SOLUTIONS MANUAL

BUSINESS INTELLIGENCE

A MANAGERIAL APPROACH

Efraim Turban | Ramesh Sharda
Jay E. Aronson | David King

CHAPTER 2 – DATA WAREHOUSING

CHAPTER 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

CHAPTER OVERVIEW

Data warehousing is at the foundation of most BI. This is *the* data warehousing chapter of the book. Later chapters will use it as they discuss DW applications such as business analytics (Chap. 3) and data mining (Chap. 4), but the guts are here. If students don't "get" the data warehouse idea here, they'll be at a big disadvantage throughout the rest of the book.

CHAPTER OUTLINE

2.1 OPENING VIGNETTE: CONTINENTAL AIRLINES FLIES HIGH WITH ITS REAL-TIME DATA WAREHOUSE

Big problems

Solution

Benefits

What we can learn from this vignette

2.2 DATA WAREHOUSING DEFINITIONS AND CONCEPTS

What is a data Warehouse?

Characteristics of data warehousing

Data marts

Operational data stores

Enterprise data warehouses (EDW)

Metadata

2.3 DATA WAREHOUSING PROCESS OVERVIEW

2.4 DATA WAREHOUSING ARCHITECTURES

Alternative architectures

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

Data integration

Extraction, transformation and load

2.6 DATA WAREHOUSE DEVELOPMENT

Data warehouse vendors

Data warehouse development approaches

The Inmon model: The EDW approach

The Kimball model: The data mart approach

Which model is best?

Additional data warehouse development considerations

Data warehouse structure: The star schema

Data warehousing implementation issues

Massive data warehouses and scalability

2.7 REAL-TIME DATA WAREHOUSING

2.8 DATA WAREHOUSE ADMINISTRATION AND SECURITY ISSUES

2.9 RESOURCES, LINKS, AND THE TERADATA UNIVERSITY NETWORK CONNECTION

Resources and links

Cases

Vendors, products and demos

Periodicals

Additional references

The Teradata University Network (TUN) Connection

TEACHING TIPS/ADDITIONAL INFORMATION

Expect to spend some time on this chapter. 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 more conceptual. 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 book 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, aren't there because some researcher thought they'd be nice, but 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.)

Section 2.3 is brief. It can be covered quickly. It's important to understand what each step does, but the details of things like ETL will be covered later in this chapter.

Section 2.4, by contrast, is eight pages long—though a lot of that is diagrams. (You might have to tell students that "MDB" in some of them stands for "Multidimensional Database." The key point here is that there are a lot of options, the correct one can only be chosen when something is known about the characteristics of the application, and that in most cases more than one will work reasonably well.

Section 2.5 discusses ETL, a process that never shows up outside a data warehouse context and that students may therefore never have heard of until now. The need for ETL arises because databases aren't all in the same format, chances are that few if any of the source databases are in the format of the DW, and even if the formats match most databases contain dirty data. This is a good place to show screen shots of ETL tools so students will really appreciate how they work.

Section 2.6, another long one, is on the development process. Students will have studied the development and implementation of information systems in general in their introductory MIS course and probably gone deeper into it in Systems Analysis and other courses. That lets you focus on *what's different about data warehouses*. For example, while training is needed for both conventional systems and data warehouses, the training needed for a DW is at a higher level: more *information* literacy than *computer* literacy. A systems analyst who is used to the correct type of training for ERP or SCM systems, and who attempts to replicate that experience for a DW user community, is courting disaster.

This section covers more than the development suggested by its title. It also discusses the internal structure of a data warehouse database. Its focus is on the *star schema*. That's as close as this book gets to how a data warehouse is actually organized. Students who've studied database management—probably most of the class in a course for MIS concentrators—will want tangible information here. You can provide this with examples showing how a simple set of historical sales data with several dimensions could be stored in an operational database and in a data warehouse organized by the star schema.

You may also wish to add a multidimensional database to this example. These were mentioned several times in Section 2.4 but are not here. While these aren't promoted for data warehouses quite as

much as they were a few years ago (the statement in the first bullet above Figure 2.5, that most DWs are built with RDBMS, is true today) there are still good products that support them, applications to which they are well suited, and a lot in use where students may find them on the job. A conceptual understanding of how they differ from a relational database organized by the star schema can be good to have.

If you laid a good foundation up to this point, Section 2.7 should go quickly. The business case for real-time DWs is well presented, and the technology introduces no new conceptual difficulties.

Section 2.8 returns to issues that students will be familiar with from a general DBMS context. The need for "establishing effective corporate security policies and procedures," for instance, is not unique to data warehouses. Therefore, you may again want to focus on what's different about them. Answers to Discussion Question 5 at the end of the chapter can lead to a good class discussion of this.

ANSWERS TO END-OF-SECTION REVIEW QUESTIONS

Section 2.1 Review 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.*

   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 their 30,000 or so most frequent flyers. When one of these "Executive Plati-

num" travelers 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 an award. If it is likely to sell out, the seat isn't offered, even to this elite group. To make this decision, telephone agents (and the Yield Management staff, which agents consult) need current information.

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.*

This screen shows the relationship among three independent and one dependent variable. The independent variables are the age are the caller's age (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 most nor the fewest 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 brings in data on an hourly or even more frequent basis.

5. *What strategic advantage can Continental derive from the real-time system as opposed 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. These decisions can affect any and all of the five competitive forces first identified by Michael Porter and probably familiar to your students from other courses.

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 is a data warehouse different 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 often 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, the information used for running the business on an on-going basis and for short-term, operational decisions.

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 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

3a. Data are loaded into the enterprise data warehouse

4a. If desired, data marts are created as subsets of the EDW

—or—

3b. Data are loaded into data marts

4b. 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. 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. 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 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.*

   - Enterprise Data Warehousing architecture
   - Data Mart architecture
   - Hub-and-Spoke Data Mart architecture
   - Enterprise Warehouse and Operational Data Store
   - Distributed Data Warehouse 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 as the answer to this review question, does not explicitly say that these are the *ten most* important factors. That allows students some leeway in choosing others.)

Section 2.5 Review Questions

1. *Describe data integration.*

Data integration is an umbrella term that covers three processes which 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.

Loading: 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 criteria listed in the text are financial strength, ERP linkages, qualified consultants, market share, industry experience, and established partnerships. These 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 (is their direction consistent with our future plans and with industry trends?) and quality of customer references. These may be so obvious that the authors (or that of the reference from which this list is taken) did not feel they needed to be mentioned, but they are still valid answers.

3. *Does a bottom-up data warehouse development approach use an enterprise data model?*

It need not. Kimball's bottom-up data mart approach usually uses dimensional data modeling, starting with tables. However, if an enterprise data model exists, its relevant parts can be used as a starting point for the tables rather than embarking on a from-the-ground-up data modeling project.

4. *Describe the major similarities and differences between the Inmon and Kimball data warehouse development approaches.*

Similarities: Both methods can produce an enterprise data warehouse and subset data marts.

Differences: Inmon's approach starts with an enterprise data warehouse, creating data marts as subsets of that EDW if appropriate. The focus is on proven, traditional methods and technologies. Kimball's starts with data marts, consolidating them into an EDW later if appropriate. It focuses in creating a useful end-user capability quickly.

5. *List the different types of data warehouse architectures.*

This is essentially the same as Section 2.4, Review Question 3.

Section 2.7 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 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), woot.com and steepandcheap.com use historical data[1] 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 operational and tactical (rarely strategic) ones. The reason: since strategic decisions affect an organization for a long time to come, they are not—and should not be—influenced by an hour's worth, or even a day's worth, of new data.

[1] An attempt by amazon.com in 2000 to create useful historical data artificially, by varying the price of the same movie to different customers and tracking the statistical effect on sales, was criticized as unethical and withdrawn. They subsequently offered refunds to customers who paid more than the lowest price offered to anyone.

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. The 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, and 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 2.8 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:

   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.
   2. Implementing logical security procedures and techniques to restrict access. This includes user authentication, access controls, and encryption.
   3. Limiting physical access to the data center environment.
   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 a 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

Section 2.9 Review Questions

Section 2.9 has no review questions.

ANSWERS TO END-OF-CHAPTER DISCUSSION QUESTIONS

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

   Data integration consists of three processes which 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 (Section 2.2) 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, non-volatile 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.

   (The first part of this question is essentially the same as Review Question 1 of Section 2.2.)

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 a 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 otherwise use a 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

   Benefits were listed in the answer to Review Question 1, Section 2.6.

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. Those skills are listed in the answer to Review Question 2, Section 2.8.

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

   A question involving the word "higher" requires us to ask "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 willingness to invest in a data warehousing infrastructure before it shows 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 potential benefits can be identified.

(This question is related to Section 2.6, Review Question 4. It would normally be redundant to assign both.)

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

Security and privacy[2] concerns are important in building a data warehouse:

1. Laws and regulations, in the U.S. and elsewhere, require safeguards on databases that contain the type of information typically found in a DW.
2. Valuable corporate data in a data warehouse can make it an attractive target for attempts to obtain those data improperly.

[2] Privacy added to the answer. The two concerns are different, but are often grouped—as they are in the relevant portion of this text. Privacy is one reason to want security. Without security, one cannot ensure privacy.

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, which is one way to limit potential violations.

Steps to address these concerns were given in the answer to Review Question 2 of Section 2.8.

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. states, but similar issues exist in any country large enough to have regional economic and labor cost differences.)

ANSWERS TO END-OF-CHAPTER APPLICATION CASE QUESTIONS

1. *Describe how Overstock.com ran its campaigns before the Teradata system was developed. How has the real-time data warehouse helped Overstock.com improve the performance of the business?*

The text does not provide information for the first part of this question. However, we can infer that decisions were made intuitively, without the benefit of a historical database or analyses based on its content. Therefore, prices may have been too high (so Overstock.com didn't achieve its objective, which was liquidating inventory in a timely fashion) or too low (in which case it did not obtain max-

imum revenue for liquidating it). The benefits are a more informed first price at which to offer close-out merchandise.

The real-time data warehouse, introduced after the earlier EDW, allows the firm to adjust pricing and other factors on the fly based on real-time information as to how they are selling. The benefit is greater ability to adjust the price of merchandise based on actual selling patterns. While the better first price that the original Teradata data warehouse made possible is an improvement over conditions that existed before it, it is still not perfect; the RDS permits fine-tuning it further.

2. *Go to the Web sites of the vendors mentioned in the case and examine the current data warehousing features and capabilities of each. Describe in detail how Overstock.com could potentially use each one.*

Vendors to look at include Teradata, GoldenGate Software and Sunopsis. The descriptions of how Overstock.com could use each one, which are open-ended, should be based on what the case says about how the firm uses them already.

(Teradata offers broad-based data warehousing capabilities that students may already be familiar with from Chapter 1. GoldenGate Software (www.goldengate.com) supplies software that provides real-time feeds of transaction data to a DW. Sunopsis (www.sunopsis.com), via its Data Conductor product, provides a business-rules-driven approach to the ETL process that has advantages over the traditional approaches, which mix business rules with the implementation of the ETL process.)