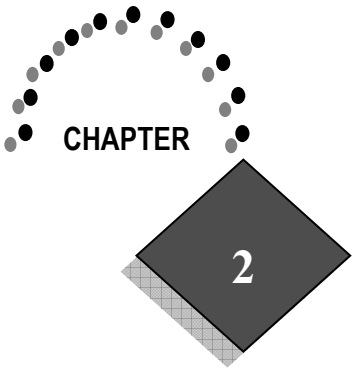

Data Warehousing

Learning Objectives for Chapter 2

- Understand the basic definitions and concepts of data warehouses
- Understand data warehousing architectures
- Describe the processes used in developing and managing data warehouses
- Explain data warehousing operations
- Explain the role of data warehouses in decision support
- Explain data integration and the extraction, transformation, and load (ETL) processes
- Describe real-time (active) data warehousing
- Understand data warehouse administration and security issues

CHAPTER OVERVIEW

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:

CHAPTER OUTLINE

2.1 OPENING VIGNETTE: ISLE OF CAPRI CASINOS IS WINNING WITH ENTERPRISE DATA WAREHOUSE

- Questions for the Opening Vignette

A. WHAT WE CAN LEARN FROM THIS VIGNETTE

2.2 DATA WAREHOUSING DEFINITIONS AND CONCEPTS

A. WHAT IS A DATA WAREHOUSE?
B. A HISTORICAL PERSPECTIVE TO DATA WAREHOUSING
C. CHARACTERISTICS OF DATA WAREHOUSING
D. DATA MARTS
E. OPERATIONAL DATA STORES
F. ENTERPRISE DATA WAREHOUSES (EDW)
- Application Case 2.1: A Better Data Plan: Well-Established TELCOs Leverage Data Warehousing and Analytics to Stay on Top in a Competitive Industry

G. METADATA
- Section 2.2 Review Questions

2.3 DATA WAREHOUSING PROCESS OVERVIEW

- Application Case 2.2: Data Warehousing Helps MultiCare Save More Lives
- Section 2.3 Review Questions

2.4 DATA WAREHOUSING ARCHITECTURES

A. ALTERNATIVE DATA WAREHOUSING ARCHITECTURES
B. WHICH ARCHITECTURE IS THE BEST?
- Section 2.4 Review Questions

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

A. DATA INTEGRATION
- Application Case 2.3: BP Lubricants Achieves BIGS Success

B. EXTRACTION, TRANSFORMATION, AND LOAD
- Section 2.5 Review Questions

2.6 DATA WAREHOUSE DEVELOPMENT

- Application Case 2.4: Things Go Better with Coke's Data Warehouse

A. DATA WAREHOUSE DEVELOPMENT APPROACHES
1. The Inmon Model: The EDW Approach
2. The Kimball Model: The Data Mart Approach
3. Which Model Is Best?
- Application Case 2.5: Starwood Hotels & Resorts Manages Hotel Profitability with Data Warehousing

B. ADDITIONAL DATA WAREHOUSE DEVELOPMENT CONSIDERATIONS
- Technology Insights 2.1: Hosted Data Warehouses

C. REPRESENTATION OF DATA IN DATA WAREHOUSE
D. ANALYSIS OF DATA IN DATA WAREHOUSE
E. OLAP VERSUS OLTP
F. OLAP OPERATIONS
1. Variations of OLAP
- Technology Insights 2.2: Hands-On Data Warehousing with MicroStrategy
- Section 2.6 Review Questions

2.7 DATA WAREHOUSING IMPLEMENTATION ISSUES

- Application Case 2.6: EDW Helps Connect State Agencies in Michigan

A. MASSIVE DATA WAREHOUSES AND SCALABILITY

- Section 2.7 Review Questions

2.8 REAL-TIME DATA WAREHOUSING

- Application Case 2.7: Egg Plc Fries the Competition in Near Real Time
- Technology Insights 2.3: The Real-Time Realities of Active Data Warehousing
- Section 2.8 Review Questions

2.9 DATA WAREHOUSE ADMINISTRATION, SECURITY ISSUES, AND FUTURE TRENDS

- Technology Insights 2.4: Ambeo Delivers Proven Data-Access Auditing Solution

A. THE FUTURE OF DATA WAREHOUSING

- Section 2.9 Review Questions

2.10 RESOURCES, LINKS, AND THE TERADATA UNIVERSITY NETWORK CONNECTION

A. RESOURCES AND LINKS
B. CASES
C. VENDORS, PRODUCTS, AND DEMOS
D. PERIODICALS
E. ADDITIONAL REFERENCES
F. THE TERADATA UNIVERSITY NETWORK (TUN) CONNECTION

Chapter Highlights
Key Terms
Questions for Discussion
Exercises

Teradata University and Other Hands-On Exercises
Team Assignments and Role-Playing Projects
Internet Exercises

- End of Chapter Application Case: Continental Airlines Flies High with Its Real-Time Data Warehouse

- Questions for the Case

References

TEACHING TIPS/ADDITIONAL INFORMATION

Expect to spend some time on this chapter, as understanding its content is crucial and many of the concepts are not intuitive to students whose major computer exposure has been at the personal level. You should prepare yourself with some technical examples of how things work, such as screen shots from some of the tools mentioned in the chapter, as most of the technical discussion here is at a more conceptual level. The business discussion is tangible, but the technical part is less so. It will be up to you to connect it to the students.

Section 2.2 defines several fundamental concepts. Students must understand that these are not definitions for the sake of definitions (every textbook has some of those, this one is no exception) but are important to anyone working in the field. The characteristics of a data warehouse in its extended definition, for example, are given not because some researcher thought they'd be theoretically nice, but, more specifically, because they've turned out to be important in practice. So, it's important to really know what each of them is about. (The concept of an operational data store should already be familiar to students, though perhaps not by that name. They should know what *metadata* are from their use with general databases, too.)

ANSWERS TO END OF SECTION REVIEW QUESTIONS

Section 2.1 Review Questions

1. Why is it important for Isle to have an EDW?

In the gaming industry, companies distinguish themselves based on their customer relationships. An enterprise data warehouse (EDW) gathers and provides the data needed to tell Isle of Capri what customers respond to, so the casinos can adapt their offerings. The information provided by the EDW lets Isle deepen its understanding of customers, so it can efficiently give them more of the kinds of entertainment they are looking for.

2. What were the business challenges or opportunities that Isle was facing?

Isle of Capri Casinos is one of the largest publicly traded gaming companies in the United States, but the gaming business is competitive. Other casinos offer essentially the same games, so Isle must find ways to make its entertainment and hospitality atmosphere one that exceeds customer expectations. Before Isle implemented the EDW, casino managers had to wait to review monthly data until the second week of the following month. The time lag made it difficult for casinos

to identify what actions were appealing to customers in time to respond. Adding to these business challenges, Hurricane Katrina set back initial efforts to set up a data warehouse at the southeastern company.

3. What was the process Isle followed to realize EDW? Comment on the potential challenges Isle might have had going through the process of EDW development.

Isle of Capri brought in two expert suppliers. Teradata provided the core solution; IBM Cognos provided expertise in business intelligence. Isle hired a management team that understood how EDW could support decision making at Isle. That team would be able to help Isle's managers with the challenge of understanding how they can frame queries and follow-up questions to figure out ways to improve the business. Most likely, the potential of using detailed, up-to-the-minute data would be unfamiliar to many of these managers.

4. What were the benefits of implementing an EDW at Isle? Can you think of other potential benefits that were not listed in the case?

The implementation of EDW brought several benefits related to the timeliness and detail of the data that became available. Instead of five week-old monthly reports, managers can now study a variety of daily, weekly, and monthly reports. The reports segment data by particular properties and customer groups, so managers can zero in on particular problems and successes, easily making comparisons among properties. Managers can submit queries about data sets and receive information within minutes. In this way, managers can find out how particular promotions are affecting customer behavior at particular casinos. The EDW also connects data about casino activity with data about customers' use of hotels and efforts by Isle's hosts. This, too, helps the company target promotions and offer customers incentives they value. Even decisions as detailed as where to locate slot machines can be adjusted to boost profits based on data from the EDW.

5. Why do you think large enterprises like Isle in the gaming industry can succeed without having a capable data warehouse/business intelligence infrastructure?

In the past, businesses in the gaming industry could succeed without a capable data warehouse/business intelligence infrastructure because their managers knew as much about customers as their competitors' management knew. They were all testing ideas for promotion or entertainment and responding to customer behavior at about the same pace. However, when one company such as Isle begins to respond to daily, property-by-property data, choosing the marketing and entertainment options that deliver exactly what most profitably lures customers, competitors without that capability will soon begin to suffer.

Section 2.2 Review Questions

1. What is a data warehouse?

A data warehouse is defined in this section as "a pool of data produced to support decision making." This focuses on the essentials, leaving out characteristics that may vary from one DW to another but are not essential to the basic concept.

The same paragraph gives another definition: "a subject-oriented, integrated, time-variant, nonvolatile collection of data in support of management's decision-making process." This definition adds more specifics, but in every case appropriately: it is hard, if not impossible, to conceive of a data warehouse that would not be subject-oriented, integrated, etc.

2. How does a data warehouse differ from a database?

Technically a data warehouse *is* a database, albeit with certain characteristics to facilitate its role in decision support. Specifically, however, it is (see previous question) an "integrated, time-variant, nonvolatile, subject-oriented repository of detail and summary data used for decision support and business analytics within an organization." These characteristics, which are discussed further in the section just after the definition, are not necessarily true of databases in general—though each could apply individually to a given one.

As a practical matter most databases are highly normalized, in part to avoid update anomalies. Data warehouses are highly denormalized for performance reasons. This is acceptable because their content is never updated, just added to. Historical data are static.

3. What is an ODS?

Operational Data Store is the database from which a business operates on an on-going basis.

4. Differentiate among a data mart, an ODS, and an EDW.

An ODS (Operational Data Store) is the database from which a business operates on an ongoing basis.

Both an EDW and a data mart are data warehouses. An EDW (Enterprise Data Warehouse) is an all-encompassing DW that covers all subject areas of interest to the entire organization. A data mart is a smaller DW designed around one problem, organizational function, topic, or other suitable focus area.

5. Explain the importance of metadata.

Metadata, "data about data," are the means through which applications and users access the content of a data warehouse, through which its security is managed, and through which organizational management manages, in the true sense of the

word, its information assets. Most database management systems would be unable to function without at least some metadata. Indeed, the use of metadata, which enable data access through names and logical relationships rather than physical locations, is fundamental to the very concept of a DBMS.

Metadata are essential to any database, not just a data warehouse. (See answer to Review Question 2 of this section above.)

Section 2.3 Review Questions

1. Describe the data warehousing process.

The data warehousing process consists of the following steps:

1. Data are imported from various internal and external sources
2. Data are cleansed and organized consistently with the organization's needs
3. a. Data are loaded into the enterprise data warehouse, or
   b. Data are loaded into data marts.
4. a. If desired, data marts are created as subsets of the EDW, or
   b. The data marts are consolidated into the EDW
5. Analyses are performed as needed

2. Describe the major components of a data warehouse.

- ***Data sources.*** Data are sourced from operational systems and possibly from external data sources.
- ***Data extraction and transformation.*** Data are extracted and properly transformed 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.*** This is the EDW that supports decision analysis by providing relevant summarized and detailed information.
- ***Metadata.*** Metadata are maintained for access by IT personnel and users. Metadata include rules for organizing data summaries that are easy to index and search.
- ***Middleware tools.*** Middleware tools enable access to the data warehouse from a variety of front-end applications.

3. Identify and discuss the role of middleware tools.

Middleware tools enable access to the data warehouse. Power users such as analysts may write their own SQL queries. Others may access data through a managed query environment. 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. All these have their own data access requirements. Those may not match with how a given data warehouse must be accessed. Middleware translates between the two.

Section 2.4 Review Questions

1. What are the key similarities and differences between a two-tiered architecture and a three-tiered architecture?

Both provide the same user visibility through a client system that accesses a DSS/BI application remotely. The difference is behind the scenes and is invisible to the user: in a two-tiered architecture, the application and data warehouse reside on the same machine; in a three-tiered architecture, they are on separate machines.

2. How has the Web influenced data warehouse design?

Primarily by making Web-based data warehousing possible.

3. List the alternative data warehousing architectures discussed in this section.

- Independent data marts architecture
- Data mart bus architecture with linked dimensional data marts
- Hub-and-spoke architecture (corporate information factory)
- Centralized data warehouse architecture
- Federated architecture

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

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

(This list from the text, while clearly intended by the authors as the answer to this review question, does not explicitly say that these *are* the ten most important factors. Students may choose others.)

5. Which data warehousing architecture is the best? Why?

See Table 2.1 Average Assessment Scores for the Success of the Architectures. What is interesting 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.

Section 2.5 Review Questions

1. Describe data integration.

Data integration is an umbrella term that covers three processes that combine to move data from multiple sources into a data warehouse: accessing the data, combining different views of the data, and capturing changes to the data.

2. Describe the three steps of the ETL process.

Extraction: selecting data from one or more sources and reading the selected data.

Transformation: converting data from their original form to whatever form the DW needs. This step often also includes cleansing of the data to remove as many errors as possible.

Load: putting the converted (transformed) data into the DW.

3. Why is the ETL process so important for data warehousing efforts?

Since ETL is the process through which data are loaded into a data warehouse, a DW could not exist without it. The ETL process also contributes to the quality of the data in a DW.

Section 2.6 Review Questions

1. List the benefits of data warehouses.

Direct benefits include:

- Allowing end users to perform extensive analysis in numerous ways.
- A consolidated view of corporate data (i.e., a single version of the truth).
- Better and more timely information. A data warehouse permits information processing to be offloaded from costly operational systems onto low-cost servers; therefore, end-user information requests can be processed more quickly.

- Enhanced system performance. A data warehouse frees production processing because some operational system reporting requirements are moved to DSS.
- Simplification of data access.

Indirect benefits arise when end users take advantage of these direct benefits.

2. List several criteria for selecting a data warehouse vendor, and describe why they are important.

Six important criteria are: financial strength, ERP linkages, qualified consultants, market share, industry experience, and established partnerships. These are important to indicate that a vendor is likely to be in business for the long term, to have the support capabilities its customers need, and to provide products that interoperate with other products the potential user has or may obtain.

One could add others, such as product functionality (Does it do what we need?), vendor strategic vision (Does their direction make sense for our future plans and/or is it consistent with industry trends?) and quality of customer references (What do their existing customers think of them?).

3. What is OLAP and how does it differ from OLTP?

Data stored in a data warehouse can be analyzed using techniques referred to as OLAP. OLAP is one of the most commonly used data analysis techniques in data warehouses. OLAP is an approach to quickly answer ad hoc questions that require data analysis.

OLTP is concerned with the capture and storage of data. OLAP is concerned with the analysis of that data.

4. What is a cube? What do drill down, roll up, and slice and dice mean?

The main operational structure in OLAP is based on a concept called cube. A **cube** in OLAP is a multidimensional data structure (actual or virtual) that allows fast analysis of data.

Using OLAP, an analyst can navigate through the database and screen for a particular subset of the data (and its progression over time) by changing the data's orientations and defining analytical calculations. These types of user-initiated navigation of data through the specification of slices (via rotations) and drill down/up (via aggregation and disaggregation) are sometimes called "slice and dice." Commonly used OLAP operations include slice and dice, drill down, roll up, and pivot.

- ***Slice:*** A slice is a subset of a multidimensional array (usually a two-dimensional representation) corresponding to a single value set for one (or more) of the dimensions not in the subset.

- ***Dice:*** The dice operation is a slice on more than two dimensions of a data cube.

- ***Drill Down/Up:*** Drilling down or up is a specific OLAP technique whereby the user navigates among levels of data ranging from the most summarized (up) to the most detailed (down).

5. What are ROLAP, MOLAP, and HOLAP? How do they differ from OLAP?

ROLAP stands for Relational Online Analytical Processing. ROLAP is an alternative to the MOLAP (Multidimensional OLAP) technology. While both ROLAP and MOLAP analytic tools are designed to allow analysis of data through the use of a multidimensional data model, ROLAP differs in that it does not require the pre-computation and storage of information. Instead, ROLAP tools access the data in a relational database and generate SQL queries to calculate information at the appropriate level when an end user requests it. MOLAP is an alternative to the ROLAP technology. MOLAP differs from ROLAP significantly in that it requires the pre-computation and storage of information in the cube—the operation known as preprocessing. MOLAP stores this data in an optimized multidimensional array storage, rather than in a relational database.

HOLAP (Hybrid Online Analytical Processing) is a combination of ROLAP and MOLAP. HOLAP allows storing part of the data in a MOLAP store and another part of the data in a ROLAP store. The degree of control that the cube designer has over this partitioning varies from product to product.

All of these are variations of OLAP.

Section 2.7 Review Questions

1. What are the major DW implementation tasks that can be performed in parallel?

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

Establishment of service-level agreements and data-refresh requirements

Identification of data sources and their governance policies

Data quality planning

Data model design

ETL tool selection

Relational database software and platform selection

Data transport

Data conversion

Reconciliation process

Purge and archive planning

End-user support

2. List and discuss the most pronounced DW implementation guidelines.

- Senior management must support development of the data warehouse. The DW needs a project champion at a high position in the organization chart. Benefits of a DW project may be difficult to measure, so management support makes it more likely the project will receive funding.
- Web-based data warehouses may need special security requirements. These ensure that only authorized users have access to the data.
- Users should participate in the development process. Their participation is essential for data modeling and access modeling. User participation ensures that the DW includes the needed data and that decision makers can retrieve the data they need.
- DW implementation requires certain skills from members of the development team: in-depth knowledge of database technology and the development tools used.

3. When developing a successful data warehouse, what are the most important risks and issues to consider and potentially avoid?

- Starting with the wrong sponsorship chain
- Setting expectations that you cannot meet
- Engaging in politically naive behavior
- Loading the data warehouse with information just because it is available

- Believing that data warehousing database design is the same as transactional database design
- Choosing a data warehouse manager who is technology oriented rather than user oriented
- Focusing on traditional internal record-oriented data and ignoring the value of external data and of text, images, and, perhaps, sound and video
- Delivering data with overlapping and confusing definitions
- Believing promises of performance, capacity, and scalability
- Believing that your problems are over when the data warehouse is up and running
- Focusing on ad hoc data mining and periodic reporting instead of alerts

4. What is scalability? How does it apply to DW?

Scalability refers to the degree to which a system can adjust to changes in demand without major additional changes or investments. DW scalability issues 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.

Section 2.8 Review Questions

1. What is an RDW?

A *real-time data warehouse*, in which decision-making data are updated on an ongoing basis as business transactions occur; same as an *active data warehouse* (ADW).

2. List the benefits of an RDW.

The RDW extends the benefits of data warehousing, in general, down into tactical, and perhaps operational, decision making. It empowers people who interact directly with customers and suppliers by providing them with information to make decisions. It can then be extended to customers and suppliers themselves, thus affecting almost all aspects of customer service, SCM, logistics, and beyond.

It can also facilitate e-business activities, as when sales outlets such as overstock.com (cited in the text) use historical data to price new close-outs.

3. What are the major differences between a traditional data warehouse and an RDW?

1. A traditional data warehouse (TDW) is used for strategic decisions (and sometimes tactical); an RDW for strategic and tactical (sometimes operational) ones.
2. The results of using a TDW can be hard to measure; results of using an RDW are measured by operational data.
3. Acceptable TDW refresh rates range from daily to monthly; RDW data must be up to the minute.
4. TDW summaries are often appropriate; RDWs must supply detailed data.
5. Small user community at upper organizational levels means a TDW supports few concurrent users; an RDW must support many, perhaps over a thousand.
6. TDWs typically use restrictive reporting to confirm or check patterns, often predefined summary tables; RDWs need flexible, ad hoc reporting.
7. TDW user community generally consists of power users, knowledge workers, managers, other internal users; RDWs are used by operational staff, call centers, perhaps external users.

4. List some of the drivers for RDW.

- A business often cannot afford to wait a whole day for its operational data to load into the data warehouse for analysis.
- Traditional 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, retaining 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.

Section 2.9 Review Questions

1. What steps can an organization take to ensure the security and confidentiality of customer data in its data warehouse?

Effective security in a data warehouse should focus on four main areas:

Step 1. Establishing effective corporate and security policies and procedures. An effective security policy should start at the top and be communicated to everyone in the organization.
Step 2. Implementing logical security procedures and techniques to restrict access. This includes user authentication, access controls, and encryption.
Step 3. Limiting physical access to the data center environment.
Step 4. Establishing an effective internal control review process for security and privacy.

2. What skills should a DWA possess? Why?

- Familiarity with high-performance hardware, software, and networking technologies, since the data warehouse is based on those
- Solid business insight, to understand the purpose of the DW and its business justification
- Familiarity with business decision-making processes to understand how the DW will be used
- Excellent communication skills, to communicate with the rest of the organization

3. What recent technologies may shape the future of data warehousing? Why?

Following are some of the recently popularized concepts and technologies that will play a significant role in defining the future of data warehousing.

Sourcing: Acquisition of data from diverse and dispersed sources

- Web, social media, and Big Data
- Open source software
- SaaS (software as a service) "The Extended ASP Model"
- Cloud computing

Infrastructure: Architectural—hardware and software—enhancements

- Columnar (a new way to score and access data in the database)
- Real-time data warehousing
- Data warehouse appliances (all-in-one solutions to DW)
- Data management technologies and practices

- In-database processing technology (putting the algorithms where the data is)
- In-memory storage technology (moving the data in the memory for faster processing)
- New database management systems
- Advanced analytics

As the world of business becomes more global and complex, the need for business intelligence and data warehousing tools also becomes more prominent. The fast improving information technology tools and techniques seem to be moving in the right direction to address the needs of the future business intelligence systems.

Section 2.10 Review Questions

(This section has no review questions.)

ANSWERS TO APPLICATION CASE QUESTIONS FOR DISCUSSION

Application Case 2.1: A Better Data Plan: Well-Established TELCOs Leverage Data Warehousing and Analytics to Stay on Top in a Competitive Industry

1. What are the main challenges for TELCOs?

To stay competitive, TELCOs must continuously refine everything from customer service to plan pricing. The major challenges faced by both entrenched and new companies in this industry include: retaining customers, decreasing costs, fine-tuning pricing models, improving customer satisfaction, acquiring new customers, and understanding the role of social media in customer loyalty.

2. How can data warehousing and data analytics help TELCOs in overcoming their challenges?

Highly targeted data analytics play an ever more critical role in helping carriers secure or improve their standing in an increasingly competitive marketplace. Argentina's Telefónica de Argentina used analytics for its "traceability project," which tracked the factors involved in customer churn, a big problem among phone service carriers. France's Bouygues Telecom used BI technologies to facilitate cost reduction through automation via its Teradata-based marketing operations management system, which automates marketing/communications collateral production. Pakistan's Mobilink uses BI to help acquire customers and grow their subscriber network, largely aided by social networking.

3. Why do you think TELCOs are well suited to take full advantage of data analytics?

TELCOs control the telecommunications infrastructure, and acquire much usage data as a result. They have the technical expertise to create, deploy, and refine plans to address their business challenges. The industry and mobile technology have expanded and improved over the years, which provides a strong foundation on which to build intelligent solutions. The data analytics solutions that have been created to meet these challenges have also improved drastically over the past few years, placing TELCOs in a good position to capitalize on their technological advantages.

Application Case 2.2: Data Warehousing Helps Multi-Care Save More Lives

1. What do you think is the role of data warehousing in healthcare systems?

This question seeks opinions, so accept any reasonable answer. The case focuses on the role of data warehousing in improving patient care, which can literally save lives. Data on patient needs, hospital practices, and patient outcomes can help health care systems identify the practices that lead to the best patient outcomes. Of course, organizations also can use data warehousing to improve efficiency and profits.

2. How did MultiCare use data warehousing to improve health outcomes?

MultiCare organized and simplified data from multiple sources across the continuum of care, so the Adaptive Data Warehouse became a single source that unified decision makers across functions. It addressed the serious problem of septicemia by defining a standard of care for septic patients, identify patients with conditions signaling they were likely to take a sudden turn for the worse, and efficiently implement the standard of care. Building on the first-year success of the effort, MultiCare intended to expand the application of data warehousing in similar ways to treat heart failure, manage the performance of the emergency department, and improve inpatient throughput.

Application Case 2.3: BP Lubricants Achieve BIGS Success

1. What is BIGS?

BIGS is the Business Intelligence and Global Standards program, a strategic initiative for management information and business intelligence. Its purpose is to deliver globally consistent and transparent management information across functions.

2. What were the challenges, the proposed solution, and the obtained results with BIGS?

The challenge facing BP Lubricants was that it had been involved in merger activity, which meant it had data held in disparate source systems. To be more agile and prepare for growth, it wanted a unified system. The proposed solution was Kalido, an adaptive enterprise data warehousing solution. The system integrates and stores information from multiple source systems to provide consolidated views for marketing, sales, and finance. The obtained results are that BIGS helps the business identify a multitude of business opportunities to improve profits and lower costs. Typical benefits include improved consistency and transparency of business data, faster and more flexible reporting, and greater ability to respond intelligently to new business opportunities.

Application Case 2.4: Things Go Better with Coke's Data Warehouse

1. How did Coca-Cola in Japan use data warehousing to improve its business processes?

The Hokuriku Coca-Cola Bottling Co. (HCCBC) hired Teradata Corp. to implement a data warehouse and analytical software to collect historical data and near-real-time data wirelessly transmitted from each vending machine. The data tells the company what is selling, and it sends alerts when a product sells out, a customer is short-changed, or a machine is malfunctioning.

2. What were the results of their enterprise active data warehouse implementation?

In a pilot test of the system, the accurate demand forecasts and rapid problem identification led to an immediate 10 percent increase in sales. More accurate machine servicing contributed to a 42 percent decline in costs while improving the efficiency of the salespeople who service the machines. Based on the results, HCCBC intended to expand the use of the active data warehouse to cover all the vending machines in Japan.

Application Case 2.5: Starwood Hotels & Resorts Manages Hotel Profitability with Data Warehousing

1. How big and complex are the business operations of Starwood Hotels & Resorts?

The business operations are very large and complex. Starwood is one of the world's leading hotel and leisure companies, with 1,112 properties in nearly 100 countries. The company has 154,000 employees at properties that it owns and manages. It also operates a customer loyalty program.

2. How did Starwood Hotels & Resorts use data warehousing for better profitability?

With data warehousing, Starwood Hotels can prepare operational reports in just four to six hours. With real-time feeds, transactions can be posted immediately to the data warehouse, and users can access the changes in just 5 to 10 minutes, or 288 times faster than before. The accelerated access to data improves profits because hotel managers can make sure rooms are available for premier customers and improve the ability to manage room occupancy rates. Marketing campaigns also have become more efficient as managers analyze results in days or weeks instead of months.

3. What were the challenge, the proposed solution, and the obtained results?

The challenge was that the company had grown significantly, especially in the Asia/Pacific region. This greatly increased the need for business critical information about the hotels and their customers. Obtaining reports from the central reservation system could take up to 18 hours, and with delays in obtaining data, managers had difficulty adjusting rates to increase revenues. The proposed solution was an Oracle Exadata system that provides near-real-time information. As a result of Starwood implementing the system, managers have accelerated access to data so they can do same-day or next-day analyses, as well as additional reporting tools. The company now operates more efficiently and profitably, and customer service has improved.

Application Case 2.6: EDW Helps Connect State Agencies in Michigan

1. Why would a state invest in a large and expensive IT infrastructure (such as an EDW)?

Like a business, a state government can operate more efficiently and make better decisions if it has access to current data. State officials can use the EDW to improve efficiency and better serve the state's residents.

2. What are the size and complexity of EDW used by state agencies in Michigan?

Michigan's EDW has almost 10,000 users in five major departments, 20 agencies, and more than 100 bureaus. Just in the Department of Human Services, the EDW contributes to nearly every function, including accurate delivery of and accounting for benefits to almost 2.5 million clients receiving public assistance.

3. What were the challenges, the proposed solution, and the obtained results of the EDW?

The challenge is not directly stated, but students should be aware that state governments struggle to balance budgets and satisfy a wide variety of needs. They also should recognize that a state operates many kinds of departments, so it likely

could benefit from a unified source of information. The proposed solution was an enterprise (statewide) electronic data warehouse linking employees across departments. The obtained results include financial benefits worth $1 million per business day. Savings come from operational efficiencies, avoidance of sanctions, improved client outcomes, integrity benefits awarded to well-performing programs, and recovery of inappropriate benefits payments. The data warehouse has yielded a 15:1 cost-effectiveness ratio and improvements in the accurate delivery and accounting of benefits.

Application Case 2.7: Egg Plc Fries the Competition in Near Real Time

1. What kind of business is Egg plc in? What is the competitive landscape?

Egg plc is an online bank that provides banking, insurance, investments, and mortgages to more than 3.6 million customers through its Internet site. Although the case does not discuss the competition, it mentions that Egg was once the world's largest online bank. Students may note that the competitive landscape for online banking has changed considerably over the past few years, with most banks now offering at least some services online. Also accept any other reasonable description of the competitive landscape.

2. How did Egg plc use near-real-time data warehousing for competitive advantage?

Near-real-time data access gave employees and customers access to data. Within the company, the data helped with the construction of sales and marketing campaigns. It allowed faster decision making about specific customers and customer classes. An interesting basis for classroom discussion might be the fact that Egg did not maintain a competitive advantage for long, as it has gone out of business. Setting up a data warehouse can give the organization an initial advantage over companies slower to adopt the technology, but how might a company maintain that advantage, using analytics?

ANSWERS TO END OF CHAPTER QUESTIONS FOR DISCUSSION

1. Compare data integration and ETL. How are they related?

Data integration consists of three processes that integrate data from multiple sources into a data warehouse: accessing the data, combining different views of the data, and capturing changes to the data. It makes data available to ETL tools and, through the three processes of ETL, to the analysis tools of the data warehousing environment.

2. What is a data warehouse and what are its benefits? Why is Web accessibility important with a data warehouse?

A data warehouse can be defined as "a pool of data produced to support decision making." This focuses on the essentials, leaving out characteristics that may vary from one DW to another but are not essential to the basic concept.

The same paragraph gives another definition: "a subject-oriented, integrated, time-variant, nonvolatile collection of data in support of management's decision-making process." This definition adds more specifics, but in every case appropriately: it is hard, if not impossible, to conceive of a data warehouse that would not be subject-oriented, integrated, etc.

The benefits of a data warehouse are that it provides decision-making information, organized in a way that facilitates the types of access required for that purpose and supported by a wide range of software designed to work with it.

Web accessibility of a data warehouse is important because many analysis applications are Web-based, because users often access data over the Web (or over an intranet using the same tools) and because data from the Web may feed the DW.

3. A data mart can replace a data warehouse or complement it. Compare and discuss these options.

For a data mart to *replace* a data warehouse, it must make the DW unnecessary. This would mean that all the analyses for which the DW would be used can instead be satisfied by a DM (or perhaps a combination of several DMs). If this is so, it can be much less expensive, in terms of development and computer resources, to use multiple DMs (let alone one DM!) instead of an overall DW.

In other situations, a data mart can be used for some analyses which would in its absence use the DW, but not all of them. For those, the smaller DM is more efficient—quite possibly, enough more efficient as to justify the cost of having a DM in addition to a DW. Here the DM complements the DW.

4. Discuss the major drivers and benefits of data warehousing to end users.

Major drivers include:

- Increased competition and pace of business, leading to increased need for good decisions quickly
- Successful pioneering experiences with data warehouses, leading to their wider user acceptance

- Decreasing hardware costs, making terabyte databases with masses of historical data economically feasible for more firms
- Increased availability of software to manage a large data warehouse
- Increased availability of analysis tools making DWs potentially more useful
- Increased computer literacy of decision makers, making them more likely to use these tools

5. List the differences and/or similarities between the roles of a database administrator and a data warehouse administrator.

Since a data warehouse is a specific type of database designed for a specific application area, a data warehouse administrator has all the roles of a database administrator—plus others. One new role is advising on decision support uses of the DW, for which a DWA needs to understand decision-making processes. Beyond that, the issue is more a need for additional skills in the same roles as a DBA—e.g., understanding high-performance hardware to deal with the large size of a DW—than it is one of additional roles. (See Review Question 2, Section 5.8, for list of skills.)

6. Describe how data integration can lead to higher levels of data quality.

A question involving the word "higher" (or any other comparative, for that matter) requires asking "higher than what?" In this case, we can take it to mean "higher than we would have for the same data, but without a formal data integration process."

Without a data integration process to combine data in a planned and structured manner, data might be combined incorrectly. That could lead to misunderstood data (a measurement in meters taken as being in feet) and to inconsistent data (data from one source applying to calendar months, data from another to four-week or five-week fiscal months). These are aspects of low-quality data which can be avoided, or at least reduced, by data integration.

7. Compare the Kimball and Inmon approaches toward data warehouse development. Identify when each one is most effective.

Inmon's approach starts with an enterprise data warehouse, creating data marts as subsets if appropriate. It is most effective when there is a recognized need for an EDW, an executive "champion" of the project, and a willingness to invest in a data warehousing infrastructure before it will show results.

Kimball's approach starts with data marts, consolidating them into an EDW later if appropriate. It is most effective when it is desired to provide a "proof of

concept" implementation before embarking on a full-scale EDW project or when a well-defined area with the greatest benefits can be identified.

8. Discuss security concerns involved in building a data warehouse.

Security and privacy concerns are important in building a data warehouse:

1. Laws and regulations, in the U.S. and elsewhere, require certain safeguards on databases that contain the type of information typically found in a DW.
2. The large amount of valuable corporate data in a data warehouse can make it an attractive target.
3. The need to allow a wide variety of unplanned queries in a DW makes it impractical to restrict end user access to specific carefully constrained screens, one way to limit potential violations.

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.

Open-ended answer to the report; it is impossible to predict what the debate will bring.

A student's position on this issue is related to his/her feelings on the relationship of national economies to the global economy. It can be argued that offshoring improves the global economy while potentially harming one or more of the national economies involved—such as the student's own. U.S. students may see primarily the damage they perceive it does to their national economy (and to their own career prospects), but students in India may take a different view. The economic, political, and philosophical issues can be pursued well beyond what is practical in a DSS course.

If you feel students are too nationalistic on this issue, you can ask them if they feel the same way about a Massachusetts or California bank processing checks in Alabama to reduce labor costs. (This example uses U.S. territories, but similar issues exist in any country large enough to have regional economic differences.)

ANSWERS TO END OF CHAPTER APPLICATION CASE QUESTIONS

1. Describe the benefits of implementing the Continental Go Forward strategy.

This strategy consisted of a number of interrelated, concurrent actions. The first version of the overall strategy had the benefit of restoring Continental (CO) to profitability and giving it first-place rankings by many airline industry metrics. The second phase of the strategy led to savings of $41 million and a reduction of

$7 million in fraud in the first year alone. Its revenue increased by over $500 million in six years. A data warehouse played a critical role in the second phase.

2. Explain why it is important for an airline to use a real-time data warehouse.

It's important for an airline to use a real-time data warehouse because many airline decisions cannot be made with week-old, or even day-old, data.

An example is frequent flyer award availability on a given flight. Airlines limit these so as not to give away too many seats that would otherwise be sold. Award seat allocation is usually automated. Travelers can check availability online. American Airlines (and probably others) offers expanded award availability to the 30,000 or so people who fly at least 100,000 miles per year with them. When one of these wants an award seat that is not available online, he or she can request it by phone. The agent must decide whether or not to make it available. If the flight is selling slowly, the traveler gets the seat, even if the computer hasn't allocated it for award. If it is likely to sell out, the seat isn't offered, even to this select group. To make this decision, telephone agents (and the Yield Management staff, which agents can consult) need up-to-the-minute, or at least up-to-the-hour, information.

3. Examine the sample system output screen at http://www.teradata.com/case-studies/Continental-Airlines-Case-Study-eb4349. Describe how it can assist the user in identifying problems and opportunities.

This screen appears to show the relationship among three independent variables and one dependent variable. The independent age are the age of a caller (column), the time the person tends to call (color of rectangle) and the number of complaints by that category of caller (height of rectangle). The dependent variable is profitability, shown by the width of the rectangle. The chart shows that weekend callers, while generating neither the highest nor the lowest number of complaints, are the airline's most profitable customers. This information can be used to find out why some other groups have more complaints, to give higher priority to weekend callers, and for many other purposes.

4. Identify the major differences between the traditional data warehouse and a real-time data warehouse, as was implemented at Continental.

A traditional data warehouse moves data from operational databases to the DW on a scheduled basis, typically daily or weekly. This provides consistent data for analyses performed during one update cycle, but does not make current information available for decisions that require it. A real-time DW, as was implemented at CO, moves data into the DW on an hourly or even more frequent basis.

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

By having real-time data available through its data warehouse, CO can make decisions using up-to-date information. While data warehousing applications, which focus on long-term decisions, aren't affected much by the last hour's, day's, or even week's data, lower-level short-term decisions are. As the use of the DW is extended to these decisions and down in the organization, current data become necessary. By having real-time (or near-real-time) data in the system, CO obtains a strategic advantage by making better decisions.