CHAPTER 1

Gaining Data Warehouse Success

You've heard about them. You may have used one. You may be asked to pay for one. But what is a data warehouse and why should you invest any time, energy, and money on one? The short answer is that a data warehouse can help your organization to be more profitable, run more efficiently, and meet the challenges of today's marketplace. Yet it is not a quick, simple, or inexpensive undertaking to build a data warehouse.

There is often a disconnect between the technical side that builds and maintains a data warehouse and the business side that will use it. This book will help bridge that gap. Both business managers and IT managers will learn what is involved with building and deploying a successful data warehouse. Executives and senior managers will also find this book helpful, especially Part 1, in order to be able to provide effective oversight and support. This book will also be beneficial for all business and technical personnel involved with a data warehouse, providing a common foundation for better communication. Managers on both sides need the knowledge and information that will enable them to help their organization build and use a data warehouse most effectively, and this book is the path to that knowledge.

This chapter explains the value of a data warehouse and highlights what is needed for success. To help frame the discussion, the chapter begins with some definitions.

The Essentials of Data Warehousing

Data warehousing is not new. Most large organizations have been investing in data warehousing for years. Currently, cost-effective technology is creating more possibilities for small and medium-size companies to build and deploy data warehouse solutions too. There are many stories about wild successes, and just as many about failed projects. With so much buzz about data warehousing, it is often assumed that everyone already knows the basics. However, many

people are being exposed to these concepts for the first time. To ensure a common understanding, it is worth taking the time to boil things down to the essence of data warehousing.

What Is a Data Warehouse?

A *data warehouse (DW)* is the collection of processes and data whose overarching purpose is to support the business with its analysis and decision-making. In other words, it is not one thing per se, but a collection of many different parts. Before looking more closely at the specific parts of a data warehouse environment, it is helpful to compare the characteristics and purpose of a data warehouse with an operational application system.

Differences Between Operational and DW Systems

Applications that run the business are called *online transaction processing systems (OLTPs)*. OLTP systems are geared toward functions such as processing incoming orders, getting products shipped out, and transferring funds as requested. These applications must ensure that transactions are handled accurately and efficiently. No one wants to wait minutes to get cash from an automated teller machine, or to enter sales orders into a company's system.

In contrast, the purpose and characteristics of a data warehousing environment are to provide data in a format easily understood by the business community in order to support decision-making processes. The data warehouse supports looking at the business data over time to identify significant trends in buying behavior, customer retention, or changes in employee productivity. Table 1-1 lays out the primary differences between these two types of systems.

The inherent differences between the functions performed in OLTP and DW systems result in methodology, architecture, tool, and technology differences. Data warehousing emerged as an outgrowth of necessity, but has blossomed into a full-fledged industry that serves a valuable function in the business community.

Now that the differences between data warehouse and OLTP systems have been reviewed, it is time to look deeper into the makeup of the data warehouse itself.

The Data Warehousing Environment

There are many different parts of a data warehouse environment, which encompasses everything from where the data lives today through where it is ultimately used on reports and for analysis. Each of the main parts of the data

warehousing environment, shown in Figure 1-1, are described in the following sections. This figure indicates how the data flows throughout the environment.

Table 1-1 Comparison of Online Transaction Processing and Data Warehousing Systems

AREA OF COMPARISON	OLTP	DATA WAREHOUSING
System purpose	Support operational processes	Support strategic analysis, performance, and exception reporting
Data usage	Capture and maintain the data	Exploit the data
Data validation	Data verification occurs upon entry	Data verification occurs after the fact
Update frequency	Data is updated when business transactions occur (e.g., client uses debit card, web order is placed)	Data is updated by periodic, scheduled processes
Historical data requirement	Current data	Multiple years of history
Data integration and balancing	Data is balanced within the scope of this one system	Data must be integrated and balanced from multiple systems

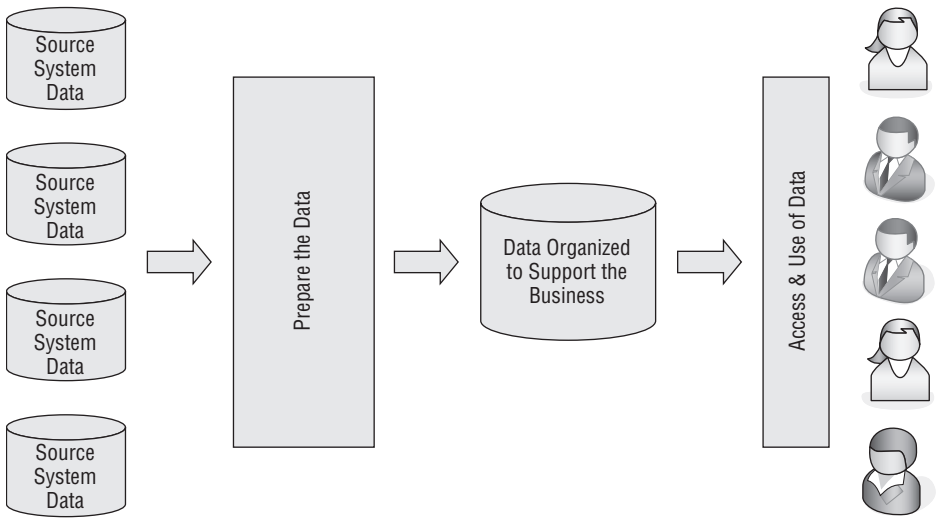

Figure 1-1 Basic data warehousing environment

Source systems, shown on the left side of Figure 1-1, are where data is created or collected by operational application systems that run the business. These are often large applications that have been in place for a long time. Examples of source systems include the following:

- Order processing
- Production scheduling
- Financial trading systems
- Policy administration
- Claims handling
- Accounts payable/receivable
- Employee payroll

The entire midsection of Figure 1-1 is devoted to the preparation and organization of data. First, the data must be extracted from the source systems. Next, the data needs to be transformed to prepare it for business use. It must be cleansed, validated, integrated together, and reorganized. Finally, the data is loaded into structures that are designed to deliver it to the business community. The entire process is referred to as the *extract, transform, and load (ETL)* process.

The database in which the data is organized to support the business is called a *data mart*. A data mart includes all of the data that is loaded into a single database and used together for analysis. Data marts are often developed to meet the needs of a business group such as marketing or finance. The key to a successful data mart is to create it in an integrated manner. It is also recommended that data be loaded into only one data mart and then shared across the organization to ensure data consistency.

Finally, an application or reporting layer is provided to facilitate access and analysis of the data. This is where business users access reports, dashboards, and analytical applications. Collections of these reports and analyses are called *business intelligence*.

There is one more critical concept that warrants some attention: the mechanism used to help organize data, which is called a *data model*.

What Is a Data Model?

A data model is an abstraction of how individual data elements relate to each other. It visually depicts how the data is to be organized and stored in a database. A data model provides the mechanism for documenting and understanding how data is organized.

There are many different types of data modeling, each with a specific goal and purpose. As organizations modified how data was structured to

support reporting and analysis, a new data modeling technique, now called *dimensional modeling*, emerged. Ralph Kimball, a pioneer in data warehousing, can be credited with crystallizing these techniques and publishing them for the benefit of the industry. (For more information about dimensional modeling, see Chapter 7.)

The data and processes to perform the work shown in Figure 1-1 are collectively called the data warehouse. These basic concepts have been fine-tuned and relabeled by many different players in the data warehousing field. The two most common philosophies are discussed in the next section.

Understanding Industry Perspectives

At the end of the day, everyone faces the same challenge: getting the data into the hands of the business user to turn it into information that can be used to make decisions. The definitions provided so far provide the backdrop for how terms are used in this book. There are many brilliant and talented people in the data warehousing industry, many of whom have different philosophies about how to design and build a data warehouse. It is worthwhile to look more closely at two of the most frequently used philosophies.

The first is from Ralph Kimball and colleagues, as described in *The Data Warehouse Lifecycle Toolkit, Second Edition* (Wiley, 2008). Ralph Kimball is a clear thought leader in the data warehousing industry and has written several books that provide detailed information essential for practitioners. That book describes the enterprise data warehouse as:

> *The complete end-to-end data warehouse and business intelligence system (DW/BI System). Although some would argue that you can theoretically deliver business intelligence without a data warehouse and vice versa, that is ill-advised from our perspective. Linking the two together in the DW/BI acronym reinforces their dependency. Independently, we refer to the queryable data in your DW/BI system as the* enterprise data warehouse*, and value-add analytics as BI (business intelligence) applications.*

A second definition worth looking at is from Bill Inmon, a prolific author and another leader in the data warehousing industry, from *Building the Data Warehouse, Fourth Edition* (Wiley, 2005):

> *The* data warehouse *is a collection of integrated subject-oriented databases designed to support the DSS (decision support system) function, where each unit of data is relevant to some moment in time. The data warehouse contains atomic data and lightly summarized data.*

Bill's definition is also described and expanded by Claudia Imhoff and colleagues in *Mastering Data Warehouse Design* (Wiley, 2003):

> *. . . It [the DW] is the central point of data integration for business intelligence and is the source of data for the data marts, delivering a common view of enterprise data.*

This second viewpoint is incomplete without also including their definition of a data mart. Again, expanding on Bill Inmon's definition, Claudia states in *Mastering Data Warehouse Design*:

> *A* data mart *is a departmentalized structure of data feeding from the data warehouse where data is denormalized [organized] based on the department's need for information. It utilizes a common enterprise view of strategic data and provides business units with more flexibility, control, and responsibility. The data mart may or may not be on the same server or location as the data warehouse.*

To bring this second viewpoint into the proper context, *Mastering Data Warehouse Design* further defines business intelligence:

> Business intelligence *is the set of processes and data structures used to analyze data and information used in strategic decision support. The components of Business Intelligences are the data warehouse, data marts, the DSS (decision support system) interface and the processes to 'get data in' to the data warehouse and to 'get information out'.*

The single definition provided by Ralph Kimball is comprehensive. You must look at the full set of definitions set forth by Bill Inmon and Claudia Imhoff to fully understand their perspective. There is much more common ground between these differing philosophies than perceived at first glance. While there are distinct differences, the common theme is that data warehousing must provide the method to prepare and deliver data to the business community to support reporting and analysis. Chapter 9 provides a comprehensive discussion about these different approaches to data warehousing.

The key point here is that there are multiple ways these terms can be interpreted. Understanding which definition is being used is critical to being able to understand what is being discussed and worked on. An organization can avoid confusion by selecting one set of definitions to be used, which enables everyone to use a common language.

Regardless of labels and terminology, all data warehouse initiatives are trying to accomplish the same thing. Now that the basic parts of the data warehouse have been defined, it is time to look at the order in which they are created.

Design and Development Sequence

Earlier in this chapter, you looked at how data flows through the data warehouse environment. While this correctly illustrates how data flows in the

completed environment, this is not the recommended sequence for designing and developing a data warehouse. A better way to design the environment is to start from the business user perspective. Figure 1-2 shows the correct order to successfully design and implement a data warehousing environment. Both the technical and business team members play a role throughout. Chapter 4 describes the different roles and responsibilities. Each step in the design process is described as follows:

1. An understanding of what the business is trying to accomplish and how success is measured should be the foundation for all data warehousing initiatives. The starting point for designing the data warehouse is with the business community. Chapter 6 covers what you need to know to effectively provide business requirements.
2. Once the business requirements are understood, the data in the underlying source systems needs to be studied. Many business people have a vision for what they want to do, but it is not always tied to the reality of the organization's actual data. In preparation for modeling data, Chapter 7 introduces techniques to help you understand your data.

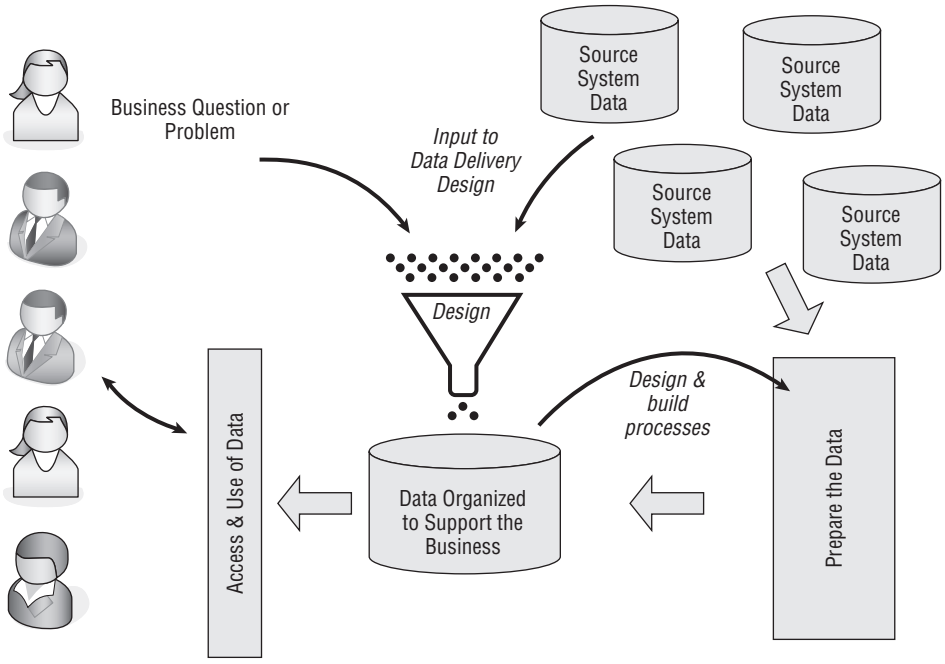

Figure 1-2 Optimal data warehouse design and development sequence

3. The foundation for successful data warehousing, now and into the future, is properly structuring the data. Data must be organized to support the business perspective. This provides ease of use and improved query performance. This design is created based on a knowledge of the business requirements, as well as the reality of the existing data. Chapter 7 focuses on how the business and technical team members can work together to develop appropriate data models for this data delivery layer.
4. After defining how the data will be organized, the design for getting the data from the source systems to the database can be created. Decisions about the architecture and tools needed to prepare the data can be made in the proper context. Too often these decisions are made before you know what is to be delivered. Chapter 9 provides the background needed to understand data warehouse architecture and technology. Chapter 10 deals with the challenges of preparing the data for business reporting and analytical use.
5. While the data is being prepared, the data access and application layer can be designed. This includes the design of basic reports, business intelligence, and analytical applications, and performance dashboards or other end user tools. Chapter 11 focuses on final delivery of data to the business community.

Many different project methodologies are available for all systems' development efforts. There are even multiple methodologies specifically targeted toward data warehousing. These have evolved over several decades. Most organizations already have adopted some type of project methodology or project life cycle. It is important to understand how your organization runs projects to ensure that the data warehouse project is adhering to the strategic direction for all information systems. Several basic building blocks are found in any methodology. These primary components, discussed throughout the book, are as follows:

- **Project definition, planning and management** is outlined in Chapters 4 and 5.
- **Defining business requirements** is discussed in Chapter 6.
- **Designing the data delivery database** is covered in Chapter 7.
- **Defining the architecture** is discussed in Chapter 9 and includes development of the database
- **Processes for building the database** are reviewed in Chapter 10.
- **Developing reports/analyses and providing education/support** is presented in Chapter 11 and includes deployment of the final results.

These basic components need to be in place regardless of the chosen methodology, approach, philosophy, or technology. The sequence outlined earlier helps to ensure overall success by tying all other activities to business requirements. This sequence also helps build a foundation that will withstand the test of time.

Why Build a Data Warehouse?

Since the first application systems were built, businesses have struggled with the proliferation of data. Systems have been implemented to automate business processes such as order processing, policy administration, shipping, and manufacturing scheduling. Each of these applications captures a lot of data. This data has the potential to be used in several ways:

- To understand what is really happening in the business
- To identify historical trends
- To predict future opportunities
- To measure performance

Originally, business requests for data were run directly against production application systems. This was not an efficient way to gather data. For example, a request for data might query which days over the last three years have the highest ATM transaction rates in the Midwest. Unfortunately, in order to gather this data, the data query slowed—or worse, interrupted—the ability of the application systems to perform their primary function. It is unacceptable to reduce the speed with which the ATMs complete transactions, even to support the corporate marketing department. Needless to say, the ability to run these reports against the production application systems was revoked.

However, understanding the results of such a request was valuable, so a copy of the production data was made to enable reporting and basic analysis to be done without affecting the actual operational systems. When working with a copy, modifications were made to how the data was stored in order to make reporting easier. In addition, if historical data was not retained in the operational systems, then the history was kept in the ''copy''. New system design techniques were developed to improve how the copy of the data was created and organized. Specific hardware, software tools, and technologies also emerged. These have grown into what we know as the data warehousing industry today.

In spite of these technical advancements, organizations still struggle with some basic data issues:

- **Direct data access:** You can't directly access the data; you have to request it from a data systems department.

- **Data usefulness:** You can't get the data you need; what is collected is not useful for your purposes.
- **Poor data quality:** The data, if and when you get it, is often not correct. You can't rely on it.
- **Facilitating exception reporting:** There is too much data; you have to dig through it to find what's important.
- **Timeliness of data:** By the time reports get to you, the window of opportunity no longer exists.
- **Flexibility:** You need total flexibility to look at data from a variety of perspectives.
- **Data integration:** You need to perform analysis across all lines of business but the data is stored all over, so you can't easily compare it.
- **Silo reporting environments:** Different departments use different reporting tools and you can't share reports.
- **Unclear definitions of data:** Everyone has their own report, but the figures do not match and you don't know why.

These basic data problems that drove the initial work in the data warehousing industry are the same core problems facing businesses today. If anything, the problems are compounded by ever-increasing mountains of data. While addressing basic data-related issues is certainly a big part of the rationale to build a data warehouse, there are also bigger, more critical business issues that need to be addressed. The data warehouse can provide a vehicle to deliver the information needed to help a business address these issues, which is what drives the true value.

The Value of Data Warehousing

Data warehousing is much bigger than simply delivering reports in a timely manner. It is not the data, the technologies, or the reports that impact the business. Rather, it is the ability of your staff to harness the information to make better, fact-based, insightful decisions. The data warehouse is simply a tool that enables your staff to be more effective. The types of things that can be done using a data warehouse include the following:

- **Tracking and trending key performance indicators:** A data warehouse can provide reports that indicate which product lines are popular in various regions, which employees have generated the most sales, or whether certain ad campaigns are correlated with successful sales.
- **Measuring business performance:** Using reports from the data warehouse, actual and forecasted performance can be compared. For example,

claims managers can see how close they are to reaching the target of making a first payment on claims within the first ten days of opening a new claim.

- **Reporting and understanding financial results:** A data warehouse can help identify departments that have exceeded their monthly budgets, highlight suppliers who have consistently met profitability goals, and single out products that have contributed the most (or least) to the bottom line.
- **Understanding customers and their behavior:** Exception reports that highlight changes in consumer purchase patterns can help identify shifts in the marketplace or erosion of brand loyalty. For example, early identification of changes in payment patterns might indicate that a customer is under financial pressure and could benefit from a courtesy call to prevent more serious problems.
- **Identifying high-value customers:** Using the data warehouse to identify the lifetime value of customers helps with the development of loyalty programs and improves customer service. Some customers may generate many business transactions, but they may not actually be profitable. Other customers may contribute consistently to the organization's profits without a lot of hands-on interaction or support.
- **Attracting and retaining high-value customers:** Data warehouse reports can help you to develop a profile of high-value customers so that initiatives can be created to seek out new customers with a similar profile. This may mean offering low-cost incentives early on so that the organization has the opportunity to develop a strong long-term relationship.
- **Better selection or development of new products:** Having integrated data in a common place, the data warehouse can help streamline the product development process by enabling all groups involved to quickly access market research test results, product packing cost scenarios, and projected product sales.
- **Understanding which products should be scaled back or eliminated:** Using the data warehouse, reports can be generated to highlight products with lagging sales. Additional analyses can be run to determine the cost effectiveness of continuing to carry these items in stores. The data warehouse can also be used to help develop plans identifying when trendy items should be marked down to clear out any remaining inventory.
- **Understanding business competitors:** The data warehouse can provide reports to compare internal sales volumes with external competitor sales

figures. This can help identify fluctuations in the overall marketplace and how well the organization is maintaining its market share.

- **Identifying opportunities to improve business flow and processes:** The data warehouse can be used to track how business transactions flow within the organization to identity bottlenecks, the need for more training, or when systems capacity can no longer keep up with demand.
- **Understanding the impact of highly qualified professionals:** A data warehouse can also be used effectively in not-for-profit scenarios. For example, data warehouse reports can help identify teachers who meet specific criteria and to track how a teacher's students perform on educational assessments over time.

These uses apply across many different industries. For example, industries that have realized data warehousing success include the following:

- Consumer packaged goods
- Financial services
- Manufacturing
- Utilities and telecommunications
- Pharmaceuticals
- Insurance
- Healthcare
- Shipping and transportation
- Educational institutions
- Nonprofit organizations

Furthermore, data warehousing has been successfully deployed across a wide variety of business functions, such as sales, marketing, finance, purchasing, manufacturing quality, human resources, inventory management, customer relationship marketing, call centers, and more.

Providing timely, reliable data enables business professionals to make informed decisions. It is not enough to provide a static report; the information must be flexible enough to make possible the identification of anomalies, so that interesting results can be further explored. This easy exploration enables business users to get to the bottom of problems, allowing corrective action to be taken quickly. Sometimes the data demonstrates unexpected positive results; these also warrant further research to determine whether they indicate extenuating circumstances or whether the results can be repeated.

The return on any individual decision may be small. A lot of small, better decisions can add up to a significant impact on the bottom line. While there have certainly been cases, often highly publicized, where a single decision

resulted in multimillion-dollar results, it is more likely for a business to see benefits from many small decisions. The data warehouse provides the keys that unlock the data's value.

There are several common expectations regarding what a data warehouse should be able to do. Many organizations fall short of meeting these expectations, so it is worth taking some time to understand these expectations and what needs to be in place in order to achieve them.

The Promises of Data Warehousing

Since its inception, data warehousing has offered the promise of helping to improve your business. A data warehouse is expected to provide both of the following:

- A single version of the truth
- The capability to access all data whenever it is needed

Unfortunately, many organizations have invested millions of dollars in data warehousing without realizing either of these goals. There are many reasons for these struggles and failures, many of which are addressed in this book. First, it is worth looking a little more closely at these expectations.

The idea of having a single version of the truth is appealing because so much time and energy is wasted in chasing down discrepancies. It is reasonable to want to trust report results so that decisions based upon those reports are sound. A data warehouse can indeed provide a single repository for all of your data, but that alone is not enough to ensure that all reports will be consistent. Clean, trusted data from the data warehouse is often pulled out, further manipulated, possibly loaded into yet another database, and finally presented on reports. The use of different formulae for calculations, and the criteria used to include or exclude data from the result set, can cause significant differences in what is shown on a report. The effort of loading data into a data warehouse is not enough to fulfill the promise of a single version of the truth. This requires organizational discipline and a commitment to data management, which is discussed in detail in Chapter 8.

The second big promise of data warehousing is that any data that is desired will be available at your fingertips whenever it is needed. Indeed, a data warehouse can make data more accessible to many different types of users across an organization, but it is too expensive to load all of the company's data into a data warehouse. The audience and business impact of some data is not significant enough to justify the expense of including it in the data warehouse. An organization must determine what data is needed to help the business decision-making processes. Then, the most useful data can be loaded

into the data warehouse over a period of time, perhaps even a number of years. Often this is viewed as the final goal: The data is available in the data warehouse.

However, data in a database does not automatically mean that it is accessible to the business community. The data must be made available through reports, dashboards, or analysis tools that are combined with appropriate education about how to leverage the data as part of the day-to-day decision-making processes.

While sound database design and the use of technology are important, many other factors need to be addressed in order to achieve success. In the next section, you'll learn about the most common roadblocks to this success, and how to remove them.

Keys to Success

There are several key factors for success in building a data warehouse. These factors are relevant regardless of the industry or size of company, and have remained constant over the years. As mentioned earlier, some organizations have invested millions of dollars to obtain the best of everything, and yet struggle to deliver any value to the business. Conversely, some organizations have built a data warehouse with little more than bailing wire and an abacus and have achieved great success. While technology and tools certainly play a role in building a data warehouse, they are clearly not the primary factors for success or failure.

PRIMARY FACTORS FOR SUCCESS

Although several factors contribute to a successful data warehouse, two are primary:

- **A strong partnership between the business and systems communities**
- **Ensuring that the data warehouse project is driven by true business requirements**

These two major success factors have more to do with people than technology. Take a look at any project plan. How many tasks are focused on communication—that is, people—compared to technical tasks? These technical tasks are indeed important, but specific actions to ensure meaningful communication are critical to the project's success. All other efforts can amount to nothing if the business community and management are not committed.

Developing and Maintaining Strong Business and Technology Partnerships

It is one thing to affirm the importance of a strong partnership between the business stakeholders and the technology units involved in a data warehouse project, but it is another thing to put that into practice.

Partnerships often start out strong in the beginning. There is joint participation in developing a project's scope and objectives. A business sponsor may be designated and may help launch the project. This is where the successful partnerships begin to emerge. Once the project is underway, however, this is *not* the time for the business community to sit back and wait for the next briefing from IT. Now is the time to dig in and help the project move forward. This includes gathering true business requirements and assisting with understanding the data and helping to make decisions along the way.

Staying involved in a project on a daily or weekly basis ensures that there is a good understanding of what is happening on a project. This regular communication also helps to identify potential problems and get the appropriate help in a timely manner. Detailed involvement helps to ensure that decisions are made within the proper business context. What may seem impossible to a technical project team may be easily accomplished by a business representative, and vice versa.

In addition, it is easier to garner support from the rest of the business stakeholders when one of their own comes to them with questions or the need to clarify something. Too often, a technical or systems representative may not be allowed the same level of access or may not be able to ask questions in a way that is meaningful to the business. By working together, the project team members can put together a strategy to get the business input that is necessary to drive the data and technical decisions. A strong partnership that includes a lot of mutual involvement also builds ownership of the data warehouse across the business group. This is important for adoption and use of the data warehouse when it is deployed.

REFERENCE **Chapter 4 provides more details about how to develop and maintain strong partnerships.**

Identifying True Business Requirements

True business requirements are not a list of data elements, data sources, or even reports. These requirements must be more fundamental to the business itself. For example, the need to better measure marketing campaign performance, understand and better manage loss ratios, and understand and track student

performance are all business reasons that explain the need for gathering data and generating a report, in contrast to a demand to ''give me the numbers.''

These clearly defined business requirements also tend to stand the test of time. The specific data elements may change, but the need to monitor competitor performance will always be important. These requirements also remain fairly constant unless an organization completely changes business industry.

REFERENCE **Chapter 6 provides more details about how to collect, understand, and communicate business requirements.**

Shifting to a Global Perspective

Many organizations structure their employee performance expectations and goals on the results of that specific group. This creates an environment where the staff members look out for their own interests, sometimes at the expense of the greater good. This culture makes it more difficult to successfully implement a data warehouse. The ability to shift the emphasis to a more global perspective increases the success of the data warehouse. Rewarding and encouraging employees, business or technical, who are looking for opportunities to collaborate, leverage, and integrate across the organization is important. The larger the organization, the more challenging this can be.

High-level managers need to look for a balance between individual department goals and enterprise goals. This is done to set priorities and direction for the organization, and happens annually during the budgeting process. This same thought process can be applied to the data warehouse environment. Looking for the largest benefit for the most number of people across the organization can determine what data is loaded is loaded first and which reports are developed. This broad perspective is also needed to create standard data definitions and calculations.

SUCCESS TIPS: ENCOURAGING AN ENTERPRISE VIEWPOINT

It may be necessary to modify individual job descriptions and annual performance goals to encourage behavior that is beneficial across the enterprise. Reward employees who fight for the greater good. Many individual contributors on a data warehouse project can recognize myopic decisions, but there may not be a way to raise these issues. Strive to provide a mechanism for raising concerns and identifying problems without fears about retribution. This can open up a dialogue that improves the decision-making process and encourages taking broader requirements or benefits into consideration.

Overcoming Unrealistic Expectations

It is unrealistic to think that many years' worth of data problems and challenges can be corrected in a matter of weeks. In many organizations, addressing fundamental data problems can actually take years. While setting forth a vision for the future, it is critical that realistic expectations are set for each individual project along the way. Informal expectations exist whether you explicitly set them or not. It is important to know what employees expect a project to provide, how much it will cost, and how long before the data warehouse will be available for use. Project expectations must be compared to what is actually planned. Organizations that regularly monitor these expectations can identify gaps and can take action sooner. This does not mean that other development problems go away or that project dates never shift; it simply means that there is a common understanding of what is happening and why.

Although there is always interest in getting things done faster, a business must be patient in order to get the right things done. Although it is a cliché, there are times when simply adding more resources will not result in faster completion of a task. You cannot assign nine people and expect a baby to be born in one month.

SUCCESS TIPS: MANAGING EXPECTATIONS

It is crucial that what a specific project will and will not deliver is documented and communicated. This *project scope* must be in business terms and should describe the business benefit. The scope needs to be communicated often and modified as appropriate. Moreover, this communication must be delivered across the entire project team, the target user audience, and associated business and systems management.

Another aspect of setting proper expectations is specifying a realistic budget and timeline for the project. Apply experience and knowledge of the organization to estimating data warehousing projects. If all other systems development projects are being completed in 10% less time than industry averages, then this might also be true for data warehousing. If the organization bureaucracy adds 15% overhead to industry average budgets or timelines, then adjust accordingly. Clearly, experiences with other data warehousing projects across the organization also need to be considered. This enables others in the organization to leverage what has already been learned.

Initial project expectations and time schedules may shift throughout the life of the project. Successful organizations communicate changes in terms of impact to the business. This helps ensure that there is a common understanding of both what the project will be delivering and when.

Providing Clear Communication

The need for regular open communication is critical to the success of a data warehouse project. This involves ensuring that there is a regular forum to discuss progress and problems. Successful communication is not simply conducting project status meetings and sending out project status reports. Those are indeed important communication vehicles, but the intended audience for those is the project team and project management personnel. Regular communication provides the best vehicle for questions and concerns to be raised and then addressed. This helps expectations to be managed.

The communication must be open to be helpful. This means that teams are encouraged to share both good news and bad news. Openly admitting that a problem exists provides the opportunity for others in the organization to help provide solutions.

REFERENCE **Chapter 4 provides tips for how to ensure effective and timely communication about data warehouse initiatives.**

SUCCESS TIPS: EFFECTIVE COMMUNICATION

Be sure to employ a multi-layered approach to communication. Organizations don't have clear communication by accident; they develop specific communication plans and then follow through. These plans need to include an approach to different levels of the organization, including senior management, project sponsors, systems management, and the business groups who are the targeted users of the data warehouse. The frequency and format for communication to each of these audiences is different.

Senior management may only need quarterly updates in an executive briefing format. Middle management, from both the business and the systems community, may need a more detailed form of an executive briefing. It is helpful to have these on a monthly basis. This is often done by including a brief time slot in regularly scheduled staff meetings. Monthly updates for the target user group may also be enough. Again, the opportunity to share progress can be easily included in regular staff meetings.

NOTE **It is critical to continue regular communication when there is no significant progress to report. The fact that work has slowed and milestones have not yet been reached is just as important as sharing significant progress. Reporting that hard work continues and that the next milestone is expected to be met by a specific date is useful.**

Treating Data As a Corporate Asset

The realization that data is more than what is collected and stored on a computer is just the beginning. Successful organizations realize that while the technical staff provides the mechanics for collecting and storing data, it is the business that really drives the meaning and use of data. Data is owned by the business itself. This drives the business to take a hands-on role in what data means, where it is collected, and how it is used.

For example, the business must determine the criteria regarding who is considered a new customer. Is it when a new customer identifier is assigned in the order entry system? Is it when a customer is issued a new credit card or when that customer uses the card for the first time? Perhaps a customer is also considered new when he or she makes a purchase after a twelve-month lack of purchases. Even seemingly simple concepts require a bit more thought and understanding. Organizations that truly leverage their data do not leave these types of decisions to the technical staff, but work together to clarify and store useful data. Treating data as a true corporate asset is much broader than the data warehouse. This encompasses all systems across the organization.

REFERENCE **Chapter 8 provides more details about what is involved in handling data as the asset it can be, with the emphasis on what this means for the data warehouse.**

Effectively Leveraging Technology

The selection and use of technology is usually very much a focus for data warehousing. Purchasing and maintaining hardware and software are some of the most concrete costs of a data warehouse project. There are a myriad of options for every component. Even after a specific vendor or technology is selected, there are more options for different levels of capability and/or configuration. Most companies have sound practices in place for the evaluation, selection, and implementation of technology. Operational applications are very much defined by construction and then deployment using a set of technologies. Successful data warehouses are implemented by investing enough in technology to empower the technical staff to be efficient in both development and ongoing maintenance and enhancements. The same investments must also ensure effective business use of the data warehouse.

The key here is that buying the least (or most) expensive technology may not be effective for your organization. Having a vision of what needs to be accomplished for the business helps drive decisions about technology.

REFERENCE **Chapter 9 contains suggestions for finding the right technology to meet your requirements.**

SUCCESSFUL TIPS: DEPLOYING TOOLS SUCCESSFULLY

When implementing technology, many organizations purchase the best-of-breed tool, but then cut corners when setting it up. Squeezing a sophisticated piece of technology onto an existing computer server is not likely to allow that tool to work effectively. It may even limit the functionality and will most certainly degrade the performance. Don't try to save money by configuring the minimum supported environment. Look for the recommended configuration that will deliver the results you desire. If the technology does not work well, the effort to build the data warehouse may have been in vain.

Roadblocks to Success

Many organizations have failed in their data warehousing efforts. Some have struggled for years, while others have failed in a big and highly visible way. Some data warehouses have not failed outright, but have never achieved their full potential. While each case has unique characteristics, several common themes regularly contribute to failure or the lack of total success.

Believing the Myth: "If You Build It, They Will Come"

There are still many companies who invest in building a data warehouse by pulling data from core operational systems into a common database, expecting that the business will then begin to use it. This is a good exercise in data movement and organization, but an approach that rarely leads to success. The data is organized in ways that are not tied to actual business needs and requirements. This is an approach that seems 'safe' from a systems perspective. All design and development work is done without venturing into the business community. The belief is that once the data is organized, the business can be consulted to determine how to deliver it to them.

This approach does not take into consideration what the business community is really doing and wants to do in the future. It is also limited to the way that data is captured and stored in the underlying operational systems. The reality of the business often differs from what is captured. Data is grouped and reported in ways that may not be reflected in the underlying systems.

Such an approach also requires a large investment in infrastructure and building this database before there is any link to business value. This was once a very popular approach to data warehousing, but it is no longer a leading one. Unfortunately, many organizations are already well on their way to building this type of data warehousing environment.

RECOMMENDATION **It is never too late to gather business requirements, a task that can identify the data needed to deliver specific business analyses and reports. Adding new data to the data warehouse and possibly restructuring and/or enhancing the existing data can derive value.**

Falling into the Project Deadline Trap

Many organizations have invested in setting up procedures and processes that follow a project methodology. Clearly, an important part of the process is defining a project's scope. Detailed project plans are developed and used to measure progress. Group and individual performance are often based upon the ability to meet project deadlines. While project planning and management are critical to coordinating multiple resources and keeping projects moving forward, this focus on the project plan itself can cause data warehousing initiatives to go awry.

The time needed to extract, transform, and load data for analytical applications (see Chapter 10) is one of the least understood and most often underestimated tasks of any systems development project. This often leads to project delays. One common way to manage these potential delays is to reduce the scope of what is to be delivered. This can lead to finally delivering something on time, but the scope has been pared down so that no business value is delivered. In other cases, testing and quality checks are reduced or eliminated in order to maintain the project schedule. Unfortunately, many organizations reward this type of behavior. Delivering something on time is a prevailing theme.

RECOMMENDATION **Reward project teams for producing quality results. This includes validated data and well-tested applications. This also means that applications indeed support the business in a valuable way. It must not be acceptable to reduce the scope so that little or no value is delivered. Part of this change in deadline focus also means that project change requests are viewed in a positive light. Project issue logs should be full of open items needing to be addressed. Teams should not be penalized for being honest about progress.**

Failing to Uphold Organizational Discipline

Despite an impression that data warehousing is glamorous, most of the work to build one is not. The type of work needed to develop and implement a successful data warehouse is detailed and painstaking. Project tasks are often

tackled with excitement, but as the realities of the project sink in, enthusiasm begins to wane. Data analysis may sound fine in a project plan, but having to crawl through hundreds of data fields is time-consuming, and can be difficult as well. Systems professionals have an aptitude and interest in this type of detailed data work. Business professionals rarely do—after all, if they did they would probably have become systems professionals!

The primary areas of effort for a data warehouse revolve around the data itself. When thinking about data, most people think about the technical aspects—where the data is stored, what computer programs may access or use it and who to call in systems if you have a problem. The technical care and feeding of data is only a part of the picture. It is also important to understand the business definition of each data element, why it matters to the organization, and how the data should or should not be used. Resources must be assigned to work on these business aspects of data. This is where true discipline comes into play. It is much easier to simply move data around than it is to figure out what data is needed and then determine where to find it. This discipline is needed not only for the initial project, but also going forward into the future. The need to develop and adhere to good data practices never ceases.

Organizations that have not been successful with other initiatives that require long-term discipline need to be up-front about this. Strategies must be developed to ensure long-term discipline. A lack of data discipline does not mean failure, it simply means that the business must be willing to accept and use data as it is. This may be sufficient to support today's business needs.

RECOMMENDATION **Adequate resources need to be allocated to build and maintain a sustainable data warehouse environment. Many organizations allocate enough systems resources for project development but neglect to allocate business resources. Business representatives must be responsible for providing guidance, answering questions, and resolving data problems during a data warehouse project. Additionally, to ensure proper organizational discipline, business processes must be put in place to address long-term maintenance of data definitions and data preparation rules. In addition, resources must be allocated to support and address ongoing business data responsibilities.**

Lacking Business Process Change

The impact that a data warehouse can have on an organization goes far beyond getting reports faster. Automating what is done today does not dramatically change business results. It is much more valuable to change how things are done. The data warehouse can be a catalyst to help make these changes. The data warehouse should be used to provide historical information and to

highlight the most important things that need attention today. This can be done at many different levels within the organization.

For example, the executive team might see a sharp increase in claims activity and evaluate the need for setting up additional financial reserves. The claims managers might need to review staffing plans in order to keep up with the increase in activity. The actuaries might need to perform analysis to identify underlying causes for the increased claims activity (unless it is a result of a natural disaster). The underwriters might need to reevaluate which policies are issued and renewed to ensure that the company continues to maintain financial stability . On an individual level, the data warehouse can also be used to have customer service representatives proactively contact claimants to communicate claim status, especially if it is taking longer than expected to resolve a claim. All of these activities are done periodically as a natural course of running the business, but changing these processes to utilize the data warehouse can influence the timing and sequence of these tasks.

Unfortunately, many organizations simply use the data warehouse as a glorified report generator. While it is true that the data warehouse can provide more accurate, integrated, and timely reports, significant long-term benefits are derived when report results and analyses drive changes to how business is conducted.

RECOMMENDATION **Use of a data warehouse can be daunting, especially for people who have not been able to get to data themselves in the past. Staff members can feel intimidated and unprepared to meet the new expectations. Appropriate education must be provided to ensure that the group has the skills necessary to use the data warehouse. This education must include both the mechanics of using the warehouse and an understanding of how to use or analyze the data effectively.**

Care must also be taken when making changes to business processes. When developing process changes, it is helpful to include employees, who often have great ideas and insight that cannot be found elsewhere. This inclusion also creates a more open attitude for embracing the changes. Business process changes may not be done in a single sweeping effort, but evolve over time, as use of the data warehouse becomes a regular part of daily work.

Narrowing the Focus Too Much

Because a broad goal of loading all of the data and then looking for opportunities to improve is rarely successful, many organizations have wedded themselves to the other end of the spectrum. With this approach, the requirements for a project are defined in a concrete, specific manner to deliver a set of key performance indicators, or perhaps the 25 most important data elements

have been identified. While this can narrow the scope of a project, it can also lead to significant challenges in the future.

With this type of narrow focus, the emphasis is indeed on the business, which is good, but often only the data specified is extracted and loaded. After looking at the requested data, it becomes obvious that some of it is as expected, some of it is not what was intended, and other data may not yield any insight even though it is was requested. This results in the need to change the data that is needed, which requires going back to the beginning.

Another common problem with too narrow a focus is apparent when follow-up questions are asked. Often, data is requested at the summarized level to represent an entire company or group. When the more detailed questions are asked to better understand why a certain number is high or low, the data is not available. This was a common downfall of the executive information systems that were popular in the 1980s. Although the development and maintenance of performance dashboards is much easier, the same potential pitfall exists.

RECOMMENDATION **Because both too broad a focus and too narrow a focus on the business is bad, there must be a middle ground that is more successful. The best balance is achieved when the underlying data needed to create performance indicators is loaded in the data warehouse and made available to support detailed analysis. Executives and high-level managers can drill down into the data themselves or rely on their employees to follow up on the more detailed analysis. Either way, the data is available to support the needed analysis.**

Another benefit of loading detailed data is that often very interesting and possibly valuable data may live alongside the specific elements requested. By pulling all useful business data, rather than just one field of immediate interest, the data warehouse supports multiple levels of analysis and future changes. The effort required to pull along these additional fields is often not significant. The most effort is needed to integrate even one element from each source, rather than multiple elements.

For example, it can be hard work aligning product codes from non-integrated systems, but the amount of effort needed to integrate the product codes is the same whether one uses one data element from each source or fifteen elements from each source. Given that the level of effort does not increase dramatically, why not bring forward other useful business data at the outset? To increase long-term benefits, it is critical that designs for data delivery and preparation *not* be limited to a specific set of business measures. The design of the data access and analysis components of the data warehouse can indeed focus on the specific requirements at hand.

Resting on Your Laurels

Some organizations have realized great business success for many years. The people in these companies have little incentive to change. The way that business has been done in the past still works well and continues to yield profits. As long as success continues, it is extremely difficult to effectively deploy a data warehousing environment. A similar culture may also exist in larger organizations, where individual employees may not feel directly connected to business results.

The process of developing a data warehouse requires that business rules regarding how data is processed must be reviewed and possibly changed. Rules and guidelines for integrating data from independent systems must be defined. Even the basic content and quality of existing databases is scrutinized. This is not easy, in any situation. When there is no real business reason to expend the effort, the chances for success dwindle.

Nonetheless, data warehouses can be effectively deployed in highly successful organizations when there is a culture and mindset to stay ahead of the competition.

RECOMMENDATION **Take an honest assessment of the need for business change. If the business can likely continue for the next three to five years as is, then perhaps the time is not right for data warehousing. If change is necessary to continue market strength or to adapt with the marketplace, then data warehousing is likely to help the organization. All data warehouse initiatives must be positioned as part of an overall strategy to take the company into the future. Data warehousing must be accompanied by other initiatives to help the organization change business processes and the culture. Individuals must see how their work contributes to the overall success of the company.**

Relying on the Technology Fix

Organizations often seek a simple solution to a set of very complex challenges. No matter how sophisticated technology becomes, it will not solve fundamental organizational and business issues. Purchasing and installing the latest technology will not overcome fundamental business process flaws. Technology does not solve data quality or integration problems either, but it can help a company address their challenges.

Businesses sometimes believe that swapping technologies will correct the problems of a data warehouse. If a data warehouse is floundering, the easiest thing to point at is a specific tool. Before changing technologies, however, it is worthwhile to evaluate the current environment. Newer versions of existing

software may resolve outstanding issues and provide additional functionality, but if the current version of a tool does not meet the organization's needs, then it may be worthwhile to survey the marketplace for other options.

Keep in mind that each tool has its own strengths and weaknesses. You may be trading in a known set of weaknesses for the promise of a new set of strengths without really understanding the new weaknesses. The cost to swap tools may be greater than the potential benefit of the new tool. Moreover, there are additional costs to retrain both technical and business users of a new tool; and still other costs associated with converting applications and uses of the existing tool.

RECOMMENDATION **Evaluate all aspects of the current data warehouse environment, not just technology. Keep in mind that technology alone does not address business, process, or data problems.**

The bottom line? Evaluate and understand all aspects of a struggling data warehouse before making drastic changes. Removing any single roadblock may not yield the desired results, as multiple roadblocks may need to be addressed with a comprehensive strategy.

Getting the Right People Involved

Once a data warehouse project is approved, the challenges of staffing the project team begin. Unfortunately, the best resources to help are usually already overbooked, so the available employees are assigned to the project. These people are often lacking the necessary experience or skills needed.

The most successful data warehouse projects are filled with the most valuable business and technical employees. These people have a lot of knowledge about the company, how it works, where data is stored, and the secret handshakes needed to gain access.

It is difficult to allocate these highly valued resources to a new project because they often play an integral role in the daily functioning of the business and other special initiatives, yet these are the people who help resolve crisis after crisis. These resources must be allowed to participate in the data warehouse project. By leveraging their knowledge and experience, the existing data problems can be addressed, enabling the data warehouse to serve as the foundation of the company's efforts to move forward.

Another common mistake is hiring new staff to fill the business and technical roles. While adding new staff members is beneficial in the long run, this can create a significant problem at first. New employees do not typically know enough about how the business is run or what data is currently used. Nonetheless, new employees do bring a wealth of ideas about what is possible,

and often have a realistic understanding of the difficulties in weaving their way around the data in its current form.

RECOMMENDATION **Even if the most experienced staff members cannot participate full-time on a data warehouse project, these valuable resources need to provide regular guidance and direction. From a business perspective, get other staff members to pick up some of the daily responsibilities for key personnel. This may mean brining in additional help—you can't expect everyone else in the group to go from 110% to 120% effort. From a technical perspective, it is important to keep your staff involved and informed about what the data warehouse is, and how and why it is built. It is risky to turn the entire project over to a third party because they will eventually leave with all the business and technical knowledge.**

At a minimum, hire contractors to backfill some of the daily operations work so that key technical personnel (DBAs, architects, and lead developers) are able to participate in the project. Another technical resource constraint is access to the key individuals who understand and intimately know the source system and data. The data warehouse team must have regular access to these individuals in order to understand the existing data. Major errors can be made in pulling data without the appropriate background knowledge. This is often not discovered until much closer to deployment.

Finally, if the appropriate resources are not being made available from a business and/or technical perspective, then the true priority of the data warehouse must be evaluated.

Finding Lost Institutional Knowledge

Over the years, application systems have been developed to automate business processes and apply business rules and logic. In addition to these systems that run the business, a series of programs, SAS routines, or complex spreadsheets have been developed to meet reporting and analytical needs. Many of these reports and analyses have now been in place for years, but over time the fundamental purpose of a report may be lost.

In addition, the business rules underlying the criteria for the report have been embedded in code that is simply run on a regular basis. Too often, a description of these rules and the business rationale behind them has been lost. Therefore, when the data warehouse team is applying pressure to understand the current business practices, the business community may honestly no longer have that knowledge. This is a fairly common situation, and one that does not warrant looking for someone to blame. It is important to acknowledge that this is the case and work from there.

RECOMMENDATION **One's first impulse is to read the programs or dissect the spreadsheets. After many hours of tedious work, the business logic can be**

found, but the rationale cannot. The result is merely a better understanding of the status quo, which is probably not meeting all of the business' needs. As an alternative, step back and figure out what is really needed to understand and run the business. This is a chance to define, or redefine, how you want to look at the group, department, or company. Starting from a clean slate seems scary, but it can be a much faster, and more direct and more useful approach than trying to replicate an existing, potentially archaic set of reports.

Summary

Building a data warehouse is challenging, but it can also provide great rewards. This chapter offered a general overview of data warehousing and illustrated how it can serve as a potential foundation for realizing business value. The next chapter has answers to some of the most frequently asked questions about data warehousing, and subsequent chapters provide more in-depth coverage about how to build and sustain a successful data warehouse.