GRAB YOUR PICKS AND SHOVELS!

there's gold in your data

BY RAYMOND LANDRY, JR.;
ROGER DEBRECENY, CMA;
AND GLEN L. GRAY, CPA

The core business processes of many organizations are becoming more dynamic and complex because of globalization and evolving technology that has enabled us to cost effectively accumulate and store vast amounts of financial and nonfinancial transaction data in digital form. Today, many organizations are assembling this data—either the detailed transaction data or summarized data—in data warehouses. The ability to mine these warehouses for new knowledge about value-adding business processes presents opportunities and significant challenges for management accountants and financial managers. Data mining can help you better understand your business by providing new insights into areas such as customer buying and return patterns, product profitability, instances of general selling and administrative (GSA) expenses, and fraud detection.

DATA MINING

Despite claims by some vendors, data mining is *not* a single tool that can be bought off the shelf and put immediately to work. Data mining incorporates a variety of tools and processes that can work independently or together to analyze and discover relationships in collections of data. Actually, data mining tools can be applied to any collection of data—it doesn't have to be a warehouse.
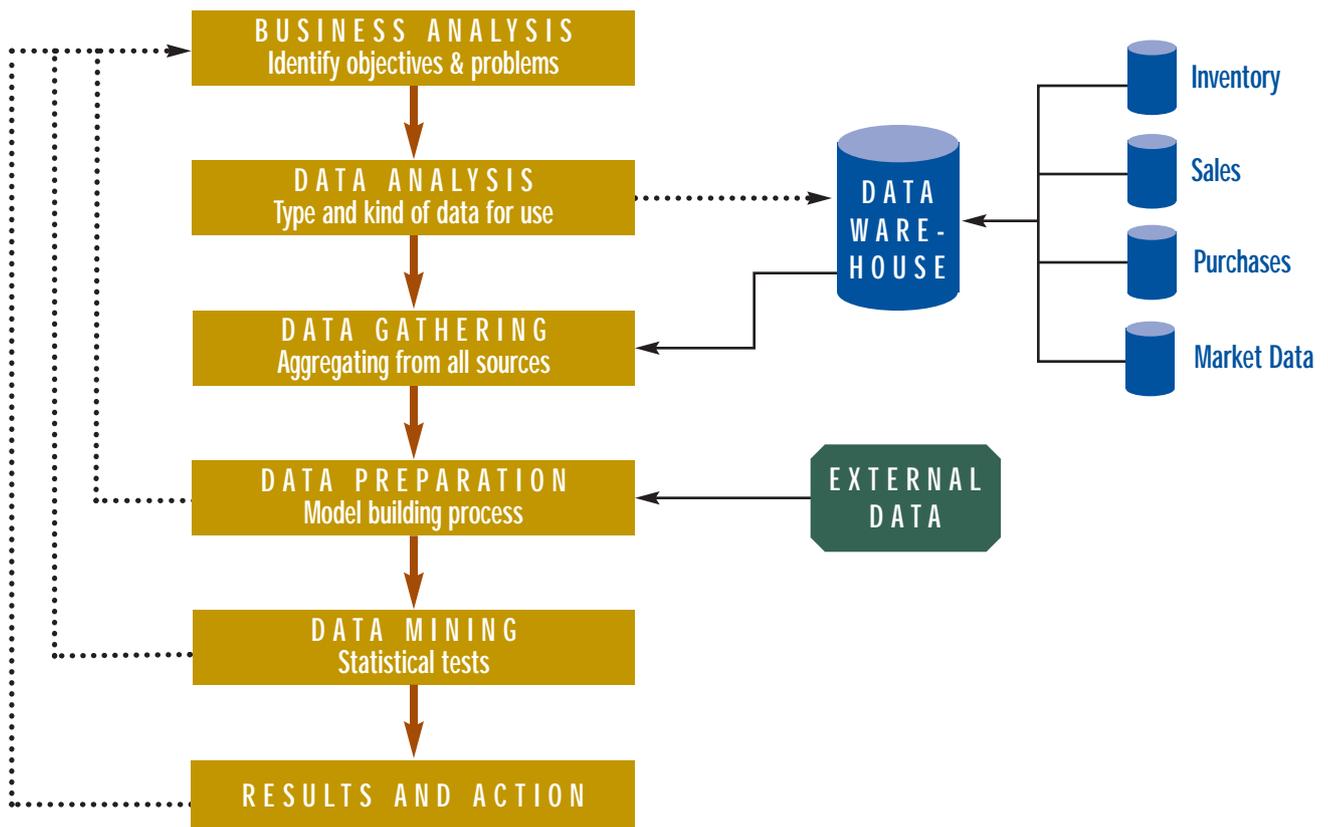
Also, data mining isn't a one-off process. It follows an iterative route along the lines shown in Figure 1. First, analysis of business processes, problems, and opportunities creates the need for data warehouses. Extensive data resources can also arise as a secondary benefit of software systems such as enterprise resource planning (ERP) or customer relationship management (CRM) packages from companies such as SAP, Oracle, or Siebel Systems. Data warehouses normally stand apart from production systems and are fed data from those systems. And they are often implemented in very large database management systems (VLDBs), such as NCR's Teradata product. The data warehouse can include internal transaction data as well as third-party data from, for example, credit rating agencies. Because search and analysis efficiency is more important than transaction processing (TP) efficiency, data warehouses aren't fully normalized and contain derived or calculated data (such as the difference between the ship date and the order date) that wouldn't be included in a TP-based database.

In the second phase, database query languages, report builders, and OLAP (online analytical processing) tools, such as Microsoft's Analyst Services, are used to screen, aggregate, and clean the data. Data that are from disparate sources are usually messy and inconsistent. For example, how many different definitions for sales revenue are there in your business? At this stage, important understandings about the entity's transactional data are built, and relationships in the data are found.

Finally, in the third phase, artificial intelligence (AI) and statistical analysis tools are used to test or discover relationships in the data. There are two primary styles of data mining. The first, *directed* data mining, is designed to test or measure expected patterns of business behavior. For example, given a particular volume discount structure, which companies will make larger orders? Or given patterns of known consumer credit fraud, which current transactions carry a high probability of risk? The second

Figure 1: Data Mining Process

BUSINESS ANALYSIS
Identify objectives & problems

DATA ANALYSIS
Type and kind of data for use

DATA GATHERING
Aggregating from all sources

DATA PREPARATION
Model building process

DATA MINING
Statistical tests

RESULTS AND ACTION

DATA WARE-HOUSE

Inventory

Sales

Purchases

Market Data

EXTERNAL DATA

Additional Information about Data Mining

SITE NAME	URL ADDRESS
DMSK	www.data-miner.com/
KDnuggets	www.kdnuggets.com/
Data Miners	www.data-miners.com/
Quest Data Mining Project	www.almaden.ibm.com/software/quest
DB2 Magazine	www.db2mag.com
National Center for Data Mining	www.ncdm.uic.edu/
Data Mining and Knowledge Discovery Journal	www.wkap.nl/journalhome.htm/1384-5810
InfoMine	www.infomine.com/

style, *undirected* data mining, seeks patterns or relationships without any preconceived expectations or hypotheses. Data mining, then, lets users search through large volumes of data for patterns that can be generalized in order to improve future decisions. As a result of the analyses made in the second and third phases, the company can make changes to different aspects of its value chain in pricing, discounting, and managing its business processes.

AREAS WHERE DATA MINING IS USEFUL

Data mining techniques have considerable potential in a variety of areas. For example, in sales and customer management, Pete Cabena and others, in their 1998 book *Discovering Data Mining,* say the technique can answer these questions:

◆ What kind of behavior pattern does your customer emulate?

◆ How can the organization make more sales to existing customers?

◆ In the sales databases, are there hidden patterns of buying?

◆ Who are the better customers, and who are the high-risk customers?

◆ How can you maintain loyalty from current customers?

◆ How can you identify unknown buying habits and specifically market to those habits?

◆ What are customer perceptions of company products?

◆ How do you improve operational and strategic business plans based on data mining results?

The organization-level driver to support the investment in a data warehouse and in data mining tools is that these investments identify value-adding business processes. In a business-process orientation, a search through the company data warehouse is based on a need to under-

stand your own business processes. In addition, management can use the data mining results to model the company's business behavior under a variety of circumstances and to provide critical business intelligence.

While data mining tools can identify information that is useful for future planning, data mining can also be very useful for detecting fraudulent (theft or invoice manipulations) or high-risk activities (bad loans, investments). Consequently, your internal auditors will also find data mining a very useful tool.

SUCCESSFUL DATA MINING PROJECTS

Data mining integrates complex information systems with an understanding of underlying business processes. Successful data mining requires considerable planning and investment of which the data warehouse is likely to be the largest component. It will often take a couple of years to plan and implement a data warehouse, but the primary benefits come in day-to-day performance reporting. The payoff that arises from mining the data warehouse is usually ancillary to these primary benefits. It's important that the design of the data warehouse facilitates data mining as well as the primary production reporting systems. It may be acceptable, for example, for production reporting systems to aggregate product sales data by week or month, but the process of aggregation may conceal important relationships that could be identified by data mining tools. In addition, top management support, open exchange of data within a company's information system, and knowledge of the company's business processes are all essential prerequisites for successful data mining.

Should companies develop data mining capabilities? There are many options for those that don't have data mining expertise or where the company has decided it isn't cost effective to develop such expertise. If your organization chooses to do its own data mining, Joerg Rein-

Data Mining Vendor Sites

VENDOR/PRODUCT NAME	URL ADDRESS
The Data Mine	www.the-data-mine.com
Salford Systems	www.salford-systems.com
RIK and EDM Software	www.data-miner.com
IBM Intelligent Miner for Data	www.software.ibm.com/data/iminer/fordata/
Oracle	http://otn.oracle.com/products/bi/odm/odmining.html
SAS Institute Enterprise Miner	www.sas.com/technologies/analytics/datamining/miner
SPSS Clementine	www.spss.com/clementine/
Alice d'Isoft	www.alice-soft.com/
Attar Software	www.attar.com/
RuleQuest	www.rulequest.com/
Angoss	www.angoss.com
Intellix	www.intellix.com/
GR-FX	www.gr-fx.com/graf-fx.htm
SGI MineSet	www.sgi.com/products/appsdirectory.dir/linux/products/m/958629.html

schmidt, in *Intelligent Miner for Data: Enhance Your Business Intelligence*, suggests the following procedures to conduct data mining successfully. These steps are also diagrammed in Figure 1.

Business analysis	◆ Identify the business processes where data mining will be useful.
Data analysis	◆ Identify the type of data. (Is there a data warehouse?)
	◆ Do the data need "cleaning"?
	◆ Identify the quality of the data.
Data gathering	◆ Identify which data are appropriate for use in the next step.
	◆ Cleaned data are aggregated with other appropriate internal and external data sources.
Data preparation	◆ Identify outliers and missing values.
	◆ Normalize values in certain fields.
	◆ Identify intercorrelations.
	◆ Do model building and testing.
Data mining	◆ Statistical techniques are applied to the data.
Interpreting results	◆ Managers and accountants analyze results and relate them to business operations to determine if results are valid.
	◆ Is further data mining necessary?
Business application	◆ Company acts on the results and incorporates them into its business plan.
Business feedback	◆ Compare predicted vs. actual results. If there are significant differences, revise the data mining model.

GOOD TIME TO START

The time is ripe for your business to consider data mining. Advances in technology have made massive data storage and data warehousing relatively inexpensive, and computing power can handle the data mining statistical techniques with ease. Why let valuable information remain untouched within your company databases? Data mining has a wide range of uses from searching for unknown patterns in company databases to identifying fraudulent activity. While data mining tools will require training and expertise to use, there is much information on the Internet, and the job can be outsourced. Successful uses of data mining have demonstrated that it's a very powerful tool that can unearth treasures among your organization's data. ■

Raymond Landry, Jr., Ph.D., is an associate professor in the Department of Accounting and Business Law at the University of Texas Pan American. You can reach him at fblandry@COBA.panam.edu.

Roger Debreceny, CMA, FCPA, Ph.D., is an associate professor in the Nanyang Business School at Nanyang Technological University, Singapore. You can contact him at adebreceny@ntu.edu.sg.

Glen L. Gray, CPA, Ph.D., is a professor in the Department of Accounting and Information Systems at California State University at Northridge. You can contact him at glen.gray@csun.edu.