3

Data Marts and Data Warehouse: Information Architecture for the Millennium

Objectives:

- The new dynamics that characterize today's global business environments are unprecedented in their scope and impact, requiring organizations to adapt their strategies and business practices far more rapidly and intelligently than ever before.
- Data warehouses and marts give organizations the ability to consolidate information from multiple, heterogeneous production systems into a common, integrated database environment.
- In this section we explain the structure and value of today's enterprise data warehouse and data mart systems, positioning them within what we call the "corporate information factory."
- We describe how these systems are used by corporate decision makers to gain and maintain a strategic edge over their competition, and suggest criteria for selecting and evaluating different data warehouse and data mart technologies for different purposes. We highlight the benefits Informix® delivers with its Decision Frontier™.
- The data mart/enterprise data warehouse environment increasingly becomes important as the data warehouse environment evolves to the corporate information factory.
- To provide the modern communications company with an environment that supports the business, an architecture has been developed to address the technological and business issues it faces.
- To reduce the complexity of the decision support systems and the development of the data warehouse, it is necessary to understand the business issues, technical infrastructure issues, and data issues.
- The telecommunications lifecycle is a holistic approach that addresses the problems at the root, long before an invoice appears.
- To review the key features required for OLAP software as referenced in the OLAP Council benchmark specification.

S. Sumathi and S.N. Sivanandam: *Data Marts and Data Warehouse: Information Architecture for the Millennium*, Studies in Computational Intelligence (SCI) **29**, 75–150 (2006)
www.springerlink.com

- Data warehousing is a management tool that enables executives to access the information they need to make informed business decisions.
- A data warehouse coupled with a focused set of decision support applications provides seamless integration of communication services globally, better understanding of profitability, and a focused marketing effort.
- The future success or failure of telecommunication companies could well depend on a company's knowledge of the market and its customers, which will, in turn, depend on the successful implementation of application tools as part of a data warehouse solution.
- The telecommunications lifecycle is a holistic approach that addresses the problems at the root, long before an invoice appears.
- The telecommunications lifecycle approach is the best practices way to address the problems before they occur and avoid overpaying.
- This section looks into the different techniques used over the years to achieve the perfect balance between performance and security and then propose a security model that is mostly based on XML and related technologies.
- A single technology for implementing security meeting the required performance target is yet to be devised, but the existing technologies can be used in combination to produce satisfactory results.
- This section explores the business and technology issues related to both buying and building data warehousing solutions.
- We discuss requirements for effective insurance decision support systems and illustrate why we feel that buying a solution is ultimately more cost effective for insurance companies.
- We present all the main data mining techniques, giving explanations on how they work, and what can they be used for.
- The expression *data mining* includes a wide range of tools and techniques to extract information from a large collection of data.
- Another important issues is to design algorithms that can handle the huge dimension of modern data warehouses.

Abstract. In this section we explain the structure and value of today's enterprise data warehouse and data mart systems, positioning them within what we call the "corporate information factory." We describe how corporate decision makers use these systems to gain and maintain a strategic edge over their competitors, and suggest criteria for selecting and evaluating different data warehouse and data mart technologies for different purposes. We highlight the benefits Informix® delivers with its Decision Frontier™. The integrated products and services that make up decision frontier provide a complete, high-performing and highly scalable data warehousing environment—including query, reporting, and data analysis capabilities, along with industry-specific templates and consulting services – which enables the rapid deployment of business-critical analytic applications.

Data warehousing is a management tool that enables executives to access the information they need to make informed business decisions. Information is the greatest weapon in the competitive arsenal of healthcare organizations. Information can

tell organizations how well providers are performing, what treatments are working, and if clients are satisfied with the services they receive. That may sound simplistic, but once the data is manipulated and interpreted, it is like cracking the enemy's code and being privy to information that reveals the lay of the battlefield and the conditions of the war. In the competitive healthcare arena, this extra information can mean the difference between survival and demise.

The raw material (data) needed for the development of an enterprise-wide data warehouse is already abundant in the deregulated telecommunications industry. Customer data is both the most plentiful and most underused resource currently available. However, data alone will not generate meaningful information and knowledge unless the company's business objectives and information technology (IT) are aligned. For technology to become a key enabler in the implementation of new strategies it must first become integrated into the company's business functions. The ability to acquire, manage, and retain customers will be driven by the speed and efficiency with which the carriers are able to utilize their architectural infrastructure. The future success or failure of telecommunication companies could well depend on a company's knowledge of the market and its customers, which will, in turn, depend on the successful implementation of application tools as part of a data warehouse solution. Effective data warehousing often results in very large databases, reaching multiple terabytes of storage. Currently, a major issue within the telecommunications companies is that the typical IT infrastructure is unable to deliver the level and quality of data to support the business unit new processes.

Data warehousing has increasingly become an important and rather essential phenomenon in the world of enterprise data storage. The idea is to allow a single infrastructure to be used in the functionality of data warehouse and to facilitate the data distribution and manipulation by the users/clients of the data in a simplified way. Different techniques are used to achieve this target with the main emphasis resting with the performance of the data warehouse. It is not easy to handle an enterprise-sized data and yet met the target performance as well. Performance may be the aim but the fact remains that with the data of an entire enterprise resting in a collection, the security has to be the critical issue. This section looks into the different techniques used over the years to achieve the perfect balance between performance and security and then propose a security model that is mostly based on XML and related technologies. In addition, the data warehousing concept for insurers is also discussed.

3.1 Data Marts, Data Warehouse, and OLAP

3.1.1 Business Process Re-engineering

The new dynamics that characterize today's global business environments are unprecedented in their scope and impact, requiring organizations to adapt their strategies and business practices far more rapidly and intelligently than ever before. Powerful market forces like mass customization in the manufacturing and retail sectors, deregulation in banking, utilities, communications, and insurance, and electronic commerce across virtually all industries are driving change within the organization at an extraordinary rate and down to its very

foundations. What is more, this trend toward the complete reinvention of core business structures and processes will only accelerate over the coming years.

In the face of such profound change, organizations are concluding that high-quality information—about markets, competitors, economic conditions, resources, and their own business—has gone beyond being a success factor and has become a survival issue. And even as decision makers struggle to make sense out of the data relevant to their business the quantity, complexity, and scope of that data is growing at an exponential rate.

3.1.2 Real-World Usage

At the same time, however, many companies—including some of Informix's leading customers—are already developing innovative ways of doing business that are creating for them unprecedented opportunities. For example:

- A bank can consolidate information from checking, savings, mortgage, and investment management applications to understand, often for the first time, who are its most valuable customers. Once these customers have been identified, the institution can proactively market to them in new, innovative ways.
- A telecommunications company can analyze call records and identify customers who are likely to churn—that is, defect to a competing vendor. With this information the company can determine how much of an investment is required to keep those customers.
- A manufacturing firm can combine customer support and shipping data in a data warehouse to identify vendors that are supplying goods that cause downstream product failures. This information enables the manufacturer to work with its suppliers to help those companies improve their processes and procedures, helping each company produce a higher quality product, reduce waste, and improve customer satisfaction. So the challenge for today's organizations has become one of getting ahead of the rapidly moving business intelligence curve through the smart acquisition and use of data and information technologies.

3.1.3 Business Intelligence

Business intelligence is a term that has emerged to define the data-analysis and decision-making environment within corporations. It is a vehicle that can enable these organizations to define, establish, and maintain competitive advantage even as the volume and complexity of critical business data continues to explode. Transforming data into business intelligence has become a trend that is increasing at a very sharp rate, as demonstrated by greater than 50% annual growth in budgets allocated by the Global 2000 corporations for data warehouses, data marts, and business intelligence software solutions. Data warehouses and marts give organizations the ability to consolidate information

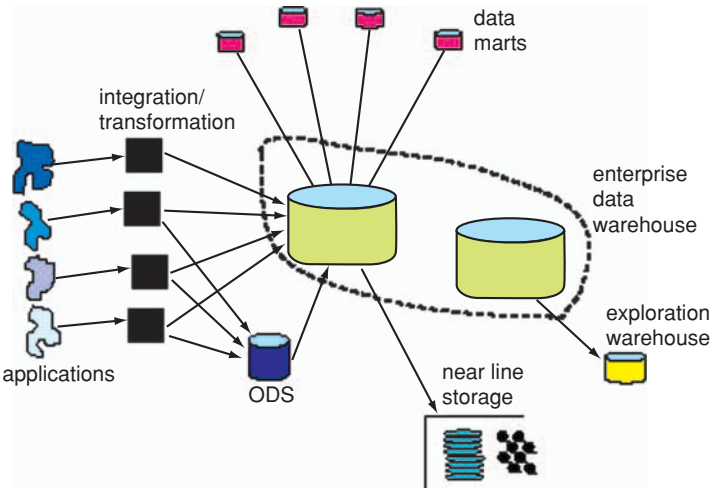

Fig. 3.1. The Corporate Information Factory

from multiple, heterogeneous production systems into a common, integrated database environment. Information in this environment can then be organized, summarized, regularly refreshed, and used to great advantage by a wide variety of powerful analytical applications. The latter give decision makers critically important tools for making strategic decisions quickly and confidently.

From the humble origins of a system that was something other than transaction processing capabilities, data warehousing has grown into a full-blown architecture. Today, data warehousing has different components that serve different purposes and communities. The architecture that has grown from the origins of data warehousing is today something that can be called the *corporate information factory*. Figure 3.1 depicts today's mature corporate information factory.

The corporate information factory provides the starting point for data warehousing—the legacy application systems. In the legacy applications, transactions are executed in which detailed transaction data is collected. The detailed data that is collected is more of a by-product of transaction processing than anything else. But this raw data forms the very detailed foundation needed for the data warehouse. The transaction data that is collected passes through a layer of programs whose purpose is to integrate and transform the data to make it fit to enter the data warehouse. The layer of programs can be called *integration and transformation* programs or *ETL*—extract/transform/load programs. These programs can be written manually or in an automated manner by program generators specifically designed to suit the needs of this interface.

Once the legacy data passes through the integration and transformation programs, the data is fundamentally changed to meet the informational needs

of the corporation. Keys are altered so that there is one corporate understanding of data, structures are changed, and encoding algorithms are unified. The data enters the layer of integration and transformation programs in an unintegrated state and passes out of the layer of programs in an integrated state. The data then passes into the enterprise data warehouse in a very granular form. Once in the enterprise data warehouse, the data is ready to serve as a basis for all sorts of decision-support systems (DSS) processing. The enterprise data warehouse is fed by the operational data store (ODS) as well. The ODS is a hybrid structure where both DSS and operational needs are met. The ODS provides standard transaction response time—two to three seconds—besides serving as a place where data can be integrated.

Emanating from the enterprise data warehouse is the collection of data marts of the corporation. In many ways the data marts appear to be the data warehouse because it is here that the end user has direct interaction with the data warehouse environment. The data marts are almost always oriented to the needs and requirements of the department. The data marts are requirements-driven in that they are shaped by the unique and specific needs of the department. The data marts are fed granular data from the enterprise data warehouse, which is then reshaped to meet the requirements of the department.

In some organizations there is more than one enterprise data warehouse. In these cases, data warehousing takes on a nontraditional form. The appearance of multiple enterprise data warehouses is a double-edged sword. On one hand, it means data warehousing has been successful. But on the other hand, multiple enterprise data warehouses present special challenges to the designer and user of the data warehouse. For example, it often happens that in a multiple enterprise data warehouse environment there is overlap of data between two or more enterprise data warehouses. Done properly, this overlap is very beneficial. Done improperly, the overlap can cause architecturally severe problems. Another feature of the corporate information factory is that of the exploration warehouse. This is a separate, specialized facility designed solely for the needs of the corporate explorer. With the exploration warehouse, the very ad hoc queries created by the explorer can be run with no damage to the normal warehouse processing.

And, finally, there is the near-line storage facility. This facility extends the capacity of the warehouse effectively to infinity. No longer is the designer constrained by the technological and economic limitations of disk storage. With near-line storage, the designer can take data down to the lowest level of detail desired—something that realistically is not a possibility with disk storage. This section focuses on the enterprise data warehouse and the data marts. In recent vintage, no other part of the warehouse environment has created as much confusion and as much controversy as these parts of the corporate information factory. The general architectural positioning of the data marts and the data warehouse is shown in Fig. 3.2, which shows how data marts are fed directly from the data warehouse.

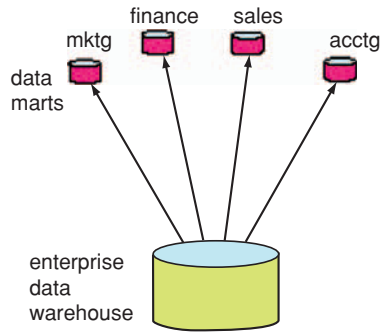

Fig. 3.2. The General structure of Data Marts and Data warehouse

The data warehouse exists to serve corporate functions, while the data marts exist to serve departmental functions. The departments that the data marts serve are typically finance, sales, marketing, and accounting. In other organizations, data marts are found in engineering, human resources, actuarial, and other places. Strictly speaking, the data marts shown in Figure 3.2 are called *dependent data marts* because they have data loaded into them ONLY from the enterprise data warehouse. Had the data marts been built independent of the enterprise data warehouse (and it is possible to build such structures) they would have been called *independent data marts*. But independent data marts are not part of the corporate information factory because there are some major architectural difficulties with them. The deficiencies of independent data marts will be discussed later in this section. Because of their deficiencies, independent data marts are not considered to be a part of the corporate information factory.

In general, enterprise data warehouses are larger than data marts. How much larger? Usually, a lot larger. There are several reasons why enterprise data warehouses are larger:

- enterprise data warehouse structures contain very detailed data; data mart structures contain aggregated and summarized data;
- enterprise data warehouse databases contain a generous amount of history; data marts contain a limited amount of history;
- enterprise data warehouse structures are very simple; data mart structures are much more complex; and,
- enterprise data warehouse structures are designed for unknown future usage; data mart structures are designed for specifically known requirements.

Figure 3.3 shows the differences between structures in the data warehouse and the data marts.

The data found in the enterprise data warehouse is for the entire corporation. The data found in the data marts is specific to the needs and the requirements of the department that owns the data mart. Because of

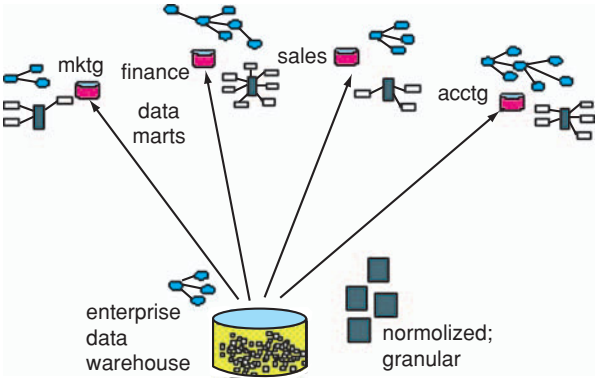

Fig. 3.3. The structure of data marts are requirements and can be called *star joins*

this basic difference, the data warehouse is "data driven" while the data marts are "requirements driven." The different types of structures found in the different environments reflect this very basic difference.

3.1.4 Different Data Structures

Data in the enterprise data warehouse is granular and simple. It needs to be this way because the enterprise data warehouse must support multiple and varied needs across the corporation. The data in the enterprise data warehouse can best be described as being normalized. For the most part, normalization fits the needs of the enterprise data warehouse quite well. However, the data found in the enterprise data warehouse is not perfectly normalized because if it were, it would not be efficient to access. The normalized data found in the enterprise data warehouse goes through a design process that can be described as *light denormalization*. Light denormalization of the enterprise data warehouse data restructures the normalized data into something less than a perfectly normalized structure. The kind of design considerations that make up light denormalization includes:

- creating a single place where multiple occurrences of data are stored when it is anticipated that the occurrences of data will normally and regularly be used together. For example, the data for the months January, February, March, . . . , December may be physically aggregated into a single location for ease of access;
- creating redundant data where it is known that the redundant data will be used regularly in conjunction with other data. For example, the description field for a part might be included with data from the bill of materials, materials requirements processing, the parts inventory file, and so forth;
- creating summarized data when the data is calculated once and used many times. For example, an employee's annual ending pay may be calculated and stored as its own field; and,

- further separating data where the record of data includes data fields where there is a great disparity of probability of access of the data. For example, a bank account record might contain date of account opened, account balance, and domicile of the account. Separating date of account opened and domicile of the account from account balance accomplishes the separation of elements of data with a wide is parity in the probability of usage.

After the data in the enterprise data warehouse goes through the process of light denormalization, it still retains its strongly normalized character. The light denormalization does not reflect the requirements of any one application or department. Instead, the light denormalization generally improves the speed and ease of access of ALL users and ALL departments.

The structure of the data mart is fundamentally different from the structure of data found in the enterprise data warehouse. Data marts contain structures of data that can best be described as snowflake structures or star join structures.

Multidimensional technology best supports these types of data structures. Star joins and snowflake structures are the result of detailed requirements' analysis for a department or a like-minded set of end users. Different departments collect their own unique requirements and the result is an aggregation of data that resembles a star or a snowflake (if we use our imagination a little!).

Different departments will have different designs for their stars or snowflakes because the requirements are different for each department. The data that is aggregated together for the star joins comes from the enterprise data warehouse. The granular data that is found at the enterprise data warehouse is gathered and reshaped to make the star and the snowflakes.

The enterprise data warehouse then provides the detailed foundation needed to create the many different flavors of stars and snowflakes. There are many ways by which the granular data found in the enterprise data warehouse is reshaped in order to create the stars and snowflakes. For example:

- Granular data is summarized
- Granular data is resequenced
- Granular data is aggregated based on data found at the granular level
- New key/foreign key relationships are found
- New fields are created as the granular data is brought together

Another aspect of the data mart/enterprise data warehouse environment is that of the data models that reflects the contents and structure inside each component. The data model that is made for the enterprise data warehouse is small and simple. The simplicity reflects the granular, normalized demeanor of the enterprise data warehouse. The data models for the data marts are much more elaborate. These data models are different for each data mart and are a reflection of the requirements that have been gathered for each data mart. There is a unique mapping from each data model to each other data model.

3.1.5 Different Users

But there is another perspective on the data mart/enterprise data warehouse environment that is relevant and has to do with the different users found in each of the environments. Figure 3.4 shows two types of users who are of interest: "farmers" and "explorers."

Farmers are those users who are the predictable, regular users of the data warehouse. Farmers know what they want before they ever submit a query. Farmers submit small queries, but they submit them frequently throughout the day. Farmers usually find what they are looking for. Farmers find small flakes of gold frequently, but seldom find huge nuggets. Farmers know what their requirements are before they start out on a pursuit of information. There usually are lots of farmers.

Explorers are the corporations' "out of the box" thinkers. Explorers are very unpredictable. Explorers may go six months submitting no queries, then may submit six queries in the same day. The queries that the explorers submit are very large. The explorers look for patterns of activity that can be exploited by the corporation. The explorers look for relationships between types of data that have never before been examined.

Explorers often find nothing for their efforts. But occasionally an explorer makes a huge and significant discovery that has been previously unexpected and has been overlooked by everyone else in the corporation. Explorers find huge nuggets of information when they find anything at all.

There are reasons why the queries the explorers submit are so large. The first reason for large queries is that explorers need details. If the explorers are going to look in-depth at data that no one else has examined, then the explorers must have detail. The second reason why the explorers' queries are lengthy is because the explorers look at large amounts of history. The most current information a corporation has may or may not prove to be of use to the explorers. In order to do a thorough analysis and to detect subtle patterns, the explorers must look at information over a long period of time.

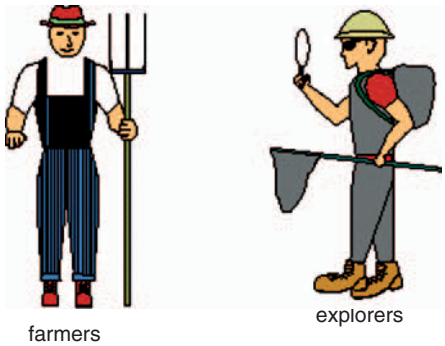

Fig. 3.4. The users in the DSS community

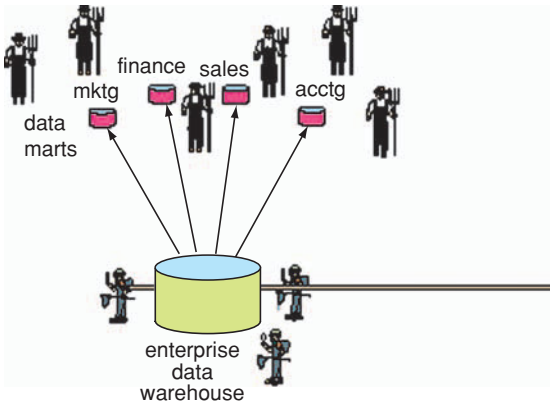

Fig. 3.5. Farmers are attracted to the data mart environment while explorers are attracted to the enterprise data warehouse

The third reason why the explorers submit lengthy queries is because the explorers look at data differently from other people. This means the data needs to be restructured. Restructuring data for the purpose of the satisfaction of a query implies that joins of data must be done—a lot of joins. So it is no accident that explorers submit what is termed the *72-hour query* or the *query from hell*. When we stop to consider the elements of the queries submitted by the explorers, we find that: *detail x history x fourteen way joins* adds up to large queries.

There is a peculiar affinity of farmers for data marts and explorers for the enterprise data warehouse, as seen in Fig. 3.5.

That farmers should find data marts attractive is no surprise. Data marts are shaped around predetermined requirements. Indeed, without requirements we cannot cogently design and build a data mart. And who are the users who know what their requirements are before they see a data warehouse database? Clearly, those are the farmers. On the other hand, which kinds of users have no idea going into an exploration of what their requirements are? These are the explorers of the world. An enterprise data warehouse is something that is versatile, if nothing else. The versatility of the enterprise data warehouse is just what the explorers are looking for. There is, therefore, a strong affinity of farmers for the data mart and of explorers for the data warehouse.

Farmers operate almost exclusively on data mart data while explorers operate on enterprise data warehouse data. Given that data marts are very requirements-driven, the attraction of the farmer community to them is very understandable. And it is likewise easy to understand the attraction of the explorer community to the enterprise data warehouse. But multiple data marts are not the only type of structure serviced by the enterprise data warehouse. Figure 3.6 shows that other types of architectural components found in the

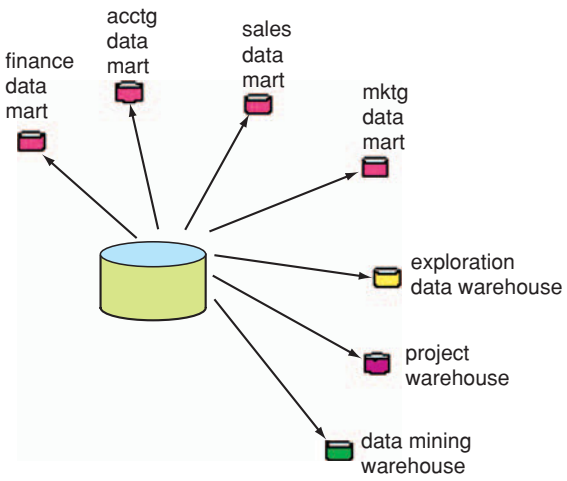

Fig. 3.6. The enterprise data warehouse needs to have a granular, corporate perspective because it supports many different kinds of DSS efforts

corporate information factory are serviced by the enterprise data warehouse as well.

The enterprise data warehouse serves as a basis for exploration warehouses, project warehouses, and data mining warehouses. Each of these types of data warehouses finds the granular data that resides in the enterprise data warehouse to be particularly attractive. One very important aspect to the notion of the enterprise data warehouse serving different communities is that there is a "single version of the truth." When the architecture looks like the one shown in Fig. 3.6, there is always a basis for reconcilability if any two analytical communities have a disagreement. In fact, the analysis created by different analytical communities will differ over time. However, when the enterprise data warehouse exists, there is the opportunity to reconcile the figures that are in disagreement. However, if the different analytical communities do not have a single basis of truth from which to operate, then there is almost no chance that there will be a reconciliation should a difference of opinion arise.

3.1.6 Technological Foundation

Not surprisingly, the world of data marts is served by different technology than the world of enterprise data warehouses. Figure 3.7 shows some of the technological considerations of each of these worlds.

The world of data marts is an on-line analytical processing (OLAP) multidimensional world. The volumes of data found in the data marts allow data to be accessed in a very flexible manner. In addition, relational technology is sometimes used for data marts when the data mart is large and/or of a general purpose nature. The essence of the data mart is flexibility and elegance of

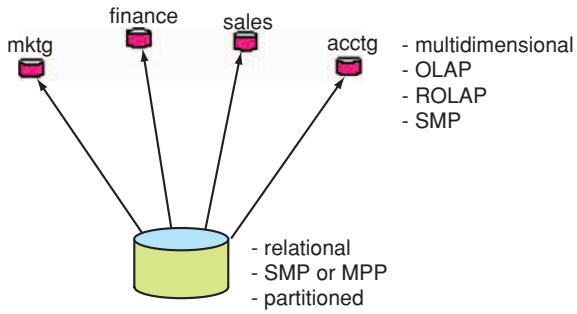

Fig. 3.7. The technologies found in the data mart are different from the technologies found in the enterprise data warehouse

presentation. To many end users, the data mart IS the data warehouse. Symmetric multiprocessor (SMP) architecture is often useful for the data mart environment because of its extreme flexibility and because of the ability to economically handle the volumes of data found in the data mart.

3.1.7 Data Warehouse

The enterprise data warehouse environment is substantively different from the data mart environment when it comes to technology. The enterprise data warehouse environment operates against very large amounts of data. Therefore, the number one priority of the database management system (DBMS) and hardware platform found at the enterprise data warehouse level is the ability to handle industrial strength amounts of data. The ability must be both economical and technological. In the data mart environment, the emphasis was on elegance of presentation of data and flexibility. It is a mistake to choose the enterprise data warehouse platforms based on the same criteria. The enterprise data warehouse platforms are typically relational with an emphasis on a massively parallel processing (MPP) hardware architecture. This combination allows the technology to be scalable up to very large volumes of data.

3.1.8 Informix Architecture

Informix offers an interesting and powerful architectural solution to the technology required for the data mart/enterprise data warehouse environment. Figure 3.8 shows this architecture.

The Informix relational DBMS provides a very solid foundation for the enterprise data warehouse. It offers the ability to manage very large amounts of data while providing the performance and scalability required by a truly enterprise-class data warehouse. Users should choose Informix as the foundation of their enterprise data warehouse principally because of the soundness of

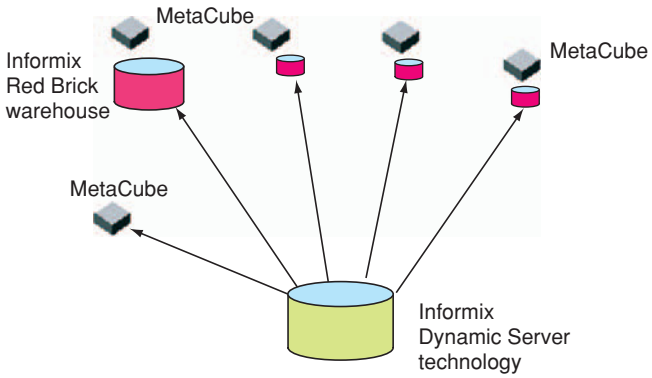

Fig. 3.8. The Informix approach to the data mart/data warehouse infrastructure

its parallel, shared-nothing relational architecture—required for highly scalable database configurations.

Informix Red Brick® Warehouse provides an ideal solution for the industrial strength data mart by virtue of its underlying star join architecture. Informix Red Brick Warehouse allows requirements to be defined then optimized in terms of design. But not only does Informix Red Brick Warehouse offer the ability to accommodate star joins and snowflake structures, but it also allows those structures to be managed on a grand scale. In this regard, Informix Red Brick Warehouse surpasses the capacity capabilities of any of the multidimensional vendors. And, finally, Informix MetaCube® rounds out the Informix offering by providing a robust solution for smaller-scale data marts that do not require all the capabilities provided by Informix Red Brick Warehouse.

The Informix solution has technology that is optimal for each component of the data mart/enterprise data warehouse environment. This is in contrast to its competitors who try to stretch a single technology into a "one-size-fits-all" solution. In fact, different organizations will need different combinations of the capabilities offered by the Informix DBMS, Informix Red Brick Warehouse, and MetaCube solutions. By having an architecturally diverse offering, Informix can provide true mix-and-match capabilities to its users. Supplying this kind of flexibility and complementary technology enables Informix to deliver an optimal solution for any organization's needs.

3.1.9 Building the Data Warehouse/Data Mart Environment

The data mart architecture/enterprise data warehouse architecture that has been described can be built in many ways. The classical way to build the architecture is to build the enterprise data warehouse first, then to build the data marts next. Figure 3.9 depicts this classical approach.

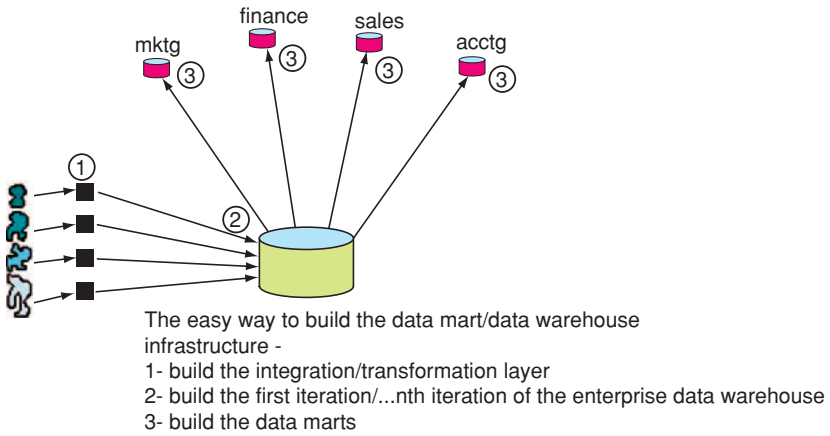

Fig. 3.9. The easy way to build the data mart/data warehouse infrastructure

In the classical approach, the integration and transformation programs are first built. Once built data is loaded into the enterprise data warehouse from the legacy applications, then the data marts are built. There is an interesting aspect to the pattern of development shown in Fig. 3.9, and that is the enterprise data warehouse is built iteratively. First, one part of the enterprise data warehouse is built, then another part of the enterprise data warehouse is populated. The enterprise data warehouse is not built in an all-at-once manner. This means the data marts can be built very quickly after the enterprise data warehouse is first populated. And there is nothing to say that farmers cannot use the enterprise data warehouse directly after it is built and before the data marts are built. As long as the data is in the enterprise data warehouse, and as long as there are not too many users, there is nothing to say the farmers cannot start their analysis on data found in the enterprise data warehouse. It is only after the enterprise data warehouse has grown and the enterprise data warehouse contains a lot of data that the farmers find the data mart attractive.

The development pattern shown in Fig. 3.9 is very straightforward. It is well documented in various methodologies for data warehouse development. But there are those who would argue that building an enterprise data warehouse and data marts, as shown in Fig. 3.9, is not a good approach. The arguments used by opponents of the classical approach are:

- the enterprise data warehouse is big and requires a lot of resources;
- the enterprise data warehouse will take two to five years to build;
- management wants immediate feedback from the data warehouse effort and does not have time to “do it right”; and,
- an enterprise data warehouse requires integration, and integration is not easy to do, and so forth.

Many reasons are given by opponents of the enterprise data warehouse for not building the data warehouse properly. Among those who do not necessarily advocate this approach are the data mart vendors who often see the enterprise data warehouse as a threat to the building of data marts. The perspective of some data mart vendors is that if they have to wait for customers to build the enterprise data warehouse then their data marts will not be built quickly. So the data mart vendor will often discourage a customer from taking an enterprise data warehouse approach in favor of building one or more data marts. What approach to development does the data mart advocate propose? The "data mart first" approach is illustrated in Fig. 3.10.

In the first scenario, a data mart is created directly from the legacy environment. At the end of the creation of the first data mart, the customer does not experience any pain. Instead, the customer has a functional data mart and is quite happy with it. In fact, the customer is so happy that another data mart is quickly built. Now the customer notices that each data mart needs the same detailed data as each other data mart. The customer remembers that redundancy of data was not a good thing. But the second data mart brings on a lot of happiness, so no one is particularly bothered by the redundancy of data between two data marts. Soon another data mart is on its way. When the third data mart arrives, it is noticed that the redundancy of detailed data between the data marts continues to increase. This has a very negative effect on the amount of money spent on each data mart because each data mart is quite large. But there is another ill side effect and that is that there is no reconcilability of data among data marts. When management asks how much revenue the company made, each department uses its data mart to produce a completely different value. The company now has stove pipe DSS applications with no integration among them. But the progress of data marts continues. A new data mart is built. Now the organization has four data marts. There is massive redundancy of detailed data among each data mart. This means that the cost of each data mart is high since each data mart is quite large. And there is no reconcilability of data among the data marts. When management asks the new data mart how much revenue was made, the answer is yet another un-integrated, irreconcilable opinion. And now someone notices that the interface between the legacy applications and the different data marts is starting to grow into its own nightmare.

The number of programs that must be built and maintained, the competition for resources in accessing the legacy data, and the inability to coordinate data among different data marts becomes unbearable. Someone stops and shouts, "This is not what data warehousing is all about!" And, indeed, the architecture described is not what data warehousing is all about. What must be done to solve the problems shown in Fig. 3.10 is to build an enterprise data warehouse, and it must become the source of data for the different data marts. So the data mart approach ends up at precisely where the data mart vendors had said a corporation did not want to be. The "data mart first" approach

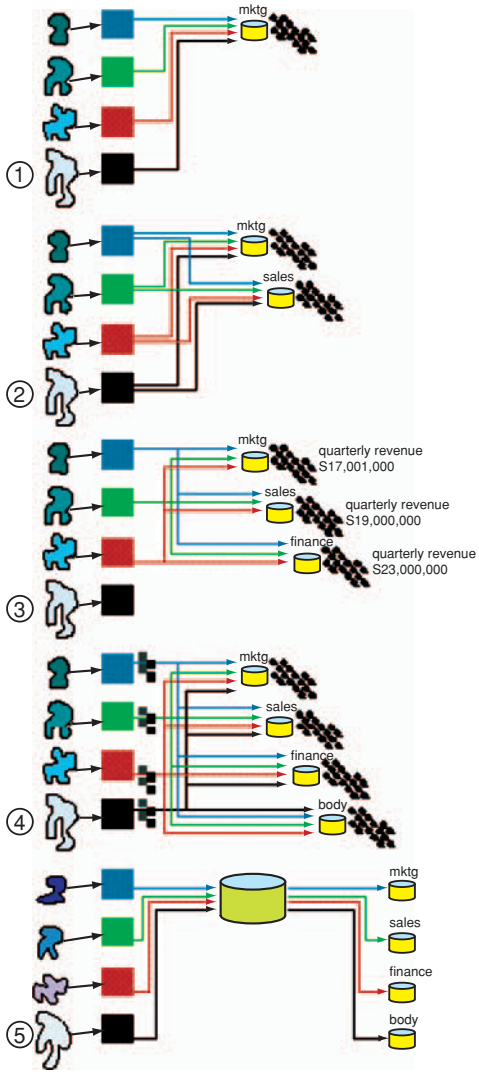

Fig. 3.10. The "long way around" to the building of the data mart/data warehouse infrastructure

caused the customer to take the long way around. There are two very different ways to arrive at the same place.

3.1.10 History

One of the major distinguishing characteristics of the data mart and the enterprise data warehouse environment is that of the amount of history stored

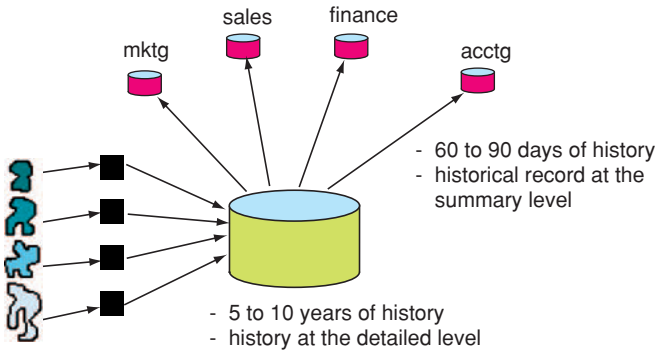

Fig. 3.11. Another major difference between the data mart and the data warehouse is in the amount history found in data mart and data warehouse

in the different environments. Figure 3.11 illustrates the differences in history as it is stored in the architectural components.

Figure 3.11 shows that the enterprise data warehouse contains a robust amount of historical data. Five to 10 years (or even longer), depending on the application, is the norm for the enterprise data warehouse. Furthermore, in order to be effective, the historical data that is stored in the enterprise data warehouse is stored at the detailed level. On the other hand, the data mart environment contains relatively little historical data. The data mart may contain from 60 to 90 days worth of data. On occasion, at a summary level, the data mart will contain more historical data. Some data marts will contain much more history than other data marts. But there simply is no comparison to the amount of history contained in the data marts and the enterprise data warehouse.

3.1.11 Nondetailed Data in the Enterprise Data Warehouse

The enterprise data warehouse can contain data that is nondetailed on occasion. As a rule, the nondetailed data found in the enterprise data warehouse makes up a very small percentage of the data found there.

Figure 3.12 shows that the enterprise data warehouse can contain corporately used, nondetailed data. In this case, the enterprise data warehouse serves as a convenient central location for the corporate data. The corporate data can take on any flavor and form. Another related type of data sometimes found in the enterprise data warehouse is public summary data or data created by a department and made available to the general public. Again, the enterprise data warehouse serves in the capacity as a convenient central location for such data. And, finally, a corporation may have benchmark data that is used widely across the corporation. The corporation will have to keep the benchmark data somewhere, and the enterprise data warehouse is a good place.

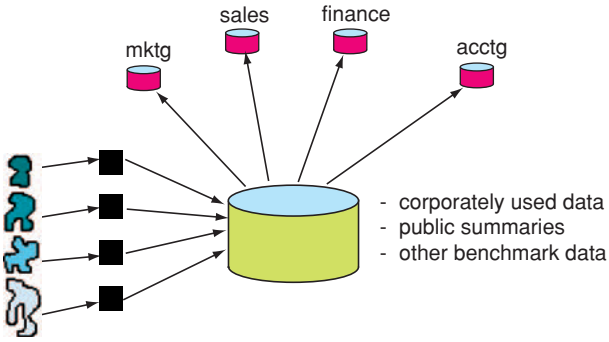

Fig. 3.12. Data other than detailed data is found in the enterprise data warehouse

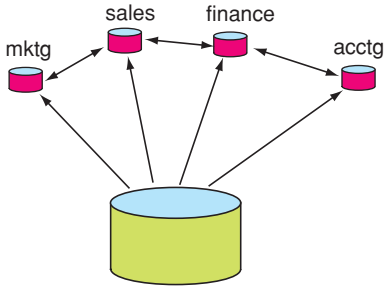

Fig. 3.13. The direct sharing of data among the departmental data marts is not a good idea

3.1.12 Sharing Data Among Data Marts

In the same vein, the question arises as to whether departments should directly share data among themselves. Figure 3.13 indicates that departments have the opportunity to share data.

The scenario shown in Fig. 3.13 is not a good idea. If there is to be sharing of data among departments, the sharing needs to be done by placing the data to be shared in the enterprise data warehouse, then allowing access from there. The scenario shown in Fig. 3.13 is the start of what has popularly become known as the "spider web" scenario. In the spider web scenario, data is passed in an almost random manner among different analysts. Soon data loses its validity and its meaning.

3.1.13 The Manufacturing Process

As data passes from the enterprise data warehouse to the data marts, the data undergoes what can be termed a *manufacturing* process. Figure 3.14 shows this process.

- the selection of data from the enterprise data warehouse;
- the aggregation of disparate data into a single structure, usually a star join or a snowflake structure;

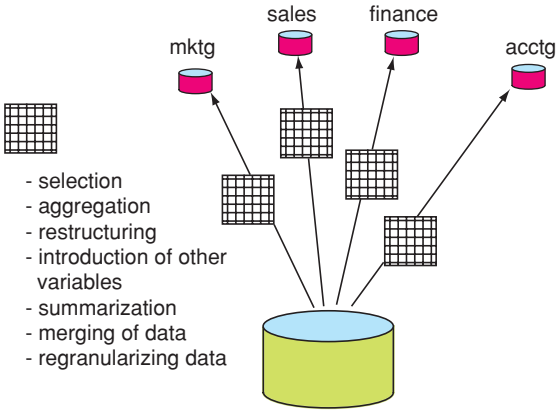

Figure 14 : There is an essen - tial manufacturing process that occurs as data goes from its granular state in the enterprise data warehouse to its refined state in the data mart.

Fig. 3.14. shows that data coming out of the enterprise data warehouse goes through a complex and tedious process. The manufacturing process includes such things as:

- the restructuring of data to meet the needs of the data mart users (such as the restructuring of data into a multidimensional format);
- the creation and introduction of other variables. For example, as data is being read into the data mart, there may be two variables: Product A revenue and Product B revenue. The database designer creates a new variable—Combined Revenue—by adding Product A and Product B revenue together. There may be many reasons for the creation of such a variable. Ease of analysis once the data mart is populated is probably the primary reason for the creation of new variables at the moment of manufacture;
- summarization of detailed data;
- merging of data as it passes into the data mart;
- granularizing data to meet the needs of the data mart user, and so forth.

In short, there is a lot of work that takes place as data moves from the enterprise data warehouse to the data marts. Furthermore, the manufacturing done for one data mart is almost always different from the manufacturing done for another data mart. It is absolutely normal, however, for there to be some amount of overlap from one data mart to the next. The same data element, the same calculation, the same definition often appear across the different data marts that a corporation creates.

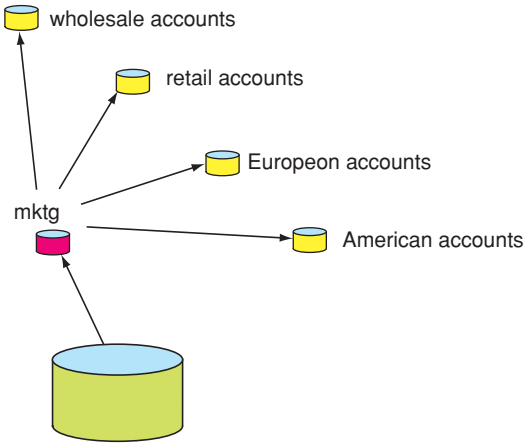

Fig. 3.15. For a large data mart, there are often subdata marts

3.1.14 Subdata Marts

On occasion a corporation will have a very large, general-purpose data mart. On these occasions there may start to appear what can be termed *subdata marts*. A subdata mart is one that emanates from a data mart.

Figure 3.15 shows how a marketing data mart has subdivided itself into several "subdata marts," one for wholesale accounts, one for retail accounts, one for European accounts, and one for American accounts. The subdata marts take on a life of their own, much like the data marts did as they sprang from the enterprise data warehouse. The subdata marts may or may not:

- share the same technology as the data mart;
- coordinate their data with other sub data marts;
- participate in a common infrastructure, such as a common metadata infrastructure, and so forth.

Subdata marts are neither good nor bad, but are simply a reality that must be dealt with. Subdata marts have the same problems as data marts, including integrity of data, cost, and performance. But, in a way, subdata marts are a sign of success since they indicate that users have adopted the concept of data marts to the point where they want to create their own "mini" version.

3.1.15 Refreshment Cycles

One of the marked differences between data marts and the enterprise data warehouse is in the refreshment cycles in which each participates. Figure 3.16 shows the differences between the two environments.

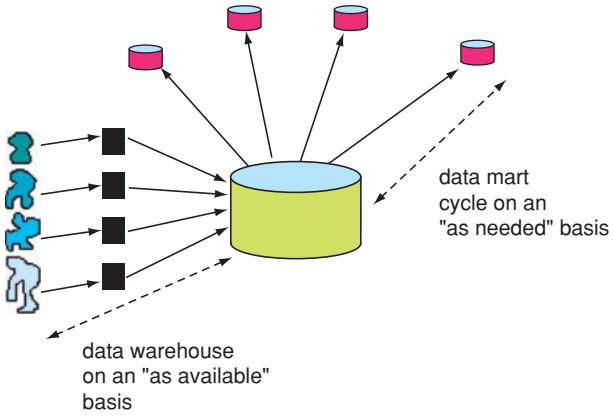

Fig. 3.16. Two very different refreshment cycles found in the data mart/data warehouse infrastructure

Figure 3.16 shows that the refreshment cycle for the data warehouse is relaxed and is "availability" driven. Availability driven refers to the movement of data into the enterprise data warehouse once the data is available. In other words, when data is transacted in the legacy environment, it is moved to the enterprise data warehouse. The enterprise data warehouse needs fresh data because it serves the needs of many types of DSS communities. However, the data refreshment in the data mart environment is "need" driven. Need driven refers to the fact that refreshment of the data mart is not made until the data is needed in the data mart.

In general, this means that the data mart schedule of refreshment is not done nearly as often as the refreshment of the enterprise data warehouse. One of the most important aspects of the data mart/data warehouse environment is that of refreshment. The amount of data needed to be refreshed, the integration and transformation that occurs during the refreshment process, the profound effect that the cycle of refreshment has on the usability of data and ultimately on the analysis that can be done on the data found in the data mart/data warehouse environment all lead to the conclusion that refreshment needs to be taken seriously as a cornerstone of the corporate information factory.

3.1.16 External Data

External data is data whose origin is outside of the corporation's systems. External data can be some of the most useful data that a corporation has. For example, it is often useful for a company to include general industry sales to be compared to internal corporate sales. The executive can sit back and compare the trends and the realities of corporate sales to the entire industry.

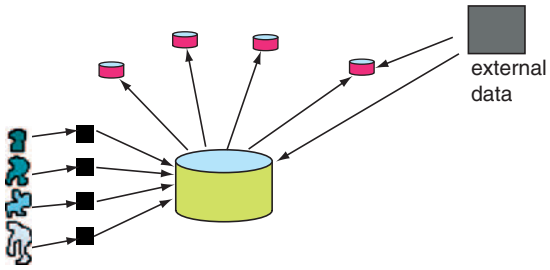

Fig. 3.17. External data is data that can be blended with the data mart or with the enterprise data warehouse

Such a benchmark is a good indication of how good or bad a corporation is doing.

As another example of the inclusion of external data with internal data, demographic growth is often useful to compare to sales or other corporate measures. Demographic changes often help to precisely address the issue of market penetration. Comparing internal results to demographic changes can lead to insights otherwise not obvious. And external data can be entered into either the data mart environment or the enterprise data warehouse environment. Figure 3.17 shows the possibilities for external data.

When external data is applicable and useful to only a single department, it can be entered into a data mart. But when external data has applicability or usefulness over more than one department, then the external data needs to be entered into the enterprise data warehouse. External data has some unique characteristics. The first is that the external data must be accepted on its face value. External data, once acquired by the corporation, is not easily altered. In the same vein, external data must have its key structure made compatible with the key structures found in the enterprise data warehouse and/or the data marts. If the external data is allowed to have different key structures, then no meaningful comparison can be made between the external data and the data found in the DSS environment.

3.1.17 Operational Data Stores (ODS) and Data Marts

Not all data goes from the legacy environment to the enterprise data warehouse, then on to the data marts. Some data makes a detour through the ODS. In particular, data needed for integrated online, real-time processing passes from the legacy environment to the ODS. Then, after it has served its purpose, it passes into the data warehouse.

Figure 3.18 shows that some data is routed through the ODS to the enterprise data warehouse. This data can ultimately arrive in the data mart environment. The data mart does not make a distinction between data that does or does not pass through the ODS. However, data never bypasses the enterprise data warehouse and goes directly to the data mart from the ODS.

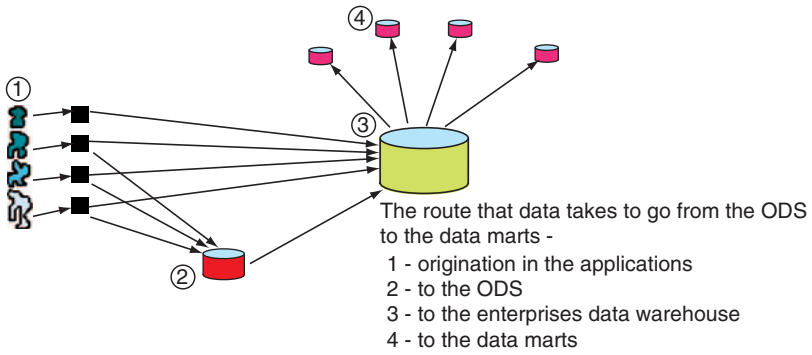

Fig. 3.18. The route that data takes to go from the application environment through ODS to the data marts

To omit placing the ODS data in the enterprise data warehouse would mean robbing the architecture of its integrity. Such a movement is the first step toward spider web systems.

3.1.18 Distributed Metadata

As we have already seen, the data mart environment and the enterprise data warehouse environment are two very different architectural structures that are often supported by different technologies. If there is to be cohesion—true integration of data and systems across the environments—how exactly is this cohesion achieved? The best way to achieve an enterprise-wide perspective is to employ metadata. Metadata is the facility by which different architectural components can exchange information.

What is metadata? The old definition of metadata is data about data. But that definition is so broad as to be almost meaningless. Instead, a more practical definition of metadata is that it is the information about the contents of systems, rather than the actual contents of the systems themselves. As a simple example, what does the number 458 represent in an inventory system? The answer is that the actual content does not mean much until you know that you are talking about 458 canisters. In this case, 458 is the content and "canisters" is the metadata that gives meaning to the content. Metadata in years past has always had a centralized orientation for the following reasons:

- metadata began in the days of the mainframe, where everything was centralized; and,
- the notion that in order for data to be shared, there needed to be a single, centralized definition and description of the terms of metadata.

But metadata in the world of data marts and the enterprise data warehouse is quite different. In this world, everything is distributed, not centralized.

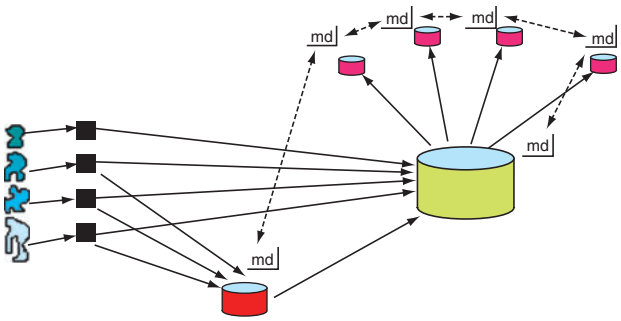

Fig. 3.19. Distributed metadata provides cohesion for the data mart/enterprise data warehouse infrastructure

Therefore, a centralized approach to metadata is not appropriate to the world of the data warehouse.

Figure 3.19 depicts a simplified view of distributed metadata and shows that metadata resides at every data mart and at the enterprise data warehouse. The first condition of distributed metadata is that there be local control and management of the metadata. What does local control and management of metadata mean? It means that each data mart and the enterprise data warehouse can:

- create, delete, and alter its own metadata;
- determine who can and cannot have access to the metadata;
- specify that metadata is private to only the data mart or enterprise data warehouse at hand;
- expect the metadata that is managed by the data mart or the enterprise data warehouse to reside in the technology that is used. For example, if one data mart were in Informix, then the metadata for that data mart would exist as an Informix table. And if another data mart had Teradata as its DBMS, then the metadata for that data mart would reside in the data mart as a Teradata table, and so forth.

The first condition then for distributed metadata is that it resides in the local architectural construct (i.e., the data mart or the enterprise data warehouse). But merely distributing metadata across the enterprise is not sufficient. All this distribution accomplishes is the solidification of the stove pipes that separate data marts and the enterprise data warehouse. The next condition of distributed metadata is the condition that metaobjects be able to be distributed across data marts and the enterprise data warehouse. This means that one data mart can easily share metadata with another data mart. Or the enterprise data warehouse can share metadata with another data mart.

As an example of the ability to share distributed metadata, consider two data marts, one in Informix and one in DB2. The Informix data mart receives a request for metadata that it manages and which resides in the data mart as

an Informix table. The request is received and the Informix data mart sees no reason why the metadata cannot be shared. The metadata that is requested is sought and located. The metadata is then sent to the DB2 data mart. By the time the metadata arrives at the DB2 data mart, the metadata has been converted into a DB2 format so that the DB2 data mart can make sense of it.

Of course, the exchange is a two-way street. When the Informix data mart requests DB2-managed metadata, the metadata is converted into an Informix format by the time it arrives at the Informix data mart. So the second condition for distributed metadata is that of metaexchange among the different architectural constructs (i.e., the data marts and the enterprise data warehouse). But there is a third important condition: namely that there be integrity of metadata across the enterprise. This principle can be expressed as the "system of record for metadata," and can be stated as follows:

- every unit of metadata can have one and only one owner across the enterprise.

In other words, if an Informix data mart owns a unit of metadata, then no other data mart or enterprise data warehouse (or anything else, for that matter) can own the unit of metadata. This means that as metadata is added to the system and before the addition can be confirmed, the distributed metadata manager must check to see if the metadata is owned elsewhere. If the metadata is owned elsewhere, ownership cannot be assigned again. Only if the metadata is not previously owned can the metadata be assigned to the new owner. Of course, the ownership of metadata can be reassigned. But a reassignment entails organizational and political decisions, where a consensus of opinion needs to be reached.

Owning metadata implies that the organization that owns it has exclusive rights to the creation, deletion, and alteration of the metadata, as well as the right to determine who can and cannot access the metadata. Once the system of record across the enterprise is introduced, sharability of metadata with integrity becomes a reality. As a unit of metadata is passed around the enterprise, the record of the ownership is carried with the metadata to clearly mark the metadata as being shared.

3.1.19 Managing the Warehouse Environment

Another interesting aspect of the data mart/enterprise data warehouse environment is that of the ongoing management and administration of the environment. Figure 3.20 illustrates a monitor that is used in the management of the environment.

The activity that passes into the data warehouse and the data marts is of keen interest to the data warehouse administrator. With an activity monitor, the data warehouse administrator can tell such important things as:

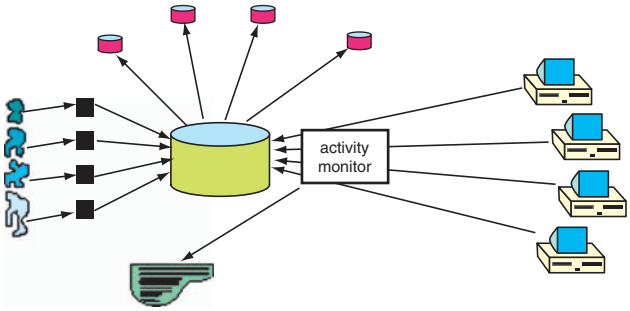

Fig. 3.20. Monitoring the enterprise data warehouse is an important part of the infrastructure

- who is using the warehouse;
- what queries are the most common;
- what resources are being used the most; and,
- when the warehouse is being used.

There are other important aspects of monitoring as well, such as the determination of what data is and is not being used. This determination is most important when considering what data is safe to be purged from the warehouse. If an organization does not take seriously the purging of data from the warehouse, then several very bad (and totally unnecessary!) things can happen:

- the cost of the warehouse goes sky high;
- the performance in the warehouse grows worse, as data that is used hides behind masses of data that is not being used; and,
- the designer must increasingly aggregate and summarize data in order to accommodate the small amount of data available for new data in the data mart or enterprise data warehouse.

Multidimensional, OLAP technology is quite suitable for the data mart environment, while industrial strength, scalable technology does well for the enterprise data warehouse environment. There are two ways to build the enterprise data warehouse/data mart environment. Build the enterprise data warehouse iteratively, then extend the data marts from the enterprise data warehouse, or build the data marts first, then go back and build the enterprise data warehouse and extend the data marts from the enterprise data warehouse. In the long run, building the enterprise data warehouse is the least expensive and the cleanest approach.

However, some organizations take what they believe to be a more expedient approach and build data marts first. In the long run, they use more development resources than if they had built the data warehouse first,

although it is not obvious that that is going to be the case until well into the DSS exercise.

Data marts and the enterprise data warehouse have different characteristics, the most important distinctions being that:

- the enterprise data warehouse contains a robust amount of history while the data marts contain a limited amount of history;
- the data marts are smaller than the enterprise data warehouse; and,
- the enterprise data warehouse contains other than granular data, such as corporate summary data.

There is an essential "manufacturing process" that occurs as data moves from the enterprise data warehouse to the data marts. The manufacturing process reshapes the corporate, granular data found in the enterprise data warehouse into a requirements-driven departmental structure. In addition, "subdata marts" are often developed off of large, actively used data marts. The subdata marts experience all the same problems and challenges that the data marts do. There is a very different refreshment cycle for the enterprise data warehouse and data marts. Data marts are refreshed on the basis of need by the departments that own the data mart. Data warehouses are generally refreshed on the basis of availability of new data ready to enter the data warehouse. External data can be loaded into either the data marts or the enterprise data warehouse, depending on the usage of the external data. The comparison of external data to internal data can lead to very important insights at the executive level.

Distributed metadata is required for the management of the distributed data mart/enterprise data warehouse environment. Informix has an architectural edge over the competition in that they have specific and unique products to fit the architecture. The Informix DBMS is well suited for the needs of the data warehouse. Informix Red Brick Warehouse is well suited for large, industrial strength data marts. And MetaCube is positioned for start up and casual data marts. No other vendor has technology that is unique and optimal across the architecture of the corporate information factory.

3.1.20 OLAP

The purpose of the section that follows is to define On-Line Analytical Processing (OLAP), who uses it and why, and to review the key features required for OLAP software as referenced in the OLAP Council benchmark specification.

What Is OLAP?

During the last ten years, a significant percentage of corporate data has migrated to relational databases. Relational databases have been used heavily in the areas of operations and control, with a particular emphasis on transaction

processing (for example, manufacturing process control, brokerage trading). To be successful in this arena, relational database vendors place a premium on the highly efficient execution of a large number of small transactions and near fault tolerant availability of data.

More recently, relational database vendors have also sold their databases as tools for building data warehouses. A data warehouse stores tactical information that answers "who?" and "what?" questions about past events. A typical query submitted to a data warehouse is: "What was the total revenue for the eastern region in the third quarter?"

It is important to distinguish the capabilities of a data warehouse from those of an OLAP (On-Line Analytical Processing) system. In contrast to a data warehouse, which is usually based on relational technology, OLAP uses a multidimensional view of aggregate data to provide quick access to strategic information for further analysis.

OLAP enables analysts, managers, and executives to gain insight into data through fast, consistent, interactive access to a wide variety of possible views of information. OLAP transforms raw data so that it reflects the real dimensionality of the enterprise as understood by the user.

While OLAP systems have the ability to answer "who?" and "what?" questions, it is their ability to answer "what if?" and "why?" that sets them apart from data warehouses. OLAP enables decision making about future actions. A typical OLAP calculation is more complex than simply summing data, for example: "What would be the effect on soft drink costs to distributors if syrup prices went up by $.10/gallon and transportation costs went down by $.05/mile?" OLAP and data warehouses are complementary. A data warehouse stores and manages data. OLAP transforms data warehouse data into strategic information. OLAP ranges from basic navigation and browsing (often known as "slice and dice"), to calculations, to more serious analyses such as time series and complex modeling. As decision makers exercise more advanced OLAP capabilities, they move from data access to information to knowledge.

Who Uses OLAP and Why?

OLAP applications span a variety of organizational functions. Finance departments use OLAP for applications such as budgeting, activity-based costing (allocations), financial performance analysis, and financial modeling. Sales analysis and forecasting are two of the OLAP applications found in sales departments. Among other applications, marketing departments use OLAP for market research analysis, sales forecasting, promotions analysis, customer analysis, and market/customer segmentation. Typical manufacturing OLAP applications include production planning and defect analysis.

Important to all of the above applications is the ability to provide managers with the information they need to make effective decisions about an organization's strategic directions. The key indicator of a successful OLAP

application is its ability to provide information as needed, i.e., its ability to provide "just-in-time" information for effective decision making. This requires more than a base level of detailed data.

Just-in-time information is computed data that usually reflects complex relationships and is often calculated on the fly. Analyzing and modeling complex relationships are practical only if response times are consistently short. In addition, because the nature of data relationships may not be known in advance, the data model must be flexible. A truly flexible data model ensures that OLAP systems can respond to changing business requirements as needed for effective decision making.

Although OLAP applications are found in widely divergent functional areas, they all require the following key features:

- Multidimensional views of data
- Calculation-intensive capabilities
- Time intelligence

Multidimensional Views

Multidimensional views are inherently representative of an actual business model. Rarely is a business model limited to fewer than three dimensions. Managers typically look at financial data by scenario (for example, actual vs. budget), organization, line items, and time; and at sales data by product, geography, channel, and time.

A multidimensional view of data provides more than the ability to "slice and dice"; it provides the foundation for analytical processing through flexible access to information. Database design should not prejudice those operations that can be performed on a dimension or how rapidly those operations are performed. Managers must be able to analyze data across any dimension, at any level of aggregation, with equal functionality and ease. OLAP software should support these views of data in a natural and responsive fashion, insulating users of the information from complex query syntax. After all, managers do not have to understand complex table layouts, elaborate table joins, and summary tables.

Whether a request is for the weekly sales of a product across all geographical areas or the year-to-date sales in a city across all products, an OLAP system must have consistent response times. Managers should not be penalized for the complexity of their queries in either the effort required to form a query or the amount of time required to receive an answer.

The OLAP Council APB-1 performance benchmark tests a server's ability to provide a multidimensional view of data by requiring queries of varying complexity and scope. Basic aggregation is performed on some of the dimensions (product, customer, and channel). More complex calculations are performed on other dimensions. The measure dimension computes ratios and averages. Variances are computed along the scenario dimension. A complex

model based on historical performance is used to compute the forecast scenario. Consistently quick response times to these kinds of queries are key to establishing a server's ability to provide multidimensional views of information.

Complex Calculations

The real test of an OLAP database is its ability to perform complex calculations. OLAP databases must be able to do more than simple aggregation. While aggregation along a hierarchy is important, there is more to analysis than simple data roll-ups. Examples of more complex calculations include share calculations (percentage of total) and allocations (which use hierarchies from a top-down perspective).

Key performance indicators often require involved algebraic equations. Sales forecasting uses trend algorithms such as moving averages and percentage growth. Analyzing the sales and promotions of a given company and its competitors requires modeling complex relationships among the players. The real world is complicated – the ability to model complex relationships is key in analytical processing applications.

OLAP software must provide a rich tool kit of powerful yet succinct computational methods. To make developers more efficient and business users more self-sufficient, the vehicle for implementing computational methods should be clear and nonprocedural. If the method for creating the desired calculations is not clear, development time and/or usage will suffer. If the calculation method is procedural, changes to the system cannot be done in a timely manner, effectively eliminating access to just-in-time information.

Whereas transaction processing systems are judged on their ability to collect and manage data, analytical processing systems are judged on their ability to create information from data. The OLAP Council APB-1 performance benchmark contains a representative selection of calculations, both simple and complex. An example of a simple calculation contained in the performance benchmark is the calculation of margin (sales minus costs). The computation of the forecast is the most complex calculation contained in the current version of the performance benchmark. Historical data is used to project the future and aggregate data is used to estimate input data. Other more complex calculations, such as allocations and trend analysis, are also often found in OLAP systems.

Time Intelligence

Time is an integral component of almost any analytical application. Time is a unique dimension because it is sequential in character (January always comes before February). True OLAP systems understand the sequential nature of time. Business performance is almost always judged over time, for example, this month vs. last month, this month vs. the same month last year.

The time hierarchy is not always used in the same manner as other hierarchies. For example, a manager might ask to see the sales for May or the sales for the first five months of 1995. The same manager might also ask to see the sales for blue shirts but would never ask to see the sales for the first five shirts. Concepts such as year-to-date and period-over-period comparisons must be easily defined in an OLAP system.

In addition, OLAP systems must understand the concept of balances over time. For example, if a company sold 10 shirts in January, five shirts in February, and 10 shirts in March, then the total balance sold for the quarter would be 25 shirts. If, on the other hand, a company had a head count of 10 employees in January, only five employees in February, and 10 employees again in March, what was the company's employee head count for the quarter? Most companies would use an average balance. In the case of cash, most companies use an ending balance.

The OLAP Council APB-1 performance benchmark contains examples of how time is used in OLAP applications. Smoothed sales are computed as a three-month moving average. Inventory is aggregated as an ending balance. The forecast calculation uses this year's vs. last year's knowledge, year-to-date knowledge, and annualization factors.

OLAP Benefits

Successful OLAP applications increase the productivity of business managers, developers, and whole organizations. The inherent flexibility of OLAP systems means business users of OLAP applications can become more self-sufficient. Managers are no longer dependent on IT to make schema changes, create joins, or worse. Perhaps more importantly, OLAP enables managers to model problems that would be impossible using less flexible systems with lengthy and inconsistent response times. More control and timely access to strategic information equal more effective decision making.

IT developers also benefit from using the right OLAP software. Although it is possible to build an OLAP system using software designed for transaction processing or data collection, it is certainly not a very efficient use of developer time. By using software specifically designed for OLAP, developers can deliver applications to business users faster, providing better service. Faster delivery of applications also reduces the applications backlog.

OLAP reduces the applications backlog still further by making business users self-sufficient enough to build their own models. However, unlike standalone departmental applications running on PC networks, OLAP applications are dependent on data warehouses and transaction processing systems to refresh their source level data. As a result, IT gains more self-sufficient users without relinquishing control over the integrity of the data.

IT also realizes more efficient operations through OLAP. By using software designed for OLAP, IT reduces the query drag and network traffic on transaction systems or the data warehouse.

Lastly, by providing the ability to model real business problems and a more efficient use of people resources, OLAP enables the organization as a whole to respond more quickly to market demands. Market responsiveness, in turn, often yields improved revenue and profitability.

3.2 Data Warehousing for Healthcare: The Greatest Weapon in your Competitive Arsenal

3.2.1 A Data Warehousing Perspective for Healthcare

The dramatic changes underway today in the healthcare industry are the result of ambitious industry and government efforts to curb sky-rocketing healthcare costs. Two key issues, cost containment and efficiency, will be the cornerstones of the healthcare industry's new infrastructure. To contain costs, payers (i.e., insurance companies, employers, and government agencies) are turning to health maintenance organizations (HMOs), preferred provider organizations (PPOs) and other managed care entities that have demonstrated the ability to cut costs. Through efficiencies that rely on economies of scale, tough bargaining methods, and the elimination of unnecessary services, these organizations have become the care providers of choice for government and private industry. As early as 1992, roughly half of all workers in the United States were enrolled in managed care plans, up from 27% five years before, according to a study by Foster & Higgins. It is no wonder in the age of cost control, capitated payments and the need for outcomes research (monitoring of clinical procedures to help determine the most clinically effective and cost effective treatments), senior executives are now looking toward information technology (IT) to give them the strategic/competitive advantage within the marketplace.

Data warehousing is a management tool that enables executives to access the information they need to make informed business decisions. A data warehousing system pulls together information from disparate sources to construct an integrated view of business activities as shown in Fig. 3.21. These systems then transform the data into a consistent, easily accessible format and distribute the data to where it is needed for decision making. Although the actual data warehouse itself may be the responsibility of the IT department, the project should be driven by business management because a data warehouse is only as good as its ability to guide management in making strategic and informed business decisions. Business users can then efficiently analyze vast quantities of data, uncover hidden opportunities, and enhance their organizations' competitive advantage.

3.2.2 Adding Value to your Current Data

Today's healthcare organizations gain competitive advantage by making smart decisions based on complete information. Developing the best method to man-

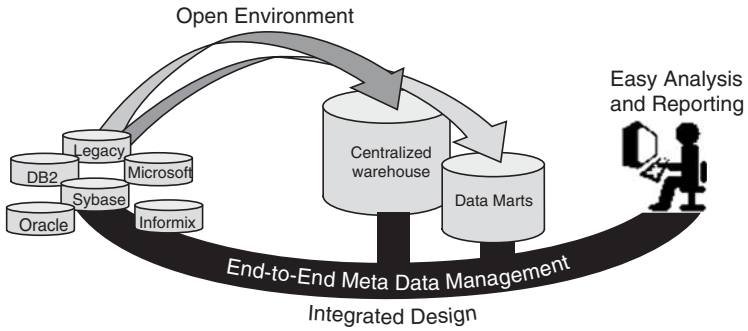

Fig. 3.21. A data warehousing system

age customer relationships or contain costs can mean sifting through massive amounts of data from multiple geographic locations, company departments and providers. The key to unlocking many healthcare issues, such as how to manage provider issues, eliminate fraud, control costs, and maintain efficiency, can be found in the data that already exists within the organization. The trick is being able to extract the relevant information from the data to build conclusions about customers and guide marketing efforts. Although data warehousing has been around for many years, healthcare organizations are just beginning to understand and realize the benefits this technology has to offer the industry. As a point of reference, it is believed that the data warehousing market will grow from a $4 billion industry in 1998 to $6.5 billion in the year 2000. Data warehousing is an effective tool for a competitive industry like healthcare. But why? What can data warehousing do to help a healthcare organization compete and survive?

- Enhance customer relationship management
- Improve provider management
- Prepare for HEDIS reporting
- Perform disease management
- Reduce fraud

3.2.3 Enhance Customer Relationship Management

"Cradle-to-grave" care is today's industry buzzword. Healthcare organizations understand that in order to keep patients for life, they need to satisfy customer groups and the individual patient. Customer relationship management is, therefore, a focus on member retention as well as customer satisfaction. Data warehousing can be used for a multitude of programs to effectively and efficiently enhance customer relationship management. For example, to keep customer groups informed of the high level of service they are receiving,

provider outcomes and cost analysis reports are made easily accessible. On a much more personal level, a data warehouse can trigger a flag so that a birthday card can be sent to individual patients. Much can be learned about a patient population from analyzing basic member information.

Such information as service utilization patterns and demographics of member groups can help organizations make targeted business decisions that help them with more strategic marketing campaigns and better customer understanding. One of the largest healthcare providers in the country is planning to enhance its data warehousing system to enable managed-care providers and members to check on claim status and member eligibility and to receive wellness information via voice response. This superior customer service feature demonstrates the organization's commitment to its members as well as its providers.

3.2.4 Improve Provider Management

The cornerstone of managed care is provider management. It is imperative that an organization accurately gauge how each provider is doing in terms of cost of care, outcomes, as well as their contract and credentialing status. Information that "slips through the cracks" in this area could mean the loss of hundreds of thousands of dollars over time. A data warehouse provides a single source where users can analyze claims-based utilization in detail. Excessive spending can be pinpointed and providers can be alerted to both negative as well as positive results, then an appropriate course of action can be developed. Currently, a national healthcare management company is using data warehousing to support provider network management and referral queries, which assists with NCQA accreditation. Other healthcare organizations have utilized data warehousing to monitor ancillary department activities, such as pharmacy cost trends and mental healthcare analysis. As data warehousing becomes prevalent throughout the enterprise, costs will be better managed, quality of care will increase and greater provider satisfaction will be realized.

3.2.5 Reduce Fraud

Fraudulent claims cost the healthcare industry billions of dollars every year. By using a data warehouse, a healthcare organization can detect potential fraud and take steps to resolve the situation in a timely manner. One healthcare organization has combined artificial intelligence with their data warehouse. Claims are compared to the organization's business rules and discrepancies are reported. The organization can then look at each discrepancy on a case-by case basis. This system has reported a savings of $1 million every month! This technology saves money for the organization and helps to manage resources more effectively.

3.2.6 Prepare for HEDIS Reporting

Not only can organizations use a data warehouse for HEDIS reporting, they can use this information to proactively evaluate how they are doing. For example, an organization can run a report six months prior to needing to file official reports. This can identify areas that need improvement. The organization can then use the six months to implement improvements in order to score even higher on the actual HEDIS report.

3.2.7 Disease Management

A data warehouse can transform a patient's medical record into an analytical tool for optimizing the care of each patient individually. This enables clinicians to continuously monitor and analyze the condition of each patient to facilitate timely and effective clinical decisions. Workflow efficiencies are enhanced when the "big picture" can be seen by integrating clinical, administrative, and financial data. For example, in one healthcare organization's kidney dialysis ward, a provider was having greater success with his patients than the other provider. His care plans and costs were analyzed and it was discovered that he had consistently found an iron deficiency when he cleansed his patient's blood, so he prescribed iron supplements on a regular basis. The care plan in the ward was adjusted and it became a clinical practice that all providers prescribe iron supplements to their patients in this ward. The quality of care throughout the ward has increased, and the costs have gone down.

3.2.8 What to Expect When Beginning a Data Warehouse Implementation

In today's world of distributed, disparate data sources, changing user needs, and demanding turn-around schedules, building and owning a data warehouse is often a complex challenge. The development, implementation, and use of a data warehouse can be eased by considering the following factors:

Integration– Technology should be chosen that works in tandem to provide a single, simple data warehouse solution. Data must be able to be extracted from a variety of sources and transformed into information that makes sense to the business user. A data warehouse project is complex enough without having to fight with proprietary systems that cannot easily retrieve data. Open technology is the key.

Fast Implementation– An integrated data warehousing solution will enable rapid deployment and lower total cost of ownership. Rapid deployment equates to a faster return on investment (ROI). If planned correctly, and implemented appropriately, ROI for a data warehouse should be very quick. Most organizations can expect to realize their return within a couple of months.

Performance– Query speed is important, so look for data stores that are optimized for decision support. Performance should be maintained regardless of what is asked and how many people are asking it. Regardless of format, data should be able to be consolidated into a single repository.

Enterprise-wide metadata management– Look for integrated tools for capturing, synchronizing, managing, and using metadata throughout the data warehouse. Quick access to information across the entire enterprise allows users to make informed business and clinical decisions and do their jobs more efficiently.

Ease of Use– Once the warehouse is populated, an intuitive query interface should open the warehouse to a wide audience. (The warehouse will lose its value if it is not easy to use.) Business users need to be able to enter queries in simple English—not complex coding. Additionally, users need to be able to easily create professional reports and use third-party reporting and analysis tools. This increases productivity because people get the information they need quicker and resources are better managed because the IT department does not have to be called.

Replication– Whatever an organization's replication requirements—efficient regular updates, data that is within seconds of real time or sophisticated change-based warehouse updates—make sure the vendor's solution can access and synchronize diverse data sources through the nonintrusive capture of the operational transactions generated.

Expertise and Experience– Buying the right technology is only the first step toward implementing a data warehousing solution. You need to confirm that the vendor you have selected has the expertise and the experience to successfully implement the data warehousing solution. Make sure the vendor has a master plan for building, using, and managing an enterprise data warehouse. Demand a road map for meeting short-term tactical goals and for extending the data warehouse to meet your organization's long-term strategic objectives.

3.2.9 Definitions

As with any technology revolution, the industry terms and acronyms run rampant. To create a level playing field, here are some definitions:

Data Warehouse– A large-scale, central database with read-only electronic files that is loaded with information from multiple operational databases for the purpose of end-user access and decision support. Data warehouses accommodate random, ad hoc queries and let users drill down to minute levels of detail. Data warehouses differ from an operational system, where data is static and updated in a scheduled manner through massive data loading procedures. Data in the warehouse generally covers a longer time period and integrates information from more sources than do operational systems. A warehouse is designed to integrate data from the operating systems while retaining a basic relational structure. Simply put, a data warehouse puts raw data into context for product and marketing strategies.

Data Mart– Data marts are "segments" of a data warehouse presenting a specific data set to individual business units, such as vertical marketing departments or finance. Data marts are a flexible approach to data warehousing that allow individual business units to have strong control over the type of information they have access to and how they can manipulate it for analysis, while still maintaining a single centralized data warehouse.

Operational Data– Data filtered from operational systems, such as billing, lead management tracking systems and customer support systems, are entered into a data warehouse for analysis.

Data Mining– Querying a data warehouse's operational data to "dig up" information about customer preferences.

MetaData– Information about data, such as a piece of data's origin. For example, did this information come from marketing or finance, or did it come from some external source, such as a purchased mailing list?

3.3 Data Warehousing in the Telecommunications Industry

There's turmoil in the telecommunications industry—worldwide turmoil. In the U.S., the Telecommunications Act of 1996 has ended government rules that maintained barriers between local and long-distance calling, cable TV, broadcasting and wireless services. In Europe, deregulation has opened markets in England, Sweden, and Finland. In 1998, 15 European countries are due to liberalize theirs. Developing nations such as Chile, Malaysia, and Peru have dismantled state-owned phone monopolies. This restructuring will create new market opportunities and will end the exclusive access to customers that has been enjoyed by the incumbent monopolies.

As competition intensifies and the network becomes a commodity, service providers are faced with the need to reduce costs, add value to their services, and ensure differentiation. Once a customer can choose among different carriers, customer care and marketing intelligence become strategic functions in retaining and gaining market share. All this change demands that a carrier's supporting infrastructure be responsive and flexible, in one word—agile. Agile in its approach to the market, the way it responds to its customers, and its delivery of new services. Making decision-support information instantly accessible to the knowledge worker creates the agile enterprise: quick, competitive, and in control of business. When done right, a data warehouse coupled with a focused set of decision-support applications provides for seamless integration of communication services globally, better understanding of profitability, and a focused marketing effort.

The introduction of data warehousing and decision-support technology addresses the following issues in the organization:

Market Expansion. As was seen in the interexchange and cellular markets, competition will result in lower prices, new services, and, ultimately, a larger available market. The introduction of new technologies, as seen in the past with cellular, faxes, paging, and the Internet, has resulted in the growth of total network usage.

Integration of New Lines of Business. The opportunity to market more services and offerings to existing and new customers is moving telecom companies into new businesses such as long-distance service, cable TV, video-on-demand, and Internet access. The delivery of a worldwide support network with the capability to connect everyone in the world for the movement of voice, video, and data on a massive scale will provide immense market opportunity.

Mergers and Acquisitions. As the deregulated markets stabilize, many of the new entrants, particularly the smaller carriers, will be acquired by either larger communications companies or new investors. Some of these new competitors will have very different business models, e.g., resellers of local and/or interexchange services, than the acquiring carrier; these new business models must be quickly and efficiently integrated into the existing structure.

Market Environment. The telecommunications industry is one of the leading users of technology in the world. Historically, this technology has provided service and competitive advantage to early adopters (e.g., British Telecom providing specific discount structures through logging calls over a rolling 15-month period). Utilizing technology can enhance a company's market share and contribute to profitable growth. This is likely to be achieved by:

Focusing on the Customer. Products will be increasingly targeted at the individual in this new customer centric era. The better understanding of customers and marketing based on that knowledge then becomes essential. Communications companies are using tools such as marketing databases, customer information systems, enhanced customer service capabilities, predicative behavior models, and integrated marketing systems to improve and solidify their relationship with the customer. This will lead to mass customization through almost infinite rating structures providing service on a customer-by-customer basis, thereby keeping the customer base while maintaining revenues and margins.

Another aspect of focusing on the customer is the need to simplify things for the consumer. To meet this need, communications companies are moving toward offering bundled services or "one-stop shopping," providing customers with all their telecom needs: local telephone, long distance, wireless, Internet, and cable TV services—all in one bill. MCI is already providing bundled services through its MCI One program.

Reducing Time-to-Market. As the market becomes more competitive, the speed with which a carrier deploys new offerings will become more critical. The time to market of products becomes crucial as any technology leadership

becomes transient. This will require activities such as prelaunch analysis, market trials, trend analysis, service roll-out structures, and lifecycle predictions. Irrespective of market type, additional capabilities at the beginning of the planning cycle will enhance the likelihood of service success, maximize return on investment, and increase the ability to respond quickly to market changes.

Differentiating Business Processes. The business processes and associated IT support systems must add value and support revenue generation while minimizing operational costs. The data stored in the IT systems must be transformed into information to introduce and enhance market opportunities. The challenge is to determine the best means to coordinate, measure, and control the different functions that deliver services to the customer.

Developing an Agile Company. The winners in the telecom industry need a number of basic elements: critical size, adequate capital, a responsive culture, a passion for customer satisfaction, market knowledge, a customer-driven service portfolio, and agility. Strategic planning is required although the company must have the ability to react to market changes or opportune situations. The agile company has the support infrastructure necessary to react quickly and flexibly within the constraints of current business practices.

Agile Decision Making. Decision making based on the analysis of customer data is a requirement across all areas of the organization, from strategic planning to customer service. If the telecom of the future is to succeed, it must use technology to address its business challenges, thus adding value directly to the business. There are a number of key activities an agile company must address:

Customer Care/Customer Retention. Carriers are facing complex challenges. They must offer multiple new services and convergent billing while presenting a single, seamless presence to the customer. Customer service is often the only point of contact the carrier has with the customer and is often the responsible element in both minimizing churn and maximizing revenue through cross-selling. To perform this task successfully, a single view of the customer is needed with rapid access to call history, both billed and nonbilled calls. This makes it possible to turn any customer query or complaint into a sales opportunity.

Churn Analysis. The objective of churn analysis is to reduce customer turnover by using advanced modeling tools that analyze customers' patterns in connects, disconnects, and changes of service. The information in a data warehouse can be used with churn prediction applications to identify high-risk targets and to provide, through outbound marketing, appropriate counter measures to retain the customer. One U.S. wireless carrier reduced churn from 30% to 25% in 6 months by using this application solution. With the advent of local exchange competition and local number portability, churn is about to become a major concern for wireline carriers as well.

Sales and Marketing Analysis (Campaign Management). The objective of sales analysis is to maximize the revenue potential of the network through the optimum mix of products and services, revenue, and profit. Accuracy and speed in tracking services and sales is the key to informed and agile decision making. Sales analysis requires a representative volume of call detail records (CDRs) to allow information based on actual data in the enterprise to be annualized. This allows business analysts to determine the direction of services and their profitability.

Database Marketing. Service providers see the need to provide customized services, targeted at each demographically dispersed customer. Data warehousing is the key component to enable this capability. Database marketing closely tracks trends so that the customer base can be profiled, segmented, and targeted. This type of marketing can maximize the short-term revenue of new and existing products and services by leveraging knowledge of existing customers. Accurate information is also key to rapidly trial new products. For example, customers who match the target profile for a new tariff can be selected. These customers are then monitored and the impact of the new tariff examined on a near real-time basis. In addition, billing can be carried out directly to prevent complex amendments to the billing system until the tariff has proven successful.

Network Capacity Planning and Utilization. Network planners require accurate information about the network utilization and the ability to link network information to the associated revenue and profit contribution of network elements. For example, a number of network elements may require upgrading. It is advantageous to prioritize the upgrades based on their potential impact on profit. The ability to analyze call detail records based on network element criteria makes this possible.

Fixed Asset Management. Many companies are not minimizing their cost base due to poor management systems, making them unable to accurately detail and track their assets. To address this problem, one major U.S. communications company uses an asset data warehouse to effectively manage their asset base. It ties financial and asset systems together via asset serial numbers which streamlines asset tracking. Managing fixed assets has a significant impact on cost structure and returns notable annual savings.

Executive Information Systems. Managers require a consistent, enterprise-wide view of business information. This requires a single data source that ensures a single view of the customer across all service areas. A top-level view is required, although the ability to "drill down" into specific information is also necessary. Trend analysis and investigation of anomalies can show marketing and service performance.

Why are Existing IT Systems Inadequate?. In the past, companies invested in proprietary systems and made significant investments to develop applications to support the business. However, these systems often failed to deliver quality

information to the business because in most cases data warehousing efforts utilized rehashed on-line transaction processing (OLTP) techniques resulting in data that was outdated, of poor quality, and not entirely relevant. With the advent of robust, scalable open systems; new applications and software tools were introduced. Today, advanced transactional analysis, workshop, and dimensional modeling techniques are available. Transaction-intensive systems are designed for the capture and processing of data—not for the provision of information.

Attempts to use them for information provision fail to address the following issues:

Query Performance without Impacting Transaction Processing. Transaction processing systems are optimized for small discrete transactions. DSS queries that are run against an OLTP system involve complex table joins and full table scans. A single query can consume all the available computing resources and prevent the critical transaction processing load from running. Restricting the queries to run at off-peak hours or on a duplicate copy of the transaction processing system, while eliminating the impact on the transaction processing load, is not an acceptable solution as the computing resource required is still too great.

Single View of the Business. Typically companies have multiple transaction processing systems to support the multiple services that they offer. Even if it is technically possible to run a distributed query against the multiple systems it is not practical to link a customer's data from one system with that of another system. A single, integrated database is required.

Data Legibility for Business Users. Transaction processing database schemas are typically very complex and consist of many hundreds of tables. Since simple business entities are stored across multiple tables, to get meaningful information, a detailed understanding of the schema is required. The average business user has neither the time nor inclination to acquire this level of knowledge. Moreover, naive use of such a system frequently leads to incorrect results being generated. It is necessary to provide a data model that reflects the business.

Integration with Appropriate Tools. To enhance competitiveness it is not only necessary to have the right information at the right time, it must also be presented in a useful way. Large paper-based reports are typically ignored with the exception of a small amount of key data that is manually entered into a spreadsheet. Data must be delivered in a format that is appropriate for its use. For managers this would typically be integrated with their standard PC-based spreadsheet. Analysts require advanced modeling utilities, while customer service representatives require integration with existing call handling systems.

Existing System Structure. The diagram shows a view of a typical telecommunications company that has several data repositories utilizing various applications. Often, the applications do not have standard interfaces and have not been integrated. This leads to a number of different views of the customer with no ability to integrate the information. Often the data is not complete and cannot be combined into one data model. From a business standpoint, it is impossible to build an accurate, comprehensive profile of the customer, which often leads to incorrect actions being taken.

The Agile Telecom. To provide the modern communications company with an environment that supports the business, an architecture has been developed to address the technological and business issues it faces. This architecture takes all the disparate systems that have been developed and overlays an infrastructure, which centralizes the market data, customer data, call data, and internal resources data. It integrates and consolidates IT systems and implements focused solutions, which address each of the businesses' issues through system and application deployment—seamlessly. This architecture consolidates and fully integrates the three key functions in the organization:

One Call: Many communications companies today have developed internal systems to support newly introduced services. The existing system structure is shown in Fig. 3.22. However, often these systems are not integrated causing problems for customer service representatives handling a customer's request. Modern applications allow the various systems to be integrated and allow the same representative one view of the customer whether involved with inbound or outbound calls. Another benefit of this system is that each agent can handle more calls, thereby improving productivity and reducing costs. Also important is the reduction in the cost of training new agents. A well-designed system can prioritize a caller. Customers can be segmented into different groups and targeted for various services and level of "VIP" treatment. Very good customers can call special numbers or the system could flag a caller as a very important customer by matching their phone number to a list.

One Bill: Technology has created the opportunity for companies to launch a variety of related communications services. Not only do customers want to obtain all their communications services from one company but they also want to receive a single all-inclusive bill. Customers' lives and businesses have become increasingly complicated, a consolidated bill is likely to ease the problem. Additionally, consolidation of this function provides the communications company with a tactical weapon for cross selling and bundled discounting. The bill in itself becomes a product differentiator.

One View: Few companies today have a single view of the customer. The development of various databases, often conflicting, poses significant problems for sales, marketing, and finance departments. Today, telecommunications business solutions address issues such as churn, network profitability, and least-cost routing. This can be achieved through the consolidation of

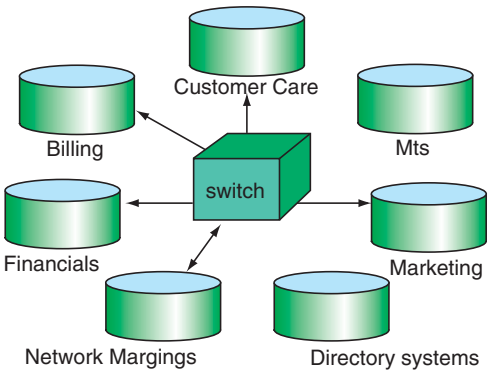

Fig. 3.22. Existing System Structure

the database, installing a data warehouse, and where appropriate, connecting satellite data marts that feed off the central data warehouse. The result of this architecture is the *agile telecom.* This architecture effectively addresses each of the following key business issues:

Minimize Customer Churn/Increase Customer Retention. This is a major issue for telecommunication companies. The agile structure is shown in Fig. 3.23 Providers offer significant incentives to get customers to move to their services, however, it costs many times more to win customers than to keep them. Between 20% and 40% of all communications customers churn annually.

Increase Services Sold To Existing Customers. Communications companies have considerable information about their customers' usage and purchase behaviors that can be leveraged to cross-sell existing services (e.g., call forwarding, single phone number) as well as new services (e.g., video-on-demand).

Gain New Customers. Telecoms have the ability to maintain an almost infinite number of rating structures, thereby allowing them to customize service almost on a customer-by-customer basis. This allows them to customize and target appealing services at new potential customers.

Increase Call Volume. Maximizing call volume allows companies to break even quicker (70% of all initial investment for new telecoms is infrastructure installation) and to utilize the fixed cost of interconnects (excluding call volume). Communications companies have considerable unused network capacity during certain periods of the day, week, and month since they must build enough capacity to handle peak loads. Understanding customer usage behaviors and unused network capacity allows communications companies to target service offerings to absorb unused network capacity.

3.3.1 Implementing One View

A key component of the agile telecom is the data warehouse. Many communications companies have been building data warehouses for years. What they

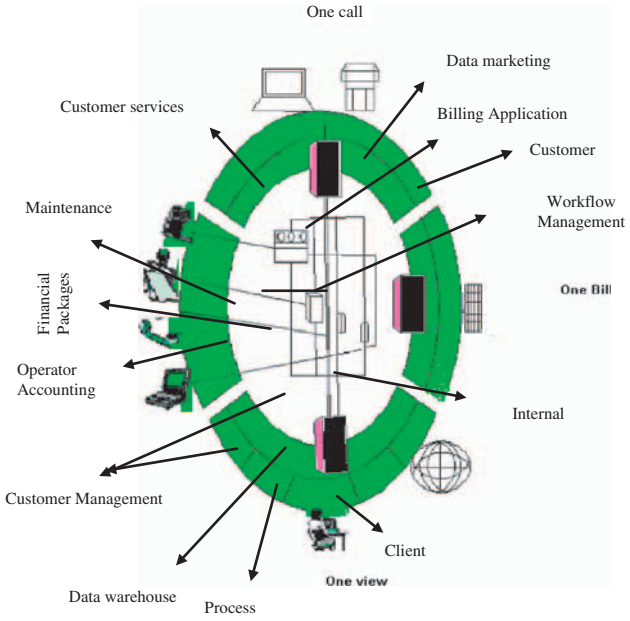

Fig. 3.23. Agile Structure

have not been doing is making them flexible, responsive, business focused—*agile*. To support the agile telecom, its data warehousing architecture must:

- have *flexibility* to cope with storing today's service transactions and the ability to accommodate tomorrow's
- give *one view* of the customer from all parts of the enterprise
- be responsive to today's business needs, not yesterday's
- use the latest tools to add intelligence and a business focus to particular issues for decision making
- be fast to implement for quick payback.

With the correct model, data schema, and integration expertise; implementation of the initial model can be achieved in as little as 3 to 6 months.

What are the components?

A business solution utilizing a data warehouse has two distinct components to it:

Physical Components

- Scalable open system platform
- Raw data
- Data model/schema

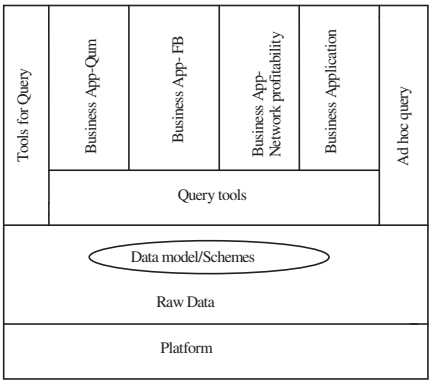

Fig. 3.24. Data warehouse components

- Query tools
- Warehouse applications
- Business applications

Human Component

- Knowledge of the data warehouse environment
- Necessary core skills to build the solution
- Sufficient telecom knowledge to apply the data and schema (providing a fast start to the project and early returns)
- Project management skills with a proven methodology

The development of a data warehouse has been depicted in Fig. 3.24. Experience has shown that a phased approach must be taken to ensure a successful implementation as illustrated in Fig. 3.25. An incremental methodology should be used, each phase addressing an area of business. This ensures that each business application is fully functional before the next phase is undertaken. This approach must also be integrated into an agile architecture, which provides support to each area of the business. Early implementation of the first phases helps generate an early return on investment.

The initial phase should involve a series of workshops or forums where the decision-support requirements of the business are discussed. These forums will help finalize and consolidate views, prioritize directions, and gain consensus on the subject at hand.

3.3.2 Business Benefit

Sequent provides a complete portfolio of Decision Advantage^{TM} Services to assist in the process of exploration, implementation, and deployment of DSS. A successful DSS strategy requires an informed analysis of a firm's current state, a strong business case (including projected financial impacts of better information), and a realistic implementation plan linked to the firm's business

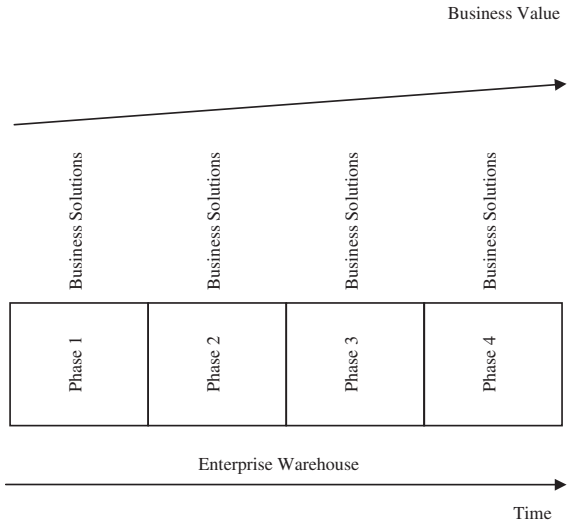

Fig. 3.25. A Phased Implementation

goals. Sequent's decision advantage business benefit analysis is an executive workshop, which meets the above goals and ultimately quantifies the benefits associated with DSS systems. Business benefit analysis combines traditional interviewing techniques and the workshop approach to answer key questions about the major opportunities and ideas, and the potential benefits and information requirements in implementing these. Each idea is plotted onto a matrix that notes technical feasibility and financial impact, suggesting high- and low-priority project phasing. The workshop concludes with a "consensus pledge" from the participants as to next steps.

3.3.3 A Holistic Approach

To reduce the complexity of the decision-support systems and the development of the data warehouse, it is necessary to understand the business issues, technical infrastructure issues, and data issues. An implementation methodology should encompass a phased approach and deliver benefits by:

- defining and *prioritizing* a high value business need that can be quickly solved
- building a cost *effective and timely* decision-support system via a phased approach that incrementally builds an organization's enterprise decision-support system
- *delivering real business value* with each phase
- providing *valuable experience* to the customer in developing decision-support systems
- providing the users with the *right information* for making critical decisions.

3.4 The Telecommunications Lifecycle

Telecommunications costs—including voice, data, and wireless services—are the second largest indirect expense at most large enterprises. The multiple service providers, service offerings, technical alternatives, contract complexities, and billing platforms combine to create massive confusion in the marketplace. This confusion has led to substantial vendor overcharges, rampant service abuse and misuse, and countless hours spent trying to manage it, creating an unnecessary drain on operating income. To make matters worse, your telecom and IT staff members, hired for their technical skills, are burdened with these administrative tasks when they should be focused on ensuring availability and quality of the voice and data network.

Enterprises are now becoming more strategic in their approach to resolving this problem. With tight budgets, small staffs, and constant pressure to cut costs even further, companies realize there is a need to apply a systematic, best practices approach to managing their telecommunications administrative functions. Therefore they are looking to outsource these noncore functions to telecom administrative outsourcing (TAO) services companies that can apply domain experience and intelligent technology while saving the enterprise money and reducing administrative hassles. They are also very carefully selecting their outsourcing partners, to ensure long-term success.

The telecommunications lifecycle is a holistic approach that addresses the problems at the root, long before an invoice appears. The terms of a telecom contract set the stage for errors, which grow exponentially through the order process. And without an up-to-date inventory of services and assets, the errors become increasingly harder to identify and correct. The telecommunications lifecycle approach addresses this by combining contract management, order management, inventory management, bill management, usage management, and information management all in one integrated solution.

It is clear that this problem is not going to go away and in fact will likely grow over the next several years. To ensure maximum success, it is imperative that the profit leaks are captured at their source and real-time tools are available to manage the complete lifecycle of telecommunications services on an ongoing basis. Therefore, enterprises need to implement a holistic, lifecycle solution to manage telecommunications administrative functions that will reduce hassles and improve the bottom line. It is imperative that the profit leaks are captured at their source and real-time tools are available to manage the complete lifecycle of telecommunications services on an ongoing basis.

3.4.1 Current Enterprise Environment

The telecommunications industry has seen dramatic change in the last ten years. The result has been a proliferation of service providers, new communi-

cations services, new communications devices, and more recently a decimation of customer service, as margins in the industry have vanished. During the same time period, telecom costs to the enterprise have been on the rise. According to Gartner Research, U.S. businesses will spend nearly $350 billion in 2003. The Aberdeen Group estimates the total cost per employee to be between $3,000 and $5,000 per year, which is double the amount it was just five years ago.

Yet most companies still have no systems or processes in place to tightly manage these costs. They still consider these indirect costs nonstrategic, therefore the purchasing and ongoing management of telecommunications services are a widely distributed (typically to someone on the IT staff who would much rather be working on more mission critical, technical tasks), poorly controlled, and heavily paper-based activity. As a result, this unwanted administrative burden does not get the level of attention commensurate with its importance to the company's bottom line. Traditionally, companies have attempted to attack the problem by hiring consultants who would identify savings opportunities and develop optimization strategies. More often than not, these consultants were paid a share of the savings achieved, so there was no motivation to fix the root of the problem, just to find immediate, one-time savings.

In more recent years, technologies and outsourcing companies have emerged to help enterprises automate and streamline the procurement process and the invoice management process. Though better than doing nothing, these initiatives have typically been shortsighted and fragmented, focusing on a narrow aspect of telecommunications costs. They have also focused more on the symptoms of the problem than the problem itself, resulting in one-time savings that are difficult to sustain over time. But the rapid decline of the global economy in 2001 was a painful reminder to the enterprise that cost control drives profitability in difficult times and supports growth in the good times. As companies have scrutinized their costs during this recession, it is becoming more and more apparent that the costs for voice, data, and wireless communications services offer one of the greatest opportunities for savings and process streamlining.

3.4.2 Getting to the Root of the Problem

No best practices approach to controlling voice, data, and wireless communications costs ever started with an invoice, yet that is what most companies focus on when they need a quick budget fix. They look to overcharges on their invoices that will likely give them a quick, one-time budget boost, but will not address the root of the problem since it begins the moment a vendor contract is signed.

Telecommunications contracts are extremely complex and heavily weighted in favor of the carriers that offer them. The sales pitch is that a custom contract means heavily negotiated, custom rates. But the reality is that buying

telecom services is like buying a car, so unless you know exactly what price to ask for and know exactly how to structure your telecom agreement, you will lose at the negotiating table and you will pay more than necessary. After the contract is signed, the unnecessary costs go up exponentially when services begin to get ordered. Since the procurement process is typically very decentralized, and handled by people who understand a T-1 in detail but do not understand or care about what a purchase order is, order compliance is virtually nonexistent and the cost to complete an order through manual processes is enormous.

This results in services being ordered from multiple vendors, services being ordered that are not in conformance with company policy, services being ordered and not properly configured so they bill inaccurately from the first invoice, and wasted time as a large number of staff members are engaged in these burdensome administrative tasks. Once services are ordered, they need to be tracked through an inventory or asset management system, or costs will continue to escalate. The ever-changing enterprise, whether that change is growth or contraction, is constantly adding and deleting services to fit the current business need. Without knowing what you have and where it resides, it is impossible to keep it optimized or even terminate it when you no longer need it. Yet the typical enterprise has no inventory whatsoever of their telecommunications services. These core problems, along with the archaic vendor billing systems that are designed to bill you at the highest possible rate, cause billing errors that are virtually always overcharges. In fact, Gartner claims these billing errors an amount to an excess of 10% if there is no ongoing cost management process in place.

Therefore, there needs to be a process in place for validating billing as it is received to ensure it is your bill and it is coming from your chosen vendor. The system must also audit billing in real time to ensure the rates being charged are in line with the contracts negotiated. It must also handle charge backs to appropriate cost centers to ensure maximum visibility of the costs by the people who generate them. The old adage, "out of sight, out of mind," plays heavily here. If they do not get a bill, they do not know the costs and cannot do anything to help control them. But even this is not enough to address the complete problem, because it is also critical to enforce proper service utilization by end users. This requires that you have monitoring systems in place that capture the invoice usage detail and analyzes it for optimization opportunities that can be reported back to the abuser or misusers. Finally, you should be using all of the information generated from the contracts, orders, invoices and usage to better manage the business. The analytics available from these processes is invaluable if you are capturing it real time, at the point of entry.

3.4.3 The Telecommunications Lifecycle

To get complete control over voice, data, and wireless costs in the enterprise, you need to take a holistic approach that addresses the problem at its root, before it ever becomes a problem. You need a lifecycle approach that manages every aspect of each of the administrative functions in the telecom department so you can focus on keeping your voice and data networks up and running. An effective telecommunications lifecycle management system should include the following six elements: contract management, order management, inventory management, invoice management, usage management, and information management.

Contract Management – You need to negotiate world-class vendor agreements through an RFP process, renegotiate existing contracts on a regular basis, and ensure vendors comply with the negotiated contract terms on an ongoing basis. To accomplish this you need:

- Vendor Benchmarking – A knowledge base of supplier performance and compliance to benchmark vendors against one another for rates, contract terms, SLAs, technical suitability, financial stability, etc. thereby insuring a more fact-based selection process.
- RFP Management – Tightly managed RFPs, using a proven methodology, for voice, data, local, long distance, and wireless services to ensure best in class contract rates, terms, and conditions.
- Rate Reviews – Tightly managed contract renegotiations, using a proven methodology, to improve contract rates and terms at every opportunity throughout the contract term.
- Contract Compliance – To track existing contracts to ensure compliance with rates, terms, and commitment levels that helps identify overcharges and optimization opportunities.
- SLA Management – The ability to track and manage SLAs on a monthly basis and file claims for recoveries when they are not met.

Order Management – You need to manage the telecommunications services procurement process to enforce compliance and provide for fully automated approvals. You need to be able to track order status on a real-time basis. To accomplish this you need:

- Order Processing System – An automated interface for placing, tracking, and managing new orders, changes, and disconnects, eliminating expensive paper-driven processes and increasing enterprise-wide compliance with selected vendors. This should include workflow and approvals to expedite all orders.
- Order Integration with Vendor – Automated bonding with your vendors, including a catalog of services or devices (cell phones, pagers), where an

order can be placed, approved, then routed automatically to the appropriate vendor.

Inventory Management – You need an up-to-date inventory of your lines and device assets. To accomplish this you need:

- Line Inventory – An inventory of all lines, maintained current on a daily basis, to which you can apply specific charge back and other coding and perform ongoing maintenance. This enables accurate charge backs and greater accountability.
- Telecom Asset Management – An inventory of all cell phones, pagers, handhelds, desktop phones, etc. maintained on a daily basis, to allow visibility, accountability, and chargeback.

Invoice Management – A system is required to automatically process and audit your invoices on a real-time basis. To accomplish this you need: - Invoice Processing – Automated systems to enter, code, validate, charge back, pay, and update your general ledger system. - Invoice Auditing – Automated systems that audit your invoices real time, based on contract terms. The system should contain intelligence on vendor and service idiosyncrasies that routinely lead to overcharges. The components of the invoice management system are shown in Fig. 3.26.

Usage Management – You need visibility into usage detail. To accomplish this you need:

- Invoice Usage Visibility and Access – To have the ability to review usage from both your paper and electronic invoices on a monthly basis, which automatically reveals fraud, abuse, misuse, and optimization opportunities.
- Call Accounting – Real-time usage tracking to the desktop phone, which also automatically reveals fraud, abuse, misuse, and optimization opportunities.

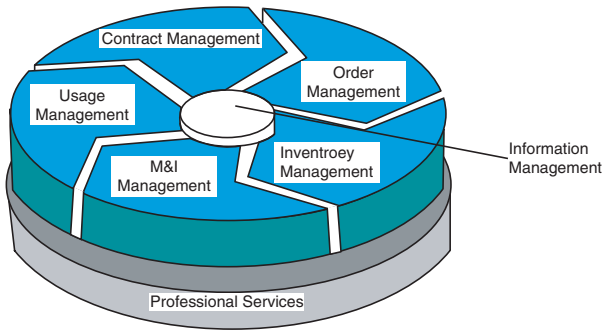

Fig. 3.26. Invoice management system

Information Management – You need to manage the extensive amount of information that comes from a telecommunications lifecycle system. To accomplish this you need:

- Reporting Functionality – A reporting tool that will create logical, easy to understand reports on each of the processes performed in the telecommunications lifecycle.
- Benchmarking – The ability to benchmark your total spend, spend by employee, rates, etc. against other firms in your industry.
- Analytics – The ability to measure this data against other business information to better manage the business.

3.4.4 Telecom Administrative Outsourcing

META Group claims that 70% of IT organizations currently outsource 10% to 50% of their IT operations and they predict the IT outsourcing market to grow by 20% in 2003. But the current trend in industry is to outsource noncore, administrative business processes, thus enabling the enterprise to focus on more strategic, mission critical tasks. But business process outsourcing ("BPO") is a long-term proposition and requires the forming of a partnership between the enterprise and the BPO provider. This warrants more detailed scrutiny of the BPO provider and an alignment of interests that should be enforced through contractual commitments. Telecommunications administrative functions are an ideal business process to consider for outsourcing. While the voice and data network are core to most enterprises, the administrative tasks required to support the networks are not. And the IT staff does not cover the tasks associated with this business process. In fact, the low man on the totem pole typically ends up with the responsibility, with hopes that they will quickly progress to a more challenging and exciting role within the department. Enterprises have attempted to develop internal systems and departments to gain control of telecommunications costs for decades, only to find that it is virtually impossible to keep pace with the rate of change. In the rare cases where companies have achieved a measure of success in this area, it has typically been as the result of one standout individual within the IT department with no great processes or systems, but a great eye for detail and a willingness to do the dirty work. However the success typically comes to a quick end when that all-star individual moves on, forcing the company to start from scratch.

3.4.5 Choose your Outsourcing Partner Wisely

The ideal partners for this business process are telecommunications administrative outsourcing companies. But with all the new entrants in this sector, you need to choose wisely. You should look for a firm with several years of experience in managing the telecommunications lifecycle. If you have been trying to solve this problem for decades, how is a company with one or two

years' of experience in the industry going to do a better job? You should look for a firm that uses leading-edge technology that addresses the complete telecommunications lifecycle. Limited solutions lead to short-term, limited results.

Additionally, you should look for a firm that shares your company's values. Since outsourcing is a long-term commitment that is expensive to change on a regular basis, it is best if you work with a firm with people you like. You should ensure you are retaining a firm that is financially secure. A solid balance sheet that shows the ability to be there for the long haul is critical in this difficult economy. You should also look for a firm that charges you an easy-to-understand, fixed fee and allows you to keep 100% of your savings. Some pricing models are unpredictable, which creates risk and could lead to unaligned motivations. But most important of all, you should expect these companies to provide a contractual commitment to reduce telecommunications costs. Eliminating administrative hassles is a clear benefit of outsourcing this business process, but in today's difficult economic climate, it must come with a guaranteed ROI.

The current problem is clear. There is enormous confusion brought on by dramatic changes in the telecommunications industry. The outlook for improvement is bleak. With telecommunications companies still struggling, margins at all-time low levels, and vendor priorities still on selling, there is no expectation that this problem will be resolved on its own or with the help of the telecom providers. A solution needs to be found. Enterprise companies that have not addressed the problem are wasting hundreds of thousands, if not millions, of dollars a year. The solution must be a holistic approach that addresses the problem at the root, not the symptoms after it is too late. The telecommunications lifecycle approach is the best practices way to address the problems before they occur and avoid overpaying. Telecommunications administrative functions are not core to the enterprise's business. Therefore, outsourcing to a firm with proven expertise will offload a burdensome job and maximize long-term savings. Your finance department figured out about twenty years ago that there are certain noncore, highly administrative jobs that just do not make sense to perform internally. So they started outsourcing the biggest source of pain—payroll—and gave birth to and industry now led by ADP. Telecommunications administrative outsourcing is the next big opportunity in this area for the enterprise. Therefore, there is no reason not to begin increasing savings and reducing administrative hassles now.

3.4.6 Security in Web-Enabled Data Warehouse

Data warehouses represent implementation to simplify the storage of data of an entire enterprise. It will be not only easy to manage the stored data but also for the users and data manipulators to use this data more effectively and without having to tackle the underlying complexity of data storage format and the different extraction procedures specialized to handle only one type of data storage and management system. Unfortunately this convenience in data

manipulation does not come without a price. The use of data after building a successful data warehouse may be much more easy and efficient, but to achieve the objectives laid down in the scope of the data warehouse, a proper design and infrastructure have to built and implemented. In the design and implementation strategy the aspect be, which is not given its due importance, is the security.

Security not only of the stored data but also of flowing data and the data transferred to the users is very important. With the introduction of Internet in accessing the data warehouse, the security becomes even more critical issue. Security roughly covers two aspects: firstly the data should be safe from the malicious party. Malicious party does not only mean the people looking to break into the data warehouse but also points to the clients or parties either accidentally or willfully get access to data that they are not meant to access. The second aspect of security deals with the stoppage of accidental spillage of data to ensure consistency and quality in data and safe conveyance of data to rightful party. This section gives a brief account of strengths and weaknesses of different techniques and practices being used and proposes a security model using XML and its related technologies to ensure a better and secure data warehousing experience.

3.5 Security Issues in Data Warehouse

In any cases of data warehouse implementation, the problem of security remains a critical issue. Without a proper infrastructure for secure distribution of information, it is like presenting all the important information required by a competitor or exploiter on a plate. It would be so easy to just put a lock on the entire data warehouse and sleep easy at night. Unfortunately the real world is much more complex then that and this complexity means inherited concerns about the security. The most common security concerns in data warehouse are the following:

Security of Static Data. The most obvious security concern lies at the core storage of data. The data should be safe from unauthorized users and access should be restricted to the users or group of users, which are "supposed" to get access to the data.

Security of Network. Another security weak point arises when the data is being transmitted on a request from the user. The static data is much more easier to secure but once the data leave the safe storage facilities its open for attack from sniffers, etc. and hence require much greater effort on the part of the data warehouse security implementer.

User Classification and Privileges. The user classification and assigning the respective privileges also represent a marked issue in data warehouse functionality and has a direct bearing on the security of the entire data warehouse. Users are arranged in groups and assigned privileges accordingly. Although it is much easier to assign a single privilege level to each user, but that is not

really a very good idea considering the security of the entire data is at stack and not every user should be allowed to access the entire data warehouse just because he wants to use say only 1% of the data stored.

Human Factor. Perhaps the most overlooked security threat is that of the human factor. If an "insider" or employee leaves the team to join some other rival then how to stop him from exploiting the security infrastructure that perhaps he himself helped erecting in the first place.

Throwing in a New Problem: Web-Enabled Data Warehouse. Things get more complicated with the introduction of the Internet. With the increasing use of Internet in every aspect of life, the task of ensuring security for the Web-enabled data warehouse does not at all get easier. Although the Internet represents a very good ready made medium to access the enterprise data from anyway in the world but like with any other easy way out, this also presents the biggest threat to the security. Implementing security in Intranet is not as complicated an exercise than tackling Internet security threat where every kind of malicious people are looking for a way to hack into the data warehouse. With the advancements in hacking techniques and ever-increasing number of industrial spies, the Internet security threat is by no means to be taken lightly.

Metadata Security Threat. A big problem with a Web-enabled data warehouse is the proper management of metadata. Metadata is, in simple words, data about data. For a Web-enabled data warehouse, it is to a great advantage to arrange metadata in a way to facilitate the searching of the required data efficiently and unfortunately this also is the biggest drawback as far as security is concerned. Metadata can be used as a marker to locate the proper information required and this can also play into the hands of a person trying to hack into the data warehouse.

3.5.1 Performance vs Security

In any data warehouse the main emphasis is laid on the two most important aspects *Security* and *Performance.* Unfortunately these two mostly are on the opposite ends of the stick. While implementing a data warehouse, security is very important and should be given due consideration in the infrastructure, but the fact remains that the main concept of building a data warehouse is to provide easy and efficient access to its user; hence its not advisable to provide security at a level that compromises the performance too much. On the other hand, there is no point in providing greater performance at the cost of exposing the data warehouse to unnecessary security risks. So in the implementation of a data warehouse a delicate balance has to be found between the performance required and the security necessary. This balance is not so easy to achieve, performance will be compromised with increasing security but depending upon the requirements the maximum possible security should be in place.

3.5.2 An Ideal Security Model

Theoretically speaking, a perfect security model is a very simple one. It should not allow any unauthorized user to access data; data should be secure and any attempts to breach the security of the data warehouse should fail without exception. Authorized users should interact with the data with the confidence that there is no way anyone can break into the warehouse and that the data they are receiving is free of errors and by no means inaccurate. Unfortunately this is only possible in a perfect world and we mortals do not live in one. Security threats will always be there. Whenever an important data about a rival company is within grasp, people will not hesitate to get their hands dirty and as long as there are such threats, malicious users will always be looking for a way to break into the data warehouse and in a way help in exposing security vulnerabilities in the so called fool proof security.

3.5.3 Real-World Implementation

In the real world although it is not possible to achieve total perfection, that never held anyone back before. There are different techniques available to come close, if not achieve, the dream of a perfectly secure data warehouse. Depending upon the security issues stated above, different techniques and technologies are used to tackle the respective problems:

Static Data Security

As pointed out the data storage should be secure enough to stop any unauthorized intrusion exploiting the data. There are different techniques to achieve that:

Traditional DBMS security

A tempting thought, it will be much easier to manage security if only the database management system security mechanism can also be used to provide security at data warehouse level and DBMS do offer some form of data security. Unfortunately the data warehouse environment is rather different from the traditional database system. Also the fact that in almost every case DBMS security can be entirely bypassed by accessing the data managed under the DBMS in a lower level native access mode. In other words, a simple disk dump that accesses and off loads data at the physical block level suffices to bypass the strongest of DBMS security. Even though the data that has been dumped is in a "raw" state, it nevertheless is easily and totally available, making it very insecure.

Application-Based Security

Another method to enforce security is to use application-based security systems. The data is stored in a normal way without much emphasis on security but the accessing mechanism for the data, e.g., UI (user interface), provides the required security. The application acts as a filter to allow only the authorized requests to be processed and carried out. Data can enter and leave the data warehouse only through the application enforcing security. But this approach limits the flexibility necessary in data warehouse design and imposes restrictions that negatively impact the overall performance and the extra overhead of developing and implementing such an application to tailor it to the requirements of a particular data warehouse is also not desired.

Encryption of Data

Keeping the inside-out paradigm in mind, a reasonable approach is to provide the security from within the data warehouse storage by encrypting the data. This involves encryption of data before it is stored in the scrambled form into the data warehouse and when user requests data, it is again decrypted first and then is delivered to the required destination. Although this is a simple and yet a very powerful approach it does involve performance compromise.

Encryption of data increases the size of the data warehouse; also encryption requires machine cycles and hence can have a negative bearing on the performance. But in a data warehouse the main performance matrix is the I/O cycles and encryption of data does not require excessive I/O cycles; and hence encryption represents a logical approach to data warehouse security.

Selective Encryption

A more reasonable approach to security is to produce selective encryption. A data warehouse features a collection of data that can be in terabytes, and not all the users are supposed to have access to the entire data. To facilitate this, the users are divided into different groups and given privileges and permissions depending upon their requirements. But to increase the security further, the data should be fragmented and each slice of data should be encrypted using a separate encryption key. Then these encryption keys are distributed to the different users depending upon their privilege level. Symmetric cryptography is an obvious candidate for this type of system to work. A single secret key can do the trick, but it is more reasonable to use asymmetric keys with private keys issued to the user used for decryption of data, while public keys used to encrypt the data. This way if a malicious user does get his hands on the data she will require a private key to decrypt the scrambled data. In either of the two cases of cryptographical techniques used, it is understood that authentication of users will be carried out first before any of the processes are allowed.

Network Security: Intranet and Internet Security

The importance of securing the static data in the data warehouse cannot be overlooked but the biggest security threat remains with the inevitable introduction of networks and their inherited venerability. The famous saying *"never trust your network"* is still as true as ever and in fact due to the advancements in the hacking field this spells the biggest nightmare of every organization working to secure their networks from intruders. Different techniques are used in making the network highways safe.

Secure Socket Layer

SSL (Secure Socket Layer) security represents the security implementation at the presentation layer. SSL is a protocol developed by Netscape for transmitting private documents via the Internet. SSL works by using a public key to encrypt data that's transferred over the SSL connection. SSL works by using authentication of the user from a trusted third party. The SSL provided by the Internet Engineering Task Force (IETF), is the industry standard, and most commonly available security mechanism on the Internet. The IETF has renamed this mechanism to be "Transport Layer Security," and it is now usually referred to as SSL/TLS. SSL/TLS also provides data security for applications that communicate across networks. This protocol sits between various application protocols and TCP/IP. The SSL/TLS protocol provides connection security that has three basic properties:

Authentication. The connection can be authenticated using asymmetric, or public key, cryptography.

Privacy. The connection is private. During the initial handshake, public key encryption is used to define a symmetric secret session key. Symmetric cryptography is used for data encryption (for example, DES, Triple DES, RC4, and so on). Different encryption strengths, including 40-bit, 56-bit, 128-bit, and 112-bit encryption, are all supported (triple DES key efficiency is 112 bits rather than 168 bits in order to protect against a theoretical "man-in-the-middle" attack).

Integrity. The connection is reliable. Message transport includes a message integrity check, using a keyed Hashed Message Authentication Code (HMAC). Secure hash functions (for example, SHA-1 and MD5) are used for HMAC computations. The SSL protocol data encryption strength depends on the length of the symmetric cryptographic algorithm keys, which range from 40 to 128 bits.

S-HTTP

Another protocol for transmitting data securely over the World Wide Web is Secure HTTP (S-HTTP). Whereas SSL creates a secure connection between

a client and a server, over which any amount of data can be sent securely, S-HTTP is designed to transmit individual messages securely. SSL and S-HTTP, therefore, can be seen as complementary rather than competing technologies.

Firewalls

A firewall is a device used to implement a security policy between networks. A firewall can have multiple network interfaces and is typically used to create a secure boundary between untrusted external networks and trusted internal networks. The security policy defines what type of access is allowed between the connected networks. There are three methods commonly used in firewalls to enforce the local security policy. The three methods are commonly used together, each complementing the others' weaknesses.

Packet Filters. Packet filters operate on the IP level, scanning the headers of each IP packet crossing the firewall and comparing its characteristics to a fixed set of rules. These rules determine whether the packet is allowed to pass unhindered or not. Characteristics recognized by packet filters are the source and destination IP addresses, the source and destination port numbers, various status bits in the header, and the direction the packet is traveling across the firewall. Packet filters do not know anything about protocols above the TCP/IP layer – they are fast and simple, but not very flexible.

Application Layer Proxies. Application layer proxies are applications running on the firewall, which users on one or both sides of the firewall can communicate with. The proxies forward the users' requests to the actual servers that can give a response, possibly imposing rules on what sort of traffic is acceptable. From the viewpoint of the "actual servers," it appears as if the firewall is making the requests – not the client. Application layer proxies are in general the most flexible type of fire walling software, but they frequently require added configuration or skills from the user. A common example of this is an HTTP proxy, which allows users to request Web pages from anywhere but may refuse some requests or rewrite pages based on rules defined by the administrator.

Network layer proxies. Network layer proxies are a cross between application layer proxies and packet filters. Like packet filters, they scan the headers of the IP packets crossing the firewall – but are able to respond in more ways than packet filters, which generally only "accept" or "reject." To make these "dynamic tunnels" as narrow as possible, a list of active communication channels (TCP, UDP, etc.) is maintained, and only packets (from the "outside"), which exactly match an active connection, are "proxied" back into the hidden network. Only internal, trusted machines can open such a tunnel through the firewall. In addition to this basic "translation" of network addresses and ports, some implementations contain built-in support for common protocols (such as File Transfer Protocol, FTP), which are dependent on external, untrusted

hosts being able to initiate connections to the client. In general though, such protocols will not work through most firewalls based on these techniques.

Another type of network layer proxy is based on cooperation between specialized application proxies and the operating system's network code. In addition to rewriting the IP addresses on the network packets, some packets may be passed for further processing to an application proxy on the firewall machine. This proxy can monitor and filter the flow of data between the client and the external server in a much more complicated (and therefore error-prone) manner than is usually allowed within an operating system's low-level network code. These proxies are generally called *transparent proxies*, because they operate in a manner completely invisible to the user.

XML Firewalls. XML firewalls are designed to protect enterprises against the unique dangers posed by Web services. These firewalls examine SOAP (Simple Object Access Protocol) headers and XML tags, and based on what they find, block any dangerous or unauthorized content or services from getting inside a corporation. Traditional firewalls cannot do this, because they can only filter on the packet level, not on the content level. XML firewalls, on the other hand, examine the XML content of the incoming traffic, understand the content, and based on that understanding, take the required action.

XML Security

The increasing use of XML has prompted the developer to look into the possibilities of using XML in order to increase security. The following XML security specifications are being used to implement security of data:

XML Signature. Defines an XML schema for cryptographically authenticating data. The authenticated data may consist of a complete XML document, individual elements in an XML document, or an external data object referenced by an XML document.

XML Encryption. Defines an XML schema for encrypting data. The encrypted data may consist of a complete XML document, individual elements in an XML document, or an external data object referenced by an XML document.

XML Key Management Specification. The XMK Key Management Specification (XKMS) specifies the protocols for distributing and registering public keys:

- XML Key Information Service Specification (X-KISS): the protocol by which an application delegates, to a service, the processing of Key Information associated with an XML signature, XML encryption, or some other public key.
- XML Key Registration Service Specification (X-KRSS) The protocol for registration of a key pair by a key pair holder, with the intent that the key pair subsequently be usable in conjunction with the XML Key Information Service Specification.

3.5.4 Proposed Security Model

It is evident that none of the above options alone is capable of providing a solution to all the security problems of a data warehouse. So to build a secure data warehouse these techniques have to be used in combination. Figure 3.27 shows a proposed model to implement security in data warehouse. XML features prominently in the proposed model.

Implementing Security

The different steps of implementing security are given below:

i) The security is implemented right from base of the data warehouse, i.e., by using encryption to scramble the data. Using asymmetric keys for encryption and decryption. This means that the size of data warehouse will increase and also an increase of data warehouse costs to cater for the increased number of hardware cycles required, but nonetheless there will not be a big depreciation in the performance. The data is stored in scrambled format using triple DES and generating the encryption – decryption keys accordingly.

ii) Firewall is put around the data warehouse to provide a further safety barrier to the secured area from the intruders. The traditional firewalls are not good enough for that matter because they can only filter at the packet level, and cannot examine the contents of messages. Traditional firewalls protect a network's perimeter by blocking incoming Internet traffic using several different means. Some block all TCP ports except for port 80 (HTTP traffic), port 443 (HTTPS traffic), and port 25 (email traffic). Some ban traffic from specific IP addresses, or ban traffic based on the traffic's usage characteristics. A better option is to use XML services in the network and allow usage of XML firewalls. XML firewalls typically work by examining SOAP message headers. The header may have detailed information put there specifically for the firewall to examine, and if so, the firewall can take an action based on that information.

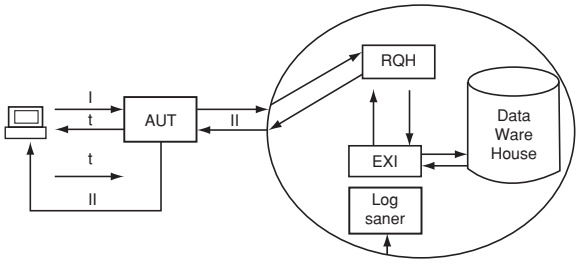

Fig. 3.27. Working steps of the Proposed Security Model

Even if the header does not have this information, however, XML firewalls can still take actions based on what is in the header. The header, for example, might have information about the recipients of the message, security of the overall message, or the intermediaries through which the message has passed. In addition, XML firewalls can look into the body of the message itself and examine it down to the tag level. It can tell if a message is an authorized one, or coming from an authorized recipient and then take action based on that – for example, blocking traffic, sending it to a secure environment where it can be further examined, or allowing it to pass through. XML firewalls have other methods of protection as well. They can understand metadata about the Web service's service requestor as well as metadata about the operation itself. They can gather information about the requesting user; such as understanding what role the user plays in the current request, for example. XML firewalls can also provide authentication, decryption, and real-time monitoring and reporting. So when a request comes from a user the XML firewall can look into the contents as well.

iii) The messages are passed using SOAP protocol over HTTP in communication between users and the data warehouse. SOAP messages are fundamentally one-way transmissions from a sender to a receiver, but SOAP messages are often combined to implement patterns such as request/response. Also the SOAP header is used to communicate information regarding the services requested, so that these headers can also be checked by the implemented XML firewall as a further security precaution.

iv) To secure the network, XML signatures are also used. Each user is identified by its signatures. SOAP Security Extensions proposes a standard way to use XML-Signature to sign SOAP 1.1 messages by defining the <SOAP-SEC:Signature> SOAP header entry. In addition, SOAP Security Extensions reuses two existing SOAP header items. These are the "actor" header item, which indicates the recipient of a header element, and the "mustUnderstand" header item, which indicates whether an application must attempt the validation of the enclosed XML Digital Signature.

Components of Security Model

The components in the security model are as follows:

Authentication Server

The authentication server is used to authenticate the users. This is the first part of security model with which the user first comes in contact with the request for manipulation of data.

XML Firewall

The firewall represents the protected boundary around the data warehouse and its core components. Authentication server is not placed inside the firewall, because a malicious user, if allowed to interact with the authentication server inside the firewall, will compromise the integrity of the firewall itself. The XML firewall works by looking into the payload of the SOAP packet, hence allow another way to stop malicious user.

Request Handler

This is a vital part of the data warehouse security model. When the authentication server authenticates the user, the authentication server hands over the request to the request handler. The request handler checks to see if the user has the privileges to access the data it is requesting. If so, it sends the request to the extractor otherwise sends back the request to the authentication server with the error message defining the user's privileges.

Extractor

Extractor represents the mechanism to extract the data depending upon the valid request handed to it by the request handler. The data extracted is no doubt in encrypted form and handed over to request handler.

Data Warehouse

This represents the collection of data for all types of users. The data is divided into different partitions and encrypted using different keys for each partition. The data in encrypted form is passed to the user and depending upon its privilege level the users are assigned private keys for the partitions for which it has access. This encrypted data is placed under a tag in XML format to be sent to the user.

User

The users are the customers, partners, clients using the data warehouse. Each user is issued the following:

- Username and password: to identify itself to authentication server in initial interaction.
- XML Signatures: to authenticate it to the authentication server.
- Private keys for each partition it has access: these private keys are used to encrypt the data when it wants to store the data back into the data warehouse (depending upon whether the user has this privilege or not) and also for decrypting the data extracted from the data warehouse.

Log Server

The log server stores all the information about the processes taking place.

Working Steps of the Security Model

The working steps of the proposed security model are given below:

1. The user interacts with the authentication server by presenting its issued user name and password as an initial step.
2. If the password is correct, as a second authentication method, the authentication server sends a document to be digitally signed by the user with its XML signatures. A part of the document is digitally signed by the authentication server itself to confirm its own identity and at the same time writes a log file with the date and time of interaction along with the user name. Otherwise the server generates an error message and logs that into the log file. The document to be sent to the user is also generated randomly every time any user interacts.
3. The user signs the document after verifying the digital XML-Signature of the Authentication Server itself. This verification of both Authentication Server and user takes place using XKMS and its two protocols X-KISS and XKRSS. After verifications the user sends the document back to the authentication server along with the request.
4. Authentication server verifies the XML signatures of the user and checks the time stamp, and if found valid, sends the request to the request handler inside the firewall along with the log information placed in the SOAP message.
5. The firewall checks the information in the header to verify that the message is from authentication server itself and allows only the request part to be transmitted to the request handler.
6. The request handler checks whether the request is within the privilege level of the user, if so it sends the request to the extractor; otherwise it generates an error and returns the request to the firewall, which in turn attaches the error information and sends it back to the authentication server.
7. The extractor extracts the required data from the data warehouse, which is in encrypted form
8. The extractor sends the data back to the request handler.
9. The request handler then creates a SOAP message placing the required data in scrambled form in it. And sends it to the firewall.
10. The firewall sends the data back to the authentication server along with a log of the requested information and the log from authentication server being sent to the log server. To achieve this the firewall must read the messages and be allowed to extract the information required by it.
11. The authentication server sends the required information to the user.
12. The user then extracts the required data, no doubt in scrambled form, and then uses his private keys to decrypt the data.

The security of a data warehouse is the most critical issue. The extra added security may increase the costs and decrease the performance by a fraction, but the fact remains that it is better to be safe than sorry. The data of the entire enterprise is at stack in a data warehouse and if a security breach does take place the effects can be devastating.

3.6 Data Warehousing: To Buy or To Build a Fundamental Choice for Insurers

3.6.1 Executive Overview

It is now well known that data warehousing offers an approach that solves the information access challenge faced both by users and information systems management. It can greatly increase the level of productivity and improve or hasten the decision-making process of knowledge workers within an organization. Business intelligence is quickly becoming a standard business practice, and the insurance industry is no exception. Given that data warehousing is "the price to play" in insurance, the focus then shifts from determining necessity to determining the best implementation method. There are two main methods in acquiring business intelligence in any firm: building a solution or buying one. Our position is that building a data warehouse from the ground up is often a risky, expensive, and resource-consuming effort. Whenever possible, it is better to buy a solution from an expert and customize it to a particular environment.

SQLiaison Inc. has been in the business of building decision-support systems for the insurance community for many years. This level of expertise is reflected in FellowDSS^{TM}, a decision-support environment for the insurance industry based on data warehousing technology. FellowDSS^{TM} will dramatically reduce the time, cost, resource requirements, and risk associated with implementing and deploying a data warehouse for different types of insurance companies. This section explores the business and technology issues related to both buying and building data warehousing solutions, discusses requirements for effective insurance decision-support systems, and illustrates why we feel that buying a solution is ultimately more cost-effective for insurance companies.

3.6.2 The Fundamental Choice

Assuming that an insurance company wants to go ahead with data warehousing, they are now faced with one fundamental choice in methodology. A company can either build their data warehouse from scratch, or they can buy a packaged solution from a vendor. Any enterprise-wide (or localized) decision-support solution should have certain requirements and properties. It must:

- Have a facilitated deployment system. This allows for a fast and simple implementation that avoids problems;
- Produce a high return on investment with a low total cost of ownership;
- Support extensibility in the architecture that allows it to grow with the business. It should perform certain critical business analyses (depending on the type of firm) such as:
- Sales analysis;
- Policy, risk, coverage, and billing analysis;
- Line-of-business segmentation;
- Marketing campaign analysis;
- Business performance analysis/underwriter profitability or results analysis by different business metrics;
- Expense control;
- Pricing (by demographic or postal code for example)
- Customer profile and loyalty analysis;
- Impact of cross- and up-selling;
- Accident year data analysis;
- Claims analysis;
- Loss analysis;
- Annuities analysis.

It should allow you to *automatically* capture key data gathered in your operational and administrative systems and move them into the data warehouse, without constant maintenance or scripting. The most important thing is to make sure that your data warehousing solution is modeled after YOUR business.

3.6.3 Analyzing the Strategic Value of Data Warehousing

To begin, let us examine the different concepts in business intelligence. Every day, a company gathers *data* from various sources: the claims system gathers data on the names, addresses, and claims of claimants. These basic elements are *data.* For example: "Claimant Bill Reed lives at 38 Green Terrace and has insurance Plan B." This is data. The fact that ABC Insurance Company collected $10 million in premiums and paid $10 million in claims is also data. What companies are looking for is a way to leverage the enormous amounts of data to gather useful *information.* Information gives you a more global idea of what is happening in some area of the business that you want to examine. "36.7% of policyholders in Iowa have Plan B, and we collected $3.67 million in premiums from them but paid out $7.34 million in claims. Plan B is our most popular." Once the information is gathered, it is now completely up to the company to turn it into *wisdom.* Wisdom allows a company to make key decisions based on the information they have at their disposal. "Our loss ratio in Iowa as a whole is 100%, but on Plan B, it's 200 percent! That's atrocious! Let's take a close look at Plan B to see why it isn't profitable." These concepts make up business intelligence.

The best way to access the information required is through the packaged solution. There are four points to consider:

1) The wisdom to be gained from the Plan B scenario will not come from the data warehouse. In fact, it will come from the methodology developed by ABC's management staff to act on certain situations. The role of the data warehousing staff, internal or external, ends when the information is delivered by the decision-support system; it is up to the knowledge workers to make the decisions.
2) The less resources involved in building an infrastructure, the more available to work out the details of the actual decision-making methodology. There is little point in waiting longer than is necessary on the construction of a data warehouse; it is simply a way to get that required information.
3) If increased market share in insurance companies is compared to the jackpot in poker, then the data warehousing infrastructure has become the "ante." The packaged data warehouse alternative simplifies and accelerates the ability to pay that ante. With the state of the insurance marketplace, data warehousing is necessary to get to that information faster.
4) Besides the money involved in building an infrastructure, the *time* required to do so can be staggering. It seems better to avoid that altogether. For a firm that wants competitive advantage through the timely delivery of information to employees, it makes sense that the implementation of the required decision-support system be done in the shortest time at the least cost, and with a measurable return on investment. In other words, it would appear that the "blood, sweat, and tears" method involved in designing, developing, and deploying such infrastructure is the less desirable option. The task at this point is to determine if it really is faster, less expensive, and ultimately more effective to buy the infrastructure rather than build it. This is discussed in the next subsection.

3.6.4 Addressing your Concerns

Conservative industries such as insurance are very apprehensive acquiring new infrastructures and embracing new business concepts. Some of these concerns lie in the reluctance to accept business intelligence as a must-have practice. Others are more basic. We attempt to uncover what some of the common issues with buying are, expose the underlying realities, and overcome them.

If you have Already Started Building

Changing direction in the middle of a project is one of the most common deterrents to accepting a packaged solution. Many insurance companies start building their solutions because they just do not know that packaged solutions exist. Usually, the central question here is: "How do we justify the money we've already spent on a custom solution if we change to a packaged one?" Companies tend to look inside to find out what technology and human

resources they have at their immediate disposal, and this is an excellent cost-saving measure. It does not preclude using an outside packaged solution by any means. Certainly there may be some unrecoverable costs incurred in the design, development, and deployment (D3) of a custom environment. These include human resources (time) spent in D3 or the opportunity cost of not spending those resources elsewhere.

When a company starts to build a decision-support system, they almost certainly acquire a number of skills and technologies that can be reused in future endeavors, such as a bought solution. For example, the knowledge gained by a company in D3 is usually reusable. Acquired hardware or software may also be reused, depending on the solution. This minimizes overlap of technology and provides a more accurate way to describe how much a packaged solution will cost in the end. Often, finding the reusable can mean the difference between a lengthy and expensive D3 process and a short, inexpensive one.

When the costs to bring a company to a certain point in their custom solution are known, the next task is to find out how much it will cost to complete the custom project, then compare that total cost to the cost of buying a new solution. Using the tallied amount of reusable assets, a firm can get a clear indication of what path to take. It is clear that when the cost of buying a solution with equivalent or better functionality and a faster time-to-market is less than it is to complete the custom project, it makes sense financially to switch to the bought solution. The things to consider are:

- What is the projected total per day D3 cost for the rest of the project?

While it is impossible to say for certain for all companies what that might be, it is fairly easy to calculate it for a particular company. If a company has been working internally on a custom project for an amount of time, t, then a full-service data-warehousing firm should be able to do the following:

1) Assess the current business and technological environment at time t,
2) Determine how much it is to complete the project from time $t$to the project's conclusion using different possible scenarios,
3) Add that cost to the cost of bringing the project to time $t$and finally,
4) Compare that total to the projection for the bought solution.

- How much longer will it take?

It is not difficult to estimate how long a packaged solution will take to deliver after a professional assessment of the environment. From beginning to end, a packaged solution takes a fraction of the time to deploy compared to a custom one. Even when a custom project is partly or mostly complete, a bought solution may still be faster to deploy. Furthermore, the amount of D3 is little to zero in the packaged solution.

- How much will extra hardware/software for the custom project cost and will that cost more than what is already offered in the packaged solution?

Again, maximizing the amount of reusables tends to equalize, or at least minimize the difference between the cost to complete a custom solution and the cost to purchase a packaged one. This is another metric that can be measured using an assessment.

- How much work will I need to do with this custom solution when the business expands? Will it be lesser/cheaper with a packaged alternative?

Unless the custom solution is being built with the extensibility of a packaged solution, it will not be cheaper to do extra work extending the custom solution when the business or data sets expand or change. Very often, companies find themselves having to rework their solution to accommodate expansion into other states, mergers, acquisitions, or corporate structure changes. This is almost the equivalent of building another warehouse from scratch. Furthermore, the kind of robust extensibility found in some packaged solutions took quite a while to develop. Doing it internally would delay the delivery of value-added decisional applications to knowledge workers. A needless delay, since that functionality is already available with a provider. Packaged solutions, by their very nature, are designed to be flexible enough to cover most existing and future business situations. Also, this flexibility is strengthened by any professional services used to customize the solution to a particular firm. Because of this, packaged solution providers have an easier time planning for and delivering such changes since they have the experience in managing many evolving business scenarios.

Does Building Save Money?

Despite all the technological advancements in the last 10–15 years, the insurance sector has historically been slow to embrace many of them. Years ago, policies, claims, and other administration systems were built by internal or outsourced development teams. A team was hired to develop it over a period of 2 years or so, a little at a time, and then implementation also took another 1–2 years. After all was said and done, final deliverables only happened after 3–4 years and millions spent.

Undertaking new technology in a business setting can be risky, but it is sometimes necessary. Insurers realized the importance of these operational systems to the running of their business, but when the opportunity to cut that development and implementation time came they were still reluctant to invest in the then-new technology. They were afraid of a lot of new infrastructure in a short amount of time. What if it did not deliver? Insurers were faced with a choice:

They could:

1) Continue to develop their own systems over a very long period of time without sufficient returns on their investments while other companies were leaping ahead in terms of vision, service, and results;

2) Forget about investing in these operational systems and continue the old way of doing things, hoping that the market will respond favorably to a company that saved money (in the short term, at least), or;
3) Buy a solution, customize it as necessary, and spend most of their efforts doing more business in a shorter amount of time using fewer resources.

In the end, the last option turned out to be the best one. Bought solutions on the operational side are standard in most of the insurance community today. On the decisional side, things are similar. The META Group reports that many data warehousing projects follow the 2:2:50 pattern. That is the average DW project takes two years to deploy, costs an average of $2 million, and has an expected failure rate of 50%. Other estimates cite overall failure rates of more than 60%.

The Data Warehouse Institute also reports rates higher than 70% across all industries. This is partly due to people not having all the technical expertise required for such a project. Companies often do not have the resources experienced in data warehousing. There is confusion in what the information systems staff can do and what they cannot. There is even confusion in what data warehousing means. In a decision-support context, the risks in building are very similar, if not identical, to the risks in building operational systems. Decision support and data warehousing are in the same position as claims and policy management systems were 10 years ago. The same arguments apply now as they did then. It no longer makes sense to develop internally. The commitment involved is not so great when you consider the alternative.

Do Internal Staff Members have Inside Knowledge not Available to Vendors?

This is partly true. While some vendors know the insurance industry very well, there may be some things in a particular company that are not standard. There may even be things that a bought solution may not address in its raw form. But there are ways around this. Using a vendor can allow more information systems projects to be completed in a shorter amount of time. An assessment where the vendor works with your company's internal business and IT staff ensures that the solution fits your business correctly. After that, your staff is free to do all the things that they have expertise in. For example, sometimes there are internal systems that need to be maintained, built, or rebuilt. Often, these projects can take precedence over data warehousing. Leveraging the vendor's expertise allows both initiatives to take place at the same time; therefore you leverage external expertise in the decision-support area.

It is always wise to focus on core competencies. Chances are the right vendor does nothing but data warehousing. It is their job to make your data useful to your knowledge workers. In that respect, this is often a good reason to allow such a vendor to support the business in decision support rather than

internal resources. Your internal resources would not become obsolete in such an environment; in fact they would be better used doing what they do best.

Is it More Beneficial to Work with Familiar Technology?

It is not necessarily the best idea to work with familiar technology, especially from a business point of view. Consider a company that uses technologies A, B, and C. The actuarial department realizes that in order to be more efficient in their rate making, they need a certain functionality that cannot be found in these technologies. It is found, however, that technology D will allow that functionality. While it may make sense initially for the IT department to favor the familiar technology, it should be noted that there may be a significant opportunity cost associated with not providing the analytical capabilities required by the actuaries. This may be measured in terms of lost revenue or time not optimized by knowledge workers. It is the business need that drives technology, not the other way around. In cases where there is an absolute requirement for certain in-house technologies to be in place, it may be possible for components in the packaged solution to be changed, depending on the flexibility of the solution chosen. Using certain technologies mainly for comfort to the exclusion of what may be better technology for the business can ultimately defeat certain business goals. The following Table 3.1 summarizes these arguments.

3.6.5 Introducing FellowDSS™

As stated, it is no longer necessary for an insurance company to use internal resources to deliver an effective administrative system to run the business (or line of business). Now the same insurers no longer have to spend the time, money, and other resources to deliver a robust decision-support system to *drive* the business.

FellowDSS™ is a decision-support system based on data warehousing technology. It has been developed with the broad needs of insurance companies in mind. It will drastically reduce the time, cost, resource requirements, and risk associated with implementing and deploying a data warehouse for insurance companies. In particular, this decision-support system has all the functionality, right out of the box, of a robust decision-support system. The next subsection demonstrates how this is done.

How FellowDSS™ Stacks Up

We mentioned that there are a number of requirements in a decision-support system; Table 3.2 summarizes how FellowDSS™ addresses these requirements and more.

Table 3.1. Summary

Problem	Origin	Reality
We have already started building a solution	Concern of not being able to cost justify the solution	The total size of all recoverable costs may be quite high compared to the size of the nonrecoverable costs. The cost to complete may be more than the cost to switch to a packaged solution with equivalent or superior functionality, when "time-to-market" is factored in. If the ability of the custom project to adapt to new business solutions is in question, then the packaged alternative is worthy of consideration, especially in this marketplace.
We will save money by doing it internally, and we do not have to allocate all of the funds immediately	Reluctance to commit	May be true in the short run, but not necessarily. Can start development and maintenance would be required. A change in the business could mean a complete reworking of the model, infrastructure and so on. Stretching out an implementation and missing out on business opportunities or not being able to effectively analyze practices can cost far more than any packaged data warehousing application.
Our developers have inside knowledge of our business that vendors could not possibly have.	Myth of internal staff obsolescence	They can work with your internal technical and business resources to customize the solution to reflect your business with little disruption to everyday operations. Then your resources can be better used in other projects.
Business intelligence is not that important anyway. Why get a whole new infrastructure anyway.	Resistance to change	Business intelligence is the price to pay in insurance today. See "the challenge of business intelligence across in the insurance industry."
We are very familiar with the existing technical environment	Resistance to change	The right vendor can teach you the technical environment, work with yours (where applicable), or a combination of both.

Table 3.2. How FellowDSS™ handles data warehousing requirements, by module

Requirements	**Fellow DSS™ Module (s)**
Customer profile, loyalty, and care analysis	Market Analysis
Marketing campaign analysis	Sales Management
Sales analysis	Sales Management
Policy, risk, coverage, and billing analysis	Market Analysis
Line of Business segmentation	Market Analysis
Business performance analysis/underwriter profitability or results analysis by business metric	Sales Management/Claims Analysis
Expense control	Agent Remuneration Analysis
Pricing (by demographic, postal code, impact of cross, and up-selling	Pricing Scenario Analysis/ Market Analysis
Accident year data Analysis	Claims Analysis
Claims Analysis	Claims Analysis
Channel Effectiveness	Distribution channel Effectiveness
Insurance Products Analysis	Insurance Products Movement
Financial Products Analysis	Annuities and Investment Products Movement
Loss Analysis	Claims Analysis

To learn more about FellowDSS™ business models, metadata definitions, database structures, transformation/loading scripts, advanced reporting and other algorithms, see the FellowDSS™ product brief.

3.7 Summary

The data mart/enterprise data warehouse environment increasingly becomes important as the data warehouse environment evolves to the corporate information factory. Data marts serve departments while the enterprise data warehouse serves the entire corporation. The enterprise data warehouse is shaped by very granular, simple data. Typically the enterprise data warehouse structure is normalized or "lightly denormalized." The data mart structures are shaped by the requirements of the departments that own them. The predominant structure for the world of data marts is the star join or the snowflake structure.

The community of users for the data mart/enterprise data warehouse environment includes farmers and explorers. Farmers and explorers have very

different characteristics. Farmers have a strong affinity for data marts while explorers have a strong affinity for the enterprise data warehouse. The enterprise data warehouse serves many other architectural structures, other than the data marts. The enterprise data warehouse is served by very different technology than the data marts.

In order to compete in today's competitive healthcare marketplace, organizations must listen to everything that is going on around them—they just cannot focus on their own corner of the world. Data warehouses, are in essence, a sophisticated hearing aid that allows healthcare organizations to hear what their customers and their providers are saying, thereby offering enhanced service and cultivating long-term customer relationships. By using data warehouses to make data-driven decisions, organizations will no longer have to rely on hit-and-miss tactics to guess at what makes effective outcomes and what customers need. In this marketplace, data warehouses are not merely a luxury, or even just a competitive advantage. They are a matter of survival.

Also in this section the data warehousing in telecommunication industry is discussed in detail. A single technology for implementing security meeting the required performance target is yet to be devised but the existing technologies can be used in combination to produce satisfactory results. In this section an attempt is made at trying to devise such a security model. XML with the proper extensions to its different technologies looks a likely candidate to be used in solving the security problems in data warehouse but still it is a long way from perfection; nonetheless the future does looks promising.

This section has explored two main methods of establishing a business intelligence environment in an insurance company: building a solution from the ground up and buying an existing solution. We conclude that buying the solution and then customizing it to a particular environment is the most effective method strategically, technologically, and financially. FellowDSS^{TM} is the method of choice for accessing that critical information. For more information, read "The Challenge of Business Intelligence Access for the Insurance Industry" as well as the FellowDSS^{TM} product brief, both available from SQLiaison representative.

3.8 Review Questions

1. Explain in detail about business process re-engineering.
2. With a star cube, explain data mart technology.
3. State some of the different structures and uses of data mart.
4. Explain about data warehouse in the enterprise technology.
5. Draw the Informix architecture and state its advantages.
6. How to build the data mart/data warehouse infrastructure?
7. Explain about the amount of history stored in data warehouse.
8. How to share data among data marts and what is the efficiency of sharing?

9. What are subdata marts and state its uses?
10. What is refreshment cycle as used in warehouse technology?
11. Give details on external data and operational data stores based on their usage.
12. What is distributed metadata?
13. How to manage warehouse environment based on applications?
14. Explain in detail on data warehousing for health care.
15. What to expect when beginning a data warehouse implementation?
16. What are the issues in industry that data warehouse addresses?
17. Write about the issues and challenges faced when data mining is applied to telecom industries.
18. Explain in detail about the various management involved in telecom life-cycle.
19. What are the security issues in data warehouse technology?
20. Compare performance versus security.
21. What are techniques and technologies used to tackle the security issues in data warehouse?
22. Explain the steps involved in implementing security with suitable model structure.
23. What are the components and the working steps involved in security model?
24. State the fundamental choices for insurers in data warehousing.
25. How to analyze the strategic value of data warehousing as applicable in insurance?
26. Explain in detail on Fellow DSS^{TM} stack up with its advantages.