NILS H. RASMUS

F I N
B U S
I N T E

WWW Companion

FINANCIAL BUSINESS INTELLIGENCE

FINANCIAL BUSINESS INTELLIGENCE

Trends, Technology,
Software Selection, and Implementation

Nils Rasmussen

Paul S. Goldy

Per O. Solli

This publication is designed to provide accurate and authoritative information in
regard to the subject matter covered. It is sold with the understanding that the
publisher is not engaged in rendering legal, accounting, or other professional services.
If legal advice or other expert assistance is required, the services of a competent
professional person should be sought.

This title is also available in print as ISBN Bookz 0-471-15555-1.

Some content that appears in the print version of this book may not be available in this
electronic version.

For more information about Wiley products, visit our web site at www.Wiley.com.

CONTENTS

FOREWORD

In the beginning were simple information systems. Then those simple systems grew into applications. Soon there were many applications. Data base technology appeared. Next came online transaction processing and soon there were online applications. As time passed applications aged, companies merged, and new applications were built or acquired. Before a company knew it there were many applications each running a separate part of the business.

One day the corporate world woke up to the dilemma of the need for information. The corporate decision makers said— "I know we have this information somewhere in our corporation, but I just can't find it." I guess I will have to make decisions based on experience and intuition.

If you asked anyone in the corporation if there was information in the corporation, the answer was a quick and definitive— "why of course we have information—we have applications." But once the surface was pricked, it was discovered that the portfolio of applications that belonged to the corporation in fact did not provide the corporation with information. The base of applications was:

- *Unintegrated.* One application thought of customers, transactions, and vendors as one thing. Another application thought of customers, transactions, and vendors as something else. And another application thought of customers, transactions, and vendors as yet something else. When it was desired to look across the corporation to see who a corporate customer was, what a corporate transaction was, and who a corporate vendor was, there simply was no way to do it because each application had its own unique interpretation. Corporate information in the face of a base of applications was nonexistent,

- *For current data only*. Corporate historical information was treated as an afterthought. In many circles historical data was shed as quickly as possible in the name of doing efficient transaction processing. The application's singular focus was on current information. The application can tell what today's bank balance is. The application can tell where a shipment is today. The application can tell what insurance coverage there is. But the application cannot tell what the average balance for an account for a month was, even though that information passed directly through the hands of the application. The application programmer saw no immediate requirement for historical information so the historical information was discarded. Ignoring history was an unwise thing to do. It is through history that perspective is gained. Without history everything looks flat. Applications thought of historical information as unnecessary baggage and got rid of it as soon as possible and in doing so preempted an entire dimension of information.

- *Not easy to get to*. The applications of yesterday were designed for efficiency of operation and compactness of storage. This approach to design made the transactions run quickly. But the designs that were optimal for efficiency of operation were anything but optimal for access and analysis of information. And the technology that the applications were written and housed in was optimal only for efficiency of operation. In order to access the data found in the applications, the end user had to fight through both technology and design. In a word, the end user was frustrated because accessing data was difficult.

It is no wonder then that the end user said, "I know the information is there in my corporate applications. I just can't get to it."

It is into this world that data warehousing was born. Data warehousing recognizes the need for corporate integration, the need for historical data, and the need to store data in a manner and in a technology that is optimal for access and analysis. Once the foundation of data warehousing is laid, whole new worlds and whole new opportunities open up.

Nowhere are there more opportunities than in the world of finance. Finance is close to the heartbeat of every corporation. If anyone needs true corporate information, it is the finance department. Top management needs to see—desperately—the larger financial picture of the corporation, and that larger picture cannot be achieved with a foundation of applications. Without integrated financial information, the corporation is flying through the clouds blind.

Likewise the corporation needs historical financial information. It is not enough to look at how finances are today (although that certainly is an inter-

esting perspective.) Deep and meaningful insight about the life of and the direction of the corporation can be derived from comparing financial positions over time. Especially useful is the ability to look across time at financial positions in different and novel ways. Being able to go back in time and restructure information in a way that has not been imagined before can be very useful. And with a properly built financial data warehouse you can do just that.

And of course putting data into a technology and in a design that is easy to access is problematic. Once the data is freed of its shackles it can be used by anyone in innovative ways. It simply goes without saying that the first step in informed decision making is unfettered access to data and information.

Data warehousing unlocks all of this for the finance community.

This book tells you what you need to know about both finance and data warehousing. It opens the door to the magic elixir of accurate, integrated, historical, and easily available financial information. Once you have that, corporate decisions become enormously easier and enormously more effective.

BILL INMON
Sept 14, 2001

INTRODUCTION

How do you measure your company's financial business intelligence (BI)? Most people do not know how to answer this question because the organizations they work for are just starting to get into this field. Financial business intelligence is the process of getting useful information out of financial data, and this book looks at both processes, trends, and tools to help you increase your company's business intelligence.

The goal with improving your organization's business intelligence processes and tools should be to support financial leadership. Before you dig into the details of BI, we therefore hope that this book can be an important building block for you to help support your corporation's growth and success. This means that your BI initiative should be part of a bigger picture, which includes the following objectives:

- Creating a vision for financial leadership
- Coaching the organization to set realistic goals
- Supporting optimal decision making

Good BI tools are needed to improve the decision-making process. Historically (and still true for the vast majority of companies today), companies have spent too much time closing their books and preparing data and financial reports, and too little time on analysis and review. This is causing a disconnect between the analysis and review process and action (decision making). (See Figure I.1)
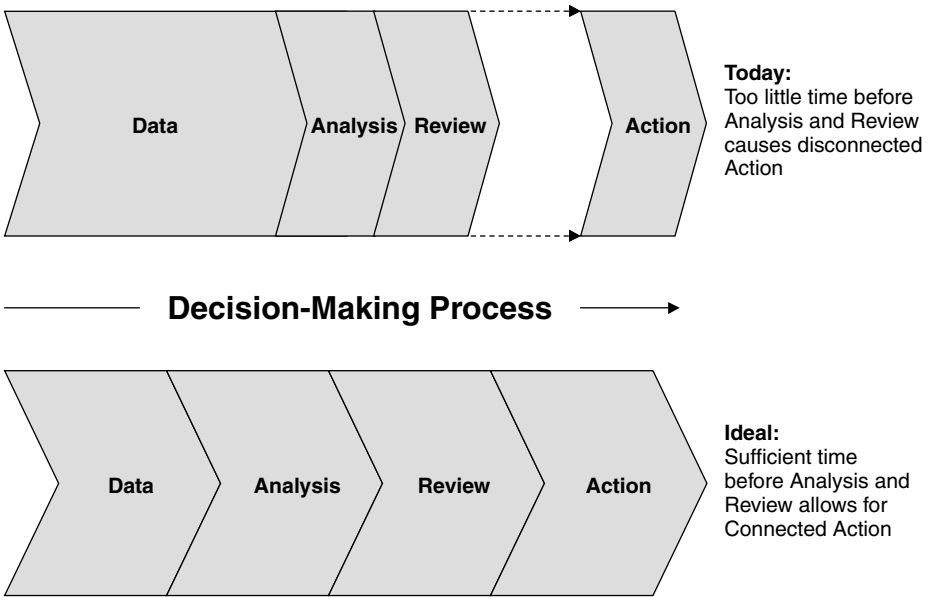
FIGURE I.1   Why business intelligence solutions are needed

After having spent the last decade implementing enterprise resource planning software and other solutions that support their operations, companies now have large databases with transactional data sitting in their computer rooms. Thousands of consulting hours and sometimes millions of dollars later, the return-on-investment scorecards range from poor to very good. However, one desire is common across all industries: to be able to quickly and easily analyze the financial data in the corporate databases in order to be able to make more intelligent decisions about the future. Transactional data is good for keeping track of *what* is happening in an organization, but is not well suited to finding out *why* things are happening or predicting future performance. In other words, after years of putting potentially valuable financial data *into* your corporate databases, it is now time to put the tools in place to get the data *out* of the same systems and organize it in useful ways to support the decision-making process.

This book will help you understand trends, technology, software selection, and implementation in regards to financial business intelligence software. Each part will arm you with knowledge that will put you in a good position to understand how your company can benefit from modern BI solutions, and how to successfully find and implement a solution that fits your needs and your technology infrastructure. Further on, because modern technology will be the

backbone of all future financial analytics processes, the book goes into detail about business intelligence software. It explains which functionality you need, how to select a new software solution, and most importantly, how to successfully implement it in your organization to excel in financial analysis in the twenty-first century.

Note that throughout the book the terms *business intelligence* (BI), *financial business intelligence*, *analytics*, and *financial analytics* are used interchangeably and unless otherwise stated will refer to the process of analyzing and reporting financial numbers and related technologies.

BOOK SUMMARY

The book consists of four parts and appendices:

- Part One—Evolution of Financial Business Intelligence
- Part Two—BI Technology
- Part Three—Software Evaluation and Selection
- Part Four—Implementing a Business Intelligence System
- Appendices

Part One, Evolution of Financial Business Intelligence, introduces you to financial business intelligence, and you can read about the trends and tools that financial managers today need to be aware of to streamline their companies' analytical processes for the future. The following topics are covered:

- Reporting and analysis trends
- Analytics processes and approaches
- Tools that revolutionize business intelligence
- Business intelligence trends

Part Two, BI Technology, examines the trends in technology that are driving modern business intelligence software. Any person planning to buy a new business intelligence solution for his or her company needs to know the strengths and weaknesses of popular online analytical processing (OLAP) and data warehouse technologies and the different platforms they support. This part also looks at some critical issues to consider for a financial data warehouse and front-end user interfaces. In addition to covering these technologies, Part Two looks at how the Web quickly is becoming a delivery vehicle and an important platform for most business intelligence solutions. Furthermore, this part of the book describes a number of key features found in many popular business

intelligence software products. Because most companies utilize spreadsheets as one of their analytical tools, there is a separate chapter about spreadsheet-based analysis.

Part Three, Software Evaluation and Selection, shows that buying and implementing business intelligence software can cost from $50,000 for a simple solution to millions of dollars for large, complex data warehouse projects, and it can be a difficult task to pick the solution and vendor with the best overall fit. This part aims at helping you to successfully evaluate and select the best analytical software for your company's particular needs. You will find basic selection criteria that should be used by the selection team, as well as guidelines for how to set up a request for proposal (RFP) and how to get the most out of a product demonstration. Part Three also gives you tips on contract negotiations and on using a software selection firm. A small but growing industry today is the use of application service providers (ASPs). This section helps you analyze if your company should use an ASP rather than hosting a business intelligence solution in-house. Last but not least, you will find a list of major business intelligence vendors and their software. The list also contains important information such as contact information and web addresses where you can find detailed information about each vendor and its solutions.

Part Four, Implementing a Business Intelligence System, shows that buying a financial business intelligence software that has a good fit with your corporate needs does not automatically mean success. Modern business intelligence software gives you powerful end-user tools, but these will often be of good use only if implemented by people with a good skill set and who understand your needs. This part discusses how to select the right people to put on the implementation team, and their desired skill sets. Part Four also covers project planning and uses a brief example of a project plan to help illustrate points in the discussion. After any implementation, there is a long-term need for maintaining and supporting your business intelligence solution. Part Four describes how you should handle these activities.

APPENDICES

We have placed a number of value-added documents in the appendices of the book. You can use these documents as examples or templates for your own documentation needs in software selection and implementation, and you can also copy information from the examples provided. In particular, it is worth mentioning the comprehensive request for proposal (RFP) appendix. It lists most of the questions a company would ask a vendor in connection with a business intelligence software selection. Utilizing questions from this list to create your own customized RFP you can save a lot of time and money, compared to

doing this yourself or hiring a consulting company to write your RFP. The appendices consists of the following documents:

- Sample nondisclosure agreement
- Sample request for proposal
- Software candidate evaluation and rating sheet
- Vendor list with contact information
- Sample implementation project plan
- Sample license agreement
- Sample support agreement
- Sample consulting agreement
- Dictionary of business intelligence terms

FILE DOWNLOADS

In order to save you time and money in your software selection and implementation process, several useful documents from the Appendices are provided on the publisher's web site (*www.wiley.com/go/fbi*) for your use.

The user password is **Financial**. These documents include:

- Sample Request for Proposal
- Software Candidate Evaluation and Rating Sheet
- Sample Licensing Agreement
- Sample Confidentiality and Non-Disclosure Agreement
- Sample Support Plan/Agreement
- Sample Project Plan
- Sample Consulting Agreement

PART ONE

EVOLUTION OF FINANCIAL BUSINESS INTELLIGENCE

1

HISTORY AND FUTURE OF BUSINESS INTELLIGENCE

As long as there has been business, there have been owners, managers, and investors trying to make sense of the numbers. The companies that understood revenue trends and customer preferences could adapt their products and services to the needs of the marketplace and would gain market share and become more successful than their competition. If a company had 10 customers and five products, and wanted to analyze revenue, cost of sales, and gross margin for the past 24 months, they would have 3,600 transactions to look at. It was hard to make sense of this amount of numbers with only a pen and paper to assist you. In modern times, the computer and databases make the job easier. However, increasingly more financial data are being gathered from customers, manufacturing, sales, web site usage information, and other front and back office systems. (See Figure 1.1.) Facing databases with hundreds of thousands and often millions of transactions being generated every month, managers and analysts are looking for effective ways to analyze all the data.

HISTORY OF BI

In the 1970s and 1980s analytical software packages started showing up in the marketplace. However, lack of computing power, poor user friendliness, and cumbersome and manual integration with the transaction systems providing the
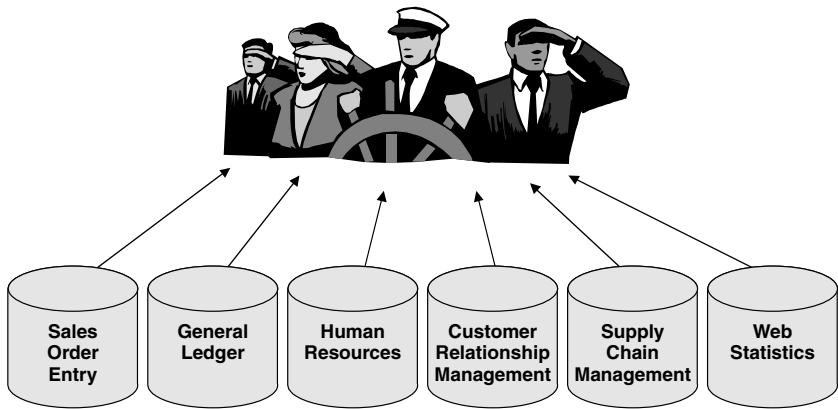
FIGURE 1.1   More data and data sources than ever before

data kept BI tools from widespread usage. The release of spreadsheet software like Lotus 1-2-3 and Excel in the 1980s opened up for end users creating their own data models for business analysis. Spreadsheets are still widely used in this area today, and probably will be for many years to come. For a period in the 1980s and early 1990s so-called executive information systems (EIS) grew in popularity with the promise that they would put key information on the desktops of executives. The idea was that colorful software screens with big buttons and sometimes touch screen monitors, so that the user did not ever have to use a mouse, should put data directly in the hands of top management to reduce the need for secretaries and assistants to write and print reports for them. However, one of the biggest problems with EIS systems was that it took a lot of manual work to convert and load data from the data sources, as well as to maintain customized versions of the user screens. Major efforts in maintaining the EIS made many implementations short-lived. In the 1990s and the new millennium, with the widespread usage of SQL (standard query language) databases; datawarehouse technologies; extraction, transformation, and loading (ETL) tools; as well as new and powerful end-user analytical software, the stage is set for fast growth in usage of BI tools in the next decade. (See Figure 1.2.)

Furthermore, most of the BI software vendors have now released web-based versions of their solutions. Companies can now easily and at a low cost give users access to large amounts of corporate data and sophisticated analytical tools. By providing access to the Internet or an intranet connection, a person can investigate and analyze data from home, when traveling, or from any other location at which they may happen to be.
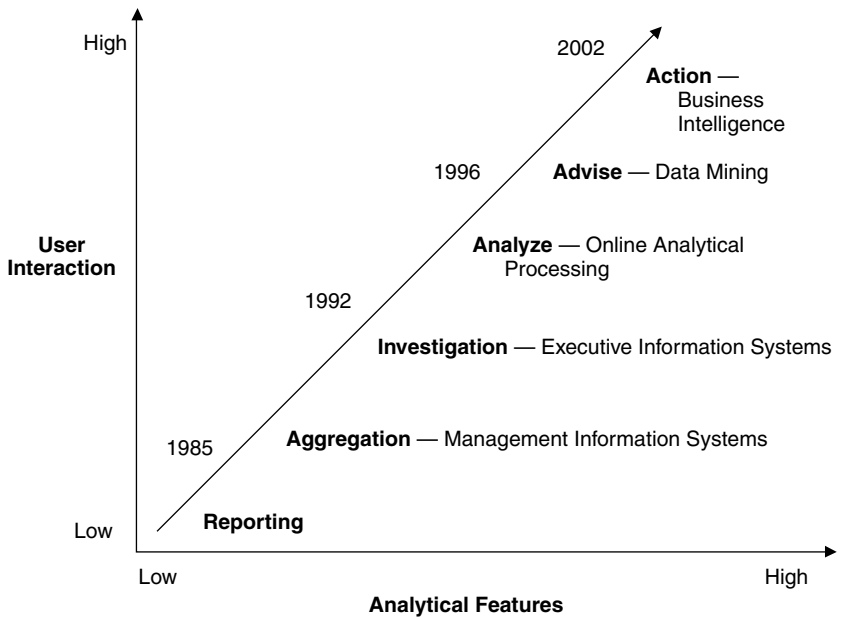
FIGURE 1.2   Evolution from static reports to business intelligence

TRENDS

Open Databases

One of the most important trends affecting BI is that vendors are moving toward commercial database platforms for their business applications. Because BI software is only as good as the data available to it, it is of key importance that data easily can be read by analytical tools or transferred to a data warehouse. There is no industry-standard definition of what an *open database platform* is. However, most technical people would agree that an open database is a software that facilitates structured data storage and that can be populated or read from by using one or more commonly known data extraction tools (you can read more about modern ETL tools later in "common data exchange formats" in this chapter and in Chapters 2, 5, 6, and 13).

For more than a decade, it has been common (and it still is) to transfer data between databases, such as between a general ledger and a consolidation and reporting or a budgeting tool, by using a standard text file (often referred to as a flat file). As long as both the source and destination database can accept the text file format, it should be possible to transfer data. Most modern databases, such as Oracle, Microsoft SQL Server, Sybase, Informix, and IBM DB2

(you can find a list of popular databases in Chapter 13), support more sophisticated data transfer procedures than simply producing and loading text files. Perhaps the most common underlying technology standard for data integration is Online Database Connectivity (ODBC). If a database supports this standard, it is relatively easy to make a direct connection between two systems (although it will usually still be necessary to do a data conversion before the data can be loaded into the source database). In addition, most modern databases that we consider "open" and easy to read from support standard query language (SQL). Using SQL, a technical consultant can write or customize an integration between two open databases and facilitate a direct and automated data transfer process.

In terms of financial business intelligence tools, they might depend on updates from the data source(s) as frequent as multiple times per day, and in the future, close to real-time access. This means that it is critical that data transfer (and possible conversions) are fully automated and that they trigger manual (human) intervention only when there are exceptions or issues that cannot be handled by the business rules in the extraction, transformation, and loading program. This type of automation lacked in the old executive information systems (EIS), and this was to a large extent the reason why most of them failed after a while. It was simply too labor intensive to keep them up to date. Today, however, an increasing number of companies have replaced their old proprietary transaction systems with new ones that allow for automated integration with datawarehouses and data marts, thereby opening up their data for sophisticated analytical tools.

Common Data Exchange Formats

The main obstacle to provide decision makers with a single BI interface that has direct access to all the information that resides in their company's financial databases is that different software vendors have never been good at extracting and delivering data to each other's databases. Historically, it has been the job of the company's information technology (IT) department to move data from one system to another. In its simplest sense, this job has consisted of creating an export file from the data source and importing it to the other system. The most common file format that has been used (and it still is today) is the text file. This is really a list of transactions, in which each transaction is a row, and each row consists of a number of information fields (such as account number, period, amount, cost center, currency code, etc.) separated by a common character such as a comma, space, or semicolon. In many cases, however, moving data from one database to another is not as simple as exporting and importing files. Often, systems have conflicting codes; for example, the same account number can refer to two different revenue or expense categories in two systems. In particular, this is a common issue in large companies with multiple

subsidiaries with systems from different vendors. In the case of conflicting code systems, a data conversion (sometimes referred to as *data cleansing* in data warehouse projects) is necessary to make it possible for two or more systems to exchange information. In a worst-case scenario, data has to be manually reentered from one system to another. (See Figure 1.3.)

In most companies, the IT department will write conversion programs to automate the conversion of data files. In a better scenario, one of the two systems exchanging data can convert the data file to the other system's required format as the file is being exported or imported. There are also a number of third party data conversion tools available today (see ETL product and vendor list in Chapter 13) that have predefined or user-defined links to popular systems, and can save a company's IT department a lot of time and money.

More and more software vendors (in particular, database/datawarehouse vendors) recognize the need of today's organizations to move large data volumes smoothly and efficiently between databases, and in recent years a number of powerful new tools have become available. Many of these tools have been referred to as extraction, transformation, and loading (ETL) tools. Setting up an effective data extraction, transformation, and loading process can take up to 70 to 75% of the time spent on a typical warehousing project.
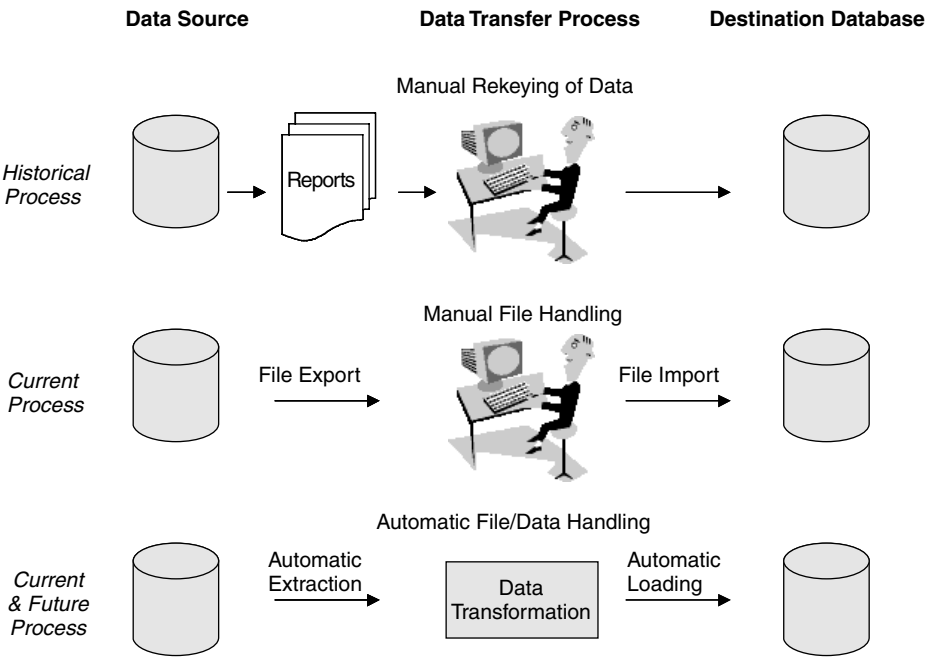
FIGURE 1.3   Evolution of data transfer

ETL tools automates the sometimes extremely laborious (and manual) job of moving data from one data source/data base (e.g., a transaction database like the general ledger) to another (e.g., a datawarehouse or data mart). In general, an ETL tool does the following:

- It reads data from an input source (relational table, flat file, etc.).
- Then, it passes the information through business rule–based process to modify, enhance, or eliminate different data elements.
- Finally, it writes the output result back out to another relational table (e.g., in a datawarehouse), flat file, and so on.

There are a number of tools that perform the ETL functions with more or less functionality in certain areas, some with more functionality in either extraction, transformation, or loading. True ETL tools are strong in all three areas.

A good example of an ETL tool is the Data Transformation Services (DTS) utility that is included with the Microsoft SQL Server database management system (DBMS). The tool comes with a number of prebuilt links to make it faster and easier to connect to other popular databases such as Oracle, Sybase, Unix-based databases, and Microsoft Access, that makes it simpler to set up integrations and it also allows for scheduling, so the ETL process can take place automatically at any given point in time and with a predefined frequency.

Recently, we have also seen the emergence of common data exchange languages in which governments, information standards organizations, software companies, and different interest groups from all major industries are cooperating. Probably one of the most significant of these data exchange languages is Extensible Markup Language (XML) and Extensive Business Reporting Language (XBRL), a subset of XML specifically developed to ease exchange of *business* information between different applications. The formal definition of XBRL is: "A standards-based electronic language for business information, financial reporting and analysis."[1] Standards like XML and XBRL take data handling a major step forward from Hypertext Markup Language (HTML), the current web standard also used for web-enabling formatted reports or other financial information. With the new XML standard, plus the related ones specifically made for accounting and financial information, it will be possible to search and extract, for example, industry or competitor information from across the web. A layer of business logic can be applied so instead of a search returning information from all of your competitors, it can for example filter out only those that have a market capitalization higher than $500 million or total revenues higher than $200 million. (See Figure 1.4.)

XBRL standards (which describe the business language) have been, or are currently being developed for the following areas:

XBRL Document on the Web or Intranet:

Data Source:

| General Ledger |
| Best-of-Breed Systems |
| CRM System |
| Custom-Built System |
| Explanatory Text |
| Third-Party Information |

XBRL →

| Tax Filing |
| US GAAP Financial Statement In German |
| US GAAP Financial Statement On the Web |
| Bank Filing |
| Internal Reports and Data Files |
| Press Releases |

XBRL →

Recipient:

| Government |
| Investors and Creditors |
| News Agencies |
| Competitors'/ Own BI System |
| Vendors |
| Employees' Cell ph./PDA |

FIGURE 1.4   XBRL simplifies exchange of data

- Assurance schedules
- Authoritative literature
- Business reporting (balanced scorecard, benchmarking, etc.)
- Credit line reporting
- Economic indicators and demographics
- Financial statements
- General ledger
- Journal entry reporting
- Mutual fund performance
- Performance-related press releases
- Risk reporting
- Regulatory filings
- Tax filings

The positive impact of XBRL for financial BI can be very significant. Imagine decision makers in a company turning on their computers (or handheld

KPI - Key Performance Indicators
Company : SPORTY Manufacturing
June 2001

	Actual Jun-01	Budget Jun-01	Variance in $	Variance In %	Industry Average	Competitor XYZ
Net Revenue	$1,300	$1,100	$200	15 %	$922	$875
Average Revenue per Employee	$260	$220	$40	15 %	$85	$80
Gross Margin	$400	$350	$50	13 %	$280	$200
Average Gross Margin per Product	$25	$20	$5	20 %	$22	$15
Online Sales as a % of Total Sales	$15	$20	($5)	-33 %	$5	$10
Cash Flow	$35	$30	$5	14 %	$28	$15
Sales Pipeline	$3,000	$4,000	($1,000)	-33 %	na	na

FIGURE 1.5   KPI Report Sample

devices) in the morning and being able to see fresh information on their BI screens that have automatically been extracted (thanks to XBRL) from competitors' or partners' press releases and regulatory filings. And because XBRL contains specific information about what type of data is being extracted, the BI software can take the data and put it into the right context. In other words, if a BI screen consists of a key performance indicator report with actual to budget variances, and actual to industry or competitor variances, the industry and competitor information can be included automatically thanks to XBRL. (See Figure 1.5.) Further on, because of the specific information contained within an XBRL tag (part of the XBRL code that is linked to financial or text information), it is possible to automatically locate and retrieve pieces of information within a report/document and pull it out to use it somewhere else, for example, in a BI system. (See Figure 1.6.)

All major industries now have international organizations that develop XBRL tags specific to their industry's financial and other information. These industry-specific tags are known as taxonomies.

Key benefits of XBRL include:

- Faster access to external information
- Easier to create integrations to third-party systems and documents
- Saves cost by reducing development time
- Saves costs and potential errors by eliminating manual data entry

Financial Report Example:

Balance Sheet
Company XYZ

	Period Ending 12/31/2001	Period Ending 11/30/2001
Cash and Cash Equivalents	$235,000	$238,000

Conversion to XBRL
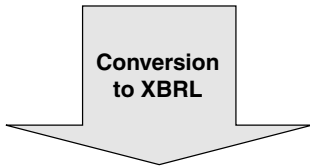
XBRL Example*:

```
<group>

<group type="currentAssets.cashandCashEquivalents">

<item period="2001-12-31">$235,000 </item>

<item period="2001-11-30">$238,000 </item>

</group>
```

* Simplified

FIGURE 1.6   XBRL example

In the coming years as legacy applications are phased out and new soft-ware and integration standards are adopted, it will be easier to move data be-tween a company's own systems or between the databases of different organizations. For business intelligence, this means that decision makers and analysts will have easier and faster access to frequently updated information, which ultimately should support quicker and better decision making.

Data Marts and Datawarehouses

Chief financial officers (CFOs), controllers, analysts, and other high-end users of financial information in a company will agree that their enterprise resource planning (ERP) software was not built with easy reporting and analysis in

mind. More than a decade ago, it was common to have isolated financial modules for each accounting area and to use multiple report writers to get information out of the different modules.

Today, most companies have an ERP system in place that provides highly integrated accounting databases. Report writers can access data from multiple modules and integrate it into information-rich reports. Many vendors now also have BI tools accessing their accounting modules directly. (See Figure 1.7.) It is now increasingly popular to gather all the significant data from the ERP system and load it into a datawarehouse or a data mart, and then connect report writers and analysis tools to these to create a more consistent and knowledge-oriented environment.

Companies have been deploying datawarehouses as a means to facilitate common data storage for better reporting and analysis for many years now. A datawarehouse can be defined as a collection of data designed to support management decision making. Typically, a majority of the implementation projects have been known to be slow, complex, and often very expensive. Dataware-

1. The Past – "Data"
Isolated Financial Systems
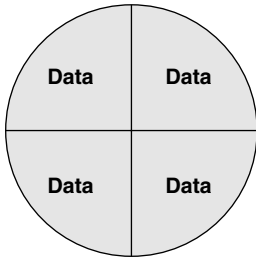
2. Today – "Information"
ERP-Focused Integrations
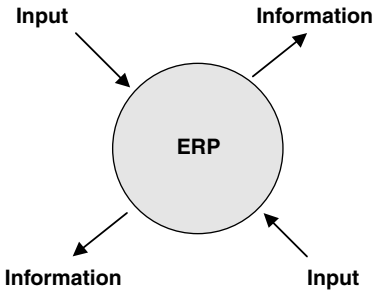
3. The Future – "Knowledge"
Datawarehouse-Focused Integrations
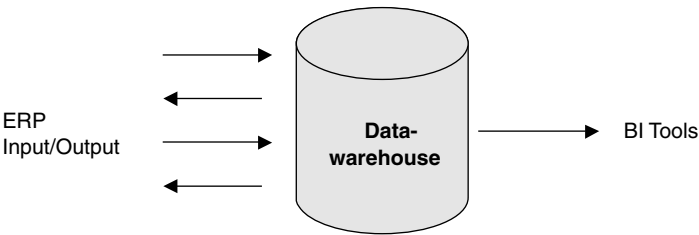
FIGURE 1.7   Evolution of financial systems

housing activities have therefore for the most part been found only in large companies with a particularly strong need for the technology and with enough resources to build and support it. However, many recent trends and technology developments are now making datawarehousing a hot topic that a large variety of companies are taking a look at. Some of these trends are:

- *BI push.* The popularity of BI tools are pushing datawarehouse initiatives to provide a common, easily accessible data source.
- *Technological advances.* Adoption of open database platforms and powerful integration tools makes it easier and less expensive to build and maintain data warehouses.
- *Reduced prices.* Increased competition (from large database vendors like Oracle and Microsoft) as well as easier to use technologies are bringing the cost of datawarehousing down.

Datawarehouses contain a wide variety of data that present a coherent picture of business conditions at a single point in time. Development of a datawarehouse includes development of systems to extract data from operating systems plus installation of a warehouse database system that provides data storage and that gives managers flexible access to the data. The term *datawarehousing* generally refers to combining many different databases across an entire enterprise.

Data mart is another frequently used term in connection with BI tools. A data mart (also spelled datamart) can be defined as a database, or collection of databases, designed to help managers make strategic decisions about their business. Whereas a datawarehouse combines databases across an entire enterprise, data marts are usually smaller and focus on a particular subject or department. Some data marts, often called *dependent data marts,* are subsets of larger datawarehouses.

In order to provide coherent, rich, and timely data to a company's BI tool, many organizations, particularly larger ones, have found it essential to gather all the data from different transactional sources in a data warehouse that can feed their BI tool. Advantages of gathering data in a datawarehouse include:

- A single source for important data
- Consistent definitions
- Well-documented relationships
- Coordinated cutoff times

Your company's BI capabilities will be only as good as the data available to the front-end analysis tools (see product and vendor list in Chapter 13) in place. Most modern financial analysis tools are using online analytical processing (OLAP) data cubes as their database to provide flexible views and fast

drill-down and querying. These OLAP cubes (often named data marts because they each contain a specific set of related data) can either be built or populated directly from the transaction source (e.g., a budgeting and reporting database, accounts payable, sales order processing, accounts receivable, customer relationship management, etc.) or they can be fed from a full-fledged datawarehouse.

Depending on the size and scope of a datawarehouse project, it can range from an hour's work and a few thousand dollars in software to build a specific data mart from a transaction system that has out-of-the-box OLAP cube generation features to multiyear projects that can cost millions of dollars and that entail loading data from all core enterprise databases into a very large datawarehouse.

Figure 1.8 shows how datawarehouses have changed to become much more effective information stores for BI tools.

Vendors and companies have realized that to link a front-end analysis tool directly to each different transactional database (such as the general ledger, accounts payable, accounts receivable, sales order entry, budget software, etc.) can be a daunting task, so many opt for putting a datawarehouse on top of the transaction systems, and then integrating the business intelligence tools with the datawarehouse. The following list shows the different options for integrating BI tools with your data:

- Integrate one BI tool directly with the transaction sources.
- Integrate several different BI tools with the different transaction sources (often based on which technology or BI tool a transaction system vendor support).
- Use several different BI tools with the different transaction sources, and provide the user with a single, portal-based interface to all the BI tools.

Past	Today
Decentralized system	Centralized and integrated
Static users	Scaleable
Low security concerns	High security concerns
Mostly batch updates	More real time
Noncritical availability	More critical availability (24 hours per day/7 days a week)
Internal data sources	Also includes external data sources

FIGURE 1.8   Evolution of datawarehouses

- Use the BI tools that come with the transaction systems you have (e.g., your customer relationship management [CRM] system might offer graphing and querying, and your financial reporting tool might come with its own analysis tools).

- Use one single BI tool (could be more) that integrates with all your data through the use of a datawarehouse that links to all transaction systems. (See Figure 1.9.)

With most companies' goals now being to take better advantage of all their rich data sources by providing employees easy and fast access to analyze data and make optimal decisions, it will in many cases make the most sense to gather and organize all the data in a single datawarehouse (often with many smaller data marts feeding off it). This allows them to standardize on a single database platform (the datawarehouse), and to standardize on one or two BI tools for reporting and analysis, and use a portal as a single, web-based front-end. (See Figure 1.10.) This will save costs, reduce training requirements in BI tools, provide faster access to corporate information, and so on.

Transaction Database   **BI Configuration**

A) A BI solution integrated directly with the transaction sources

B) Several BI solutions integrated directly with the different transaction sources

C) Same as B) but with a single portal as a user interface

Web Portal

D) Use of possible integrated BI solution that comes from the ERP vendor

E) Use a datawarehouse as the data source for the end-user BI tool

Data-warehouse

FIGURE 1.9   Datawarehouse and data mart configurations
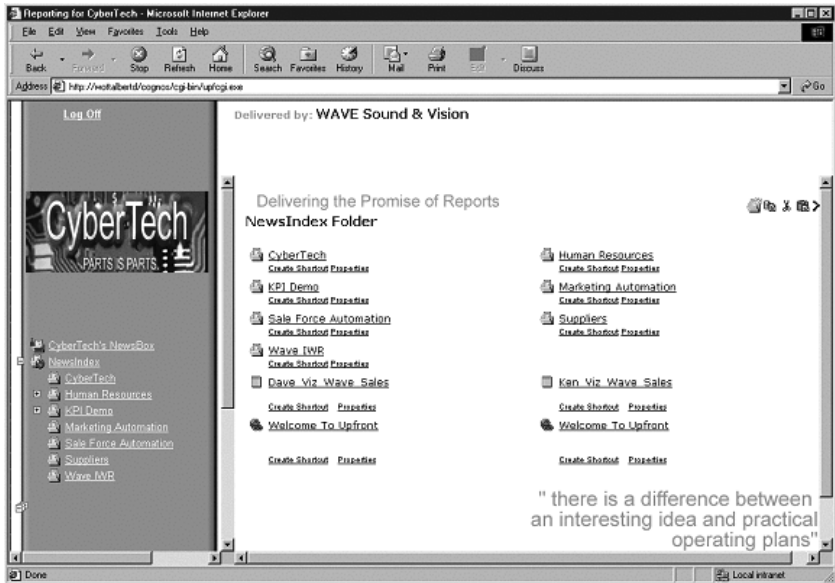
FIGURE 1.10   BI Portal Example

Source: Used with permission. Cognos, 2001.

Merging of Report Writers and Analysis Tools

Financial reporting has gone through several phases since its inception half a century ago. Today, this has resulted in three different types of reporting tools that all are still in use. These can be categorized by the type of end-user interaction they allow:

- Standard reports (static, specific format)
- Semi-custom reports (customizable with parameters and variables)
- Ad-hoc reports (allows drill-down and customization on the fly) (see Figure 1.11)

For years, the common way of retrieving and organizing data from a financial database has been to write (or code, as many people call it due to the technical expertise required to use many report writers) reports. If your system has a strong report writer, this will take you pretty far in supplying the organization with good, informative reports on which to base control and decision making. However, the problem with a typical financial report, such as a profit and loss report, a balance sheet, or a sales report, is that it is static. This means that you can read from your report that salary expenses for the company are

Low   **High**

Ad-hoc
Reports

Semi-Custom
Reports

Formatting
and
Layout
Control

User
Interactivity
and
Analytical
Features

Static Reports

High   **Low**
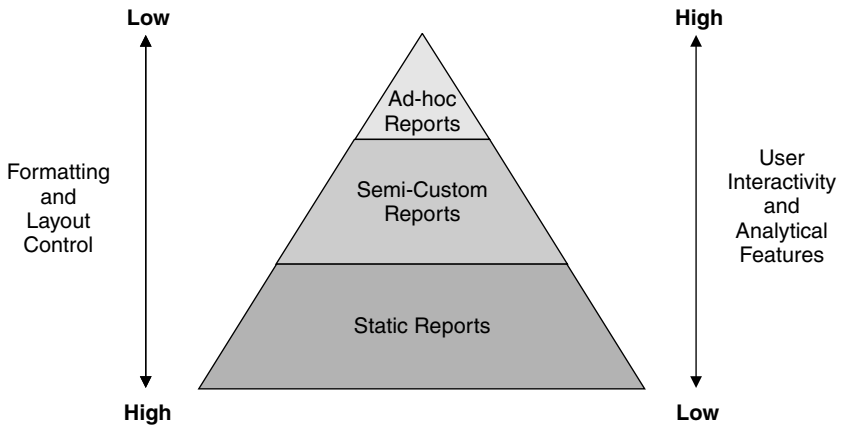
FIGURE 1.11   Categories of report writers

20% above budget this month, but the question that immediately should follow is *why*? With the typical financial report writer at your disposal, you will in this case have to run a more detailed report for each department to find out which department(s) have caused the high salary expense variance. That is, if someone has already written such a report for you, otherwise you hopefully know how to write reports so you can create one yourself. Once you run the report for each department and discover that department 100 caused the high variance, the next step is to figure out what hides behind the salary account in this department. At this point, most users will have to move over to the payroll system and run a detailed employee level report there, that is, if you know how to run reports in the payroll software.

All the wasted time and effort exemplified above, and true for so many financial software users today, is becoming history for an increasing number of companies. A new breed of BI tools that can be categorized as financial ad-hoc report writers are entering the market. (see Figure 1.12) The tools offer functionality such as:

- Drill-down, drill-around, and drill-through—so report users can get their questions answered without running additional reports

- Direct spreadsheet integration—to provide a high level of formatting and user friendliness

- Fully web-based interface—for report writing, report processing, and drill-down from a web browser

- Intelligent trees—tree hierarchies are increasingly popular to automate consolidation report writing and drill-down. Intelligent trees can also provide formatting, automatic maintenance of hierarchies, and more.

- Automatic generation of OLAP cubes—for low-cost and easy-to-implement OLAP analysis that complements highly formatted financial reports.

Drill-down means that you can double-click on a row, column, or number in a report and the system automatically generates a new report/view for you that provides the underlying detail. With *drill-around*, you can also drill "sideways" (instead of down in a hierarchy) to a neighboring department, account, period, and so on. When you are at the lowest level of information in your database (e.g., salary expense for department 100 in February 2002), some report writers now offer a direct link to the other database(s) that supplied the detail underlying the lowest transaction level in the current report and database. This is referred to as a *drill-through*. This means that users with little training can look at reports and investigate numbers of interest without using multiple reports or report writers and multiple software packages. The result should be better, faster decision making and time and cost savings for the company.
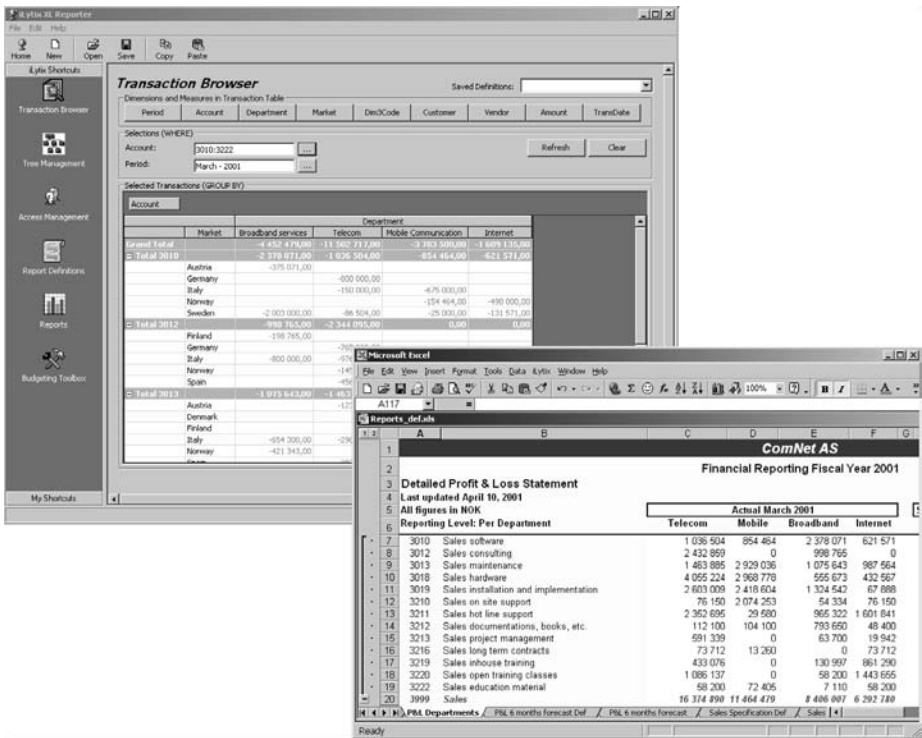
FIGURE 1.12   Modern Financial Report Writer Example

Source: Used with permission, iLytix.

Traditionally, interactive functionality such as the drill-down features described above has been the domain of executive information systems, spreadsheet pivot tables, and OLAP. While a new breed of report writers are providing strong analytical features, OLAP software packages are becoming better at producing formatted reports, as compared to more rigid data grids with little financial reporting formatting features. The result is a blurring of the traditional lines between report writers and analysis tools, and increasingly end users can enjoy the best of two worlds: both powerful formatting and flexible report writing, as well as strong analysis features. The ultimate result is what analytical software vendors have been preaching for several years: You should be able to spend 80% of your time on analysis and decision making and 20% on preparing the information (e.g., uploading data, writing reports, etc.), rather than the opposite. Most companies do not yet have the systems in place to get anywhere close to this ratio, but as the different technologies used move forward, it should be reality for most people in the coming years. The companies that first enable themselves to spend most time on analysis and decision making will certainly come out on top in tomorrow's marketplace.

Reengineering the Financial Reporting Process

Many companies are now taking a second look at how their financial reporting is structured, not only from an information standpoint, but with regard to the entire reporting process and the value provided to the organization in terms of planning, management, and control. A number of initiatives are taken in order to achieve the following goals:

- Better measure profitability of products and services
- Develop proven methods of measuring the profitability
- Improve the consistency in reporting financial results
- Increase visibility of new services and solutions
- Improve speed of the reporting process (read more about this in the section "Real time reporting")

Examples of initiatives taken to reach the goals are:

- Standardize on tools and technology platforms
  - ERP solutions
  - Datawarehousing
  - BI tools (including consolidation and reporting software)
  - One single web-based interface (enterprise information portal)

- Standardize data models
  - Chart of accounts
  - Other dimension codes
  - Organization-wide support and control that ensures that the model is complied with
- Business process improvement (BPI)
  - Create a business planning environment rather than a financial planning environment (don't be so focused on the numbers that you forget the big picture).
  - Improve the process by realigning the ownership of business drivers (using a balanced scorecard approach is one way of achieving this— see BI and Balanced Scorecards section later in this chapter).
  - Leverage best practices across all organizational entities.
  - Improve the utilization of personnel human resources through better motivation.
- Actively manage the financial knowledge process
  - Identify the target audience of the financial information.
  - Define their knowledge requirements.
  - Establish a process to transfer data into knowledge (meaningful information).
  - Deliver the knowledge personalized to each audience (e.g., through custom and ad-hoc reporting capabilities and OLAP analysis).

With the new BI tools, standardization on open platforms, and more organizational focus on the analysis of business information, it is becoming easier and more realistic than ever before to successfully put the initiatives listed above into action.

Real-Time Reporting

Especially in financial management, you will start to hear this term more often. *Real-time reporting* refers to a process in which a user will get a report at the exact same time as the underlying transactional data is entered and saved somewhere else. Per definition, we will most likely never be able to experience "real" real-time reporting, but we will get so close that the seconds or milliseconds we wait for our financial reports will not matter.

We are accustomed to thinking about the urgency of receiving financial information in connection with the stock market. Up-to-the-second updated in-

formation has usually been the privilege of only professional stock brokers and traders, but many web-based information services now deliver stock quotes and other information updated every 15 minutes or less, straight to your office or home computer.

With regard to a company's own internal financial reporting activities, it is still not uncommon to wait days for weekly or bimonthly sales reports or a week or more for monthly or quarterly financial statements (such as profit and loss, balance sheet, and cash flow reports). This is a far cry from real-time reporting, and it is the goal of most financial managers to drastically reduce this delay from the time information is entered until it is available to decision makers in both summarized and detailed reports.

There are several reasons why most organizations wait days or weeks for their vital financial information. One major factor is the lack of good integration between software solutions to, for example, get information from a sales order entry system and into a BI data mart or into an Excel spreadsheet (integration issues are described in more detail in the section Common Data Exchange Formats earlier in this chapter). Another reason is the need for human intervention to make adjustments to data or to check that the data is correct before it is given to decision makers.

Data entry to most financial systems is moving online as systems are web-enabled or by other means directly connected to each other. This is seen across a number of applications, from point-of-purchase scanners in stores to customers entering orders themselves directly through an e-commerce store front, and it is automatically updating the ERP system.

Many software packages are becoming better at providing automatic entries (e.g., allocations, currency adjustments, or elimination entries) to reduce manual entries as financial information is rolled up to company or corporate group levels. An increasing number of companies have also introduced rules for accepting certain, less significant, variances or errors in preliminary reports to avoid human-related delays. If, for example, a sales or inventory report in Company A is 97% accurate and they can get it one hour after the last data was entered, it means that the company can make decisions more quickly and get a clear competitive advantage on Company B, which may read the same type of report two days later.

This preclosing of the accounting books is often referred to as real-time reporting (if the time delay is very short) or *virtual closing*. It will normally mean that the necessary control procedures and adjustments that are manual are done later, and a final closing for legal and control purposes takes place later without delaying initial key decision making. John Chambers, chief executive officer (CEO) at Cisco, a company known to be an industry leader in fast turnaround in their financial reporting, made the following statements during Cisco's year-end earnings release call on August 10, 1999: ". . . what excites the CEO is the ability to know what the business is doing at any given point in

time, react quickly to market shifts and competitive threats and remain in tight control while empowering employees to make informed decisions more quickly." He also stated the following: "The Virtual Close . . . has, in my opinion, just as much impact on a company's future success or lack thereof as the well-published e-commerce area." An increasing number of organizations will weigh the cost of potential poor decisions based on lacking or erroneous information against the cost of late decisions. As the world goes more online every year, and the global marketplace becomes more competitive, we will see that a large number of companies will strive toward achieving real-time reporting.

BI and Balanced Scorecards

What is a balanced scorecard? A balanced scorecard provides a means for linking the strategies of the different business areas within a corporation to the overall corporate vision. The term and methodology was introduced by David Norton and Harvard Business School professor Robert Kaplan back in 1992. They had found that the typical financial reports most companies were churning out did not provide management with enough information to run their companies. Instead of just reporting of the account numbers in the general ledger, Norton and Kaplan suggested that management focus on the key performance indicators, whether financial figures or statistical figures, that really drive their business.

In recent years, the balanced scorecard methodology has become a well-accepted school of thought. A number of well-known management consulting companies offer implementation services, and software vendors are improving their applications to handle balanced scorecard reporting. Even very successful organizations could probably be more profitable by utilizing the same resources as they have today by developing a set of key performance indicators and a continuous follow-up process. Examples of such indicators could be:

- Average revenue per employee
- Customer satisfaction (e.g., as a number on a scale from 1 to 5)
- Employee retention rate

Most modern BI systems now have the capability to allow for loading or entering all the data needed to create a balanced scorecard and they have the report-writing features (for advanced calculations), as well as the analytical functionality (such as drill-down and graphics) needed. As balanced scorecards are being offered by BI software vendors, large organizations with applications from multiple vendors in use around the world have found it necessary to integrate their scorecards. Vendors are now solving this by implementing XML (read more about this earlier in this chapter) and a set of standards for integration.

The Hackett Group (a consulting company that maintains the world's largest and most comprehensive ongoing benchmark studies of different knowledge-worker functions) has found that the information that eventually reaches top management is often in too much general detail and lacks specific focus on the key financial and statistical drivers of the business.

The idea behind the balanced scorecard is to address this issue and to give decision makers a performance measurement and planning tool that is easy to understand and focused, and that can help the company link its strategic plan with key performance indicators.

Why use balanced scorecards in the reporting and analysis process? The way we do business is changing. It used to be simpler to keep track of the performance of a business and to create future plans. However, as key business conditions are changing, performance measurement, reporting, and analysis also need to change. The idea of the balanced scorecard is to focus in on the financial and statistical measurement that really drives the company and not waste time and money on planning and reporting for every possible line item in every corner of the business. The balanced scorecard approach also encourages management to include general industry comparisons as well as competitive comparisons in the planning and reporting. (See Figure 1.13.)

BETTER MANAGEMENT TOOL THAN STANDARD REPORTS

The majority of companies utilize the general ledger as their key information source, so management reporting remains largely driven by the accounting close. One of the benefits of the balanced scorecard is that the reports are very easy to read and every piece of data contains key information about the business. This means that management does not get distracted with nonpertinent details and they can spend their time analyzing the real drivers of the business.

Maybe one of the most important aspects of the balanced scorecard approach is that it is a tool that is directly linked with the corporate strategy.

Business Driver	Yesterday	Tomorrow
Competitors	Few/Big	Many/Small
Markets	National	Global
Manufacturing	Mass Production	Flexible
Products	Standardized	Innovative
Service	Standardized	Custom

FIGURE 1.13   The way we do business is changing

Whereas most companies today have a reporting process that is not tightly integrated with their strategic objectives and thus has much less value as a business management tool, the metrics in the balanced scorecard should be developed following the setting of strategic targets and the related tactical plans.

MOTIVATES AND INVOLVES EMPLOYEES

Another key focus in the balanced scorecard approach is to link employee compensation to the performance metrics. This increases the focus and effort spent on achieving the targets set forth in the balanced scorecard. And because the reports focus on the key performance indicators, it is easy to follow where the company is going every period when actual figures are measured against budgets and forecasts.

POTENTIAL PROBLEMS WITH THE BALANCED SCORECARD

Almost half of all large companies use some type of balanced scorecard, combining operational and financial measures. However, the impact is not pronounced. For those companies claiming to use the faddish tool, almost three-quarters of performance measures still are financial. This is little better than companies not utilizing the balanced scorecard, which rely on slightly more than 80% financial measures. Furthermore, if too many measures are chosen, the amount of time necessary to capture the data may outweigh the value of the information.

Impact of the Internet

Business intelligence applications will empower users by giving them access to easily analyze and visualize essential financial and statistical data from the company's different transactional databases. The Internet is making BI applications even more powerful for several reasons:

- Data sources such as exchange rates, competitive or industry measures, and so on can be accessed online and used by the BI applications.
- Convenience—the BI application itself can be web-based, allowing users to access the system from home, while traveling, or from remote offices.
- Speed of data distribution—information is available immediately from anywhere.
- Speed of data collection—budgets, statistics, comments, and the like can be entered from anywhere.
- There is no need for software installation and maintenance on users' computers.
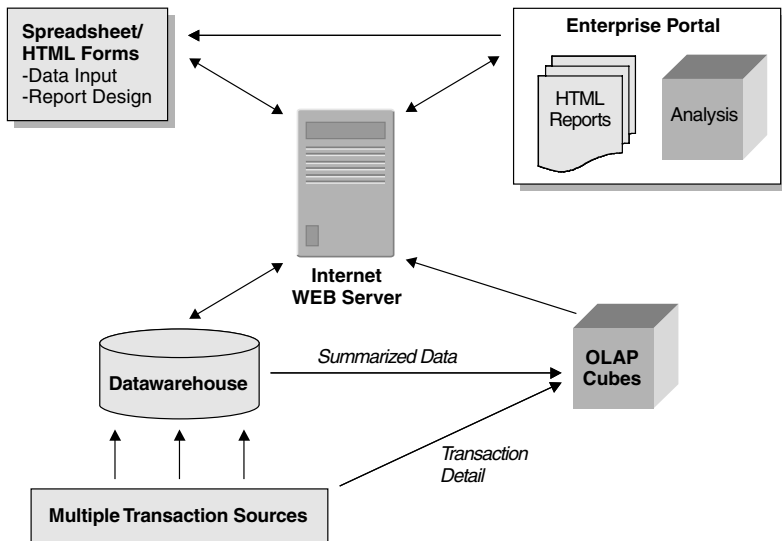
FIGURE 1.14 Web-based BI System Configuration Example

Most BI vendors now have a web version of their software, and as long as you have the infrastructure in place (web server and Internet connection for all users), you should seriously evaluate making your BI information available through a web-based application. (See Figure 1.14.)

Make sure you check that speed, security, printing quality, and functionality are not below what you and your users can accept. Although most BI vendors still require some setup and maintenance work to take place on a regular client software installation, rich functionality is quickly becoming available in their Web versions.

As not only your BI application but also many of your information systems become available on the Web, the next natural step is to integrate your BI application into a corporate web portal (also referred to as an enterprise information portal [EIP], or often in the area of financial BI, a digital dashboard). (See Figure 1.15.)

Over the next decade most employees in any modern organization will access all corporate information through a single Internet or intranet-based portal. EIPs provide employees with a single point of access to corporate information and resources. The technology provides a web-based interface that channels content so that individuals need to see only the underlying business application modules or information that is most relevant to them. For example, in the context of a company's financial staff members, a CFO can have a customized portal with sales information, cash positions, and actual to budget variance reports, while an accounting clerk can have a portal view that provides
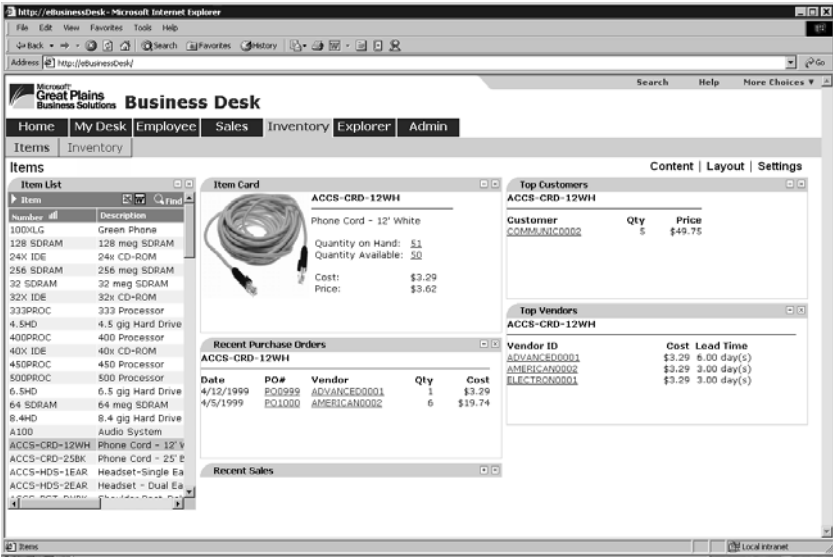
FIGURE 1.15   Enterprise Information Portal Example

Source: Used with permission, Microsoft Great Plains Software, Inc.

access to accounts receivable and general ledger journal entries. Applications and information that can make up a portal include:

- BI applications
- ERP application (accounts payable, accounts receivable, general ledger, sales order processing, etc.)
- CRM applications
- Company news and policies
- External information

Today, a number of organizations already have EIPs up and running. This means that they have a common web site that provides access to all the different web-based software applications in the company. However, in the first generation of EIPs these were just *links*, and not live information or summaries fed directly from the underlying applications. Most software vendors have created or will need to create a special front end for their solution to plug straight into a portal, instead of being just a link from a portal to a proprietary web interface for their package.

Recently, a new generation of portal software has arrived that provides functionality like:

- Display of many sources of data in a single user interface (*dashboard*)
- A single log-in to the portal with user security linked to all the applications in the portal, to avoid multiple log-ins to different data sources
- Very simple (drag and drop, and menu-based) customization of the portal so users see only the information most relevant to them

A portal category that might grow quickly in companies where BI is a key focus is a corporate portal with built-in report writer, graphics, and business rules, and with the additional ability to directly pull information from the transactional data sources. The latter ensures that users always get access to the latest data because there is no temporary storage system (like a datawarehouse or data mart). However, a portal that uses third-party, specialized BI applications to provide data and analytical functionality might be faster and offer more powerful BI features.

With the many exciting and useful applications the Internet has for the field of business intelligence and related software, IT and financial managers will have a number of new technologies to stay abreast of in the coming years.

eLearning: The New Way of Bringing BI to End Users

What does it help if you implement a powerful new BI solution to make valuable information available to users if they don't know how to properly use the software? If this book had been written just a few years ago, this topic would not have been included because eLearning was not even in our vocabulary. What is eLearning? The "e" in eLearning, like so many other new "e" words in our vocabulary, stands for *electronic*. The word generally refers to an educational process that is carried out with the information and instructions typically residing on a data device enabled with sound and animation and accessed through a computer. Examples include:

- Multi-media CD-ROM courses (have been around for many years)
- Self-run or instructor-led web-based courses

Here, we will focus on eLearning as a web-based process, as this is clearly the trend of the future. Companies are constantly introducing new technologies to their workforce, as well as demanding that employees stay up to date on new or changing information, such as budgeting requirements, financial reports, and so on. Training costs are in many cases skyrocketing and this

has forced corporate managers and training providers to look for alternative technologies and methods to support their educational initiatives.

It is not uncommon that companies get a poor return on investment on their software applications due to lack of training. Maybe John Doe went to a three-day training class, but what does that help if John left the company last month, and Lisa, who was trained by John half a day before he left, is now sitting there with the frustrating task of running the system? Or what if Company A has 80 users at 30 locations who need the one-day end-user class to start using the great new BI tool you just installed and implemented? Sending a trainer to 30 locations or sending 80 users to a training session can quickly become a $100,000 investment. Not only that, but a massive amount of scheduling must take place to fit this into everyone's busy calendars, and even then, there is a good chance that a few attendees can't come due to sickness or other unexpected events. Last but not least, even for all the people who come to a training session, many will miss out on important information due to:

- Poor attention span
- Leaving class to check messages or make calls
- Slow learning abilities

Now, imagine that the one-day training, which probably in effective presentation time consisted of about two hours of continuous presentation, was available on the Internet. If your new BI application was ready to run this week, you could send an e-mail to all your end users and ask them to go to the web site where the course files reside during the next week and they can take the courses as many times as they want to and whenever it works best for their own schedules. Each user will go to the given web site, probably download the same manual as would have been handed out during the regular training class, and take the course by clicking on the right course module on the eLearning web site. Normally, eLearning courses are broken up into short modules consisting of 2- to 10-minute sessions with text slides, graphics, screen videos, sound, and the like) that is easy to find and quick to finish or to move forward or backward in.

In order for web-based eLearning to provide a satisfactory learning environment, it is necessary to make available to users' computers with sound capability (including headset or loudspeakers) and an Internet connection with reasonable speed (at least 50 kilobytes per second for many eLearning solutions).

eLearning is also a great tool for repetitive training. Even if users were trained at some point in time, chances are that they do not frequently use all the features of the software and they will forget certain parts of it. With all the course modules residing on a web site, the users can log on and retake specific course modules at any time.

Most eLearning providers will offer courses on a per-time basis or on a subscription basis. If you have a customized or home-grown BI solution and no eLearning provider yet offers courses for your software, you can create your own online courses and put them on your web site, intranet, or local area network (see Chapter 13 for a list of eLearning software and vendors). This option can also be of great value if you need to train a large number of users in a few modules, and you need to show them specific features and functionality. Depending on the resources available in your organization and their skill level, course development can be handled by your own team or outsourced to the eLearning software vendor or a consultant. Depending on the quality required and the amount of custom animations, graphics, and the like that are used, one hour of online courses can cost from a few thousand to $50,000 to develop. The key is to evaluate the cost/benefit and then pick the best alternative. There is also some very inexpensive and easy-to-use software you or a power user can utilize to create the courses yourself. It all depends on how many people will use the courses and how long they can be used before they are out of date, as well as the time and resources you have available.

Software training (like BI software) lends itself very well to online courses because you will see the actual screens you will be using in real life, with information, buttons, and menus residing on a screen, and what can be better than showing the real screens and a mouse clicking on items while an instructor's voice is explaining in the background?

A large number of companies are planning to provide BI applications to their users over the next few years to unlock the information "gold" residing deep in their corporate databases, and the new eLearning devices can help make the BI software investment a great decision, and it can help end users quickly get up to speed on the application and help them stay that way in the long term.

ENDNOTE

1. www.XBRL.org.

2

LEVERAGING THE POWER OF BUSINESS INTELLIGENCE TOOLS

By now, most companies have implemented software to support all major financial and accounting activities in their organizations. There is also a strong trend toward implementation of powerful systems for managing a company's relationship with its customers (customer relationship management software) and its suppliers (supplier management software). However, these systems' main purpose is to record, organize, and retrieve (by lookups or reports) information that resides in their specific database, and they are not built as analysis tools. Even though some of the systems offer limited analytical features like graphs and drill-down, these features cover only their own application's database and cannot be used with the other systems in the company. This is the reason for one of today's great frustrations of corporate managers and analysts: It is very hard to get access to summarized as well as detailed information through a single user interface that spans across all the transactional databases in the company. The result is that the people responsible for operating the different systems spend a great deal of their time writing, maintaining, and printing their reports for management.

There are many limitations and issues relating to decision making based on static reports from different systems:

- *Information overload.* It is not uncommon that hundreds of pages are printed every month to provide information about the different areas of

the business. For a decision maker with limited time available, it can be hard to find time to read through every line of every report (and usually they don't).

- *Lack of information.* Almost contradictory to the "information overload" described above, although managers get a lot of pages of reports, there is also a lot of information they do not get. Even with hundreds of reports produced every month, it is virtually impossible to print enough reports to provide all the different views a manager might need to see to understand why certain variances or trends occur. Even though the information is available in the underlying production databases, it would simply be too many reports if every possible combination of information were printed.

- *No interactivity.* When managers find a number that needs to be investigated further (beyond what is showing in the printed report), they are dependent on more reports to be produced to find the answer they are looking for. And often, underlying systems are not capable of creating the reports asked for. Rather than the person finding the information him- or herself in a few seconds, at best it can take hours and the involvement of other people that are experts in the underlying report writer(s).

- *Lack of unified cross-database analysis tools.* Even if a company has created combined databases (or spreadsheets that hold data from several systems) to simplify reporting, these systems are still for the most part used to create static and often high-level reports (e.g., monthly financial statements). They lack interactive analysis tools that can drill to detailed data from many sources and that do not require predefined reports.

NEW BREED OF BUSINESS INTELLIGENCE TOOLS

The solution to the problem of a poor analytical environment in a company with multiple data sources, different report writers, and lack of analytical tools can be to implement a datawarehouse with a modern business intelligence software as a front end for the users (see Chapter 13 for a list of datawarehouse and business intelligence products). With modern ETL (extraction, transformation, and loading) tools and most database vendors now supporting open standards protocols, it is finally becoming feasible for companies to implement datawarehouses that can be updated and maintained with relative ease. The result is a common data repository, which provides decision makers with endless possibilities for investigating (data mining) and analyzing variances, trends, and exceptions. Because a modern BI solution feeds off a frequently updated datawarehouse that includes detailed information, it becomes much more than just a tool for executives (like the old executive information systems [EIS]

were), but it can become a tool for any person within the organization or related to it who needs easy and fast access to summarized and detailed information from across the company's databases. In the rest of this chapter we will look more in detail at a number of uses of BI software and popular features offered.

Examples of BI Software Functionality

Business intelligence software is getting increasingly more powerful, and although most of the popular solutions on the market share much common functionality, some might be easier to use or offer more or fewer features than others. The following list is some of the key features offered by many BI tools (for a comprehensive description of BI functionality, refer to Chapter 7):

- Drill-down (on dimensions such as time, company trees, products, etc.)
- Graphs, charting, and trees
- Exception highlighting
- Pivot rows and columns
- Drag and drop dimensions into the current view or to the background
- Custom calculations (calculate new measures based on existing ones)
- Queries
- Comments
- Combo views
- Dashboards
- BI and web portals
- Distribution of cubes/reports
- Other popular analytical features (ranking, filtering, sorting etc.)

Examples of Business Views Provided by BI software

The different views you can generate with a good BI tool is for the most part limited only by the data available to you. If the data mart or datawarehouse can provide rich, timely, and well-structured and cleansed information to the BI tool, it should take only a few seconds to use BI software to generate highly valuable views of your business. Before we look at different examples of financial data views, it is important to understand a bit more about the dimensions (data fields) that can be provided by different financial data sources (most of which would come from your company's ERP modules). Figure 2.1 shows

Data Source	Typical Dimensions	Sample Analytical Viewpoints
Sales order entry	Employee (sales person) Sales territory Customer Customer class Customer location Document (invoice) Department/Company Item (product) Item class (product group) Time (date) Amount Currency	• Sales by top customers (Fig. 2.2) • Sales by top customers with time comparison (Fig. 2.3) • Sales by customer class (Fig. 2.4) • Sales by customer and salesperson (Fig. 2.5) • Sales decrease by customers (Fig. 2.6) • Quarterly sales by salesperson (Fig. 2.7) • Sales growth by salesperson (Fig. 2.8) • Rolling sales by U.S. region (Fig. 2.9) • Sales by customer by location (Fig. 2.10) • Top items sold (Fig. 2.11) • Rolling sales by item (Fig. 2.12) • Sales drill-down (Fig. 2.13)
Accounts receivable	Customer/Prospect Customer location Collection manager Salesperson Sales territory Department/Company Product Market Time (date) Amounts outstanding	• Aging periods—bar (Fig. 2.14) • Aging periods by salesperson (Fig. 2.15) • Aging periods by collection manager (Fig. 2.16) • Collections by customer—91 days and over (Fig. 2.17)
Bank reconciliation	Bank accounts Customer Department/Company Beginning balances Ending balances Current transactions time (date)	• Cash in the bank (Fig. 2.18) • Cash on First National (Fig. 2.19)
General ledger	Department/Company Time (date) Expense accounts Revenue accounts Actual/Budget/Forecast Currency	• Sales and profit by channel (Fig. 2.20) • P&L, actual, budget, and variance (Fig. 2.21) • Actual, budget, and variance by division (Fig. 2.22) • Example of a dashboard (Fig. 2.23)

FIGURE 2.1   Financial data sources with examples of dimensions and analytical views

typical dimensions (data fields) that can be analyzed (*note:* not all accounting modules are included in these examples).

On the following pages you will find a number of examples showing different business views you can create in a BI tool. Note that in real life the business views are dynamic and interactive, and if given access an end user can change dimensions and selection parameters on the fly. Also, Figures 2.2 through 2.22 are from one BI tool, and it does not necessarily reflect the features in all other tools on the market.
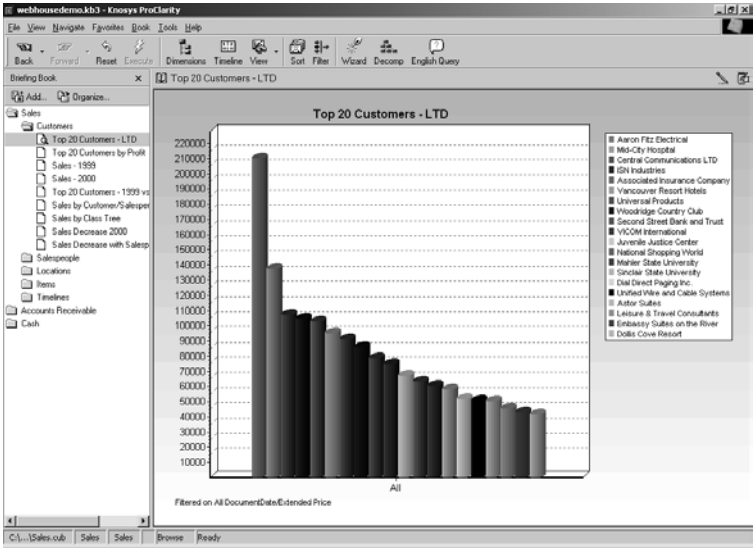
FIGURE 2.2   Sales by top customers
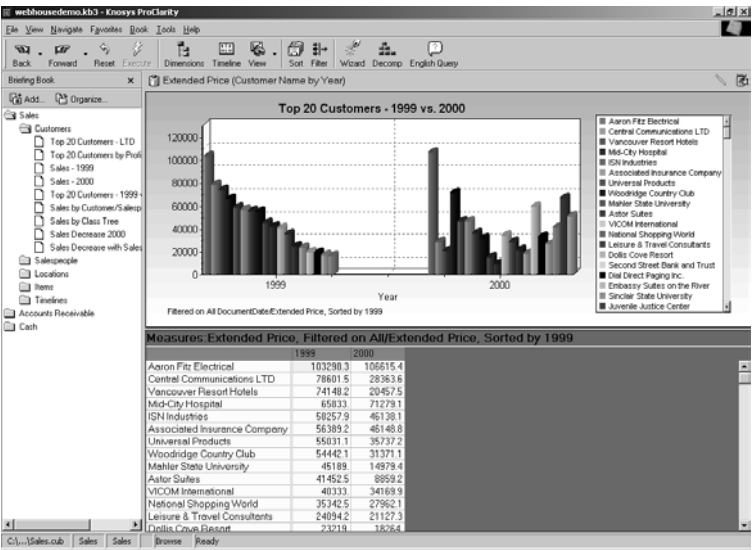
FIGURE 2.3   Sales by top customers, with time comparison

Source: Used with permission. Professional Advantage Pty Limited and 1999–2001 ProClarity Corporation. All rights reserved.
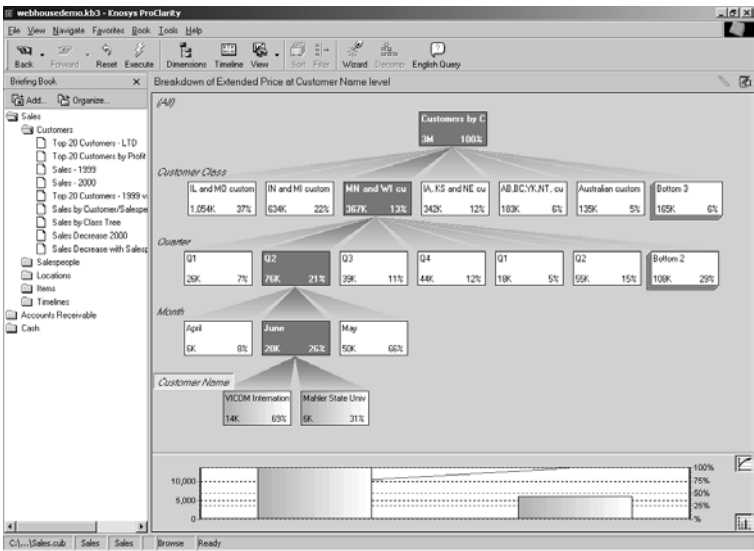
FIGURE 2.4   Sales by customer class

Source: Used with permission. Professional Advantage Pty Limited and 1999–2001 ProClarity Corporation. All rights reserved.
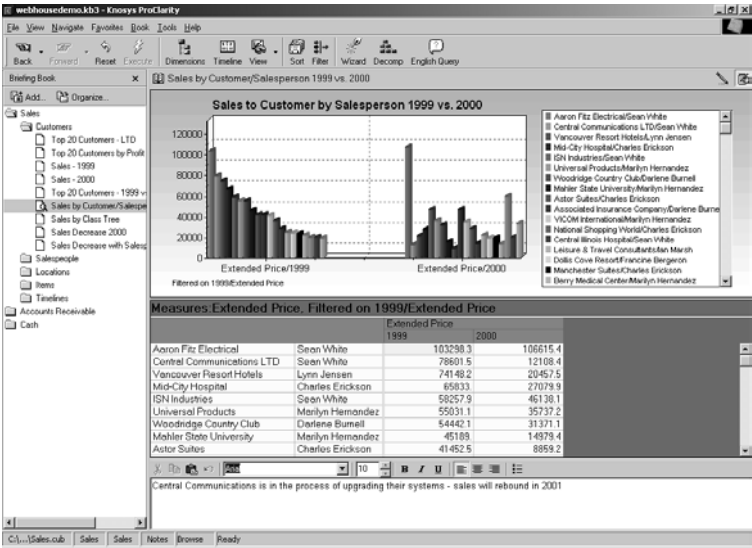
FIGURE 2.5   Sales by customer and salesperson
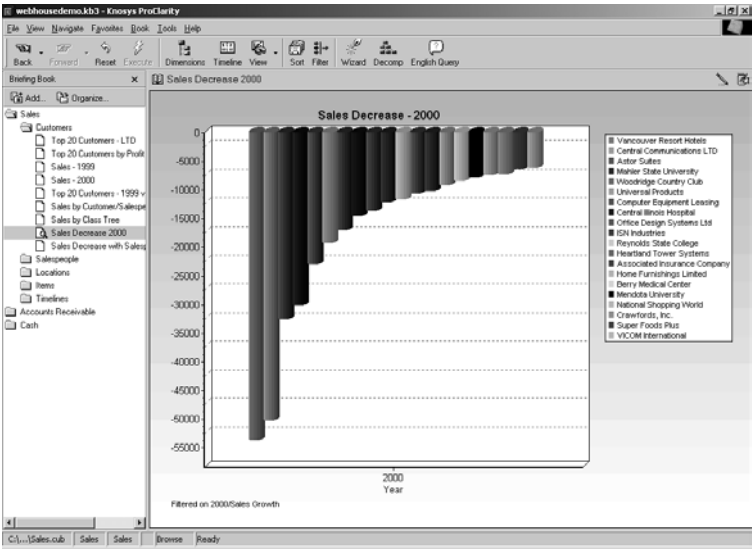
FIGURE 2.6   Sales decrease by customer
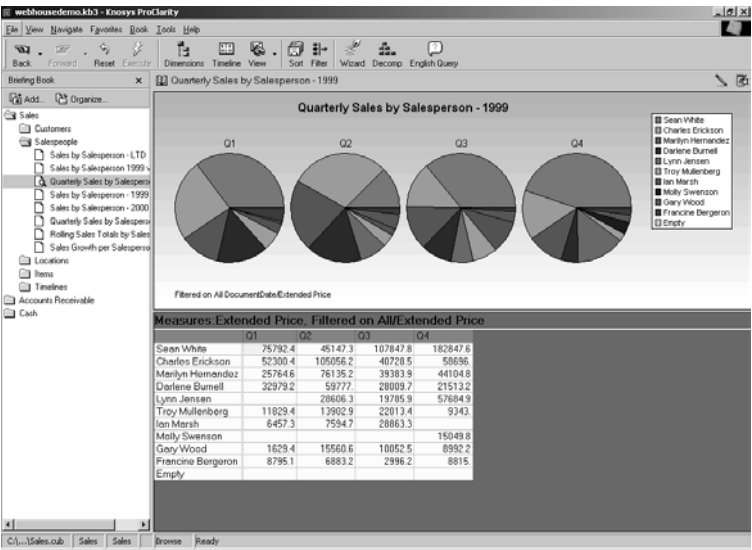
FIGURE 2.7   Quarterly sales by salesperson
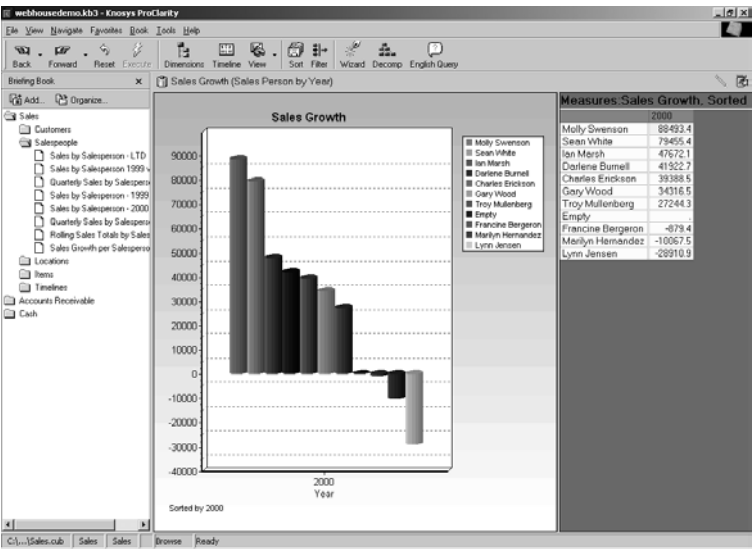
FIGURE 2.8   Sales growth by salesperson
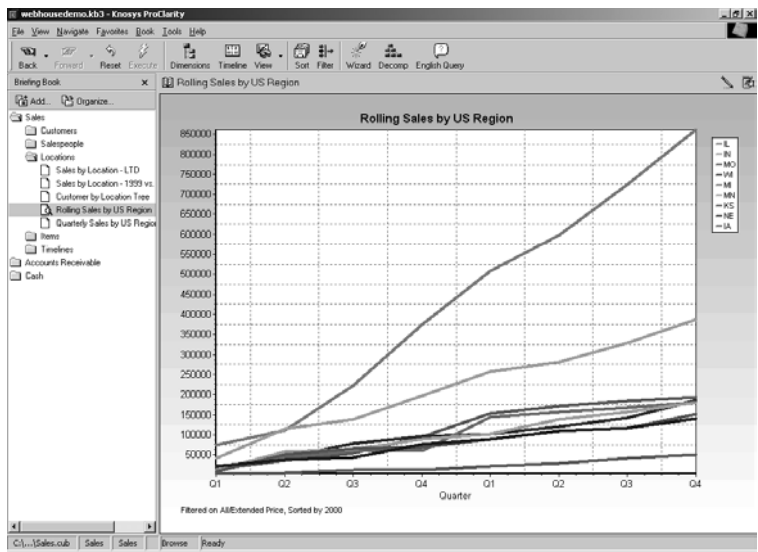
FIGURE 2.9   Rolling sales by region
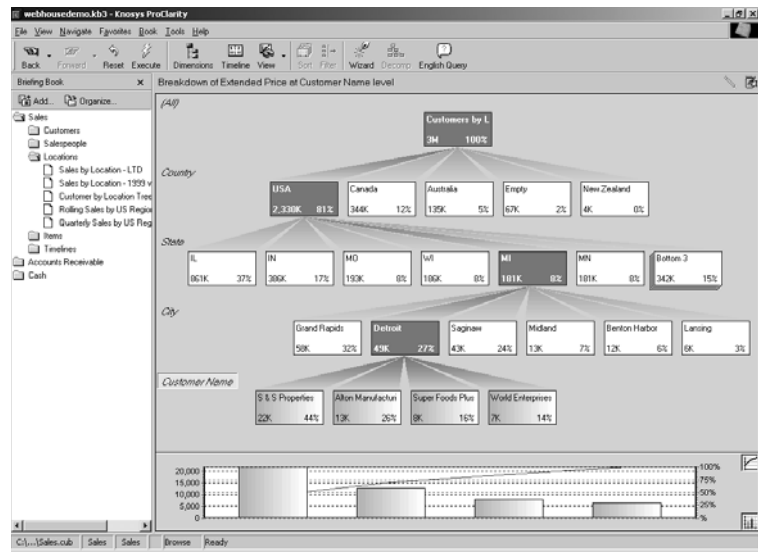
FIGURE 2.10   Sales by customer by location
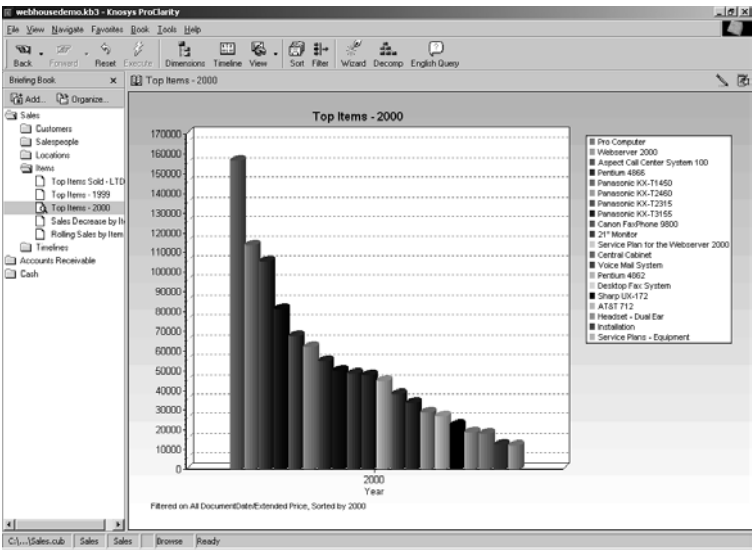
FIGURE 2.11   Top items sold
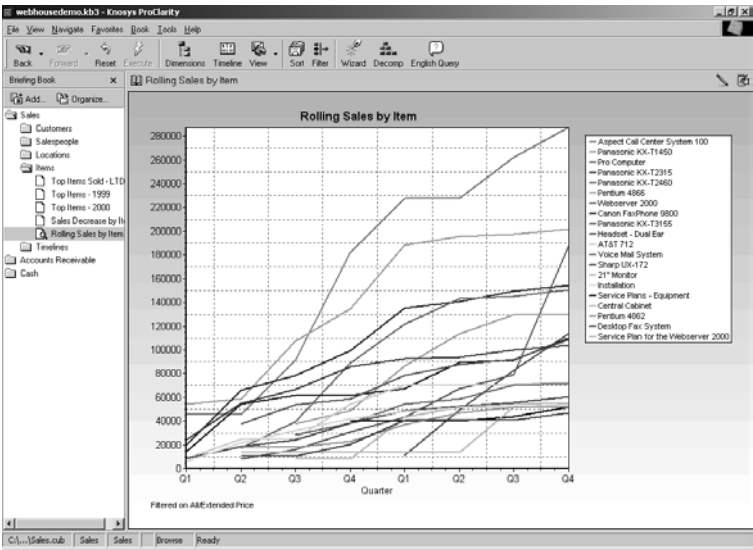
FIGURE 2.12   Rolling sales by item
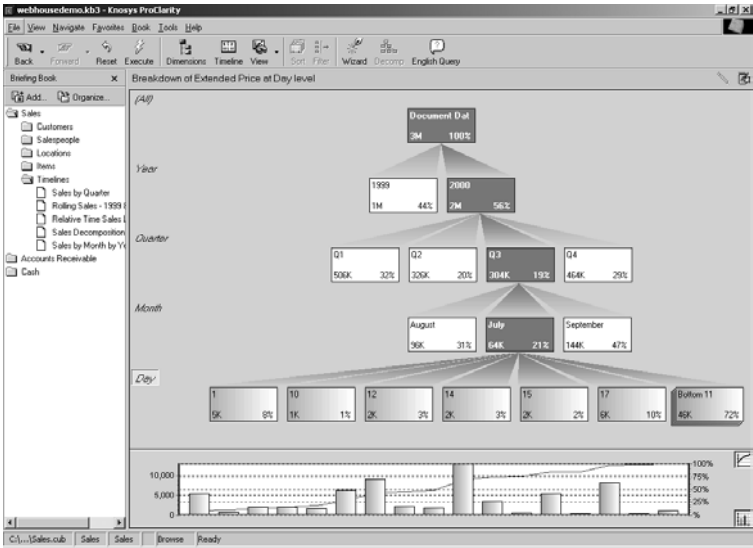
FIGURE 2.13   Sales drill-down
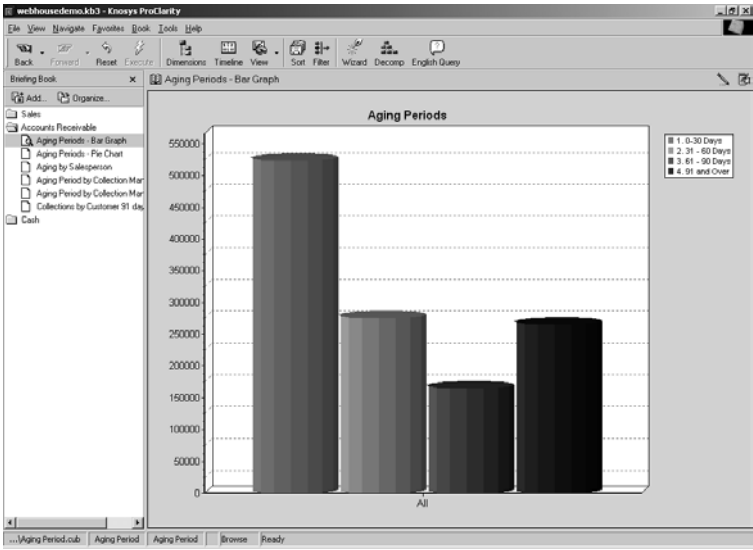
FIGURE 2.14   Aging periods: bar graph
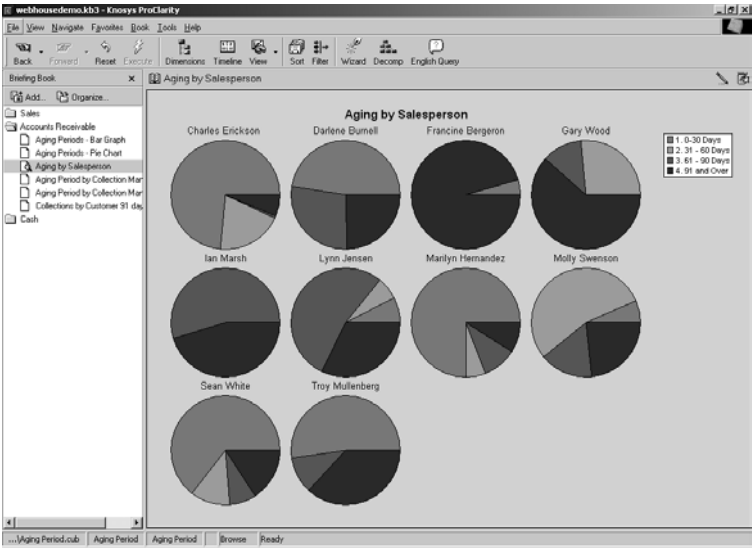
FIGURE 2.15   Aging periods by salesperson
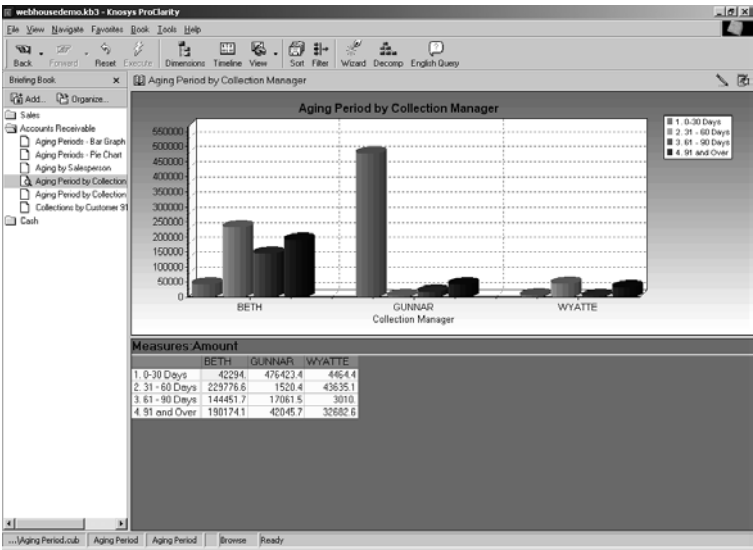
FIGURE 2.16   Aging periods by collection manager
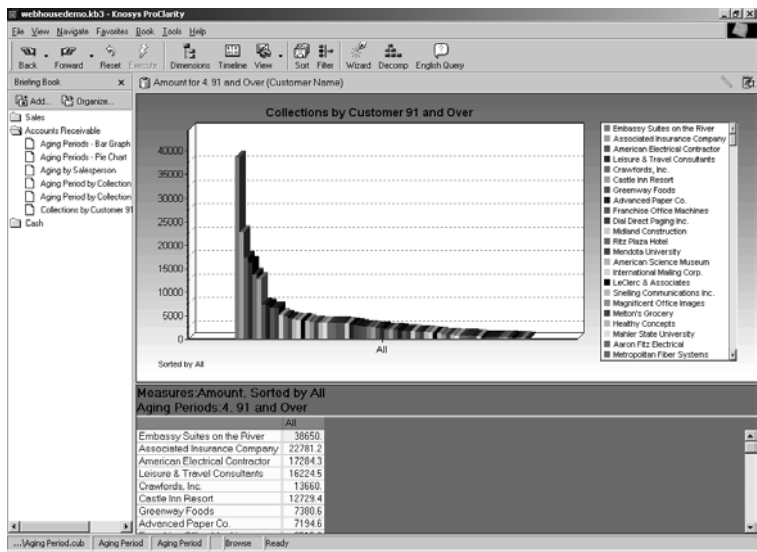
FIGURE 2.17   Collections by customer, 91 days and over

Source: Used with permission. Professional Advantage Pty Limited and 1999–2001 ProClarity Corporation. All rights reserved.
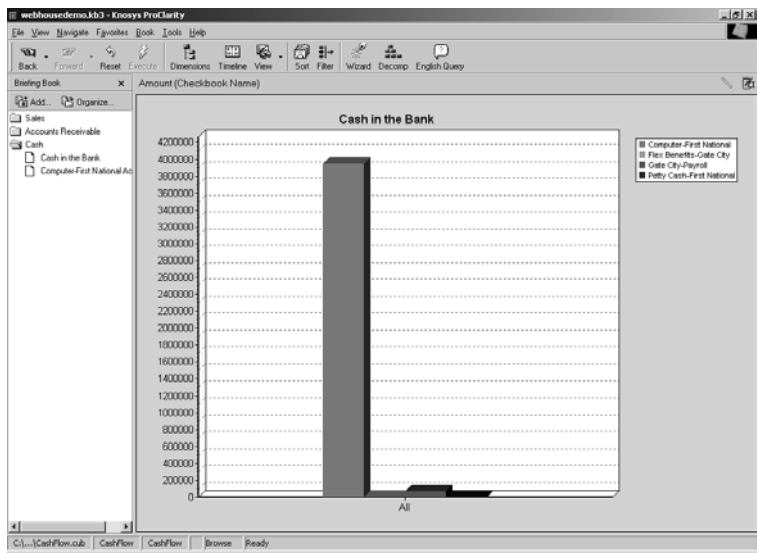
FIGURE 2.18   Cash in the bank

Source: Used with permission. Professional Advantage Pty Limited and 1999–2001 ProClarity Corporation. All rights reserved.
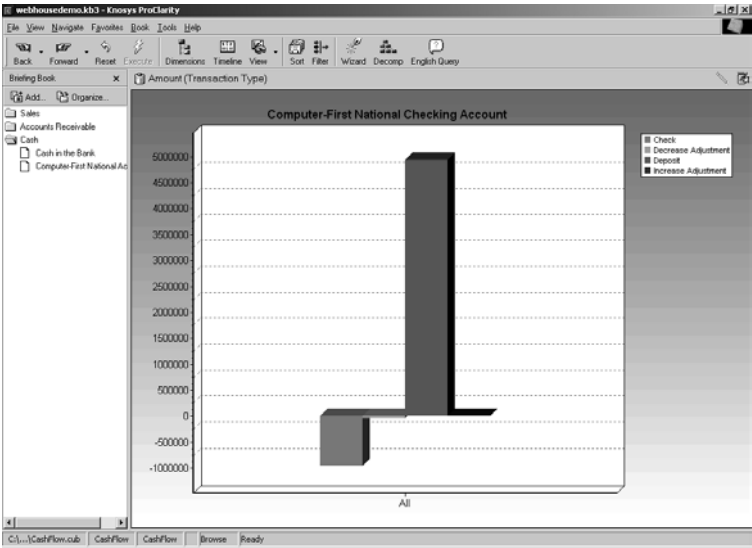
FIGURE 2.19   Check payments and account deposits
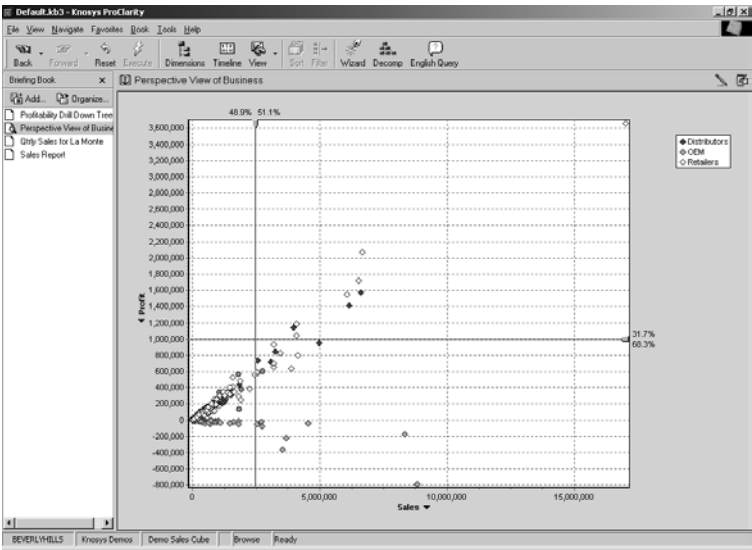
FIGURE 2.20   Sales and profit by channel
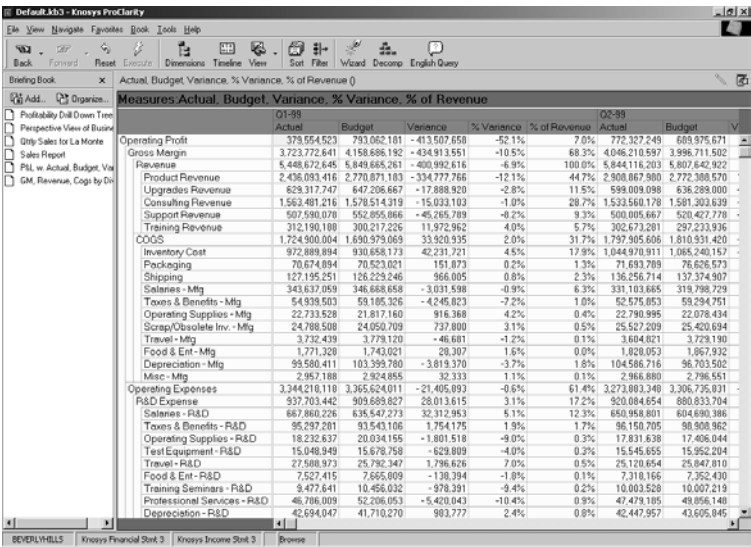
FIGURE 2.21   P&L, actual, budget, and variance
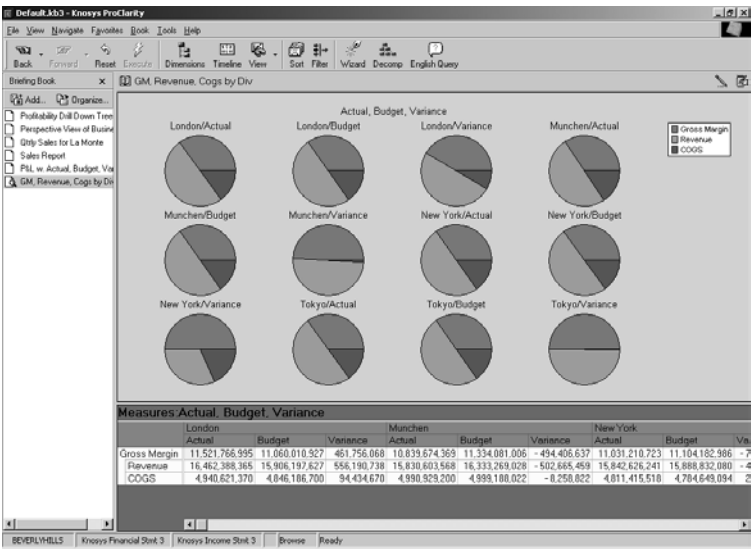
FIGURE 2.22   Actual, budget, and variance by division

FIGURE 2.23   Dashboard example

Source: Used with permission. Cognos, 2001.

3

WHY CONSIDER A
FINANCIAL BI TOOL
IN YOUR ORGANIZATION?

Every financial manager has felt the pain of having to ask the information sys-
tems (IS) department about new data from one or more of the corporate infor-
mation systems. And what usually comes back from them is rarely what you
asked for. Everything would have been perfect if only the people in the IS de-
partment would understand the difference between debit and credit, year-to-
date (YTD), or periodic balances, but this has not traditionally been considered
as a necessity of IS support personnel. Some larger organizations are specializ-
ing some of their IS staff by hiring IS personnel into the accounting department
trying to tear down these barriers. This is an expensive solution and doesn't nec-
essarily solve the problems for which they were intended. The protocol between
the accounting department and the IS department seems like a hard nut to crack.

Dealing with the never-ending requests for new, better, and more de-
tailed information from every area of your company seems impossible, doesn't
it? Senior management, members of the board, and department heads always
seem to have something new up their sleeve, don't they?

The proactive approach for the financial department seems to be to be-
come self-supplied with timely financial information and better quality on the
data in order to meet the never-ending request for more financial information.
Sounds incredible, but impossible? It's probably not! Many factors will affect
how hard or easy this will be to accomplish, but the fact that databases and on-
line analytical processing (OLAP) tools are becoming more easily available to

end users has actually made day-to-day tasks a lot easier for both the financial department as well as the IS department.

In order to initiate the process of taking ownership over a financial data-warehouse, you need to be a little proactive and analytical—listen to your "customers" while being a little "clairvoyant". Many companies have succeeded over the last few years by creating datawarehouses that host data from disparate sources and making them available to end users through tools they know like pivot tables in Excel, Internet browsers, and OLAP or managed query tools.

The availability of more powerful servers and the price drop per mega-hertz has contributed to the fact that this is no longer for only the elite of companies and managers, but is affordable and effective for the smaller and mid-market companies as well. Servers today can execute many times the transactions of the old mainframe systems. Computing power used to be a scarce commodity but this is not true any longer. So, the technology is there, the computing power is available, and the challenge is basically to collect all the information that you will need in order to be the "crystal ball" of your organization—being able to answer every question ever asked!

This chapter provides an overview of the steps you need to take in order to build your financial datawarehouse that hopefully will enable end users to make better decisions, your company to grow, and you to become the hero! The challenge you are facing is the fact that the dynamics of the queries and data structure in a datawarehouse will stop changing only when your business stops changing: *never!* This is where we will try to give you a helping hand!

The following chapters will discuss in more detail the different methodologies you can utilize, listen to other people's "best practices," and provide you with useful vendor listings as well as sample project plans and legal agreements.

DEFINING THE FINANCIAL DATAWAREHOUSE

What Is It That You Are Trying to Do?

Taking on the process of building a financial datawarehouse is the proactive approach to better *reporting, analysis, availability,* and *financial control* for your organization. What if you as the financial manager of a large corporation could have your own datawarehouse containing 80% of all the data requested on a day-to-day basis, without having to query different systems, paste all your findings into a spreadsheet, and then create it into some kind of report. Wishful thinking? Not really. In Chapter 19, you will see what some companies have achieved by implementing a datawarehouse. Not all have created a specific financial datawarehouse, but they all responded that participating in building the corporate data warehouse has made their life a lot easier; providing cleaner and reconciled data with ease of use and faster turnaround times to business problems.

Customers of the Financial Department

The ultimate target after a fully working model has been developed, tested, and in production for a while in the financial department is to make the financial analytics application available to your customers. Customers? Yes, by thinking of users of the financial department services as customers, you will actually be able to improve the products and services delivered with regard to timeliness, accuracy, and presentation. When the sales rep of the Western Regions asks for sales numbers for a specific salesperson, do not just look up that number from your sales order processing system, but send him the whole list for all the salespeople for his region, month by month and YTD. In the beginning he will be surprised and probably respond with an intimidating e-mail such as "Please just provide me what I asked for." Later, he will start realizing that he already has all of this good information, thus saving himself a couple of phone calls. He can now start spending time doing some analysis and projections with the data, eventually resulting in better business decisions. This is how the financial department can help its customers make better business decisions.

The business processes involved in getting your datawarehouse on the air are discussed below, and some of the more important issues are highlighted and discussed in the following chapters of this book.

The ultimate goal for your project is trying to cut back on expenses for your corporation, making your workday easier and more enjoyable, as well as breaking the cycle in which decision makers request information, and you spend valuable time collecting all data only to realize that they actually need more of the same data. If you can accomplish this within the boundaries of a budget and while keeping tight time frames, there is no doubt that you will be the corporate hero for a long time. You are not making the most important decisions for your company, but you are really helping out.

Find your Corporate Sponsor

The most important thing you can do when initiating a datawarehouse process is to find a corporate sponsor for the project. Like anything else, when changing a fundamental process, such as how decision makers throughout the organization receive their data, you need the anchoring of that decision from somebody at the top level. If you try it on your own, it is very likely that it will meet resistance.

Corporate sponsorship is not about hard currency only, but also all the other assets needed for a successful project:

- Providing financial funding for the process
- Defining business impacts
- Active communication to the end-user group, as well as communications objectives and progress to top-level management

Your corporate sponsor should have an active role in the project. As important as sponsoring the necessary resources for the project is the role for the corporate sponsor to attending status meetings and to be a catalyst for business processes and project milestones that have come to a halt. Defining the overall objectives and tying the business impact to other business projects are of crucial importance.

If your company is implementing a new general ledger (GL) system, redesigning their budget and forecast process as well as conducting a due-diligence process when acquiring a large competitor, it should be pretty obvious that this may not be the best time for implementing and designing a new datawarehouse. The project will be better off if all these business processes are completed one at a time. Since the design of your datawarehouse is dependent on all the processes mentioned earlier, it is obvious that it should be postponed. If a datawarehouse project was initiated by an eager soul in the IS department at the same time, he wouldn't necessarily be aware that all these other tasks already were underway, and failure would be the best possible outcome.

On the same note it could be that not even all the processes were initiated, but management had decided to acquire a competitor or initiate other activities that very likely would change the way information is being used or the analyses that were being performed. The responsibilities of a corporate sponsor should ensure that all these business processes have the blessing from high-level management and do not collide with other corporate goals and objectives.

Ask for what you need—not for the double amount, expecting it to be cut in half.

Find Your Team Members

As mentioned earlier, building a datawarehouse is a process. Once you have started it, you need to keep it alive by adding and modifying the data inputs as well as the deliverables. As dynamic as your business, your datawarehouse should always try to support and develop your business and managers' needs for information; therefore, it will have to change dynamically with your business and your business drivers. The most important resource in this process is the team members selected to participate in your project. You need to pay attention to the following items when selecting your contributors:

- They should have a diversified set of skill levels.
- They need to be able to devote time and resources necessary.
- They should represent different levels of your organization.
- They should have teamwork experience.

In successful financial datawarehouse projects, we have seen that developing a "skills needed" matrix can help you put together your team. When your

	DW Project Team	IS Department	Support Department	External Consultant
Understanding the business problem	●	○	○	○
Knowledge of how information consumers will perform analyses on the data	●	●	○	○
Knowledge about ETL tools and OLAP dimensions	○	●	●	●
Data modeling expertise	○	●	●	●
Prior experience with DW projects	●	●	○	●

FIGURE 3.1   Skills matrix

team is almost complete, match the skills requested with the skills of the people of your group. This will help you to adjust the members of the group to how your dream team was composed. The most important skills you should cover are:

- Standard query language (SQL) server experience
- Data modeling
- Understanding of financial reporting and analytics
- Interviewing skills, to work with the end-user group

This will give you the skills matrix shown in Figure 3.1.

Define Your Needs for Analysis + Reporting = Business Intelligence

The purpose of your datawarehouse (DW) is to provide better and faster information to key personnel of your organization, helping them make better decisions, as well as providing management with a better tool for understanding how the organization is performing. Since you are then left with the impossible role of describing the models of your complex business, we need to try to understand what makes up the rules of your complex business. This is where your DW architects along with the users of the data will play a crucial role.

The process in trying to describe your business can be summarized with the following processes:

- Data source
- Data input
- Processes
- Data output
- Analysis
- Vision

The most intuitive process would be to start at the top and proceed along to the bottom. This is known as traditional system design and a secure road that will bring you to your target. Not necessarily!

For BI purposes, we need to turn the process upside-down, so that the most important topic to first ask your user group about is their vision for the business. What type of analysis do they need to perform to execute this vision? This determines what type of output will be needed. The outcome of the questions asked above will again provide answers for what type of data you need to feed to the system in order to support the vision! By taking this controversial approach to designing your needs for data, it is given what data sources you need in order to complete the datawarehouse. (See Figure 3.2.)

When the information pyramid is complete, you should reiterate the process to make sure the financial business drivers have been accounted for, like:

Vision
Analysis
Data Output
Process
Data Input
Data Source
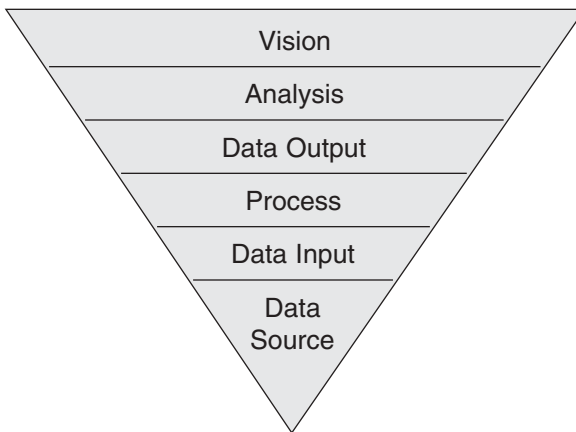
FIGURE 3.2   Information pyramid

- Is it the number of items sold or produced that is generating most of your expenses?
- What expenses are needed to generate the expected revenue?

These questions will be unique from business to business.

Model Financial Drivers

In order for the users of the financial datawarehouse to analyze historical performance of corporate operations, you need to make sure that these vital business rules are modeled into your DW. This is called business intelligence (BI). BI is what adds value from a printed report, to where the model itself can explain exceptions and variance in your data.

A financial DW developed with a focus on human resources (HR) data will be able to answer questions about Monday sick leaves, trends in wages for certain departments, and other questions related to providing financial analytics for employees in the HR Department. However, this business intelligence is in most companies mainly located in the heads of salespeople, senior managers, accounting departments, and other key people in your company. The challenge is to transform all this valuable knowledge into rule sets in your DW.

Users of the system are never a very homogeneous group; requests are coming to the financial department today from east and west for a wide variety of information. There are few reasons why this will change. On the contrary, you will probably see an explosion of requests to the system, and to what data it will contain when it goes live. Thus, it is important to do your homework and conduct in-depth research when modeling the system. We have already discussed the need for good data sources. Highly related to this is the financial drivers/business drivers that are driving your business forward and determine whether new projects will succeed or not.

When the end-user group is queried for such information, they are likely to answer with: "Show me what information you have, and I'll tell you what you are missing." This is based on the fact that it is always easier to modify something that some thought has already been put into, but don't fall into that trap!

With a sales manager, you should sit down and have him issue statements like:

- I want to know how many products I sold in every region in every channel by salesperson.
- What is the total revenue by product in a specific region?

Statements may seem vague and complicated, but they are excellent starting points for starting to model your datawarehouse. From his statement you can read that he wants to analyze:

- *Products*
- By *region*
- By type of *store* (channel)
- By *salesperson*

DIMENSIONS

The preceding highlighted items are dimensions. From this statement, you can see that you will need to have a dimension for products (e.g., code and name and group), region (e.g., state, North or Southwest), channel (e.g., store, e-commerce, mail order) by salesperson (e.g., code, name, years with the company). These are examples of properties/attributes of every dimension.

If you start listening to corporate executives and how they talk about their operations, you will be better able to model their financial world. There are a lot of indicators you can pick up by just talking to them, or you may also want to schedule more formal interviews with what they would like to see coming from the DW initiative. Sometimes a less formal setting will have them talk more openly about what they really would like to see. It does require some training, but if you start to analyze what you really were discussing in the hallway with the vice president of sales, soon you will be able to find yourself converting conversations to valuable business intelligence for the corporation.

MEASURES

From the interview with the sales manager in the preceding section, we can see that he needs to see *numbers of products* as well as *dollars sold*. These are both measures. Based on our own knowledge we can probably double the number of measures here, by anticipating that the sales manager would like to see the two measures above for both actual and budget numbers. We have now modeled with four dimensions and four measures:

1. Budgeted product sales
2. Actual product sales
3. Budgeted dollar revenue
4. Actual dollar revenue

By asking more and more of these questions you will be able to come up with a model that will serve the purpose of why you initiated this process. The more complex the sentences get, and the more of them you have to model together, the more important it is that you have somebody to assist you with data modeling. This is actually a field on its own, and it is recommended that you acquire this knowledge in your team, or have somebody responsible for data denormalization. Normalization of data models is too comprehensive a topic to bring

up in this book. But you will find a lot of good material by searching in the on-line bookstores or asking somebody in the IS department.

Model Your Data and Prototype

At this point you should have a better idea of what data you will need to provide the information requested. Now is when the real footwork begins. In this step you will start collecting, modeling, and prototyping the data model. It is important to seek assistance from internal resources or outsource this step if such knowledge is not available in-house. It's time to check the knowledge map you created when putting your team together to see who is best fitted to be responsible for this step, and to start some research on data-normalization theories. The most important factor is not that the data are completely optimized, but that you don't have to do redesign when adding more modules to your data warehouse. If you decided to store account numbers in common accounting system format, like a long text string, that is, xxxxx-xx-xxx-xx-xxxx, in a datawarehouse it is better to divide the different segments, such as natural account number, division, department, and subaccount into separate tables with code, description, and other properties. The long text string approach that GL systems are using works for GL purposes, but is not well suited for datawarehouses. It is very likely that you will add data from another data source, like headcount by department. This will be a lot easier if you have the data divided up into different dimensions, rather than trying to store information from position 5 to position 7 in a 25-character text string. (See Chapter 18 for in-depth discussion about account number design.)

Another important decision is how you will store data for your different business units. Do you want to use a centralized solution, or will it be more effective to maintain a distributed solution? In order to decide for your data storage you should consider the following elements:

- Will they be uploading local data in order to perform analyses that you don't have access to at HQ?
- Can your IT infrastructure support a centralized model? Important elements are server capacity, line speeds, and powerful client PCs.

Define Your Data Sources

By discussing what information managers need, and by modeling that data in the steps outlined above, you should have a pretty clear picture of where these data reside. An important source for financial data is your GL (trial balance) and the underlying detail for this. This might be a challenge if they reside in different databases on different operating systems from different vendors. Many companies try to solve this by uploading the different data sources into

a spreadsheet or a database (DB) program like Access Oracle, SQL, or DB2, depending on the requirements to clean, map, and convert numbers during the load and the area of use afterward. Other companies already have these trial balances in a financial reporting and consolidation system like Hyperion, Enterprise Reporting, FDC Commander, or other specialized consolidation and reporting systems. Any way you look at these data, they are and will always be a financial data mart of financial GL data, which is a great starting point for your datawarehouse. Your challenge is more in putting the data together, mapping, cleaning, and reconciling, so that the data has "financial validity." Why is financial validity of numbers so important? It is important that a financial manager analyzes and adjusts the numbers so they become comparable. For example, a university may book revenue in a 5,2,5 type of cycle in which the spring and fall semester tuition can consist of five months of data, while summer classes only last for about two months. Revenue from dorm rental may come in on a monthly basis. To make the numbers comparable, a person must tell the system how to split numbers out on a monthly basis to find the relationship between rental revenues and tuition. This can apply to businesses as well, where sale of service agreements may not have the same cycles as regular product sales. When defining your data sources, issues like these are important to take into consideration for number comparisons.

Define Your ETL Needs

There are many tools available to help you with cleaning numbers with regard to different charts of accounts, levels of aggregation, and other challenges that may arise through the business analytics of your data sources. In Chapter 13, you will find a list of the market-leading Extraction, Transformation, and Loading (ETL) tools.

Very few ETL tools are equipped to handle the special functionality you encounter when working with financial trial-balance data (GL data). You may find yourselves creating additional mapping tables that will assist the ETL tools in translating your data to a normalized form. Some companies create strict rules on how subsidiaries can create their account numbers, which makes the whole cleaning process a lot easier. Other companies don't have any rules, which leads to headquarters having to map such account numbers. There is no golden rule for the right answer, but a centralized solution is better in the long run. A centralized solution for creating new account numbers has many advantages when trying to keep a clean chart of accounts, but is often conceived as a more bureaucratic process. You will need to analyze your organization to see what procedure will fit best for you. What type of policies you have for creating number series for account numbers, departments, and customer numbers will help you decide the need for an ETL process. (See Figure 3.3.)

Company	SourceFile	From_AC	To_AC	Add_AC	Corp_AC	Debit_Credit_Flag	Factor	CurrCode	RateType	TransType
01	01*.imp	1010	1030		1010	1	1	USD	BSC	A,B
01	01*.imp	4000	4999		4010	−1	1	USD	ISC	A,B
01	01*.imp	5000	5010	6780, 8010	6710	1	1	USD	ISC	A,B
10	10-B*.imp	1010	1099	1101	1010	1	1	GBP	BSC	B
10	10-B*.imp	2000	2999		2010	1	1	GBP	BSC	B
10	10-B*.imp	4000	4999	8010, 8110	4010	−1	1	GBP	ISC	B
10	10-B*.imp	5000	6799		6010	1	1	GBP	ISC	B
10	10-A*.imp	1010	1099	1101	1011	1	1	GBP	BSC	A
10	10-A*.imp	2000	2999		2180	1	1	GBP	BSC	A
10	10-A*.imp	4000	4999	8010, 8110	4011	−1	1	GBP	ISC	A
10	10-A*.imp	5000	6799		6011	1	1	GBP	ISC	A

FIGURE 3.3 Transaction Mapping Table

In Figure 3.3, the first column describes the company, and the second column describes the naming convention on the upload files. From_AC, To_AC describes the account numbers in the source file. In this example, we have added a column for additional account numbers to include if the first two columns do not describe the selection. The Corp_AC will tell the ETL tool where to put the balance. The Debit_Credit_Flag indicates whether we want to change the default sign. Factor is used for multiplying numbers for inflation or other risk factors we want to apply to the data loaded. The CurrCode and Rate-Type tell us how to currency convert the data. TransType comes in handy if we want different mapping for actual, forecast, and budget numbers.

Selecting Your Tools

Selecting the right tools *might* be the difference between success and failure in a datawarehouse project. Multiple factors will influence the choice of technologies:

- Performance needs
- Number of transactions in the database
- Frequency of use
- Distributed/centralized solution
- Number of queries

HARDWARE

Many projects have failed because the client or server hardware didn't scale to the final solution. Who wants to be the user of a system that takes several minutes to come up with the answers to your queries. Probably not your end users. Instead of planning using the existing equipment, you should budget to purchase to run the new application on a new hardware configuration. Of course, you can develop your system on existing equipment, but you do want to make sure that the resources have been allocated to purchase necessary hardware when you are going live. If testing reveals that you really don't need to beef it up, you can even present a savings from your project.

Figure 3.4 shows the configuration a minimum requirement will consist of. In addition, you should consider the need for equipment to perform processor-intensive tasks like:

- *Report server*. Batch running of reports
- *Load server*. To handle all upload of data
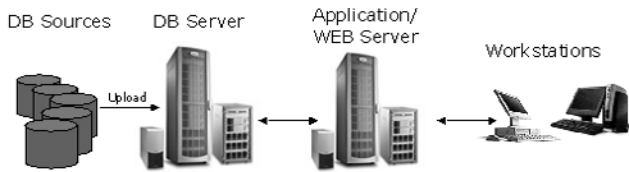
FIGURE 3.4   Hardware configuration

Security doesn't come for free either, so you might want to talk to the security personnel if they have the required routers and software to handle your rollout. It could be that you would need to have extra firewalls to handle the traffic of confidential information.

SOFTWARE

You will need to purchase some software tools for your project. You will at a minimum need:

- A database server license
- ETL tool
- Managed query/OLAP tool

You should look into whether you need a portal tool in order to make the data available to your users. Other factors may highly influence the total picture of your investment:

- What platforms does the OLAP tool run under?
- Do we need to upgrade the operating system for the servers?
- Do we need to buy more memory for the workstations?

Factors for Tool Selection

Chapter 10 is dedicated to how to deal with vendors, request for proposal (RFP), legal agreements, and other tools to ease the purchase software and consulting services.

Following is a list of the main topics of interest when considering your purchase and setup of the hardware needed for your solution.

DATABASE SERVER

The requirements of the database server are dependent on two factors:

1. Number of transactions present in the database in a query
2. Average number of expected requests per user per day

It is impossible to state something generic about these requirements. They must be analyzed on a case-to-case basis. When prototyping your model, you will have to simulate the number of transactions in an average query and try to simulate how many users will be running these queries.

WEB SERVER

Requirements for your web server(s) are dependent on the software selected to serve your end-user base. In today's marketplace, there are solutions ranging from totally web-based to thick clients reading data over a wide area network (WAN) or an intranet. If you go for a thin browser-based client, you may trade some of the presentation features of a thicker client, but you may have a lighter workload on the web servers. The more bells and whistles you add to your presentation and portal tools, the more load you will put on the client or the web server. It is important to have focus on printing capabilities of your presentation software, as the major browsers today do have limited functionalities here. One example is that none of the browsers in the marketplace today have the capability of reading the orientation of a report from Hypertext Markup Language (HTML) and adjust the printer properties accordingly. If you choose to go for a pure HTML version, your end users will print reports either in landscape or portrait, independent of the report on screen—not very user friendly! There are many other limitations to take into consideration with a web-based solution that will favor a traditional client–server solution that will solve all the traditional challenges with web-based applications.

BANDWIDTH

In disfavor of a thick client is the fact that the more features you add to the presentation tool, the longer it will take to download, and this usually puts a higher workload on the web servers. You need to ask yourself if you have the infrastructure required to support all the features you want to add to a user-friendly tool.

Example:

For a renowned *Fortune* 500 company, Excel was the preferred front end for the budget users to submit their budgets. Every feature of Excel was enabled. In order to enable input checks for valid codes like account numbers and product codes, these were automatically updated and checked toward any input cell in the spreadsheets.

Problem:

Reasonable? Yes, but the problem was the spreadsheets grew to 4 MB, and there were 400 concurrent users who should open these spreadsheets on a daily basis and save them back to a local web server over a WAN. The math here is 1.6 GB (x2 for Load and Save) of data flying back and forth over the network as users were opening and saving their workbooks. The only solution was to reduce the size of the spreadsheets (or start the search for a totally web-based solution).

As a rule of thumb, the more features you add to your solution, and the thicker the client download is, the more load you will put on your web servers. You will then need to come up with the extra dollars, contact the purchasing department, and get more hardware.

Plan Your Project

Every successful datawarehouse project has a detailed project plan with detailed tasks with assigned responsibilities. It is important to note that the plan in itself is not enough, but needs to be a steering tool for the process. Some of the most successful plans in real life consist of a scheduled weekly follow-up meeting in which progress over the last week is being updated.

As important as assigning tasks to attendees in a project is managing delegated tasks and progress. What do you do when you realize there is 0% progress on every delegated task? They were assigned to Joe, Jane, and Jill! Weekly status meetings in which every task is checked for status and progress are very important to every project. Project management is hard to learn; it's simply a lot of hard tedious work.

Execute!

If the average employee is asked about his or her relationship to a datawarehouse project, the employee may have worked for a company when the datawarehouse project failed, or attended one or knows somebody who was responsible for a major failure. You even have consulting companies that specialize in getting failed or derailed projects back on track. The major problem with all of these projects is that they were not executed according to plan, usually due to budget, timing, or resource constraints. The common denominator for projects that succeed is that they were all tightly monitored, personnel stayed motivated, and budgeted resources were within planned boundaries. It sounds easy, and it's the key for success.

There are a lot of previous experiences provided in written form out there that can help you to get your project on track; these are called best practices. The good thing is that you don't have to reinvent the wheel, but rather learn and succeed from other experiences. It could be called ignorance, not to do some

research on some of these road maps to success before heading out yourself. For recommended literature, see Appendix I.

Define Your Support Needs!

As with any other software implementation, you will need to have a support function for your application. As the skill sets of your user groups grow, they will require more and they will inquire about advanced features. Skilled end users will quit and need to be replaced by new employees and trained. Networks will go down, uploads will fail, and very few DBs have 100% uptime. Somebody will have to be responsible to answer for all these user requests and support needs.

DEDICATED SUPPORT PERSONNEL

It is important to budget and plan for a support organization that can be available and answer all these support requests. Don't expect the corporate support department to be capable of responding to all these type of questions for your highly specialized application after sitting in on a couple of days of training. Make sure you get dedicated personnel throughout the development/implementation process that takes ownership and acquires great knowledge about the background of the application, as well as the technical configuration, user demography, deployment, and training plans.

USE OF TECHNOLOGY

A well of technologies and systems are available on the market to ease the support function. These range from complicated knowledge-based systems to discussion groups to something as simple as possibilities for searching office documents. Many of them are web-based and some require complicated implementation, while others can be implemented on a low budget, basically overnight. Don't try to overkill with the most fancy technologies to assist end users in using the system—they are very likely to not be used. The most ingenious and most frequently used solution might be right there in front of you. Look to other implementations you have attended, and pull other resources to compare what works and what doesn't work.

There might be support systems already in place in your organization that with minor modifications can be of great help in collecting user feedback from the user community. This might be support systems, help desks, or knowledge bases as well as discussion forums. In the support system, there might be a list box that lists type of support. The options can be: Accounting, Finance, HR, and Manufacturing. Make sure that Datawarehouse is added as one of the categories in order for the end users to feel that this has been customized for their problem. Small changes can make the difference.

If you plan to have a discussion forum, not only should you let end users try to answer end users' requests, but be sure to have a moderator to keep the threads of discussion alive. Nobody will ever use a system when there are a few weeks between every response.

Make sure you establish a knowledge base for commonly asked questions and responses from experienced users that have overcome problems or flaws in the application. A searchable Word document can sometimes be your best tool.

You will never hit the bulls-eye on your first try (version 1.0). Many projects have gone south because key personnel have gotten 90% of what they asked for, but concluded it was totally useless. This can be avoided by having a database in which new user requests are added. These requests are then reviewed by the datawarehouse project team every month, or weekly if required. Every user request is then discussed and accepted and implemented or rejected. It is important that the end user who submitted the request get some feedback in order not to feel like he is shouting into the night! This is a proven technique that can keep the appreciation and acknowledgment of your group high up on the scale.

Train End Users

Many approaches can be taken to training end users. Train the trainer, coaching, and centralized and decentralized solutions are some of the options possible. Not as important as how you provide training to your end-user group is the fact that you take this task seriously! An unskilled user group that is being blamed for not understanding your masterpiece is your worst enemy. It's like a painter with an empty reception in a gallery. Training end users is not a thing you *have to do*, it is your opportunity to sell your solution to your audience. It's the chance for your project to succeed!

A good planned training program should have electronic documentation that is easy to search and print. Cover basic as well as advanced features. Make sure your teachers are well prepared and have technical skills, knowledge about the area, and pedagogical skills. Neither an introverted person with a low voice nor a technical whiz-kid that overwhelms your audience with technical terms will leave a very good impression with your user group. Your audience will probably have a very negative attitude toward the new application when they are back at their desk, and you have increased your chance of failure. The voice of the herds will sooner or later reach the members of the board and you will be to blame.

eLearning is another technology in which there has been an explosion over the last few years. This is basically web-based technologies for transferring knowledge from the datawarehouse project team to the end users, or a great method of training new employees in the in-house systems. See Chapter 1 for further discussions on this topic.

Promote Your DW: Deliver Something New Every Third Month

A worn phrase is that datawarehousing is not a product, it's a process—a process that never ends. You have taken on the task of describing your company with more or less metric measures. Your organization is an organic creature that constantly changes its character, sometimes every month, overnight, or the next hour. Thus, it's given that you will need to change the metrics of the DW. Users will have new requests, members of the board will sell off parts of the company, and senior management will want to see A compared to B, not X compared to Y. That's the nature of doing business. You are left with the difficult task of satisfying and complying with all these requests and requirements. It's not impossible, but be proactive: Deliver something new at least every third month! Review all of the user requests and implement the most popular feature, add some new graphs, add more reports, and add exporting capabilities. These are all measures that can be taken for the user base to stay satisfied and appreciative of doing business with you and your application.

Top 10 Factors of Success

10. *Integrate your business, not just your data.* Make sure everybody is involved and have his or her opinions discussed throughout the life cycle of the DW project. A DW project is different in nature from other projects. You are trying to collect data across departments, operating units, and officers. This has shown to sometimes be a major obstacle and a showstopper for a successful project. Break some barricades if you have to!

9. *Focus on deployment.* Another external factor is to do internal marketing of the project. This is a thin line to walk; you don't want to set the expectations too high, and you don't want to reveal too little. To keep everybody motivated it is very important to tell your team what's going on, whether it's better progress than you expected or the hurdles you are trying to solve. Think of yourself as a salesperson and the rest of the organization as your customers. What do they want to hear from you?

8. *Be focused on your project management and milestones.* Internally, you want to make sure that your project members stay highly motivated. You can do this by providing good project management and regular status updates in which you get the chance to give praise for outstanding performance and also let the group know where they need to improve. But, most importantly, celebrate every victory, and use it for what it is worth to build motivation in your team.

7. *Understand what's adding value to your customers.* Put your own priorities away and try to really listen to your users. What you think is cool is not necessarily as cool for the end users. There are multiple "Dilbert" sto-

ries that come to mind when thinking of the challenges that lie between the datawarehouse analyst and the developer. If you manage this part of your project, you have come a long way.

6. *Evaluate your data sources.* There are many criteria for evaluating your data sources: timeliness, reliability, frequency, and availability. Your DW is 100% dependent on feeding from other sources. It's like picking a vendor for your deli store. You want it to be fresh and on time.

5. *Datawarehousing is a dynamic process.* The scope will change and most likely increase!

4. *Set expectations and meet the objectives.* Deliver what you first predicted and what's set forth in the project plan!

3. *Not teaming up with the right people.* Sometimes, having the right people on your team is more important than having the right tools. People can make the tools work. You should try to optimize both, of course. Your selected vendors are also selected team members, so don't ignore the importance of selecting the vendor you believe in.

2. *Don't forget to listen to the end users!*

Question	Score (–10 to +10)
1. With the assumption that a BI implementation (including datawarehouse) will cost 10% of an ERP software implementation, is your company able and willing to make the investment?	
2. Do you have the IT staff to maintain a datawarehouse/data mart?	
3. Does your current transaction system allow (is it open or proprietary) for integration with a BI tool?	
4. Are your users "hungry" for easy access and analysis of data (i.e., can you "sell" it to your users and will they use the system?)?	
5. Do you have or will you set up an intranet or Internet infrastructure that allows users access to the BI tool from any desktop via a browser?	
6. Are you willing to make a continuous commitment to training your users?	
7. Have you identified a clear need for (or do you already have) a central repository of corporate data to use for reporting and analysis?	

Question	Score (−10 to +10)
8. Have you identified key drivers (metrics) to measure your company's performance after?	
9. Will you or other users be able to make better decisions if you have a BI tool with updated and easy-to-analyze and understand information in front of you?	
10. Will easier access to report and analyze corporate information improve your employees' motivation and quality of work life?	
11. Will easier access to report and analyze corporate information save significant expenses by streamlining information processes?	
Grand Total	

1. *Don't get blinded by all the details.* Nothing is as important for your project, as being able to see the forest despite all the trees! A lot of people get blinded by all the details in a project. They either lose control or they don't have an actual perception because they can't get a birds-eye view perspective of what's really going on. If you lose your view, you need to resurface and make sure you stay on track. A lot of projects look great the first week or two, but when there are too many loose ends, it's a different ballgame.

WHAT IS YOUR COMPANY'S BUSINESS INTELLIGENCE READINESS?

The purpose of this exercise is to help you gauge if your organization is ready for a business intelligence implementation. Use the "Score" column to grade each question from −10 (strongly disagree) to +10 (strongly agree). Interpretation of score:

Less than −25: Your company is not ready for or has no need of a BI tool. Investigate your answers to see if any changes would be worthwhile.

−25 to +25: You need to make changes before moving forward with a BI project.

Above +25: Your company is almost ready or ready to start a BI project. Investigate low scores and take action to improve them and then kick off your project.

PART TWO

BI TECHNOLOGY

4

PLATFORMS AND DIFFERENCES IN TECHNOLOGY

In this chapter we first look at the trends in technology and the mix of tools available for business intelligence (BI) with a special focus on the financial market. Identifying the players in the market is a key part of the selection process to weed out tools that may or may not help provide value to a financially focused business. Drilling down into the different platform choices significantly assesses the costs related to the different technologies. There are trade-offs between cost, scalability, functionality, and performance. Any person planning to buy a new BI solution for his or her company needs to know the strengths and weaknesses of popular online analytical processing (OLAP) and datawarehouse technologies and the different platforms they support. "The web is king" has become a realistic statement in today's business world. Assessing where BI fits into the web and corporate portals will be critical in deciding on technology, tools, and vendors for any BI solution. OLAP and datawarehouse are mentioned in the same sentence above. It is worth discussing what those terms mean before proceeding. *OLAP* is used to refer to the tools and structures associated with cube building and information delivery. This is a contrast to *datawarehouses* where data is stored in preparation for use in the OLAP tools and structures.

FINANCIAL AND NONFINANCIALLY FOCUSED TOOLS

Tools available in the BI market range from the new arrivals to the tools that have been evolving for 15 to 20 years. Along with age come the considerations of the issues of cost, scalability, performance, and probably most important, functionality, or the focus and intent of the tool. The tools in today's BI market were designed and implemented with specific target markets and audiences. Identifying the focus of these tools will simplify the selection of the tools appropriate for specific business needs, and what additional work will be required to make the selected tools meet the business needs.

Many of the historical tools available in the BI market grew directly from financial analysis needs. So, it is assumed that most tools that have been available for the past 10-plus years had, or still have, a financial focus. It's not safe to say that all tools historically available in the BI market can deliver all financial analysis needs for a business, but they could typically deliver 80% of those needs. When looking at the tools available, examination of both the data and delivery components is necessary. Data components identifies the data processing and storage tools needed to provide the multidimensional (OLAP) data model. These can be server- or desktop-based systems. Delivery components are the means by which information is delivered to the decision maker. A spreadsheet has been the traditional delivery component, but there have been recent additions to the market providing additional information analysis through visualization tools.

The historically financial focused tools are: Powerplay (Cognos), Essbase (Hyperion), Express (Oracle), Business Objects 5i (Business Objects), and iTm1(Applix). These tools are also the primary data component for these vendors and typically use a spreadsheet as the primary delivery component to place information in the hands of the decision maker. Most of these vendors have grown their business over the past few years to include graphical information delivery components and web-based solutions. For the spreadsheet components, information is brought into a spreadsheet as the result of a "query." The decision maker must define the information desired and arrange it on the columns and rows of the spreadsheet. The decision maker typically then performs some analysis on the information in the spreadsheet, perhaps running though a series of calculations and charting the resulting numbers—a typical "number-crunching" scenario. These tools usually come with predefined financial metrics, making the implementation of a financial analysis system simpler. The downside of these predefined metrics is that they may not be flexible enough to meet business-specific analysis needs.

Other tools that have come to market in the last few years and may not have a financial focus are: Analysis Services (Microsoft), Olap@Work (Business Objects), and ProClarity (ProClarity Corp.). These tools don't cover both the data components and the delivery components. For example, Microsoft's

Analysis Services is the data component, which relies on Excel as the delivery component. Microsoft also promotes a datawarehousing alliance in which other partners have developed delivery components. This is where tools like Olap@Work and ProClarity fit into the picture. Even though these tools are not financially focused, they still require serious consideration when choosing the tools appropriate for a solution. Information delivery components range from spreadsheets to graphical visualization. There has been a large step forward in delivering graphical information in the last few years. The benefit of using a graphical delivery component stems from the classic statement, "A picture is worth a thousand words." Even in spreadsheet-based delivery systems, the information is usually translated to a chart prior to the final delivery. These newer tools simply start out with a visual display to provide "instant" information and often support a spreadsheet delivery, too. They are typically adequate (or more than adequate) to deliver information necessary for financial analysis. But they may require more work on the implementation phase to ensure that the required metrics are delivered for financial analysis. The upside to this is that these tools are more flexible, allowing business-specific metrics that may not be constructible in financially focused tools that deliver predefined metrics and don't offer as much flexibility. These tools can also offer a solution to other segments of business analysis such as sales and marketing, production efficiency, supply-chain analysis, and others.

WHO ARE THE PLAYERS?

The primary vendors in today's BI market range from the historical ones that have been in the BI segment for 10-plus years, and some newcomers. Some of the top players and a brief summary of their offerings, and business focus are:

- *Hyperion.* Essbase is the flagship product from Hyperion. Essbase information was traditionally accessed via an Excel spreadsheet with a connection to the Essbase data source. They also have started offering web-based solutions, customer relationship management (CRM), and other horizontal markets.

- *Cognos.* Powerplay has been the flagship product in the past with strength of delivery in a spreadsheet. But in recent years Cognos has started to offer a wide variety of products to meet a variety of business needs. There are too many to recount here, but they are trying to cover all the bases from web-based to desktop delivery in spreadsheets and graphical information.

- *Oracle.* Famously known as the database vendor, they have also offered BI solution with Oracle Express. Again, this has been a traditional

spreadsheet delivery mechanism. Express will be discontinued, and BI will be incorporated as part of future database solutions for Oracle.

- *Microsoft.* Analysis Services is a data component for the BI solution. Microsoft is a recent entry into the BI mix. They have relied on a spreadsheet (Excel) as the primary delivery component. They also promote a partner channel to foster the delivery component, and the leading vendor in this category is ProClarity Corporation with their ProClarity product line.

- *Applix.* iTM1 server is the data component from Applix. In the past they relied on an Excel spreadsheet as the primary delivery component. However, in the past year they have expanded to offer solutions oriented around CRM, planning (financial), and others.

UNIX VERSUS MICROSOFT AS A PLATFORM CHOICE

Financial analysts should be asking themselves, "Why is this topic important for me? What difference does it make choosing the operating system for a BI solution?" Well, the Unix versus Microsoft choice is important because the tools chosen for implementation will limit what platforms are supported. And if there are limitations within an organization such as corporate standards or approved vendors, this will be an important issue and narrows the field of tool choices early in the selection process.

Unix has been the historical operating system of choice for BI solutions. This was driven by the horsepower requirements (processor speed and disk storage requirements) in early BI implementations. Referencing Unix in this context does not specify a particular flavor of Unix, but is used as a general term applying to historical minicomputer operating systems. *Minicomputer* refers to a class of machines that fell between mainframe computers and personal computers. Today, this class of machines has a number of synonyms such as *workstation*, *minicomputer*, and *super-server*. Some of the common flavors of Unix are *HP-UX*, *Sun-Solaris*, *AIX,* and *Linux.*

Microsoft is considered a new player in the operating system arena suitable for BI solutions. Microsoft has based its operating systems on the Intel microprocessor family. Until the introduction of the Intel Pentium processor, along with corresponding advances in inexpensive disk storage and random-access memory (RAM), the Microsoft operating systems couldn't produce enough horsepower to support a BI solution. Today, there's no question that Microsoft's operating systems offering (Win NT Server, Windows 2000 series) can match, if not exceed, the scalability and horsepower traditionally available only on a computer that supported a flavor of Unix.

Vendor/Tool	Supported Operating Systems
Cognos/Powerplay	Windows NT Windows 2000 Unix
Oracle/Warehouse Builder	Any Oracle 8*i* platform
Hyperion/Essbase	AIX v4.3.3 Sun Solaris 7/8 HP-UX 11.0 Redhat Linux 6.1 for Intel Windows NT 4 Windows 2000
Microsoft/Analysis Services	Windows NT 4 Server Windows 2000
Business Objects/Business Objects	Unix
Applix/iTM1	Windows NT Windows 2000 Unix

FIGURE 4.1   Supported operating systems

Most of today's tool vendors offer systems for different flavors of Unix and the Microsoft operating systems. Figure 4.1 provides a sampling of the tool vendors and the supported operating systems.

BI as Part of Corporate Portals

There are many reasons to drive BI information delivery to the portal (or "web") solution on a company's internal network or secure intranet system. There is even opportunity for data vendors to offer BI delivery to their customers via secure web sites. These are organizations that typically sell information such as market research, investment data, and health care information. Some of the most compelling reasons to drive a portal solution are:

- *Familiar interface.* Information delivered in a web browser is typically simple and intuitive to use. Nearly everyone in business today uses a web browser to gain information from the Internet. Yet, how many of those individuals have attended training on how to use that web browser?

None. Information delivered via a corporate portal is done in a web browser and should be as simple to consume as other information gained via the web.

- *Low maintenance costs.* Corporate portals centralize the location of the software used in the information delivery. With one location of software to maintain, this substantially lowers maintenance cost of software when compared to a desktop system that would need to be installed throughout an organization. Controlling the versions of desktop applications in today's world of "personal" computing is a substantial cost. Check with any support desk in an organization and they will identify software maintenance close to the top of their list of major costs.

- *"Thin" client solutions.* A "thin" client solution's primary benefit is lowering the requirements for the machine used by the information consumer. The target machine doesn't need to have substantial processor and disk storage requirements. A "thin" client solution displaces the processor and speed requirements to the server providing the web-based BI solution. The server does the majority of the work for creating graphical images, formatting displays, and managing the queries against the data.

Corporate portal solutions usually consist of more than BI information delivery. BI is part of a larger picture in the corporate portal. Most large organizations have existing internal web sites to deliver general information within the organization. Some organizations also use these sites to deliver reporting information throughout the organization. Incorporating BI information to one of these corporate portals is simply an extension of their current functionality.

5

PERFORMANCE: GETTING INFORMATION ON TIME

Performance for getting information is a key consideration when examining BI tools. The performance concern is not how fast the records in a transaction database can be individually updated or retrieved. The performance concern is in two primary areas:

1. The speed for data cleansing and staging after an accounting period has closed

2. Retrieval performance for the information consumers

The first area, speed for data cleansing and staging, is determined by the functional process within the business. There are many factors that will influence this process within the business. The speed is typically not limited by technology, but by a process within the organization.

The second area, retrieval performance, is directly affected by the technology and warrants a good look at the tools available. Performance is also influenced by the delivery mechanism. Organizations must ask themselves, "Do we want information delivery controlled by a central reporting section within our organization, or should it be at everyone's fingertips, or is there a point between those two extremes that make sense for us?"

FAST ENOUGH FOR YOUR BUSINESS

Closing an accounting period is a rigorous task for even the most experienced finance organization. The rigor doesn't necessarily stem from extra labor

involved during the end-of-month or end-of-year closings, but from the importance and meticulousness applied to the resulting numbers. Many people have been in a position in which upper management within an organization is waiting in anticipation to get the final details on the financial performance figures at the end of an accounting period.

Adding business intelligence and online analytical processing (OLAP) technology to the existing pressures when closing an accounting period might appear as a burden you could do without. However, the benefits will outweigh any additional work, and with proper planning and staffing, the cleansing and staging of the data can be a smooth process.

At the start of this chapter it was stated that the first performance consideration was the speed for data cleansing and staging after an accounting period has closed. In an ideal world this would be a simultaneous event. As the books were closed, so was the data staged clean and ready for further consumption. In reality, there is usually some time lag between these two milestones. The time lag between closing the books and having cleaned and staged data will be dependent on the following factors:

- When does the OLAP information need to be ready for consumption by the business user?
- The process for cleansing and staging—commonly referred to as extraction, transformation, and loading (ETL)
- Delivering reports, or desktop information availability

The time when an organization requires the OLAP information to be available varies depending on the business. An organization may have a mandate that the information may need to be ready within some specific time period (e.g., two days) after closing the books. More typically, this deadline is driven by the information consumers who always want information "yesterday."

The ETL process will determine the speed of this segment. Most organizations will simply "append" the newly closed periods' information to the OLAP system. This is typically a small period of time (hours) since there is only a new month's worth of information added to an existing store of historical information. Even as the information is rolled up across quarters and fiscal years, there is still only a new month's worth of information added. Occasionally, history needs to be rewritten to reflect a product or organizational change (see Chapter 17 for more information on *slowly changing dimensions*), but if these are isolated changes they shouldn't significantly affect processing time.

Information delivery typically takes one of two forms: standard reports or desktop information. If standard report delivery is the primary means for information dissemination, then the time consumed by this segment of the process will need to be taken into account. These reports can range from paper copy to formatted Excel worksheets available from a shared network location,

to an internal web site where information is formatted into HTML and available online within an organization. Desktop information delivery is more efficient from a distribution point of view. It also offers individuals the opportunity to format information so it is meaningful to them. Desktop information delivery typically requires a more educated user who is willing to learn the tools and spend a small amount of time taking advantage of their power. There are a number of BI tools that offer desktop information delivery today. Some of those tools come from Cognos, Hummingbird, ProClarity Corporation, Business Objects, and others.

PERIODICITY CONSIDERATIONS

Financial organizations operate on a clock that is driven by the closing period. Every organization has its own closing periodicity. Most fall into two categories: monthly or weekly. The closing period is usually the "right" time to add a new set of data to the OLAP information set. However, it's not the "only" time. Adding data to the OLAP information set should be driven by the organization's business needs, not the accounting rules. If an organization closes the accounting period monthly and waits until the end of each month to add new data to the OLAP information, there will be a segment of time in which the business demands recent information, but it is not available. If the accounting period is monthly, that segment of time is usually about 20 days. For example, if an organization closes the accounting period on the last working day of the month, and updates the OLAP data with that accounting period information, then there will be about 10 days wherein users can get relatively recent information. After that 10 days, the information is "stale," meaning not recent, and therefore not relevant in day-to-day information quests. If the information remains static over the next 20 days, it will not be used and information consumers will simply go elsewhere to source what they need, typically consuming larger and larger amounts of time from information technology (IT) or accounting staff personnel.

The effective method to avoid this pitfall is to update the OLAP data set with a higher periodicity than the close of the accounting period. A weekly update will satisfy most information appetites, but daily updates are sometimes required. Again, this is driven by the business need, not the accounting rules. With the OLAP information updates that don't fall on the close of an accounting period, a caveat must be issued that the information has not been audited as would a "closed" accounting period and that the information may be updated when the accounting period has closed. The recommended course of action is to revisit the OLAP information set when an accounting period does close and "rewrite" the intermediate updates with data from the closed accounting period. This tactic typically requires two ETL processes: an intermediate ETL process

that updates the OLAP information with the daily or weekly information, and another ETL process that removes (or overwrites) the intermediate information with the closed accounting period information.

A final periodicity consideration is the ETL processing time. Most organizations can get their daily or weekly processing time to occur within a few hours, but it is dependent on three factors:

1. Size of the data update
2. Platform for the processing (hardware and software)
3. Information delivery mechanism

With a daily or weekly update, the additional data set added to the OLAP information is not large in comparison to the existing OLAP information set. However, the size of the data update will vary depending on the organization and the level of detail contained in the information.

The platform for the processing will have the greatest impact on the ETL processing time. Platform is defined as the hardware and software configuration of machines(s) performing the ETL process. The processing impact is broken into two segments: hardware and software. Dwelling on the variations in hardware is another topic of discussion, but ETL processes are data intensive, which translates into a lot of disk input/output (I/O), and random-access memory (RAM) usage. ETL processes are almost never bound by the processor speed. Obtaining faster mass storage (disk) and using lots of RAM (to cache the pages from disk) will result in large—and very large—data sets. The software selected for the ETL process will have a greater impact on the ETL process speed than the hardware. Keep in mind that software is only as efficient as its design and quality of implementation. Since all ETL software vendors claim their tool is the best, how can you make a wise choice? Two things separate the very good tools from the mediocre:

1. Historical success of the ETL tool(s)
2. Independent references

Vendors that have been in the ETL tool business have stayed in that business by delivering quality software and providing the other attributes desired in a software vendor (support, product updates, responsiveness). Independent references are always a value when examining unknown products. Much like selecting a real estate agent or a medical specialist, an independent reference will go a long way in giving the ETL tool vendor credibility.

If the information delivery mechanism is a set of static reports that must be produced, then the time for report production must be factored into the time for information delivery. Static report production is often tacked on to the end

of an ETL process as an additional process. These static reports are typically delivered to an internal shared location in an organization's intranet, or to a secure web site. However, if the information delivery mechanism is desktop software that reads information directly from the OLAP data set, or web-based delivery, then there is no production time to factor into the time for information delivery. Information consumers can access information directly from the OLAP data set without waiting for formatted reports produced from a reporting system.

DATA STORAGE METHODOLOGY (ROLAP, MOLAP, OR HOLAP?)

ROLAP and MOLAP refer to the storage mechanism for the OLAP information. These subjects are touched on in this chapter on performance because the storage mechanism will affect performance in two ways:

1. ETL processing time
2. Query performance against the OLAP information

Before examining these two performance topics, it is important to understand the definitions of MOLAP and ROLAP.

The "M" in MOLAP refers to "multidimensional." Multidimensional OLAP (MOLAP) stores the aggregated values in a multidimensional structure on the OLAP server. (See Figure 5.1.) An aggregated value is a precalculation of a cell within an OLAP cube defined by its multidimensional definition. Aggregations are precalculated summaries of data that improve query response time by having the answers ready before the questions are asked. For example, when a datawarehouse fact contains hundreds of thousands of rows, a query requesting the weekly sales totals for a particular product line can take a long
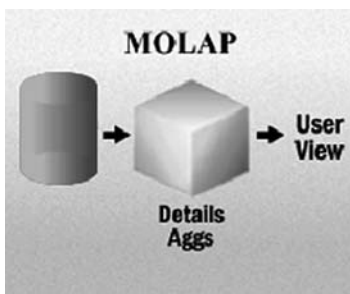
FIGURE 5.1   Aggregated values in a multidimensional structure

time to answer if the fact table has to be scanned to compute the answer. However, the response can be almost immediate if the summarization data to answer this query has been precalculated. Fact tables are central tables containing all the transaction records in a multidimensional model. See Chapter 16 for more information on fact tables. Precalculation of summary data is the foundation for the rapid response times of MOLAP storage mechanism. Most cell values in an OLAP cube do not exist in the source data. The source data typically has information only at the detail level. The OLAP structure allows the aggregation (precalculation) of the values along each of the dimensions in the OLAP cube. This aggregation results in a phenomenon called *data explosion*. It is called data explosion because the OLAP cube precalculates every possible aggregation value when the cube is built and stores that value in the cube structure. As a simple example, starting with a relational source data set with a size of 30MB, a resulting cube would become 120MB. This is an oversimplified example because exploring the factors that contribute to data explosion—the biggest of which is the dimensional definitions associated with an OLAP cube—is another topic altogether. Another important point to note in the data explosion is that the growth from the relational data store to the OLAP cube is not linear. In our simple example, the size goes from 30MB to 120MB, which is a growth of 300%. It is not safe to apply that same percentage growth to a relational data store of 100MB and state that the resulting cube would be 400MB. The data explosion is typically exponential in growth and is influenced by the dimensionality of the cube.

The benefit of a MOLAP storage mechanism is query speed. The retrieval of information from a MOLAP cube is fast, especially when compared to a ROLAP storage mechanism that is described later. The penalty for this speed is in two parts. First, the additional disk space required to handle the data explosion. Second, the processing time to construct the cube can be large.

A recent addition to the OLAP world has changed the data explosion problem that MOLAP generates. Microsoft Analysis Services is from Microsoft and is part of their SQL Server tool set (not sold separately). The Microsoft solution is worth mentioning here because they have introduced a concept in which the OLAP administrator can make a trade-off between query performance and size of the data explosion, all within a MOLAP storage mechanism. Microsoft Analysis Services incorporates a sophisticated algorithm to select aggregations for precalculation so that other aggregations can be quickly computed from precalculated values. It has a Storage Design Wizard that provides options for the administrator to specify storage and percentage constraints to the algorithm to achieve a satisfactory trade-off between query response time and storage requirements. Most often, the MOLAP implementation with Microsoft Analysis Services results in a cube(s) smaller than the original relational data store and have very fast query performance. This is in contrast to other tools using MOLAP in which data explosion is the norm.

The "R" in ROLAP refers to "relational." Relational OLAP (ROLAP) stores aggregations in a relational structure and leaves the partition's source data in its existing relational structure. (See Figure 5.2.)

This architecture places a heavy burden on the relational database system (RDBS) where the source tables for the OLAP cubes is stored. Every query issued against the OLAP information first passes through the OLAP definition, then is processed by the underlying RDBS prior to returning a result. This seems like a laborious process with serious performance considerations and raises the question of why ROLAP would be considered for the storage mechanism. The primary driver to use ROLAP is data storage limitations. Our discussion on MOLAP outlined the downside of data explosion as all the possible data aggregations are stored in the MOLAP structure (with the exception of Microsoft's Analysis Services). With ROLAP, data explosion is not an issue because the information is not preaggregated as in the MOLAP storage. This is especially important when the original data set fed into the OLAP mechanism is large. The MOLAP data explosion would exceed most data storage mechanisms if starting with gigabytes of data from the relational store.

The benefit of ROLAP is twofold. First, there is very little additional disk storage needed to accommodate a ROLAP implementation. Unlike MOLAP, the ROLAP storage mechanism doesn't cause data explosion. Second, the processing time is very short when compared to MOLAP. ROLAP does not aggregate (precalculate) the data available in a cube, so there is very little processing that occurs when creating the cube.
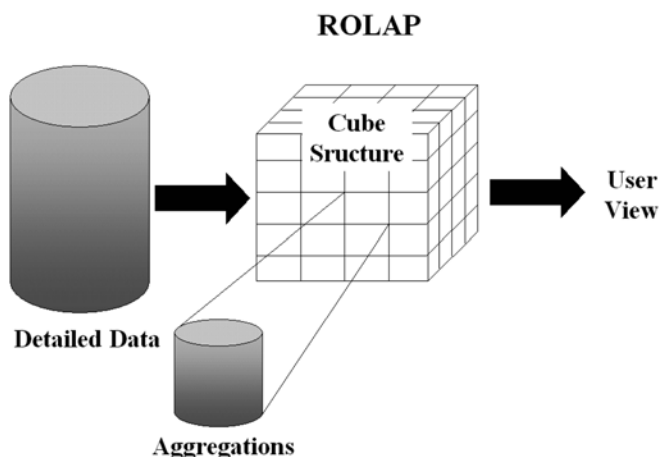
FIGURE 5.2   ROLAP
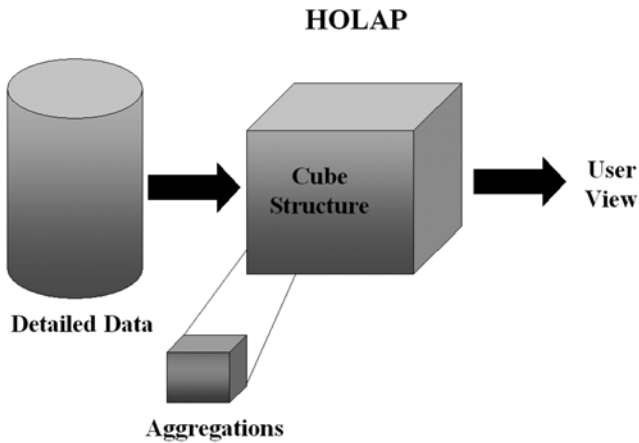
HOLAP

FIGURE 5.3   HOLAP

Each OLAP server vendor has its own storage aggregation. Some only support one method or the other (MOLAP or ROLAP). Other vendors give you a choice, and a new addition to this arena is Microsoft's Analysis Services. Microsoft has introduced an additional storage mechanism they have named HOLAP. The "H" is for "hybrid." Hybrid OLAP (HOLAP) stores aggregations in a multidimensional structure on the OLAP server and leaves the partition's source data in its existing relational structure. (See Figure 5.3.)

Microsoft can implement this "hybrid" because of their implementation of setting aggregation performance discussed in the MOLAP section above. In the hybrid model, the aggregated (or precalculated) values are stored in the MOLAP section and queries that can be answered by the aggregated values have fast response times. When a query retrieves information below the aggregated values and accesses the ROLAP section, response time slows down as the relational system retrieves the data prior to passing it back through the OLAP system. On the surface, the HOLAP implementation from Microsoft appears to solve the issues of data explosion and large relational data sets, but since Microsoft is a relative newcomer to the OLAP field this should warrant further investigation prior to accepting Microsoft as the vendor of choice.

6

BUILDING THE DATAWAREHOUSE/MART

This chapter is titled "Building the Datawarehouse/Mart" because the issues outlined can apply to both the datawarehouse and the data mart. Our discussion will lean more toward the data mart because a datawarehouse is typically a longer-term and larger-scoped project than discussed in this book. There are books and institutions (Data Warehousing Institute, *http://www.dw-institute .com*) built solely on the concepts of datawarehouses, and this book will not eclipse the breadth of information available from those sources in this chapter. This chapter will focus on some key issues for delivering valuable information when building the datawarehouse/mart.

The primary difference between a datawarehouse and a data mart is the scope of the problem it is intended to solve. A *datawarehouse* can be defined as a collection of data designed to support decision making on an enterprise scale, whereas a *data mart* can be defined as a database, or collection of databases, designed to help make tactical decisions about a business.

IMPORTANT SUCCESS FACTORS

Don't Wait; It Won't Be Right the First Time

The most common (and costly) mistake datawarehouse/mart projects incur is trying to deliver a "finished" product or "final" set of information the first time. Objectives are defined, a project plan is laid out, staff is hired, and work

progresses. No one will take issue with this process because it is how nearly all technology projects happen and is a known recipe for success. The disconnect occurs in two areas:

1. The assumption that there is an endpoint
2. Information needs are static

There should be no endpoint, or finished deliverable, in a dataware-house/mart project. Yes, there are concrete objectives, but to plan it as a typi-cal software deliverable that is "finished" at a point in time is not true. This is a project that should live as long as the business needs information. It will not follow the traditional software cycle of birth, life, and death. The dataware-house/mart will live and continue evolving over time. It will use new tech-nologies as they become available and mature, and change over time to meet the business needs.

The information needs in a business are not static. These needs continu-ally change. This financial quarter's key issues will not be the same issues for the same quarter next year (if the issues are the same quarter to quarter, that points to bigger problems in a business because issues are not being addressed). The information about a business is under constant change—the products and services we sell, markets we operate in, and organizational structure. All the major facets of our business change over time, and some are rapid changes.

Delivering some information from the datawarehouse/mart into users' hands quickly is the key to avoiding this shortcoming. Quickly should be within two to three months (one financial quarter). Most enterprise-wide datawarehouse projects are originally scoped in range from 18 months to three years. This is too long to wait prior to getting feedback from users on the in-formation available from the datawarehouse/mart. The initial information ex-posed to users should not be some complete set to replace existing systems. The information does need to provide value and can start as augmenting existing re-porting, allowing it to start small and build over time, eventually phasing out the system(s) it is designed to replace. Think of the datawarehouse/mart as an *evolution*, not a *revolution*.

Consumers Will Help You Understand Information Value and Quality

Information quality is not data quality. One of the primary purposes of a datawarehouse/mart is to ensure that the data is accurate and credible. This is required, but an area where a datawarehouse/mart typically falls short is in in-formation value. Data is not information. The correct formula for information is shown in Figure 6.1.

Data
+ Relevance
+ Context
───────────
= Information

FIGURE 6.1 Definition of information

Getting information from the datawarehouse/mart into consumers' hands quickly is a key part in providing information value. Too often, a datawarehouse/mart project will wait until its longer-term scheduled delivery date before getting information to the consumers. This is too long to wait for feedback from the information consumers. Phasing the rollout of information from a datawarehouse/mart to consumers and establishing an effective feedback cycle can ensure the success of a datawarehouse/mart project. The effective feedback cycle is critical because only the consumers of the information can tell you if your datawarehouse/mart is successful. The task of cleaning, staging, and presenting information can continue for many days, but in the end if the information provides no value to the consumer, then the datawarehouse/mart has failed. Placing information into the consumers' hands early and maintaining a feedback loop will be a key to success.

Plan for Iterations

Don't define a single endpoint and expect the datawarehouse/mart to be finished at some point in time. Plan to iterate on the project often. The three key parts to drive the iteration are:

1. Information-consumer feedback
2. Change in business information needs
3. Business structure changes

Information-consumer feedback is driven by the topic directly above this section discussing how information consumers help define the quality and usefulness of information. Prioritize and incorporate the changes defined by the consumers to better meet their informational needs. As a business evolves, so will its informational needs. If a datawarehouse/mart does not evolve to meet those needs, it will be used less and less until falling into obsolescence. Structure to a business will change. New products or services will be sold, the internal organization will change, markets will shift. All these factors, and others, will contribute the always changing models used in delivering information from the datawarehouse/mart. (See Figure 6.2.)
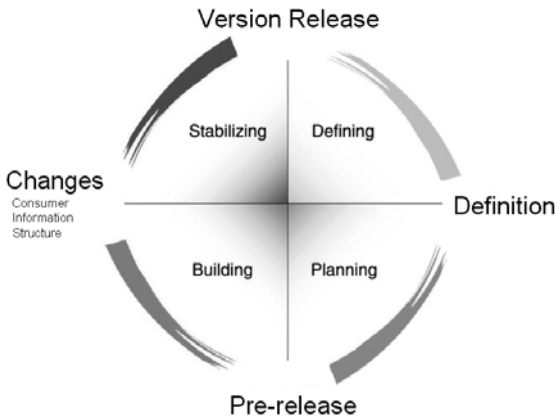
FIGURE 6.2 Always changing models

USING FINANCIAL DATA SOURCES

Financial data sources revolve around the general ledger (GL), accounts receivable (AR), accounts payable (AP), sales order entry, and so on—all the modules normally found within an enterprise resource planning (ERP) system. All data recorded in the ERP system typically falls into different GL hierarchies so it can roll up into the standard financial reporting formats for a balance sheet, profit and loss, and the like (refer to Chapter 18 for more detail on financial data modeling relative to the different ERP components). The data capture in the ERP system and GL focus can present some benefits and limitations. Three of the benefits that come from the ERP systems and GL focus of a financial data source are:

1. The data is already captured in the day-to-day business.
2. Information is categorized by the business process into a GL account (or accounts).
3. The focus for the data is to make it "reportable" for financial statements in which the processes and rules are clearly defined.

The capture of data to place into an OLAP reporting or analysis structure cannot be trivialized. If the data is not currently getting captured, the cost to go and capture that data can be high. That's where a real cost savings occurs—by using information an organization is already capturing. The process to record and capture the data is in place. We are simply going to redirect its usage to an OLAP structure for use by the BI analysis tools.

A process already exists within the business to categorize the information into its appropriate GL account(s). The business produces profit and loss, balance sheets, and the like. This process structure forces information into the GL account structure, saving up-front time in trying to define how the information should be categorized. This is also a real cost savings.

The OLAP structure can be used to provide more flexibility to the already "reportable" nature of the GL information. Classifying the data into the GL account structure provides a definition of an analysis starting point. Using the OLAP structure, users can "drill" into the underlying information that drives the values rolled up into the GL structure.

The focus of a GL structure in financial data can also have some limitations when used in an OLAP structure. Three of those limitations are:

1. Preaggregation of the data
2. Inability to view, or drill through, to the detail records
3. Analysis by nonfinance consumers

Preaggregation of the data means that there has been some summary of the information prior to its inclusion in the OLAP structure. A table or set of tables in the relational data store has had a summary, or resolution, of values processed, and the resulting value ends up in the OLAP structure instead of the base records that make up the processed value. It is always better to include the lowest level of detail to the OLAP structure and not a preaggregated value because it allows an analyst to "drill" down to the level of detail he or she needs to find information. If values are preaggregated, the exposure to the detail level is not available to the analyst.

The topic of preaggregation leads us to the issue of drill-through. Drill-through allows an information consumer to view the source data for a single cell value in an OLAP cube and get the detailed information provided by the transaction records that make up that cell value. If the values provided to the OLAP structure are preaggregated, the power exposed by drill-through is lost, and information consumers won't be able to go to that level of detail.

The GL account structure is a great way to view information by finance-oriented consumers and executives that understand the importance of financial reporting. This leaves out a large audience that does not understand, or does not need to learn, GL account structures. If the information presented in an OLAP structure is driven by a GL account structure, it is likely that nonfinancial personnel will not use the information, or at least it may be difficult to understand. This is an issue that cube designers should understand, and perhaps offer other cubes (from the same source data) that don't have the GL account structure. As a simple example, a typical cube would have the GL account structure as one dimension. Other common dimensions would be time, customers, and

organization (internal to the business). Finally, a single metric for "monetary amount" would be in the cube. The value of monetary amount would be driven by the GL account structure, and as the account structure rolls up, the OLAP portion does the math to add/subtract values based on the classification of the GL accounts. This organization of the information makes perfect sense to someone accustomed to examining financial statements. However, it can be confusing to the nonfinance information consumer. The nonfinance-oriented consumer would ask questions like, "What were revenue and gross margin for products X, Y, and Z over the past two quarters compared to the same two quarters in the previous year within sales region 1?" This question can be answered using the GL account structure, but it would be easier for the nonfinance information consumer to have two measures to choose from (probably more)—revenue and gross margin—then to use the other dimensions to answer the rest of his question. This is a cube design issue and is typically addressed by providing multiple cubes in an OLAP system, with each cube designed for a specific set of business users. The benefit for the business is that the information included in these cubes comes from the same data source providing consistent information across the organization.

Flexibility in viewing different GL account structures is a primary benefit for financial data in an OLAP structure. This is the ability to use different GL account structures to look at the same information. One of the most powerful features of an OLAP structure allows one to apply different dimensionality to the same set of information. Different dimensionality is applied by changing the GL account structure dimension, and not the fact data. The GL account entry that a data value belongs to does not change, but the arrangement in which information is viewed and how those values "roll up" into the GL account structure changes.

STAGING DATA FOR FINANCIAL ANALYSIS

Staging data is the term applied in the ETL process for making the data ready for consumption. This is a task covered in great detail by other books about datawarehousing. This chapter does not cover the ETL details for staging data, but focuses on some key items important to financial data, or sometimes not considered by standard staging discussions. The few items for focus here are intermediate data updates, updates when closing a financial period, and considering two sets of information.

Earlier in this chapter it was discussed how often information consumers' needs do not synchronize with the close of a financial period. A recommendation to help overcome this information gap is to update the OLAP structure with intermediate sets of data. If the books are closed monthly, then a good in-

termediate update might be weekly. This intermediate update periodicity should be driven by the business analysis and reporting needs.

The intermediate data update also raises questions about the ETL process. Should the same ETL process be used for the intermediate updates as for the updates after the books have closed? Most organizations use two separate ETL processes and even two sets of data: one ETL process and data set for the intermediate/unaudited data, and another ETL process for the closed-book/audited ETL process. Most of the time, the final result is the same OLAP cube(s), but there are cases in which it makes sense for a business to use different OLAP cube(s) that result from both ETL processes. The use of the same or separate cube(s) is driven by the business analysis and reporting needs and is an internal decision within the organization.

Data updates in between closing financial periods will present us with two sets of information:

1. Information *in flux*
2. Information for *closed periods*

The special challenge with information *in flux* is that the data won't be audited. This should be a key concern for most organizations. At a minimum, a high-level glance at the numbers should be done to ensure that there are not tremendous errors in sales, costs, or other key performance indicators. There should also be information-consumer education that the *in flux* updates are not audited figures and to exercise judicious use of the information. Information for *closed periods* is the same information used for public reporting of financial data and other external use. It is considered reliable, consistent, and accurate.

7

FRONT-END
ANALYTIC TOOLS

Part One discussed the reasons why today's highly competitive organizations need front-end analytic tools. It also laid out the primary features that should be sought after when looking for those tools. This chapter will look at the front-end tools available today and their vendors, plus a brief list of their strengths and weaknesses. Keep in mind that the software market changes rapidly, especially in the business intelligence (BI) space as a shake-out continues and inadequate tools fall by the wayside and others rise to the top.

SPECIALIZED BI TOOLS

We call these "specialized" BI tools because they are available only with OLAP as the data source. These are not general reporting tools, which are discussed in the "New Breed of Business Intelligence Tools" section of Chapter 2. We call these tools because they are one more piece available in our tool kits to solve today's business problems. Specialized BI tools are not an end in themselves, but only one more means to gather information. A key part of these tools is their delivery mechanism. There are two primary delivery mechanisms offered by BI tool vendors today:

1. Desktop applications
2. Web delivery

Desktop applications typically offer a rich client interface. This is where a large number of features are available at the desktop for analysis, information exploring, formatting, and sharing of that information. This is an important feature set for analysts and information "publishers." These are the individuals who prepare information for a much wider audience that doesn't have time or doesn't need to perform analysis.

Web delivery means information is available in web browser from a web site. This doesn't mean the Internet, and is typically not the Internet, but isolated to a company's internal network or VPN (virtual private network). The information exposed for BI tools and by OLAP technology is usually sensitive in nature and not for general consumption on the World Wide Web. However, there are organizations that broker information to customers via the Internet and others that make information available to internal people via secure web sites. Web delivery is a key mechanism for these organizations. The downside to web delivery is that the interface and user experience is not as rich as a desktop application. In fact, some web-base delivery is simply serving up static formatted pages over the web while others have a high level of interactivity. As tools and technology progress, web delivery will become more the norm, with most of the functionality available only in desktop application shifting to web-delivered information.

Desktop applications and web delivery are not the same primary delivery mechanisms offered five years ago, and probably won't be the same primary delivery mechanisms offered five years from now. The BI tool space evolves with and exploits the rapid change of technology. See Chapter 13 for vendor lists and references on where to find more information.

Features and Functionality: A Technical Perspective

The key features of BI tools mentioned in Chapter 3 should be examined closely as to what that *feature* means in detail and from the information consumer's perspective. Business intelligence software is getting increasingly more powerful, and many of the popular solutions share common functionality. The following list details what some of the key features really mean:

- *Drill-down.* This is an action that allows a user to navigate lower within a dimensional hierarchy. Hierarchies are the structural arrangement of a dimension. As an example, time can have a hierarchical structure of year, quarter, month, and day. If I am looking at a data point in the year 1998, and if I drill down, then I should see the quarters of 1998. Drill-down enables us to see what information is driving a single aggregated value.

- *Standard chart/graphs.* Support of standard business charts is a key part to information delivery in BI. The statement, "A picture is worth a thou-
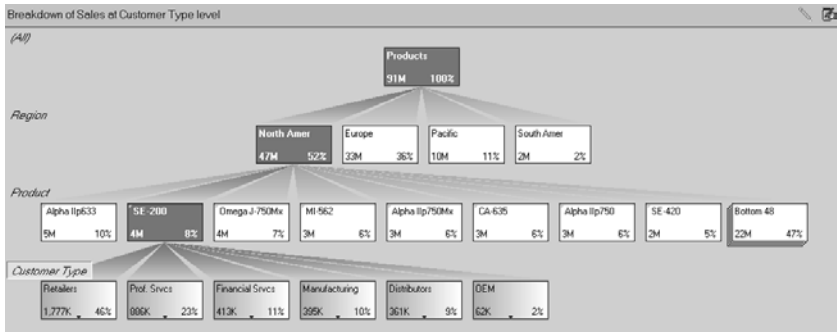
FIGURE 7.1   Decomposition tree

sand words," holds even more weight with OLAP information. A tool
that supports standard business charts (e.g., bar, horizontal bar, pie, area,
line, point, etc.) is more valuable than a tool that simply puts numbers
into a grid forcing the information consumer to sort through a table of
numbers to get information.

- *Advanced visualization.* This is a feature that is defined by the vendor as
  advanced. It typically refers to a part of the tool that has a patent by the
  vendor and that they tout as a better way to get information. An example
  would be the Decomposition Tree from ProClarity Corporation in which
  the information consumer is allowed to drill down into any dimension
  from a given data point in a tree structure. (See Figure 7.1.)

- *Exception highlighting.* This is also sometimes called *alerts*, wherein in-
  formation consumers are allowed to specify their own parameters for
  highlighting values and how those values are to be displayed. This is use-
  ful when looking at a large table of numbers and values in a certain range
  should stand out. An example would be to color all the revenue numbers
  less than $100,000 as red because dropping below $100,000 is an un-
  healthy sign. This is the same concept as conditional formatting, avail-
  able in Excel.

- *Combo views.* This refers to the ability to make a combination chart in
  which a line and bar chart can be on the same graph and be able to show
  two different Y-axes that have differently scaled values. One of the most
  common examples is a chart that shows an absolute value for revenue
  and a percentage value for revenue growth. The values for revenue and
  percentage growth won't scale if shown on the same axis because the
  revenue numbers will dwarf the percentage values. The solution is to
  place the revenue numbers on one side of the chart (y-axis one), and the
  percentage growth number on the other side of the chart (y-axis two).
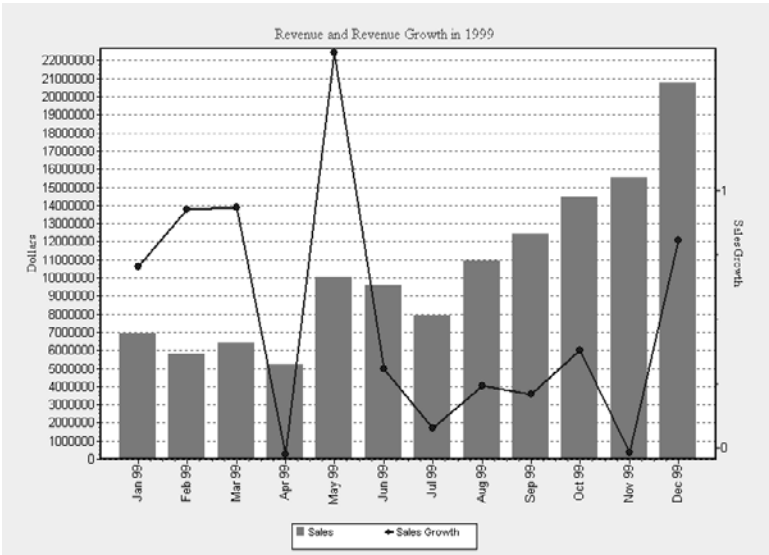
FIGURE 7.2 Chart example

Then apply a different chart type to each value such as a line chart for growth and a bar chart for revenue. (See Figure 7.2.)

- *Pivot rows and columns.* This is a fundamental feature for most BI tools. It allows the values that are shown on the rows and columns to be interchanged. The values on the rows are moved to the columns, and the values on the columns are moved to the rows.

- *Drag and drop dimensions.* This is a key part to simplifying the interaction of an information consumer with the BI tool. The action expected by today's information consumer is to click on an item and move it to a desired location. It should be this easy to pull desired dimensions into position on rows, columns, and so on, and to arrange their order—all by just clicking and dragging the mouse. Forcing the information consumer to open a different dialog and specify dimensional positioning using other techniques is an unnecessary burden.

- *User-defined custom calculations.* This means placing the ability to create metrics that are important to information consumers into their hands. Anyone who has used Excel and created formulas understands the power of this feature. The individual creates metrics driven from other values that give them more information. This is a powerful feature for a BI tool to place this ability into the information consumers' hands. Otherwise, creating metrics not normally found in the information set will be an ad-

ministrative task performed by those responsible for maintaining the OLAP structure.

- *Expose queries.* This refers to the ability to "hand-modify" the language or script that defines a query in a BI tool. Some tools expose this and the usage of it would be by power users—individuals who need to "tweak" an existing query to get exactly what they want and what they can't get from the user interface of that tool. Examples of this would be the scripts defined by Oracle Sales Analyzer, or the MDX exposed by the ProClarity Analytic Platform, from ProClarity Corporation.

- *Qualitative comments.* Augmenting a view of information with qualitative comments can offer the extra credit needed for real information value. Think of this as postanalysis reporting. Looking at numbers is a good start, but if an analyst can add words that relate to the how and why for those numbers, the information value increases. The qualitative comments should be entered by users for a view of information that they can construct, and the comments are saved as part of that view. The objective is to share the comments with others interested in the same information.

- *Dashboards.* This is a term used to define a few key metrics that point to the health of a business, and placed in a location that is commonly viewed by the information consumer monitoring those metrics. An example of a commonly viewed location would be Microsoft Outlook, wherein individuals can see their inbox for e-mail, calendar, and task list. This would be an ideal location for the key metrics in a dashboard. Other examples usually involve a web site internal to an organization that acts as an information "portal." Placing a dashboard in this location makes it accessible from any Internet browser within the organization. The dashboard implementation can range from something as simple as a grid with a few numbers for the key metrics to a graphic speedometer-type dial that changes rapidly as underlying information changes. The choice of implementation depends on the information consumers' needs and what can be produced by the BI tool.

- *Distribution of cubes/reports.* This is critical to sharing information. The two items of cubes and reports are tied together because with this combination the information consumer has the ability to do more than simply see a static report. Most of us are accustomed to seeing a static snapshot of information in a report. The report is separated from the data set. To get more information, we need to derive another report and see its results. This cycle continues until we get to the information we need. Contrast this to a situation in which a report is tied to the underlying cube that derived the information. To get more detail the information consumer simply needs to "drill down" into the cube information—all from within the

same report. The benefits from this solution are saved time, information consumer–defined context and relevance, and information is derived faster.

- *Other popular analytical features such as sorting and filtering.* Data filtering is an easy way to limit the records you are viewing at any one time. It is primarily used to eliminate the bottom, top, or a specified range of values. This makes it easy to remove those 50 clients that account for only a total of 2% of your business, and focus on the 20 clients that account for 98% of your business. By filtering the data, you can work with a smaller subset of the members and reduce the time required to return a response from a query. Sorting is often useful to see the items sorted by the measure being viewed. If you are viewing sales, you might want to see the items sorted from highest to lowest sales.

REAL TIME ANALYSIS: MYTH OR REALITY (TECHNOLOGY PERSPECTIVE)

"Real"-time analysis refers to a ROLAP (relational OLAP, see Chapter 6 for more details) implementation in which the cube data comes directly from a transactional system. The data fed into the ROLAP definition is the same data set that is the OLTP (online transaction processing) system. This is a direct contradiction to the traditional OLAP implementation wherein data is cleaned and staged prior to inclusion into the cube.

The reported benefit of using a real-time OLAP implementation is that the information is as up to date as the information in the OLTP system. There is no time lag between when data is entered into the OLTP system and when information is available from the OLAP cube(s). With the obvious benefit of no time lag for information availability, why isn't every OLAP implementation a real-time implementation? There are two key reasons why real-time OLAP is typically not implemented:

1. *Information accuracy.* The credibility of information in an OLAP cube is only as reliable as the cleaning and staging process that typically occurs in the extraction, transformation, and loading (ETL) process. There's a good reason those phases are called *cleaning* and *staging*. A percentage of the data entered or maintained in an OLTP system is inaccurate, not correctly related to other tables, or simply incorrect. This is why a typical OLAP implementation will draw its data from a datawarehouse, or data mart, where the data has been through a process of cleaning and staging providing credibility. When a real-time OLAP implementation draws directly from the OLTP tables, there is no ETL process to clean

and stage the data. No audit or accuracy check has been performed, and the information accuracy should be questioned.

2. *Performance.* A key part of OLAP implementations is to provide fast access to information. This is typically accomplished by providing aggregations of the information. Aggregations are precalculated summaries of data that improve query response time by having the answers ready before the questions are asked (see Chapter 6 for more details). With a real-time OLAP implementation, aggregations are eliminated. The cube information is drawn directly from the tables that make up the OLTP system. Aggregation of the data in those tables is simply not done. OLAP query performance will suffer. The degree of performance suffering is affected by several factors:

   - *Size of the OLTP tables.* This issue almost speaks for itself. If the tables are large, and the OLAP implementation asks for data summarized over time, say by financial quarter, then we can expect the summarization of information over those quarters to consume some time. This lag in returning results is a direct effect of not having aggregations in the OLAP implementation.

   - *Number of concurrent users.* This is not only the number of concurrent OLAP users, but also of the users entering, modifying, and managing the OLTP data. This includes individuals running reports and those adding, modifying, or deleting existing records to keep the information current.

   - *Scalability of the system managing the OLTP tables.* The system managing the OLTP tables needs to be scalable enough to handle this level of concurrent users. Another scalability impact is that the OLAP users are not simply asking for a record, changing its content, and replacing it in the system (a typical OLTP action), but are issuing queries against the OLTP system that will gather large amounts of data, summarize it "on the fly," and present the summarized results back to the OLAP system. This action by a number of concurrent users will tax the scalability of a system.

Real-time OLAP implementations are not typical, primarily for the reasons of information accuracy and performance listed above. These restrictions are typically prohibitive enough that they negate the original benefits for which the OLAP system was touted. The reality is that real-time OLAP systems almost never get implemented. However, in the future, real-time OLAP systems may become common as tools and technologies overcome the challenges outlined above.

EXCEL ADD-INS

The first part of this chapter outlined some specific tools for BI and their key features. In contrast, there is another set of tools to retrieve information from OLAP data sources and place them directly into an Excel spreadsheet. These are Excel add-ins, which are specifically designed to work in Excel and do not function as stand-alone solutions, or integrated into other environments such as a web site or custom application.

Moving information directly into Excel from an OLAP data source is a powerful asset, especially for information consumers that already perform analysis in Excel. The real benefit of an Excel add-in that allows access to OLAP data is the creation of a link between Excel and the OLAP data. This is more than simply extracting data from the OLAP source and placing a static set of values into the spreadsheet. A "link" between Excel and the OLAP data exists so that as the OLAP data is updated so is the spreadsheet. The update occurs at the information consumers' discretion, but they only need to define the link once and refresh the data when it is important to them, such as when the OLAP data source is updated.

Another key benefit of bringing OLAP data into Excel is to take advantage of the formatting capabilities. Many organizations use Excel as a reporting tool because they can format the information to meet their exact reporting needs. Some Excel workbooks have large investments of time to create complex templates that format and present information. Bringing the OLAP data directly into Excel helps protect this investment.

TRADITIONAL REPORT WRITER VERSUS
OLAP-BASED ANALYSIS TOOLS

There is a lot of discussion in today's BI world that reports will simply "go away" once we put online information into everyone's hands. That thought vector is incorrect. Reports continue to play a valuable role in today's business even as we move to online analysis. The standard business reports of a balance sheet and a profit and loss report provide many analysts with the brief view they need to understand the financial health of a business. To expect that these reports would no longer be used is naïve.

However, there is another group of reports that are continually refined and customized to provide the information decision makers need to make more well-informed decisions. The objective of these reports is to provide finer granularity of information to identify trends, show relationships, or cause-and-effect instances in information. This group of reports is destined for obsolescence as BI tools become more commonplace and place this "exploration" of information into the decision maker's hands. The fundamental issue is that re-

ports are static. Reports are a snapshot of information at a point in time. A report typically generates more questions than answers, yet it does not provide the means to answer those questions. A decision maker then needs to formulate another request for a report (either to himself or someone else), generate, and examine its results.

Online analysis overcomes this time-consuming cycle by placing the decision maker online with the information. As questions are asked, decision makers can explore the facts driving summarized information. They can identify trends, explore relationships, and see cause-and-effect instances in the information. Reports provide the necessary functionality of summarizing details and snapshots of information. Online analysis puts decision makers in touch with information that they can explore and format for their use, eliminating the time-consuming cycle of requesting reports.

Report-writing tools are focused on highly formatted output. Report definers can specify items such as the titles, column layouts, column widths, subtotal breaks, fonts, font sizes, and a long list of layout/content features. The key functionality points for a good report-writing tool are:

- *Flexibility in report layout.* Flexibility translates to customizable. Can the format of the report be customized to the extent required to meet the business need? Most report-writing tools are highly customizable for formatting output. This functionality is the "ticket in the door"; if a report-writing tool is not highly customizable then a product evaluator won't even look at it.

- *Customizable to meet the business needs.* This is more than flexibility in the report layout. A profit and loss statement for business A will be different than for business B. A good report-writing tool will enable both businesses to customize that profit and loss statement to reflect their business need. A profit and loss statement is just one example. The ability to customize should cover a wide range of reports in the report-writing tool.

- *Automatable.* The ultimate automation feature for a report-writing tool is twofold: (1) generate the report and (2) distribute it. The automated distribution is where some report-writing tools fall short. At that point, some custom development or other tool needs to be used for distribution. A tool that can perform automatic generation and distribution that is easily maintained can be very valuable.

8

SECURITY

The information in online analytical processing (OLAP) structures is typically very sensitive. The information sensitivity can range from highly confidential internal data to information that has a high level of intellectual capital investment. An example of highly confidential internal data is the historical general ledger information for a privately held company. On the other end of the spectrum is information that has a high level of intellectual capital investment such as grocery store retail sales data that is collected, massaged, and "brokered" to grocery product manufacturers.

Security is an important issue that each OLAP implementation must address that concerns the level of sensitivity for that information. There are two areas that should be focused on for security: role-based access and Internet security.

ROLE-BASED ACCESS

Role-based access means associating a user ID with a role. The role then has certain restrictions for information visibility. As an example, we can define a role named "Midwest Managers." This role accesses the nationwide profit and loss OLAP cube. We set up "Midwest Managers" to see details of the Midwest geographic region, but only summary information for the other geographic regions. We would then have corresponding roles for the other geographic regions with the same type of restriction specific to their regions. After the roles are established we would add user IDs to each role that corresponds to their geographic responsibility.

Role-based access is not restricted to a geographic limitation. A role could be defined that narrows the visibility to very specific areas across all the dimensions within an OLAP cube. A typical plan for role-based security has roles defined for different levels of responsibility within an organization. The individuals who fall into those areas of responsibility then become members of those roles. As their responsibilities change, those individuals change roles or become members of other roles.

INTERNET SECURITY: ONLY AS GOOD AS YOUR BARRIERS

Making BI information available across the Internet can be an incredible advantage for a business. This can range from a secure connection available to employees only to information that is "brokered" across the Internet to specific customers to information you choose to make publicly available. The advantage of using the Internet is that it offloads all the infrastructure responsibilities from your information technology (IT) organization. This can be a significant savings. The downside of using the Internet is the possibility of theft or vandalism. Once you make information available on the Internet, it is exposed to hackers or those bent on stealing your intellectual capital. The sophistication of security mechanisms for Internet access grows and changes every day. A lengthy discussion on this topic is beyond the scope of this chapter and this book, but Internet security for your information is not a trivial subject, and there are plenty of resources (tools and people) that specialize in this arena available on the public market. Do your homework and make an investment in security if you plan to provide Internet access to BI information. Then make sure it is kept up to date. The time and energy spent making sure your Internet access to information is secure is well worth the investment.

9

THE INTERNET'S IMPACT
ON BUSINESS INTELLIGENCE

The Internet has forced two issues to the forefront in information delivery:

1. The possibility of obtaining up-to-date information from nearly anywhere in the world, when you want it, and at a low cost, is a realistic possibility.

2. Vendors of business intelligence (BI) solutions and BI tools are pressured to provide real world information delivery mechanisms for the Internet.

Prior to widespread use of the Internet, BI solution vendors could pick and choose the delivery mechanism that suited their particular strengths. Now, with the Internet as a primary information delivery mechanism, solution vendors must offer Internet delivery as part of their solution. If they do not have an Internet solution, they usually don't even get their foot in the door.

EVERYONE'S A PLAYER

Internet-based delivery for BI information appears to be a "must have" in today's world. All the tool and solution vendors offer products in this area. If they don't, their competition soon passes them by. The question an organization must ask itself is, "Do we need Internet access to BI information?" Often the knee-jerk reaction is to say, "Yes, of course, we do. Everyone is making their information available on the Internet." However, as our mothers used to say when we were young, "Well, if everyone else jumped off a cliff, would you

too?" This isn't saying that making information available on the Internet is like jumping off a cliff. The point is that just because everyone else is doing it doesn't mean it's the right decision for your business.

Some of the factors that need to be considered prior to making BI information available on the Internet are:

- *Defining the business benefit.* Answer the question, "Where will this impact my business, and what is the return on investment?" Internet information access to internal people usually gives them faster and easier information availability. Speed of information typically gives a competitive edge in the market. This speed could give field salespeople more information when they need it to offer better deals than the competition.

- *Addressing security.* First, ensure that the risk of Internet information access is acceptable. Second, put in place good security measures to minimize the risk.

- *Examining the additional costs.* Detail the additional infrastructure, time, and personnel costs this may incur. Plan for these costs and ensure that you can get a return on the investment.

WHAT IS A PORTAL?

Portal is a common term used to describe a web site that points visitors to information. This is where the commonality of the term *portal* stops. What goes into the portal and what the portal looks like or behaves like can vary widely from one web site to another. A typical example of a portal for public use is *Yahoo.com.* On *Yahoo.com*, we can find references to current event topics, media happenings, home shopping, travel, and more. It is a consumer site.

Many large organizations have portals for internal use. The intended audience for these portals is employees, and the portals are typically available only on a company's intranet. Although some portals intended for internal company use are also available on the Internet through a secure logon to an Internet web site, the purpose of an internal portal is for organizational information dissemination. It is a common point where current information can be found about the organization. This information can range from the company calendar with a reference to the upcoming company picnic to a link for the latest financial statements about the organization.

Another, somewhat newer, aspect for portals are enterprise information portals (EIP) and business intelligence portals (BIP). There are very few commercial EIP applications available, so companies build their own EIPs with consultants or internal staff. Some vendors now offer tools to create EIPs. See Chapter 13 for more details on the vendors and links for more information.

EIPs are meant to interface with any generic set of information, but there is a new class of emerging portals called BIPs. These are specific to business intelligence, and geared toward serving up information particular to business intelligence. Again, look to Chapter 13 for more details on vendors.

HOW DO I DELIVER BI INFORMATION FOR THE INTERNET?

Using a portal as the starting point for Internet BI information delivery for use within an organization is probably a smart choice, especially since most medium to large organizations already have a portal established for internal use. The definition of *internal use* here means that the portal is accessible only by members of the organization, and that it has some level of security to screen out nonmembers of the organization. The "internal" part doesn't mean that it is available only from within the company's infrastructure. In fact, that is the power of using a portal on the Internet—to make the information available to organization members from wherever they happen to be.

What if my organization doesn't have a portal or existing web site? Then the first step will be in establishing a web site as the access point for BI information. This will entail setting up the web server and all its accompanying infrastructure. This is not a daunting task, but it's also not trivial. There are plenty of experts in the world today who offer the service of setting up and managing web sites. There are typically two types of resources:

1. *Independent service providers* (ISPs). These are providers that have established Internet infrastructure and "host" your web site at their location and on their systems. This is their business—they are the "host," or service provider, for many organizations' web sites. Typically, there is a process, or mechanism, wherein you provide your web site content to the web servers hosted at the ISP. Then visitors to your web site never know that the site is physically hosted by an ISP. Delivering BI information in this arrangement does present two key challenges. The first challenge is that the BI information you are offering is sensitive. Can the ISP provide the level of security required, and can the ISP be trusted not to provide unauthorized access? The second challenge is the physical access to the BI information. The OLAP solution providing the information will need to be accessible to the web site for delivery. Can this access be provided to the ISP, and will it be secure?

2. *Solution providers.* These organizations can contract with your business to set up a web site internally within your organization for use via the Internet. This places the burden of infrastructure and maintenance on your

organization unless you contract with the solution provider to provide ongoing maintenance.

The next step in information delivery over the Internet is choosing the right BI tool or BI solution. Often, the BI solution you are using for desktop applications will have some flavor of a solution for Internet delivery. This is a good starting point to begin evaluating solutions. If their solution meets the business needs, or can be customized to meet those needs, then it should be adequate. If your current BI solutions' Internet flavor doesn't meet the business need, then the process for finding the correct solution will need to be started.

PART THREE

SOFTWARE EVALUATION AND SELECTION

10

SELECTING A BUSINESS INTELLIGENCE SOLUTION

A business intelligence system selection is similar to the purchasing process for other accounting and financial software packages. However, business intelligence (BI) software is typically less complex than an accounting or enterprise resource planning (ERP) solution, so the whole selection process should be shorter and should require fewer people. There is a wide variety of BI tools available (see Chapter 13), some of which will fit your company's needs better than others. Although the process need not be scientific, many companies could definitely make better choices and save time and frustration by taking an organized approach to their software acquisition process. Make sure you have people from both the user side and information technology (IT) involved in the software selection process. BI software is typically dependent on support from the IT department for building and maintaining datawarehouses and OLAP cubes, and end users must be involved to ensure that financial analysis needs and requirements are met. This section discusses the key issues and items you should incorporate into your BI software planning and evaluation.

CREATE A PLAN

The first thing you should do with regard to a decision to buy a new BI package is to sit down and create a plan. (See Figure 10.1.) At the very least, you should:
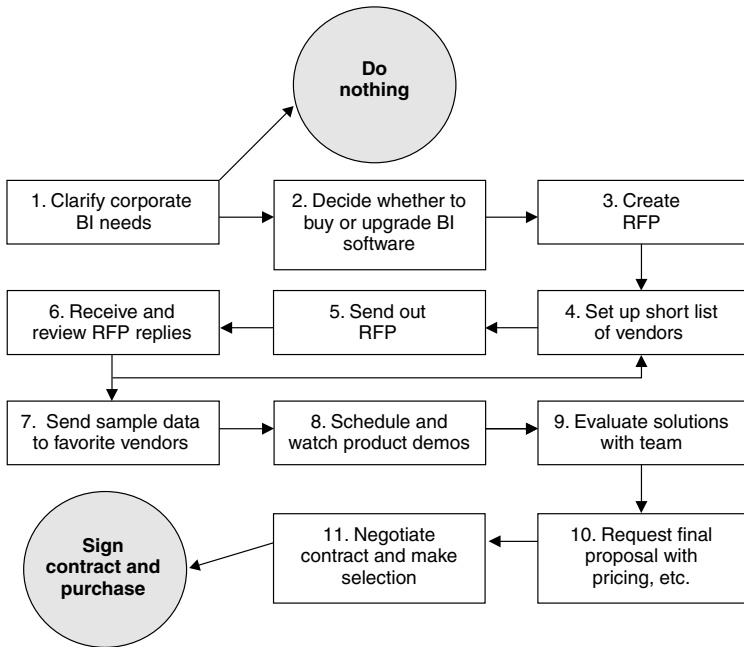
FIGURE 10.1 BI software selection process example

Review the Current Situation

Review your organization's current information needs, and research and document information needs that might surface over the next few years.

Choose the Type of Solution That Best Fits Your Needs

Keep in mind that different types of users have different needs. This means that you might need several types of BI tools (such as a report writer and an OLAP-based query tool) to cover all the user requirements. It is therefore important to categorize your users by their information needs and skill levels.

You typically have three choices with regard to BI software:

1. Upgrade your current BI system (if you already own a system).

2. Build a new system based on your organization's needs.

3. Purchase a commercial BI solution. This chapter focuses on commercial BI solutions, and it assumes that you are moving in this direction.

Create a Request for Proposal

If you decide to look at a commercial BI solution, it is strongly suggested that you approach potential vendors with an RFP (request for proposal) before you view any demonstrations or undertake any other steps in the selection process. Creating this document also serves as a very good organizational learning exercise to activate key people and make them think about what they are looking for in a product and a vendor. The RFP should describe thoroughly the functionality and requirements you need in the following areas:

- Integration to your data sources and productivity tools
- Current and future data volumes
- Areas of usage (e.g., financial reporting, sales and marketing analysis, key performance indicators, components of a digital dashboard, etc.)
- Key features
- User security, access, and roles
- Implementation and training (e.g., involved parties, timelines, roles, etc.)
- Deployment infrastructure (e.g., web, local network, single user, etc.)
- Hardware and software requirements (e.g., platforms, versions, etc.)
- Itemized cost (e.g., software, hardware, implementation, training, etc.)

There is a detailed RFP example in the back of this book (see Appendix A) that you can use as a starting point for your own RFP. *Note:* Don't overstate your requirements. The result could be that you end up with only a few high-end tools in your final selection process, which may have functionality you don't need, be overly complicated, and may cost your company a lot of extra money. The following are suggested steps to follow when you create an RFP:

- Create a selection timetable.
- Create a list of all related parties in your organization.
- Write a short document outlining the scope of the project.
- Start with a sample RFP (see Appendix A) and modify it. Keep it as short and simple as possible. It is better to spend most of your time actually looking at vendor products and discussing your needs with them.
- Distribute your version of the RFP to key people and ask them to make additions and changes. Some people might need only parts of the RFP. For example, information system (IS) staff can focus on platforms, integration, infrastructure, and general technology issues. Management, financial, and accounting staff can focus on financial information and analytical features.
- Review, clarify, clean up, and finish the RFP.

Creating an RFP and later evaluating all the responses can be a time-consuming process. Sometimes, it can be a better and more efficient process to engage directly in in-depth demos and needs analysis with a handful of vendors. Still, it is recommended to create at the minimum a "mini" RFP, in which you describe the key requirements and ask the most important questions. This approach is best if you have experience in software selection, you have good knowledge of the BI software market, and you are good at communicating your company's needs to vendors.

If you want to outsource the RFP and/or other parts of the selection process to a third party, refer to the section Using a Software Selection Company later in this chapter.

Research Vendors, Create a Short List, and Send Out the RFP

A decade ago, before business intelligence became a hot topic for businesses everywhere, only a few BI software vendors existed. Today, however, there are a number of vendors with solutions of different complexity, prices, and technology platforms. Because there are many more choices than ever before, you need to thoroughly research the vendors and their solutions after you receive the replies to your RFP. Depending on the time available, you should try to narrow the field down to three to five vendors whom you invite to make a presentation to your organization.

The following list provides resources you can use to gather information about vendors and their solutions (see Figure 10.2):

- RFP
- Buyer's guides (in accounting/financial, datawarehouse, and other software and industry publications)
- Product brochures
- Current users
- Company background brochures
- White papers
- Third-party evaluations
- Vendor web sites
- Third-party web sites

If you don't know the specific web address you are going to, you can search the web for key words, such as "business intelligence," "OLAP software," or "analytical software." Be aware that you might not be able to get a comprehensive list of vendors with a web search. It is dependent on what web
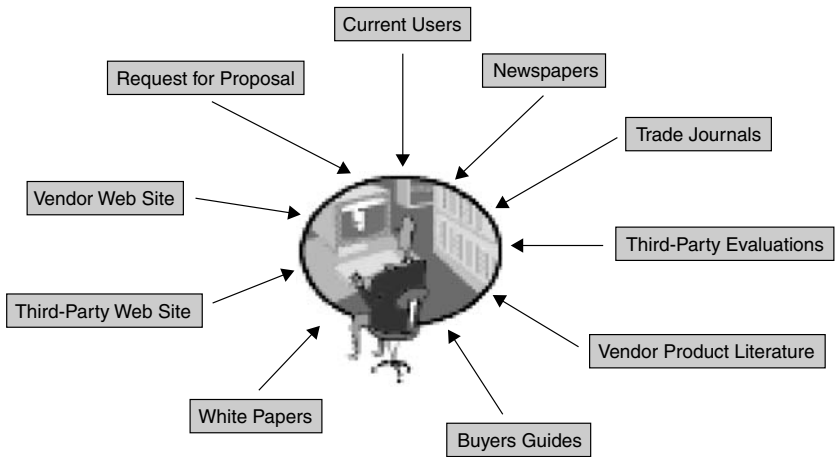
FIGURE 10.2   Sources of information for vendor and product research

search engine you use, your key words, and indexing of vendors' web sites on the particular search engine.

Keep in mind that many buyer's guides, publications, and other available third-party information are not necessarily comprehensive as far as including all vendors. Companies often allow salespeople to educate them about the market. Often, this information will include only the select solution the salesperson would like his or her product compared with. In addition, feature lists can be flavored by:

- Employee bias (used a package before, personal connections)
- Consulting company bias (experts in implementation of a particular solution)
- Vendor bias (such as an ERP vendor promoting its own BI package)
- A journalist's limited understanding of complex business intelligence software

Therefore, your own evaluation of a software (see "Getting the Most Out of a Product Demonstration" later in the chapter) as well as interviews with current customers are the best basis for a software selection.

Distribute RFPs

Some companies send out their RFP to all the vendors they have listed and then let the vendors themselves find out (based on the content of the RFP) if their

software is qualified or not. This might still create a large number of replies with many pages of information you will have to go through in order to find the interesting candidates. Other companies do some vendor research in advance (and check software prices, database platforms, vendor solidity, key features, etc.), and then send their RFPs to a limited number of qualified vendors. Do not forget to put a reply deadline on the RFPs you send out, so your evaluation process does not get delayed because of long vendor response time.

Review RFP Replies

After you have received all the information from the vendors, you and your team need to analyze and rank each reply. Normally, companies pick a list of three to five vendors to come on-site to present their companies, services, and solutions. If you choose too few, you might end up missing potentially good candidates (remember that an RFP is a written document and doesn't show the full picture of a vendor and a solution), and if you invite a large number of vendors, you and your team can easily end up spending weeks in presentations. The latter is not only exhausting and costly, but it also easily creates biases toward vendors based on external factors such as where in the row of presentations they were, how busy the evaluation team was on a particular day, and so on. It is therefore recommended that you choose a handful of vendors. If you are in doubt about certain vendors, give them a call and get the answers you are looking for.

When you study the vendors' replies to your RFP, you need to take it for what it is, a biased answer to a list of questions. To make sure the vendors are as correct as possible with their answers, your RFP and final software contract should state that the vendor is liable for the information they provide should you choose to purchase their solution.

Send Out Demo Material to the Selected Vendors

This step is optional but strongly recommended. In order to get the most out of the upcoming product demonstrations and to provide vendors with a better understanding of your information needs, it can be a good idea to send out to the vendors sample reports, key figures, roll-up structures, and examples of the different dimensions you will use. This will allow them to show you how some of your BI model will look and work in their software package. It is also easier for your evaluation team to ask good questions because they will see a familiar model in the presentation. If you plan to completely revamp the way you analyze and look at your company's information, it might be a good idea to let the vendors show their own examples so you can get an impression of their standard capabilities without too much customization toward your old model.

Meet with Vendors and Review Their Company and Product Demonstrations

Arrange for the presentations to be held and carefully evaluate each vendor and their solution. Try to have all the key people on your evaluation team present at all the demonstrations, so that everyone can get a complete picture before they choose their favorites (in Chapter 11 we will cover the evaluation process in more detail).

A product demonstration is your chance to get a firsthand look at the interface and features of a software package. If a skilled and experienced sales team is doing the presentation, you should be able to get a lot of valuable information from the demonstration. Normally, if you sent out an RFP, you already have information from the vendor about all of the key areas to look at, but now is the time to see as much as possible with your own eyes. Also, if you have been unclear about anything with regard to the vendor or the product until now, the demonstration is a good forum to dig into it. Alternatively, you should request a workshop if you need to spend more time with the software package to verify whether it has a good fit for your business needs. Some things to look for during the demonstration are:

- Vendor/reseller background
- Product history and future releases
- User Interface
- Features
- Customization/flexibility
- Extraction, transformation, and loading of data from different data sources
- Technology platform
- Documentation
- Implementation
- Pricing

For detailed information in each area listed above, see the RFP example in Appendix A.

Post-Demonstration Evaluation

After the demonstrations, sit down and evaluate each one and select the favorite(s), or if more detail is needed or additional areas need to be covered, request a new demonstration. Use a Candidate Evaluation form to score each vendor. (See Appendix B.)

Request Final Proposal(s)

When you have found the vendor you feel has the right fit with your company's needs, request a final proposal with detailed maintenance, implementation, and support prices; implementation time estimates; sample project plan; and so on. (See Appendix F for a sample project plan.) If there are several vendors that might fulfill your needs at a similar level, do not hesitate to ask all of them for this information, so you can make a final, detailed comparison between them. In this case, you will also have some information to help leverage your negotiations.

Negotiate Contracts/License Agreements

Once you have picked a favorite vendor and software solution, it is time to request a contract. This document will normally include:

- Exact software pricing, based on your number of users, number of sites, modules needed, and the like
- Consulting and training rates and time estimates
- Software modules, based on the functionality you have requested
- Project outline, with main activities that need to take place prior to model completion (see Appendix F for an example)
- Hardware and software requirements
- License agreement (see Appendix C for an example)
- Support plan (see Appendix E for an example)
- Other information that you have requested in writing in the RFP

Make Final Vendor Selection

Based on the response(s) to your RFP, you will usually need some further conversations to discuss or clarify certain items before you feel comfortable with the information. One of the negotiation points is often pricing (see in Chapter 11, "Cost/Benefit Analysis" and "Return on Investment Analysis"). Once your company and the vendor have straightened out everything, the last step is to sign a license agreement, and then to pay (based on payment terms) and to receive the software and documentation. Sign the contract and facilitate payment.

USING A SOFTWARE SELECTION COMPANY

Many companies solicit the services of third-party firms to help them select the optimal BI solution. Sometimes these firms consists of single individuals with prior experience from the BI software industry that now work as consultants helping their clients to pick the package that is right for them. Other times, companies engage one of the Big 5 consulting companies. Be aware that many of the Big 5 have local, regional, or global strategic sales and/or consulting alliances with specific BI vendors. In any case, you want to do some advance research to be assured that their recommendations are not biased in one way or another.

In general, you should consider using a software selection company if one or more of the following are true:

- You don't have time to perform a thorough vendor/solution analysis yourself.
- You are considering technologies unfamiliar to you.
- Your corporate environment is very politically loaded, so an internal selection process alone can cause conflict between different decision makers.
- You want a third party to assist in doing a needs analysis for your company and in matching these needs with available BI products.

Before you engage a selection company, you should ask for a price quote and what the deliverables are. This will help prevent any costly surprises at the end of the engagement. Typically, costs range from $5,000 to $25,000, and can take from a few days to many weeks, depending on the time they spend analyzing your needs and participating in demonstrations and evaluations. Deliverables can be one or more of the following:

- Needs analysis
- Development of RFP document
- Selection of vendor finalists
- Participation in demonstrations

11

SOFTWARE EVALUATION: FACTORS TO CONSIDER

In general, the evaluation and purchase process for financial business intelligence (BI) software is similar to that of other niche software solutions, such as consolidation and reporting software or a budgeting software. In other words, it should require much less time for evaluations and presentations than a typical enterprise resource planning (ERP) solution with numerous modules (e.g., general ledger, accounts receivable, accounts payable, payroll).

EXPECTED USE NOW AND IN THE FUTURE

One of the first questions you should ask yourself during the software evaluation phase is whether you are looking for a short- or a long-term solution. There are cases in which a company needs to solve its analytical issues immediately, without having time to do an in-depth analysis of long-term organizational needs or long-term platform requirements. Sometimes this can be related to an in-house analytical model that no longer works properly or that is too hard to maintain. Other times it can be related to organizational politics and you cannot afford to wait any longer for parties in the organizations to provide their input to your software needs. However, in most cases, companies that invest in best-of-breed BI software are planning to stay with it for a long time, not only because it usually represents a significant investment in terms of money, consulting, and user training, but also because of the total time involved in setting up data extraction, conversion and loading procedures, and business views. It is not a process you and your co-workers want to go through every year.

In other words, if you are going for a short-term solution, you should not be looking at the most expensive and most complex BI packages. However, if this is a long-term investment, you should spend plenty of time looking into your own needs, product and vendor capabilities, and long-term maintenance cost, and focus somewhat less on acquisition and implementation cost.

Another element to consider is the *scope* of usage in the near future versus that expected in the long run. If your organization needs only a single location implementation and the BI model will be fairly simple with few planned changes, the software selection process should be fairly simple. If, however, you are planning to expand the BI model and organizational usage down the road, you should spend more time looking into the required functionality of software packages and their vendors. If you are acquiring a new software package today and you are investing in related training and implementation, you want to make sure that the software is scalable and can grow with you as the organization grows.

BI technology platforms are also important to consider if you are hoping to make your new application a long-term investment. Many companies attempt to standardize on a single BI server platform, such as Microsoft SQL Server, Oracle, or Hyperion Essbase, in order to reduce the need for human knowledge to manage the datawarehouse (backups, troubleshooting, maintenance, data cleansing, etc.) and to simplify integrations and the need for third-party tools. Sometimes a current or coming corporate technology/platform standardization can make your BI software selection harder, especially because the software you like the most might not be available on the required hardware platform or operating system. Today, the BI server platform supported by the largest number of vendors is Windows NT and Windows 2000, and fewer support Unix and AS400. The latter two platforms have traditionally been chosen for their ability to run on hardware that can handle larger data volumes (not necessarily true anymore), but these software/hardware solutions are also a bit more expensive than, for example, a Windows 2000 platform running on an Intel-based server. This type of conflict of interest will often force you to perform a cost/benefit analysis to figure out what is most important to your company: software features or database platform. No BI software is best at everything. If you are planning to use your new BI application in several different areas of the company, it can therefore be a good idea to chose a BI platform (e.g., Microsoft Analysis Services or Hyperion Essbase) that is supported by a number of different BI end-user tools from various vendors, so you are not locked into one specific end-user tool.

GETTING THE REST OF THE COMPANY TO BUY IN TO THE PROCESS

If you have made up your mind that your company needs a new BI package, you will often still have to convince the rest of the people that will be affected by the new software. If you move ahead on your own without conferring with

the key users in the organization, chances are that few of them will support you if anything goes wrong during or after implementation, or if you have to suggest work-arounds and the like.

There is no sure formula for success in achieving company-wide consensus for a new software acquisition, but there are several things you can do to improve your chances. Today, most companies that decide to look for a new BI solution put together a project team. This team normally consists of representatives of end users, information systems specialists, and key decision makers. The advantages are that each team member will bring special expertise and insight from his or her own area of the organization. The team will set up evaluation criteria (see RFP in Appendix A) to help screen the software packages and vendors with the best fit. They will participate in software demonstrations to see firsthand the different products. Do not expect end users to be able to list every possible requirement from their area when creating the RFP, and do not push them to predict every possible need. This could easily result in an unmanageable and too detailed RFP. Through a point scoring system (see Appendix B) or other means of evaluation, the project team reaches a consensus and then makes the software acquisition. Whether the decision leads to the best possible solution or not, the organizational support behind the project will be much stronger than if one person alone had made the evaluation. They have now been part of the decision process themselves and they provided input in the areas they deemed the most important.

One of the most typical problems with an organizational rollout of a new BI software is most frequently found in larger, multisite organizations. Remote offices are often not involved in a headquarters' software evaluation and acquisition, but they are required to use the software as it is rolled out. This will often cause less-than-optimal use of a new software package at remote sites, and not always ideal cooperation and goodwill to resolve ongoing analysis and reporting issues. Most companies that are successful in achieving organization-wide support do their own sales pitch to their divisions before a new BI software is rolled out. By explaining the advantages of the new solution and by providing sufficient training and other assistance when the software is rolled out, they improve general understanding and goodwill toward the new solution.

COST/BENEFIT ANALYSIS

As a selection tool, a cost/benefit analysis is very helpful to assist you in focusing on the most suitable category of software, rather than trying to differentiate between similar BI solutions. For example, it can help you find out if your needs are better covered by a particular type of solution, such as:

- Proprietary versus open database platforms
- High-end software versus mid-market and low-end software
- Web-based versus client–server software

In addition to helping to distinguish between different software categories, a cost/benefit analysis can help you make the decision of whether to stay with what you have or to go ahead with a new software acquisition.

If you were to look at a range of companies using the different BI packages on the market, you would see that their level of satisfaction ranges from highly satisfied to dissatisfied. The success with a particular solution is dependent on several factors:

- Software features
- Skills of implementation partner
- Training, skills, and involvement of key employees
- Long-term support and software upgrades from the vendor

In the software selection phase, matching organizational needs and constraints with software features is key to success. One of the best ways to do this is to set up a list that weighs benefits against costs. (See Figure 11.1.)

Benefits	Score	Costs	Score
Saves x man-hours for end users	6	$X in software and hardware required	–8
Improves control	7	X days of training/customization needed	–6
Offers better reports/analysis	9	Unfamiliar user interface	–7
Saves $X and time in report distribution	5	Need integration to ERP modules	–2
$X less than in-house programming	7	Takes time away from other projects	–4
Comes with good documentation	4	Might be replaced by new technology in a few years, thus requiring new investment	–4
Score	**+38**	**Score**	**–28**
TOTAL SCORE			**+10**
Scoring is from 10 to –10 for each item, where 10 is the highest score and –10 is the lowest score. A total score above 0 shows that the company most likely will achieve an overall benefit by going with a new software package.			

FIGURE 11.1   Cost/benefit analysis for a new software acquisition example

A tool like the one above can often become rather subjective because it is hard to assign a value to many intangible items. However, because each item is weighted, it is usually better than just listing items with no assigned value, for and against a new software.

RETURN ON INVESTMENT ANALYSIS

While a cost/benefit analysis can help you quickly establish whether the overall benefits of a new BI solution will exceed the costs, it does not put quantitative numbers on the investment. Although a more time-consuming and difficult task, a return on investment (ROI) analysis will help you estimate the costs and the returns of a BI software implementation, to decide whether or not the investment is worthwhile. It is very easy to assume that installing flashy, new technology cannot possibly be a bad decision, but unfortunately it has been for many companies due to some of the following reasons:

- Excessive cost of BI tool compared to actual use in organization
- Purchased the wrong BI tool for the job (poor fit versus need)
- Poor data quality and/or detail (due to lack of integration with data sources)
- Slow performance (due to lack of hardware investment)
- Lack of usage due to poor training and/or not intuitive user interface

To ensure the best possible investment decision, an ROI analysis can be a helpful tool. Figure 11.2 is an example of a simple ROI analysis.

Estimated Investment Life (3 years)	
Item:	
• BI Software (all modules)	$200,000
• Annual enhancement fee ($40,000/yr)	$120,000
• Implementation	$100,000
• Maintenance and support of BI Model	$100,000
	$520,000
Estimated Return:	
• Cost savings in report writing ($30,000/yr)	$ 90,000
• Cost savings in report printing/distrib. ($50,000/yr)	$150,000
• Value of better decision making ($200,000/yr)	$600,000
	$840,000
ROI	$840,000/$520,000 = 162%

FIGURE 11.2   ROI analysis example

In Figure 11.2, the return on a BI investment over a three-year period is estimated at 162%. Unless the same capital can be employed with a higher return elsewhere, this example tells us that the investment seems to be well worth it. A follow-up analysis should be carried out after each year to verify that the original estimates are accurate.

FEATURES AND FLEXIBILITY

This is almost always the area where companies spend most of their time during the software evaluation phase. The key is to try to identify the functionality that your company needs and then try to match it against what the different software vendors have to offer. Seldom will you find a perfect match, because most of the BI solutions on the market are built to fulfill the needs of a broad range of companies in the marketplace rather than all of your specific requirements. However, you might find that the lack of some of the features you are looking for might be balanced out by other functionality offered by a particular software package.

At the core of a feature study is the creation of the RFP (request for proposal) document that specifies all the key features and requirements a company is looking for rather than going into detail on specific features (see Chapter 10 and Appendix A for detailed information on RFPs). See the following sections for some general advice on evaluation of software functionality.

No matter how you chose to execute your software selection process, keep in mind that it must fit the different categories of users in the organization. User categories typically include:

- End users—will need basic drill-down and charting features
- Power users—will need full-blown functionality to set up brand new views/reports and apply filters and other business rules
- Information recipients—might need to receive only static reports by e-mail or web links on a periodical basis

COMPATIBILITY WITH EXISTING SOFTWARE

It would make software evaluation a whole lot simpler if we could ignore all current systems in place and just focus on the features and functionality of a BI package. But there are usually several other software packages that need to integrate with a company's new BI software. (See Figure 11.3.)

For example:

- General ledger(s)
- Current spreadsheet models
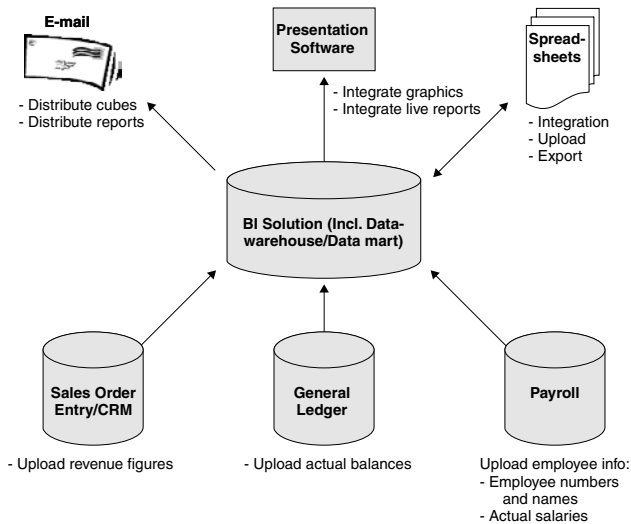
FIGURE 11.3 Compatibility issues to consider with a new BI solution

- Payroll/human resource system(s)
- Sales order entry system(s)
- Datawarehouse/data mart
- E-mail program(s)
- Presentation program

Traditionally, exporting and importing text files have been the medium of exchanging data between different applications. Some of the downsides of this are the extra manual work with possible human errors and mistakes and slow data transfer speeds. This is rapidly changing as BI and accounting applications as well as database vendors build custom interfaces to each other's applications and databases, and as open standards such as XML/XBRL and ODBC are adopted (see Chapter 1 for more information). When you are evaluating a new BI package, you first need to identify which of your own current and planned applications it should integrate with, and then ask the vendor to explain how this will be handled.

EASE OF USE

As with all other software out there, the most powerful BI applications on the market are not necessarily the easiest ones to use. Similarly, a simple application with just a few screens and menus can probably be mastered in a few hours, but it will also not have the same amount of features and flexibility as more

advanced BI software. In particular, you will find this to be true in the area of designing different user views (such as charts, reports, etc.) and applying business rules, such as filtering, ranking, subtotals, variance columns, calculated measures, and so on. The more you want to customize end-user views to fit your company's particular business model and issues, the more powerful BI tool you will need and you should expect to spend more time learning it. While most high-end applications offer customization of charts, labels, reports, and the like, lower-end software usually has limited functionality in these areas.

All modern BI applications have graphical user interfaces in which you can use a pointing device, drill-down in charts, reports, and so on, and screens offer a variety of different fonts and colors. Online help files and context-sensitive help have also become common features to improve user friendliness. Each vendor has come up with its own unique screen layouts and user-related functions. The best way to get a good feel for it is to watch an in-depth demonstration or, even better, to request a demo copy or a workshop where you can get some hands-on experience with the ease of use of the application. The key issue is to find out if power users as well as end users will be comfortable enough with the application to adopt it and learn all key features. The last thing a company wants is an application that is completely consultant dependent or that nobody knows how to use. Often, a company will find that they need the power of an advanced BI solution, and they accept that there is a ramp-up period to get users up to speed. In this case, the key is to provide good training and to motivate key employees to learn the software in detail.

SOFTWARE STABILITY

Software stability is often an overlooked item during a software evaluation. Often, the decision to purchase one software over another can come down to small differences in features or how well the vendor presented itself and its product. However, the closest a company usually comes to checking out the stability of the current version of the software is when they call up references and ask what their opinions about the package are. The same software can run better on one database or hardware platform than on another, and this is often ignored during a reference check. Also, different customers using the same application on the same database and hardware platform rarely use all of the same features. Certain features might cause crashes, memory leaks, or other stability problems, and this might be a problem for only a handful of customers.

To get the best possible understanding of different stability issues, you should ask for references with similar hardware/software and datawarehouse/data mart configurations as the one your company plans to use. You should also determine if the vendor has a reputation of producing reasonably stable software and get an understanding of how the vendor will handle it if you experience that type of problem after you have purchased the software.

VENDOR-RELATED ITEMS

Sometimes it is easy to get buried in small details of a product and forget that more important than the product itself are the company and people behind it. A number of times in the past, excellent software products have disappeared from the market in a few years because the company behind it ran into problems. Such problems could be:

- Financial mismanagement
- Human resource–related issues (e.g., poor retention of key employees)
- Poor strategic vision and business execution skills
- Large-scale lawsuits or legal problems

Some of the qualities to look for in a vendor are:

- A good reputation in the marketplace
- Happy users for the past several years
- Availability of high-quality training and support in different regions and time zones
- A strong distribution channel and/or strong direct sales and service force
- Utilizing up-to-date technology in development, products, implementation, and support
- Online customer access to support-history database
- Professional development and testing methodologies
- Strong sales and good profitability (*Note:* Do not fall into the trap of believing the companies with the highest revenues or most installations have the best products. Many great products come from companies that initially were small in size.)
- Good future direction and strategies
- A well-developed list of third-party products that work with their software (e.g., web browser versions, databases, general ledgers, spreadsheet add-ins)
- Offer hosting services either themselves or through application service providers

WORKING WITH AN IMPLEMENTATION PARTNER

Whether you utilize implementation services from the product vendor or a reseller or other partner, make sure that you will be working with a person(s) who is not only trained on the product, but has prior implementation experience. In other words, before engaging a consultant, ask how long his or her experience

with the product is and how many other implementations of this particular product the person has done. Some companies even ask for references from the assigned consultant. Other things to look for are:

- Type of implementation methodology used (e.g., project plans, proto-typing, testing)
- Average implementation time for similar projects
- Knowledge of database platform
- Knowledge of integration tools (ETL) used to extract, transform, and load data from your other applications

HOW TO SELECT: SUMMARY

The process of selecting a new software is handled in a wide variety of ways at different companies, from a single person doing a quick evaluation to large project groups doing an in-depth analysis of product, vendor, consultants, and existing customers. The following list provides some tips before you start your software evaluation process:

- Does your company have the skills to handle the evaluation, or should you engage a software selection company? (See Chapter 10 for more details.)
- Do a thorough analysis of your current and future analytics needs before you start looking at vendors.
- Create a request for proposal (RFP) document (see Chapter 10 and Appendix A for more information) to communicate your company's needs to vendors, or, as an alternative, engage directly in an in-depth dialogue with a group of preselected vendors.
- Have all key people present at the product demonstrations.
- If needed, call vendors in for repeat demos or specifically focused demos.
- If you believe your BI model and process is unique in any way, provide the vendors with specific information and request a customized demon-stration.
- If you are in doubt about key feature areas and how your users will handle them, request a hands-on workshop.

See Chapter 10 for more in-depth information on the software selection process.

12

OUTSOURCING: THE NEW ALTERNATIVE

Today, an increasing number of businesses seek to focus on their core functions instead of their software, creating a shift toward outsourcing that is leading businesses to application service providers (ASPs) for their information technology (IT) needs.

One of the popular trends in the software marketplace is application outsourcing, or *hosting services* as many people refer to it. What this means is that instead of installing and servicing your new software application in-house, you contract with an ASP to host the application for you. Having a third party host an application for you is by no means a new concept. For more than a decade, there have been ASPs that have worked as if they were your own information systems (IS) department, except that they are not on your payroll and the application usually does not reside in your own facilities, and not on your own hardware either. During the time when mainframes were the key platform for most business applications, many companies chose to let an ASP host their application. One of the key drawbacks was high communication cost (usually dedicated phone lines) to link the user's computer/terminal to the application. Today, that is quickly becoming less of an issue with the use of the Internet, corporate wide area networks (WANs), and much lower telecommunications rates.

HOW IT WORKS

Utilizing the Internet or other means of telecommunication (such as dial-up links, virtual private networks, dedicated lines, and a number of other options), you connect to the software application residing at the ASP's site. (See Figure 12.1.)

Everything else you do inside of the software works just as if you had the software on one of your own servers in your office. More and more BI applications come with Web interfaces, which means that the user needs only a Web browser (e.g., MS Explorer or Netscape) and Internet/intranet access to work with the application. Backups, application upgrades, server hardware maintenance, and the like are handled by the ASP. For this service, you normally pay a fixed or variable (or combination of the two) subscription fee. In this new era of application outsourcing, vendors and ASPs are still analyzing how to best position themselves with pricing and software-related services with regard to application hosting. In some cases, you might purchase the application and then let the ASP host it. In other cases, you might never own the ap-
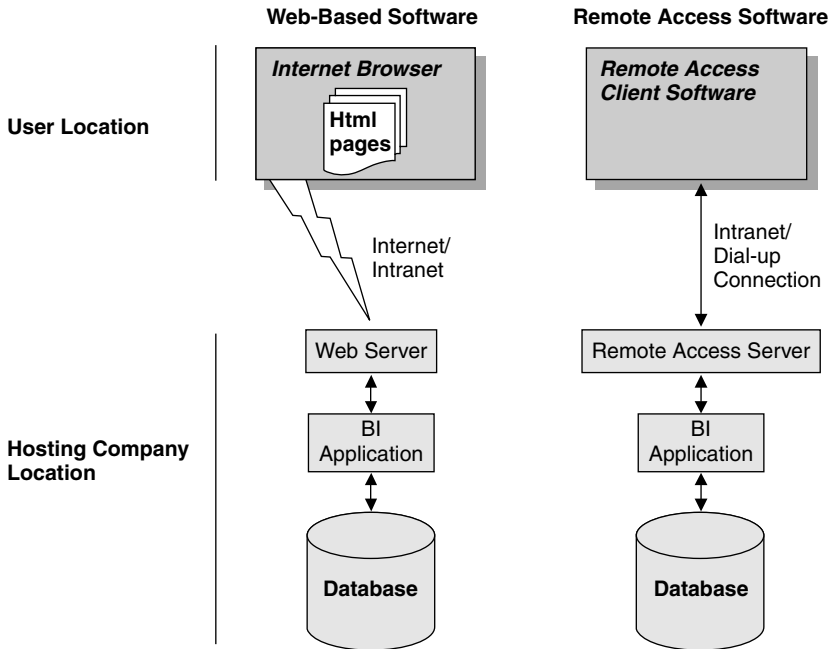
FIGURE 12.1   Configurations used with application hosting example

plication, but rather pay per usage. For example, you might be charged for the length of time logged in, amount of processing power used (typically measured by central processing unit [CPU] time used), number of transactions entered/processed, and number of registered users.

WHEN YOU SHOULD CONSIDER AN APPLICATION SERVICE PROVIDER

Before you make the decision to install your BI application in-house or to outsource it to an ASP, you need to analyze your organization's needs and capabilities. If the bottom line is that the company will reduce costs or increase revenue by outsourcing the application, this should be a key deciding factor. Of course, there are also other elements involved, such as security concerns and user satisfaction, that will weigh the decision one way or the other. Some examples of factors indicating that application outsourcing might be beneficial are:

Internal Factors

- *Lack of in-house technical support and skills.* Recruiting and retaining skilled IT professionals is often difficult and expensive. Your current IT staff might not have any available time to support an additional in-house application.

- *No in-house power users.* Most advanced BI applications are not learned overnight. A company needs power users who can set up reports, graphs, and the like, as well as maintain the datawarehouse/data mart when the chart of accounts, organizational structures, and information requirements change. If you don't have any people with the skills or time to become power users, you will need frequent visits from consultants or your BI model will quickly become obsolete.

- *No budget for necessary infrastructure.* An in-house BI application usually requires a powerful database server, and if you have remote offices that need to access the application remotely, you need a Web server and proper security hardware/software installed. In addition, the cost of the BI application itself can be higher than outsourcing it and instead leasing it or paying for the amount of usage.

- *Rapid organizational change can outdate software.* Most ASPs offer arrangements in which you don't have to buy the application, but you lease it and/or pay per usage. This means that if your company's needs outgrow the capabilities of the BI application, it might be quicker and less costly to move to a new software if you are leasing it.

External Factors

- *New BI application technology.* Until recently, BI software packages have not been built with technologies that allow for dynamic Internet/intranet access through, for example, a browser. Today, however, this is rapidly changing and remote access to a central server is becoming part of the standard capabilities. This means that an ASP can host the whole application, while all users access it remotely without any type of client software installed on their machines.

- *Lower telecommunications costs.* Many years ago when data centers where hosting mainframe applications for corporations, telecommunications were one of the major expenses. Today, however, increased competition among telecom companies as well as the emergence of the Internet as an inexpensive information highway, linking up with an application residing at the ASP does not have to be a major expense.

- *Better support.* If the ASP offers full support services (including BI model support), their support staff will have immediate access to look at your issue first hand, because the application resides in their location. This should translate into better and faster support. If you receive support from a third party, such as the software vendor or a reseller, your infrastructure when using an ASP can also let their support people get access to your model from any location (obviously only with proper required security).

- *Lower consulting costs.* Just as your infrastructure when using an ASP can let remote support people gain direct access to your application, consultants can do the same. This can translate into cost savings in reduced travel, hotel, and per-diem expenses, as well as quicker response from consultants because of zero travel time.

SELECTION CRITERIA FOR AN APPLICATION SERVICE PROVIDER

Because the application outsourcing trend is relatively new, there is little history to refer to as far as customer satisfaction rates and cost and service offerings. It is therefore of importance to carefully study your potential ASPs before you move ahead. The players in this market consist of:

- Firms that have traditionally (in the mainframe era) provided outsourcing services
- Software vendors and resellers of software companies
- Hardware vendors

- Internet service providers (ISPs)
- Telecommunications companies
- Business process consultants

If you have decided to outsource your application, you need to select an ASP. The following list provides items to consider when evaluating an application service provider:

- *Experience.* The prospective ASP should have some experience with the application. Are they trained to understand the software package and the technology platform it is based on? Some ASPs will take responsibility only for backups and making sure the application runs satisfactorily. In this case, you will also deal with the software vendor or reseller with regard to software consulting and support services. Other ASPs are full-service providers, and they will offer you implementation and support through a staff of consultants that are trained on the application.

- *Knowledge.* The ASP should have basic knowledge of the analytical process and related procedures, so they can be prepared to handle heavy usage time periods (such as at the end of a reporting cycle).

- *Up-to-date technologies.* Does the ASP have a modern technology infrastructure? For example, the Web is emerging as the key delivery mechanism for accessing remote applications. The outsourcing vendor you consider should offer the hardware, software, and knowledge to satisfy your requirements in this area.

- *Vendor experience.* Because application outsourcing is a relatively new field, few ASPs have long experience with the technology and services involved. However, this does not mean that most vendors offer the same level of quality. Find out who the vendor has hosted applications for and for how long, and check references.

- *ASP flexibility.* Does the ASP have business strategy that is aligned with your own company's goals for future growth and change? Try to discern the most likely software and organizational changes that your company will go through down the road, and then get a feel for how the ASP will cope with these changes.

- *Customer service.* As with all your other business partners, you want to deal with people who are friendly and service minded. If you run into problems with the application, you do not want to increase your potential aggravation and stress level by having to deal with people who are not on their toes to help solve the problem and to put you back on track.

ENSURING CONTINUOUS SUCCESS

When you have found an ASP that fits the requirements of your company, the next step is to structure a contractual agreement that protects the interests of both parties and helps ensure mutual understanding and long-term success.

Make sure the contract properly covers key areas such as minimum acceptable processing speeds, data access security, backups, and disaster recovery planning. Because your own company's sensitive data will reside in the database at the ASP's site, it is important to include nondisclosure agreements (in case the ASP needs to get access to your data). If the ASP is also going to provide you with application consulting and support services, details for these also need to be specified.

Work to achieve a mentality of partnership. Hosting a BI application has traditionally been an in-house arrangement, and by outsourcing it, you are looking to achieve additional benefits. To ensure that your decision to outsource leads to long-term benefits, it is important to nurture the relationship with the ASP through regular status meetings and an open line of communication.

13

BUYER'S GUIDE

Business intelligence (BI) software used in the finance and accounting area, has evolved from report writers and query tools, to OLAP databases, data mining, and enterprise information portals. The product and vendor lists on the following pages include a variety of different software solutions. The goal is to give you a good starting point for your own in-depth research. Remember that both vendors and their products evolve, and names, platforms, and functionality will often change over time, so always be prepared to also look for new names and technologies that can support your BI needs.

There are several ways to categorize and group BI tools (see Figure 13.1). We have broken them down into three major categories:

1. Query and Reporting Systems
2. Decision Support Systems
3. Enterprise Information Portals

General Characteristics

Easy-to-use interface (typically web based), with single point of access to information from across the organization → Enterprise Information Portals

Sophisticated analysis of higher-level data, usually from specific data sources → Decision Support Systems

User-friendly and flexible report writing and analysis → Query and Reporting Tools
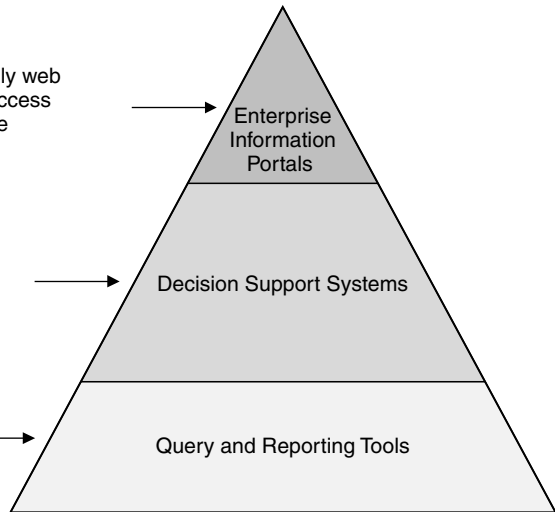
FIGURE 13.1   Categories of BI Tools

QUERY AND REPORTING SYSTEMS

Query and reporting systems allow you to look at transactional data on a row-by-row, column-by-column, and cell-by-cell basis. You can search and sort your data in a myriad of ways, perform everything from simple to advanced calculations on your data and output to various levels of summarization of the data, including subtotals and grand totals in reports.

Query solutions are often divided into two categories: (1) *managed query* tools and (2) *user-defined query* tools. Managed query tools present the data to the end users in a predefined business-oriented manner, and the information technology (IT) department typically handles the back-end logic of managing and describing the data. User-defined query tools require the end user to understand the layout of the database. Although it provides more flexibility in terms of reporting, this category is mostly suited for highly technical users or programmers.

Managed query and reporting tools provide flexible, user-oriented report-writing capabilities while shielding individuals from complex programming tasks. Report writers, in particular, are ideal for producing highly formatted, presentation-quality financial reports, while many managed query tools are faster (because they often run on OLAP databases with data summarized at different levels) and easier to use, and excellent for ad-hoc reporting and analysis.

Query Tools

Major vendors of query tools include:

Vendor Name	Product Name	Web Site
Brio	Brio.Report	*www.brio.com*
Business Objects	Businessobjects	*www.businessobjects.com*
Microstrategy	DSS Agent	*www.microstrategy.com*
ProClarity Corp.	ProClarity	*www.proclarity.com*
Oracle	Oracle Discoverer	*www.oracle.com*
Cognos	Impromptu	*www.cognos.com*
Microsoft	Excel 2000	*www.microsoft.com*

Report Writers

Financial report writers include:

High-End and Mid-Market Report Writers (with consolidation functionality)

Vendor Name	Product Name	Web Site
Comshare	FDC Commander	*www.comshare.com*
FRx Corporation	FRx	*www.frxsoftware.com*
Hyperion	Hyperion Enterprise	*www.hyperion.com*
iLytix	XL Reporter	*www.ilytix.com*
Khalix	Longview Solutions	*www.longview.com*
SAS	CFO Vision	*www.sas.com*
Solver	Enterprise Reporting	*www.solverusa.com*
SRC	Information Advisor	*www.srcsoftware.com*

Low-End Report Writers

Vendor Name	Product Name	Web Site
Seagate Software	Crystal Reports	*www.crystaldecisions.com*
Synex Systems	F9	*www.f9.com*

DECISION SUPPORT SYSTEMS

Decision support systems (DSS) are used for different types of analysis of business data. Typically, DSS applications use summarized data derived from transactional information to allow users to analyze past results and forecast future trends. Two common types of DSS technologies are: (1) online analytical processing (OLAP) and (2) data mining.

OLAP

OLAP solutions are fast and powerful tools for reporting on data (as compared to data mining tools that specialize in finding patterns in data) stored in a database. OLAP enables users to analyze different dimensions of multidimensional data. For example, it provides time series and trend analysis views. It provides sophisticated data analysis capabilities to end users enabling interactive, dynamic analysis of business information.

The main component of OLAP is the OLAP server, which sits between a client and a database management system (DBMS). The OLAP server understands how data is organized in the database and has special functions for analyzing the data. There are OLAP servers available for nearly all the major database systems. Many of the OLAP servers come with their own user interface, and some (e.g., Microsoft Analysis Services and Hyperion Essbase) also have a number of third-party OLAP front-end tools (see "Query Tools") that integrate with their OLAP server.

Major vendors of OLAP solutions include:

Vendor Name	Product Name	Web Site
Applix	TM1	*www.applix.com*
Brio	Brio Enterprise	*www.brio.com*
Cognos	Powerplay	*www.cognos.com*
Crystal Decision (formerly Seagate)	Holos	*www.crystaldecisions.com*
Hummingbird	BI/Analyze	*www.hummingbird.com*
Hyperion	Essbase	*www.hyperion.com*
IBM	DB2 OLAP Server	*www.ibm.com*
Microsoft	Analysis Services	*www.microsoft.com*

Microstrategy	Microstrategy	*www.microstrategy.com*
Oracle	Oracle 9i OLAP Services	*www.oracle.com*
PricewaterhouseCoopers	Carat	*www.pwcglobal.com*
Sagent Technologies	Sagent	*www.sagent.com*
Whitelight Systems	Whitelight	*www.whitelight-tech.com*

Data Mining Tools

Data mining is a *decision support* process in which you search for patterns of information in data. It identifies hidden relationships in large volumes of data that provide valuable insights into key business drivers such as buying behavior, credit usage patterns, and sources of risk.

Typical users of data mining tools have specialized skills both to use the mining tools and to understand the data and patterns they are analyzing. There are typically only a few "data miners" in a corporation.

Major vendors of data mining tools include:

Vendor Name	Product Name	Web Site
IBM	IBM Business Miner	*www.ibm.com*
Microsoft	Microsoft Data Mining	*www.microsoft.com*
SAS	SAS Enterprise Miner	*www.sas.com*
SPSS	SPSS	*www.spss.com*

ENTERPRISE INFORMATION PORTALS

Enterprise information portals (EIPs) are the highest level of business intelligence applications. Typical characteristics of modern EIP applications are that they provide a highly user-friendly and customizable web-based interface. Until recently, very few commercial EIP applications have been available, so companies built their own EIPs using their IT department and/or hired developers. Vendors now offer portal software that a consultant or a trained in-house resource can set up and customize to provide easy access to all the key information sources in the company.

As portal solutions become increasingly popular, several categories of portals are emerging. Some are more generic and are meant to interface to any type of data source, and on the other side there are portals that are specialized

for business intelligence. This first type of corporate portal is called an enterprise information portal (EIP), while the second is often categorized as a business intelligence portal (BIP). Because EIPs and BIPs are built to present information from a large variety of data sources with a single user-friendly interface, there will typically be a large number of users utilizing this technology in a company. An increasing number of BI software vendors (see list below) now also offer a BIP, and these portals can often plug into an EIP as key vehicles for delivery of BI specific information. For more information about BIP portal vendors, you can check out the individual web sites of the BI vendors listed in this chapter.

Major vendors of enterprise information portal software include:

Vendor Name	Product Name	Web Site
Bowstreet	Enterprise Portal	*www.bowstreet.com*
Computer Associates	MyEureka	*www.ca.com*
Hummingbird	Enterprise Information Portal	*www.hummingbird.com*
IBM	IBM Enterprise Information Portal	*www.ibm.com*
Microsoft	Digital Dashboard	*www.microsoft.com*
Oracle	Oracle Portal	*www.oracle.com*
Plumtree	Plumtree Corporate Portal	*www.plumtree.com*
SageMaker	SageWave	*www.sagemaker.com*
Viador	E-Portal	*www.viador.com*

DATAWAREHOUSE SOFTWARE

In order to successfully employ a BI tool in an organization, it is critical that the data to be analyzed and reported on is easily accessible and that it is organized in a logical manner. A datawarehouse (can be very large databases that involve implementation projects and software worth millions of dollars, or smaller, less complicated databases that will cost in the tens of thousands and up to buy and implement) will provide the accessibility and clarity of data that will allow the users to get the full "bang for the buck" with their BI tool. In short, a datawarehouse is a database specially suited for decision support and business intelligence. Major vendors of datawarehouse software include:

Vendor Name	Product Name	Web Site
IBM	DB2	*www.ibm.com*
Microsoft	SQL Server	*www.microsoft.com*
Oracle	Oracle	*www.oracle.com*
Sybase	Adaptive Server IQ	*www.sybase.com*
SAS	SAS Intelligent Warehousing Solution	*www.sas.com*

EXTRACTION, TRANSFORMATION, AND LOADING TOOLS VENDORS

Although just a "middleware" (a software used between two other software applications) needed to move and convert data from one database to another, ETL software can range from relatively inexpensive to very costly. In addition, the time and expense of implementing the ETL tool can make the entire investment fairly large. However, a commercial ETL tool that works with your databases will typically provide a faster and simpler way to transform and move data than something you build in-house. It is therefore important to do some research up front to make sure you choose the tool that has the best fit, for both your functionality needs and your budget. Different ETL tools include:

Vendor Name	Product Name	Web Site
Ab Initio	mPower	*www.abinitio.com*
Acta Technologies	Acta Works	*www.acta.com*
Ascential	Data Stage	*www.aboutnewco.com*
Coglin Mill	Rodin	*www.coglinmill.com*
Cognos	Decision Stream	*www.cognos.com*
Computer Associates	Decision Base	*www.ca.com*
Constellar	DataMirror	*www.constellar.com*
Data Junction	Data Junction	*www.datajunction.com*
DataMirror	Transformation Server	*www.datamirror.com*
Decisionism	Broadbase	*www.decisionism.com*
Evolutionary Technologies	ETI Extract	*www.evtech.com*
Hummingbird Communications	Genio	*www.hummingbird.com*

Vendor Name	Product Name	Web Site
IBM	Visual Warehouse/ DataPropagator	*www.ibm.com*
Informatica	Powercenter/ PowerMart	*www.informatica.com*
Metagon Technologies	Dqtransform	*www.metagon.com*
Microsoft	DTS	*www.microsoft.com*
Oracle	Oracle Warehouse Builder	*www.oracle.com*
Sagent Technology	Sagent Solution	*www.sagenttech.com*
SAP	Business Warehouse	*www.sap.com*
SAS	SAS/Warehouse Administrator	*www.sas.com*
Sybase	Industry Warehouse Studio	*www.sybase.com*
Systemfabrik	Warehouse Workbench	*www.systemfabrik.com*
Taurus Software	Warehouse	*www.taurus.com*

eLEARNING TOOLS VENDORS

Unless users are given proper and timely training, the return on BI tool invest-ment will drop dramatically. Attending product training classes can be expen-sive and often hard to coordinate with everybody's schedules. An electronic training medium that today often is referred to as *eLearning* allows companies (both software vendors and their customers) to develop and utilize the Internet or CD-ROM as a channel to deliver the training on demand. As organizations have started to embrace the benefits of online education and training, an in-creasing number of eLearning software vendors have appeared. Well-known eLearning solutions providers include:

Vendor Name	Product Name	Web Site
Centra	Centra 99	*www.centra.com*
Click2learn.com	ToolBook II Instructor	*www.click2learn.com*
Docent	Docent Enterprise	*www.docent.com*

Interwise	Millennium	*www.interwise.com*
Knowledgeplanet.com	KP 2000	*www.knowledgeplanet.com*
Learn2.com, Inc.	StreamMaker	*www.learn2.com*
MacroMedia	Macromedia Authorware	*www.macromedia.com*
RWD Technologies	RWD WILS	*www.rwd.com*
Saba Software	Saba Learning Enterprise	*www.saba.com*
WBT Systems	TopClass Publisher Studio	*www.wbtsystem.com*

PART FOUR

IMPLEMENTING A BUSINESS INTELLIGENCE SYSTEM

14

PROJECT PLANNING

Lack of project planning will set the stage for failure on a business intelligence (BI) project. This is obvious for any business endeavor. However, for BI projects, overplanning can also set the stage for failure. This is a surprise for most newbies to the BI world. Overplanning hasn't killed a BI project, but it can delay its delivery to the point that the problems it was originally designed to address no longer exist. A system is delivered that is not useful, which is one definition of failure.

The most common (and costly) mistake a BI project can incur is trying to deliver a finished product or final set of information the first time. Refer to Chapter 6 for more details on why attempting to deliver a "finished" product can set the stage for failure. To summarize what is said in the introduction to Chapter 6, there should be no endpoint, or finished deliverable, in a BI project. Think of the BI project as an *evolution*, not a *revolution*.

BUSINESS INTELLIGENCE PROJECT DOs

There are some key issues that often are not addressed properly when pulling together a BI project. At a high level, these issues are apparent in any information technology (IT) project, but details for these issues are different in a BI project.

Get the Right People Involved

This is not an IT-only project. BI projects need to deliver business value, and leaving the definition of business value to the IT people will not deliver the expected results. Here's a quick story to illustrate that point. I was once in a meeting with a large health care provider. This was a business needs definition meeting for a BI pilot project. We had progressed to the point where our discussion was revolving around the business definition of some of the data sources. I asked a very pointed question about some specific columns in one of the data tables. The question was phrased something like this, "What are these two columns?" An individual replied, "Well, those are numeric columns representing dollar amounts." He was contradicted by another individual who said, "No, those are ASCII values that represent the dollar amounts." A heated discussion ensued in which they argued about the data type of the columns for a few minutes. As their discussion cooled a bit, I was able to interrupt and ask, "No, I don't want to know the data type. What I need to know is what do those values mean for the business?" Both replied, "We don't know." This example may be a bit extreme, but it helps illustrate that leaving a BI project definition to the IT group alone may not deliver the desired business value.

A good mix of individuals that form the BI project definition team will come from different segments of the business. As the BI project progresses, the mix of individuals needed to meet project milestones will change. This section will focus on the individuals who will provide the most value during the initial definition of the BI project. The skills of the individuals that would best comprise the initial team are listed below. These do not need to be separate individuals, and in most cases are not. Often, one person will fill two or more roles.

- *Business problem understander(s)*. This person (or people) understands the business problem(s) the BI project is targeted to solve.

- *OLAP expert*. Understands the capability and limitations of the OLAP technology. Manages the expectations on functionality that can be expected from the OLAP implementation.

- *Source data expert*. Knows the transaction system(s) from which the datawarehouse, or data mart information will be drawn. Manages expectations on what data is available, and if it's not available, what it will take to acquire that data.

- *Infrastructure advisor*. Familiar with network topology and potential routes for data transference. Capable of addressing questions about server capacity, network throughput, and available bandwidth.

- *Application or information delivery specialist*. Building the datawarehouse/mart and the OLAP cubes is not enough. The information must be

delivered into the user's hands or the project will be pointless. This individual can advise on the delivery mechanisms available, or possible. Drives the selection of delivery applications or tools.

- *Project leader/manager.* Steers the project to closure points and milestones. This is almost always done along with one of the other roles.

Plan

Plan, but don't overplan. A good pilot project, or initial BI project plan should look roughly like these five major steps:

1. *Determine KPIs.* Get organization, corporate, and departmental level objectives. Ensure that they align. Answer the question, "What are Key Performance Indicators?" Answer the question, "How do we measure the outcome?" This drives us to the final question, "What metrics (measures) fulfill the answers to the above questions, and what are their attributes?

2. *Plan delivery vehicles.* Define who will be using the information. These will typically fall into groups of job responsibilities. Examining common usage patterns and access methods will point to the best mechanism for the delivery vehicle to those groups. Some typical usage patterns are:

   - *Monitoring.* This is a group that simply "looks" at the information in a post-analysis format. They have little or no need to explore and discover more information than what is presented. An example is a warehouse manager who simply wants to inspect how many units of a particular product line were shipped over the previous 12 months. A simple display often served by a static web page can meet this group's information needs.

   - *Periodic/episodic inspection.* This group accesses information more frequently and drills to the details of summarized numbers. An example is a salesperson who looks at last quarter's sales trying to find the products with the highest profit margins that are being underserved in certain regions. Some level of information exploration is required. This can be done in a web browser interface, or perhaps a "dumbed-down" version of a desktop application.

   - *Strategic analysis and planning.* This group produces the post-analysis reports. Their job is to drive strategic and tactical direction based on information analysis. A rich client desktop is often required to meet their "power-user" needs.

3. *Build the data marts and interfaces.* The bulk of the initial work is in this stage. It consists of:

- *Schema designs.* This references the schema designs specific to OLAP, and are usually a *star schema* or a *snowflake schema.* This design is the direct result of the business needs definition and should be done prior to examining the source data or other potential limitations on the model. Think of it as a theoretical goal of what is required to meet the business needs. As the source data is examined and other limitations defined, a reality check is forced on this theoretical model. The schema design defines the dimensional definition of the data marts and their metrics. See Figure 14.1 as an example of a star schema for a retail sales sample data mart.

- *Locate source data.* The data populated in the data mart star schema or snowflake schema is usually derived from a transaction system. That source data needs examination to see if it will meet the require-
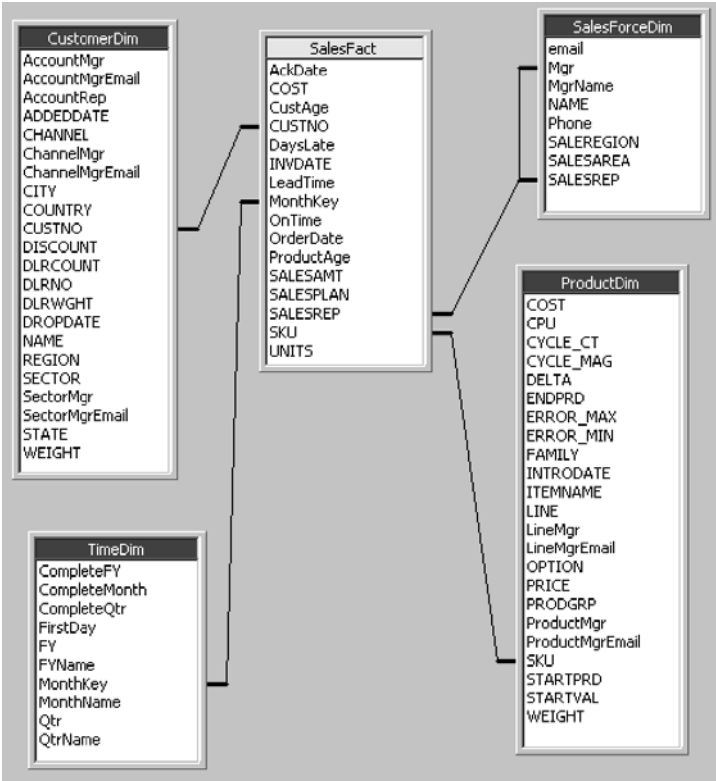
FIGURE 14.1   Star schema example

ments laid out in the schema definition, and to address the issue of data cleanliness. Is the source data accurate and reliable?

- *Map ETL (extraction, transformation, and loading) processes.* We are focused on the ETL for a data mart, not the complete dataware-house process. (See Figure 14.2.)

- *User interfaces built and planned for deployment.* Define the delivery mechanism and build if necessary. Often, an off-the-shelf solution will work, but sometimes a level of customization is required. Internet delivery of BI information almost always requires some level of cus-tomization/integration for a web site.

- Audit processes and results. Verify that the information in the BI sys-tem is accurate and credible through an audit process.

4. *Deployment*. This section will include training, material plans, software deployment plans, infrastructure details for production data servers/con-nectivity, and security plans.

5. *Evolution and maintenance*. Definition of rolling changes and user feed-back into the data mart and cubes definition. Plans for how the informa-tion delivery application is kept up to date, and information load cycles.
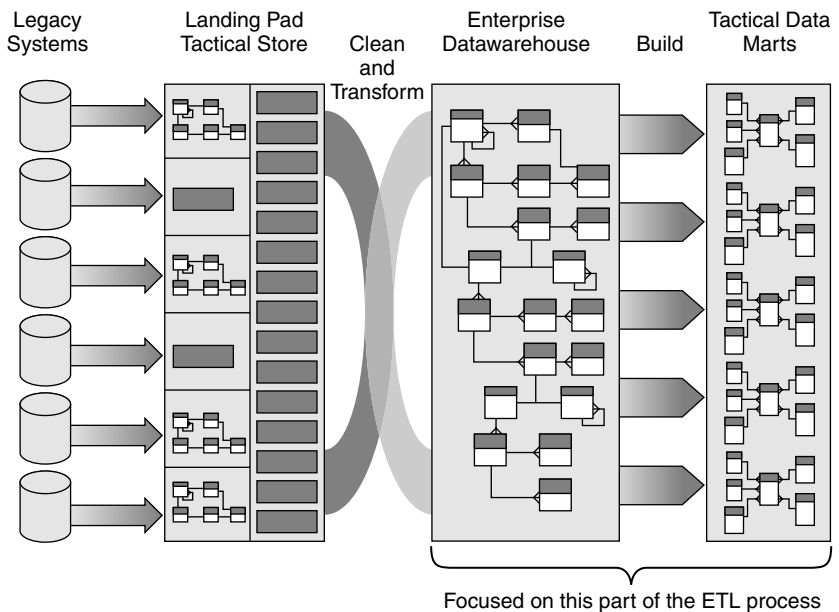
FIGURE 14.2   ETL process

Prototype

Too often, a BI project will wait until a longer-term scheduled delivery date before getting information to the consumers. This is too long to wait for feedback from the information consumers. Prototyping a release to consumers and establishing an effective feedback cycle can ensure the success of a BI project. The effective feedback cycle is critical because only the consumers of the information can tell you if your information is useful.

Iterate

Cycle through the information consumer feedback cycle often. This point can't be emphasized enough. Don't define a single endpoint and expect the BI project to be finished at some point in time. Plan to iterate on the project often. The key parts to drive the iteration are:

- Information-consumer feedback
- Change in business information needs
- Business structure changes

BUSINESS INTELLIGENCE PROJECT DON'TS

There are common pitfalls for BI projects that are carried over from traditional IT projects. The following sections examine some of those key issues in more detail and offer some recommendations about those issues.

Don't Overplan

Scaling to an enterprise-wide BI system can be a large undertaking. Organizations have spent millions of dollars in enterprise-wide systems that never effectively delivered information. As the informational needs definition grows in a BI system, so will its scope. It is always best to serve a departmental level business need at the beginning of a project. Measure its effectiveness, improve on it, then expand to a wider audience of information consumers.

Don't Wait

Any BI system project that has a lag time longer than three months from definition to delivery will have some level of failure. Facets of business change so rapidly today that a system definition more than a few months old will likely

not meet the business needs it was defined to address, likely because the business needs will have changed, but the definition won't reflect that change. Don't wait to get prototypes into the information consumers' hands. A prototype will identify shortcomings and needs much faster than any level of interviews and studies.

Don't Leave It Up to IT

Defining a BI system and "throwing it over the wall" to an IT shop will deliver less than the desired results. IT shops by nature usually do not have the same objectives as the operational side of a business. The organizations in which these groups have closely aligned objectives are rare. Keeping the information consumers and business definition analysts involved in the BI project through the cycle of implementation, review, and prototype will help ensure that the business informational needs are met.

Don't Initially Feature Creep

The tendency at the beginning of any systems project (not isolated to BI) is to continually add more features before delivering. That issue doesn't go away because we are focused on a BI project. Resist the urge to add more features prior to getting a prototype into information consumers' hands. The BI system should be an evolving system in which new features and changes are planned for and done often. This iterative cycle can help alleviate the desire to add features to an initial release.

DRAFTING AND EXECUTING THE PROJECT PLAN

If you develop a project plan, and don't actively use it, you are wasting your time. Develop a project plan and actively monitor project activities, one hour planned is one hour saved during project execution. In other words, it is essential to make sure you plan and coordinate activities—it's a great investment. A plan should be an active part of your documentation and be developed and maintained by every participant in the process. When a secretary is updating the plan based on your "Status e-mail" and then the e-mail is forwarded to the project attendees, the plan is not going to help much. If you have weekly status meetings in which all project attendees participate and contribute with their input, and then create a new revision based on the goals and objectives set during your meeting, you have a much greater chance of succeeding with your project. Make every participant update their status, and adjust their time

estimates after it has been revised by the project steering committee. This creates ownership over the process as well as for the contribution to the project tasks.

When creating your project plan, make sure you make room for planning of the project. When inserting tasks like status meetings, kick-off meetings, and vendor demonstrations, make sure to also make room for *planning* these meetings. If you or your project members are assigned a task, it is important that you have time to plan it out. It is important that the person assigned a task doesn't feel that preparing for it is something expected by him or her on extra hours or in addition to everything else employees have to do. This will kill the spirit on productivity of all of his or her assigned tasks. If planning becomes as natural as completing a task, you have come a long way.

The planning process is also an excellent opportunity to make sure all corporate initiatives are coordinated and avoid overlapping projects. Depending on who's developing the project plan, it usually (read: always) requires approval from a level above. If it's a larger project, maybe the Chief Technology Officer (CTO) is involved in developing the high-level project plan, and he will then need approval from another person above him. If the VP of marketing has initiated a CRM datawarehouse (DW) project while the CTO's project is an integration between a customer relationship management (CRM) system and the general ledger (GL), they should sit down together and make sure their projects don't overlap and find out how they can obtain joined benefits. In larger corporations, there are always projects that collide or are partially covering the same material. Creating a small database (DB) application over all the existing and historical DW projects can help avoid future project collisions and duplication of work.

SAMPLE PROJECT PLAN

We have provided a sample project plan (see Appendix D) in which we have outlined the most important steps in your overall project plan. In addition, you will need to develop detailed project plans for every important task of your high-level plan. As a rule of thumb, you should always develop a detailed plan for every task that lasts for more than three days, as well as other tasks of shorter duration but of crucial importance of the overall success rate for your project. The plan is divided into the following tasks:

- *Business research.* Gather information about your business and the problems that this project is trying to address. For converting data to information and information to knowledge, you need to know as much as possible about your company's business processes.

- *Architecture selection.* Rather early in the project, it is important to get an idea of the architecture on which you will base your solution: Do we need a datawarehouse or data marts, do we have remote users, where are the data located, and how can they be collected?

- *Project scoping.* In order to get acceptance from senior management to go forward with your project, you need to be able to present them with the realm of your project. Are you going to make a new space shuttle or are you trying to figure out $2 \times 2 = ?$ To sign on and become supporters of your team, you need to inform them about what it is that you are trying to build, and most importantly, how much is it going to cost?

- *Project planning.* When management has bought into your idea, there is no point of return. From now on the most important thing you can do is to execute and do your best. The project plan will lay out the available resources of your team, teach you where you have gaps in the knowledge of your team members, which will help you plan for hiring external resources. You need to come up with a deadline and important milestones. When the project plan is complete, you have every major task assigned to a project attendee and you should be able to get in front of the team, present the solution to them, make them believe in it, and become committed to reaching these goals.

- *Materialize plan.* From here to the finish line, you only have to materialize what's being set forth in the project plan. If you follow the map, you know it will get you through all the hurdles.

- *Technology definition and technology selection.* Two important issues have been left out until after the project plan has been finished: selecting your tools, and selecting your team. It is important that the team be part of the tool selection process. Together with end users, the team should exercise the end-user tools used in the data presentation. More technical tools, like whether you should standardize on SQL or Oracle or DB2, should be left as a decision of the project team internally.

- *Execute!* Follow the project plan and regularly conduct project status meetings. You have to trust what you have done so far. Adjust to major deviation, and celebrate every victory. Leave the details alone!

A few things to keep in mind when planning out your project are:

- Why does the organization need a BI application?
- What are the objectives of the BI application?
- What are the major obstacles for succeeding with this project?

- Do the project members possess the right knowledge to succeed?
- Does each member of the project team understand his or her role and responsibilities?
- Do we have access to all the tools, resources, and personnel we will need?
- Is the timing right? Are there other major events that will jeopardize our project or tasks in our project?
- Are the deadlines and milestones realistic?
- Do we know what the users want?

15

DATAWAREHOUSE
OR DATA MART?

Choosing whether to implement a datawarehouse or a data mart is really making a decision about the scope of the information that needs to be addressed. A datawarehouse is typically a longer-term and larger-scoped project than a data mart. A datawarehouse is usually scoped to support management decision making at an enterprise level within an organization, whereas a data mart is typically scoped to help information consumers, at some lower level in an organization, to make tactical decisions. Often, the data populated into a data mart has come from a datawarehouse. (See Figure 15.1.) The main functions for a datawarehouse are to:

- *Collect.* Collecting the data means that it probably comes from disparate sources. Different systems may not have the same database management system (DBMS). Bringing the data together in a single DBMS simplifies the subsequent processes of cleaning and staging the data.

- *Clean.* Cleaning the data really means ensuring its integrity, addressing the issues of accuracy and validity.

- *Stage.* Staging the data means moving it to tables that make the data readily available to create data marts or for other reporting systems to draw directly from the datawarehouse.

A detailed discussion of datawarehousing is beyond the scope of this book. There are books and institutions (Data Warehousing Institute,
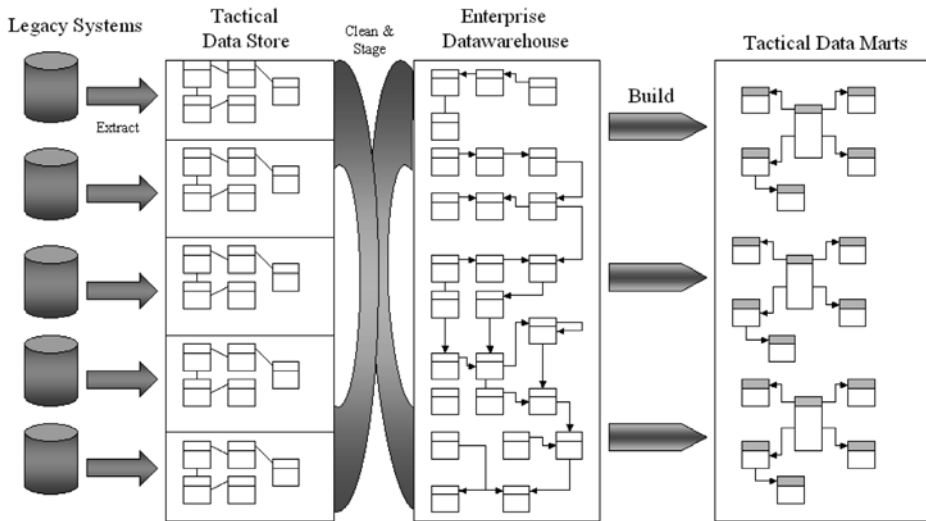
FIGURE 15.1 Enterprise-wide datawarehouse

http://www.dw-institute.com) built solely on the concepts of datawarehouses that can provide much more detailed information.

This discussion implies that we need a datawarehouse before we can have a data mart. That's not true. Using a datawarehouse to build the data marts is a real advantage because the datawarehouse ensures the data is cleaned and staged. If a data mart is built directly from the transactional tables, then it is the data marts' responsibility to ensure the data is staged and cleaned. This is not an overwhelming task, especially when the scope of the informational needs is not huge and isolated to a set of specific business issues. Creating a data mart from the transactional tables is often the right thing to do for the scope of a BI project. (See Figure 15.2.) The factors to consider when choosing a datawarehouse or a data mart are:

- *Scope of informational needs.* If the information is meant to meet enterprise-wide informational needs, then a datawarehouse is probably the right choice. Information needs that are specific to particular business segments, projects, markets, or other smaller scoped informational projects are suited for a data mart.

- *Budget.* Datawarehouse projects for large organizations run to the millions of dollars. Most initial BI projects are not prepared to spend to this
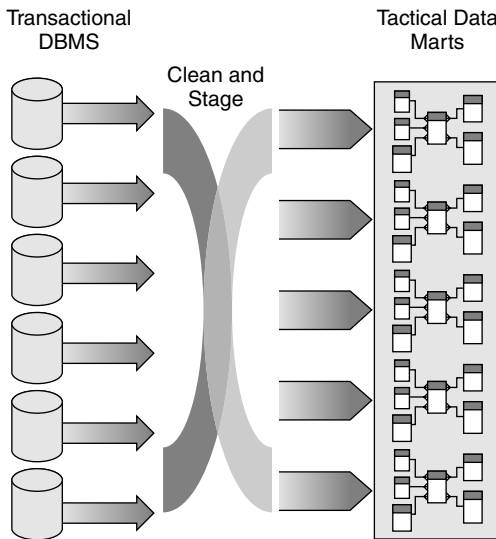
FIGURE 15.2   Creating a data mart from transactional tables

level. A pilot project certainly shouldn't. Budget constraints for new BI projects will push most new projects to use a data mart.

- *Time.* An enterprise-wide datawarehouse can be a project that consumes long periods of time, months at a minimum, and sometimes years. This is even before the data becomes available to the information consumer. Drawing data to include in a data mart should be a small part of the project comprising only a few weeks at most.

- *Source data accessibility.* If the source data is readily accessible, then implementing a data mart directly from the database management system (DBMS) is greatly simplified. However, if obtaining the source data is difficult, then creating a datawarehouse may be the best route for a couple of reasons. First, there will be some investment to get source data not readily available. The investment will come from time spent collecting the data or money spent directly purchasing it. Second, this is the primary goal of a datawarehouse—to collect data from disparate sources and bring it together in a common format.

- *Data quality issues.* Identifying and addressing data quality issues is an important part of BI projects. Determining the level of data quality required for the BI project will help define whether a datawarehouse or a

data mart is the best choice. A datawarehouse is best used when the quality of the data must be very high, mainly because ensuring accuracy and quality is one of the primary objectives for a datawarehouse. If a datawarehouse is not available and a data mart can be built directly from the transactional DBMS, then we should question the level of data quality required for the project. This scenario is suitable when data quality issues don't have a critical level. The data mart (and subsequent cubes) will be built with some data quality issues. Identifying the quality issues to the information consumers is often an adequate precaution for them to still use the information effectively.

16

MULTIDIMENSIONAL MODEL DEFINITION

The multidimensional model definition is a collection of tables, or table definitions, that drive the cube construction. This is the data mart, or tactical, information definition. There are specific objectives, or information needs, that are defined as the requirement for the multidimensional model. This chapter describes the process to define the multidimensional model and the different vectors that definition can follow. We will discuss the model definition process, resulting schemas, and dimension types.

We will draw a fictitious example as we go through the multidimensional model definition process. Some of the features in this fictitious model definition are drawn from real-world examples, while others are manufactured to adequately describe parts of the process. The fictitious company is N&P Industries. They manufacture and sell framing materials to window manufacturers and have been in business for 25 years. N&P is the second largest manufacturer of framing materials in the world. Our example draws from a meeting with the vice president of sales to discuss a new business intelligence (BI) initiative. By discussing the initiative with the vice president of sales, we have already segmented our informational needs to meet the sales organization, and not N&P as a whole. Our model may draw information from other parts of N&P, but the information is meant for consumption by the sales segment of N&P Industries. The process to define a multidimensional model involves the following steps:

- *Defining, or obtaining, objectives.* These objectives are typically driven from a higher set of goals outlined for the organization. They reflect the organization's objectives to be successful and are usually broad in scope. N&P Industries has overall company objectives to *increase profitability* and maintain *market share.*

- *Interviewing decision makers.* This task uncovers the informational needs of those that will interact with the multidimensional model. The results of the interview should be:

  - *A general definition of the tactical information needs.* These are the key short-term objectives that were driven from the organizational objectives.

  - *Key performance indicators (KPIs) to meet those tactical information needs.* KPIs are the *what* part of the question someone asks when they want to get information, "I want to see *dollar sales* for quarter one by state . . ." KPIs are almost always a numeric value and can cover a wide range of needs. In the finance world, KPIs are typically monetary amounts.

  - *The attributes of the KPIs.* The what, when, where that describe the KPIs. These almost always translate directly into the dimensions that will be applied to the multidimensional model. The attributes, or dimensions, are the *how* or *by* part of the question, "I want to see sales for *quarter one* by *state* . . ."

- The interview with the vice president of sales at N&P Industries has boiled the organizational goals down into these specific informational needs of the sales section:

  - *Improve ability to negotiate prices.* Examine pricing and sales history with customers to effectively negotiate. Place the correlation between sales and prices into the hands of the salespeople on the telephone with the customers.

  - *Manage customer accounts to ensure higher levels of profitability.* Profitability is more important than revenue, and examining profit is as simple as subtracting costs to examine the gross profit. Profitability also varies by products since there are cost differences between the products. Another impact on profitability is the manufacturing plant. Examining profit by customer, product, and manufacturing facility is ideal. Another aspect affecting profitability are returns and canceled orders. These need to be tracked by customer, and product.

- *Defining the model.* The two key pieces that must come from the model definition are the *measures* and *dimensions.* The measures are typically numeric values that are derived directly from the KPIs resulting from the

Measures	Dimensions
Prices—directly from the interview	Customer accounts
Gross profit—directly from the interview	Products
Sales (dollars)—needed to calculate profit	Manufacturing plant
Costs (dollars)—needed to calculate profit	History (time)
Units sold—needed to derive average selling price for products	Transactions

FIGURE 16.1   Measures and dimensions

interview. The dimensions are the information attributes, or the *how/by* part of the informational questions. The interview results provide the starting point for the measures and dimensions, yet these are never "finished." The measures and dimensions of a multidimensional model will evolve over time as the business changes and the informational needs change. The measures and dimensions that result from the brief informational needs of N&P Industries can be seen in Figure 16.1.

- *Refining the model.* The model that results from the initial interviews and definition is only the starting point for the multidimensional model. For the model to continue to provide value, it must be refined as the business and informational needs change. The refinement should be viewed as an ongoing process without an endpoint. (See Figure 16.2.)
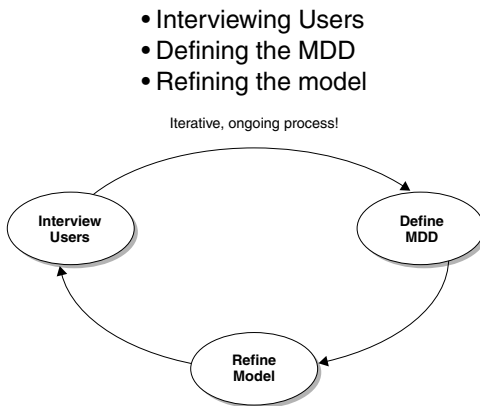
FIGURE 16.2   Refining the model: A continuous process

DEFINING DIMENSION HIERARCHIES AND DIMENSION TYPES

Identifying the dimensions from the interview process is not the end of dimensional definition. Most dimensions have a built-in hierarchical structure that is a logical representation of how information in the dimension is organized. However, some dimensions don't have any depth and serve more as an attribute of the information. An example of a dimension with a hierarchical structure is time (referred to as *history* in our example with N&P Industries). Time is the most common dimension used in multidimensional models because nearly every business model has time associated with it. A simple example time in a hierarchical structure can be seen in Figure 16.3. You can see that Figure 16.3 has some additional annotation defining *levels* and *members*:

- *Levels* are each step in the hierarchy. In Figure 16.3, the time example has the levels of Year, Quarter, and Month. The levels could just as easily been defined as Year and Week where there is a 52/53-week year. Most online analytical processing (OLAP) tools are flexible in allowing the schema builder to design dimensions that fit the informational needs.

- *Members* are the values within the levels of the hierarchy. The example shows that the Year level has only one member (1999), and the Quarter level has four members: Quarter 1, Quarter 2, Quarter 3, and Quarter 4.

The hierarchical structure is not specific to the time dimension. Each dimension in the model will usually have a hierarchical structure. There are some dimensions that serve more as attributes in the multidimensional model. Some examples are size, color, and flavor. These are all examples in which a structure doesn't exist, but are seen more as a flat list that has a lengthy set of values. For example, color would be a list consisting of red, blue, green, purple, brown, black, and so on. This is still considered a dimension, but it has only one level and the members are all part of that level.

There are three distinct types of hierarchies in multidimensional models: (1) *balanced*, (2) *unbalanced*, and (3) *ragged* hierarchies.

In a *balanced* hierarchy, all branches of the hierarchy descend to the same level, and each member's logical parent is the level immediately above
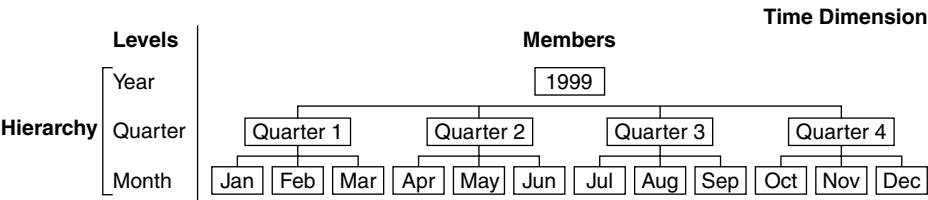
FIGURE 16.3 Time hierarchy example

the member. In our time example, the quarter level members are Quarters 1, 2, 3, and 4, and the dimension's hierarchy is balanced. All branches of the hierarchy terminate at months at the Month level. The parent of each month is either Quarter 1, 2, 3, or 4, and 1999 is the parent of all the quarters. A balanced hierarchy is built from a dimension table or view that will become part of the multidimensional schema (more on schemas later in "Star and Snowflake"). If a dimension's hierarchy has three levels, you will need a table or view that has one column for each of the hierarchy levels. There will be repetition of values in the columns that represent the higher levels in the hierarchy, but the column that represents the leaf level should have unique values. Figure 16.4 represents the table for the simple time hierarchy outlined in Figure 16.3.

Unbalanced hierarchies are characterized by the branches descending to different levels. For example, a Net Income GL Account dimension contains a member for Taxes. Beneath Taxes are the subaccounts of Current Tax, Deferred Tax, and so on. But a sibling for the general category of taxes is Group Contributions, and Group Contributions has no children. So, Taxes has descendent members, but Group Contributions does not. (See Figure 16.5.)

Year	Quarter	Month
1999	Quarter 1	January
1999	Quarter 1	February
1999	Quarter 1	March
1999	Quarter 2	April
1999	Quarter 2	May
1999	Quarter 2	June
1999	Quarter 3	July
1999	Quarter 3	August
1999	Quarter 3	September
1999	Quarter 4	October
1999	Quarter 4	November
1999	Quarter 4	December

FIGURE 16.4   Tables in a simple time hierarchy

Unbalanced Hierarchy
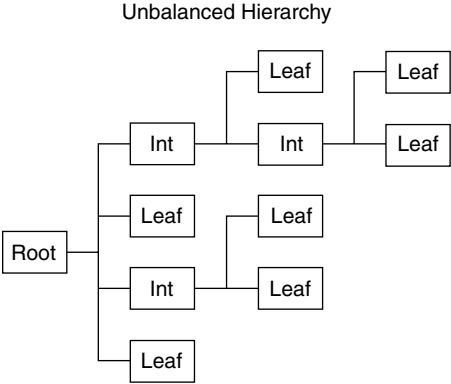
FIGURE 16.5   Unbalanced hierarchies

In a *ragged* hierarchy, the logical parent member of at least one member is not in the level immediately above the member. This can cause branches of the hierarchy to descend to different levels. For example, in a Geography dimension defined with the levels Continent, Country, Province, and City, in that order, the member Asia appears in the top level of the hierarchy, the member India appears in the Country level, the member West Bengal appears in the Province level, and the member Calcutta appears in the bottom level. When we look at the country Sri Lanka, it has no members at the Province level and goes directly to cities. The parent of the city Colombo is the country Sri Lanka which is not in the level immediately above Colombo. It may be impossible for information consumers to distinguish between unbalanced and ragged hierarchies. However, it is important for implementers to distinguish the difference because different techniques can be employed to support these two types of hierarchies. (See Figure 16.6.)
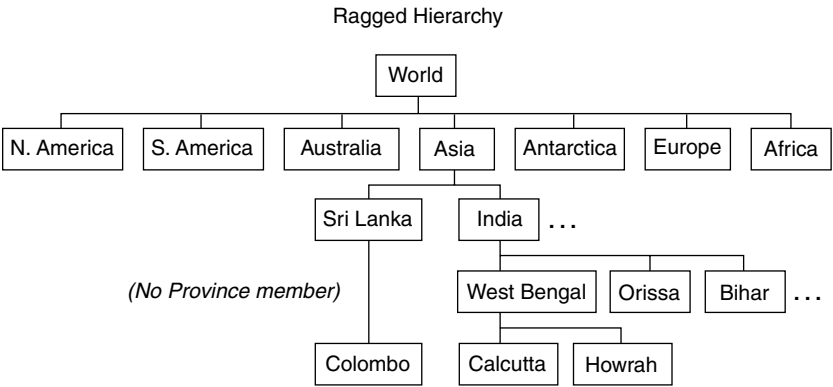
Ragged Hierarchy

FIGURE 16.6   Ragged hierarchy

EmployeeName	EmployeeNumber	ManagerEmployeeNumber
Bob Tilzey	1	3
Dave Young	2	4
Clay Hallmen	3	4
Tammy Patterson	4	4
Michael Erickson	5	3
Pete Raimey	6	5
Elaine Adams	7	5
Paul Whitney	8	3

FIGURE 16.7   Employee table example

Unbalanced and ragged hierarchies can be implemented with a simple table mechanism called *parent–child dimension.* A parent-child dimension table is based on two dimension table columns that together define the lineage relationships among the members of the dimension. One column, called the *member key column,* identifies each member; the other column, called the *parent key column,* identifies the parent of each member. This information is used to create parent–child links, which are then combined into a single-member hierarchy that represents a single hierarchy level.

For example, in Figure 16.7, the column that identifies each member is EmployeeNumber. The column that identifies the parent of each member is ManagerEmployeeNumber. (This column stores the employee number of each employee's manager.) (See also Figure 16.8.)

Organization Chart from Parent–Child Table
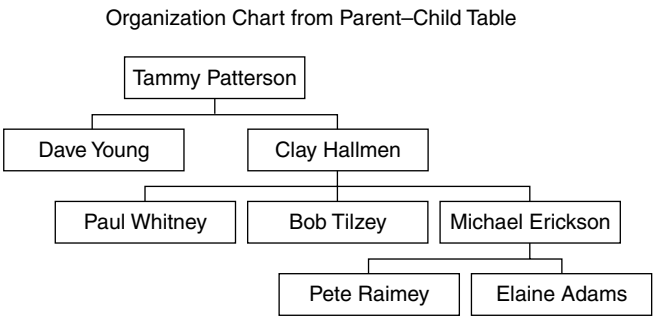
FIGURE 16.8   Employee hierarchy

Dimension	Levels
Products	* Line ** Family *** SKU
Customer Accounts	* Regions ** Reps *** Accounts
Manufacturing Plant	* Plant
Transactions	* Type
Time	* Year ** Quarter *** Month **** Day

FIGURE 16.9   Dimensions with Level Definitions

N&P Industries identified four dimensions from the interview with the vice president of sales. Each of those dimensions has a hierarchy with level definitions. The level definitions were driven by internal organization of those dimensions in current reporting systems. The members of the dimension levels will be driven from the contents of the source data. Figure 16.9 outlines the dimensions and their levels.

All the dimensions from N&P Industries are balanced dimensions and driven by balanced dimension tables. As an example, Figure 16.10 provides a few records from the dimension tables for Products and Manufacturing Plant.

MULTIDIMENSIONAL SCHEMAS: STAR AND SNOWFLAKE

Schemas define the relationships between tables, or views, that comprise a data mart and subsequently get built into an OLAP cube. The tables in a schema are the dimension tables and the fact table. The dimension tables are the result of the information attributes, or the *how/by* part of the informational questions. A fact table contains the specific numeric measurable events defined by the KPIs specified in the objective definition and interview process for defining the multidimensional model. The numeric values are called measures and can range from values for product sales, votes by precinct, room temperatures, absences, product defects, to hours worked. In addition to the specific numeric KPIs, the fact table also contains key values that map each record in the fact table to the

ProductDim Table

	LINE	LineMgr	LineMgrEmail	FAMILY	ProductMgr	ProductMgrEmail	PRODGRP	ITEMNAME	SKU	PRICE	UNITCOST	INTRODATE
1	PWC	Beth Miller	bmiller@galaxy.com	Square Tube Frame	Bill Frost	bfrost@galaxy.com	Alpha IIp	PWC-ST303	PCB1409303	5510.625	1029.3948	3/21/1995
2	PWC	Beth Miller	bmiller@galaxy.com	Square Tube Frame	Bill Frost	bfrost@galaxy.com	Alpha IIp	PWC-ST304	PCB1409304	9009.375	1687.5448	9/26/1994
3	PWC	Beth Miller	bmiller@galaxy.com	Square Tube Frame	Bill Frost	bfrost@galaxy.com	Alpha IIp	PWC-ST305	PCB1409305	6479.375	1255.36	7/7/1995
4	PWC	Beth Miller	bmiller@galaxy.com	Square Tube Frame	Bill Frost	bfrost@galaxy.com	Alpha IIp	PWC-ST306	PCB1409306	6915.625	1384.25	9/4/1995
5	PWC	Beth Miller	bmiller@galaxy.com	Square Tube Frame	Bill Frost	bfrost@galaxy.com	Alpha IIp	PWC-ST307	PCB1409307	12813.125	2600.8648	8/17/1996
6	PWC	Beth Miller	bmiller@galaxy.com	Square Tube Frame	Bill Frost	bfrost@galaxy.com	Alpha IIp	PWC-ST308	PCB1409308	10606.875	1983.6548	2/15/1996
7	PWC	Beth Miller	bmiller@galaxy.com	Square Tube Frame	Bill Frost	bfrost@galaxy.com	Alpha IIp	PWC-ST309	PCB1409309	6241.875	1199.74	8/9/1996
8	PWC	Beth Miller	bmiller@galaxy.com	Square Tube Frame	Bill Frost	bfrost@galaxy.com	Alpha IIp	PWC-ST310	PCB1409310	14840.625	2698.3848	5/16/1996
9	PWC	Beth Miller	bmiller@galaxy.com	Square Tube Frame	Bill Frost	bfrost@galaxy.com	Alpha IIp	PWC-ST311	PCB1409311	5165.625	962.9248	2/14/1997
10	PWC	Beth Miller	bmiller@galaxy.com	Square Tube Frame	Bill Frost	bfrost@galaxy.com	Alpha IIp	PWC-ST312	PCB1409312	7469.375	1476.81	11/20/1996
11	PWC	Beth Miller	bmiller@galaxy.com	Trapezoid Frame	Suzan Wells	swells@galaxy.com	Omega J-	PWC-TR313	PCM1409313	5360.625	976.32	9/27/1994

PlantsDim Table

	PlantName	PlantNo
1	Stuttgart	EUR1
2	Long Beach	NA1
3	Houston	NA2
4	Singapore	PAC1

FIGURE 16.10 Records from dimensional tables example

dimension tables. These are foreign keys in the fact table, and typically a primary key in each of the dimension tables. Fact tables are often large, containing millions of records and sometimes requiring large amounts of disk storage. The fact table is often derived from a core file in the transactional system with additional columns added and populated to act as foreign keys mapping to the dimension tables. It's customary to show the fact table at the center of the schema with the dimension tables radiating from it.

Figure 16.11 shows what is called a *star schema*. A star schema has one fact table and one or more dimension tables. The dimension tables are directly linked to the fact table via their respective keys. A contrast to the star schema is the *snowflake schema*. A snowflake schema is an extension of a star schema. One or more dimensions are defined by multiple tables. In a snowflake schema, only primary dimension tables are joined to the fact table. Additional dimension tables are joined to primary dimension tables. This is particularly useful when using tables from an existing datawarehouse in which the tables already exist. The tables can be used in place and don't need to be combined into single dimension tables. This can provide a savings in maintenance and data mart construction. However, in some systems there is a penalty when using snowflake schemas. For the most part, this has to do with query complexity— snowflake schemas involve an extra "join" for every decomposed dimension— and performance consequences can result depending on the underlying OLAP technology that builds the cubes.

Our example company, N&P Industries, uses a snowflake schema to implement the cube for its analysis. The snowflake section comes from the relationship between the CustomerDim and SalesForceDim tables. The CustomerDim table has information about the specific customer account, which includes

Fact Table with Supporting Dimension Tables – Star Schema
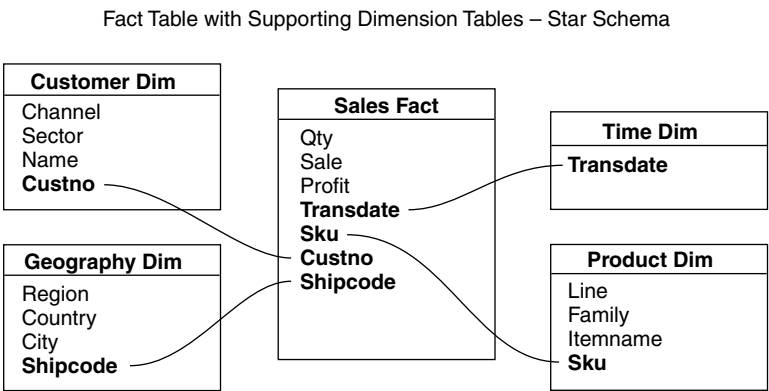
FIGURE 16.11   Star schema

Snowflake Schema for N&P Industries

dbo.CustomerDim
CUSTNO
AccountName
CorpName
CITY
STATE
COUNTRY
REGION
CHANNEL
SECTOR
DISCOUNT
AccountRep
ADDEDDATE
DROPDATE

dbo.SalesForceDim
AccountRep
FirstName
LastName
SalesRegion
SalesArea
Mgr
MgrName
Phone
email

dbo.SalesFact
TransDate
CUSTNO
SKU
SALESAMT
UNITS
COST
REP
OrderDate
AckDate
TransType
PlantNo

dbo.TimeDim
TransDate
FirstDayofWeek

dbo.ProductDim
LINE
LineMgr
LineMgrEmail
FAMILY
ProductMgr
ProductMgrEmail
PRODGRP
ITEMNAME
SKU
PRICE
UNITCOST
INTRODATE

dbo.TransactionDim
transtype
TransOp
TransDescrip

dbo.PlantsDim
PlantName
PlantNo

FIGURE 16.12   Snowflake schema

a reference to the AccountRep. Our dimensional definition called for sales force information to be included in the Customer dimension. That sales force information is linked in with the AccountRep link between the CustomerDim table (primary dimension table) and the SalesForceDim table (secondary dimension table) via the same AccountRep key. (See Figure 16.12.)

The end result of the multidimensional model definition is a collection of tables or table definitions. This may seem like a sterile information technology end result for the dynamic process of defining information needs and objectives. It's not an end, but really the beginning of an evolution of information for an organization. The process to define the multidimensional model is a discipline to really define the informational needs for a business in today's rapidly changing environment.

17

MODEL MAINTENANCE

There are many facets for ensuring the BI project's "evolution," and multidimensional model maintenance is one of the more important facets. It is important because the model definition is the central point where information consumer feedback, business changes, and source data changes come together to be implemented as changes to the data marts and online analytical processing (OLAP) cubes. This chapter will not cover all parts of model maintenance, but touch on key points of periodicity considerations, slowly changing dimensions, and general ledger account maintenance.

PERIODICITY CONSIDERATIONS

Adding data to the BI project data mart should be driven by the organization's business needs, not necessarily when the closing period is reached. This doesn't mean that they can't be the same, but the closing period doesn't have to drive the BI project data updates. Financial organizations typically have a closing period of monthly or weekly. The closing period is usually the "right" time to add a new set of data to the datamart. However, it's not the "only" time. This topic was discussed in more detail in Chapter 6. We'll recap it here by saying that information consumers expect BI information that is up to date and relevant to the task at hand. If the data mart is updated only monthly, there will be a period of time when the information is considered "stale" by the consumers and it simply won't get used.

Updating the data mart with a higher periodicity than the close of the accounting period is the solution to avoiding "stale" information. A weekly

update will satisfy most information appetites, but daily updates are sometimes required. These are "intermediate" updates to the data mart. The periodicity of these intermediate updates should be driven by the business need. Data mart updates that don't fall on the close of an accounting period should also carry the caveat they are unaudited. Then, when the accounting period has closed, follow up with a data mart update of the audited data. This tactic typically requires two extraction, transformation, and loading (ETL) processes: an intermediate ETL process that updates the data mart with the daily or weekly information and another ETL process that removes (or overwrites) the intermediate information with the closed accounting period information.

An alternative tactic, to keep information current, is to use a duplicate data mart and corresponding OLAP cube(s). This duplicate set is used to store the intermediate and unaudited information. Information consumers would need education on the difference between these data sets because the structure would be identical. The duplicate data mart and OLAP cube(s) will have the same dimensions, KPIs, and schema structure. The difference will lie in the data set used to build the data mart. The duplicate data mart will have the unaudited intermediate data updates, while the original data mart and OLAP cube(s) contain the audited closing period data. This tactic can cause some confusion with the information consumers because the dimensions and KPIs will be identical. The only way to tell the data sets apart is with a naming convention for the data mart and OLAP cube(s), and the time dimensions will differ because the duplicate data mart will have more recent time periods.

SLOWLY CHANGING DIMENSIONS

A common issue in maintaining data marts is that dimensions change over time. For the most part these changes are slow—hence the name *slowly changing dimensions.* It is rare to see a dimension change radically overnight. The typical case is that small dimensional changes will be made as we move through time. Examples include:

- Addition of new dimensions members (i.e., a new product is launched)
- Deletion of dimension members (a general ledger account is no longer used)
- Changing of relationships between members within a dimension (a general ledger subaccount is moved to another account)
- Dimension member attributes change (a general ledger account was renamed)

Slowly changing dimensions are a normal issue for all data marts that exist over some period of time. Even though it is normal, slowly changing dimensions pose challenges.

Drawing from our earlier example for N&P Industries, a common change is for a sales rep to move from one region to another. Going forward all of the rep's activity will belong within the recently attached to region. But what do we do with the historical information associated with that sales rep? We have two clear choices, but they also have challenges to overcome:

1. Change the history for this sales rep so that all the historical information follows him into his new region. This will have some issues; probably the most severe is that the numbers within his old sales region will suddenly drop. The manager for that region will probably not be happy to see the sudden decrease in the historical numbers for that region.
2. Leave the historical information intact. All the information will remain associated with the prior sales region, but who do they belong to within that region? What about the transferred sales rep's history? Does it no longer belong to the sales rep since he has changed regions?

These two choices outline the first two types of slowly changing dimensions. Type 1 is "rewriting history," and type 2 is "keeping track of history."

Type 1: Rewriting History

In Type 1 dimensions in which we want to "rewrite history," a property of a member has changed and the change should be retroactively reflected in history as if the new property value is the original value. This is outlined in the first example above. In a simple normalized system, all you have to do is overwrite the values in the row of the appropriate dimension table. This is actually the technique Kimball suggests.[1]

This technique is typically not this simple in a modern datawarehouse. Managing this situation requires complex management of the transactions. The main problem we encounter with Type 1 dimensions lies in the maintenance of the aggregations. If the sales rep moves from Region A to Region B, the appropriate adjustments need to be reflected in all of the regions' aggregations. This can be a formidable task in the datawarehouse environment. The choice of datawarehouse and OLAP tools will impact how easy or difficult this task is. Assessing the techniques in how a tool handles slowly changing dimensions should be a key part of the tools evaluation.

Type 2: Keeping Track Of History

With Type 2 dimensions, we would like to keep a record of the old dimension data. In our example, when the sales rep is moved from region to region, we'd like to keep all of the old transactions in the old region and relate only the new transactions to the new region.

A suggested solution is creating a new record for the sales rep in the Sales_Rep dimension table in this scenario. In using this technique, the main problem we will face is the fragmentation of the data. The sales rep will now have two members associated with the transaction history in the dimension table. This causes fragmentation of the data. When browsing the transaction history, the sales rep will show up with transactions associated with his track record appearing in two places in the hierarchy. This is a challenge that will not be addressed in detail in this book.

MORE HELP IN MAINTAINING DIMENSIONS

This book does not go into the details on how to maintain the datawarehouse for these slowly changing dimension types, however, there are a number of references to get more detailed help in maintaining slowly changing dimensions. One of the best references is a text from Ralph Kimball titled, *The Data Webhouse Toolkit.*[2] This text discusses different techniques to manage type 1 and type 2 slowly changing dimensions and many other issues for managing a datawarehouse.

ENDNOTES

1. Kimball, Ralph, and Richard Merz, *The Data Webhouse Toolkit: Building the Web-Enabled Data Warehouse,* New York: John Wiley & Sons, Inc., 2000.
2. *Ibid.*

18

FINANCIAL
DATA MODELING

Many companies are struggling with data structure problems. Their chart of accounts in the general ledger (GL) system has been maintained by different employees that are inconsistent in the way department codes and natural account numbers are created. In many cases, restructuring the chart of accounts has been driven by efforts trying to ease reporting and mapping problems, but in many cases just made bad worse!

When faced with creating a financial datawarehouse, there are typically two options when it comes to structuring the corporate chart of accounts:

1. Restructure the chart of accounts in the GL system, with a new structured numbering system.
2. Map it and clean it up by using corporate reporting lines (CRLs) in order to make reporting and drill-down easier.

There are pros and cons of both approaches. The first one is the more extensive one, and will have consequences for a lot of supporting systems like sales order processing and payroll systems that are uploading data to the GL system. While the latter method is less comprehensive, it will probably provide a better structure for drill-down since you can map multiple account numbers into one that makes more sense for the end users, but has the drawback that you have to maintain a mapping table with the challenge of missing new account numbers or department series, which will cause rejections or errors in the ETL process.

DATA COLLECTION

The final outcome of your datawarehouse (DW) depends heavily on how good the roll-up structures fit your business rules. In DW terminology this is known as creating parent–child, snowflake, or star schemas. The challenge is dependent on two items:

1. How your data is stored during data collection (as discussed in Chapter 16)
2. How to describe your business using properties for the dimensions stored during your data collection

Let's assume you have the privilege of setting up new, clean data structures for your accounting system. We will try to create a model that will serve the requirements of the GL system and for further DW analysis.

Business Case

You are the vice president (VP) of accounting for a multibillion-dollar broadcasting conglomerate that is in the process of reviewing its account structures, business intelligence (BI) tools, and reporting structures. Its old data collection system and structures are outdated and you don't even want to spend time looking into that one. Since you have the opportunity to rearrange the chart of accounts, we assume that you will take this opportunity to get a tidier, better structured, and more meaningful number series. One challenge you are facing with such an approach is the fact that sooner or later you will have to remap historical data to your new "Number Series Master Piece." The information systems (IS) department usually possesses the knowledge in order to upload all the data to a DB server, create a mapping table, and then renumber the new structure.

You have been told that the company owns 500 stations in 25 markets, one market being major cities or areas of dense population. Very few markets span over more than one state. When a station is sold or moved, it moves out of the hierarchy or into another market. There are five VPs responsible for approximately five markets each. Each station tracks its expenses in 8 to 12 different departments.

The situation described above should be a familiar one to a lot of people. Often when new brushes are hired, management tends to want to sweep out the good with something new. The problem management often ignores is that they don't initiate the process for all business processes involved, and the result is a half-hearted attempt that often ends with failure.

Challenge

You have been asked to put together a numeric structure that will serve the needs of the GL system and other disparate systems as well as financial reporting and consolidations, budgeting, and a datawarehouse for financial analysis. What is your recommendation?

Data Model

There are multiple ways to approach a model like this. One model many financial managers would jump into is to start modeling a new chart of accounts with all dimensions discussed earlier. (See Figure 18.1.)

Figure 18.1 is definitely a viable alternative, but from a data modeling standpoint this approach is not optimized for the purposes you are trying to solve. The major problems with the model are:

- If one station is moved from one market to the other, or responsibility for one market is moved between VPs, you will face challenges when trying to report across markets or VP's even with the most flexible report writer.
- The number of dimensions having to be keyed in for every voucher is way too many!

With many of today's accounting systems we can easily cut the number of digits keyed on every transaction dramatically.

Natural Account	VP	Market	Station	Department
xxxx	xx	xx	xxxx	xx
Creating sample account numbers like these:				
4010	01	80	3300	05
4020	01	80	3300	05
4010	01	80	3400	10
4020	01	80	3400	10
"x" indicates a numeric position of the dimension.				

FIGURE 18.1 Chart of accounts modeling

In order to achieve an optimal dimension design, it is important to distinguish between: What are the *dimensions* of a transaction and what are the *properties* or attributes of the dimension? A dimension is defined as code that is stored with the transaction in order to effectively classify, store, and retrieve data from a database. Properties are used to *classify* or describe the dimensions and categorize them into groups of data that add value for a user of these transactions.

Is there any reason why you need to track both the VP and the market? One VP is responsible for five markets. Thus, there is a very tight relationship between the VP and the market. You will not see many transactions in their GL database where the VP and the market are *always* the same! Since we have this tight relationship, also called a one-to-one relationship between the data, we do not have to store them both. From one market code, you can tell the VP by looking at the transaction. So, we have eliminated one dimension. But wait, isn't the same also true between the station and the market. By looking at the station code, you can figure out which market it belongs to, and thus the VP. This is called *normalization of a data structure.* We are left with the dimensions shown in Figure 18.2. As you can see, we have normalized the data and created a much better structure that will better serve the need for the users of our system:

- When booking entries to the GL, we only have to collect natural account, station, and department code.

- For reporting purposes in our BI system, we can create a tree that describes the relationships between the entities above in a very flexible way.

Figure 18.3 is an example in which the relationship among all the entities with whom our company is conducting business is described using flexible reporting trees. There is no business intelligence without the use of trees. Time

Natural Account	Station	Dep
XXXX	XXXX	XX
Creating sample account numbers like these:		
4010	3300	05
4020	3300	05
4010	3400	10
4020	3400	10

FIGURE 18.2   Normalized data structure

Consolidated	VP	Market	Station	Department
	01	70	2100	01
				02
				03
			2200	01
				02
				03
		80	3300	01
				02
				03
			3400	01
				02
				03
	02	90	6500	01
				02
				03
			6600	01
				02
				03

FIGURE 18.3   Use of flexible reporting trees

periods are the simplest way of describing a tree-relationship. Weeks roll into months, months into quarters, and quarters into years. If we had chosen the solution of storing all the dimensions (VP, market, station, and department) on every transaction as discussed in our first approach to the account number setup in Figure 18.1, it would be comparable to storing the year in a separate dimension, in addition to the month or day.

There are a few other pitfalls you could fall into, like trying to concatenate the station codes and the department codes, or the station codes and the market codes. This will create a more rigid pattern for your reporting since you don't have a unique identifier for your station. (Your station code would then be expressed like 3300?? or 3300-00:3300-99.) The struggle for an optimized and normalized data model is a tug of war between the fact that you do want to minimize the number of dimensions stored on every transaction, and the objective of being able to answer any business problem from one single source of information. However, you can add business intelligence to the reporting process by thinking out of the box and add BI for management by the use of intelligent number series and especially utilizing the power of trees.

Analysis and Business Intelligence

Now, when you have created the most normalized data capture model, and you can fulfill the requests from senior management to create reports by markets and VPs, it is time to start thinking about how you can add additional business value to your data.

From the last minutes of the board meeting, you observed that "VP 3"— Mark Listener—asked a question about whether profit was down depending on the different formats of the stations or listener demographics for his markets. (Radio stations have customers/listeners who provide ratings for the stations, which allows stations to charge more or less for advertising.)

You start investigating some old data sources and come across information that breaks out all the 500 stations into what type of radio station they are classified as. You see groups like Rock Pop, News, Talk Radio, Adult Contemporary, Christian, and Sports. So now is the time to ask yourself: How can I utilize all this great information in order to provide VP 3 with the analyses he requested.

You wish you were aware of this when you decided on your data collection model in your GL earlier. Your datawarehouse is the place to take care of challenges like these. We assume that you already have an upload of your GL data going to a datawarehouse that can produce OLAP databases. All you have to do now is to have your datawarehouse designer upload the information and create a parent–child relationship between the format and your stations. The next time you generate a cube, you will be able to report on your data and slice your data per market per format. (See Figure 18.4.) A little further down in the document, you find that all stations were also classified by age of the listeners:

15–18
19–25
26–35
36–50
51–65
65 and over

All the stations were grouped underneath each age class. All you have to do is to send this information over to your datawarehouse designer again and ask him to create another dimension. You don't have to change the data model in your GL system, because they will all very easily adapt to your new request, due to your flexible design in the first place!

You are the next board meetings hero! Mark Listener is really impressed when you show him your new drill-down feature and is astonished when you provide him with an Excel Pivot table for his own analysis.

This may sound too good to be true, but the technologies are here today to provide you with these great tools.
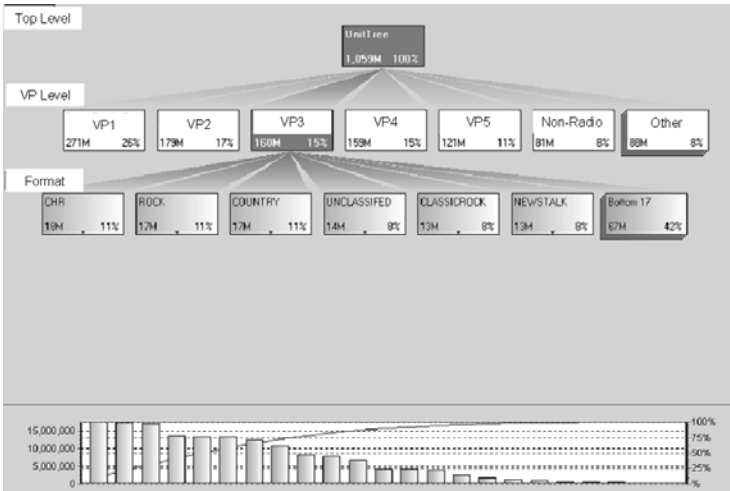
FIGURE 18.4   Multi-dimensional slicing of data example

EXTENDING YOUR FINANCIAL VISION!

So far in this chapter, we have covered the spine of the financial business in-
telligence application: the setup and maintenance of your financial data struc-
tures and dimensions. Now is the time for broadening our financial vision to
see where we proactively can add even more valuable information to your
datawarehouse.

Traditional financial reporting and analysis is usually taking the profit
and loss (P&L) reports at a closer look, while at least smaller and mid-sized
companies are forgetting to make their analysis beyond these limited "State-of-
the-union" reports. It's a known fact (or at least it should be) that your P&L re-
ports can show great net incomes, while your creditors are about to start
moving all your furniture out of the building, due to lack of liquidity to support
the high growth of your expansion. Therefore, it is extremely important to
broaden the corporate understanding of all financial aspects of your business.
We have developed the following model to help you better understand how you
with fairly simple methodologies and a set of tools can increase the under-
standing of your business.

Information can come at a high price if you need to start tracking new
data in order to produce your reports and business analyses. When a dataware-
house is in place, it can be very easy to add data that already exists, enabling
you to create advanced cubes for management to drill around in. Many extrac-
tion, transformation, and loading (ETL) tools (Microsoft DTS Services) enable
you to upload directly from a somewhat structured spreadsheet or a comma
separated values (CSV) file. This is great news for personnel in the financial

department, who can start uploading data to the datawarehouse. Of course, you don't want everybody to start creating tables, links, and indexes and join in a datawarehouse. But if the process is streamlined together with an IS-knowledgeable person, this can really shorten the time from when the data are captured to the moment when that information is available to the information consumer. Also financial departments have the feel for the quality of the data, whereas IS people usually do not. We are decreasing the lead time and increasing data quality—two important measures of a good business analytics system.

These additional information sources are often referred to as subledgers and serve the purpose of providing more analysis for the information consumer. As mentioned earlier in this chapter, the P&L and the balance sheet are the heart of every business. Now we can increase the value of these data by adding subledgers to our datawarehouse. In some cases this could be other OLAP cubes from the enterprise datawarehouse, but they may also come from manual input or simple upload procedures from spreadsheets or CSV files as mentioned earlier.

In Figure 18.5, we have put focus on the balance sheet and the P&L, and expanded this traditional view of business operations with a list of the most important subledgers you could consider adding to your financial datawarehouse. Every business is different and requires different measurements, so you should analyze your industry, look at what competitors are doing, or try to get ideas to
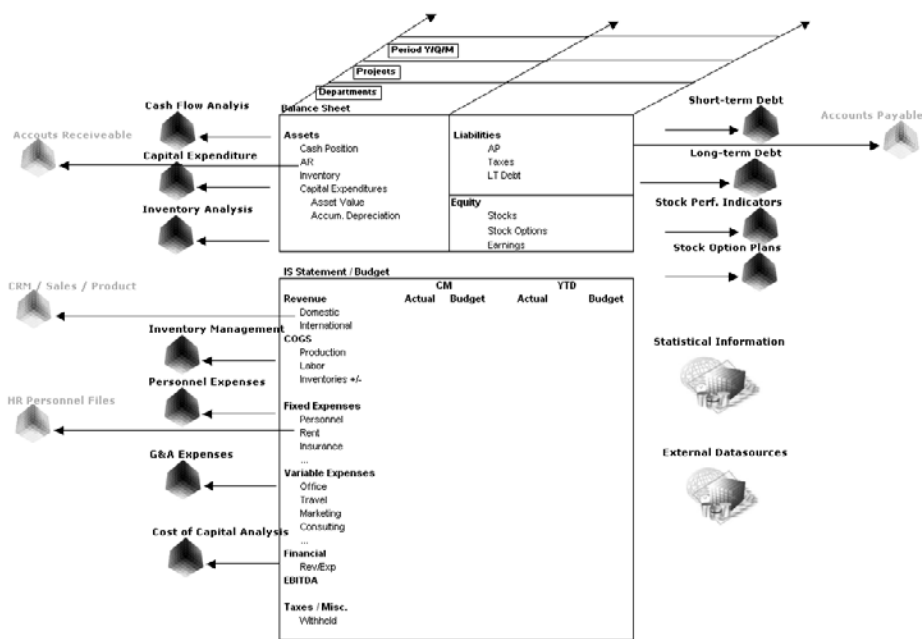

FIGURE 18.5   Common data sources for subledgers

similar industries. Following the figure, you will find an explanation of the most important items to watch out for, as well as a table that describes the most common data sources for these subledgers.

BALANCE SHEET

For many companies, analyzing the balance sheet has been a long forgotten task. Nonetheless, this is a source of vital information about your business.

The balance sheet contains vital information about how healthy your business is running. For financial managers with extensive backgrounds from accounting like CPA degrees or majors in accounting, this may sound too basic to be ignored. However, after conversations with financial managers and personnel across many different companies, it is surprising to see how low the concern with the most important balance sheet items is. Some research of your company, along with a little study about how the P&L statement ties to the balance sheet and how the balance sheet ultimately will affect your cash flow, will be an eye-opener for senior management and members of the board. Analyzing printouts from your GL system may not be enough in detecting the pitfalls of your balance sheet. If you design your datawarehouse correctly, you should be able to perform in-depth analysis on the most important indicators and items to look out for. A good approach is to create subledgers that contain more detailed information about the items you want to follow up more closely. There is always a balance between trying to capture it all or start off with something. Don't let the excitement get to you, not making you satisfied until you have captured it all, but try to start off with some of the more vital elements in your industry, and start adding more as you see the needs arise.

The following sections outline some of the items most companies should pay extra attention to, and give some samples on where and how to track and implement this information into your datawarehouse.

Assets

On the asset side, you can tell how well you are performing in converting accounts receivable into cash. If none of your customers are paying your invoices, it's a matter of time before you run out of cash in order to pay fixed expenses like salaries, insurance, and rent. You will soon run into a cash crisis. Many items are not shown on the income statement like purchases of production machinery or cars. According to generally accepted accounting principles (GAAP) rules, you cannot expense such items over the income statement, but you need to activate them in your balance sheet and depreciate it over its "estimated time of life." Being aware of how hard this is hitting your cash position is of vital character to a healthy business.

Liabilities

Your liability side of the balance sheet is made up of whom you owe money. These are usually classified into short- and long-term departments. Your accounts receivable is a subledger of every vendor you owe money, usually short term. Vital information is found here in planning your cash position. If you collect the right data, you can find information regarding who is your largest vendor. When is your accounts receivable balance due—in 30, 60, or 90 days? Other account numbers usually hold information about withheld taxes due to federal or state requirements.

Another important part is how your long-term investments are doing:

- Are they so long term that almost everybody has forgotten about them?
- Are there potential losses that need to be analyzed or realized?
- How is goodwill affecting your future value analysis?

These items will vary from industry to industry. You need to take a printout of your balance sheet to start analyzing the items affecting your value analysis and the performance of your company.

Equity

This part of your balance sheet contains mostly information about the stocks of your company. How much would you owe your stockholders if you terminated operations today? It is very easy to pull external sources of information about your current stock price. Mix this with internal information about stock options and earnings per share and add inside information about how expected earning will go, and you have a great tool for analyzing stock evaluations and future values of your company.

Subledgers

As mentioned earlier, the information contained in your regular GL system is not enough to accurately analyze and assess how these different items are actually affecting your operations. You need to collect more information! In the past, there were no other ways of getting data into a database than having to call the IS department and wait in line in order to get your "simple" spreadsheet uploaded to your datawarehouse or another database. With the evolution of ETL tools like DataJunction or Microsoft DTS (Data Transformation Services), almost anybody can upload an Excel spreadsheet as a database.

These tools can be found in any flavor and price range. See Chapter 13 for a list of the most common tools in the marketplace. The character of these tools is that they can read from any data source, transform your data, and write the output into any data source. This makes it easy to take data from an Oracle Database and write it to an IBM DB2 database, or an Excel spreadsheet (pre-

formatted and structured, not any mess) and upload it into a SQL server DB. Sounds like a dream? It's not. It's probably your best bet in order to be able to collect information from any data source and write it virtually to anywhere. Make somebody in your organization responsible for checking out the websites of the ETL vendors in the appendix, and start evaluating the tools you will be using. These tools are very helpful when you need to add more data to your financial datawarehouse, for creating subledgers as mentioned above.

Figure 18.6 lists the most common subledgers you will find in a full-sized datawarehouse. Start out by targeting a few of these subsystems and see if you are able to collect, upload, and connect this data to the other information in your DW, making it useful for drill-down and analysis of the other data already contained in your database.

Assets
Cash Flow Analysis: analyze how many more months we can survive!
Typical Acc.# All balance sheet
Source: GL system
Method: Classify all BS accounts in whether they affect cash position or not. Increase in accounts receivable will increase future cash stream, while depreciations are "dead" items in regards to cash flow; cash position changed when the asset was purchased Change in bank acct. balances
Source File: XL-Sheet uploaded to DW tables
Capital Asset Analysis: what's the capital cost of running our business?
Typical Acct.# 1850,1860,1870 (per asset class)
Source: Fixed assets system
Type of Data: Asset class Asset ID number: Depreciation per month Purchase cost and period Time of life Department/Project
Inventory Analysis: how well do we plan our production needs?
Typical Acct.# Stock evaluation
Source: Production system
Type of Data: Product item group Product item number/name Location Items in stock per period Value per item Cost to produce

FIGURE 18.6 Common subledgers example

Liabilities
Short-term Debt–Vendor Info: analyze vendor information by amount and items

Typical Acct.#	2010
Source:	Procurement system
Type of Data:	Vendors
	Vendor groups
	Purchases by dept./proj./account dimension
	Product
	Subclass by product groups
	Purchase amount per vendor per Y/Q/M
Source File:	Direct upload from procurement system

Liabilities (cont.)
Long-term Debt–Investments Analysis: how are investments going?

Typical Acct.#	2200–2400
Source:	Multiple sources—balance sheet
Type of Data:	Number of stocks and stock options
	Value per entity
	Purchase price
	Current price
	Free/Bond
	Use "Other" for nonclassifiable data
Source File:	Excel sheet for upload to DW

Stock Performance Indicators: how are we doing?

Typical Acct.#	3000–3500
Source:	External systems
Type of Data:	Number of stocks
	Value per unit per month
	Stats like EPS/PE
	Group of owners (large)
Source File:	Excel sheet for upload to DW

Employee Stock Options: be on alert about incr./decr. in employee's wealth

Typical Acct.#	3500–3999
Source:	HR system
Type of Data:	Number of stocks per employee
	Value per unit per month
	Vesting
	Department
Source File:	Feed from HR system

FIGURE 18.6   (*continued*)

Liabilities (*continued*)
Statistical information: ratio analysis for every other measure
Typical Acct.#   Statistical accounts
Source: Manual input/maintenance
Type of Data: Number of employees Number of sq.ft. Number of FTE Number of cars By dept./proj./accounting dimension
Source File: Feed from HR system.
External Data Sources: enables advanced statistical comparison with competitors and other external indicator
Typical Acct.#   Statistical accounts
Source: Manual input/maintenance–Upload–XML
Type of Data: Business specific statistics like: Nielsen ratings for broadcasting Market share Production indicators Governmental statistics Stock market indicators By Y/Q/M
Source File: XL–XML feeds to DW

FIGURE 18.6   (*continued*)

19

SURVEY OF DATAWAREHOUSE USERS

Wouldn't it be great to have an insider's perspective on another company's datawarehousing project? What if you could follow the progress of that project as described by the datawarehouse project manager? What if you could learn from his successes and failures?

We have interviewed key personnel who have previously had experience managing or playing an important role in a datawarehouse implementation project in order to uncover some of the most important factors for success as well as the most obvious traps to avoid. We believe other managers' experiences can be great wisdom to bring into your project if you are relatively new to datawarehouse projects. If you have already had experience with datawarehouse implementations, then read the following.

THE SURVEY

The professionals interviewed represent a wide variety of industries that have been exercising different types of datawarehouse implementations that include different approaches, successes, and experiences. A questionnaire was e-mailed to key personnel representing different functions and roles in a datawarehouse implementation. The survey was conducted in the first quarter of 2001.

ANALYSIS OF RESPONSES

Types of Datawarehouse Initiatives

From our selection of respondents, 62% said they had experiences from the enterprise resource planning (ERP) type of datawarehouses, 31% responded customer relationship management (CRM), and only 7% said they were using financial/GL data as their data sources. (See Figure 19.1.)

The majority of our respondents are working on implementing a datawarehouse with the ERP system as their major data source. Chapter 18 stated that traditionally unexplored areas of financial information are starting to grow. Those unexplored areas of financial information include the balance sheet/profit and loss (P&L) statement; add subledgers for the specific areas you want to explore in your datawarehouse. Traditionally, it has been easier to discover patterns, relationships, and hierarchies in customer demographics, product properties, and purchase patterns. However, financial data has been thought of as fairly two-dimensional data: a time dimension going across and an account dimension on the rows. By thinking a little outside the box, there is a great potential for adding valuable information for your organization by adding more dimensions to your traditionally old-fashioned balance sheet and P&L system.

Overall Satisfaction

From our survey, 60% responded that the functionality of their datawarehouse was below expectations. Thirty percent of these said they were not using their datawarehouse any longer. This accounted for 20% of all respondents. (See Figure 19.2.)
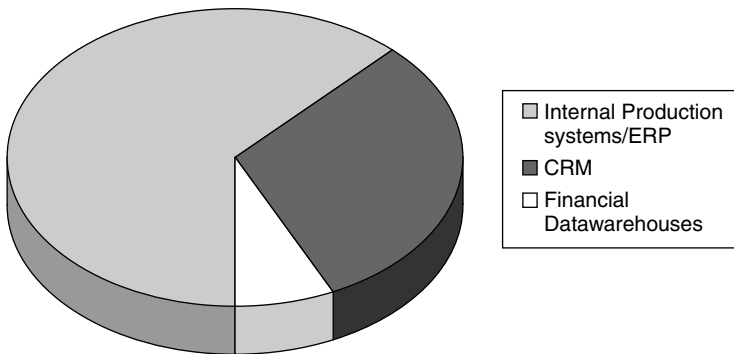

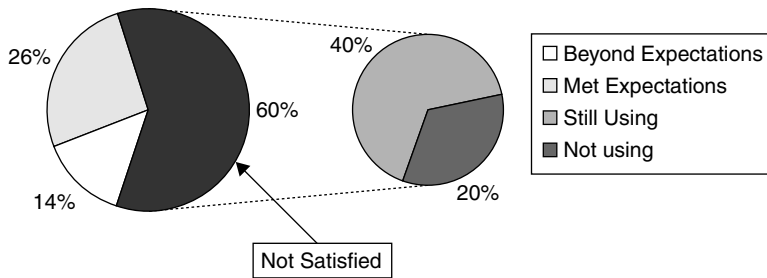**FIGURE 19.1**   Types of warehouse initiatives

FIGURE 19.2 Overall satisfaction

The majority mentioned the "personalization factor" as the single most important thing they would change for their next datawarehouse process from a question regarding the improvement of their datawarehouse implementation. Communication between the personnel developing the datawarehouse and the end users is critical in order to be in the success section of the pie in Figure 19.2. There might be multiple factors influencing the feedback from our respondents.

We didn't ask how dissatisfied they were. Sometimes it's just a thin line between being dissatisfied and being satisfied. Forty percent of the dissatisfied group were actually still using it. This means that 80% of our respondents were still using their internal datawarehouse, which is a good indication of the over-all degree of satisfaction. This only supports the fact that you should continu-ously improve the functionality of the datawarehouse. You need to keep up the good work in order to keep the end users happy, and involve them even more whenever possible.

Knowledgeable Vendors?

Another obvious finding we detected in the material from our respondents was an overall concern about the knowledge of the vendor when giving demonstra-tions of their products. (See Figure 19.3.)

Fifty-eight percent of the respondent were not satisfied with the knowl-edge that vendors had about their products during the sales presentations. Ven-dors can definitely learn a few things: Educate the salespeople or add a product engineer to the sales process. As a participant in a project, this shows that it is important to do homework in order to select the right tools for your end users. It is also important to study the tools you will use as your front end to the datawarehouse and the online analytical processing (OLAP) cubes. Selecting the wrong tools could make you fall into the "Not Met Expectations" in the "Overall Satisfaction" category above.
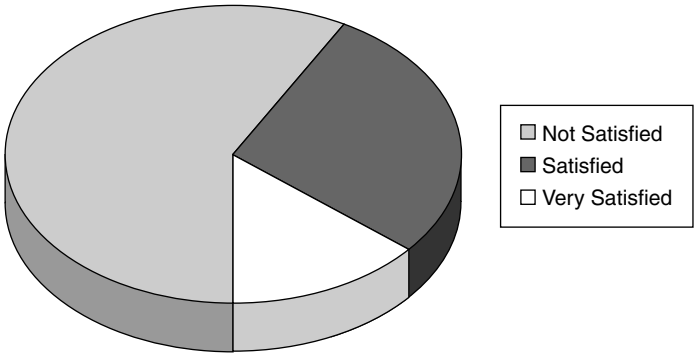
FIGURE 19.3   Knowledgeable vendors?

External Consultants

What role do external consultants play in your search for excellence in your datawarehouse? (See Figure 19.4.) When asked what role the consultants played in the development and project participation of the development of the datawarehouse, 60% said that they played "No Particular Role" or a "Minor Role" in the overall outcome of the project.

Overall Knowledge of External Consultants

The majority of respondents said that they perceived external consultants as "Not Knowledgeable" about the datawarehousing processes. Many of them stated that they would have liked to participate more in the project. (See Figure 19.5.)
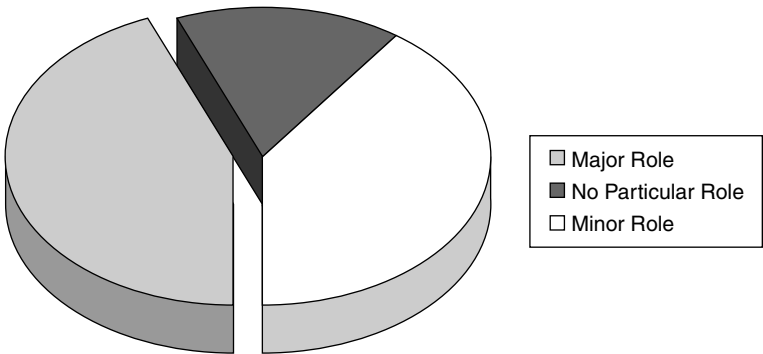
FIGURE 19.4   What role did the consultants play in success?
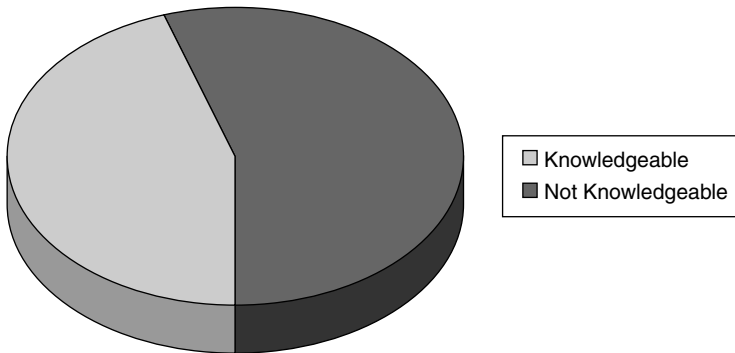
FIGURE 19.5 Overall knowledge of external consultants?

One lesson learned is that you need to make sure that the external consultants are knowledgeable about the type of datawarehouse that you will be implementing. Act like the consultant is applying for a full-time job: Ask for a short summary of experience and check his or her references. If they fail, you will hear about it; however, if the consultants do a great job then somebody else seems to be taking the praise!

What Would You Change for Your Next Project?

Another quality measure is the overall experience of the process. If a survey was conducted of triathlons after passing the finishing line, a few of them would probably be willing to do it all over again. However, they all would have an opinion about what they wanted to change next time. (See Figure 19.6.)

Our datawarehouse triathletes responded that 4% would never do it again, but only 5% would not change anything. Thirty-eight percent said that they would not implement any major changes from their last project. Fifty-
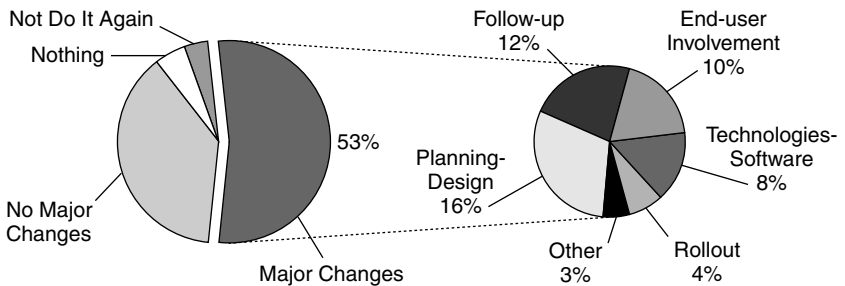
FIGURE 19.6 What would you change in your next project?

three percent would implement major changes to their process. Of these 53%, 16% (30% of the ones who wanted major changes) would change the planning and design process, 12% (22%) would put more efforts into the follow-up process, and 10% (18%) would have more end-user involvement.

Conclusion

On a final note, make sure the project is properly planned and that the end users are included in every aspect of the design. The most important thing is to follow up with the end users after the product is delivered to make sure their satisfaction remains high or to take corrective measures to change their process to their satisfaction!

APPENDICES

APPENDIX A

SAMPLE RFP

REQUEST FOR PROPOSAL

Vendor: ABC Software, Inc.

XYZ CORPORATION

Business Intelligence System

TABLE OF CONTENTS

INSTRUCTIONS

Use the enclosed file containing this document. The document is written in MS Word 2000 and the file is called <filename>.

1. Enter your answers next to each question. Be short and straight to the point wherever possible.
2. You can use attachments to include any additional relevant information you believe is important to disclose.
3. If you have any questions, contact <contact persons name, phone number and e-mail address>.
4. Include the completed requirements document with your proposal.

VENDOR IDENTIFICATION

Software Name	\<software name\>
Vendor Name	ABC Software, Inc.
Authorized Signature	\<contact person's signature\>
Print Name	\<name of contact person\>
Address	\<street\>, \<city\>, \<state\>, \<zip code\>, \<country\>
Telephone Number	\<vendor's phone number\>
Fax Number	\<vendor's fax number\>
E-Mail Address	\<contact person's e-mail address\>
Web Address	\<vendor's web address\>
Date	\<date of RFP completion\>

INTRODUCTION

XYZ Corporation currently utilizes a number of different transactional systems for different accounting and financial reporting purposes. No modern business intelligence systems are currently in place. The company has 900 employees, with 64 cost centers spread across Europe, Asia, and the United States. More than 95% of all revenues and business activities are based on the following lines of business: \<Line of Business A\>, \<Line of Business B\>, and \<Line of Business C\>.

XYZ is seeking a user friendly, robust business intelligence tool to meet our present and anticipated future analytical. Over the last several months the requirements in the enclosed Request for Proposal (RFP) have been developed to evaluate commercial business intelligence solutions. Your system has been selected as a candidate to receive this RFP.

The following is a summary of the process that will be utilized to select the successful bidder:

A. Vendor Response to RFP

Please complete your responses to the following requirements questionnaire, based on the latest version of your software that is currently installed and in use at customer sites, and submit two copies of your responses to:

<contact name>

<phone number>

<fax number>

<e-mail>

<address>

Any questions may also be directed to the person listed above.

The responses to this RFP must be received no later than <date>.

B. Vendor Product Review

From the initial evaluation of the vendor responses, XYZ will select no more than three vendors who will be invited to conduct a product demonstration. Demonstrations will be conducted in early <month>; each company will be given up to four hours to conduct the demonstration. The primary emphasis of the demonstration should be for the vendor to introduce their company and demonstrate the capabilities of their product. The vendor's ability to communicate succinctly will be an important consideration in the selection process.

C. Review of Vendor Installations

An important factor in the evaluation process will be the review of one or more installations of your software. Each vendor will be requested to provide a contact at several sites that are most comparable to XYZ's environment, and are already using your product. We will be looking to discuss with the contact how the software is being used, the effectiveness of the installation, and the relative success of the implementation process.

D. Implementation Proposals

Subsequent to product demonstrations and reference evaluation, XYZ will solicit implementation proposals from the finalists and from other independent implementation vendors. The cost and quality of the implementation support will be critical factors in vendor selection.

E. Contract Development

XYZ reserves the right to reject any and all proposals and alternatives. XYZ intents to negotiate and enter into a contract with the selected vendor. Negotiations must be conducted between the selected vendor firm and XYZ. The content of the RFP and successful vendor proposals will become an integral

part of all contracts, but either may be altered by the final negotiated provisions of the contract.

If, in the opinion of XYZ, a contract, for any reason, cannot be negotiated with the selected vendor firm, XYZ may so notify the vendor. XYZ, within its discretion, may select other vendors from the proposals submitted or reject all proposals and/or may re-issue the RFP.

F. Final Selection

Final selection of a vendor and product will be decided based on the outcome of the above mentioned process. The selection will be made by <date>, and all parties that have made it to this point of the RFP process will be notified of the decision during the week of <date>.

Note: Use the following codes (where it applies) in the vendor reply section:

Y—Yes

N—No

P—Partly

1.0 TECHNICAL ARCHITECTURE

1.1 Platform and Hardware

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
1.1.1	Integration	Tightly integrated with Windows 2000 and MS SQL Server 2000.		
1.1.2	Integration	Integrate with Windows XT		
1.1.3	Information delivery	Use Office 2000 to deliver knowledge management solutions (i.e., digital dashboards, Excel 2000)		
1.1.4	Desktop/ Windows application integration	Facility to export to the following: (Please specify version as well.) ASCII Excel Spreadsheet Word Document ActiveX Object Linking and Embedding (OLE) Hypertext Markup Language (HTML) XML/XBRL		

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
1.1.5	Use of browser technology	Client tools utilizes web browser–based technology (JAVA or ActiveX)		
1.1.6	Spreadsheet access	Can the database be accessed directly from a spreadsheet? Is this interface delivered? What additional software is required on the client?		
1.1.7	Read/Write to Spread-sheets	Can the product read and write to the following spreadsheets? And which versions? Excel Other (e.g., Lotus)		
1.1.8	Cut and paste	Can the product cut and paste data from the following spreadsheets? Excel Other (e.g., Lotus)		
1.1.9	Ability to export data and common standards supported	Does the product provide . . . Export to ASCII files? Export to Excel spreadsheet? Export to Word Document? OLE embedding? ActiveX support?		
1.1.10	Application customization	Does the product allow application development with standard 3rd and 4th generation languages using the following: DLLs OCX Others (please specify)		
1.1.11	Standard application interfaces	Does the product allow application development using the following: Microsoft Visual Studio C/C++ Oracle Designer/Developer Others (please specify)		

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
1.1.12	Required hardware configuration	Can the product run efficiently on Intel (or compatible) processor based PC? What are the minimum requirements? Processor HDD RAM		
1.1.13	Operating system— server	What operating system does the product support? Windows NT Windows 2000 Windows XT Windows CE HP-UNIX Any other? (please specify)		
1.1.14	Operating system— client	What operating system does the product support? Windows NT Windows XT Windows 2000 Windows CE Any other? (please specify)		
1.1.15	TCP/IP	Does the product support a TCP/IP network environment running on ethernet?		
1.1.16	Novell Intranetware	Is the product compatible with Novell Intranetware client?		
1.1.17	Novell Directory Services	Is the product compatible with Novell Directory Services (NDS)?		

1.2 System Architecture and Data Management

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
1.2.1	Queries	Query data in both star and snowflake ROLAP schemas.		
1.2.2	Develop-ment environment	Provide good visual development and debugging environment.		
1.2.3	ASP and COM	Support ASP and COM development.		
1.2.4	Client–server architecture	Does the tool support a client–server architecture?		
1.2.5	Two-tier client server	If two-tier architecture, is the focus of processing and data manipulation on the server or is there a heavy reliance on the client?		
1.2.6	Leverage multi-processor capabilities	Can the architecture leverage multiprocessor capabilities?		
1.2.7	Designation of location of executables	Does the product allow the user to designate where custom applications are to run? (This ability would allow code to run on the server where more processing capacity exists.)		
1.2.8	Three-Tier Client Server	Does the tool support three-tier architecture? Does three tier support a client, application server, and database server layers?		
1.2.9	Dimension relationships	Does the product allow the dimensions (aspects of the business) to be related and data retrievals based on these relationships?		

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
1.2.10	Dynamic modifications to dimensions	Does the product support dynamic modifications to these business dimensions?		
1.2.11	Use of multiple data sources in queries and reports	Tool should make it easy to create queries and reports by providing features such as automatic table linking of tables from multiple databases and files, within one query.		
1.2.12	Data volume, grouping, and sorting	Ability to handle large amounts of data and to group, sort, and set criteria on data fields using large pick lists, free form entry or external file containing criteria. Allows several fields and variables to be included in reports and queries.		

1.3 Data Extraction, Transformation, and Loading

The following list items relate to ETL services. For each item, describe the capabilities of your products. References to capabilities of unreleased versions of your product must be clearly noted.

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
1.3.1	Data extraction	Change data capture for incremental loads		
1.3.2	Data extraction	Transaction events (new, updates, deletes)		
1.3.3	Data extraction	Full refresh		
1.3.4	Data extraction	Replication (continuous and batch)		

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
1.3.5	Data extraction	Compression/Decompression		
1.3.6	Data extraction	Encryption/Decryption		
1.3.7	Data transformation	Integration (generate surrogate keys, map and code keys, master key lookup)		
1.3.8	Data transformation	Changing dimensions (identify changed values and create surrogate keys)		
1.3.9	Data transformation	Referential integrity validation		
1.3.10	Data transformation	Denormalization/Renormalization		
1.3.11	Data transformation	Data type conversion		
1.3.12	Data transformation	Calculation, derivation, and allocation functions (string, date/time arithmetic, conditional, math)		
1.3.13	Data transformation	User extensibility to functions		
1.3.14	Data transformation	Aggregation		
1.3.15	Data transformation	Content auditing		
1.3.16	Data transformation	Lineage auditing		
1.3.17	Data transformation	Null value handling		

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
1.3.18	Data transformation	User exits Pre- and post-step		
1.3.19	Data transformation	Utilization of metadata in ETL processes		
1.3.20	Data transformation	Integration of metadata with Microsoft Metadata Repository		
1.3.21	Data transformation	Enterprise management of multiple sources to multiple data marts		
1.3.22	Data transformation	Support both star schemas and OLAP data structures		
1.3.23	Data loading services	Job definition		
1.3.24	Data loading services	Job scheduling		
1.3.25	Data loading services	Monitoring		
1.3.26	Data loading services	Logging		
1.3.27	Data loading services	Exception handling		
1.3.28	Data loading services	Error handling		
1.3.29	Data loading services	Notification		
1.3.30	Data loading services	Transformation and loading volume capacity		

1.4 Performance

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
1.4.1	Scalability	Ability to scale up. Add servers to reporting architecture.		
1.4.2	Hardware upgrades	Is support for concurrent users linear with hardware upgrades?		

1.5 Other

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
1.5.1	Maximum rows retrieved	Does the product have the ability to restrict the number of rows returned to a query? Is there a maximum number of rows returned?		
1.5.2	Optimization of queries	Does the product provide functions to support optimization of queries and database cubes?		
1.5.3	Read/Write	Read/Write capability to allow forecasting, budgeting, modeling, "what if" analysis.		

2.0 USER INTERFACE

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
2.1	Ease of installation	How long does it take to install the product for the required configuration?		
2.2	Cue cards and wizards	Does the product support cue cards and wizards to help a novice user?		

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
2.3	MS Windows–like interface	Does the product have common MS Windows like interface with the ability to have Window Control, Minimize and Maximize buttons, short cut keys, vertical and horizontal scroll bars, etc.? Use of right click in various screens and menus?		
2.4	Help functions	Does the product have delivered online help functions and are they context sensitive?		
2.5	Pick list	Does the product have pick lists available for each dimension? Can you search for a string? Can you type in the member name?		
2.6	Shared analysis and reports	Can users of the tool share analysis and reports?		
2.7	Views based on user role	Initial views of tools/information based on user role or user preference.		
2.8	Web presentation	Web presentation of data, graphs, and reports similar to nonweb presentation?		

3.0 USER MAINTENANCE AND ACCESS

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
3.1	Security	Provide security based on reports and data content.		
3.2	Security	Integrated security with NT domain and IIS authenticated users.		
3.3	Data security	Is data security enforced in criteria selection at cube, dimension, and cell levels?		

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
3.4	Multiple users	Can multiple users run the same report at the same time? If so, can you control this? Please explain.		
3.5	Support mult-iple concur-rent users	How many concurrent users can the product support?		
3.6	Available security for the product	What security features does the product provide? Cell security Dimension security Database security Group security User restriction Other security features. If so, provide details.		

4.0 REPORTING FEATURES

4.1 Layout

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
4.1.1	Printing	Ability to print formatted reports from client. (WYSIWYG)		
4.1.2	Standard report types	Does the product support various report types e.g.. free form, tabular, control break, cross-tab, combination of all above?		
4.1.3	Report headers and footers	Does the product have the ability to define headers and footers at page level or group level (control break level)? Different headers, footers and content for the first and last page of the report? Manual page breaks?		

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
4.1.4	Format	Can you control the format of the report? Can you modify titles?		
4.1.5	Formatting flexibility	Flexibility to format (e.g., a profit and loss statement) or modify as desired (page breaks, print ranges, % reduction so data fits on one page, etc.) to produce professional quality reports.		
4.1.6	Graphics and notes	Can you add graphics or notes to a report?		

4.2 Calculations, Formulas, and Report Design

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
4.2.1	Queries and reports	Support ad-hoc queries and report customization.		
4.2.2	Extraction and aggregation	Extract and aggregate data from multiple data sources (such as ERP databases, datawarehouses, data marts, spreadsheets, etc. on single report.)		
4.2.3	Wizards	Provide wizards and intelligent tools for analytical ad-hoc reporting.		
4.2.4	Calculations	Perform statistical calculations and rolled-up totals on aggregated data, including automatic calculation of rolled up % and multi-pass SQL for complex computations.		
4.2.5	Aggregations	Aggregate awareness, which precludes the need for users to understand summarization hierarchies.		

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
4.2.6	Roll-up	Can members of a dimension be easily added, subtracted, multiplied, divided, or not aggregated when calculating the parent level for creating financial statements (such as income statements and cash flow statements)?		
4.2.7	Suppress null values	Can users suppress null rows? Null columns?		
4.2.8	Report wizard	Does the product have a report wizard to aid novice users in report creation?		
4.2.9	Report templates	Does the product have standard report templates that can be used to create reports easily?		
4.2.10	Nested dimensions	Can you stack dimensions on one or both axis of a report?		
4.2.11	Filtering	Can you limit the result set to a specific subset? Can you further display only those parts that meet even stricter criteria, such as a decline of more than 10% from last year? Can you highlight these records?		
4.2.12	Ad-hoc variables	Can temporary variables be created and calculated on the fly?		
4.2.13	Flexibility	Does the end user have the ability to dynamically change data selection criteria?		
4.2.14	User-defined calculations	Does the product allow the user to specify user defined calculations and data types? (i.e., the user should be able to add his own calculations)		
4.2.15	Macros	Tool should have advanced features such as allowing for the use of macros.		
4.2.16	Report conversion	Mechanism to automatically convert pre-built reports from other reporting tools		

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
4.2.17	Single screen for reports and queries	Ability to create queries or reports on one screen.		
4.2.18	Exceptions	Exception reporting		

4.3 Report Management

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
4.3.1	Scheduling	Schedule the creation of new reports and data cubes.		
4.3.2	Notifications	Notify defined users of report generation triggered by events. (E-mail)		
4.3.3	Batch execution	Run reports as a group?		
4.3.4	Query status	Query status persistently available, ability to see estimated run time, see if a job is running, failed with error, or is finished. If there is an error, there should be good error messages that tell you what to do. Ability for user, data center, or tool (based on parameters) to cancel query.		

4.4 Report Viewing and Distribution

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
4.4.1	Report execution	Generate both cached and dynamic reports.		

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
4.4.2	Ad-hoc web reports	Create ad-hoc reports through web client interface.		
4.4.3	Ad-hoc web reports	View ad-hoc reports and data cubes through Web browser interface.		
4.4.4	Export to Excel	Export reports to Microsoft Excel.		
4.4.5	Printing	Can the report be sent directly to the printer? Can you specify any printer? Can you specify margins? Page breaks? Print style? Does it have a WYSIWYG print preview?		
4.4.6	Lotus Notes	Can the product send and receive reports using Lotus Notes?		
4.4.7	Microsoft Exchange	Can the product send and receive reports using MS Exchange?		
4.4.8	Saving	Option to save or not to save. Ability to save queries or reports to common drive for sharing or reuse and include version information (date and time query/report created, modified).		

4.5 Other Features

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
4.5.1	Background execution	Does the product have the ability to execute a query in background?		
4.5.2	Multiple requests	Can a user submit multiple requests at one time? Can you control this? If so, explain.		

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
4.5.3	Expert rules	Ability to create expert rules (triggers creating alerts)		
4.5.4	Alerts	Alerts by e-mail and on-screen		
4.5.5	Report customization	Report "building blocks" or code must be customizable.		
4.5.6	Changing of data selection	Does the product have the ability to dynamically change data selection criteria at runtime?		

5.0 ANALYTICAL FEATURES

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
5.1.1	Viewing and navigating data	Interface supports viewing and navigating multi-dimensional data. Example: pivot table controls with drill-down and -up capabilities.		
5.1.2	Drill down	Drill down into more granular data down to the lowest transaction level.		
5.1.3	Drill anywhere	Can the tool drill down to greater detail, drill up to consolidated data or across dimensional relationships? Do you double click to drill down?		
5.1.4	Charts and graphs	Generate charts and graphs based on aggregations.		
5.1.5	Pivot rotation	Can you reorient the axis of one or more dimensions on the fly? How is this accomplished? Drag and drop? With a button?		

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
5.1.6	Transform to graph	Can you instantly transform a report into a graph of choice?		
5.1.7	Support standard charting features	Does the package allow standard charting features such as the following? Pie charts Bar charts Columns Lines Area graphs Any other (please specify)		
5.1.8	Graph format	Can you control the format of the graph? Can you add graphics or captions? Can you modify legends and titles?		
5.1.9	What-if	"What if" features for modeling, budgeting, and forecasting scenarios.		
5.1.10	Graphical drill-down	Can you drill down graphically within a chart view of a report?		
5.1.11	Statistical functions support	Does the product support standard financial and statistical functions? If so, list the functions supported.		
5.1.12	Data sort and ranking	Can you sort the output of a query (alphabetically and numerically)?		
5.1.13	Filtering	Use of filters to suppress zeros, empty dimension members, etc.		
5.1.14	Exceptions	Exception highlighting		

6.0 EXTENSIONS (ADD-ONS)

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
6.1	Data mining package	Is the access tool integrated with a data mining package such as: SAS Enterprise Miner Microsoft Data Mining SPSS IBM Business Miner Any Other (please specify)		

7.0 ADDITIONAL INFORMATION FROM VENDOR

Item	Feature	Description	Vendor Reply (Y/N/P)	Vendor Comments
7.1.1	Training	Vendor provides various types of training		
7.1.2	Documen-tation	Vendor provides various types of documentation		
7.1.3	Support	Vendor provides various types of tech support		
7.1.4	Consulting	Vendor provides various types of consulting		

8.0 VENDOR'S PROPOSED SOLUTION

8.1 Please describe below, options your company can provide to help meet our business intelligence objectives.

8.2 Please summarize below why XYZ Corp., should select your company's services and products.

9.0 XYZ CORP.: TECHNICAL ARCHITECTURE (SAMPLE INFORMATION)

Desktop Environment

- Windows 2000 on 90% of all PCs
- Laptops run Windows XT
- Microsoft Office 2000 is installed as a standard
- Outlook 2000 is the enterprise standard mail/collaboration client
- Internet Explorer 5.0 is the primary browser

Server/Back Office

- Exchange 5.5 e-mail and collaboration services
- IIS 4 Web Server
- We maintain NT 4 domain (directory), NTLM security, and the Exchange5.5 directory
- We develop and operate custom HTML/ASP applications
- SQL 2000 and Oracle 9i RDBMS
- IBM mainframe with DB2

Current Strategy

- XYZ Corp.'s IT strategy is to use Microsoft as our primary technology provider
- We will be deploying PeopleSoft ERP and Microsoft Great Plains ERP in different divisions in phases over the next several years

APPENDIX B

SOFTWARE CANDIDATE EVALUATION AND RATING SHEET

COMPANY NAME: _____

EVALUATOR NAME: _____ DATE: _____
COMPLIED WITH INSTRUCTIONS?   ____ YES ____ NO (explain on reverse) MAXIMUM   TOTAL
SELECT FOR INTERVIEW?   ____ YES ____ NO   POINTS: SCORE:

Company qualifications, related work last three years, pertinence of experience, related work with our industry, financial stability and capability of company—comments: 10 points
_____
_____
_____

Consulting qualifications, experience, organization of team, expertise and breadth of experience, availability during <date> and <date>, location of nearest office—comments: 10 points
_____
_____
_____

Project approach, thorough, detailed, well-organized proposal; adequate implementation schedule, effective acceptance testing plan, understanding of scope of work and our company's needs, approach to training, quality and quantity of documentation—comments: 10 points
_____
_____
_____

Quality of product, system architecture, flexibility of product, expected performance level and response time estimates—comments: 12 points
_____
_____
_____

Features, response to key requirements, additional features offered, custom design flexibility, printing and other publishing features, web readiness, system security elements—comments: 12 points
_____
_____
_____

Convenience of use, data input screens, on-screen help and warning messages, "user-friendly" features—comments:
                                                                                                          12 points
_____
_____
_____

Fit with current hardware and software, cost of additional hardware, need for modifications, network interface, networking features—comments: 12 points
_____
_____
_____

Maintenance and technical support, types of support, available hours, problem resolution—comments: 7 points
_____
_____
_____

Additional benefits, other extraordinary or unique qualifications—comments: 5 points
_____
_____
_____

Reasonableness of fees, payment schedule, agreement terms—comments: 10 points
_____
_____
_____

TOTAL SCORE: _____

APPENDIX C

SAMPLE LICENSE AGREEMENT

LICENSE AGREEMENT FOR DATA WAREHOUSE SOFTWARE

The License Agreement and Commercial Terms and Conditions herein cover the licensing of the ABC Data Warehouse Software (herein described as "DWS"), associated technical support, and any optional classroom training that the Licensee receives from the Licensor (herein referred to as ABC, Inc.). The licensee, also referred to as Purchaser in this document, is XYZ Corporation of «City», «State».

Licensee:

1. Is granted a nonexclusive license to use DWS only as set forth in this License.

2. May not rent or lease the software but may transfer the software and accompanying materials providing the recipient agrees to the terms of this agreement, and the recipient re-registers the software with ABC, Inc.

3. Acknowledges that DWS, which includes, without limitation, any form of DWS software, user manuals thereto, copies, modifications, enhancements, revisions, or updates thereof, and magnetic media, as defined herein ("DWS") and underlying ideas, algorithms, concepts, procedures, processes, principles, and methods of operation, is confidential and contains trade secrets, and Licensee shall use its best efforts to maintain

confidentiality. Licensee agrees that this obligation shall survive termination of this License.

4. ABC, Inc., grants Licensee the unlimited right to use the software at the number of sites set forth in this contract. Licensee may access the software from a hard disk, over a network, or by any other method Licensee chooses.

5. May not decompile DWS from object code to source code or cross-compile or otherwise adapt or modify DWS, and may not reverse-engineer in any manner DWS.

6. MAY NOT USE, COPY, MODIFY, OR TRANSFER DWS OR USER MANUALS OR ANY COPIES EXCEPT AS EXPRESSLY PROVIDED IN THIS LICENSE.

Copyright

Licensee acknowledges that all intellectual property rights in DWS are owned by ABC, Inc. Without limiting the generality of the preceding sentence, DWS and user manuals are copyrighted and may not be copied except as specifically allowed by this License for backup purposes and to load DWS into the computer as part of executing DWS. All other copies of DWS and related user manuals are in violation of this License. The copyright protection claim also includes all forms and matters of copyrightable material and information now allowed by statutory or common law or hereinafter granted, including, without limitation, material generated from the software programs which are displayed on the screen, such as icons, screen displays, etc. ABC, Inc. indemnifies and holds the Licensee harmless from all claims, losses, damages, etc. arising out of copyright infringement claims by third parties.

Restriction on Use and Disclosure

ABC, Inc., is supplying the Licensee, directly or indirectly, software to be used by the Licensee for financial consolidation and reporting. The Licensee agrees not to duplicate such materials, except as permitted in this License; provided that such restrictions shall not apply with respect to any portion of such materials: (i) which corresponds in substance to that developed by the Licensee and in the Licensee's possession prior to the Licensee's receipt of same from ABC, Inc., (ii) which at the time of disclosure thereof by ABC, Inc. to the Licensee is, or thereafter becomes through no act or failure to act on the Licensee's part, part of the public domain by publication or otherwise, or (iii) which corresponds in substance to that furnished to the Licensee by others as a matter of

right without restriction on disclosure; and provided further that the occurrence of (i), (ii), or (iii) above shall not be construed as granting any rights, either express or implied, under ABC, Inc.'s copyrights which relate to the materials furnished to the Licensee. ABC, Inc. materials provided under this agreement shall not be deemed to be within the foregoing exceptions merely because such materials are embraced by more general materials in the public domain or in the Licensee's possession. In addition, any combination of features shall not be deemed to be within the foregoing exceptions merely because individual features are in the public domain or in the Licensee's possession, but only if the combination itself and its principle of use are in the public domain or in the Licensee's possession. It is expressly understood and agreed that the Licensee shall not disclose any commercial/business information that ABC, Inc. may furnish to the Licensee in writing or otherwise. ABC, Inc. makes no warranties, representations, or guarantees with respect to information supplied hereunder, the use thereof, or the results obtained therefrom.

Disclaimer and Limitation of Liability

ABC, Inc., does not warrant that DWS will meet your requirements or that the operation of DWS will be uninterrupted or error free. ABC, Inc., shall not be liable to the Licensee or to any third party for any direct, indirect, incidental, special, or consequential damages (including business interruption, loss of anticipated profits, loss of good will, or other economic or commercial loss), or for personal injury (including death) resulting from any cause whatsoever arising out of the Licensee's use of or reliance on such deliverables and services, and regardless of the form of action, whether in contract or tort, including breach of any warranty, even if ABC, Inc., has been advised of the possibility of such damages. DWS is licensed in perpetuity "as is" and without any warranty of merchantability or fitness for a particular purpose. Furthermore, neither ABC, Inc., nor any of its employees, subcontractors, consultants, or other assigns, make any warranty, expressed or implied, or assume any liability or responsibility for any use, or the results of such use, of any information, product, process, or other service provided with this software.

Ownership

ABC, Inc., is, and shall remain, the sole and exclusive owner of the magnetic media, DWS, any copies of DWS, and the user manuals. ABC, Inc., has all proprietary rights in DWS and user manuals, including but not limited to all processes, ideas, and data supplied by ABC, Inc. to Licensee. Any back-up copies made by Licensee shall be the sole and exclusive property of ABC, Inc.

Hiring of Employees

While this Agreement is in force, and for a period of six months after the termination of this Agreement for any reason, neither party will employ or offer employment to any person employed by or acting on behalf of the other party, without the prior written permission of the other party.

Termination

IMMEDIATE TERMINATION

Either party may terminate this Agreement forthwith if the other party:

1. assigns its rights or obligations under the Agreement otherwise than stated in this agreement;
2. enters into a composition with its creditors, is declared bankrupt, goes into liquidation, or a receiver, or a receiver and manager, or statutory receiver is appointed in respect of it.

TERMINATION BY EITHER PARTY ON NOTICE

If one party defaults in the performance of any of its obligations under this Agreement and: the default is capable of being remedied, and, within ten (10) working days of notice by the nondefaulting party specifying the default, is not remedied; or the default is not capable of being remedied, the nondefaulting party may immediately terminate, or temporarily suspend the operation of this Agreement, at its sole discretion.

TERMINATION BY THE PURCHASER

In addition to any other remedy, the Purchaser may terminate this Agreement upon written notice if the services do not meet the agreed acceptance criteria in this agreement. If the Purchaser gives notice to ABC, Inc., to terminate this Agreement, the Purchaser may, in addition to terminating this Agreement and any other specific rights:

1. Recover any sums paid to ABC, Inc., for consulting services that have not been performed;
2. Request ABC, Inc., at its own expense, to return all property belonging to the Purchaser;
3. Pursue any additional or alternative remedies provided by law.

Waiver

No delay, neglect, or forbearance by either party in enforcing against the other any provision of this Agreement will be a waiver, or in any way prejudice any right, of that party.

Notices

Any notice given pursuant to this Agreement will be sufficiently given if it is in writing and delivered, or sent by prepaid post or facsimile to the other party at the address as shown below:
The Purchaser's contact name and address:

See Attachment 1 of this document.
ABC Inc.'s contact name and address:
John Doe, 270 N. Canon Drive, #1166, Beverly Hills, CA 90210

Entire Agreement

This License Agreement constitutes the entire understanding between ABC, Inc., and the Licensee and supersedes any previous communications, representations, or agreements by either party, whether oral or written. The terms and conditions contained herein take precedence over the Licensee's additional or different terms and conditions that may be contained in any acknowledgment form, manifest, or other document forwarded by the Licensee to ABC, Inc., to which notice of objection is hereby given. No change of any of the terms or conditions herein shall be valid or binding on either party unless in writing and signed by an authorized representative of each party. If any of the provisions hereof are invalid under any applicable statute or rule of law, such provisions are, to that extent, deemed omitted, but these terms and conditions shall remain otherwise in effect. There are no understandings, agreements, representations, or warranties, expressed or implied, that are not specified herein respecting the subject matter hereof. The agents, employees, distributors, and dealers of ABC, Inc., are not authorized to modify this Agreement nor to make statements, representations, or terms binding on ABC, Inc. Accordingly, any statements, representations, or terms not made or given directly and expressly by ABC Inc., whether oral or written, are not binding on ABC Inc.

General

If any of the provisions or portions thereof of this License are invalid or unenforceable under any applicable statute or rule or law, they are to that extent to be deemed omitted. It is agreed that the laws of the State of California shall govern without reference to the place of execution or performance, and the parties agree to submit to the courts of the State of California. Licensee acknowledges that Licensee has read this license, understands it, and agrees to be bound by its terms and conditions. Licensee also agrees that its use of DWS acknowledges that Licensee agrees to the terms and conditions of this license. Licensee further agrees that the license is a complete and exclusive statement of the agreement between Licensee and ABC Inc., and that it supersedes all proposals or prior agreements, oral or written, and any and all other communications relating to the subject matter of this License.

A P P E N D I X   D

SAMPLE CONFIDENTIALITY AND NONDISCLOSURE AGREEMENT (SALES/DEMO PROCESS)

In connection with discussions on July 21, 1998, and thereafter, between ABC Software, Inc. of 270 S. Peck Drive, Los Angeles, CA 91212, including but not limited to all affiliates of ABC, Inc., and XYZ, LTD of 3110 Johnson Rd., Edison City, FL 33090 regarding information on themselves, their relationships and transactions, and certain of their products, the parties hereto propose to provide to each other, with either in the role as "Donor" or "Recipient" (dependent upon direction of information flow as exchanged from time to time), certain confidential and proprietary information ("Confidential Information" as further defined below) for "Evaluation" or for commercial use. In consideration of Donor's providing this information, and for other good and valuable consideration, the extent and sufficiency of which is acknowledged by Recipient hereby **AGREES** to the terms and conditions as follow with respect to the treatment and use of such confidential information.

1. *Definition:*
   a. Confidential information shall mean materials, information, data, and like items which have been or which may be disclosed, whether orally, in writing or within media used in electronic data processing and programs, including but not limited to information concerning

(1) collection, tabulation, and analysis of data, (2) computer programming methods, designs, specifications, plans, drawings, and similar materials, (3) programs, data bases, inventions, (whether or not eligible for legal protection under patent, trademark or copyright laws), (4) research and development, and (5) work in progress. Additionally so defined is all information on any or all aspects of the business of Donor and its affiliate(s), including without limitation (6) any and all financial statements, (7) business and/or marketing plans or programs, (8) pending or threatened litigation, (9) prospective or existing contractual relations, (10) customer, vendor or affiliate lists or identification(s), and any other information of like nature, value, meaning or significance to categories described as (1) through (10) herein. Confidential information includes the fact of disclosure or evaluation, as well as any tangible or intangible material or information-conveying aspect, whether disclosed in oral or written form or otherwise, when and as involved in any disclosure by Donor or Recipient, and whether or not any such material or aspect is specifically marked, labeled or described as "confidential." This definition shall be subject to exclusions as appear below:

b. Confidential information does not include any information which

   i)   is in the public domain at the time disclosed or communicated to Recipient; or which enters the public domain at the time disclosed or communicated to Recipient; or which enters the public domain through no act, fault, or responsibility of Recipient.;

   ii) is lawfully obtained by the Recipient, with permission to disclose, from a third party who is not, to Recipient's knowledge, subject to any contractual or fiduciary duty not to disclose;

   iii) has been independently derived or formulated by the Recipient without reference to the confidential information given either before or after the effective date of this Agreement;

   iv) the Recipient can demonstrate was lawfully in its possession, free of any duty not to disclose, before the date of disclosure by Donor to the Recipient.

2. *Duty of Confidentiality.* Recipient agrees to receive the confidential information in confidence, to keep the same secret and confidential, to make no additional copies of same without the express written consent of Donor, and not to disclose Donor's confidential information to any party whatsoever, save for its officers, directors, employees and agents as required for evaluation hereunder or for commercial use under a coexistent Agreement, without the prior written approval of Donor. Further, Recipient shall ensure that each of its officers, directors, employees, and

agents, to whom Donor's confidential information might be disclosed under the terms of this Agreement, shall be held to subscribe to all the terms of this Agreement and to agree to be bound by all Agreement terms and conditions.

3. *Duty of Care*. Recipient shall, in its protection of Donor's confidential information from risk of unauthorized disclosure, exercise the same level of care as it does and/or would exercise to safeguard its own confidential and proprietary information, and in no event afford any less such protection and safeguarding as may be regarded as reasonable in view of the premises of this Agreement and Recipient's undertakings pursuant hereto.

4. *No License*. Nothing in this Agreement shall be construed as granting any license or right under any patent, trade secret, copyright or otherwise, nor shall this Agreement impair the right of either party to contest the scope, validity or alleged infringement of any patent, copyright, or trade secret. Nor shall the parties hereto use any part of the confidential information disclosed for any purpose other than furtherance of the relationship between the parties.

5. *Termination of Evaluation; Termination of Commercial Agreement.* If Evaluation is terminated for any reason, or if a commercial agreement under whose operation Donor's information has been divulged shall terminate, and in any event upon the reasonable request of Donor at any time, Recipient shall return or destroy, at Donor's option, all copies of Donor's confidential information and all notes, in documents or on media of any type, related thereto, as may be in Recipient's possession or under the Recipient's direct or indirect control. All such written or media-carried notes shall be subject to such return or destruction, regardless of the authorship of such notes. Recipient shall confirm in writing that it has retained no copies, notes, or other records of the Confidential Information in any medium whatsoever.

6. *Injunctive Relief: Enforcement Costs*. Recipient acknowledges that any breach of this Agreement would cause Donor irreparable harm, which would be difficult if not impossible to quantify in terms of monetary damages. Recipient consents to a granting of immediate injunctive relief to Donor upon any breach of this Agreement, in addition to all other remedies available to Donor at law or in equity. Recipient waives any requirement that Donor post a bond in connection with any application for or order granting injunctive relief. Further, in event of legal process for breach or other cause, the prevailing party shall be entitled to recover its costs, including costs of suit, expenses and any Attorney's fees involved in any action to enforce the terms of this Agreement.

7. *Miscellaneous:*

   A. This Agreement shall be governed by the laws of the State of California. Each party irrevocably submits to the jurisdiction of the courts of the State of California for resolution of any dispute hereunder, without regard to the conflicts of laws principles thereof.

   B. This Agreement may be modified only by a writing executed by both parties. The parties agree that this Agreement represents the entire agreement between the parties as to the subject hereof, and that all prior understanding, agreements, & representations related to the subject matter hereof have been incorporated into this Agreement. This Agreement rescinds and supplants any prior Agreement(s) as to the subject matter hereof.

   C. This Agreement shall become effective on the date first written above, and remain in effect for the duration of the period of Evaluation or until this Agreement has been formally supplanted by any Subscriber, vendor-relationship, financial relationship or similar agreement between the parties which specifically absorbs, rescinds or replaces this Agreement. Failing such specific absorption, rescission or replacement, this Agreement shall remain in full force and effect for the period of five (5) years following the completion of the period of Evaluation or the expiry of any consequent Subscriber or other agreement between the parties.

IN WITNESS WHEREOF, each party agrees to and accepts this Agreement and its terms, by signature of its authorized official below.

_____	_____
ABC, Inc.	XYZ, Ltd.
by: _____	*by:* _____
Title:	*Title:*
Date:	*Date:*

APPENDIX E

SAMPLE SUPPORT PLAN/AGREEMENT

ECONOMY SUPPORT PLAN

(for customers who need two hours or less support per month)

- One registered contact on your staff
- Prioritized incoming call
- Two hours included per month* (balance cannot be accumulated)
- Same day response guaranteed

$175 per month

NORMAL SUPPORT PLAN

- Two registered contacts on your staff
- Prioritized incoming call
- Three hours guaranteed response time
- Limited to five hours per month*

$400 per month

PREMIER SUPPORT PLAN

- Five registered contacts on your staff
- Two assigned support engineers assigned to your account
- Unlimited use
- Two hours guaranteed response time
- Access to the online consultant (WWW)
- A consultant on site by the next day

$1,500 per month

DAY BY DAY–ACCOUNT ATTENTION PLAN

On five working days' notification, a senior support engineer is assigned to your account for special attention to you during days of heavy workloads. For example, you can reserve three days when receiving all monthly reporting files from subsidiaries at the end of your budgeting process, etc.

Immediate Response Time!

1 day	$ 500
3 days	$ 1,250
1 week	$ 2,000

CASE-BY-CASE SUPPORT PLAN

On a per call/case basis, the charge will be $50 for the first 15 minutes and $35 for each additional 15-minute interval.

ADDITIONAL OPTIONS

24 by 7

Special Support Plan: "Premier 24/7" may be purchased for varying time periods. Price: On request.

All services may be purchased for a limited period of time, based on your workloads, monthly reporting, etc.
If time limit is exceeded, calls will be invoiced in 15-minute intervals, at an hourly rate of $140.

APPENDIX F

SAMPLE PROJECT PLAN
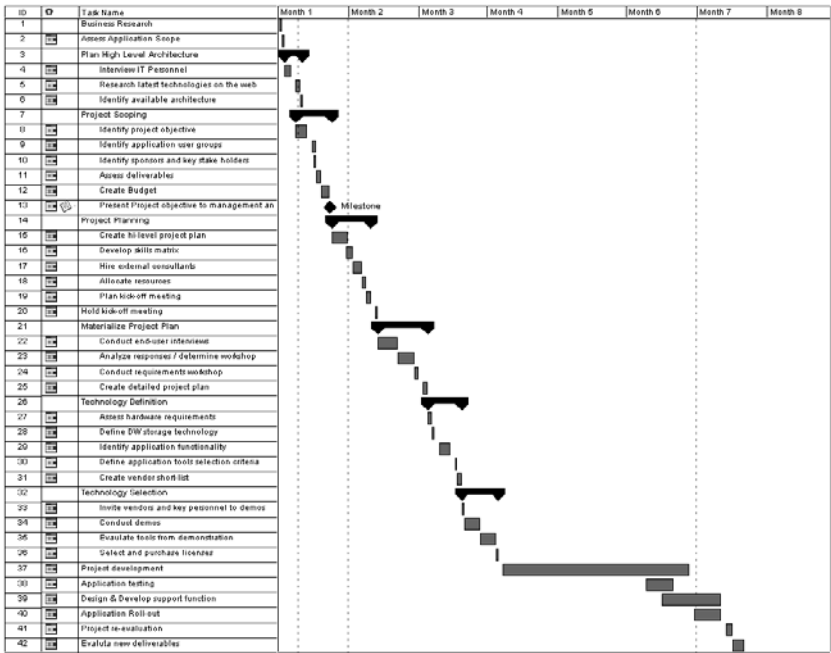
FIGURE F.1 Sample project plan

APPENDIX G

SAMPLE CONSULTING AGREEMENT

CLIENT NAME _____

CLIENT ADDRESS _____

CITY, STATE ZIP CODE_____

This Consulting Agreement (the "Agreement") is entered into by and between **CLIENT** and **CONSULTANT**. The effective date of this agreement is the same date as the software agreement contract is signed.

  Whereas the Client wishes to obtain the consulting services of Consultant to assist Client in connection with the design and implementation of Datawarehouse and Consultant has negotiated the terms of such an agreement with Client and has agreed to the terms as set forth hereafter.

  Now therefore, the parties hereby agree as follows:

1. TERM OF AGREEMENT: The Client hereby hires Consultant and Consultant accepts such contract work for a term for as long as PROFESSIONAL SERVICES is requested by Client, terminating when the Datawarehouse Project is completed, according to project plan, unless sooner terminated as hereinafter provided.

2. SURVIVAL OF AGREEMENT: This Agreement shall not be terminated by a restructuring of the company or of the Consultant. If either of the parties restructures but remains in the business, the contract shall survive.

3. LEGAL REPRESENTATION: Each party acknowledges that they were advised that they were entitled to separate counsel and they have either employed such counsel or voluntarily waived their right to consult with counsel.

237

4. NOTICES: All notices and other communications provided for or permitted hereunder shall be in writing and shall be made by hand delivery, first-class mail, telex, or telecopier, addressed as follows:

   All such notices and communications shall be deemed to have been duly given when delivered by hand, if personally delivered; three (3) business days after deposit in any United States Post Office in the continental United States, postage prepaid, if mailed; when answered back, if telexed; and when receipt is acknowledged, if telecopied.

5. ATTORNEY'S FEES: In the event that a dispute arises with respect to this Agreement, the party prevailing in such dispute shall be entitled to recover all expenses, including, without limitation, reasonable attorneys' fees and expenses, incurred in ascertaining such party's rights or in preparing to enforce, or in enforcing, such party's rights under this Agreement, whether or not it was necessary for such party to institute suit.

6. COMPLETE AGREEMENT OF THE PARTIES: This is the complete agreement of the parties and it supersedes any agreement that has been made prior to this Agreement.

7. ASSIGNMENT: This Agreement is of a personal nature and may not be assigned.

8. BINDING: This Agreement shall be binding to both of the parties hereto.

9. NUMBER AND GENDER: Whenever the singular number is used in this Agreement and when required by the context, the same shall include the plural. The masculine gender shall include the feminine and neuter genders, and the word "person" shall include a corporation, firm, partnership, or other form of association.

10. GOVERNING LAW: The parties hereby expressly acknowledge and agree that this Agreement is entered into in the State of California and, to the extent permitted by law, this Agreement shall be construed and enforced in accordance with the laws of the State of California.

11. FAILURE TO OBJECT NOT A WAIVER: The failure of a party to object to, or to take affirmative action with respect to, any conduct of the other which is in violation of the terms of this Agreement shall not be construed as a waiver of the violation or breach or of any future violation, breach, or wrongful conduct until 180 days since the wrongful act or omission to act has passed.

12. SEVERABILITY: If any term of this Agreement is held by a court of competent jurisdiction to be invalid, void, or unenforceable, the remainder of the provisions of this Agreement shall remain in full force and effect and shall in no way be affected, impaired, or invalidated.

13. FURTHER ASSISTANCE: From time to time each party shall execute and deliver such further instruments and shall take such other action as

any other party may reasonably request in order to discharge and perform their obligations and agreements hereunder and to give effect to the intentions expressed in this Agreement.

14. INCORPORATION BY REFERENCE: All exhibits referred to in this Agreement are incorporated herein in their entirety by such reference.

15. CROSS REFERENCES: All cross-references in this Agreement, unless specifically directed to another agreement or document, refer to provisions in this Agreement, and shall not be deemed to be references to any overall transaction or to any other agreements or documents.

16. MISCELLANEOUS PROVISIONS: The various headings and numbers herein and the grouping of provisions of this Agreement into separate divisions are for the purpose of convenience only and shall not be considered a part hereof. The language in all parts of this Agreement shall in all cases be construed in accordance to its fair meaning as if prepared by all parties to the Agreement and not strictly for or against any of the parties.

17. WORK TO BE PERFORMED: As requested by you, CONSULTANT will develop a Datawarehouse Model using VENDOR SOFTWARE. This Datawarehouse Model will be based upon the existing spreadsheet model system used by CLIENT (Attachment X) and/or additional design specifications agreed upon by CLIENT and CONSULTANT. The Datawarehouse Model created by CONSULTANT must be capable of meeting the specifications agreed upon by CLIENT and CONSULTANT, listed in Attachment X.

18. PERFORMANCE OF DUTIES: Consultant agrees to perform at all times faithfully, industriously, and to the best of their ability, experience, and talents all of the duties that may reasonably be assigned to them hereunder, and shall devote such time to the performance of such duties as may be necessary.

19. FEES & EXPENSES: The fee for the services provided by CONSULTANT to CLIENT will be $X/hour (not to exceed $X for all services provided). This service cost does not include any time for work performed that is not identified in Attachment X. Additionally, this cost does not include travel, lodging, meals or incidentals. These will be billed at actual and/or per diem rates.

20. INDEPENDENT CONTRACTOR: In performing services and duties hereunder, Consultant and any person acting on Consultant's behalf shall do so as independent contractors and are not, and are not to be deemed, employees or agents of Client or any other person acting on behalf of Client. Consultant shall be responsible for meeting any legal requirements imposed on Consultant or any person acting on his behalf as a result of this Agreement, including but not limited to the filing of income tax returns and the payment of taxes; and Consultant agrees to indemnify

Client for the failure to do so, if Client is required to make any such payment otherwise due by Consultant or any such person acting on Consultant's behalf.

21. REMEDY FOR BREACH: Consultant acknowledges that the services to be rendered by Consultant hereunder are of a special, unique, and extraordinary character which gives this Agreement a peculiar value to Client, the loss of which cannot be reasonably or adequately compensated in damages in an action at law, and that a breach by Consultant of this Agreement shall cause Client irreparable injury. Therefore, Consultant expressly acknowledges that this Agreement may be enforced against him by injunction and other equitable remedies, without bond. Such relief shall not be exclusive, but shall be in addition to any other rights or remedies Client may have for such breach.

22. CAUSES FOR TERMINATION: This Agreement shall terminate immediately upon the occurrence of any one of the following events:

    a. The expiration of the term hereof;

    b. The written agreement of the parties;

    c. The occurrence of circumstances that make it impossible for the business of Client to be continued;

    d. The occurrence of circumstances that make it impossible for the business of Consultant to be continued;

    e. Consultant's breach of his duties hereunder, unless waived by Client or cured by Consultant within 30 days after Client's having given written notice thereof to Consultant;

23. COMPENSATION UPON TERMINATION: Unless otherwise mutually agreed in writing by the parties, the termination of this Agreement due to any cause other than that specified in Paragraph 22.4 shall not relieve Client of its obligation to make any payment of money or any delivery of shares or securities which would have been required, or could have been required by Consultant, pursuant to Paragraph 19, if this Agreement had not been so terminated.

IN WITNESS WHEREOF, the parties have executed this Agreement on the date above.

_____   _____
signature   signature

_____   _____
(Printed Name)   (Printed Name)
   For CLIENT   For CONSULTANT

APPENDIX H

VENDOR ADDRESSES

See the Buyer's Guide in Chapter 13 for more information on the products offered by each vendor.

Vendor Name	Web Site	Addresses
Ab Initio	*www.abinitio.com*	2001 Spring Street Lexington, MA 02421 Tel: (781) 301-2000 Fax: (781) 301-2001
Acta Technologies	*www.acta.com*	1667 Plymouth Street Mountain View, CA 94043-1203 Tel: (650) 230-4200 Fax: (650) 230-4201
Applix	*www.applix.com*	112 Turnpike Road Westboro, MA 01581 Toll Free: (800) 8-APPLIX Tel: (508) 870-0300 Fax: (508) 366-2278
Ascential	*www.aboutnewco.com*	50 Washington Street Westboro, MA 01581 Toll Free (800) 966-9875 Tel. (508) 366-3888 Fax. (508) 366-3669

Vendor Name	Web Site	Addresses
Bowstreet	*www.bowstreet.com*	343 Sansome St. 9th Floor, Suite 950 San Francisco, CA 94104 Tel: (415) 262-4300 Fax: (415) 262-4333
Brio	*www.brio.com*	4980 Great America Parkway Santa Clara, CA 95054 Toll Free: (800) 879-2746 Tel: (408) 496-7400 Fax: (408) 496-7420
Business Objects	*www.businessobjects.com*	3030 Orchard Parkway San Jose, CA 95134 Toll Free: (800) 527-0580 Tel: (408) 953-6000 Fax: (408) 953-6001
Centra	*www.centra.com*	430 Bedford Street Lexington, MA 02420 Tel: (781) 861-7000 Fax: (781) 863-7288
Click2learn.com	*www.click2learn.com*	110 110th Avenue NE Suite 700 Bellevue, WA 98004 Toll free: (800) 448-6543 Tel: (425) 462-0501 Fax: (425) 637-1504
Coglin Mill	*www.coglinmill.com*	421 First Ave SW Suite 204 Rochester, MN 55902 Tel: (507) 282-4151 Fax: (507) 282-4727
Cognos	*www.cognos.com*	67 South Bedford Street Burlington, MA, 01803-5164 Toll Free: (800) 426-4667 Tel: (781) 229-6600 Fax: (781) 229-9844

Vendor Name	Web Site	Addresses
Computer Associates	*www.ca.com*	One Computer Associates Plaza Islandia, NY 11749 Toll Free: (800) 225-5224 Tel: (631) 342-6000 Fax: (631) 342-6800
Comshare	*www.comshare.com*	555 Briarwood Circle Ann Arbor, MI 48108 Toll Free: (800) 922-7979 Tel: (734) 994-4800 Fax: (734) 213-2161
Constellar	*www.constellar.com*	3100 Steeles Avenue East, Suite 1100 Markham, Ontario Canada, L3R 8T3 Toll Free: (800) 362-5955
Crystal Decisions (formerly Seagate)	*www.crystaldecisions.com*	895 Emerson Street Palo Alto, CA 94301-2413 Toll Free: (800) 877-2340 Tel: (604) 681-3435 Fax: (604) 681-2934
Data Junction	*www.datajunction.com*	2201 Northland Drive Austin, TX 78756 Tel: (800) 580-4411 or (512) 452-6105 Fax: (512) 459-1309
DataMirror	*www.datamirror.com*	3100 Steeles Avenue East Suite 1100 Markham, Ontario Canada L3R 8T3 Toll-Free: 1 (800) 362-5955 Tel: (905) 415-0310 Fax: (905) 415-0340
Decisionism	*www.decisionism.com*	

Vendor Name	Web Site	Addresses
Docent	*www.docent.com*	2444 Charleston Road Mountain View, CA 94043 Tel: (650) 934-9500 Fax: (650) 962-9411
Evolutionary Technologies	*www.evtech.com*	816 Congress Avenue Suite 1300 Frost Bank Plaza Austin, Texas 78701 Toll Free: (800) 856-8800 Tel: (512) 383-3000 Fax: (512) 383-3300
FRx Corporation	*www.frxsoftware.com*	4700 S. Syracuse Parkway Suite 700 Denver, Colorado 80237 Toll Free: 800-FRx-8733 Tel: (303) 741-8000 Fax: (303) 741-3335
Hummingbird Communications	*www.hummingbird.com*	480 San Antonio Rd Suite 200 Mountain View, CA 94040 Toll Free: 1-877-FLY-HUMM Tel: (416) 496-2200
Hyperion	*www.hyperion.com*	1344 Crossman Avenue Sunnyvale, CA 94089 Tel: (408) 744-9500 Fax: (408) 744-0400
IBM	*www.ibm.com*	1133 Westchester Avenue White Plains, New York 10604 Toll Free: 1-888-SHOP-IBM
ILytix	*www.ilytix.com*	United States: 270 N. Canon Drive, #1166 Beverly Hills, CA 90210 Tel: (800) 281-6351
Informatica	*www.informatica.com*	3350 West Bayshore Rd Palo Alto, CA 94303 Toll Free: 1 (800) 653-3871 Tel: (650) 687-6200 Fax: (650) 687-0040

Vendor Name	Web Site	Addresses
Interwise	*www.interwise.com*	2334 Walsh Avenue Santa Clara, CA 95051 Phone: (408) 748-7800 Fax: (408) 748-7801
Khalix	*www.longview.com*	1974 Sproul Road, Suite 402 Broomall, PA 19008 Tel: (610) 325-3295 Fax: (610) 325-8851
Knowledgeplanet.com	*www.knowledgeplanet.com*	11490 Commerce Park Drive Suite 400 Reston,VA 20191 Toll Free: (800) 646-3008 Ph: (703) 262-6600 Fax: (703) 716-0237
Learn2.com, Inc.	*www.learn2.com*	1311 Mamaroneck Avenue, Suite 210 White Plains, NY 10605 Toll Free (888) 339-8898 Ph: (914) 682-4300 Fax: (914) 682-4440
MacroMedia	*www.macromedia.com*	600 Townsend Street San Francisco, CA 94103 Tel: (415) 252-2000 Fax: (415) 626-0554
Metagon Technologies	*www.metagon.com*	PO Box 2810 Matthews, NC 28106-2810 Tel: (704) 847-2390 Fax: (704) 847-4875
Microsoft	*www.microsoft.com*	One Microsoft Way Redmond, WA 98052-6399 Toll Free: (800) 426-9400 Tel: (425) 882-8080
Microstrategy	*www.microstrategy.com*	861 International Drive McLean, VA 22102 Toll Free: (888) 537-8135 Tel: (703) 848-8600 Fax: (703) 848-8610

Vendor Name	Web Site	Addresses
Oracle	*www.oracle.com*	500 Oracle Parkway Redwood Shores, CA 94065 Toll Free: 1-800-ORACLE-1
Plumtree	*www.plumtree.com*	500 Sansome Street San Francisco, California 94111 Tel: (415) 263-8900 Fax: (415) 263-8991
Pricewaterhouse-Coopers	*www.pwcglobal.com*	1301 Avenue of the Americas New York, New York 10019 Tel: (646) 471-4000 Fax: (646) 394-1301
ProClarity Corporation	*www.proclarity.com*	500 S. 10th Street, Suite 100 Boise, ID 83702 Tel: (208) 344-1630 Fax: (208)343-6128
RWD Technologies	*www.rwd.com*	10480 Little Patuxent Parkway RWD Building, Suite 1200 Columbia, MD 21044-3530 Toll Free: (888) RWD-TECH Tel: (410) 730-4377 Fax: (410) 964-0039
Saba Software	*www.saba.com*	2400 Bridge Parkway Redwood Shores, CA 94065-1166 Tel: (650) 581-2500
SageMaker	*www.sagemaker.com*	1301 N. Elston, Chicago, IL 60622 Toll Free: (888) 412-7177 Tel: (773) 394-6600 Fax: (773) 394-6601
Sagent Technologies	*www.sagent.com*	800 W. El Camino Real, 3rd Floor Mountain View, CA 94040 Tel: (650) 815-3100 Toll Free: (800) 782-7988
SAP	*www.sap.com*	3999 West Chester Pike Newtown Square, PA 19073 Tel: (610) 661-1000 Fax: (610) 355-3106

Vendor Name	Web Site	Addresses
SAS	*www.sas.com*	SAS Campus Drive Cary, NC 27513-2414 Tel: (919) 677-8000 Fax: (919) 677-4444
Seagate Software (now Crystal Decisions)	*www.crystaldecisions.com*	895 Emerson Street Palo Alto, CA 94301-2413 Toll Free: (800) 877-2340 Tel: (604) 681-3435 Fax: (604) 681-2934
Solver	*www.solverusa.com*	270 N. Canon Drive #1166 Beverly Hills, CA 90210 Toll Free: (800) 281-6351 Tel: (310) 444-9633 Fax: (310) 477-7113
SPSS	*www.spss.com*	233 S. Wacker Drive, 11th floor Chicago, IL 60606-6307 Toll Free: (800) 543-2185 Tel: (312) 651-3000 Fax: (800) 841-0064
SRC	*www.srcsoftware.com*	2120 SW Jefferson Street Portland, OR 97201 Toll Free: 1-800-544-3477 Tel: (503) 221-0448 Fax: (503) 223-7922
Sybase	*www.sybase.com*	6475 Christie Ave. Emeryville , CA 94608 Tel. (510) 922-3500
Synex Systems	*www.f9.com*	1444 Alberni St., Fourth Floor Vancouver, British Columbia Canada V6G 2Z4 Toll Free: (800) 663-8663 Fax: (604) 688-1286
Systemfabrik	*www.systemfabrik.com*	Tel. (310) 689-7555 Fax. (310) 689-7515
Taurus Software	*www.taurus.com*	1032 Elwell Court Palo Alto, CA 94303 Tel. (650) 961-1323 Fax. (650) 961-1454

Vendor Name	Web Site	Addresses
Viador	*www.viador.com*	2000 Charleston Road Suite 1000 Mountain View, CA 94043 Tel: (650) 645-2000 Fax: (650) 645-2001 Toll Free 877-VIADOR1
Whitelight Systems	*www.whitelight-tech.com*	
X.HLP	*www.xhlp.com*	610 Lincoln Street Waltham, MA 02451 Ph:(781) 663-7500 Fax:(781) 890-4505

A P P E N D I X I

REFERENCES AND FURTHER READING

PRINT

Inmon, William H. *Building the Data Warehouse*. New York: John Wiley & Sons, Inc., 1996. ISBN: 0-471-14161-5

Kimball, Ralph and Richard Mertz. *The Data Webhouse Toolkit: Building the Web-Enabled Data Warehouse*. New York: John Wiley & Sons, Inc., 2000. ISBN: 0-471-37680-9

Dyche, Jill. *e-Data: Turning Data into Information with Data Warehousing*. Boston: Addison-Wesley, 2000. ISBN: 0-201-65780-5

Reingruber, Michael C. and William W. Gregory. *The Data Modeling Handbook: A Best-Practice Approach to Building Quality Data Models*. New York: John Wiley & Sons, Inc., 1994. ISBN: 0-471-05290-6

INTERNET

Blahunka, Richard. "An Intelligible Approach to Enterprise Business Discovery and Information Deployment." *DM Review* (April 1999). *http://www.dmreview.com/master.cfm?NavID=198&EdID=63*

(Richard Blahunka is a data warehouse architect for Energy Services Inc. He can be reached at *rblahun@entergy.com*)

Gill, Harjinder S. "The Case for Enterprise Business Model Management." *DM Review* (December 2000).
http://www.dmreview.com/master.cfm?NavID=198&EdID=2789

The Data Warehousing Institute. Best Practices and Leadership in Data Warehousing awards.
http://www.dw-institute.com/pr/bestpractices-leadership-main.htm

The Data Warehousing Institute. "How to Build an Enterprise Data Warehouse: Implement Incrementally, Prove the ROI." *What Works?* 11, lesson no. 6.
http://www.dw-institute.com/whatworks11/lessons/lessons6/lessons6.html

GLOSSARY

Access Provider   A company that provides its customers a service whereby they can access the Internet. The user normally connects to the access provider's computer via a modem using a dial up connection.

Accounts Payable   Amounts owed by a business for purchases received and accepted.

Accounts Receivable   Amounts owed to a business by customers for goods and services received and accepted.

Accumulated Depreciation   The total depreciation of a fixed asset from its purchase to the present time.

Ad-Hoc Query   Any query that cannot be determined prior to the moment the query is issued. A query that consists of dynamically constructed SQL, which is usually constructed by desktop-resident query tools.

Ad-Hoc Query Tool   An end-user tool that accepts an English-like or point-and-click request for data and constructs an ad-hoc query to retrieve the desired result.

Administrative Budget   A formal and comprehensive financial plan through which management may control day-to-day business affairs and activities.

Administrative Data   In a datawarehouse, the data that helps a warehouse administrator manage the warehouse. Examples of administrative data are user profiles and order history data.

Aggregate Data   Data that is the result of applying a process to combine data elements. Data that is taken collectively or in summary form.

Aggregator   This is an e-commerce business model in which the web site sells products or services which it does not produce or warehouse. An aggregator creates an environment where multiple providers (sellers) must compete on terms determined by the use.

Alerts   A notification from an event that has exceeded a predefined threshold.

Allocated Cost   Cost of one type that is assigned or charged to costs of other types.

Analysis Of Variances   Analysis and investigation of causes for variances between standard costs and actual costs. A variance is considered favorable

if actual costs are less than standard costs. It is unfavorable if actual costs exceed standard costs.

Annual Budget   A budget prepared for a calendar or fiscal year.

ASCII (American Standard Code for Information Interchange)   An eight-bit code for character representation; includes seven bits plus parity.

ASP (Application Service Provider)   A company that offers access over the Internet to application programs and related services that would otherwise have to be located in other own personal or enterprise computers.

Asset   Anything owned that has monetary value.

Atomic Data   Data elements that represent the lowest level of detail. For example, in a daily sales report, the individual items sold would be atomic data, while rollups such as invoice and summary totals from invoices are aggregate data.

Attribute   A field represented by a column within an object (entity). An object may be a table, view, or report. An attribute is also associated with an SGML (HTML) tag used to further define the usage.

Authorization Rules   Criteria used to determine whether or not an individual, group, or application may access reference data or a process.

B2B (business to business)   Business-to-business commerce conducted over the web.

B2C (business to consumer)   Business-to-consumer commerce conducted over the Internet. It links consumers to commercial entities in one-way networks.

Balance Sheet   A financial statement showing the assets, liabilities, and equity of a business as of a certain date.

Balanced Budget   A budget in which total expenditures equal total revenue. An entity has a budget surplus if expenditures are less than tax revenues. It has a budget deficit if expenditures are greater than tax revenues.

Base Tables   The normalized data structures maintained in the target warehousing database. Also known as the detail data.

Benefit Segmentation   The process of grouping customers into market segments according to the benefits they seek from the product. Refers to their needs and wants only.

Best Practices   A case study considered to be a good example of a business discipline.

Book Value   The current accounting worth of a business or a balance sheet item. For a fixed asset, it equals cost minus accumulated depreciation. For a business, it is the same as equity, net worth, or shareholders' equity.

Braking Mechanism   A software mechanism that prevents users from querying the operational database once transaction loads reach a certain level.

Break-Even Chart   The chart where sales revenue, variable costs, and fixed costs are plotted on the vertical axis while volume, x, is plotted on the horizontal axis. The break-even point is the point where the total sales revenue line intersects the total cost line.

Break-Even Analysis   Analysis that determines the break-even sales which is the level of sales wherein total costs equal total revenues.

Bricks-and-Mortar   Refers to businesses that exist in the real world as opposed to just the cyber world such as bricks-and-mortar retail outlets, bricks-and-mortar warehouses, etc.

Budget   A quantitative plan of activities and programs expressed in terms of the assets, equities, revenues, and expenses that will be involved in carrying out the plans, or in other quantitative terms such as units of product or service. The budget expresses the organizational goals in terms of specific financial and operating objectives.

Budget Control   Budgetary actions carried out according to a budget plan. Through the use of a budget as a standard, an organization ensures that managers are implementing its plans and objectives and their activities are appraised by comparing their actual performance against budgeted performance. Budgets are used as a basis for rewarding or punishing them, or perhaps for modifying future budgets and plans.

Budget Fund   Annual budgets of estimated revenues and expenditures prepared for most governmental funds. The approved budgets of such funds are recorded in budgetary accounts in the accounting system to provide control over revenues and expenditures.

Budgeting Models   Mathematical models that generate a profit-planning budget. The models help planners and budget analysts answer a variety of what-if questions. The resultant calculations provide a basis for choice among alternatives under conditions of uncertainty.

Bulk Data Transfer   A software-based mechanism designed to move large data files. It supports compression, blocking, and buffering to optimize transfer times.

Burden Rate   The percentage rate at which a cost burden is added to particular other costs. The most common burden rates are the various overhead rates.

Business Architecture   One of the four layers of an information systems architecture. A business architecture describes the functions a business performs and the information it uses.

Business Data   Information about people, places, things, business rules, and events, which is used to operate the business. It is not metadata. (Metadata defines and describes business data.)

Business Drivers   The people, information, and tasks that support the fulfillment of a business objective.

Business Intelligence (BI)   Knowledge gained about a business through the use of various hardware/software technologies that enable organizations to turn data into information.

Business Model   A view of the business at any given point in time. The view can be from a process, data, event, or resource perspective, and can be the past, present, or future state of the business.

Business Transaction   A unit of work acted upon by a data capture system to create, modify, or delete business data. Each transaction represents a single valued fact describing a single business event.

Consumer to Business (C2B)   The financial interaction, initiated by a consumer, between a consumer and business.

Cache   Pronounced "cash." The storage of recently visited sites and data that can be accessed from computer memory instead of linking the server each time you return to the site. This speeds the access time, but does not reflect any changes to the site while in the cache. On rapidly changing sites you may need to click the reload button in order to read the most recent changes.

Call Center   The part of an organization that handles inbound/outbound communications with customers.

Capital   The investment in a business, ordinarily equal to equity plus debt.

Capital Expenditure Budget   A budget plan prepared for individual capital expenditure projects. The time span depends upon the project. Capital expenditures to be budgeted include replacement, acquisition, or construction of plants and major equipment.

Capital Rationing   The problem of selecting the mix of acceptable projects that provides the highest overall net present value (NPV), where a company has a limit on the budget for capital spending. The profitability index is used widely in ranking projects competing for limited funds.

CASE Management   The management of information between multiple computer aided software engineering (CASE) encyclopedias, whether the same or different CASE tools.

Cash Accounting   An accounting basis in which revenue, expense, and balance sheet items are recorded when cash is paid or received.

Cash Flow   The increase or decrease in the cash of a business over a particular period of time.

Cash Flow Forecasting   Forecasts of cash flow, including cash collections from customers, investment income, and cash disbursements.

Central Warehouse   A database created from operational extracts that adheres to a single, consistent, enterprise data model to ensure consistency of decision-support data across the corporation. A style of computing in which all the information systems are located and managed from a single physical location.

Classic Datawarehouse Development   The process of building an enterprise business model, creating a system data model, defining and designing a datawarehouse architecture, constructing the physical database, and lastly populating the warehouses database.

Clicks-and-Mortar   A business that has successfully integrated its online existence with its offline, real-world existence. For example, a retail store that allows customers to order products online or purchase products at its store location.

Click-through   The percentage of advertisements a user clicks on or chooses to view.

Client   A software program used to contact and obtain data from a server software program on another computer. Each client program is designed to work with one or more specific kinds of server programs, and each server requires a specific kind of client.

Client/Server   A distributed technology approach in which the processing is divided by function. The server performs shared functions—managing communications, providing database services, etc. The client performs individual user functions—providing customized interfaces, performing screen to screen navigation, offering help functions, etc.

Collection   A set of data that resulted from a DBMS query.

Continuous Budget   An annual budget that continues to the earliest one month or period and adds the most recent one month or period, so that a 12-month or other periodic forecast is always available.

Contribution Margin   A percentage measure of profitability, equal to revenue minus variable or direct costs, divided by revenue. (The term is sometimes used for only the dollar amount of revenue minus variable or direct costs, although the latter is more often called *contribution*).

Contribution Margin (CM) Variance   The difference between actual contribution margin per unit and the budgeted contribution margin per unit, multiplied by the actual number of units sold. If the actual CM is greater than the budgeted CM per unit, a variance is favorable. Otherwise, it is unfavorable. CM variance = (actual CM per unit – budgeted CM per unit) × actual sales.

Contribution Margin Analysis   Also called Cost-Volume-Profit (CVP) analysis that deals with how profits and costs change with a change in volume. More specifically, it looks at the effects on profits of changes in such factors as variable costs, fixed costs, selling prices, volume, and mix of products sold. By studying the relationships of costs, sales, and net income, management is better able to cope with many planning decisions.

Control Concept   A concept that ensures that actions are carried out or implemented according to a plan or goal.

Control Data   Data that guide a process, for example, indicators, flags, counters, and parameters.

Cooperative Processing   A style of computer application processing in which the presentation, business logic, and data management are split among two or more software services that operate on one or more computers. In cooperative processing, individual software programs (services) perform specific functions that are invoked by means of parameterized messages exchanged between them.

Copy Management   The analysis of the business benefit realized by the cost of expenditure on some resource, tool, or application.

Corporate Planning Model   An integrated business planning model in which marketing and production models are linked to the financial model. Corporate planning models are the basic tools for risk analysis and what-if experiments.

Correlation   The degree of relationship between business and economic variables such as cost and volume. Correlation analysis evaluates cause/effect relationships. It looks consistently at how the value of one variable changes when the value of the other is changed. A prediction can be made based on the relationship uncovered. An example is the effect of advertising on sales.

Cost Allocation   The process of assigning or charging one type of cost to other costs.

Cost Benefit Analysis   The analysis of the business benefit realized by the cost of expenditure on some resource, tool, or application development.

Cost Of Goods Sold   The direct costs of producing revenue, burdened by closely associated indirect costs. Also, beginning inventory plus purchases minus ending inventory. Often called cost of sales or cost of revenue.

Cost Pool   A grouping of costs for the purpose of allocation to, or identification with, particular cost centers, products, or services. Common cost pools are various overhead costs, general and administrative costs, and corporate costs.

Cost-Benefit Analysis   An analysis to determine whether the favorable results of an alternative are sufficient to justify the cost of taking that alternative. This analysis is widely used in connection with capital expenditure projects.

Cost-Volume-Profit Analysis   See Contribution Margin Analysis.

Critical Success Factors   Key areas of activity in which favorable results are necessary for a company to reach its goal.

CRM   Customer Relationship Management.

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

Currency Date   The date the data is considered effective. It is also known as the "as of" date or temporal currency.

Current Ratio   The ratio of current assets to current liabilities.

Customer Relationship Management   The idea of establishing relationships with customers on an individual basis, then using that information to treat different customers differently. Customer buying profiles and churn analysis are examples of decision support activities that can affect the success of customer relationships.

Data   Items representing facts, text, graphics, bit-mapped images, sound, analog, or digital live-video segments. Data is the raw material of a system supplied by data producers and is used by information consumers to create information.

Data Access Tools   An end user–oriented tool that allows users to build SQL queries by pointing and clicking on a list of tables and fields in the dataware-house.

Data Analysis and Presentation Tools   Software that provides a logical view of data in a warehouse. Some create simple aliases for table and column names; others create data that identify the contents and location of data in the warehouse.

Data Dictionary   A database about data and database structures. A catalog of all data elements, containing their names, structures, and information about their usage. A central location for metadata. Normally, data dictionaries are designed to store a limited set of available metadata, concentrating on the information relating to the data elements, databases, files, and programs of implemented systems.

Data Directory   A collection of definitions, rules, and advisories of data, designed to be used as a guide or reference with the datawarehouse. The directory includes definitions, examples, relations, functions, and equivalents in other environments.

Data Element   The most elementary unit of data that can be identified and described in a dictionary or repository that cannot be subdivided.

Data Extraction Software   Software that reads one or more sources of data and creates a new image of the data.

Data Flow Diagram   A diagram that shows the normal flow of data between services as well as the flow of data between data stores and services.

Data Loading   The process of populating the datawarehouse. Data loading is provided by DBMS-specific load processes, DBMS insert processes, and independent fastload processes.

Data Management   Controlling, protecting, and facilitating access to data in order to provide information consumers with timely access to the data they need. The functions provided by a database management system.

Data Management Software   Software that converts data into a unified format by taking derived data to create new fields, merging files, summarizing, and filtering data; the process of reading data from operational systems. Data management software is also known as data extraction software.

Data Mapping   The process of assigning a source data element to a target data element.

Data Mart   A subset of the data resource, usually oriented to a specific purpose or major data subject, that may be distributed to support business needs.

Data Mining   A technique using software tools geared for the user who typically does not know exactly what he's searching for, but is looking for particular patterns or trends. Data mining is the process of sifting through large amounts of data to produce data content relationships. It can predict future

trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. This is also known as data surfing.

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

Data Modeling   A method used to define and analyze data requirements needed to support the business functions of an enterprise. These data requirements are recorded as a conceptual data model with associated data definitions. Data modeling defines the relationships between data elements and structures.

Data Owner   The individual responsible for the policy and practice decisions of data. For business data, the individual may be called a business owner of the data.

Data Partitioning   The process of logically and/or physically partitioning data into segments that are more easily maintained or accessed. Current RDBMS systems provide this kind of distribution functionality. Partitioning of data aids in performance and utility processing.

Data Pivot   A process of rotating the view of data.

Data Producer   A software service, organization, or person that provides data for update to a system of record.

Data Propagation   The distribution of data from one or more source datawarehouses to one or more local access databases, according to propagation rules.

Data Replication   The process of copying a portion of a database from one environment to another and keeping the subsequent copies of the data in sync with the original source. Changes made to the original source are propagated to the copies of the data in other environments.

Data Repository   A logical partitioning of data where multiple databases that apply to specific applications or sets of applications reside. For example, several databases that support financial applications could reside in a single financial data repository.

Data Scrubbing   The process of filtering, merging, decoding, and translating source data to create validated data for the datawarehouse.

Data Store   A place where data is stored; data at rest. A generic term that includes databases and flat files.

Data Transfer   The process of moving data from one environment to another environment. An environment may be an application system or operating environment. See Data Transport.

Data Transformation   Creating "information" from data. This includes decoding production data and merging of records from multiple DBMS formats. It is also known as data scrubbing or data cleansing.

Data Transport   The mechanism that moves data from a source to target environment. See Data Transfer.

Data Visualization   Techniques for turning data into information by using the high capacity of the human brain to visually recognize patterns and trends. There are many specialized techniques designed to make particular kinds of visualization easy.

Datawarehouse   An implementation of an informational database used to store sharable data sourced from an operational database of record. It is typically a subject database that allows users to tap into a company's vast store of operational data to track and respond to business trends and facilitate forecasting and planning efforts.

Datawarehouse Architecture   An integrated set of products that enable the extraction and transformation of operational data to be loaded into a database for end-user analysis and reporting.

Datawarehouse Engines   Relational databases (RDBMS) and multidimensional databases (MDBMS). Datawarehouse engines require strong query capabilities, fast load mechanisms, and large storage requirements.

Datawarehouse Infrastructure   A combination of technologies and the interaction of technologies that support a datawarehousing environment.

Datawarehouse Management Tools   Software that extracts and transforms data from operational systems and loads it into the datawarehouse.

Datawarehouse Network   An integrated network of datawarehouses that contain sharable data propagated from a source datawarehouse on the basis of information consumer demand. The warehouses are managed to control data redundancy and to promote effective use of the sharable data.

Database Schema   The logical and physical definition of a database structure.

Days Payables   The amount of payables relative to total material purchased, expressed in "days" typically as 365 times accounts payable divided by annual material purchases.

Days Receivables   The amount of accounts receivable relative to revenue, expressed in "days" typically as 365 times accounts receivable divided by annual revenue.

Decentralized Database   A centralized database that has been partitioned according to a business or end-user defined subject area. Typically, ownership is also moved to the owners of the subject area.

Decentralized Warehouse   A remote data source that users can query/access via a central gateway that provides a logical view of corporate data in terms that users can understand. The gateway parses and distributes queries in real time to remote data sources and returns result sets back to users.

Decision Support Systems (DSS)   Software that supports exception reporting, stop light reporting, standard repository, data analysis, and rule-based analysis. A database created for end-user ad-hoc query processing.

Delta Update   Only the data that was updated between the last extraction or snapshot process and the current execution of the extraction or snapshot.

Denormalized Data Store   A data store that does not comply to one or more of several normal forms. See Normalization.

Dependent Data Mart   Also called an architected data mart. Shares common business rules, semantics, and definitions. A dependent data mart reads meta data from a central meta data repository to define local meta data. This ensures that all components of the architecture are linked via common meta data.

Depreciation   The gradual decline in value of an asset because of use or age; the expense arising therefrom.

Derived Data   Data that is the result of a computational step applied to reference of event data. Derived data is the result either of relating two or more elements of a single transaction (such as an aggregation), or of relating one or more elements of a transaction to an external algorithm or rule.

Design Interface   The mechanism by which customers specify exactly what they need.

Desktop Applications   Query and analysis tools that access the source database or datawarehouse across a network using an appropriate database interface. An application that manages the human interface for data producers and information consumers.

Dimension   A dimension is a structural attribute of a cube that is a list of members, all of which are of a similar type in the user's perception of the data. For example, all months, quarters, years, etc., make up a time dimension; likewise all cities, regions, countries, etc., make up a geography dimension. A dimension acts as an index for identifying values within a multi-dimensional array. If one member of the dimension is selected, then the remaining dimensions in which a range of members (or all members) are selected defines a sub-cube. If all but two dimensions have a single member selected, the remaining two dimensions define a spreadsheet (or a "slice" or a "page"). If all dimensions have a single member selected, then a single cell is defined. Dimensions offer a very concise, intuitive way of organizing and selecting data for retrieval, exploration, and analysis.

Dirty Data   Inconsistent, missing, incomplete, or erroneous data. Source data often contains a high percentage of "dirty" data.

Disbursement   An amount of cash paid out.

Discounted Cash Flow (DCF) Techniques   Methods of selecting and ranking investment proposals, such as the net present value (NPV) and internal rate of return (IRR) methods where time value of money is taken into account.

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

DSS   See Decision Support System.

DWA   Data Warehouse Administrator.

Dynamic data exchange   An industry standard accepted by most horizontal application software for exchanging data among different software programs.

Dynamic Dictionary   A data dictionary that an application program accesses at run time.

Dynamic Queries   Dynamically constructed SQL that is usually constructed by desktop-resident query tools. Queries that are not pre-processed and are prepared and executed at run time.

E-Business   E-business is simply doing business electronically which in most cases means doing business on the Internet. The two main types of e-business are business-to-consumer (B2C) and business-to-business (B2B). A third, and less talked about, type of e-business is consumer-to-consumer (C2C).

E-Commerce   E-commerce deals with using the Internet, digital communications, and IT applications to enable the buying or selling process.

Efficiency Variance   Difference between inputs (materials and labor) that were actually used and inputs that should have been used (i.e., standard quantity of inputs allowed for actual production), multiplied by the standard price per unit.

EIS   Executive Information System.

Electronic data interchange (EDI)   EDI is the computer-to-computer exchange of normal business transactions including payments, information exchange, and purchase order requests. The most basic EDI line consists of a computer-to-computer link. The second level incorporates an application-to-application design where individual companies link a minimum of one of their in- house systems to the EDI interface. The most elaborate version of EDI actually transforms the way business procedures are executed to gain optimal productivity. These involve trend-institutions that evolve into centralized EDI based functions.

End User Data   Data formatted for end-user query processing; data created by end users; data provided by a datawarehouse.

Enterprise   A complete business consisting of functions, divisions, or other components used to accomplish specific objectives and defined goals.

Enterprise Data   Data that is defined for use across a corporate environment.

Enterprise Datawarehouse   An enterprise datawarehouse is a centralized warehouse that services the entire enterprise.

Enterprise Modeling   The development of a common consistent view and understanding of data elements and their relationships across the enterprise.

Enterprise Portal   The enterprise portal offers a web-like solution to the problem of distributing business information, consolidating business intelligence objects (reports, documents, spreadsheets, data cubes, etc.) gener-

ated anywhere in the enterprise by any application and making them easily accessible, subject to security authorization, to nontechnical users via standard browser technology.

Enterprise Resource Planning (ERP)   ERP systems are comprised of software programs which tie together all of an enterprise's various functions—such as finance, manufacturing, sales, and human resources. This software also provides for the analysis of the data from these areas to plan production, forecast sales, and analyze quality. Today, many organizations are realizing that to maximize the value of the information stored in their ERP systems, it is necessary to extend the ERP architectures to include more advanced reporting, analytical, and decision support capabilities. This is best accomplished through the application of datawarehousing tools and techniques.

Enterprise Storage   A shared central repository for information, connected to disparate computer systems, which provides common management, protection, and information-sharing capabilities.

Entity Identification   The identification of the entities involved in the subject area. Entity identification is the process of giving data entities unique data elements by which they can be identified.

Entity Relationship Diagramming   A process that visually identifies the relationships between data elements.

Equity   The accounting value of a business, equal to assets minus liabilities. Commonly used interchangeably with book value, net asset value, net worth, and shareholders' equity.

ERP   See Enterprise Resource Planning.

ETL   See Extract, transform, and load.

Event Analysis   A process of analyzing notifications and taking action based on the notification content.

Event Data   Data about business events (usually business transactions) that have historic significance or are needed for analysis by other systems. Event data may exist as atomic event data and aggregate data.

Event-Based Execution Rules   The process of identifying those tasks that must be successfully executed to completion, or the system events that must occur, before a given task is to be triggered for processing.

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

Extendibility   The ability to easily add new functionality to existing services without major software rewrites or without redefining the basic architecture.

Extract Date   The date data was extracted.

Extract Frequency   The latency of data extracts, such as daily versus weekly, monthly, quarterly, etc. The frequency that data extracts are needed

in the datawarehouse is determined by the shortest frequency requested through an order, or by the frequency required to maintain consistency of the other associated data types in the source datawarehouse.

Extract Specification   The standard expectations of a particular source datawarehouse for data extracts from the operational database system of record. A system of record uses an extract specification to retrieve a snapshot of shared data, and formats the data in the way specified for updating the data in the source datawarehouse. An extract specification also contains extract frequency rules for use by the data access environment.

Extract, Transform, and Load (ETL)   Powerful tools that help today's organizations move large data volumes smoothly and efficiently between databases.

Extranet   An internal network or intranet opened to selected business partners. Suppliers, distributors, and other authorized users can connect to a company's network over the Internet or through private networks.

Fastload   A technology that typically replaces a specific DBMS load function. A fastload technology obtains significantly faster load times by preprocessing data and bypassing data integrity checks and logging.

FIFO   A method of posting a transaction in first-in, first-out order. In other words, transactions are posted in the same order that the data producer entered them.

Filters   Saved sets of chosen criteria that specify a subset of information in a datawarehouse.

Financial Budget   A budget that embraces the impacts of the financial decisions of the firm. It is a plan including a budgeted balance sheet, which shows the effects of planned operations and capital investments on assets, liabilities, and equities. It also includes a cash budget, which forecasts the flow of cash and other funds in the business.

Financial Projection   An essential element of planning that is the basis for budgeting activities and estimating future financing needs of a firm (financial projections (forecasts) being with forecasting sales and their related expenses).

Financial Statement   An accounting document showing the financial status of a business or the results of business activity.

Finished Goods Inventory   The portion of inventory that consists of goods and products ready for sale.

Firewall   A combination of specialized hardware and software set up to monitor traffic between an internal network and an external network (i.e., the Internet). Its primary purpose is for security and is designed to keep unauthorized outsiders from tampering with or accessing information on a networked computer system.

First Normal Form   The result of normalizing to ensure that every attribute is dependent upon the key of an entity. In practice, this is achieved by

removing repeating groups and multivalued attributes and modeling them as distinct entities.

Fiscal Year   The 12-month period for which financial results are prepared and reported. It may be different, by company choice, from the calendar year.

Fixed Assets   Assets whose life will extend beyond the current accounting period: machinery, land, buildings, and the like.

Fixed Cost   A cost that does not vary with revenue over a relevant range of revenue amount.

Fixed Overhead Variance   The difference between actual fixed overhead incurred and fixed overhead applied to production.

Flash Report   A report that provides the highlights of key information promptly to the responsible nonfinancial manager. An example is an exception report, such as performance reports, that highlight favorable or unfavorable variances. A flash report allows managers to take a corrective action for an unfavorable variance.

Foreign Key   A foreign key is the primary key of one data structure that is placed into a related data structure to represent a relationship among those structures. Foreign keys resolve relationships, and support navigation among data structures.

Frequency   The timing characteristics of the data.

Gateway   A software product that allows SQL-based applications to access relational and nonrelational data sources.

General and Administrative Expense (G&A)   Cost necessary to operate a business but not associated directly with the production of revenue, such as the cost of accounting.

Global Business Models   Provides access to information scattered throughout an enterprise under the control of different divisions or departments with different databases and data models. This type of datawarehouse is difficult to build because it requires users from different divisions to come together to define a common data model for the warehouse.

Granularity   Degree of summarization of data.

Gross Margin   A measure of profitability, equal to revenue minus cost of goods sold divided by revenue. (The term is sometimes used for only the dollar amount of revenue minus costs of goods sold, although the latter is more often called *gross profit*).

Hash   Data allocated in an algorithmically randomized fashion in an attempt to evenly distribute data and smooth access patterns.

Hierarchical Relationships   Any dimension's members may be organized based on parent–child relationships, typically where a parent member represents the consolidation of the members which are its children. The result is a hierarchy, and the parent–child relationships are hierarchical relationships.

High-level Enterprise Model   A formal, high-level description in business terms of the information needed and used to manage the entire business, understood by business users and IS personnel.

Historical Database   A database that provides an historical perspective on the data.

HOLAP (Hybrid OLAP)   A product that can provide multidimensional analysis simultaneously of data stored in a multidimensional database and in an RDBMS. Becoming a popular architecture for server OLAP.

Horizontal Partitioning   Horizontal partitioning divides a single logical table into multiple physical tables based on the rows. All columns generally appear in the new tables, but each new table contains a subset of the original table's rows. The resultant tables may contain either discrete or overlapping subsets of the original table's data. Horizontal partitioning is employed when there is a regular need to access or to isolate a readily identifiable subset of the "parent" table's rows. This technique may be effective to meet security, distribution, and performance optimization needs.

Host-Driven   A processing method in which the host computer controls the session. A host-driven session typically includes terminal emulation, front ending, or client–server types of connections. The host determines what is displayed on the desktop, receives user input from the desktop, and determines how the application responds to the input.

Hypercube   An OLAP product that stores all data in a single cube that has all the application dimensions applied to it.

Increment   Datawarehouse implementation can be broken down into segments or increments. An increment is a defined datawarehouse implementation project that has a specified beginning and end. An increment may also be referred to as a departmental datawarehouse within the context of an enterprise.

Indirect Cost   Cost not directly related or assignable to the production of specific revenue, products, or services.

Info-Glut in Cyberspace   Too much data! (30+ million electronic mailboxes, 7,000 CD-ROMs with 650 Megs, 5000+ online databases, 500 cable channels, etc.)

Information   Data that has been processed in such a way that it can increase the knowledge of the person who receives it. Information is the output, or finished goods, of information systems. Information is also what individuals start with before it is fed into a data capture transaction processing system.

Information Consumer   A person or software service that uses data to create information.

Information Needs Analysis   The identification and analysis of the needs for information required to satisfy a particular business driver.

Information Systems Architecture   The authoritative definition of the business rules, systems structure, technical framework, and product backbone for business information systems. An information systems architecture consists of four layers: business architecture, systems architecture, technical architecture, and product architecture.

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

Intelligent Agent   A software routine that waits in the background and performs an action when a specified event occurs. For example, agents could transmit a summary file on the first day of the month or monitor incoming data and alert the user when certain transactions have arrived.

Interoperability   The ability of various types of computers and programs to work together.

Interviews   A procedure to obtain prioritized information needed to generate warehouse increments.

Intranet   The subset of the Internet used internally by a company or organization. Unlike the Internet, intranets are private and accessible only from within the organization.

Inverted File Indexes   A more efficient method to access data in an ad-hoc or analysis environment. It maintains indexes to all values contained in an indexed field. Those values, in turn, can be used in any combination to identify records that contain them, without actually scanning them from disk.

Local Directory   A data dictionary propagated from the repository to the desktop containing metadata used for developing desktop applications and for generating transactions. A local directory is also used to bind definitions of local data structures used by desktop applications to the data requested from servers.

Location Transparency   A mechanism that keeps the specific physical address of an object from a user. The physical location is resolved within the system so that operations can be performed without knowledge of the actual physical location.

Logical Data Model   Actual implementation of a conceptual module in a database. It may take multiple logical data models to implement one conceptual data model.

Long-Range Budget   Projections that cover more than one fiscal year. It is also called strategic budgeting. The five-year budget plan is the most commonly used.

Management by Exception   A management concept or policy by which management devotes its time to investigating only those situations in which actual results differ significantly from planned results. The idea is that management should spend its valuable time concentrating on the more important items (such as shaping of the company's future strategic course).

Management Control System   A system under which managers assure that resources are obtained and used effectively and efficiently in the accomplishment of the organization's goals.

Margin   A percentage measure of profitability relative to revenue (also sometimes used as the dollar amount of revenue minus selected costs).

Market Share   A company's sales expressed as a percentage of the sales for the total industry.

Master (Comprehensive) Budget   A plan of activities expressed in monetary terms of the assets, equities, revenues, and costs which will be involved in carrying out the plans. Simply put, a master budget is a set of projected or planned financial statements.

Metadata   Metadata is data about data. Examples of metadata include data element descriptions, data type descriptions, attribute/property descriptions, range/domain descriptions, and process/method descriptions. The repository environment encompasses all corporate metadata resources: database catalogs, data dictionaries, and navigation services. Metadata includes things like the name, length, valid values, and description of a data element. Metadata is stored in a data dictionary and repository. It insulates the datawarehouse from changes in the schema of operational systems.

Metadata Synchronization   The process of consolidating, relating, and synchronizing data elements with the same or similar meaning from different systems. Metadata synchronization joins these differing elements together in the datawarehouse to allow for easier access.

Methodology   A system of principles, practices, and procedures applied to a specific branch of knowledge.

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

Mid-Tier Datawarehouse   To be scalable, any particular implementation of the data access environment may incorporate several intermediate distribution tiers in the datawarehouse network. These intermediate tiers act as source datawarehouses for geographically isolated sharable data that is needed across several business functions.

Mini Mart   A small subset of a datawarehouse used by a small number of users. A mini mart is a very focused slice of a larger datawarehouse.

MIP-O-Suction   A query that consumes a high percentage of CPU cycles.

MIPS   An acronym for millions of instructions per second. MIPS is mistakenly considered a relative measure of computing capability among models and vendors. It is a meaningful measure only among versions of the same processors configured with identical peripherals and software.

Moving Average   For a time series, an average that is updated as new information is received. With the moving average, the manager employs the most recent observations to calculate an average, which is used as the forecast for the next period.

Massive Parallel Processing (MPP)   The "shared nothing" approach of parallel computing.

Multidimensional   Data structure with three or more independent dimensions.

Multidimensional Array   A group of data cells arranged by the dimensions of the data. For example, a spreadsheet exemplifies a two-dimensional array with the data cells arranged in rows and columns, each being a dimension. A three-dimensional array can be visualized as a cube with each dimension forming a side of the cube, including any slice parallel with that side. Higher dimensional arrays have no physical metaphor, but they organize the data in the way users think of their enterprise. Typical enterprise dimensions are time, measures, products, geographical regions, sales channels, etc.

Multidimensional Database (MDBS and MDBMS)   A powerful database that lets users analyze large amounts of data. An MDBS captures and presents data as arrays that can be arranged in multiple dimensions.

Multidimensional Query Language   A computer language that allows one to specify which data to retrieve out of a cube. The user process for this type of query is usually called slicing and dicing. The result of a multidimensional query is either a cell, a two-dimensional slice, or a multidimensional subcube.

Multidimensional Analysis   The objective of multidimensional analysis is for end users to gain insight into the meaning contained in databases. The multidimensional approach to analysis aligns the data content with the analyst's mental model, hence reducing confusion and lowering the incidence of erroneous interpretations. It also eases navigating the database, screening for a particular subset of data, asking for the data in a particular orientation, and defining analytical calculations. Furthermore, because the data is physically stored in a multidimensional structure, the speed of these operations is many times faster and more consistent than is possible in other database structures. This combination of simplicity and speed is one of the key benefits of multidimensional analysis.

Naïve Forecast   Forecasts obtained with a minimal amount of effort and data manipulation and based solely on the most recent information available. One such naïve method would be to use the most recent datum available as the future forecast.

Net Present Value (NPV)   The difference between the present value (PV) of all cash inflows generated by the project and the amount of the initial value (I).

Normalization   The process of reducing a complex data structure into its simplest, most stable structure. In general, the process entails the removal of redundant attributes, keys, and relationships from a conceptual data model.

Object   A person, place, thing, or concept that has characteristics of interest to an environment. In terms of an object-oriented system, an object is an entity that combines descriptions of data and behavior.

Object Description All the properties and associations that describe a particular object.

Open Database Connectivity (ODBC) A standard for database access coopted by Microsoft from the SQL Access Group consortium.

OLAP Client End-user applications that can request slices from online analytical processing (OLAP) servers and provide two-dimensional or multidimensional displays, user modifications, selections, ranking, calculations, etc., for visualization and navigation purposes. OLAP clients may be as simple as a spreadsheet program retrieving a slice for further work by a spreadsheet-literate user or as high functioned as a financial modeling or sales analysis application.

Online Transaction Processing (OLTP) OLTP describes the requirements for a system that is used in an operational environment.

Online Analytical Processing Processing that supports the analysis of business trends and projections.

Online Transaction Processing Processing that supports the daily business operations.

Open Architecture When a manufacturer publicly publishes the specifications for their computer, the computer is said to have an open architecture. This allows other companies to create add-ons to enhance and customize the machine, and to make peripheral devices that work properly with it. With a closed architecture, only the original manufacturer can make add-ons and peripherals.

Operational (Operating) Budget A budget that embraces the impacts of operating decisions. It contains forecasts of sales, net income, the cost of goods sold, selling and administrative expenses, and other expenses.

Operational Data Store (ODS) An ODS is an integrated database of operational data. Its sources include legacy systems and it contains current or near term data. An ODS may contain 30 to 60 days of information, while a datawarehouse typically contains years of data.

Operational Database The database of record, consisting of system-specific reference data and event data belonging to a transaction-update system. It may also contain system control data such as indicators, flags, and counters. The operational database is the source of data for the datawarehouse. It contains detailed data used to run the day-to-day operations of the business. The data continually changes as updates are made, and reflect the current value of the last transaction.

Operational Systems Applications that run the business on a day-to-day basis using real-time data.

Overhead Indirect costs that are often used to describe indirect cost in a particular function or activity, closely associated with production of revenue but not assignable to a particular product or service. Typical classes of overhead are manufacturing labor overhead, manufacturing material overhead, engineering overhead, and service overhead.

Parallelism   The ability to perform functions in parallel.

Partitioning   Splitting of target data into smaller units.

Payback Period   The length of time required to recover the initial amount of a capital investment.

Percentage Of Completion   A method of accounting, used for large and long contracts, that recognizes revenue during the course of the contract in accordance with the proportion of work that has been completed, or cost that has been incurred.

Period Cost   Cost expensed during the same period in which it is incurred.

Population   See Data Loading and Data Replication.

Portal   A web site that is the first place people visit when using the web. Typically, a portal site has a catalog of sites, a search engine, or both. A portal site may also offer e-mail and other services to entice people to use that site as the main point of entry or portal to the web. Portals are designed to be the "front door" through which a user accesses links to relevant sites.

Primary Key   A column or combination of columns whose values uniquely identify a row or record in the table. The primary key(s) will have a unique value for each record or row in the table.

Pro Forma Balance Sheet   A budgeted balance sheet.

Pro Forma Income Statement   A budgeted income statement.

Process Management   A set of functions supporting the definition of inter-related process steps and the management of their execution across a variety of hardware and software platforms; used mainly by data replication.

Product Architecture   One of the four layers of an information systems architecture. It describes standards to be followed in each portion of the technical architecture and vendor-specific tools and services to apply in developing and running applications.

Production Budget   A schedule for expected units to be produced. It sets forth the units expected to be manufactured to satisfy budgeted sales and inventory requirements. Expected production volume is determined by adding desired ending inventory to planned sales and then subtracting beginning inventory.

Production Data   Source data which is subject to change. It is a data capture system, often on a corporation's mainframe

Profit Center   The unit in an organization that is responsible for revenues earned and costs incurred. The manager of a profit center has control over revenues and costs, as well as attempts to maximize profit.

Profit Planning   A process of developing a profit plan that outlines the planned sales revenues and expenses and the net income or loss for a time period. Profit planning requires preparation of a master budget and various analyses for risk and "what-if" scenarios. Tools for profit planning include the cost-volume-profit (CVP) analysis and budgeting.

Profit Variance   A difference between actual profit and budgeted profit. Profit, whether it is gross profit in absorption costing or contribution margin in direct costing, is affected by sales price, sales volume, and costs.

Profitability Index   The ratio of the total present value (PV) of future cash inflows to the initial investment (I).

Profit-Volume Chart   A chart that determines how profits vary with changes in volume. Profits are plotted on the vertical axis while units of output are shown on the horizontal axis.

Projected (Budgeted) Balance Sheet   A schedule for expected assets, liabilities, and stockholders' equity. It projects a company's financial position at the end of the budgeting year. A budgeted balance sheet discloses unfavorable financial conditions that management may want to avoid, serves as a final check on the mathematical accuracy of all other budgets, and highlights future resources and obligations.

Projected (Budgeted) Income Statement   A summary of various component projections of revenues and expenses for the budget period. It indicates the expected net income for the period.

Propagated Data   Data that is transferred from a data source to one or more target environments according to propagation rules. Data propagation is normally based on transaction logic.

Protocol   A set of conventions that governs the communications between processes. Protocol specifies the format and content of messages to be exchanged.

Quality Assurance   The process of ensuring a correct result.

Quantitative Forecasting   A technique that can be applied when information about the past is available, if that information can be quantified and if the pattern included in past information can be assumed to continue into the future.

Query   A (usually) complex SELECT statement for decision support. See Ad-Hoc Query or Ad-Hoc Query Software.

Query Governor   A facility that terminates a database query when it has exceeded a predefined threshold.

Query Response Times   The time it takes for the warehouse engine to process a complex query across a large volume of data and return the results to the requester.

Query Tools   Software that allows a user to create and direct specific questions to a database. These tools provide the means for pulling the desired information from a database. They are typically SQL-based tools and allow a user to define data in end-user language.

Raw Materials Inventory   The portion of inventory that consists of purchased material that will be used to make revenue-producing products, and the purchase cost of that material.

Real Time   Refers to the utmost level of timeliness regarding the use of information.

Real-time Data   Up-to-the-second, detailed data used to run the business and accessed in read/write mode, usually through predefined transactions.

Redundancy   The storage of multiple copies of identical data.

Redundancy Control   Management of a distributed data environment to limit excessive copying, update, and transmission costs associated with multiple copies of the same data. Data replication is a strategy for redundancy control with the intention to improve performance.

Reference Data   Business data that has a consistent meaning and definition and is used for reference and validation (e.g., process, person, vendor, and customer). Reference data is fundamental to the operation of the business. The data is used for transaction validation by the data capture environment, decision support systems, and for representation of business rules. Its source for distribution and use is a datawarehouse.

Replicated Data   Data that is copied from a data source to one or more target environments based on replication rules. Replicated data can consist of full tables or rectangular extracts.

Repository Environment   The repository environment contains the complete set of a business's metadata. It is globally accessible. As compared to a data dictionary, the repository environment not only contains an expanded set of metadata, but can be implemented across multiple hardware platforms and database management systems (DBMS).

Residual   A synonym for error. It is calculated by subtracting the forecast value from the actual value to give a residual or error value for each forecast period.

Responsibility Accounting   The collection, summarization, and reporting of financial information about various decision centers (responsibility centers) throughout an organization.

Responsibility Center   A unit in the organization which has control over costs, revenues, or investment funds. Responsibility centers are classified as cost centers, revenues centers, profit centers, and investment centers.

Return on Assets   Profit divided by assets, a measure of the percentage of the value of its assets that is earned by the business.

Return on Capital   Profit divided by capital, a measure of the percentage of the total investment earned by the business.

Return on Equity   Profit divided by equity, a measure of the percentage of the owners' investment earned by the business.

Return on Investment   For an entire business, synonymous with return on capital. For a given capital investment within a business, the ratio of the profit or cash flow that will result to the amount of the investment.

Risk Analysis   The process of measuring and analyzing the risks associated with financial and investment decisions. Risk refers to the variability of expected returns (earnings or cash flows).

ROLAP (Relational OLAP)   A product that provides multidimensional analysis of data, aggregates, and metadata stored in an RDBMS. The multidimensional processing may be done within the RDBMS, a mid-tier server, or the client. A merchant ROLAP is one from an independent vendor which can work with any standard RDBMS.

Roll-Up Queries   Queries that summarize data at a level higher than the previous level of detail.

Sales Budget   An operating plan for a period expressed in terms of sales volume and selling prices for each class of product or service. Preparation of a sales budget is the starting point in budgeting, since sales volume influences nearly all other items.

Sales Forecasting   A projection or prediction of future sales. It is the foundation for the quantification of the entire business plan and a master budget. Sales forecasts serve as a basis for capacity planning, budgeting, production and inventory planning, manpower planning, and purchasing planning.

Sales Price Variance   The difference between the actual number of units sold and the budgeted number, multiplied by the budgeted selling price per unit. It is also called sales quantity variance.

Scalability   The ability to scale to support larger or smaller volumes of data and more or less users. The ability to increase or decrease size or capability in cost-effective increments with minimal impact on the unit cost of business and the procurement of additional services.

Schema   A diagrammatic representation of the structure or framework of something. It is the logical and physical definition of data elements, physical characteristics, and interrelationships.

Second Normal Form   The result of normalizing to ensure that a data model contains no partial key dependencies. In practice, when entities have compound keys, seek out any attribute that is dependent on only part of the key. Whatever business thing is identifiable by that part of the key is an entity, and the 2NF violation is an attribute of that entity.

Secondary Key   A secondary key is a set of one or more attributes whose value identifies a set of occurrences in a data structure that share common characteristics. Access by secondary keys may return multiple occurrences, where access by a primary key is assured to find no more than one occurrence.

Securability   The ability to provide differing access to individuals according to the classification of data and the user's business function, regardless of the variations.

SELECT   An SQL statement (command) that specifies data retrieval operations for rows of data in a relational database.

Semantic Mapping   The mapping of the meaning of a piece of data.

Server   A service that provides standard functions for clients in response to standard messages from clients. *Note:* A commonly used definition of server also refers to the physical computer from which services are provided.

Simulation   An attempt to represent a real life system via a model to determine how a change in one or more variables affects the rest of the system. It is also called "what-if" analysis.

Simulation Model   A "what-if" model that attempts to simulate the effects of alternative management policies and assumptions about the firm's external environment. It is basically a tool for management's laboratory.

Slice and Dice   A term used to describe a complex data analysis function provided by MDBMS tools.

Symmetrical Multiprocessing (SMP)   The "shared everything" approach of parallel computing.

Snowflake Schema   A set of tables composed of a single, central fact table surrounded by normalized dimension hierarchies. Each dimension level is represented in a table. Snowflake schemas implement dimensional data structures with fully normalized dimensions. Star schemas are an alternative to snowflake schema.

Source Database   An operational, production database or a centralized warehouse that feeds into a target database.

Structured Query Language (SQL)   A structured query language for accessing relational, ODBC, DRDA, or nonrelational compliant database systems.

SQL Query Tool   An end-user tool that accepts SQL to be processed against one or more relational databases.

SQL Compliant   Conformity to ANSI standards for Structured Query Language specifications.

Standard Cost   Production or operating costs that are carefully predetermined. A standard cost is a target that should be attained.

Standard Cost System   A system by which production activities are recorded at standard costs and variances from actual costs are isolated.

Standard Query   A stored procedure of a recently executed query. Technically, a standard query may be stored on the desktop as "canned " SQL and passed as dynamic SQL to the server database to execute. This is undesirable unless the stored query is seldom executed.

Star Schema   A set of tables composed of a single, central fact table surrounded by denormalized dimensions. Each dimension is represented in a single table. Star schemas implement dimensional data structures with denormalized dimensions. Snowflake schemas are an alternative to star

schemas. A relational database schema for representing multidimensional data. The data is stored in a central fact table, with one or more tables holding information on each dimension. Dimensions have levels, and all levels are usually shown as columns in each dimension table.

Static Query   A stored, parameterized procedure, optimized for access to a particular datawarehouse.

Stoplighting   A technique using colored circles to identify the content of a data element. The colors are defined by a set of predefined thresholds.

Strategic Planning   The implementation of an organization's objectives. Strategic planning decisions will have long-term impacts on the organization while operational decisions are day to day in nature.

Subject-Oriented Databases   Rather than build one massive, centralized datawarehouse, most companies are building numerous subject-oriented warehouses to serve the needs of different divisions.

Summarization Tables   These tables are created along commonly used access dimensions to speed query performance, although the redundancies increase the amount of data in the warehouse. See Aggregated Data.

Surrogate Key   A surrogate key is a single-part, artificially established identifier for an entity. Surrogate key assignment is a special case of derived data—one in which the primary key is derived. A common way of deriving surrogate key values is to assign integer values sequentially.

Syntactic Mapping   The mapping required to unravel the syntax of information.

Systems Architecture   One of the four layers of the information systems architecture. The systems architecture represents the definitions and inter-relationships between applications and the product architecture.

Tactical Datawarehouse Development   The process of selecting a portion of an enterprise and implementing a datawarehouse. The process includes constructing a data model for the area, determining the datawarehouse architecture, constructing the physical model, and populating the warehouse database. It also includes creating or buying the applications to access the datawarehouse, prototyping the tactical warehouses (access definitions, views, etc.) and incorporating end-user feedback.

There Ain't No Such Thing as a Free Lunch! (**TANSTAAFL**)   In developing a datawarehouse, there is work involved, and there is no "free lunch."

Target Database   The database in which data will be loaded or inserted.

Technical Architecture   One of the four layers of the information systems architecture. The technical architecture defines and describes the interfaces, parameters, and protocols used by the product and systems architecture.

The End-User Mindset   "Give me what I say I want, then I can tell you what I really want." To build a successful datawarehouse, end users must be able to explore the possibilities.

Tool Encyclopedias   Encyclopedias, repositories, or dictionaries used by application development tools. The nondefinable "repository" used by a tool.

Transformers   Rules applied to change data.

Triggering Data   Data that selects and loads data on a scheduled basis

Unit of Work Consolidation   The process of consolidating multiple updates to a single row image into a single update.

Update   Not allowed in a datawarehouse.

Variable Cost   Cost that varies with revenue.

Variable Overhead Efficiency Variance   The difference in actual and budgeted variable overhead costs that are incurred due to inefficient use of indirect materials and indirect labor.

Variable Overhead Spending Variance   The difference in actual and budgeted overhead costs that result from price changes in indirect materials and indirect labor and insufficient control of costs of specific overhead items.

Variance   The difference of revenues, costs, and profit from the planned amounts. One of the most important phases of responsibility accounting is establishing standards in costs, revenues, and profit and establishing performance by comparing actual amounts with the standard amounts. The differences are calculated for each responsibility center and analyzed, and unfavorable variances are investigated for possible remedial action.

Versioning   The ability for a single definition to maintain information about multiple physical instantiations.

Vertical Partitioning   Vertical partitioning divides a single logical table into multiple physical tables based on the columns. All rows may appear in the new tables, but each new table contains a subset of the original table's columns. The set of columns may be redundant across tables, and will necessarily be so for the columns that implement keys and indexes. Columns for row-level metadata are also implemented in all resultant tables. Vertical partitioning is employed when there is a regular need to access or to isolate a readily identifiable subset of the "parent" table's columns. This technique may be effective to meet security, distribution, and usability requirements.

VITAL Compliance   Conformance to the design objectives and principles, distributed computing styles, development approaches, standards, and data-distribution and access techniques; functional compatibility with the vertically integrated technical architecture life cycle (VITAL) technical architecture.

Vortal   Targeted vertical portals that are aimed at specific interest groups and focus on providing consumers with a gateway to information from other sources.

Warehouse Business Directory   Provides business professionals access to the datawarehouse by browsing a catalog of contents.

Warehouse Technical Directory   Defines and manages an information life cycle, a definition of warehouse construction, change management, impact analysis, distribution, and operation of a warehouse.

Work-in-Progress Inventory   The portion of inventory that consists of partially completed products, and the associated burdened labor and material costs.

Working Capital   Current assets minus current liabilities.

XML   eXtensible Markup Language; a subset of SGML defined by W3C.

Zero-Base Budgeting (ZBB)   A planning and budgeting tool that uses cost/benefit analysis of projects and functions to improve resource allocation in an organization. Traditional budgeting tends to concentrate on the incremental changes from the previous year. It assumes that the previous year's activities and programs are essential and must be continued. Under zero-base budgeting, however, cost and benefit estimates are built up from scratch, from the zero level, and must be justified.

ENDNOTE

The information in this Glossary is based on the Glossary found at *www.dmreview.com* (under General Resources), which is maintained and edited by Bill Inmon.

INDEX