructed with the multiple dimensions pre-filled in hyper dimensional “cubes” of data rather than the traditional two dimensional tables of Relational Databases.

On-Line Analytical Processing—Processing that supports the analysis of business trends and projections. It is also known as decision support processing and OLAP.

On-Line Transaction Processing—Processing that supports the daily business operations. Also known as operational processing and OLTP.

Operational Data—Data used in the operational processing of business transactions that support day-to-day business operations. They are detailed, largely primitive data necessary to keep the organization operating.

Operational Applications—Applications which support the daily operations of the enterprise. Usually included in this class of applications are Order Entry, Accounts Payable, Accounts Receivable, etc.

Query—A request for information from the Data Warehouse posed by the user or tool operated by the user.

Relational Database Management System (RDBMS)—A database system built around the relational model based on tables, columns and views.

““ *Ryder's Enterprise Data Warehouse strategy required a flexible administration tool that could quickly retrieve data from a wide range of sources, manage the design, construction and operation of distributed datamarts from a central design center, and start small then scale to process high volume operational data as required. We selected Visual Warehouse for Windows NT because it has the architecture to do the job at a reasonable cost, and its ROI is very positive indeed.* ””

*Jim Sutter, Data Warehouse Architect
Ryder System, Inc.*

““ *IBM's Visual Warehouse is the first affordable production quality tool to address all of the major warehouse implementation, generation and administration issues in an integrated and unified product. Visual Warehouse is also one of the first products to emphasize the importance of driving the warehouse implementation process logically by means of the extensive use of meta data. Visual Warehouse's wide ranging functionality combined with a usage metaphor which stresses a logical and meta driven development process define a product which is ideally positioned to address market needs.* ””

*Stephen Hendrick
International Data Corporation*

This Technology Guide is one of a series of guides, published by ATG, designed to put complex data warehousing concepts into practical and understandable terms.

Each guide provides objective, non-biased information to assist in the internal education, evaluation and decision making process.

This Technology Guide, as well as the other Data Warehousing Technology Guides in the series, are available on ATG's Web Site.

www.techguide.com

GC26-9030-00



Printed in USA

Produced and Published by

ATG
The Applied Technologies Group

One Apple Hill, Suite 216, Natick, MA 01760

Tel: (508) 651-1155 Fax: (508) 651-1171 E-mail: info@techguide.com