From
Decision Support and Business Intelligence Systems - Efraim Turban

CHAPTER

8

Data Warehousing

LEARNING OBJECTIVES

1 Understand the basic definitions and concepts of data warehouses

2 Understand data warehousing architectures

3 Describe the processes used in developing and managing data warehouses

4 Explain data warehousing operations

5 Explain the role of data warehouses in decision support

6 Explain data integration and the extraction, transformation, and load (ETL) processes

7 Describe real-time (active) data warehousing

8 Understand data warehouse administration and security issues

The concept of data warehousing has been around since the late 1980s. This chapter provides the foundation for an important type of database, called a *data warehouse*, which is primarily used for decision support and provides improved analytical capabilities. We discuss data warehousing in the following sections:

8.1 OPENING VIGNETTE: DIRECTV THRIVES WITH ACTIVE DATA WAREHOUSING

As an example of how an interactive data warehousing and business intelligence product can spread across the enterprise, consider the case of DirecTV. Using software solutions from Teradata and GoldenGate, DirecTV developed a product that integrates its data assets in near real time throughout the enterprise. The company's data warehouse director, Jack Gustafson, has said publically that the product has paid for itself over and over again through its continuous use. For DirecTV, a technical decision to install a real-time transactional data management solution has delivered business benefits far beyond the technical ones originally anticipated.

DirecTV, which is known for its direct television broadcast satellite service, has been a regular contributor to the evolution of TV with its advanced HD programming, interactive features, digital video recording services, and electronic program guides. Employing more than 13,000 people across the United States and Latin America, DirecTV's 2008 revenues reached $20 billion, with total subscriber numbers approaching 50 million.

Problem

Midst of a continuing rapid growth, DirecTV faced the challenge of dealing with high transactional data volumes created by an escalating number of daily customer calls. Accommodating such a large data volume, along with rapidly changing market conditions, was one of DirecTV's key challenges. Several years ago, the company began looking for a better solution to providing the business side with daily reports on its call-center activities. Management wanted reports that could be used in many ways, including measuring and maintaining customer service, attracting new customers, and preventing customer churn. Equally important, the technical group at DirecTV wanted to reduce the resource load that its current data management system imposed on its CPUs.

Even though an early implementation of the data warehouse was addressing the company's needs fairly well, as business continued to grow its limitations became clear. Before the active data warehouse solution, the data was pulled from the server every night in batch mode, a process that was taking too long and straining the system. A daily batch-data upload to the data warehouse had long been (and for many companies, still is) the standard procedure. If the timeliness of the data is not a part of your business competitiveness, such a daily upload procedure may very well work for your organization. Unfortunately, this was not the case for DirecTV. Functioning within a highly dynamic consumer market, with a very high call volume to manage, DirecTV's business users needed to access the data from customer calls in a timely fashion.

Solution

Originally, the goal of the new data warehouse system was to send fresh data to the call center at least daily, but once the capabilities of the integrated solutions became apparent, that goal dropped to fresh data every 15 minutes. "We [then] wanted data latency of less than 15 minutes across the WAN running between different cities," Gustafson explains.

A secondary goal of the project was to simplify changed data capture to reduce the amount of maintenance required from developers. Although data sourcing across multiple platforms was not part of the initial requirement, that changed once DirecTV saw the capabilities of the GoldenGate integration system. GoldenGate allows the integration of a range of data management systems and platforms. At DirecTV, that included Oracle, the HP NonStop platform, an IBM DB2 system, and the Teradata data warehouse. "With GoldenGate, we weren't tied to one system," Gustafson says. "That also appealed to us. We're sourcing out of call logs, but we're also sourcing out of NonStop and other data sources. We thought, if we're going to buy a tool to do this, we want it to work with all the platforms we support."

Results

As the capabilities of the system became increasingly clear, its potential benefits to the business also became apparent. "Once we set it up, a huge business benefit [turned out to be] that it allowed us to measure our churn in real time," Gustafson says. "We said, 'Now that we have all these reports in real time, what can we do with them?'" One answer was to use the data to immediately reduce churn by targeting specific customers. With fresh data at their fingertips, call center sales personnel were able to contact a customer who had just asked to be disconnected and make a new sales offer to retain the customer just hours later the same day. Once the IT group set up the necessary reporting tools, sales campaigns could target specific customers for retention and prioritize them for special offers. That sort of campaign has clearly worked: "Our churn has actually gone down since we've implemented this program," Gustafson says. "Analysts are just raving about how great we're doing compared to our competitors in this area. A lot of it comes down to using this real-time copy to do analysis on customers, and to [make a fresh] offer to them the same day."

The system has also been set up to log customer service calls, reporting back constantly on technical problems that are reported in the field. That allows management to better evaluate and react to field reports, improving service and saving on dispatching technicians. Real-time call-center reports can also be produced to help manage the center's workload based on daily information on call volumes. Using that data, management can compare daily call volumes with historical averages for exception reporting.

In another business-centric use that was not originally anticipated, the company is using real-time operational reports for both order management and fraud detection. With access to real-time order information on new customers, fraud management experts can examine the data, and then use that information to weed out fraudulent orders. "That saves us rolling a truck, which drives [labor] and product costs down," Gustafson points out.

Questions for the Opening Vignette

1. Why is it important for DirecTV to have an active data warehouse?
2. What were the challenges DirecTV faced on its way to having an integrated active data warehouse?
3. Identify the major differences between a traditional data warehouse and an active data warehouse, such as the one implemented at DirecTV.
4. What strategic advantage can DirecTV derive from the real-time system as opposed to a traditional information system?
5. Why do you think large organizations like DirecTV cannot afford not to have a capable data warehouse?

What We Can Learn from This Vignette

The opening vignette illustrates the strategic value of implementing an active data warehouse, along with its supporting BI methods. DirecTV was able to leverage its data assets spread throughout the enterprise to be used by knowledge workers wherever and whenever they are needed. The data warehouse integrated various databases throughout the organization into a single, in-house enterprise unit to generate a single version of the truth for the company, putting all employees on the same page. Furthermore, the data was made available in real time to the decision makers who needed it, so they could use it in their decision making, ultimately leading to a strategic competitive advantage in the industry. The key lesson here is that a real-time, enterprise-level active data warehouse combined with a strategy for its use in decision support can result in significant benefits (financial and otherwise) for an organization.

Sources: L. L. Briggs, "DirecTV Connects with Data Integration Solution," *Business Intelligence Journal*, Vol. 14, No. 1, 2009, pp. 14–16; "DirecTV Enables Active Data Warehousing with GoldenGate's Real-Time Data Integration Technology," *Information Management Magazine*, January 2008; directv.com.

8.2 DATA WAREHOUSING DEFINITIONS AND CONCEPTS

Using real-time data warehousing in conjunction with DSS and BI tools is an important way to conduct business processes. The opening vignette demonstrates a scenario in which a real-time active data warehouse supported decision making by analyzing large amounts of data from various sources to provide rapid results to support critical processes. The single version of the truth stored in the data warehouse and provided in an easily digestible form expands the boundaries of DirecTV's innovative business processes. With real-time data flows, DirecTV can view the current state of its business and quickly identify problems, which is the first and foremost step toward solving them analytically. In addition, customers can obtain real-time information on their subscriptions, TV services, and other account information, so the system also provides a significant competitive advantage over competitors.

Decision makers require concise, dependable information about current operations, trends, and changes. Data are often fragmented in distinct operational systems, so managers often make decisions with partial information, at best. Data warehousing cuts through this obstacle by accessing, integrating, and organizing key operational data in a form that is consistent, reliable, timely, and readily available, wherever and whenever needed.

What Is a Data Warehouse?

In simple terms, a **data warehouse (DW)** is a pool of data produced to support decision making; it is also a repository of current and historical data of potential interest to managers throughout the organization. Data are usually structured to be available in a form ready for analytical processing activities (i.e., online analytical processing [OLAP], data mining, querying, reporting, and other decision support applications). A data warehouse is a subject-oriented, integrated, time-variant, nonvolatile collection of data in support of management's decision-making process.

Characteristics of Data Warehousing

A common way of introducing data warehousing is to refer to its fundamental characteristics (see Inmon, 2005):

- ***Subject oriented.*** Data are organized by detailed subject, such as sales, products, or customers, containing only information relevant for decision support. Subject orientation enables users to determine not only how their business is performing, but why. A data warehouse differs from an operational database in that most operational databases have a product orientation and are tuned to handle transactions that update the database. Subject orientation provides a more comprehensive view of the organization.
- ***Integrated.*** Integration is closely related to subject orientation. Data warehouses must place data from different sources into a consistent format. To do so, they must deal with naming conflicts and discrepancies among units of measure. A data warehouse is presumed to be totally integrated.
- ***Time variant (time series).*** A warehouse maintains historical data. The data do not necessarily provide current status (except in real-time systems). They detect trends, deviations, and long-term relationships for forecasting and comparisons, leading to decision making. Every data warehouse has a temporal quality. Time is the one important dimension that all data warehouses must support. Data for analysis from multiple sources contains multiple time points (e.g., daily, weekly, monthly views).
- ***Nonvolatile.*** After data are entered into a data warehouse, users cannot change or update the data. Obsolete data are discarded, and changes are recorded as new data.

These characteristics enable data warehouses to be tuned almost exclusively for data access. Some additional characteristics may include the following:

- ***Web based.*** Data warehouses are typically designed to provide an efficient computing environment for Web-based applications.
- ***Relational/multidimensional.*** A data warehouse uses either a relational structure or a multidimensional structure. A recent survey on multidimensional structures can be found in Romero and Abelló (2009).

- ***Client/server.*** A data warehouse uses the client/server architecture to provide easy access for end users.
- ***Real time.*** Newer data warehouses provide real-time, or active, data access and analysis capabilities (see Basu, 2003; and Bonde and Kuckuk, 2004).
- ***Include metadata.*** A data warehouse contains metadata (data about data) about how the data are organized and how to effectively use them.

Whereas a data warehouse is a repository of data, data warehousing is literally the entire process (see Watson, 2002). Data warehousing is a discipline that results in applications that provide decision support capability, allows ready access to business information, and creates business insight. The three main types of data warehouses are data marts, operational data stores (ODS), and enterprise data warehouses (EDW). In addition to discussing these three types of warehouses next, we also discuss metadata.

Data Marts

Whereas a data warehouse combines databases across an entire enterprise, a **data mart** is usually smaller and focuses on a particular subject or department. A data mart is a subset of a data warehouse, typically consisting of a single subject area (e.g., marketing, operations). A data mart can be either dependent or independent. A **dependent data mart** is a subset that is created directly from the data warehouse. It has the advantages of using a consistent data model and providing quality data. Dependent data marts support the concept of a single enterprise-wide data model, but the data warehouse must be constructed first. A dependent data mart ensures that the end user is viewing the same version of the data that is accessed by all other data warehouse users. The high cost of data warehouses limits their use to large companies. As an alternative, many firms use a lower-cost, scaled-down version of a data warehouse referred to as an *independent data mart*. An **independent data mart** is a small warehouse designed for a strategic business unit (SBU) or a department, but its source is not an EDW.

Operational Data Stores

An **operational data store (ODS)** provides a fairly recent form of customer information file (CIF). This type of database is often used as an interim staging area for a data warehouse. Unlike the static contents of a data warehouse, the contents of an ODS are updated throughout the course of business operations. An ODS is used for short-term decisions involving mission-critical applications rather than for the medium- and long-term decisions associated with an EDW. An ODS is similar to short-term memory in that it stores only very recent information. In comparison, a data warehouse is like long-term memory because it stores permanent information. An ODS consolidates data from multiple source systems and provides a near-real-time, integrated view of volatile, current data. The exchange, transfer, and load (ETL) processes (discussed later in this chapter) for an ODS are identical to those for a data warehouse. Finally, **oper marts** (see Imhoff, 2001) are created when operational data needs to be analyzed multidimensionally. The data for an oper mart come from an ODS.

Enterprise Data Warehouses (EDW)

An **enterprise data warehouse (EDW)** is a large-scale data warehouse that is used across the enterprise for decision support. It is the type of data warehouse that DirecTV developed, as described in the opening vignette. The large-scale nature provides integration of data from many sources into a standard format for effective BI and decision support applications. EDW are used to provide data for many types of DSS, including CRM, supply-chain management (SCM), business performance management (BPM), business activity monitoring (BAM), product lifecycle management (PLM), revenue management, and sometimes even knowledge management systems (KMS). Application Case 8.1 shows the enormous benefits a large company can realize from EDW, if it is designed and implemented correctly.

APPLICATION CASE 8.1

Enterprise Data Warehouse Delivers Cost Savings and Process Efficiencies

Founded in 1884 in Dayton, Ohio, the NCR Corporation is now a $5.6 billon NYSE-listed company providing technology solutions worldwide in the retail, financial, insurance, communications, manufacturing, and travel and transportation industries. NCR solutions include store automation and automated teller machines, consulting services, media products, and hardware technology.

When acquired by AT&T in 1991, NCR operated on an autonomous country- and product-centric structure, in which each country made its own decisions about product and service offerings, marketing, and pricing and developed its own processes and reporting norms. Under the country model, dozens of different financial and operational applications were required to capture the total results of the company, by no means an enterprise solution.

In 1997, when NCR was spun off on its own again, company operations were losing substantial amounts of money every day. The spin-off provided NCR with the much-needed funds to engage in the deep process changes required to maintain and strengthen its competitive position in the global market and to undertake the transformation to a truly global enterprise.

The goal was to move from a primarily hardware-focused and country-centric organizational model to an integrated, solution-oriented business structure with a global focus. To do this, NCR needed to globalize, centralize, and integrate its vast store of information resources. Only then could it gain effective control over the necessary business changes. NCR's EDW initiative was critical to the company's successful transformation and would be vital to the successful deployment of a new worldwide, single-instance, enterprise resource planning (ERP) system planned for several years later.

NCR Finance and Worldwide Customer Services (WCS) led the drive for implementation of the EDW. Business teams from Finance and WCS, Financial Information Delivery (FID), and Global Information Systems (GIS), respectively, worked closely with the EDW team to ensure that IT understood the business requirements for the new structure. The Teradata system was chosen for its scalability, its flexibility to support unstructured queries and high numbers of concurrent users, and its relatively low maintenance costs.

The enormous potential of the EDW spread throughout the company, driving organizational and process changes in Finance, where the financial close cycle was reduced from 14 days to 6 and worldwide reporting integrity standards were established; in WCS, where individual customer profitability profiles and improvement plans were made possible; and in Sales and Marketing, Operations and Inventory Management, and Human Resources. ERP operational standardization and a dramatic improvement in the business of serving its customers mean that NCR is poised for the future. Internally and externally, NCR has become a global solution provider, supported by global business processes.

The returns have already been superb. Not only has the EDW project proved to be more than self-funding at the project cost level, but revenue generation is around the corner. Some of the benefits include $100 million in annual savings in inventory-carrying costs, a $200 million sustainable reduction in accounts receivable, a $50 million reduction in annual finance costs, and $22 million in cost savings over the first 5 years of the EDW implementation for WCS.

There is still much to be done and significant value to be realized by the project. Beyond cost savings and process efficiencies, the strategy going forward is to use the EDW to drive growth.

Although the EDW project was not undertaken as a profit-producing opportunity, it was self-funding. The cost savings far exceeded the expense of implementation. As the EDW matures, growth-focused goals are developing, and the EDW will drive profits in the future. The quantified benefits of the EDW speak for themselves. There are many more benefits of a qualitative nature. A sampling of both follows.

Qualitative Benefits

- Reduced financial close cycle from 14 days to 6
- Heightened reporting integrity to corporate standards
- Created individual customer profitability profiles and improvement plans
- Provided consistent worldwide reporting processes
- Improved on-time delivery
- Decreased obsolescence due to enhanced inventory management

Quantified Benefits

- $50 million reduction in annual finance controllership costs
- $200 million sustainable reduction in accounts receivable, which translates into $20 million per year savings in accounts receivable carrying costs
- $100 million sustainable reduction in finished inventory, which, in turn, equals a $10 million per year savings in inventory carrying costs
- $22 million cost savings over the first 5 years of the EDW implementation for WCS, including automation of the SLA reporting to customers, headcount savings, and lower customer maintenance costs
- $10 million for improved supply chain management
- $6.1 million net present value of cost reductions over 5 years as a result of reducing headcount from the finance and auditing (F&A) reporting function
- $3.5 million reduction in telecommunications costs
- $3 million in savings through the reduction of ERP transition costs
- $1.7 million saved on report development costs in the rollout from Oracle 10.7 and 11 to 11i, for reports that do not have to be custom-written for Oracle

Source: Teradata, "Enterprise Data Warehouse Delivers Cost Savings and Process Efficiencies," **teradata.com/t/resources/case-studies/NCR-Corporation-eb4455/** (accessed June 2009).

Metadata

Metadata are data about data (e.g., see Sen, 2004; and Zhao, 2005). Metadata describe the structure of and some meaning about data, thereby contributing to their effective or ineffective use. Mehra (2005) indicated that few organizations really understand metadata, and fewer understand how to design and implement a metadata strategy. Metadata are generally defined in terms of usage as technical or business metadata. Pattern is another way to view metadata. According to the pattern view, we can differentiate between syntactic metadata (i.e., data describing the syntax of data), structural metadata (i.e., data describing the structure of the data), and semantic metadata (i.e., data describing the meaning of the data in a specific domain).

We next explain traditional metadata patterns and insights into how to implement an effective metadata strategy via a holistic approach to enterprise metadata integration. The approach includes ontology and metadata registries; enterprise information integration (EII); extraction, transformation, and load (ETL); and service-oriented architectures (SOA). Effectiveness, extensibility, reusability, interoperability, efficiency and performance, evolution, entitlement, flexibility, segregation, user interface, versioning, versatility, and low maintenance cost are some of the key requirements for building a successful metadata-driven enterprise.

According to Kassam (2002), business metadata comprise information that increases our understanding of traditional (i.e., structured) data. The primary purpose of metadata should be to provide context to the reported data; that is, it provides enriching information that leads to the creation of knowledge. Business metadata, though difficult to provide efficiently, release more of the potential of structured data. The context need not be the same for all users. In many ways, metadata assist in the conversion of data and information into knowledge. Metadata form a foundation for a metabusiness architecture (see Bell, 2001). Tannenbaum (2002) described how to identify metadata requirements. Vaduva and Vetterli (2001) provided an overview of metadata management for data warehousing. Zhao (2005) described five levels of metadata management maturity: (1) ad hoc, (2) discovered, (3) managed, (4) optimized, and (5) automated. These levels help in understanding where an organization is in terms of how and how well it uses its metadata.

The design, creation, and use of metadata—descriptive or summary data about data—and its accompanying standards may involve ethical issues. There are ethical considerations involved in the collection and ownership of the information contained in metadata, including privacy and intellectual property issues that arise in the design, collection, and dissemination stages (for more, see Brody, 2003).

Section 8.2 Review Questions

1. What is a data warehouse?
2. How does a data warehouse differ from a database?
3. What is an ODS?
4. Differentiate among a data mart, an ODS, and an EDW.
5. Explain the importance of metadata.

8.3 DATA WAREHOUSING PROCESS OVERVIEW

Organizations, private and public, continuously collect data, information, and knowledge at an increasingly accelerated rate and store them in computerized systems. Maintaining and using these data and information become extremely complex, especially as scalability issues arise. In addition, the number of users needing to access the information continues to increase as a result of improved reliability and availability of network access, especially the Internet. Working with multiple databases, either integrated in a data warehouse or not, has become an extremely difficult task requiring considerable expertise, but it can provide immense benefits far exceeding its cost (see the opening vignette and Application Case 8.2).

APPLICATION CASE 8.2

Data Warehousing Supports First American Corporation's Corporate Strategy

First American Corporation changed its corporate strategy from a traditional banking approach to one that was centered on CRM. This enabled First American to transform itself from a company that lost $60 million in 1990 to an innovative financial services leader a decade later. The successful implementation of this strategy would not have been possible without its VISION data warehouse, which stores information about customer behavior, such as products used, buying preferences, and client-value positions. VISION provides:

- Identification of the top 20 percent of profitable customers
- Identification of the 40 to 50 percent of unprofitable customers
- Retention strategies
- Lower-cost distribution channels
- Strategies to expand customer relationships
- Redesigned information flows

Access to information through a data warehouse can enable both evolutionary and revolutionary change. First American Corporation achieved revolutionary change, moving itself into the "sweet 16" of financial services corporations.

Sources: Adapted from B. L. Cooper, H. J. Watson, B. H. Wixom, and D. L. Goodhue, "Data Warehousing Supports Corporate Strategy at First American Corporation," *MIS Quarterly*, Vol. 24, No. 4, 2000, pp. 547–567; and B. L. Cooper, H. J. Watson, B. H. Wixom, and D. L. Goodhue, "Data Warehousing Supports Corporate Strategy at First American Corporation," SIM International Conference, Atlanta, August 15–19, 1999.

Many organizations need to create data warehouses—massive data stores of time-series data for decision support. Data are imported from various external and internal resources and are cleansed and organized in a manner consistent with the organization's needs. After the data are populated in the data warehouse, data marts can be loaded for a specific area or department. Alternatively, data marts can be created first, as needed, and then integrated into an EDW. Often, though, data marts are not developed, but data are simply loaded onto PCs or left in their original state for direct manipulation using BI tools.

In Figure 8.1, we show the data warehouse concept. The following are the major components of the data warehousing process:

- ***Data sources.*** Data are sourced from multiple independent operational "legacy" systems and possibly from external data providers (such as the U.S. Census). Data may also come from an OLTP or ERP system. Web data in the form of Web logs may also feed a data warehouse.
- ***Data extraction.*** Data are extracted using custom-written or commercial software called ETL.
- ***Data loading.*** Data are loaded into a staging area, where they are transformed and cleansed. The data are then ready to load into the data warehouse.
- ***Comprehensive database.*** Essentially, this is the EDW to support all decision analysis by providing relevant summarized and detailed information originating from many different sources.
- ***Metadata.*** Metadata are maintained so that they can be assessed by IT personnel and users. Metadata include software programs about data and rules for organizing data summaries that are easy to index and search, especially with Web tools.
- ***Middleware tools.*** Middleware tools enable access to the data warehouse. Power users such as analysts may write their own SQL queries. Others may employ a managed query environment, such as Business Objects, to access data. There are many front-end applications that business users can use to interact with data stored in the data repositories, including data mining, OLAP, reporting tools, and data visualization tools.

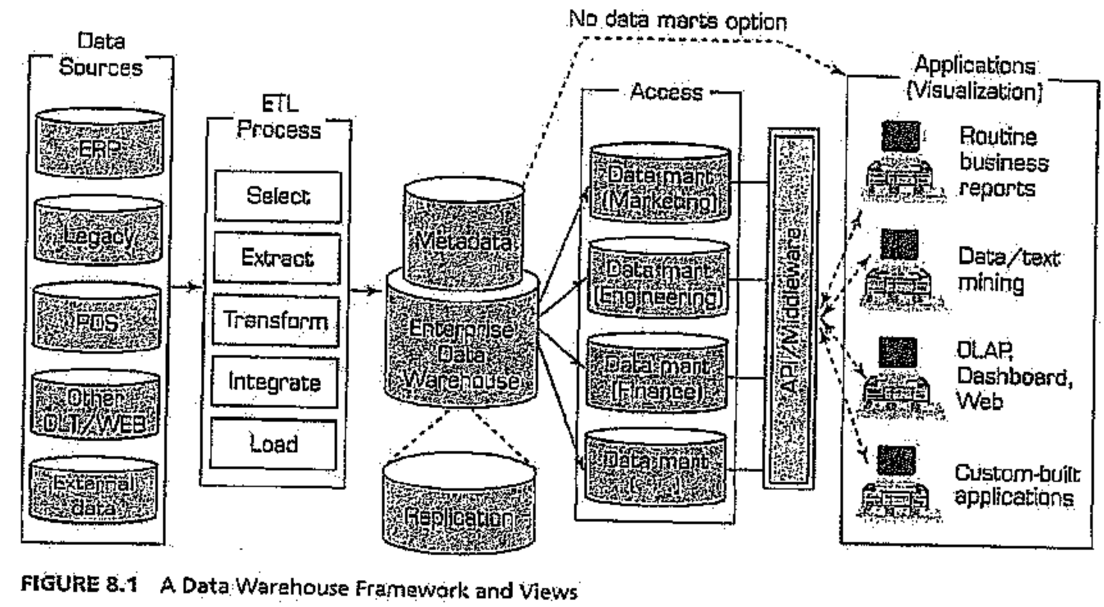


FIGURE 8.1 A Data Warehouse Framework and Views

Section 8.3 Review Questions

1. Describe the data warehousing process.
2. Describe the major components of a data warehouse.
3. Identify the role of middleware tools.

8.4 DATA WAREHOUSING ARCHITECTURES

There are several basic architectures for data warehousing. Two-tier and three-tier architectures are common (see Figures 8.2 and 8.3), but sometimes there is simply one tier. Hoffer et al. (2007) distinguished among these architectures by dividing the data warehouse into three parts:

1. The data warehouse itself, which contains the data and associated software
2. Data acquisition (back-end) software, which extracts data from legacy systems and external sources, consolidates and summarizes them, and loads them into the data warehouse
3. Client (front-end) software, which allows users to access and analyze data from the warehouse (a DSS/BI/business analytics [BA] engine)

In a three-tier architecture, operational systems contain the data and the software for data acquisition in one tier (i.e., the server), the data warehouse is another tier, and the third tier includes the DSS/BI/BA engine (i.e., the application server) and the client (see Figure 8.2). Data from the warehouse are processed twice and deposited in an additional multidimensional database, organized for easy multidimensional analysis and presentation, or replicated in data marts. The advantage of the three-tier architecture is its separation of the functions of the data warehouse, which eliminates resource constraints and makes it possible to easily create data marts.

In a two-tier architecture, the DSS engine physically runs on the same hardware platform as the data warehouse (see Figure 8.3). Therefore, it is more economical than the three-tier structure. The two-tier architecture can have performance problems for large data warehouses that work with data-intensive applications for decision support.

Much of the common wisdom assumes an absolutist approach, maintaining that one solution is better than the other, despite the organization's circumstances and unique needs. To further complicate these architectural decisions, many consultants and software vendors focus on one portion of the architecture, therefore limiting their capacity and motivation to assist an organization through the options based on its needs. But these aspects are being questioned and analyzed. For example, Ball (2005) provided decision criteria for organizations that plan to implement a BI application and have already determined their need for multidimensional data marts but need help determining the appropriate tiered architecture. His criteria revolve around forecasting needs for space and speed of access (see Ball, 2005, for details).

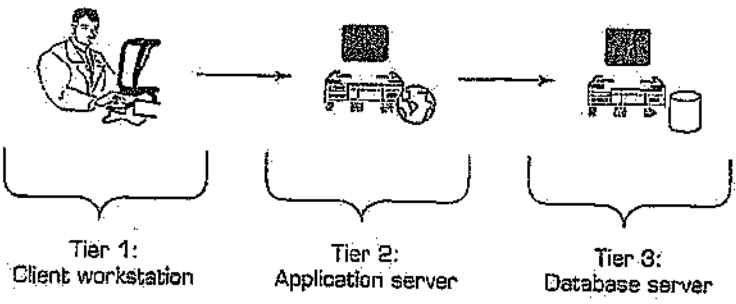


FIGURE 8.2 Architecture of a Three-Tier Data Warehouse

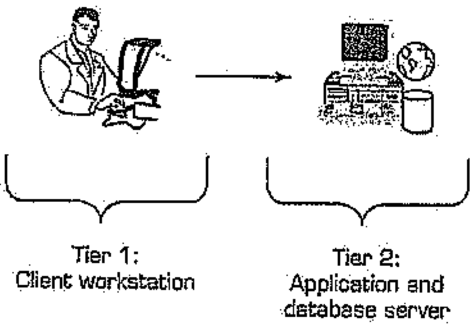


FIGURE 8.3 Architecture of a Two-Tier Data Warehouse

Data warehousing and the Internet are two key technologies that offer important solutions for managing corporate data. The integration of these two technologies produces Web-based data warehousing. In Figure 8.4, we show the architecture of Web-based data warehousing. The architecture is three tiered and includes the PC client, Web server, and application server. On the client side, the user needs an Internet connection and a Web browser (preferably Java enabled) through the familiar graphical user interface (GUI). The Internet/intranet/extranet is the communication medium between clients and servers. On the server side, a Web server is used to manage the inflow and outflow of information between client and server. It is backed by both a data warehouse and an application server. Web-based data warehousing offers several compelling advantages, including ease of access, platform independence, and lower cost.

The Vanguard Group moved to a Web-based, three-tier architecture for its enterprise architecture to integrate all its data and provide customers with the same views of data as internal users (Dragoon, 2003). Likewise, Hilton migrated all its independent client/server systems to a three-tier data warehouse, using a Web design enterprise system. This change involved an investment of \$3.8 million (excluding labor) and affected 1,500 users. It increased processing efficiency (speed) by a factor of six. When it was deployed, Hilton expected to save \$4.5 million to \$5 million annually. Finally, Hilton experimented with Dell's clustering (i.e., parallel computing) technology to enhance scalability and speed (see Anthes, 2003).

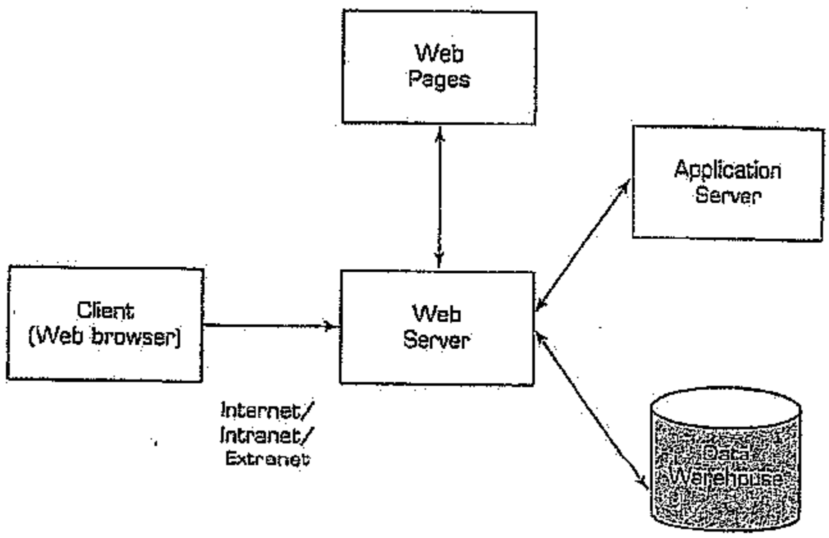


FIGURE 8.4 Architecture of Web-Based Data Warehousing

Web architectures for data warehousing are similar in structure to other data warehousing architectures, requiring a design choice for housing the Web data warehouse with the transaction server or as a separate server(s). Page-loading speed is an important consideration in designing Web-based applications; therefore, server capacity must be planned carefully.

Several issues must be considered when deciding which architecture to use. Among them are the following:

- ***Which database management system (DBMS) should be used?*** Most data warehouses are built using relational database management systems (RDBMS). Oracle (Oracle Corporation, **oracle.com**), SQL Server (Microsoft Corporation, **microsoft.com/sql/**), and DB2 (IBM Corporation, **306.ibm.com/software/data/db2/**) are the ones most commonly used. Each of these products supports both client/server and Web-based architectures.
- ***Will parallel processing and/or partitioning be used?*** Parallel processing enables multiple CPUs to process data warehouse query requests simultaneously and provides scalability. Data warehouse designers need to decide whether the database tables will be partitioned (i.e., split into smaller tables) for access efficiency and what the criteria will be. This is an important consideration that is necessitated by the large amounts of data contained in a typical data warehouse. A recent survey on parallel and distributed data warehouses can be found in Furtado (2009). Teradata (**teradata.com**) has successfully adopted and often commented on its novel implementation of this approach.
- ***Will data migration tools be used to load the data warehouse?*** Moving data from an existing system into a data warehouse is a tedious and laborious task. Depending on the diversity and the location of the data assets, migration may be a relatively simple procedure or (in contrary) a months-long project. The results of a thorough assessment of the existing data assets should be used to determine whether to use migration tools, and if so, what capabilities to seek in those commercial tools.
- ***What tools will be used to support data retrieval and analysis?*** Often it is necessary to use specialized tools to periodically locate, access, analyze, extract, transform, and load necessary data into a data warehouse. A decision has to be made on (i) developing the migration tools in-house, (ii) purchasing them from a third-party provider, or (iii) using the ones provided with the data warehouse system. Overly complex, real-time migrations warrant specialized third-party ETL tools.

Alternative Architectures

Data warehouse architecture design viewpoints can be generally categorized into enterprise-wide data warehouse design and data mart design (Golfarelli and Rizzi, 2009). In Figure 8.5 (parts a–e), we show some alternatives to the basic architectural design types that are neither pure EDW nor pure DM, but in between or beyond the traditional architectural structures. Notable new ones include hub-and-spoke and federated architectures. The five architectures shown in Figure 8.5 (parts a–e) are proposed by Ariyachandra and Watson (2006b). Previously, in an extensive study, Sen and Sinha (2005) identified 15 different data warehousing methodologies. The sources of these methodologies are classified into three broad categories: core-technology vendors, infrastructure vendors, and information-modeling companies.

The data warehousing literature provides additional discussions about a variety of architectures, such as independent data marts, data mart bus architecture with linked dimensional data marts, and federated data marts (see Ariyachandra and Watson, 2005,

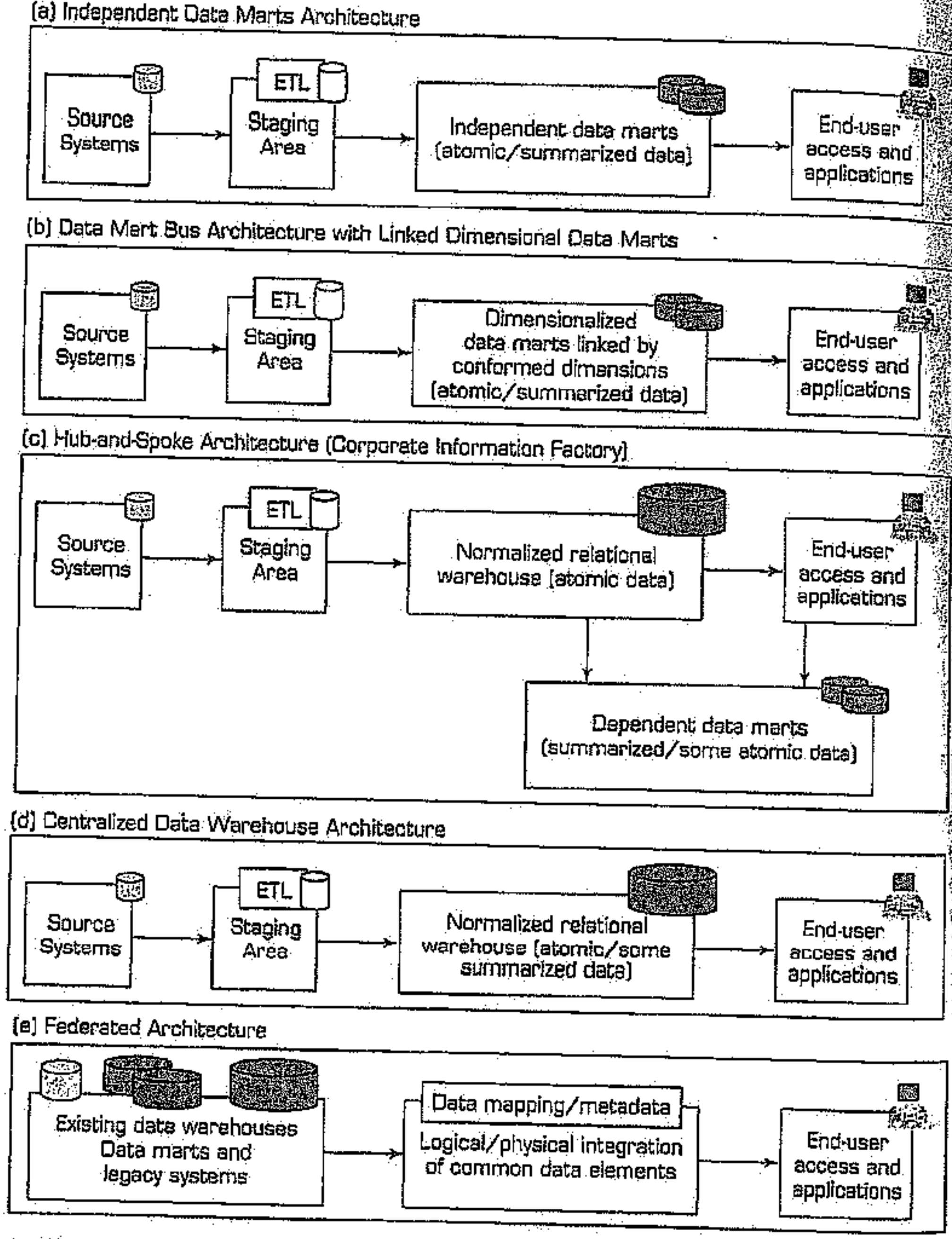
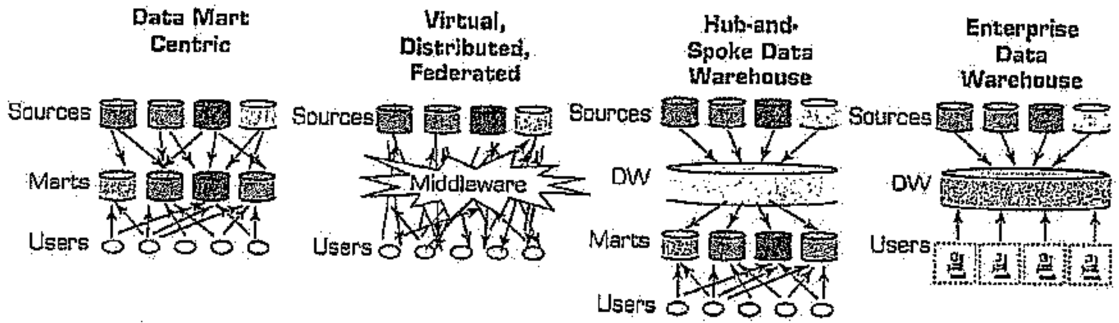


FIGURE 8.5 Alternative Data Warehouse Architectures *Source:* Adapted from T. Ariyachandra and H. Watson, "Which Data Warehouse Architecture Is Most Successful?" *Business Intelligence Journal*, Vol. 11, No. 1, First Quarter, 2006, pp. 4–6.

2006a); see Figure 8.6. In independent data marts, the marts are developed to operate independently of each another. Thus, they have inconsistent data definitions and different dimensions and measures, making it difficult to analyze data across the marts (i.e., it is difficult, if not impossible, to get to the "one version of the truth"). In a hub-and-spoke architecture, attention is focused on building a scalable and maintainable infrastructure; it is developed in an iterative way, subject area by subject area, and dependent data marts are developed. A centralized data warehouse is similar to the hub-and-spoke architecture



	Independent Data Marts	Leave Data Where It Lies	Dependent Data Marts	Centralized Integrated Data With Direct Access
Pros	• Easy to Build Organizationally • Easy to Build Technically	• No need for ETL • No need for separate platform	• Allows easier customization of user interfaces and reports	• Business enterprise view • Design consistency and data quality • Data reusability
Cons	• Business Enterprise view unavailable • Redundant data costs • High ETL costs • High App costs • High DBA and operational costs	• Only viable for low volume • Meta-data issues • Network bandwidth and join complexity issues • Workload typically placed on workstation	• Business Enterprise view challenging • Redundant data costs • High DBA and operational costs • Data latency	• Requires corporate leadership and vision

FIGURE 8.6 Alternative Architectures for Data Warehousing Efforts *Source:* Based on W. Eckerson, "Four Ways to Build a Data Warehouse," *What Works: Best Practices in Business Intelligence and Data Warehousing,* Vol. 15, The Data Warehousing Institute, Chatsworth, CA, June 2003, pp. 46–49. Used with permission.

except that there are no dependent data marts. The central data warehouses architecture, which is advocated mainly by Teradata Corp., advises using data warehouses without any data marts (see Figure 8.7). This centralized approach provides users with access to all data in the data warehouse instead of limiting them to data marts. In addition, it reduces the amount of data the technical team has to transfer or change, therefore simplifying data management and administration.

The federated approach is a concession to the natural forces that undermine the best plans for developing a perfect system. It uses all possible means to integrate analytical resources from multiple sources to meet changing needs or business conditions. Essentially, the federated approach involves integrating disparate systems. In a federated architecture, existing decision support structures are left in place, and data are accessed from those sources as needed. The federated approach is supported by middleware vendors that propose distributed query and join capabilities. These eXtensible Markup Language (XML)–based tools offer users a global view of distributed data sources, including data warehouses, data marts, Web sites, documents, and operational systems. When users choose query objects from this view and press the submit button, the tool automatically queries the distributed sources, joins the results, and presents them to the user. Because of performance and data quality issues, most experts agree that federated approaches work well to supplement data warehouses, not replace them (see Eckerson, 2005).

Ariyachandra and Watson (2005) identified 10 factors that potentially affect the architecture selection decision:

1. Information interdependence between organizational units
2. Upper management's information needs

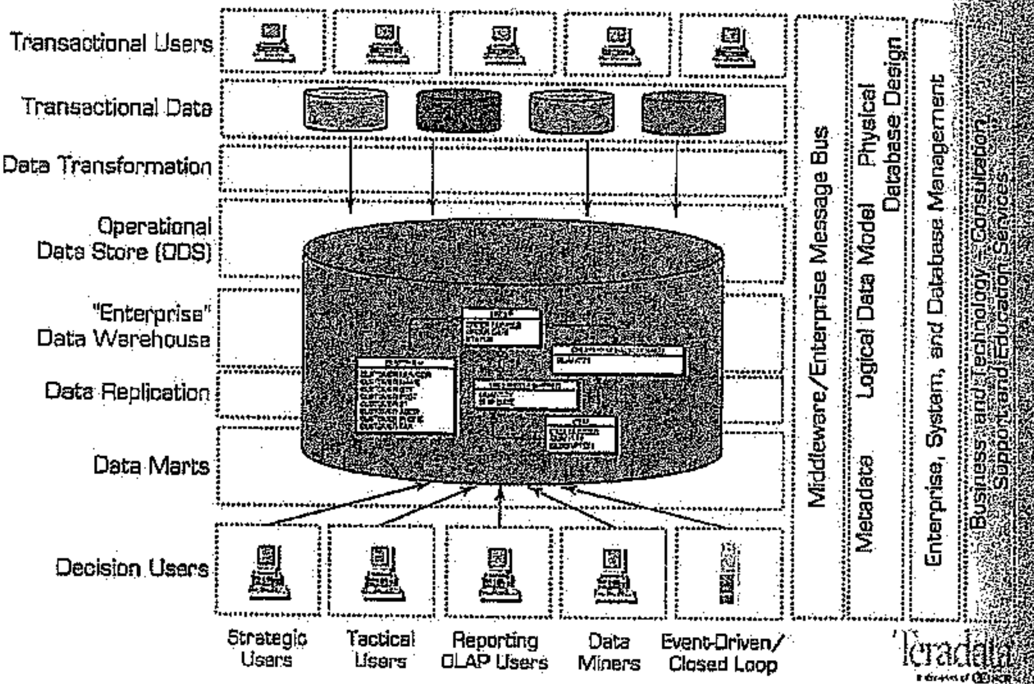


FIGURE 8.7 Teradata Corp.'s Enterprise Data Warehouse *Source:* Teradata Corporation (teradata.com). Used with permission.

3. Urgency of need for a data warehouse
4. Nature of end-user tasks
5. Constraints on resources
6. Strategic view of the data warehouse prior to implementation
7. Compatibility with existing systems
8. Perceived ability of the in-house IT staff
9. Technical issues
10. Social/political factors

These factors are similar to many success factors described in the literature for information systems projects and DSS and BI projects. Technical issues, beyond providing technology that is feasibly ready for use, is important, but often not as important as behavioral issues, such as meeting upper management's information needs and user involvement in the development process (a social/political factor). Each data warehousing architecture has specific applications for which it is most (and least) effective and thus provides maximal benefits to the organization. However, overall, the data mart structure seems to be the least effective in practice See Ariyachandra and Watson (2006a) for some additional details.

Which Architecture Is the Best?

Ever since data warehousing became a critical part of modern enterprises, the question of which data warehouse architecture is the best has been a topic of regular discussion. The two gurus of the data warehousing field, Bill Inmon and Ralph Kimball, are at the heart of this discussion. Inmon advocates the hub-and-spoke architecture (e.g., the Corporate Information Factory), whereas Kimball promotes the data mart bus architecture with conformed dimensions. Other architectures are possible, but these two options are fundamentally different approaches, and each has strong advocates. To shed light on this controversial question, Ariyachandra and Watson (2006b) conducted an empirical study. To collect the data, they used a Web-based survey targeted at individuals involved in data warehouse implementations. Their survey included questions about the respondent, the respondent's company, the company's data warehouse, and the success of the data warehouse architecture.

In total, 454 respondents provided usable information. Surveyed companies ranged from small (less than $10 million in revenue) to large (in excess of $10 billion). Most of the companies were located in the United States (60%) and represented a variety of industries, with the financial services industry (15%) providing the most responses. The predominant architecture was the hub-and-spoke architecture (39%), followed by the bus architecture (26%), the centralized architecture (17%), independent data marts (12%), and the federated architecture (4%). The most common platform for hosting the data warehouses was Oracle (41%), followed by Microsoft (19%) and IBM (18%). The average (mean) gross revenue varied from $3.7 billion for independent data marts to $6 billion for the federated architecture.

They used four measures to assess the success of the architectures: (1) information quality, (2) system quality, (3) individual impacts, and (4) organizational impacts. The questions used a seven-point scale, with the higher score indicating a more successful architecture. Table 8.1 shows the average scores for the measures across the architectures.

As the results of the study indicate, independent data marts scored the lowest on all measures. This finding confirms the conventional wisdom that independent data marts are a poor architectural solution. Next lowest on all measures was the federated architecture. Firms sometimes have disparate decision-support platforms resulting from mergers and acquisitions, and they may choose a federated approach, at least in the short run. The findings suggest that the federated architecture is not an optimal long-term solution. What is interesting, however, is the similarity of the averages for the bus, hub-and-spoke, and centralized architectures. The differences are sufficiently small that no claims can be made for a particular architecture's superiority over the others, at least based on a simple comparison of these success measures.

They also collected data on the domain (e.g., varying from a subunit to company-wide) and the size (i.e., amount of data stored) of the warehouses. They found that the hub-and-spoke architecture is typically used with more enterprise-wide implementations and larger warehouses. They also investigated the cost and time required to implement the different architectures. Overall, the hub-and-spoke architecture was the most expensive and time-consuming to implement.

Section 8.4 Review Questions

1. What are the key similarities and differences between a two-tiered architecture and a three-tiered architecture?
2. How has the Web influenced data warehouse design?
3. List the alternative data warehousing architectures discussed in this section.

TABLE 8.1 Average Assessment Scores for the Success of the Architectures

	Independent Data Marts	Bus Architecture	Hub-and-Spoke Architecture	Centralized Architecture (No Dependent Data Marts)	Federated Architecture
Information Quality	4.42	5.16	5.35	5.23	4.73
System Quality	4.59	5.60	5.56	5.41	4.69
Individual Impacts	5.08	5.80	5.62	5.64	5.15
Organizational Impacts	4.66	5.34	5.24	5.30	4.77

4. What issues should be considered when deciding which architecture to use in developing a data warehouse? List the 10 most important factors.
5. Which data warehousing architecture is the best? Why?

8.5 DATA INTEGRATION AND THE EXTRACTION, TRANSFORMATION, AND LOAD (ETL) PROCESSES

Global competitive pressures, demand for return on investment (ROI), management and investor inquiry, and government regulations are forcing business managers to rethink how they integrate and manage their businesses. A decision maker typically needs access to multiple sources of data that must be integrated. Before data warehouses, data marts, and BI software, providing access to data sources was a major, laborious process. Even with modern Web-based data management tools, recognizing what data to access and providing them to the decision maker are nontrivial tasks that require database specialists. As data warehouses grow in size, the issues of integrating data grow as well.

The business analysis needs continue to evolve. Mergers and acquisitions, regulatory requirements, and the introduction of new channels can drive changes in BI requirements. In addition to historical, cleansed, consolidated, and point-in-time data, business users increasingly demand access to real-time, unstructured, and/or remote data. And everything must be integrated with the contents of an existing data warehouse (see Devlin, 2003). Moreover, access via PDAs and through speech recognition and synthesis is becoming more commonplace, further complicating integration issues (see Edwards, 2003). Many integration projects involve enterprise-wide systems. Orovic (2003) provided a checklist of what works and what does not work when attempting such a project. Properly integrating data from various databases and other disparate sources is difficult. But when it is not done properly, it can lead to disaster in enterprise-wide systems such as CRM, ERP, and supply-chain projects (see Nash, 2002). Also see Dasu and Johnson (2003).

Data Integration

Data integration comprises three major processes that, when correctly implemented, permit data to be accessed and made accessible to an array of ETL and analysis tools and data warehousing environment: data access (i.e., the ability to access and extract data from any data source), data federation (i.e., the integration of business views across multiple data stores), and change capture (based on the identification, capture, and delivery of the changes made to enterprise data sources). See Sapir (2005) for details. See Application Case 8.3 for an example of how BP Lubricant benefits from implementing a data warehouse that integrates data from many sources. Some vendors, such as SAS Institute, Inc., have developed strong data integration tools. The SAS enterprise data integration server includes customer data integration tools that improve data quality in the integration process. The Oracle Business Intelligence Suite assists in integrating data as well.

APPLICATION CASE 8.3

BP Lubricants Achieves BIGS Success

BP Lubricants established the BIGS program following recent merger activity to deliver globally consistent and transparent management information. As well as timely business intelligence, BIGS provides detailed, consistent views of performance across functions such as finance, marketing, sales, and supply and logistics.

BP is one of the world's largest oil and petrochemicals groups. Part of the BP plc group, BP Lubricants is an established leader in the global automotive lubricants market. Perhaps best known for its Castrol brand of oils, the business operates in over 100 countries and employs 10,000 people. Strategically, BP Lubricants is concentrating on further improving its customer focus and increasing its effectiveness in automotive markets. Following recent merger activity, the company is undergoing transformation to become more effective and agile and to seize opportunities for rapid growth.

Challenge

Following recent merger activity, BP Lubricants wanted to improve the consistency, transparency, and accessibility of management information and business intelligence. In order to do so, it needed to integrate data held in disparate source systems, without the delay of introducing a standardized ERP system.

Solution

BP Lubricants implemented the pilot for its Business Intelligence and Global Standards (BIGS) program, a strategic initiative for management information and business intelligence. At the heart of BIGS is Kalido, an adaptive enterprise data warehousing solution for preparing, implementing, operating, and managing data warehouses.

Kalido's federated enterprise data warehousing solution supported the pilot program's complex data integration and diverse reporting requirements. To adapt to the program's evolving reporting requirements, the software also enabled the underlying information architecture to be easily modified at high speed while preserving all information. The system integrates and stores information from multiple source systems to provide consolidated views for:

- ***Marketing.*** Customer proceeds and margins for market segments with drill down to invoice-level detail
- ***Sales.*** Sales invoice reporting augmented with both detailed tariff costs and actual payments
- ***Finance.*** Globally standard profit and loss, balance sheet, and cash flow statements—with audit ability; customer debt management supply and logistics; consolidated view of order and movement processing across multiple ERP platforms

Benefits

By improving the visibility of consistent, timely data, BIGS provides the information needed to assist the business in identifying a multitude of business opportunities to maximize margins and/or manage associated costs. Typical responses to the benefits of consistent data resulting from the BIGS pilot include:

- Improved consistency and transparency of business data
- Easier, faster, and more flexible reporting
- Accommodation of both global and local standards
- Fast, cost-effective, and flexible implementation cycle
- Minimal disruption of existing business processes and the day-to-day business
- Identifies data quality issues and encourages their resolution
- Improved ability to respond intelligently to new business opportunities

Sources: Kalido, "BP Lubricants Achieves BIGS, Key IT Solutions," **keyitsolutions.com/asp/rptdetails/report/95/cat/1175/** (accessed August 2009); Kalido, "BP Lubricants Achieves BIGS Success," **kalido.com/collateral/Documents/English-US/CS-BP%20BIGS.pdf** (accessed August 2009); and BP Lubricant homepage, **bp.com/lubricanthome.do** (accessed August 2009).

A major purpose of a data warehouse is to integrate data from multiple systems. Various integration technologies enable data and metadata integration:

- Enterprise application integration (EAI)
- Service-oriented architecture (SOA)
- Enterprise information integration (EII)
- Extraction, transformation, and load (ETL)

Enterprise application integration (EAI) provides a vehicle for pushing data from source systems into the data warehouse. It involves integrating application functionality and is focused on sharing functionality (rather than data) across systems, thereby enabling flexibility and reuse. Traditionally, EAI solutions have focused on enabling application reuse at the application programming interface (API) level. Recently, EAI is accomplished by using SOA coarse-grained services (a collection of business processes or functions) that are well defined and documented. Using Web services is a specialized way of implementing an SOA. EAI can be used to facilitate data acquisition directly into a near-real-time data warehouse or to deliver decisions to the OLTP systems. There are many different approaches to and tools for EAI implementation.

Enterprise information integration (EII) is an evolving tool space that promises real-time data integration from a variety of sources, such as relational databases, Web services, and multidimensional databases. It is a mechanism for pulling data from source systems to satisfy a request for information. EII tools use predefined metadata to populate views that make integrated data appear relational to end users. XML may be the most important aspect of EII because XML allows data to be tagged either at creation time or later. These tags can be extended and modified to accommodate almost any area of knowledge (see Kay, 2005).

Physical data integration has conventionally been the main mechanism for creating an integrated view with data warehouses and data marts. With the advent of EII tools (see Kay, 2005), new virtual data integration patterns are feasible. Manglik and Mehra (2005) discussed the benefits and constraints of new data integration patterns that can expand traditional physical methodologies to present a comprehensive view for the enterprise.

We next turn to the approach for loading data into the warehouse: ETL.

Extraction, Transformation, and Load

At the heart of the technical side of the data warehousing process is **extraction, transformation, and load (ETL)**. ETL technologies, which have existed for some time, are instrumental in the process and use of data warehouses. The ETL process is an integral component in any data-centric project. IT managers are often faced with challenges because the ETL process typically consumes 70 percent of the time in a data-centric project.

The ETL process consists of extraction (i.e., reading data from one or more databases), transformation (i.e., converting the extracted data from its previous form into the form in which it needs to be so that it can be placed into a data warehouse or simply another database), and load (i.e., putting the data into the data warehouse). Transformation occurs by using rules or lookup tables or by combining the data with other data. The three database functions are integrated into one tool to pull data out of one or more databases and place them into another, consolidated database or a data warehouse.

ETL tools also transport data between sources and targets, document how data elements (e.g., metadata) change as they move between source and target, exchange metadata with other applications as needed, and administer all runtime processes and operations (e.g., scheduling, error management, audit logs, statistics). ETL is extremely important for data integration as well as for data warehousing. The purpose of the ETL process is to load the warehouse with integrated and cleansed data. The data used in ETL processes can come from any source: a mainframe application, an ERP application, a CRM tool, a flat file, an Excel spreadsheet, or even a message queue. In Figure 8.8, we outline the ETL process.

The process of migrating data to a data warehouse involves the extraction of data from all relevant sources. Data sources may consist of files extracted from OLTP databases, spreadsheets, personal databases (e.g., Microsoft Access), or external files. Typically, all the input files are written to a set of staging tables, which are designed to facilitate the load process. A data warehouse contains numerous business rules that define such things as

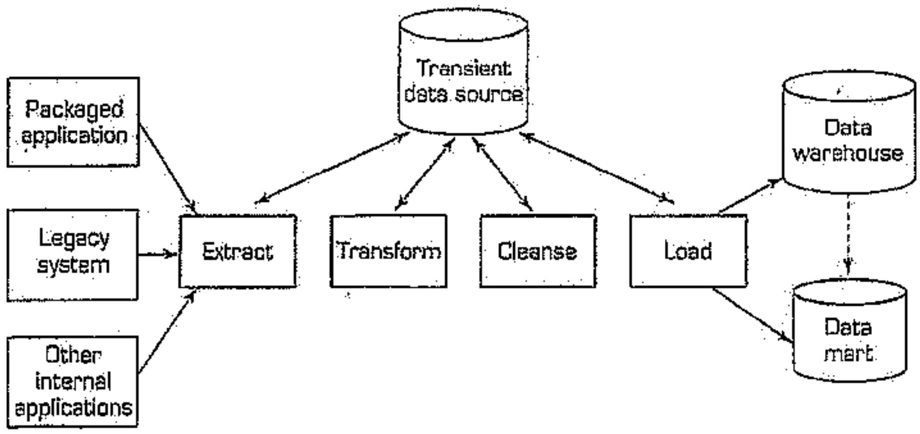


FIGURE 8.8 The ETL Process

how the data will be used, summarization rules, standardization of encoded attributes, and calculation rules. Any data quality issues pertaining to the source files need to be corrected before the data are loaded into the data warehouse. One of the benefits of a well-designed data warehouse is that these rules can be stored in a metadata repository and applied to the data warehouse centrally. This differs from an OLTP approach, which typically has data and business rules scattered throughout the system. The process of loading data into a data warehouse can be performed either through data transformation tools that provide a GUI to aid in the development and maintenance of business rules or through more traditional methods, such as developing programs or utilities to load the data warehouse, using programming languages such as PL/SQL, C++, or .NET Framework languages. This decision is not easy for organizations. Several issues affect whether an organization will purchase data transformation tools or build the transformation process itself:

- Data transformation tools are expensive.
- Data transformation tools may have a long learning curve.
- It is difficult to measure how the IT organization is doing until it has learned to use the data transformation tools.

In the long run, a transformation-tool approach should simplify the maintenance of an organization's data warehouse. Transformation tools can also be effective in detecting and scrubbing (i.e., removing any anomalies in the data). OLAP and data mining tools rely on how well the data are transformed.

As an example of effective ETL, Motorola, Inc., uses ETL to feed its data warehouses. Motorola collects information from 30 different procurement systems and sends them to its global SCM data warehouse for analysis of aggregate company spending (see Songini, 2004).

Solomon (2005) classified ETL technologies into four categories: sophisticated, enabler, simple, and rudimentary. It is generally acknowledged that tools in the sophisticated category will result in the ETL process being better documented and more accurately managed as the data warehouse project evolves.

Even though it is possible for programmers to develop software for ETL, it is simpler to use an existing ETL tool. The following are some of the important criteria in selecting an ETL tool (see Brown, 2004):

- Ability to read from and write to an unlimited number of data source architectures
- Automatic capturing and delivery of metadata
- A history of conforming to open standards
- An easy-to-use interface for the developer and the functional user

Performing extensive ETL may be a sign of poorly managed data and a fundamental lack of a coherent data management strategy. Karacsony (2006) indicated that there is a direct correlation between the extent of redundant data and the number of ETL processes. When data are managed correctly as an enterprise asset, ETL efforts are significantly reduced, and redundant data are completely eliminated. This leads to huge savings in maintenance and greater efficiency in new development while also improving data quality. Poorly designed ETL processes are costly to maintain, change, and update. Consequently, it is crucial to make the proper choices in terms of the technology and tools to use for developing and maintaining the ETL process.

A number of packaged ETL tools are available. Database vendors currently offer ETL capabilities that both enhance and compete with independent ETL tools. SAS acknowledges the importance of data quality and offers the industry's first fully integrated solution that merges ETL and data quality to transform data into strategic valuable assets. Other ETL software providers include Microsoft, Oracle, IBM, Informatica, Embarcadero, and Tibco. For additional information on ETL, see Golfarelli and Rizzi (2009), Karakson (2006), and Songini (2004).

Section 8.5 Review Questions

1. Describe data integration.
2. Describe the three steps of the ETL process.
3. Why is the ETL process so important for data warehousing efforts?

8.6 DATA WAREHOUSE DEVELOPMENT

A data warehousing project is a major undertaking for any organization and is more complicated than a simple, mainframe selection and implementation project because it comprises and influences many departments and many input and output interfaces and it can be part of a CRM business strategy. A data warehouse provides several benefits that can be classified as direct and indirect. Direct benefits include the following:

- End users can perform extensive analysis in numerous ways.
- A consolidated view of corporate data (i.e., a single version of the truth) is possible.
- Better and more-timely information is possible. A data warehouse permits information processing to be relieved from costly operational systems onto low-cost servers; therefore, many more end-user information requests can be processed more quickly.
- Enhanced system performance can result. A data warehouse frees production processing because some operational system reporting requirements are moved to DSS.
- Data access is simplified.

Indirect benefits result from end users using these direct benefits. On the whole, these benefits enhance business knowledge, present competitive advantage, improve customer service and satisfaction, facilitate decision making, and help in reforming business processes, and therefore they are the strongest contributions to competitive advantage. (For a discussion of how to create a competitive advantage through data warehousing, see Parzinger and Frolick, 2001.) For a detailed discussion of how organizations can obtain exceptional levels of payoffs, see Watson et al. (2002). Given the potential benefits that a data warehouse can provide and the substantial investments in time and money that such a project requires, it is critical that an organization structure its data warehouse project to maximize the chances of success. In addition, the organization must, obviously, take costs into consideration. Kelly (2001) described a ROI approach that considers benefits in the categories of keepers (i.e., money saved by improving traditional decision support functions); gatherers (i.e., money saved due to automated collection and dissemination of information); and users (i.e., money saved or gained from decisions made using the data warehouse). Costs include those related to hardware, software, network bandwidth, internal development, internal support, training, and external consulting. The net present value (NPV) is calculated over the expected life of the data warehouse. Because the benefits are broken down approximately as 20 percent for keepers, 30 percent for gatherers, and 50 percent for users, Kelly indicated that users should be involved in the development process, a success factor typically mentioned as critical for systems that imply change in an organization.

Application Case 8.4 provides an example of a data warehouse that was developed and delivered intense competitive advantage for the Hokuriku (Japan) Coca-Cola Bottling Company. The system was so successful that plans are under way to expand it to encompass the more than 1 million Coca-Cola vending machines in Japan.

APPLICATION CASE 8.4

Things Go Better with Coke's Data Warehouse

In the face of competitive pressures and consumer demand, how does a successful bottling company ensure that its vending machines are profitable? The answer for Hokuriku Coca-Cola Bottling Company (HCCBC) is a data warehouse and analytical software implemented by Teradata Corp. HCCBC built the system in response to a data warehousing system developed by its rival, Mikuni. The data warehouse collects not only historical data but also near-real-time data from each vending machine (viewed as a store) that could be transmitted via wireless connection to headquarters. The initial phase of the project was deployed in 2001. The data warehouse approach provides detailed product information, such as time and date of each sale, when a product sells out, whether someone was short-changed, and whether the machine is malfunctioning. In each case, an alert is triggered, and the vending machine immediately reports it to the data center over a wireless transmission system. (Note that Coca-Cola in the United States has used modems to link vending machines to distributors for over a decade.)

In 2002, HCCBC conducted a pilot test and put all its Nagano vending machines on a wireless network to gather near-real-time point of sale (POS) data from each one. The results were astounding because they accurately forecasted demand and identified problems quickly. Total sales immediately increased 10 percent. In addition, due to the more accurate machine servicing, overtime and other costs decreased 46 percent. In addition, each salesperson was able to service up to 42 percent more vending machines.

The test was so successful that planning began to expand it to encompass the entire enterprise (60,000 machines), using an active data warehouse. Eventually, the data warehousing solution will ideally expand across corporate boundaries into the entire Coca-Cola Bottlers network so that the more than 1 million vending machines in Japan will be networked, leading to immense cost savings and higher revenue.

Sources: Adapted from K. D. Schwartz, "Decisions at the Touch of a Button," *Teradata Magazine*, **teradata.com/t/page/117774/index.html** (accessed June 2009); K. D. Schwartz, "Decisions at the Touch of a Button," *DSS Resources*, March 2004, pp. 28–31, **dssresources.com/cases/coca-colajapan/index.html** (accessed April 2006); and Teradata Corp., "Coca-Cola Japan Puts the Fizz Back in Vending Machine Sales," **teradata.com/t/page/118866/index.html** (accessed June 2009).

Clearly defining the business objective, gathering project support from management end users, setting reasonable time frames and budgets, and managing expectations are critical to a successful data warehousing project. A data warehousing strategy is a blueprint for the successful introduction of the data warehouse. The strategy should describe where the company wants to go, why it wants to go there, and what it will do when it gets there. It needs to take into consideration the organization's vision, structure, and culture. See Matney (2003) for the steps that can help in developing a flexible and efficient support strategy. When the plan and support for a data warehouse are established, the organization needs to examine data warehouse vendors. (See Table 8.2 for a sample list of vendors; also see The Data Warehousing Institute [**twdi.com**] and *DM Review* [**dmreview.com**].) Many vendors provide software demos of their data warehousing and BI products.

Data Warehouse Vendors

McCloskey (2002) cited six guidelines that need to be considered when developing a vendor list: financial strength, ERP linkages, qualified consultants, market share, industry experience, and established partnerships. Data can be obtained from trade shows and corporate Web sites, as well as by submitting requests for specific product information.

TABLE 8.2 Sample List of Data Warehousing Vendors

Vendor	Product Offerings
Computer Associates (**cai.com**)	Comprehensive set of data warehouse (DW) tools and products
DataMirror (**datamirror.com**)	DW administration, management, and performance products
Data Advantage Group (**dataadvantagegroup.com**)	Metadata software
Dell (**dell.com**)	DW servers
Embarcadero Technologies (**embarcadero.com**)	DW administration, management, and performance products
Business Objects (**businessobjects.com**)	Data cleansing software
Harte-Hanks (**harte-hanks.com**)	Customer relationship management (CRM) products and services
HP (**hp.com**)	DW servers
Hummingbird Ltd. (**hummingbird.com**)	DW engines and exploration warehouses
Hyperion Solutions (**hyperion.com**)	Comprehensive set of DW tools, products, and applications
IBM (**ibm.com**)	DW tools, products, and applications
Informatica (**informatica.com**)	DW administration, management, and performance products
Microsoft (**microsoft.com**)	DW tools and products
Oracle (including People Soft and Siebel) (**oracle.com**)	DW, ERP, and CRM tools, products, and applications
SAS Institute (**sas.com**)	DW tools, products, and applications
Siemens (**siemens.com**)	DW servers
Sybase (**sybase.com**)	Comprehensive set of DW tools and applications
Teradata (**teradata.com**)	DW tools, products, and applications

Van den Hoven (1998) differentiated three types of data warehousing products. The first type handles functions such as locating, extracting, transforming, cleansing, transporting, and loading the data into the data warehouse. The second type is a data management tool—a database engine that stores and manages the data warehouse as well as the metadata. The third type is a data access tool that provides end users with access to analyze the data in the data warehouse. This may include query generators, visualization, EIS, OLAP, and data mining capabilities.

Data Warehouse Development Approaches

Many organizations need to create the data warehouses used for decision support. Two competing approaches are employed. The first approach is that of Bill Inmon, who is often called "the father of data warehousing." Inmon supports a top-down development approach that adapts traditional relational database tools to the development needs of an enterprise-wide data warehouse, also known as the EDW approach. The second approach is that of Ralph Kimball, who proposes a bottom-up approach that employs dimensional modeling, also known as the data mart approach.

Knowing how these two models are alike and how they differ helps us understand the basic data warehouse concepts (e.g., see Breslin, 2004). Table 8.3 compares the two approaches. We describe these approaches in detail next.

TABLE 8.3 Contrasts Between the Data Mart and EDW Development Approaches

Effort	Data Mart Approach	EDW Approach
Scope	One subject area	Several subject areas
Development time	Months	Years
Development cost	$10,000 to $100,000+	$1,000,000+
Development difficulty	Low to medium	High
Data prerequisite for sharing	Common (within business area)	Common (across enterprise)
Sources	Only some operational and external systems	Many operational and external systems
Size	Megabytes to several gigabytes	Gigabytes to petabytes
Time horizon	Near-current and historical data	Historical data
Data transformations	Low to medium	High
Update frequency	Hourly, daily, weekly	Weekly, monthly
Technology		
Hardware	Workstations and departmental servers	Enterprise servers and mainframe computers
Operating system	Windows and Linux	Unix, Z/OS, OS/390
Databases	Workgroup or standard database servers	Enterprise database servers
Usage		
Number of simultaneous users	10s	100s to 1,000s
User types	Business area analysts and managers	Enterprise analysts and senior executives
Business spotlight	Optimizing activities within the business area	Cross-functional optimization and decision making

Sources: Adapted from J. Van den Hoven, "Data Marts: Plan Big, Build Small," in *IS Management Handbook*, 8th ed., CRC Press, Boca Raton, FL, 2003; and T. Ariyachandra and H. Watson, "Which Data Warehouse Architecture Is Most Successful?" *Business Intelligence Journal*, Vol. 11, No. 1, First Quarter 2006, pp. 4–6.

THE INMON MODEL: THE EDW APPROACH Inmon's approach emphasizes top-down development, employing established database development methodologies and tools, such as entity-relationship diagrams (ERD), and an adjustment of the spiral development approach. The EDW approach does not preclude the creation of data marts. The EDW is the ideal in this approach because it provides a consistent and comprehensive view of the enterprise. Murtaza (1998) presented a framework for developing EDW.

THE KIMBALL MODEL: THE DATA MART APPROACH Kimball's data mart strategy is a "plan big, build small" approach. A data mart is a subject-oriented or department-oriented data warehouse. It is a scaled-down version of a data warehouse that focuses on the requests of a specific department, such as marketing or sales. This model applies dimensional data modeling, which starts with tables. Kimball advocated a development methodology that entails a bottom-up approach, which in the case of data warehouses means building one data mart at a time.

WHICH MODEL IS BEST? There is no one-size-fits-all strategy to data warehousing. An enterprise's data warehousing strategy can evolve from a simple data mart to a complex data warehouse in response to user demands, the enterprise's business requirements, and the enterprise's maturity in managing its data resources. For many enterprises, a data mart is frequently a convenient first step to acquiring experience in constructing and managing a data warehouse while presenting business users with the benefits of better access to their data; in addition, a data mart commonly indicates the business value of data warehousing. Ultimately, obtaining an EDW is ideal (see Application Case 8.5). However, the development of individual data marts can often provide many benefits along the way toward developing an EDW, especially if the organization is unable or unwilling to invest in a large-scale project. Data marts can also demonstrate feasibility and success in providing benefits. This could potentially lead to an investment in an EDW. Table 8.4 summarizes the most essential characteristic differences between the two models.

APPLICATION CASE 8.5

HP Consolidates Hundreds of Data Marts into a Single EDW

In December 2005, HP planned to consolidate its 762 data marts around the world into a single EDW. HP took this approach to gain a superior sense of its own business and to determine how best to serve its customers. Mark Hurd, HP's president and chief executive, stated that "there was a thirst for analytic data" inside the company that had unfortunately led to the creation of many data marts. Those data silos were very expensive to design and maintain, and they did not produce the enterprise-wide view of internal and customer information that HP wanted. In mid-2006, HP started to consolidate the data in the data marts into the new data warehouse. All the disparate data marts will ultimately be eliminated.

Sources: Adapted from C. Martins, "HP to Consolidate Data Marts into Single Warehouse," *Computerworld*, December 13, 2005; C. Martins, "HP to Consolidate Data Marts Into Single Warehouse," *InfoWorld*, December 13, 2005; and C. Martins, "HP to Consolidate Data Marts into One Warehouse," December 14, 2005.

TABLE 8.4 Essential Differences Between Inmon's and Kimball's Approaches

Characteristic	Inmon	Kimball
Methodology and Architecture		
Overall approach	Top-down	Bottom-up
Architecture structure	Enterprise-wide (atomic) data warehouse "feeds" departmental databases	Data marts model a single business process, and enterprise consistency is achieved through a data bus and conformed dimensions
Complexity of the method	Quite complex	Fairly simple
Comparison with established development methodologies	Derived from the spiral methodology	Four-step process; a departure from relational database management system (RDBMS) methods
Discussion of physical design	Fairly thorough	Fairly light
Data Modeling		
Data orientation	Subject or data driven	Process oriented
Tools	Traditional (entity-relationship diagrams [ERD], data flow diagrams [DFD])	Dimensional modeling; a departure from relational modeling
End-user accessibility	Low	High
Philosophy		
Primary audience	IT professionals	End users
Place in the organization	Integral part of the corporate information factory	Transformer and retainer of operational data
Objective	Deliver a sound technical solution based on proven database methods and technologies	Deliver a solution that makes it easy for end users to directly query the data and still get reasonable response times

Sources: Adapted from M. Breslin, "Data Warehousing Battle of the Giants: Comparing the Basics of Kimball and Inmon Models," *Business Intelligence Journal*, Vol. 9, No. 1, Winter 2004, pp. 6–20; and T. Ariyachandra and H. Watson, "Which Data Warehouse Architecture Is Most Successful?" *Business Intelligence Journal*, Vol. 11, No. 1, First Quarter 2006.

Additional Data Warehouse Development Considerations

Some organizations want to completely outsource their data warehousing efforts. They simply do not want to deal with software and hardware acquisitions, and they do not want to manage their information systems. One alternative is to use hosted data warehouses. In this scenario, another firm—ideally, one that has a lot of experience and expertise—develops and maintains the data warehouse. However, there are security and privacy concerns with this approach. See Technology Insights 8.1 for some details.

Data Warehouse Structure: The Star Schema

A typical data warehouse structure is shown in Figure 8.1. Many variations on data warehouse architecture are possible; the most important one is the star schema. The data warehouse design is based on the concept of dimensional modeling. **Dimensional modeling** is a retrieval-based system that supports high-volume query access. The star schema is the means by which dimensional modeling is implemented. A star schema contains a central fact table surrounded by several **dimension tables** (Adamson, 2009). The fact table contains a large number of rows that correspond to observed facts. A fact table contains the attributes needed to perform decision analysis, descriptive attributes used for query reporting, and foreign keys to link to dimension tables. The decision analysis attributes consist of performance measures, operational metrics, aggregated measures, and all

TECHNOLOGY INSIGHTS 8.1 Hosted Data Warehouses

A hosted data warehouse has nearly the same, if not more, functionality as an onsite data warehouse, but it does not consume computer resources on client premises. A hosted data warehouse offers the benefits of BI minus the cost of computer upgrades, network upgrades, software licenses, in-house development, and in-house support and maintenance.

A hosted data warehouse offers the following benefits:

- Requires minimal investment in infrastructure
- Frees up capacity on in-house systems
- Frees up cash flow
- Makes powerful solutions affordable
- Enables powerful solutions that provide for growth
- Offers better quality equipment and software
- Provides faster connections
- Enables users to access data from remote locations
- Allows a company to focus on core business
- Meets storage needs for large volumes of data

Despite its benefits, a hosted data warehouse is not necessarily a good fit for every organization. Large companies with revenue upward of $500 million could lose money if they already have underused internal infrastructure and IT staff. Furthermore, companies that see the paradigm shift of outsourcing applications as loss of control of their data are not prone to use a business intelligence service provider (BISP). Finally, the most significant and common argument against implementing a hosted data warehouse is that it may be unwise to outsource sensitive applications for reasons of security and privacy.

Sources: Partly adapted from M. Thornton and M. Lampa, "Hosted Data Warehouse," *Journal of Data Warehousing*, Vol. 7, No. 2, 2002, pp. 27–34; and M. Thornton, "What About Security? The Most Common, but Unwarranted, Objection to Hosted Data Warehouses," *DM Review*, Vol. 12, No. 3, March 18, 2002, pp. 30–43.

the other metrics needed to analyze the organization's performance. In other words, the fact table primarily addresses what the data warehouse supports for decision analysis.

Surrounding the central fact tables (and linked via foreign keys) are dimension tables. The dimension tables contain classification and aggregation information about the central fact rows. Dimension tables contain attributes that describe the data contained within the fact table; they address how data will be analyzed. Dimension tables have a one-to-many relationship with rows in the central fact table. Some examples of dimensions that would support a product fact table are location, time, and size. The star schema design provides extremely fast query-response time, simplicity, and ease of maintenance for read-only database structures. According to Raden (2003), setting up a star schema for real-time updating could be a straightforward approach, as long as a few rules are followed. We show an example star schema in Figure 8.9.

The **grain** (also known as granularity) of a data warehouse defines the highest level of detail that is supported. The grain indicates whether the data warehouse is highly summarized or also includes detailed transaction data. If the grain is defined too high, then the warehouse may not support detail requests to **drill down** into the data. Drill-down analysis is the process of probing beyond a summarized value to investigate each of the detail transactions that comprise the summary. A low level of granularity will result in more data being stored in the warehouse. Larger amounts of detail may affect the performance of queries by making the response times longer. Therefore, during the scoping of a data warehouse project, it is important to identify the level of granularity that will be needed. See Tennant (2002) for a discussion of granularity issues in metadata.

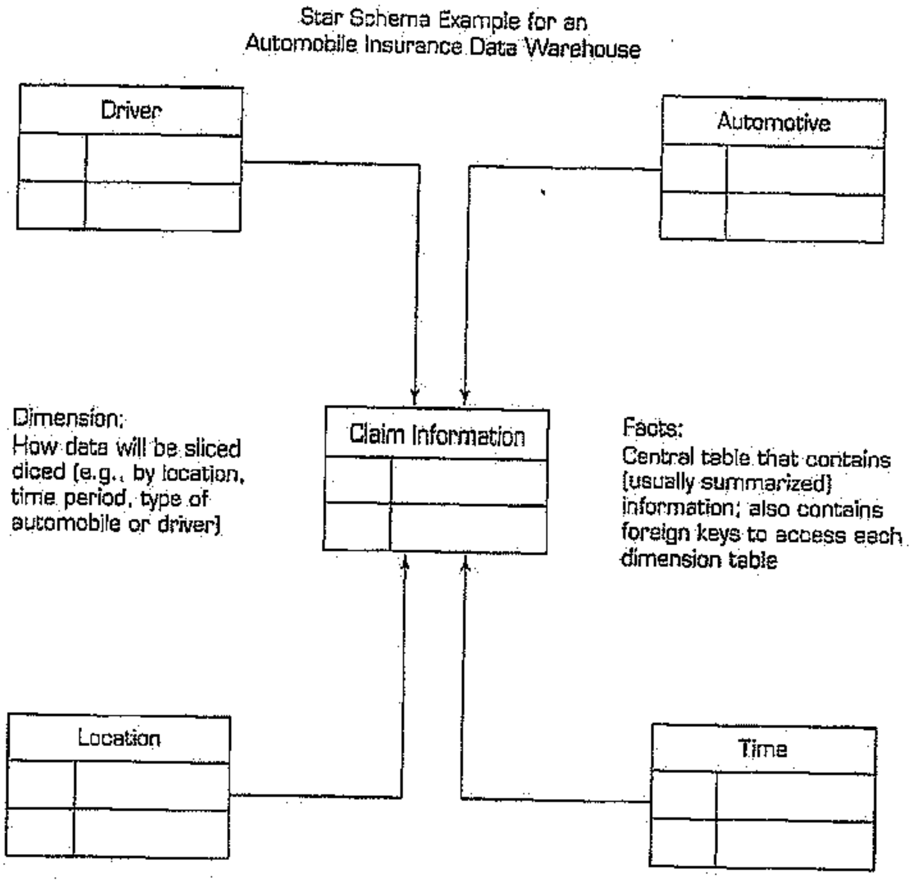


FIGURE 8.9 Star Schema

Data Warehousing Implementation Issues

Implementing a data warehouse is generally a massive effort that must be planned and executed according to established methods. However, the project lifecycle has many facets, and no single person can be an expert in each area. Here we discuss specific ideas and issues as they relate to data warehousing. Inmon (2006) provided a set of actions that a data warehouse systems programmer may use to tune a data warehouse.

Reeves (2009) and Solomon (2005) provided some guidelines regarding the critical questions that must be asked, some risks that should be weighed and some processes that can be followed to help ensure a successful data warehouse implementation. He compiled a list of 11 major tasks that could be performed in parallel:

1. Establishment of service-level agreements and data-refresh requirements
2. Identification of data sources and their governance policies
3. Data quality planning
4. Data model design
5. ETL tool selection
6. Relational database software and platform selection
7. Data transport
8. Data conversion
9. Reconciliation process
10. Purge and archive planning
11. End-user support

Following these guidelines should increase an organization's chances for success. Given the size and scope of an enterprise-level data warehouse initiative, failure to anticipate these issues greatly increases the risks of failure.

Hwang and Xu (2005) conducted a major survey of data warehousing success issues. The results established that data warehousing success is a multifaceted construct, and Hwang and Xu proposed that a data warehouse be constructed while keeping in mind the goal of improving user productivity. Extremely significant benefits of doing so include prompt information retrieval and enhanced quality information. The survey results also indicated that success hinges on factors of different dimensions.

People want to know how successful their BI and data warehousing initiatives are in comparison to those of other companies. Ariyachandra and Watson (2006a) proposed some benchmarks for BI and data warehousing success. Watson et al. (1999) researched data warehouse failures. Their results showed that people define a "failure" in different ways, and this was confirmed by Ariyachandra and Watson (2006a). The Data Warehousing Institute (**tdwi.org**) has developed a data warehousing maturity model that an enterprise can apply in order to benchmark its evolution. The model offers a fast means to gauge where the organization's data warehousing initiative is now and where it needs to go next. The maturity model consists of six stages: prenatal, infant, child, teenager, adult, and sage. Business value rises as the data warehouse progresses through each succeeding stage. The stages are identified by a number of characteristics, including scope, analytic structure, executive perceptions, types of analytics, stewardship, funding, technology platform, change management, and administration. See Eckerson et al. (2009) and Eckerson (2003) for more details.

Saunders (2009) provided an easy-to-understand cooking analogy to developing data warehouses. Weir (2002) specifically described some of the best practices for implementing a data warehouse, which include the following guidelines:

- The project must fit with corporate strategy and business objectives.
- There must be complete buy-in to the project by executives, managers, and users.
- It is important to manage user expectations about the completed project.
- The data warehouse must be built incrementally.
- Adaptability must be built in.
- The project must be managed by both IT and business professionals.
- A business–supplier relationship must be developed.
- Only load data that have been cleansed and are of a quality understood by the organization.
- Do not overlook training requirements.
- Be politically aware.

Data warehouse projects have many risks. Most of them are also found in other IT projects, but data warehousing risks are more serious because data warehouses are expensive, large-scale projects. Each risk should be assessed at the inception of the project. Adelman and Moss (2001) described some of these risks, including the following:

- No mission or objective
- Quality of source data unknown
- Skills not in place
- Inadequate budget
- Lack of supporting software
- Source data not understood
- Weak sponsor
- Users not computer literate
- Political problems or turf wars

- Unrealistic user expectations
- Architectural and design risks
- Scope creep and changing requirements
- Vendors out of control
- Multiple platforms
- Key people leaving the project
- Loss of the sponsor
- Too much new technology
- Having to fix an operational system
- Geographically distributed environment
- Team geography and language culture

Practitioners have unearthed a wealth of mistakes that have been made in the development of data warehouses. Watson et al. (1999) discussed how such mistakes could lead to data warehouse failures (also see Barquin et al., 1997). Turban et al. (2006) listed the following reasons for failure: cultural issues being ignored, inappropriate architecture, unclear business objectives, missing information, unrealistic expectations, low levels of data summarization, and low data quality.

When developing a successful data warehouse, it is important to carefully consider the following issues:

- ***Starting with the wrong sponsorship chain.*** You need an executive sponsor who has influence over the necessary resources to support and invest in the data warehouse. You also need an executive project driver, someone who has earned the respect of other executives, has a healthy skepticism about technology, and is decisive but flexible. You also need an IS/IT manager to head up the project.
- ***Setting expectations that you cannot meet.*** You do not want to frustrate executives at the moment of truth. Every data warehousing project has two phases: Phase 1 is the selling phase, in which you internally market the project by selling the benefits to those who have access to needed resources. Phase 2 is the struggle to meet the expectations described in Phase 1. For a mere $1 million to $7 million, hopefully, you can deliver.
- ***Engaging in politically naive behavior.*** Do not simply state that a data warehouse will help managers make better decisions. This may imply that you feel they have been making bad decisions until now. Sell the idea that they will be able to get the information they need to help in decision making.
- ***Loading the warehouse with information just because it is available.*** Do not let the data warehouse become a data landfill. This would unnecessarily slow down the use of the system. There is a trend toward real-time computing and analysis. Data warehouses must be shut down to load data in a timely way.
- ***Believing that data warehousing database design is the same as transactional database design.*** In general, it is not. The goal of data warehousing is to access aggregates rather than a single or a few records, as in transaction-processing systems. Content is also different, as is evident in how data are organized. DBMS tend to be nonredundant, normalized, and relational, whereas data warehouses are redundant, not normalized, and multidimensional.
- ***Choosing a data warehouse manager who is technology oriented rather than user oriented.*** One key to data warehouse success is to understand that the users must get what they need, not advanced technology for technology's sake.
- ***Focusing on traditional internal record-oriented data and ignoring the value of external data and of text, images, and, perhaps, sound and***

video. Data come in many formats and must be made accessible to the right people at the right time and in the right format. They must be cataloged properly.
- ***Delivering data with overlapping and confusing definitions.*** Data cleansing is a critical aspect of data warehousing. It includes reconciling conflicting data definitions and formats organization-wide. Politically, this may be difficult because it involves change, typically at the executive level.
- ***Believing promises of performance, capacity, and scalability.*** Data warehouses generally require more capacity and speed than is originally budgeted for. Plan ahead to scale up.
- ***Believing that your problems are over when the data warehouse is up and running.*** DSS/BI projects tend to evolve continually. Each deployment is an iteration of the prototyping process. There will always be a need to add more and different data sets to the data warehouse, as well as additional analytic tools for existing and additional groups of decision makers. High energy and annual budgets must be planned for because success breeds success. Data warehousing is a continuous process.
- ***Focusing on ad hoc data mining and periodic reporting instead of alerts.*** The natural progression of information in a data warehouse is: (1) Extract the data from legacy systems, cleanse them, and feed them to the warehouse; (2) support ad hoc reporting until you learn what people want; and (3) convert the ad hoc reports into regularly scheduled reports.

  This process of learning what people want in order to provide it seems natural, but it is not optimal or even practical. Managers are busy and need time to read reports. Alert systems are better than periodic reporting systems and can make a data warehouse mission critical. Alert systems monitor the data flowing into the warehouse and inform all key people who have a need to know as soon as a critical event occurs.

Sammon and Finnegan (2000) revealed the outcome of a study of four mature users of data warehousing technology. Their practices were captured in an outline of 10 organizational requisites for applying data warehousing. Organizations can use this representation to internally evaluate the chances of the success of a data warehousing project and to recognize the parts that need attention prior to beginning implementation. A summary of their prerequisites model is as follows:

- A business-driven data warehousing initiative
- Executive sponsorship and commitment
- Funding commitment based on realistically managed expectations
- A project team
- Attention to source data quality
- A flexible enterprise data model
- Data stewardship
- A long-term plan for automated data extraction methods/tools
- Knowledge of data warehouse compatibility with existing systems
- Hardware/software proof of concept

Wixom and Watson (2001) defined a research model for data warehouse success that identified seven important implementation factors that can be categorized into three criteria (i.e., organizational issues, project issues, and technical issues):

1. Management support
2. Champion
3. Resources
4. User participation

5. Team skills
6. Source systems
7. Development technology

In many organizations, a data warehouse will be successful only if there is strong senior management support for its development and if there is a project champion. Although this would likely be true for any IT project, it is especially important for a data warehouse. The successful implementation of a data warehouse results in the establishment of an architectural framework that may allow for decision analysis throughout an organization and in some cases also provides comprehensive SCM by granting access to an organization's customers and suppliers. The implementation of Web-based data warehouses (called *Webhousing*) has facilitated ease of access to vast amounts of data, but it is difficult to determine the hard benefits associated with a data warehouse. Hard benefits are defined as benefits to an organization that can be expressed in monetary terms. Many organizations have limited IT resources and must prioritize projects. Management support and a strong project champion can help ensure that a data warehouse project will receive the resources necessary for successful implementation. Data warehouse resources can be a significant cost, in some cases requiring high-end processors and large increases in direct-access storage devices (DASD). Web-based data warehouses may also have special security requirements to ensure that only authorized users have access to the data.

User participation in the development of data and access modeling is a critical success factor in data warehouse development. During data modeling, expertise is required to determine what data are needed, define business rules associated with the data, and decide what aggregations and other calculations may be necessary. Access modeling is needed to determine how data are to be retrieved from a data warehouse, and it assists in the physical definition of the warehouse by helping to define which data require indexing. It may also indicate whether dependent data marts are needed to facilitate information retrieval. The team skills needed to develop and implement a data warehouse include in-depth knowledge of the database technology and development tools used. Source systems and development technology, as mentioned previously, reference the many inputs and the processes used to load and maintain a data warehouse.

Application Case 8.6 presents an excellent example for a large-scale implementation of an integrated data warehouse in the insurance industry.

APPLICATION CASE 8.6

A Large Insurance Company Integrates Its Enterprise Data with AXIS

A large U.S. insurance company developed an integrated data management and reporting environment to provide a unified view of the enterprise performance and risk, and to take a new strategic role in planning and management activities of the large number of business units.

XYZ Insurance company (the actual name is not revealed) and its affiliated companies constitute one of the world's largest financial services organizations. Incorporated a century ago, XYZ Insurance has grown and diversified to become a leading provider of domestic property and casualty insurance, life insurance, retirement savings, asset management, and strategic investment services. Today the firm is an industry powerhouse with over $150 billion in statutory assets, over $15 billion in annual revenue, 20,000+ employees, and more than 100 companies operating under the XYZ Insurance umbrella.

Problem

For most of its years in business, the growing family of XYZ Insurance companies enjoyed considerable independence and autonomy. Over time, as the enterprise got bigger, such a decentralized management style produced an equally diverse reporting and decision-making environment. With no common view of enterprise performance, corporate reporting was shortsighted, fragmented, slow, and often inaccurate. The burden of acquiring, consolidating, cleaning, and validating basic financial information crippled the organization's ability to support management with meaningful analysis and insight.

In order to address the pressing needs for integration, in January 2004 XYZ Insurance launched a needs analysis initiative, which resulted in a shared vision for having a unified data management system. The integrated system, called AXIS, was envisioned to be capable of supporting enterprise-level planning, capital management, risk assessment, and managerial decision making with state-of-the-art reporting and analytical services that were timely, accurate, and efficient.

Solution

XYZ Insurance decided to develop AXIS using a best-of-breed approach. As opposed to buying all of the components from a single vendor, it chose the best fit for each module as determined by the needs analysis. The following tools/vendors were selected:

- The data warehouse: The AXIS environment has a hub-and-spoke architecture with a Teradata data warehouse at the center.
- Extraction, transportation, integration, and metadata management: All data movement from originating systems into the AXIS environment (and between systems within AXIS) is handled by Informatica PowerCenter.
- Reporting and analysis: Virtually all reporting and analytical functionality in the AXIS environment is provided through a suite of Hyperion tools that includes Essbase, Planning, Reporter, Analyzer, and Intelligence.
- Master data management: Reference data hierarchies and dimensions and business rules for interface translation and transformation are developed and maintained using master data management software from Kalido.

Results

Implementing the AXIS environment was a monumental undertaking, even for an organization with XYZ Insurance's resources. More than 200 operational source system interfaces had to be created. At its peak, the development team employed 280 people (60% from internal IT and the business departments and 40% external contractors) who dedicated 600,000 man-hours to the project. The system with full functionality was released in April 2006.

By standardizing the information assets along with the technology base and supporting processes, XYZ Insurance was able to consolidate much of the labor-intensive transactional and reporting activities. That freed up people and resources for more strategic, high-value contributions to the business. Another benefit is that business units across the organization now have consistent and accurate operating information on which to base their decisions. Probably the most important benefit of the AXIS environment is its ability to turn XYZ Insurance into an agile enterprise. Because they have access to corporate level data in a timely manner, the business units can react to changing conditions (address problems and take advantage of opportunities) accurately and rapidly.

Source: Teradata, "A Large US-based Insurance Company Masters Its Finance Data," Teradata Industry Solutions, **teradata.com/t/WorkArea/DownloadAsset.aspx?id=4858** (accessed July 2009).

Massive Data Warehouses and Scalability

In addition to flexibility, a data warehouse needs to support scalability. The main issues pertaining to scalability are the amount of data in the warehouse, how quickly the warehouse is expected to grow, the number of concurrent users, and the complexity of user queries. A data warehouse must scale both horizontally and vertically. The warehouse will grow as a function of data growth and the need to expand the warehouse to support new business functionality. Data growth may be a result of the addition of current cycle data (e.g., this month's results) and/or historical data.

Hicks (2001) described huge databases and data warehouses. Wal-Mart is continually increasing the size of its massive data warehouse. Wal-Mart is believed to use a warehouse with hundreds of terabytes of data to study sales trends and track inventory and other tasks. IBM recently publicized its 50-terabyte warehouse benchmark (IBM, 2009). The U.S. Department of Defense is using a 5-petabyte data warehouse and repository to hold medical records for 9 million military personnel. Because of the storage required to archive its news footage, CNN also has a petabyte-sized data warehouse.

Given that the size of data warehouses is expanding at an exponential rate, scalability is an important issue. Good scalability means that queries and other data-access functions will grow (ideally) linearly with the size of the warehouse. See Rosenberg (2006) for approaches to improve query performance. In practice, specialized methods have been developed to create scalable data warehouses. Scalability is difficult when managing hundreds of terabytes or more. Terabytes of data have considerable inertia, occupy a lot of physical space, and require powerful computers. Some firms use parallel processing, and others use clever indexing and search schemes to manage their data. Some spread their data across different physical data stores. As more data warehouses approach the petabyte size, better and better solutions to scalability continue to be developed.

Hall (2002) also addressed scalability issues. AT&T is an industry leader in deploying and using massive data warehouses. With its 26-terabyte data warehouse, AT&T can detect fraudulent use of calling cards and investigate calls related to kidnappings and other crimes. It can also compute millions of call-in votes from TV viewers selecting the next American Idol.

For a sample of successful data warehousing implementations, see Edwards (2003). Jukic and Lang (2004) examined the trends and specific issues related to the use of offshore resources in the development and support of data warehousing and BI applications. Davison (2003) indicated that IT-related offshore outsourcing had been growing at 20 to 25 percent per year. When considering offshoring data warehousing projects, careful consideration must be given to culture and security (for details, see Jukic and Lang, 2004).

Section 8.6 Review Questions

1. List the benefits of data warehouses.
2. List several criteria for selecting a data warehouse vendor, and describe why they are important.
3. Does a bottom-up data warehouse development approach use an enterprise data model?
4. Describe the major similarities and differences between the Inmon and Kimball data warehouse development approaches.
5. List the different types of data warehouse architectures.

8.7 REAL-TIME DATA WAREHOUSING

Data warehousing and BI tools traditionally focus on assisting managers in making strategic and tactical decisions. Increased data volumes and accelerating update speeds are fundamentally changing the role of the data warehouse in modern business. For many businesses, making fast and consistent decisions across the enterprise requires more than a traditional data warehouse or data mart. Traditional data warehouses are not business critical. Data are commonly updated on a weekly basis, and this does not allow for responding to transactions in near-real-time.

More data, coming in faster and requiring immediate conversion into decisions, means that organizations are confronting the need for real-time data warehousing. This is because decision support has become operational, integrated BI requires closed-loop analytics, and yesterday's ODS will not support existing requirements.

In 2003, with the advent of real-time data warehousing, there was a shift toward using these technologies for operational decisions. **Real-time data warehousing** (RDW), also known as **active data warehousing (ADW)**, is the process of loading and providing data via the data warehouse as they become available. It evolved from the

EDW concept. The active traits of an RDW/ADW supplement and expand traditional data warehouse functions into the realm of tactical decision making. People throughout the organization who interact directly with customers and suppliers will be empowered with information-based decision making at their fingertips. Even further leverage results when an ADW provides information directly to customers and suppliers. The reach and impact of information access for decision making can positively affect almost all aspects of customer service, SCM, logistics, and beyond. E-business has become a major catalyst in the demand for active data warehousing (see Armstrong, 2000). For example, online retailer Overstock.com, Inc. (**overstock.com**), connected data users to a real-time data warehouse. At Egg plc, the world's largest purely online bank, a customer data warehouse is refreshed in near-real-time. See Application Case 8.7.

APPLICATION CASE 8.7

Egg Plc Fries the Competition in Near-Real-Time

Egg plc (**egg.com**) is the world's largest online bank. It provides banking, insurance, investments, and mortgages to more than 3.6 million customers through its Internet site. In 1998, Egg selected Sun Microsystems to create a reliable, scalable, secure infrastructure to support its more than 2.5 million daily transactions. In 2001, the system was upgraded to eliminate latency problems. This new customer data warehouse (CDW) used Sun, Oracle, and SAS software products. The initial data warehouse had about 10 terabytes of data and used a 16-CPU server. The system provides near-real-time data access. It provides data warehouse and data mining services to internal users, and it provides a requisite set of customer data to the customers themselves. Hundreds of sales and marketing campaigns are constructed using near-real-time data (within several minutes). And better, the system enables faster decision making about specific customers and customer classes.

Sources: Compiled from "Egg's Customer Data Warehouse Hits the Mark," *DM Review*, Vol. 15, No. 10, October 2005, pp. 24–28; Sun Microsystems, "Egg Banks on Sun to Hit the Mark with Customers," September 19, 2005, **sun.com/smi/Press/sunflash/2005-09/sunflash.20050919.1.xml** (accessed April 2006); and ZD Net UK, "Sun Case Study: Egg's Customer Data Warehouse," **whitepapers.zdnet.co.uk/0,39025945,60159401p-39000449q,00.htm** (accessed June 2009).

As business needs evolve, so do the requirements of the data warehouse. At this basic level, a data warehouse simply reports what happened. At the next level, some analysis occurs. As the system evolves, it provides prediction capabilities, which lead to the next level of operationalization. At its highest evolution, the ADW is capable of making events happen (e.g., activities such as creating sales and marketing campaigns or identifying and exploiting opportunities). See Figure 8.10 for a graphic description of this evolutionary process. A recent survey on managing evolution of data warehouses can be found in Wrembel (2009).

Teradata Corp. provides the baseline requirements to support an EDW. It also provides the new traits of active data warehousing required to deliver data freshness, performance, and availability and to enable enterprise decision management (see Figure 8.11 for an example).

An ADW offers an integrated information repository to drive strategic and tactical decision support within an organization. With real-time data warehousing, instead of extracting operational data from an OLTP system in nightly batches into an ODS, data are assembled from OLTP systems as and when events happen and are moved at once into the data warehouse. This permits the instant updating of the data warehouse and the elimination of an ODS. At this point, tactical and strategic queries can be made against the RDW to use immediate as well as historical data.

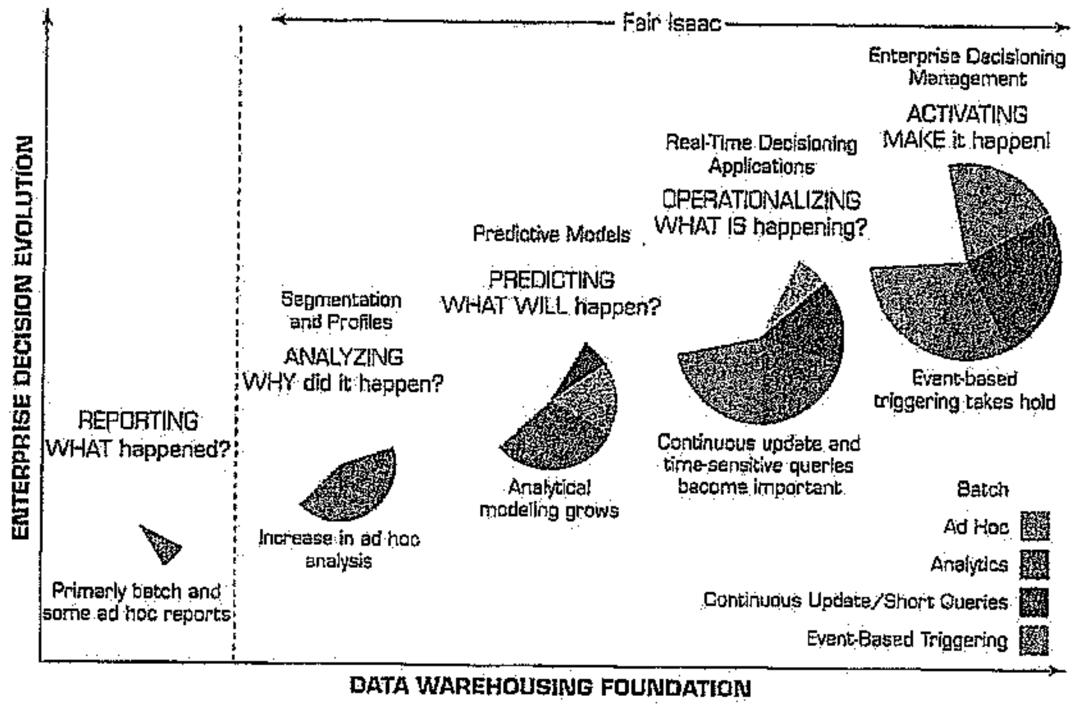


FIGURE 8.10 Enterprise Decision Evolution *Source:* Courtesy of Teradata Corporation. Used with permission.

According to Basu (2003), the most distinctive difference between a traditional data warehouse and an RDW is the shift in the data acquisition paradigm. Some of the business cases and enterprise requirements that led to the need for data in real time include the following:

- A business often cannot afford to wait a whole day for its operational data to load into the data warehouse for analysis.
- Until now, data warehouses have captured snapshots of an organization's fixed states instead of incremental real-time data showing every state change and almost analogous patterns over time.
- With a traditional hub-and-spoke architecture, keeping the metadata in sync is difficult. It is also costly to develop, maintain, and secure many systems as opposed to one huge data warehouse so that data are centralized for BI/BA tools.
- In cases of huge nightly batch loads, the necessary ETL setup and processing power for large nightly data warehouse loading might be very high, and the processes might take too long. An EAI with real-time data collection can reduce or eliminate the nightly batch processes.

Despite the benefits of an RDW, developing one can create its own set of issues. These problems relate to architecture, data modeling, physical database design, storage and scalability, and maintainability. In addition, depending on exactly when data are accessed, even down to the microsecond, different versions of the truth may be extracted and created, which can confuse team members. For details, refer to Basu (2003) and Terr (2004).

Real-time solutions present a remarkable set of challenges to BI activities. Although it is not ideal for all solutions, real-time data warehousing may be successful if the organization develops a sound methodology to handle project risks, incorporate proper planning, and focus on quality assurance activities. Understanding the common challenges and applying best practices can reduce the extent of the

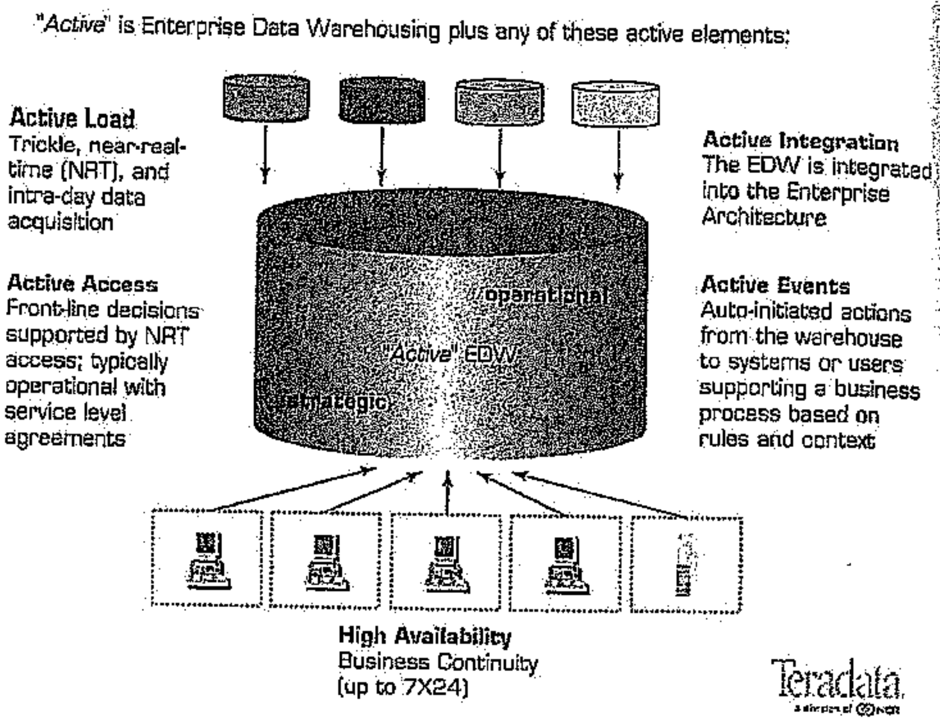


FIGURE 8.11 The Teradata Active EDW *Source:* Courtesy of Teradata Corporation.

problems that are often a part of implementing complex data warehousing systems that incorporate BI/BA methods. Details and real implementations are discussed by Burdett and Singh (2004) and Wilk (2003). Also see Akbay (2006) and Ericson (2006).

See Technology Insights 8.2 for some details on how the real-time concept evolved. The flight management dashboard application at Continental Airlines (see the end of chapter application case) illustrates the power of real-time BI in accessing a data warehouse for use in face-to-face customer interaction situations. The operations staff uses the real-time system to identify issues in the Continental flight network. As another example, UPS invested $600 million so it could use real-time data and processes. The investment was expected to cut 100 million delivery miles and save 14 million gallons of fuel annually by managing its real-time package-flow technologies (see Malykhina, 2003). Table 8.5 compares traditional and active data warehousing environments.

Real-time data warehousing, near-real-time data warehousing, zero-latency warehousing, and *active data warehousing* are different names used in practice to describe the same concept. Gonzales (2005) presented different definitions for ADW. According to Gonzales, ADW is only one option that provides blended tactical and strategic data on demand. The architecture to build an ADW is very similar to the corporate information factory architecture developed by Bill Inmon. The only difference between a corporate information factory and an ADW is the implementation of both data stores in a single environment. However, an SOA based on XML and Web services provides another option for blending tactical and strategic data on demand.

One critical issue in real-time data warehousing is that not all data should be updated continuously. This may certainly cause problems when reports are generated in real time, because one person's results may not match another person's. For example, a company using Business Objects Web Intelligence noticed a significant problem with real-time

TECHNOLOGY INSIGHTS 8.2 The Real-Time Realities of Active Data Warehousing

By 2003, the role of data warehousing in practice was growing rapidly. Real-time systems, though a novelty, were the latest buzz, along with the major complications of providing data and information instantaneously to those who need them. Many experts, including Peter Coffee, *eWeek*'s technology editor, believe that real-time systems must feed a real-time decision-making process. Stephen Brobst, CTO of the Teradata division of NCR, indicated that active data warehousing is a process of evolution in how an enterprise uses data. *Active* means that the data warehouse is also used as an operational and tactical tool. Brobst provided a five-stage model that fits Coffee's experience (2003) of how organizations "grow" in their data utilization (see Brobst et al., 2005). These stages (and the questions they purport to answer) are reporting (What happened?), analysis (Why did it happen?), prediction (What will happen?), operationalizing (What is happening?), and active warehousing (What do I want to happen?). The last stage, active warehousing, is where the greatest benefits may be obtained. Many organizations are enhancing centralized data warehouses to serve both operational and strategic decision making.

Sources: Adapted from P. Coffee, "'Active' Warehousing," *eWeek*, Vol. 20, No. 25, June 23, 2003, p. 36; and Teradata Corp., "Active Data Warehousing," **teradata.com/t/page/87127/index.html** (accessed April 2006).

intelligence. Real-time reports produced at slightly different times differ (see Peterson, 2003). Also, it may not be necessary to update certain data continuously (e.g., course grades that are 3 or more years old).

Real-time requirements change the way we view the design of databases, data warehouses, OLAP, and data mining tools, because they are literally updated concurrently while queries are active. But the substantial business value in doing so has been demonstrated, so it is crucial that organizations adopt these methods in their business processes. Careful planning is critical in such implementations.

TABLE 8.5 Comparison of Traditional and Active Data Warehousing Environments

Traditional Data Warehouse Environment	Active Data Warehouse Environment
Strategic decisions only	Strategic and tactical decisions
Results sometimes hard to measure	Results measured with operations
Daily, weekly, monthly data currency acceptable; summaries often appropriate	Only comprehensive detailed data available within minutes is acceptable
Moderate user concurrency	High number (1,000 or more) of users accessing and querying the system simultaneously
Highly restrictive reporting used to confirm or check existing processes and patterns; often uses predeveloped summary tables or data marts	Flexible ad hoc reporting, as well as machine-assisted modeling (e.g., data mining) to discover new hypotheses and relationships
Power users, knowledge workers, internal users	Operational staffs, call centers, external users

Sources: Adapted from P. Coffee, "'Active' Warehousing," *eWeek*, Vol. 20, No. 25, June 23, 2003, p. 36; and Teradata Corp., "Active Data Warehousing," **teradata.com/t/page/87127/index.html** (accessed April 2006).

Section 8.7 Review Questions

1. What is an RDW?
2. List the benefits of an RDW.
3. What are the major differences between a traditional data warehouse and an RDW?
4. List some of the drivers for RDW.

8.8 DATA WAREHOUSE ADMINISTRATION AND SECURITY ISSUES

Data warehouses provide a distinct competitive edge to enterprises that effectively create and use them. Due to its huge size and its intrinsic nature, a data warehouse requires especially strong monitoring in order to sustain satisfactory efficiency and productivity. The successful administration and management of a data warehouse entails skills and proficiency that go past what is required of a traditional database administrator (DBA). A **data warehouse administrator (DWA)** should be familiar with high-performance software, hardware, and networking technologies. He or she should also possess solid business insight. Because data warehouses feed BI systems and DSS that help managers with their decision-making activities, the DWA should be familiar with the decision-making processes so as to suitably design and maintain the data warehouse structure. It is particularly significant for a DWA to keep the existing requirements and capabilities of the data warehouse stable while simultaneously providing flexibility for rapid improvements. Finally, a DWA must possess excellent communications skills. See Benander et al. (2000) for a description of the key differences between a DBA and a DWA.

Security and privacy of information are main and significant concerns for a data warehouse professional. The U.S. government has passed regulations (e.g., the Gramm-Leach-Bliley privacy and safeguards rules, the Health Insurance Portability and Accountability Act of 1996 [HIPAA]) instituting obligatory requirements in the management of customer information. Hence, companies must create security procedures that are effective yet flexible to conform to numerous privacy regulations. According to Elson and LeClerc (2005), effective security in a data warehouse should focus on four main areas:

1. Establishing effective corporate and security policies and procedures. An effective security policy should start at the top, with executive management, and should be communicated to all individuals within the organization.
2. Implementing logical security procedures and techniques to restrict access. This includes user authentication, access controls, and encryption technology.
3. Limiting physical access to the data center environment.
4. Establishing an effective internal control review process with an emphasis on security and privacy.

See Technology Insights 8.3 for a description of Ambeo's important software tool that monitors security and privacy of data warehouses. Finally, keep in mind that accessing a data warehouse via a mobile device should always be performed cautiously. In this instance, data should only be accessed as read-only.

In the near term, data warehousing developments will be determined by noticeable factors (e.g., data volumes, increased intolerance for latency, the diversity and complexity of data types) and less noticeable factors (e.g., unmet end-user requirements for dashboards, balanced scorecards, master data management, information quality). Given these drivers, Moseley (2009) and Agosta (2006) suggested that data warehousing trends will lean toward simplicity, value, and performance.

TECHNOLOGY INSIGHTS 8.3 Ambeo Delivers Proven Data Access Auditing Solution

Since 1997, Ambeo (**ambeo.com**; now Embarcadero Technologies, Inc.) has deployed technology that provides performance management, data usage tracking, data privacy auditing, and monitoring to *Fortune* 1000 companies. These firms have some of the largest database environments in existence. Ambeo data access auditing solutions play a major role in an enterprise information security infrastructure.

The Ambeo technology is a relatively easy solution that records everything that happens in the databases, with low or zero overhead. In addition, it provides data access auditing that identifies exactly who is looking at data, when they are looking, and what they are doing with the data. This real-time monitoring helps quickly and effectively identify security breaches.

Sources: Adapted from "Ambeo Delivers Proven Data Access Auditing Solution," *Database Trends and Applications*, Vol. 19, No. 7, July 2005; and Ambeo, "Keeping Data Private (and Knowing It): Moving Beyond Conventional Safeguards to Ensure Data Privacy," **am-beo.com/why_ambeo_white_papers.html** (accessed May 2009).

Section 8.8 Review Questions

1. What steps can an organization take to ensure the security and confidentiality of customer data in its data warehouse?
2. What skills should a DWA possess? Why?

8.9 RESOURCES, LINKS, AND THE TERADATA UNIVERSITY NETWORK CONNECTION

The use of this chapter and most other chapters in this book can be enhanced by the tools described in the following sections.

Resources and Links

We recommend looking at the following resources and links for further reading and explanations:

- The Data Warehouse Institute (**tdwi.com**)
- *DM Review* (**dmreview.com**)
- DSS Resources (**dssresources.com**)

Cases

All major MSS vendors (e.g., MicroStrategy, Microsoft, Oracle, IBM, Hyperion, Cognos, Exsys, Fair Isaac, SAP, Information Builders) provide interesting customer success stories. Academic-oriented cases are available at the Harvard Business School Case Collection (**harvardbusinessonline.hbsp.harvard.edu**), Business Performance Improvement Resource (**bpir.com**), Idea Group Publishing (**idea-group.com**), Ivy League Publishing (**ivylp.com**), ICFAI Center for Management Research (**icmr.icfai.org/casestudies/icmr_case_studies.htm**), KnowledgeStorm (**knowledgestorm.com**), and other sites. For additional case resources, see Teradata University Network (**teradatauniversitynetwork.com**). For data warehousing cases, we specifically recommend the following from the Teradata University Network (**teradatauniversitynetwork.com**): "Continental Airlines Flies High with Real-Time Business Intelligence," "Data Warehouse Governance at Blue

Cross and Blue Shield of North Carolina," "3M Moves to a Customer Focus Using a Global Data Warehouse," "Data Warehousing Supports Corporate Strategy at First American Corporation," "Harrah's High Payoff from Customer Information," and "Whirlpool." We also recommend the Data Warehousing Failures Assignment, which consists of eight short cases on data warehousing failures.

Vendors, Products, and Demos

A comprehensive list of vendors, products, and demos is available at *DM Review* (**dmreview.com**). Vendors are listed in Table 8.2. Also see **technologyevaluation.com**.

Periodicals

We recommend the following periodicals:

- *Baseline* (**baselinemag.com**)
- *Business Intelligence Journal* (**tdwi.org**)
- *CIO* (**cio.com**)
- *CIO Insight* (**cioinsight.com**)
- *Computerworld* (**computerworld.com**)
- *Decision Support Systems* (**elsevier.com**)
- *DM Review* (**dmreview.com**)
- *eWeek* (**eweek.com**)
- *InfoWeek* (**infoweek.com**)
- *InfoWorld* (**infoworld.com**)
- *InternetWeek* (**internetweek.com**)
- *Management Information Systems Quarterly* (*MIS Quarterly*; **misq.org**)
- *Technology Evaluation* (**technologyevaluation.com**)
- *Teradata Magazine* (**teradata.com**)

Additional References

For additional information on data warehousing, see the following.

- C. Imhoff, N. Galemmo, and J. G. Geiger. (2003). *Mastering Data Warehouse Design: Relational and Dimensional Techniques*. New York: Wiley.
- D. Marco and M. Jennings. (2004). *Universal Meta Data Models*. New York: Wiley.
- J. Wang. (2005). *Encyclopedia of Data Warehousing and Mining*. Hershey, PA: Idea Group Publishing.

For more on databases, the structure on which data warehouses are developed, see the following:

- R. T. Watson. (2006). *Data Management*, 5th ed. New York: Wiley.

The Teradata University Network (TUN) Connection

TUN (**teradatauniversitynetwork.com**) provides a wealth of information and cases on data warehousing. One of the best is the Continental Airlines case, which we require you to solve in a later exercise. Other recommended cases are mentioned earlier in this chapter. At TUN, if you click the Courses tab and select Data Warehousing, you will see links to many relevant articles, assignments, book chapters, course Web sites, PowerPoint presentations, projects, research reports, syllabi, and Web seminars. You will also find links to active data warehousing software demonstrations. Finally, you will see links to Teradata (**teradata.com**), where you can find additional information, including excellent data warehousing success stories, white papers, Web-based courses, and the online version of *Teradata Magazine*.

Chapter Highlights

- A data warehouse is a specially constructed data repository where data are organized so that they can be easily accessed by end users for several applications.
- Data marts contain data on one topic (e.g., marketing). A data mart can be a replication of a subset of data in the data warehouse. Data marts are a less expensive solution that can be replaced by or can supplement a data warehouse. Data marts can be independent of or dependent on a data warehouse.
- An ODS is a type of customer-information-file database that is often used as a staging area for a data warehouse.
- Data integration comprises three major processes: data access, data federation, and change capture. When these three processes are correctly implemented, data can be accessed and made accessible to an array of ETL and analysis tools and data warehousing environments.
- ETL technologies pull data from many sources, cleanse them, and load them into a data warehouse. ETL is an integral process in any data-centric project.
- Real-time or active data warehousing supplements and expands traditional data warehousing, moving into the realm of operational and tactical decision making by loading data in real time and providing data to users for active decision making.
- The security and privacy of data and information are critical issues for a data warehouse professional.

Key Terms

active data warehousing (ADW) *359*
data integration *342*
data mart *330*
data warehouse (DW) *329*
data warehouse administrator (DWA) *364*
dependent data mart *330*
dimension table *351*
dimensional modeling *351*
drill down *352*
enterprise application integration (EAI) *343*
enterprise data warehouse (EDW) *330*
enterprise information integration (EII) *344*
extraction, transformation, and load (ETL) *344*
grain *352*
independent data mart *330*
metadata *332*
oper mart *330*
operational data store (ODS) *330*
real-time data warehousing (RDW) *359*

Questions for Discussion

1. Compare data integration and ETL. How are they related?
2. What is a data warehouse, and what are its benefits? Why is Web accessibility important with a data warehouse?
3. A data mart can replace a data warehouse or complement it. Compare and discuss these options.
4. Discuss the major drivers and benefits of data warehousing to end users.
5. List the differences and/or similarities between the roles of a database administrator and a data warehouse administrator.
6. Describe how data integration can lead to higher levels of data quality.
7. Compare the Kimball and Inmon approaches toward data warehouse development. Identify when each one is most effective.
8. Discuss security concerns involved in building a data warehouse.
9. Investigate current data warehouse development implementation through offshoring. Write a report about it. In class, debate the issue in terms of the benefits and costs, as well as social factors.

Exercises

TERADATA STUDENT NETWORK (TSN) AND OTHER HANDS-ON EXERCISES

1. Consider the case describing the development and application of a data warehouse for Coca-Cola Japan (a summary appears in Application Case 8.4), available at the DSS Resources Web site. **dssresources.com/cases/cocacolajapan/index.html**. Read the case and answer the nine questions for further analysis and discussion.
2. Read the Ball (2005) article and rank-order the criteria (ideally for a real organization). In a report, explain how important each criterion is and why.
3. Explain when you should implement a two- or three-tiered architecture when considering developing a data warehouse.

4. Read the full Continental Airlines case (summarized in the end-of-chapter case) at **teradatastudentnetwork.com** and answer the questions.
5. At **teradatastudentnetwork.com**, read and answer the questions to the case "Harrah's High Payoff from Customer Information." Relate Harrah's results to how airlines and other casinos use their customer data.
6. At **teradatastudentnetwork.com**, read and answer the questions of the assignment "Data Warehousing Failures." Because eight cases are described in that assignment, the class may be divided into eight groups, with one case assigned per group. In addition, read Ariyachandra and Watson (2006a), and for each case identify how the failure occurred as related to not focusing on one or more of the reference's success factor(s).
7. At **teradatastudentnetwork.com**, read and answer the questions with the assignment "Ad-Vent Technology: Using the MicroStrategy Sales Analytic Model." The MicroStrategy software is accessible from the TUN site. Also, you might want to use Barbara Wixom's PowerPoint presentation about the MicroStrategy software ("Demo Slides for MicroStrategy Tutorial Script"), which is also available at the TUN site.
8. At **teradatastudentnetwork.com**, watch the Web seminars titled "Real-Time Data Warehousing: The Next Generation of Decision Support Data Management" and "Building the Real-Time Enterprise." Read the article "Teradata's Real-Time Enterprise Reference Architecture: A Blueprint for the Future of IT," also available at this site. Describe how real-time concepts and technologies work and how they can be used to extend existing data warehousing and BI architectures to support day-to-day decision making. Write a report indicating how real-time data warehousing is specifically providing competitive advantage for organizations. Describe in detail the difficulties in such implementations and operations and describe how they are being addressed in practice.
9. At **teradatastudentnetwork.com**, watch the Web seminars "Data Integration Renaissance: New Drivers and Emerging Approaches," "In Search of a Single Version of the Truth: Strategies for Consolidating Analytic Silos," and "Data Integration: Using ETL, EAI, and EII Tools to Create an Integrated Enterprise." Also read the "Data Integration" research report. Compare and contrast the presentations. What is the most important issue described in these seminars? What is the best way to handle the strategies and challenges of consolidating data marts and spreadsheets into a unified data warehousing architecture? Perform a Web search to identify the latest developments in the field. Compare the presentation to the material in the text and the new material that you found.
10. Consider the future of data warehousing. Perform a Web search on this topic. Also, read these two articles: L. Agosta, "Data Warehousing in a Flat World: Trends for 2006," *DM Direct Newsletter*, March 31, 2006; and J. G. Geiger, "CIFe: Evolving with the Times," *DM Review*, November 2005, pp. 38–41. Compare and contrast your findings.
11. Access **teradatastudentnetwork.com**. Identify the latest articles, research reports, and cases on data warehousing. Describe recent developments in the field. Include in your report how data warehousing is used in BI and DSS.

TEAM ASSIGNMENTS AND ROLE-PLAYING PROJECTS

1. Kathryn Avery has been a DBA with a nationwide retail chain (Big Chain) for the past 6 years. She has recently been asked to lead the development of Big Chain's first data warehouse. The project has the sponsorship of senior management and the CIO. The rationale for developing the data warehouse is to advance the reporting systems, particularly in sales and marketing, and, in the longer term, to improve Big Chain's CRM. Kathryn has been to a Data Warehousing Institute conference and has been doing some reading, but she is still mystified about development methodologies. She knows there are two groups—EDW (Inmon) and architected data marts (Kimball)—that have equally robust features.

   Initially, she believed that the two methodologies were extremely dissimilar, but as she has examined them more carefully, she isn't so certain. Kathryn has a number of questions that she would like answered:

   a. What are the real differences between the methodologies?
   b. What factors are important in selecting a particular methodology?
   c. What should be her next steps in thinking about a methodology?

   Help Kathryn answer these questions. (This exercise was adapted from K. Duncan, L. Reeves, and J. Griffin, "BI Experts' Perspective," *Business Intelligence Journal*, Vol. 8, No. 4, Fall 2003, pp. 14–19.)
2. Jeet Kumar is the administrator of data warehousing at a big regional bank. He was appointed 5 years ago to implement a data warehouse to support the bank's CRM business strategy. Using the data warehouse, the bank has been successful in integrating customer information, understanding customer profitability, attracting customers, enhancing customer relationships, and retaining customers.

   Over the years, the bank's data warehouse has moved closer to real time by moving to more frequent refreshes of the data warehouse. Now, the bank wants to implement customer self-service and call center applications that require even fresher data than is currently available in the warehouse.

   Jeet wants some support in considering the possibilities for presenting fresher data. One alternative is to entirely commit to implementing real-time data warehousing. His ETL vendor is prepared to assist him make

EAI and EII technologies and wonders how they might fit into his plans.

In particular, he has the following questions:

a. What exactly are EAI and EII technologies?
b. How are EAI and EII related to ETL?
c. How are EAI and EII related to real-time data warehousing?
d. Are EAI and EII required, complementary, or alternatives to real-time data warehousing?

Help Jeet answer these questions. (This exercise was adapted from S. Brobst, E. Levy, and C. Muzilla, "Enterprise Application Integration and Enterprise Information Integration." *Business Intelligence Journal*, Vol. 10, No. 2, Spring 2005, pp. 27–32.)

3. Interview administrators in your college or executives in your organization to determine how data warehousing could assist them in their work. Write up a proposal describing your findings. Include cost estimates and benefits in your report.
4. Go through the list of data warehousing risks described in this chapter and find two examples of each in practice.
5. Access **teradata.com** and read the white papers "Measuring Data Warehouse ROI" and "Realizing ROI: Projecting and Harvesting the Business Value of an Enterprise Data Warehouse." Also, watch the Web-based course "The ROI Factor: How Leading Practitioners Deal with the Tough Issue of Measuring DW ROI." Describe the most important issues described in them. Compare these issues to the success factors described in Ariyachandra and Watson (2006a).
6. Read the article K. Liddell Avery and Hugh J. Watson, "Training Data Warehouse End-users," *Business Intelligence Journal*, Vol. 9, No. 4, Fall 2004, pp. 40–51 (which is available at **teradatastudentnetwork.com**). Consider the different classes of end users, describe their difficulties, and discuss the benefits of appropriate training for each group. Have each member of the group take on one of the roles and have a discussion about how an appropriate type of data warehousing training would be good for each of you.

INTERNET EXERCISES

1. Search the Internet to find information about data warehousing. Identify some newsgroups that have an interest in this concept. Explore ABI INFORM in your library, e-library, and Google for recent articles on the topic. Begin with **tdwi.com**, **technologyevaluation.com**, and the major vendors: **teradata.com**, **sas.com**, **oracle.com**, and **ncr.com**. Also check **cio.com**, **dmreview.com**, **dssresources.com**, and **db2mag.com**.
2. Survey some ETL tools and vendors. Start with **fairisaac.com** and **egain.com**. Also with **dmreview.com**.
3. Contact some data warehouse vendors and obtain information about their products. Give special attention to vendors that provide tools for multiple purposes, such as Cognos, Software A&G, SAS Institute, and Oracle. Free online demos are available from some of these vendors. Download a demo or two and try them. Write a report describing your experience.
4. Explore **teradata.com** for developments and success stories about data warehousing. Write a report about what you have discovered.
5. Explore **teradata.com** for white papers and Web-based courses on data warehousing. Read the former and watch the latter. (Divide up the class so that all the sources are covered.) Write up a report on what you have discovered.
6. Find recent cases of successful data warehousing applications. Go to data warehouse vendors' sites and look for cases or success stories. Select one and write a brief summary to present to your class.

END OF CHAPTER APPLICATION CASE

Continental Airlines Flies High with Its Real-Time Data Warehouse

As business intelligence (BI) becomes a critical component of daily operations, real-time data warehouses that provide end users with rapid updates and alerts generated from transactional systems are increasingly being deployed. Real-time data warehousing and BI, supporting its aggressive Go Forward business plan, have helped Continental Airlines alter its industry status from "worst to first" and then from "first to favorite." Continental Airlines is a leader in real-time BI. In 2004, Continental won the Data Warehousing Institute's Best Practices and Leadership Award.

Problem(s)

Continental Airlines was founded in 1934, with a single-engine Lockheed aircraft in the Southwestern United States. As of 2006, Continental was the fifth largest airline in the United States and the seventh largest in the world. Continental has the broadest global route network of any U.S. airline, with more than 2,300 daily departures to more than 227 destinations.

Back in 1994, Continental was in deep financial trouble. It had filed for Chapter 11 bankruptcy protection twice and was heading for its third, and probably final, bankruptcy. Ticket sales were hurting because performance on factors that are important to customers was dismal, including a low percentage of on-time departures, frequent baggage arrival problems, and too many customers turned away due to overbooking.

Solution

The revival of Continental began in 1994, when Gordon Bethune became CEO and initiated the Go Forward plan, which consisted of four interrelated parts to be implemented simultaneously. Bethune targeted the need to improve customer-valued performance measures by better understanding customer needs as well as customer perceptions of the value of services that were and could be offered. Financial management practices were also targeted for a significant overhaul. As early as 1998, the airline had separate databases for marketing and operations, all hosted and managed by outside vendors. Processing queries and instigating marketing programs to its high-value customers were time-consuming and ineffective. In additional, information that the workforce needed to make quick decisions was simply not available. In 1999, Continental chose to integrate its marketing, IT, revenue, and operational data sources into a single, in-house EDW. The data warehouse provided a variety of early, major benefits.

As soon as Continental returned to profitability and ranked first in the airline industry in many performance metrics, Bethune and his management team raised the bar by escalating the vision. Instead of just performing best, they wanted Continental to be their customers' favorite airline. The Go Forward plan established more actionable ways to move from first to favorite among customers. Technology became increasingly critical for supporting these new initiatives. In the early days, having access to historical, integrated information was sufficient. This produced substantial strategic value. But it became increasingly imperative for the data warehouse to provide real-time, actionable information to support enterprise-wide tactical decision making and business processes.

Luckily, the warehouse team had expected and arranged for the real-time shift. From the very beginning, the team had created an architecture to handle real-time data feeds into the warehouse, extracts of data from legacy systems into the warehouse, and tactical queries to the warehouse that required almost immediate response times. In 2001, real-time data became available from the warehouse, and the amount stored grew rapidly. Continental moves real-time data (ranging from to-the-minute to hourly) about customers, reservations, check-ins, operations, and flights from its main operational systems to the warehouse. Continental's real-time applications include the following:

- Revenue management and accounting
- Customer relationship management (CRM)
- Crew operations and payroll
- Security and fraud
- Flight operations

Results

In the first year alone, after the data warehouse project was deployed, Continental identified and eliminated over $7 million in fraud and reduced costs by $41 million. With a $30 million investment in hardware and software over 6 years, Continental has reached over $500 million in increased revenues and cost savings in marketing, fraud detection, demand forecasting and tracking, and improved data center management. The single, integrated, trusted view of the business (i.e., the single version of the truth) has led to better, faster decision making.

Continental is now identified as a leader in real-time BI, based on its scalable and extensible architecture, practical decisions on what data are captured in real time, strong relationships with end users, a small and highly competent data warehouse staff, sensible weighing of strategic and tactical decision support requirements, understanding of the synergies between decision support and operations, and changed business processes that use real-time data. (For a sample output screen from the Continental system, see **teradata.com/t/page/139245/**.)

Questions for the Case

1. Describe the benefits of implementing the Continental Go Forward strategy.
2. Explain why it is important for an airline to use a real-time data warehouse.
3. Examine the sample system output screen at **teradata.com/t/page/139245/**. Describe how it can assist the user in identifying problems and opportunities.
4. Identify the major differences between the traditional data warehouse and a real-time

data warehouse, as was implemented at Continental.

5. What strategic advantage can Continental derive from the real-time system as opposed to a traditional information system?

Sources: Adapted from H. Wixom, J. Hoffer, R. Anderson-Lehman, and A. Reynolds, "Real-Time Business Intelligence: Best Practices at Continental Airlines," *Information Systems Management Journal*, Winter 2006, pp. 7–18; R. Anderson-Lehman, H. Watson, B. Wixom, and J. Hoffer, "Continental Airlines Flies High with Real-Time Business Intelligence," *MIS Quarterly Executive*, Vol. 3, No. 4, December 2004, pp. 163–176 (available at **teradatauniversitynetwork.com**); H. Watson, "Real Time: The Next Generation of Decision-Support Data Management," *Business Intelligence Journal*, Vol. 10, No. 3, 2005, pp. 4–6; M. Edwards, "2003 Best Practices Awards Winners: Innovators in Business Intelligence and Data Warehousing," *Business Intelligence Journal*, Fall 2003, pp. 57–64; R. Westervelt, "Continental Airlines Builds Real-Time Data Warehouse," August 20, 2003, **searchoracle.techtarget.com**; R. Clayton, "Enterprise Business Performance Management: Business Intelligence + Data Warehouse = Optimal Business Performance," *Teradata Magazine*, September 2005, **teradata.com/t/page/139245/**; and The Data Warehousing Institute, "2003 Best Practices Summaries: Enterprise Data Warehouse," 2003, **tdwi.org/display.aspx?ID=6749**.

References

Adamson, C. (2009). *The Star Schema Handbook: The Complete Reference to Dimensional Data Warehouse Design*. Hoboken, NJ: Wiley.

Adelman, S., and L. Moss. (2001, Winter). "Data Warehouse Risks." *Journal of Data Warehousing*, Vol. 6, No. 1.

Agosta, L. (2006, January). "The Data Strategy Adviser: The Year Ahead—Data Warehousing Trends 2006." *DM Review*, Vol. 16, No. 1.

Akbay, S. (2006, Quarter 1). "Data Warehousing in Real Time." *Business Intelligence Journal*, Vol. 11, No. 1.

Ambeo. "Keeping Data Private (and Knowing It): Moving Beyond Conventional Safeguards to Ensure Data Privacy." **am-beo.com/why_ambeo_white_papers.html** (accessed May 2009).

Ambeo. (2005, July). "Ambeo Delivers Proven Data Access Auditing Solution." *Database Trends and Applications*, Vol. 19, No. 7.

Anthes, G. H. (2003, June 30). "Hilton Checks into New Suite." *Computerworld*, Vol. 37, No. 26.

Ariyachandra, T., and H. Watson. (2005). "Key Factors in Selecting a Data Warehouse Architecture." *Business Intelligence Journal*, Vol. 10, No. 2.

Ariyachandra, T., and H. Watson. (2006a, January). "Benchmarks for BI and Data Warehousing Success." *DM Review*, Vol. 16, No. 1.

Ariyachandra, T., and H. Watson. (2006b). "Which Data Warehouse Architecture Is Most Successful?" *Business Intelligence Journal*, Vol. 11, No. 1.

Armstrong, R. (2000, Quarter 3). "E-nalysis for the E-business." *Teradata Magazine Online*. **teradata.com**.

Ball, S. K. (2005, November 14). "Do You Need a Data Warehouse Layer in Your Business Intelligence Architecture?" **datawarehouse.ittoolbox.com/documents/industry-articles/do-you-need-a-data-warehouse-layer-in-your-business-intelligencearchitecture-2729** (accessed June 2009).

Barquin, R., A. Paller, and H. Edelstein. (1997). "Ten Mistakes to Avoid for Data Warehousing Managers." In R. Barquin and H. Edelstein (eds.). *Building, Using, and Managing the Data Warehouse*. Upper Saddle River, NJ: Prentice Hall.

Basu, R. (2003, November). "Challenges of Real-Time Data Warehousing." *DM Review*.

Bell, L. D. (2001, Spring). "MetaBusiness Meta Data for the Masses: Administering Knowledge Sharing for Your Data Warehouse." *Journal of Data Warehousing*, Vol. 6, No. 2.

Benander, A., B. Benander, A. Fadlalla, and G. James. (2000, Winter). "Data Warehouse Administration and Management." *Information Systems Management*, Vol. 17, No. 1.

Bonde, A., and M. Kuckuk. (2004, April). "Real World Business Intelligence: The Implementation Perspective." *DM Review*, Vol. 14, No. 4.

Breslin, M. (2004, Winter). "Data Warehousing Battle of the Giants: Comparing the Basics of Kimball and Inmon Models." *Business Intelligence Journal*, Vol. 9, No. 1.

Brobst, S., E. Levy, and C. Muzilla. (2005, Spring). "Enterprise Application Integration and Enterprise Information Integration." *Business Intelligence Journal*, Vol. 10, No. 2.

Brody, R. (2003, Summer). "Information Ethics in the Design and Use of Metadata." *IEEE Technology and Society Magazine*, Vol. 22, No. 2.

Brown, M. (2004, May 9–12). "8 Characteristics of a Successful Data Warehouse." *Proceedings of the Twenty-Ninth Annual SAS Users Group International Conference* (SUGI 29). Montreal, Canada.

Burdett, J., and S. Singh. (2004). "Challenges and Lessons Learned from Real-Time Data Warehousing." *Business Intelligence Journal*, Vol. 9, No. 4.

Coffee, P. (2003, June 23). "'Active' Warehousing." *eWeek*, Vol. 20, No. 25.

Cooper, B. L., H. J. Watson, B. H. Wixom, and D. L. Goodhue. (1999, August 15–19). "Data Warehousing Supports Corporate Strategy at First American Corporation." SIM International Conference, Atlanta.

Cooper, B. L., H. J. Watson, B. H. Wixom, and D. L. Goodhue. (2000). "Data Warehousing Supports Corporate

Strategy at First American Corporation." *MIS Quarterly*, Vol. 24, No. 4, pp. 547–567.

Dasu, T., and T. Johnson. (2003). *Exploratory Data Mining and Data Cleaning*. New York: Wiley.

Davison, D. (2003, November 14). "Top 10 Risks of Offshore Outsourcing." META Group Research Report, now Gartner, Inc., Stamford, CT.

Devlin, B. (2003, Quarter 2). "Solving the Data Warehouse Puzzle." *DB2 Magazine*.

Dragoon, A. (2003, July 1). "All for One View." *CIO*.

Eckerson, W. (2003, Fall). "The Evolution of ETL." *Business Intelligence Journal*, Vol. 8, No. 4.

Eckerson, W. (2005, April 1). "Data Warehouse Builders Advocate for Different Architectures." *Application Development Trends*.

Eckerson, W., R. Hackathorn, M. McGivern, C. Twogood, and G. Watson. (2009). "Data Warehousing Appliances." *Business Intelligence Journal*, Vol. 14, No. 1, pp. 40–48.

Edwards, M. (2003, Fall). "2003 Best Practices Awards Winners: Innovators in Business Intelligence and Data Warehousing." *Business Intelligence Journal*, Vol. 8, No.4.

"Egg's Customer Data Warehouse Hits the Mark." (2005, October). *DM Review*, Vol. 15, No. 10, pp. 24–28.

Elson, R., and R. LeClerc. (2005). "Security and Privacy Concerns in the Data Warehouse Environment." *Business Intelligence Journal*, Vol. 10, No. 3.

Ericson, J. (2006, March). "Real-Time Realities." *BI Review*.

Furtado, P. (2009). "A Survey of Parallel and Distributed Data Warehouses." *International Journal of Data Warehousing and Mining*, Vol. 5, No. 2, pp. 57–78.

Golfarelli, M., and Rizzi, S. (2009). *Data Warehouse Design: Modern Principles and Methodologies*. San Francisco: McGraw-Hill Osborne Media.

Gonzales, M. (2005, Quarter 1). "Active Data Warehouses Are Just One Approach for Combining Strategic and Technical Data." *DB2 Magazine*.

Hall, M. (2002, April 15). "Seeding for Data Growth." *Computerworld*, Vol. 36, No. 16.

Hicks, M. (2001, November 26). "Getting Pricing Just Right." *eWeek*, Vol. 18, No. 46.

Hoffer, J. A., M. B. Prescott, and F. R. McFadden. (2007). *Modern Database Management*, 8th ed. Upper Saddle River, NJ: Prentice Hall.

Hwang, M., and H. Xu. (2005, Fall). "A Survey of Data Warehousing Success Issues." *Business Intelligence Journal*, Vol. 10, No. 4.

IBM. (2009). *50 Tb Data Warehouse Benchmark on IBM System Z*. Armonk, NY: IBM Redbooks.

Imhoff, C. (2001, May). "Power Up Your Enterprise Portal." *E-Business Advise*.

Inmon, W. H. (2005). *Building the Data Warehouse*. 4th ed. New York: Wiley.

Inmon, W. H. (2006, January). "Information Management: How Do You Tune a Data Warehouse?" *DM Review*, Vol. 16, No. 1.

Jukic, N., and C. Lang. (2004, Summer). "Using Offshore Resources to Develop and Support Data Warehousing Applications." *Business Intelligence Journal*, Vol. 9, No. 3.

Kalido. "BP Lubricants Achieves BIGS, Key IT Solutions." **keyitsolutions.com/asp/rptdetails/report/95/cat/1175/** (accessed August 2009).

Kalido. "BP Lubricants Achieves BIGS Success." **kalido.com/collateral/Documents/English-US/CS-BP%20BIGS.pdf** (accessed August 2009).

Karacsony, K. (2006, January). "ETL Is a Symptom of the Problem, Not the Solution." *DM Review*, Vol. 16, No. 1.

Kassam, S. (2002, April 16). "Freedom of Information." *Intelligent Enterprise*, Vol. 5, No. 7.

Kay, R. (2005, September 19). "EII." *Computerworld*, Vol. 39, No. 38.

Kelly, C. (2001, June 14). "Calculating Data Warehousing ROI." *SearchSQLServer.com Tips*.

Malykhina, E. (2003, January 3). "The Real-Time Imperative." *InformationWeek*, Issue 1020.

Manglik, A., and V. Mehra. (2005, Winter). "Extending Enterprise BI Capabilities: New Patterns for Data Integration." *Business Intelligence Journal*, Vol. 10, No. 1.

Martins, C. (2005, December 13). "HP to Consolidate Data Marts into Single Warehouse." *Computerworld*.

Matney, D. (2003, Spring). "End-User Support Strategy." *Business Intelligence Journal*, Vol. 8, No. 2.

McCloskey, D. W. (2002). *Choosing Vendors and Products to Maximize Data Warehousing Success*. New York: Auerbach Publications.

Mehra, V. (2005, Summer). "Building a Metadata-Driven Enterprise: A Holistic Approach." *Business Intelligence Journal*, Vol. 10, No. 3.

Moseley, M. (2009). "Eliminating Data Warehouse Pressures with Master Data Services and SOA." *Business Intelligence Journal*, Vol. 14, No. 2, pp. 33–43.

Murtaza, A. (1998, Fall). "A Framework for Developing Enterprise Data Warehouses." *Information Systems Management*, Vol. 15, No. 4.

Nash, K. S. (2002, July). "Chemical Reaction." *Baseline*.

Orovic, V. (2003, June). "To Do & Not to Do." *eAI Journal*.

Parzinger, M. J., and M. N. Frolick. (2001, July). "Creating Competitive Advantage Through Data Warehousing." *Information Strategy*, Vol. 17, No. 4.

Peterson, T. (2003, April 21). "Getting Real About Real Time." *Computerworld*, Vol. 37, No. 16.

Raden, N. (2003, June 30). "Real Time: Get Real, Part II." *Intelligent Enterprise*.

Reeves, L. (2009). *Manager's Guide to Data Warehousing*. Hoboken, NJ: Wiley.

Romero, O., and A. Abelló. (2009). "A Survey of Multidimensional Modeling Methodologies." *International Journal of Data Warehousing and Mining*, Vol. 5, No. 2, pp. 1–24.

Rosenberg, A. (2006, Quarter 1). "Improving Query Performance in Data Warehouses." *Business Intelligence Journal*, Vol. 11, No. 1.

Sammon, D., and P. Finnegan. (2000, Fall). "The Ten Commandments of Data Warehousing." *Database for Advances in Information Systems*, Vol. 31, No. 4.

Sapir, D. (2005, May). "Data Integration: A Tutorial." *DM Review*, Vol. 15, No. 5.

Saunders, T. (2009). "Cooking Up a Data Warehouse." *Business Intelligence Journal*, Vol. 14, No. 2, pp. 16–22.

Schwartz, K. D. "Decisions at the Touch of a Button." *Teradata Magazine*. **teradata.com/t/page/117774/index.html** (accessed June 2009).

Schwartz, K. D. (2004, March). "Decisions at the Touch of a Button." *DSS Resources*, pp. 28–31. **dssresources.com/cases/coca-colajapan/index.html** (accessed April 2006).

Sen, A. (2004, April). "Metadata Management: Past, Present and Future." *Decision Support Systems*, Vol. 37, No. 1.

Sen, A., and P. Sinha (2005). "A Comparison of Data Warehousing Methodologies." *Communications of the ACM*, Vol. 48, No. 3.

Solomon, M. (2005, Winter). "Ensuring a Successful Data Warehouse Initiative." *Information Systems Management Journal*.

Songini, M. L. (2004, February 2). "ETL Quickstudy." *Computerworld*, Vol. 38, No. 5.

Sun Microsystems. (2005, September 19). "Egg Banks on Sun to Hit the Mark with Customers." **sun.com/smi/Press/sunflash/2005-09/sunflash.20050919.1.xml** (accessed April 2006; no longer available online).

Tannenbaum, A. (2002, Spring). "Identifying Meta Data Requirements." *Journal of Data Warehousing*, Vol. 7, No. 2.

Tennant, R. (2002, May 15). "The Importance of Being Granular." *Library Journal*, Vol. 127, No. 9.

Teradata Corp. "A Large US-based Insurance Company Masters Its Finance Data." **teradata.com/t/WorkArea/DownloadAsset.aspx?id=4858** (accessed July 2009).

Teradata Corp. "Active Data Warehousing." **teradata.com/t/page/87127/index.html** (accessed April 2006).

Teradata Corp. "Coca-Cola Japan Puts the Fizz Back in Vending Machine Sales." **teradata.com/t/page/118866/index.html** (accessed June 2009).

Teradata. "Enterprise Data Warehouse Delivers Cost Savings and Process Efficiencies." **teradata.com/t/resources/case-studies/NCR-Corporation-eb4455/** (accessed June 2009).

Terr, S. (2004, February). "Real-Time Data Warehousing: Hardware and Software." *DM Review*, Vol. 14, No. 2.

Thornton, M. (2002, March 18). "What About Security? The Most Common, but Unwarranted, Objection to Hosted Data Warehouses." *DM Review*, Vol. 12, No. 3, pp. 30–43.

Thornton, M., and M. Lampa. (2002). "Hosted Data Warehouse." *Journal of Data Warehousing*, Vol. 7, No. 2, pp. 27–34.

Turban, E., D. Leidner, E. McLean, and J. Wetherbe. (2006). *Information Technology for Management*, 5th ed. New York: Wiley.

Vaduva, A., and T. Vetterli. (2001, September). "Metadata Management for Data Warehousing: An Overview." *International Journal of Cooperative Information Systems*, Vol. 10, No. 3.

Van den Hoven, J. (1998). "Data Marts: Plan Big, Build Small." *Information Systems Management*, Vol. 15, No. 1.

Watson, H., J. Gerard, L. Gonzalez, M. Haywood, and D. Fenton. (1999). "Data Warehouse Failures: Case Studies and Findings." *Journal of Data Warehousing*, Vol. 4, No. 1.

Watson, H. J. (2002). "Recent Developments in Data Warehousing." *Communications of the ACM*, Vol. 8, No. 1.

Watson, H. J., D. L. Goodhue, and B. H. Wixom. (2002). "The Benefits of Data Warehousing: Why Some Organizations Realize Exceptional Payoffs." *Information & Management*, Vol. 39.

Weir, R. (2002, Winter). "Best Practices for Implementing a Data Warehouse." *Journal of Data Warehousing*, Vol. 7, No. 1.

Wilk, L. (2003, Spring). "Data Warehousing and Real-Time Computing." *Business Intelligence Journal*, Vol. 8, No. 2.

Wixom, B., and H. Watson. (2001, March). "An Empirical Investigation of the Factors Affecting Data Warehousing Success." *MIS Quarterly*, Vol. 25, No. 1.

Wrembel, R. (2009). "A Survey of Managing the Evolution of Data Warehouses." *International Journal of Data Warehousing and Mining*, Vol. 5, No. 2, pp. 24–56.

ZD Net UK. "Sun Case Study: Egg's Customer Data Warehouse." **whitepapers.zdnet.co.uk/0,39025945,60159401p-39000449q,00.htm** (accessed June 2009).

Zhao, X. (2005, October 7). "Meta Data Management Maturity Model." *DM Direct Newsletter*.