

Why Companies Should Outsource Data Warehousing



by Shawn Burke, Senior Network Engineer,
Certified Cisco Design Associate

Introduction

In the past, it has been a problem to provide all key people in an enterprise with access to the information necessary to enable the enterprise to survive and prosper in an increasingly competitive world. Recently, a set of significant new concepts and tools have evolved into a technology that makes it possible to attack this problem.

The term that has come to characterize this new technology is 'data warehousing'. With this technology comes the threat of many companies leaping into the field with few skills, little training, and little expertise.

Building a data warehouse is a complex venture. A data warehouse is not a technology that can be bought off the shelf as a unified whole. The costs to a company can be overwhelming, and internal resources scarce.

Data warehouse projects are much harder to complete successfully than traditional systems development projects. As companies are forced to face this new challenge, they should also be aware of how outsourcing can overcome the obstacles associated with this technology.

This paper will focus primarily on providing companies with enough insight to seriously consider outsourcing data warehousing. Topics to be examined include:

- A tutorial and discussion on key attributes of both outsourcing in general, and data warehouses in particular, will solidify the winning formula of combining the two.

- A summary of how outsourcing provides the expertise required to succeed in data warehousing, and an examination of the resource-intensive factors involved.

- Finally, a brief overview on ways that USi, the leading Internet Managed Application ProviderSM (iMAPSM), can offer comprehensive solution suites for successful enterprise data warehousing.

1. Outsourcing

The concept of outsourcing gained credibility in the 1970s with the focus largely on the Information Technologies area in Fortune 1000 companies. These outsourcing arrangements were initiated and implemented by giants such as EDS, Arthur Andersen, E&Y, Price Waterhouse, and others. Today, outsourcing is a \$100 billion per year industry which, according to recent research, will grow to \$282 billion per year worldwide by the year 2000.

What is Outsourcing?

Simply put, outsourcing is the practice of subcontracting work to other companies or individuals. But the difference between basic subcontracting and true outsourcing is that outsourcing involves the wholesale restructuring of the corporation around core competencies and outside relationships. Outsourcing is a management strategy by which an organization outsources major non-core functions to specialized and efficient service providers.

The term 'outsourcing' has been applied to many different relationships across a variety of business areas, and has existed in many forms well before Ross Perot was credited with starting the industry over 30 years ago.

Companies have always hired special contractors for particular types of work, or to level off peaks in their workload. They have always partnered, forming long-term relationships with firms whose capabilities complement

their own, and have always contracted for shared access to resources that were beyond their individual reach — whether it be buildings, technology, or people. The necessity for outsourcing has been realized in most organizations.

Why Companies Should Outsource

Statutory requirements, such as IEEE standards or ISO9001 certification, are rapidly increasing the drain on employer financial resources. Recent studies by business and government organizations conclude that it is no longer cost-effective to divert resources away from core business functions into non-core essential obligations, such as a data mart, an intranet, or a sales force automation solution.

Outsourcing enables companies to focus on core activities and increase flexibility by knowing when to farm out functions best done by others says the US Chamber of Commerce. In other words, anyone whose workload exceeds their permanent staff's ability to meet deadlines, as well as the company's financial bottom line, is a prime candidate for outsourcing.

Ten Reasons Companies Outsource

1. Reduce and control operating costs
2. Improve company focus
3. Gain access to world-class capabilities
4. Free internal resources for other purposes
5. Resources are not available internally
6. Accelerate reengineering benefits
7. Function is difficult to manage or out of control
8. Make capital funds available
9. Share risks
10. Cash infusion

Source: Survey of Current and Potential Outsourcing End-Users, The Outsourcing Institute Membership, 1998

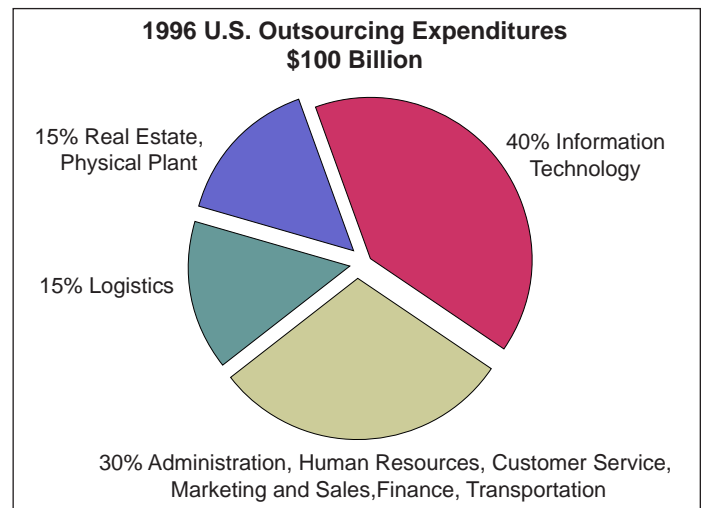


Figure 1: U.S. Outsourcing Expenditures

Summary of 'Reasons to Outsource'

Outsourcing has become the preferred method of doing business by companies for the reasons outlined above, and other reasons as well. Companies have geared toward outsourcing information technologies more than any other business function. It is predicted by the year 2000 the worldwide IT outsourcing market will grow to over \$100 billion.

In keeping with the theme of this paper, emphasis will be placed on how companies need the world-class capabilities, expertise, reduced costs, and resources to obtain and keep a successful data warehouse.

2. Data Warehousing

A data warehouse is a structured extensible environment designed for the analysis of non-volatile data which is logically and physically transformed from multiple source applications to align with business structure, is updated and maintained for a long time period, is expressed in simple business terms, and can be summarized for quick analysis.

Companies have data. Most companies have a lot of data that is stored in computer systems. This data has come from sales, production information, recording purchases, and customer information. Often this data is stored in many different places. In data warehousing a company gathers its information and puts it in one place - the data warehouse. Data warehousing is the process of organizing that data in a way that creates new business knowledge and enables new insights into business processes.

The Evolution of Data Warehousing

In reviewing the evolution of data warehousing, we need to begin with an overview of traditional approaches to historical data, followed by an insight on the emergence of key enabling technologies.

Traditional Approaches to Historical Data

Throughout the history of systems development, the primary emphasis had been given to the operational systems and the data they process. It is not practical to keep data in the operational systems indefinitely, and only as an afterthought was a structure designed for archiving the data that the operational system has processed.

In the 1970s virtually all business system development was done on the IBM mainframe computers using tools such as Cobol, CICS, IMS, DB2, etc. The 1980s brought in the new mini-computer platforms such as AS/400 and VAX/VMS. The late eighties and early nineties made UNIX a popular server platform with the introduction of client/server architecture.

Despite all the changes in the platforms, architectures, tools, and technologies, a remarkably large number of business applications continue to run in the mainframe environment of the 1970s. By some estimates, more than 70% of business data for large corporations still resides in the mainframe environment. These legacy systems continue to be the largest source of data for analysis systems.

During the past decade, the sharply increasing popularity of the personal computer on business desktops has introduced many new options and compelling opportunities for business analysis. The gap between the programmer and end user has started to close as business analysts now have at their fingertips many of the tools required to gain proficiency in the use of spreadsheets for analysis and graphic representation. The only problem is that each individual user has obtained only the information that he or she requires. The time and cost involved in addressing the requirements of only one user prove prohibitive.

Today's data warehousing systems provide the analytical tools afforded by their precursors. But their design is no longer derived from the specific requirements of analysts or executives; and, as we will see later, data warehousing systems are most successful when their design aligns with the overall business structure rather than specific requirements.

Emergence of Key Enabling Technologies

Many factors have influenced the quick evolution of the data warehousing discipline. The most significant set of factors has been the enormous forward movement in the hardware and software technologies. Sharply decreasing prices and the increasing power of computer hardware, coupled with ease of use of today's software, has made possible quick analysis of hundreds of gigabytes of information and business knowledge.

The most important factor in the evolution of data warehousing has been the sharply increasing power of computer hardware. Along with the increase in this power, their prices have fallen just as sharply. Gordon Moore, co-founder of Intel, predicted that the capacity of a microprocessor will double every 18 months. This has not only held true for the processor but also for other components of the computer. The Pentium II and Alpha processors have brought incredible power to the commodity computer market. Higher capacity memory chips, a key component influencing the performance of a data warehouse system, are now available at low prices.

With the increasing power of the desktop hardware and software, data warehousing has evolved at a rapid pace. But the most important development in computing since the advent of the personal computer is the explosion of Internet- and web-based applications.

One of the most exciting fields in the computing industry today is the development of Internet applications. Intranets are private business networks that are based on Internet standards, although the designs are to be used internally. The Internet/intranet trend has very important implications for data warehousing applications. First,

data warehouses can be available worldwide on public/private network at much lower cost. This availability minimizes the need to replicate data across diverse geographical locations. Second, this standard has allowed the web server to provide a middle tier where all the heavy-duty analysis takes place before it is presented to the web-browsing client to use.

The increasing powers of software and hardware, along with the availability of affordable and easy-to-use reporting and analysis tools have played the most important role in evolution of data warehouses.

The Framework of the Data Warehouse

As data warehouses have gradually evolved, organizations have increasingly accepted their viability. One of the reasons that data warehousing has taken such a long time to develop is that it is actually a very comprehensive technology. In order to understand all how all of the components involved in a data warehousing strategy are related, it is essential to have an Data Warehouse Architecture.

Architecture of the Data Warehouse

The overall structure of data, communication, processing and presentation that exists for end user computing within the enterprise represents the data warehouse architecture. There are a number of interconnected parts that make up the architecture:

- Information Access Layer
- Operational Data Base / External Data Base Layer
- Data Access Layer
- Data Directory Layer
- Process Management Layer
- Application Messaging Layer
- Data Warehouse Layer
- Data Staging Layer

Information Access Layer

End users deal directly with the information access layer. It represents the tools that the end user normally uses day to day, e.g. Lotus 1-2-3, Focus, Access, Excel, and SAS, etc. This layer also includes the hardware and software involved in displaying and printing reports, graphs spreadsheets, and charts for analysis and presentation. Today, more and more sophisticated tools exist on the desktop for manipulating, analyzing, and presenting data.

Operational Data Base/External Data Base Layer

Operational systems process data to support critical operational needs. Operational data bases have been historically created to provide an efficient processing structure for a relatively small number of well-defined business transactions. The goal of data warehousing is to free the information that is locked up in the operational databases and to mix it with information from other, often external, sources of data. This information includes econometric, demographic, competitive, and purchasing trends.

Data Access Layer

The data access layer of the data warehouse architecture is involved with allowing the information access layer to the operational layer. The common data language that has emerged to become today's de facto standard for data interchange is SQL. The data access layer is then responsible for interfacing between information access tools and operational databases.

Data Directory Layer

For the provision of universal data access, it is absolutely necessary to maintain some form data directory or repository of meta-data information. Meta-data is the data about data within the enterprise. In order to have a fully functional warehouse, it is necessary to have a variety of meta-data available, data about the end user views of data and data about the operational databases. End users will access data from the data warehouse without having to know where that data resides or the form in which it is stored.

Process Management Layer

The process management layer is involved in scheduling the various tasks that must be accomplished to build and maintain the data warehouse and data directory information. The process management layer can be thought of as the scheduler or the high-level job control for the many processes that must occur to keep the data warehouse up to date.

Application Messaging Layer

The Application Message layer has to do with transporting information around the enterprise-computing network. Application Messaging is also referred to as “middleware”, but it can involve more than just networking protocols. Application messaging is the transport system underlying the data warehouse.

Data Warehouse Layer

Actual data used primarily for informational uses occurs at the data warehouse layer. In a physical data warehouse, copies of operational and or external data are actually stored in a form that is easy to access and is highly flexible. Data warehouses may be stored on client/server platforms, but they are often stored on mainframes as well.

Data Staging Layer

The final component of this architecture is data staging. Data staging includes all the processes necessary to select, summarize, edit, combine, and load data warehouse and information access data from operational and external databases. Data staging involves data quality analysis programs and filters that identify patterns and data structures within operational data.

Summary

After the rapid acceptance of data warehousing systems during the past three years, there will continue to be many more enhancements and adjustments to the data warehousing system model. With this fact in mind, the focus will now turn towards integrating the necessity of outsourcing data warehouses.

3. The Data Warehouse Process

Warehouse projects are different from traditional relational design. Do not assume that a skilled traditional developer will be able to lead a warehouse project. The data warehouse process starts with a successful strategy for developing the warehouse, followed by a design that brings forth the actual building of the warehouse. After analyzing the complexity of this process, many companies should choose to outsource data warehouse needs because of the level of specialization and variation of expertise that is needed.

Developing a Data Warehouse Strategy

It is critical to develop a balanced data warehousing strategy that is appropriate for its needs and user population. Questions to be answered are: Who is the audience? What is the scope? What type of data warehouse should we build? The answers could only come from an experienced professional.

One strategy is to establish a virtual data warehouse environment. This is created by following several steps:

- (1) install a set of data access, data directory, and process management facilities
- (2) train end users
- (3) monitor how the data warehouse facilities are actually used
- (4) based on actual usage, create a physical data warehouse to support the high-frequency requests

A second strategy is to build a copy of the operational data from a single operational system and enable the data warehouse from a series of information access tools. This approach can create a number of problems if the existing data is of poor quality and the access to the data has not been thought through carefully.

Ultimately, the optimal data warehousing strategy is to select a user population based on value to the enterprise and do an analysis of their issues, questions, and data access needs.

In the final analysis, there is no single approach to building a data warehouse that will fit the needs of every enterprise. Since data warehouse technology is evolving as we learn more about developing data warehouses, the complexity leads to the conclusion that the only practical approach to data warehousing success is to outsource it.

Designing and Implementing Data Warehouses

Designing data warehouses is very different from designing traditional systems. Designing a data warehouse often involves thinking in terms of much broader and more difficult to define business concepts than does designing an operational system. The fact that end users have difficulty defining their minimum needs complicates the matter. Data warehouse designers find that they have to use their finesse to help their users visualize their requirements. In this respect, robust working prototypes are essential.

Warehouse designs must work with the following factors:

- large volumes of data
- ability to integrate multiple forms of storage media
- ability to freely index and monitor data for maintenance reasons
- ability to assign CPU prioritization to users
- ability to provide levels of security and visibility based on role entity
- ability to monitor query requests and performance
- ability to monitor system level I/O
- performance and resource utilization
- deliver a meta-data access vehicle
- ability to manage data in parallel devices
- ability to optimally place data on storage media
- deliver robust tools for ad hoc DSS query, ad hoc DSS reporting, and customized EIS applications
- ability to enforce index-only queries; provide for bulk loading, provide for disabling locking/concurrency schemes; handle many data types such as BLOB, bitmap, varchar text, etc.
- provide for distributed warehousing and remote access

The fact is that there are too many considerations to list for warehouse designers. Only a truly experienced professional can actually implement a data warehouse project.

One of the biggest mistakes made in this field is to focus on the building of a data warehouse, viewing it as an accomplishment to be completed and achieved, rather than focusing on the ongoing process of data warehousing. The construction of a data warehouse will never be finished. The process of data warehousing is a continuing journey that requires an abundance of technical expertise.

When the data warehouse is properly designed and constructed, it extracts the information value from data and puts it into a form that can be easily understood and interpreted by the people who need it to make decisions. In the past, an organization could maintain its competitive advantage just by using information technology to automate repetitive tasks and control complex processes. Now, however, these traditional uses for data processing are no longer enough to keep a company competitive. It is essential to access the informational value of data in order to provide true decision support capabilities. By leveraging world-class capabilities through outsourcing, a

properly designed and constructed data warehouse will maximize the decision-making power that is hidden in every organization's data.

Summary

Excellence and expertise are now recognized as the keys to competitiveness, and they are rooted in specialization. Organizational specialization takes many forms. One of the most important is a keen understanding and aggressive investment in an organization's core competencies. Corporations realize that to gain competitive advantage in today's business world, they must transform themselves into streamlined entities concentrated on the things they do best. Outsourcing is the answer for excellence and expertise.

The real key to success in data warehousing lies in the people and management processes surrounding it. Without the people and the management processes, data warehouse technology by itself can not insure success.

4. Associated Costs for Data Warehousing

According to a report by the Insight Research Corporation, (<http://www.insight-corp.com>, Data Warehousing in Networks 1998-2003), the cost of implementing a data warehouse can run anywhere from \$40,000 to \$3 million, depending on the size of the carrier and the number of customers involved. In today's world of global competition, businesses are trying desperately to boost, or at least maintain, profits and market share while reducing spending costs.

Funding of the Data Warehouse

The purchase of hardware, software, and services requires a serious investment, which hopefully will increase the capitalization for the company and return profits to the investors and stock holders. The decision for acquiring the data warehouse versus other projects within the company should be based on sound financial analysis.

While financial analysis should drive decision making, more often the people advocating the data warehouse are driven by Generally Accepted Accounting Procedures (GAAP) which charges someone's budget with the depreciation and operating expense for the data warehouse. A company can avoid these harmful effects upon its budget but seeking the benefits of the outsourcing.

Funding is not a one-time issue. A tremendous amount of money will be spent on pilot projects. Expect the first pilot project to fail, along with others. If the data warehouse is successful, then additional money will be required for additional capacity. As the growth of the data warehouse demands more capacity, the systems organization will need more funding.

A factor that will always vary is the allocated floor space needed for the data warehouse. As the need for more space surfaces as the data warehouse grows, scalability becomes key. Meanwhile, precious floor space is used up, and the high price of real estate per square foot places an increasing burden on the company attempting this implementation.

The total costs of the data warehouse will include a number of support personnel, who will require staffing authorizations and budgets to cover their costs. Technologies vary greatly in the amount of resources needed to manage the data warehouse. (See figure 4)

Summary

Access to the provider's lower cost structure, which may be the result of a greater economy of scale or some other advantage based on specialization, is clearly and simply one of the most compelling tactical reasons for outsourcing. Companies that

	2 Systems Personnel	5 Systems Personnel*
Year 1	\$500,000	\$800,000
Year 2	\$650,000	\$950,000
Year 3	\$650,000	\$950,000
Year 4	\$650,000	\$950,000
Year 5	\$650,000	\$950,000
Total	\$2,900,000	\$4,500,000

Figure 4: Effect of support costs

attempt to build data warehouses on their own will incur vastly higher research, development, marketing and deployment expenses. These expenses will be passed onto the ultimate customer, who may then choose to do business with a competitor offering more reasonable prices.

5. Classic Resources for Data Warehousing

It is beneficial for many companies to allow a vendor to perform all the work and then provide the company with a collection of integrated subject-oriented databases designed to enable decision support systems and manage the overwhelming volumes of data collected. As warehouses are being rolled out almost daily in almost every industry, new obstacles to managing and maintaining the warehouse are appearing.

Managing the Data Warehouse

Data warehouses do not appear as if by magic; they take a great deal of work. A data warehouse is a good investment only if it is an improvement over existing technology, enabling end users to locate vital information faster and more cost-effectively than they could using their old familiar technology. Consequently, management has to think seriously about how they want their warehouses to perform and how they are going to get the word out to the end user community.

Management also must know that the maintenance of the data warehouse is critical. Experience has shown that data warehouses quickly become one of the most used systems in any organization. Management, especially in the field of IT, must also be aware that if they embark on a data warehousing program, they will create new demands upon their operational systems; demands for better data, demands for consistent data, demands for different kinds of data. These demands can really only be met sufficiently by outsourcing.

Maintenance of Data Warehouses

With more and more warehouse projects now reaching the maintenance phase, the temptation to assign blame for data warehousing failures or difficulties with vendors is growing. However, there are far more overwhelming technical problems that can undermine the warehouse maintenance efforts.

- Spending excessive amounts of time writing data cleansing routines as more situations are uncovered
- Struggling to keep up with current technology
- Keeping pace with user demands for development and data mart implementation, relying upon a static number of staff in a changing corporate environment
- Taking raw operational data and transforming it into semantically consistent information that is meaningful to the end user
- Capturing and effectively managing meta-data, from simple audit trails to tracking complex operations such as data primacy and old key/new key correspondence
- Making descriptions of meta-data elements available in a natural language format that users can readily understand
- Exchanging meta-data among tools in the warehouse environment
- User expectations increasingly geared toward 24x7x365 availability

As projections for the future and progress toward long-term goals are being made in tandem in all areas of data warehousing, it is clear that the perfect world is still ahead of us. Long-term maintenance problems await creative solutions. However, by outsourcing, companies can be assured of finding integrated solutions for maintenance of a data warehouses.

Summary

Until there are better ways to meaningfully understand and process the volumes of data available within a company, any company considering implementing a data warehouse will have to anticipate a growing monster that will require more IT/IS staff than they currently employ. A clear choice to accommodate this lack of support is obviously an outsourcing strategy.

Data warehousing is not a new phenomenon. All large organizations already have data warehouses, but they are not managing or maintaining them in-house. Over the next few years, the growth of data warehousing is going to be enormous with new products and technologies coming out frequently. In order to get the most out of this period, it is going to be important that data warehouse planners and developers have a clear idea of what they are looking for, and then choose strategies and methods that will provide them with performance today and flexibility for tomorrow.

6. A Unique Solution Provider for Outsourcing

USi, the leading iMAP, and Sagent Technology, Inc., provider of turnkey solutions for bringing data marts to the Web, have combined to form a partnership to deliver data mart-based decision support solutions over the Internet based on a fixed monthly subscription price model. This end-to-end solution is the USi Rapid Data Mart Powered by Sagent.

USi combines Sagent software with rapid deployment methods and data warehousing consulting expertise, backed by a world-class continuous operations data center environment. The benefits are evident within months, without the risk of funding a warehouse. For their flat monthly fee, a company receives round-the-clock, mission-critical, customized support.

Conclusion

A company should not be relied upon to provide all the data warehouse components or expertise. A company cannot afford to internally sustain the technical expertise required. It is therefore important that organizations realize that outsourcing will allow them to leverage new technologies, incur low costs, and hurdle all obstacles put forth when starting a data warehouse. As this white paper has shown, there remains a need for companies to seriously consider outsourcing data warehousing. The outsourcing process should be applied with the understanding that it is an exercise to determine how to best source a function or service under scrutiny. The answer presented here was to outsource data warehousing to an experienced provider of end-to-end solutions, such as the USi Rapid Data Mart Powered by Sagent.

Copyright USinternetworking, Inc. 1999

All Rights Reserved