IBM DB2 OLAP Miner:
an opportunity-discovery
feature of DB2 OLAP Server

Table of Contents

IBM

Distinguishing between OLAP and data mining

Rapid response to queries with DB2 OLAP Server.

Do you need OLAP and data mining?

While day-to-day business decisions are based on facts, strategic business decision-making hinges on insights into patterns or trends that can shape your business. To gain such insights, many businesses turn to data mining—a data analysis method that involves extracting predictive information tucked away in a data warehouse.

Data mining complements more traditional analysis methods such as online analytical processing (OLAP). The key difference between the two methods is that data mining is discovery-driven, while OLAP analysis is hypothesis-driven. In other words, data mining predicts or discovers insights that may be hidden or counterintuitive. In contrast, OLAP allows the user to navigate the data to find insights into specific questions the user has already identified as important.

Analysts predominantly use OLAP cubes to identify anomalies that may represent problems or opportunities. Currently, OLAP servers support high-speed, hypothesis-driven exploration of multidimensional data by using slice, dice and drill operations. Navigating through large data sets that encompass many aspects of the business, analysts can use a solution such as IBM DB2® OLAP Server™ to rapidly investigate areas of interest, frame ideas and build what-if scenarios. DB2 OLAP Server supports the rapid query response and simple navigation that makes hypothesis-driven analysis feasible.

Multidimensional analysis is well suited to answering "what" questions:

- What areas of my business are doing well?
- What products are failing to meet forecasts?

Highlights

Answering "why" questions with data mining

Finding deviations in business data.

Mining, on the other hand, is better suited for the "why" questions. For example, multidimensional analysis could quickly determine that car loans exceeded profit goals in the first quarter. But mining would be more effective in identifying the reason for this, for instance, increased activity from women aged 18 to 25 who are the single heads of households. Consider another example: multidimensional analysis quickly identifies that sales of outdoor leisurewear in the Northeast territory were down by five percent. Mining may be the best way to find out that this decline stemmed from a combined effect of lower revenues from vests, pants and specialty items during different time periods. These examples demonstrate that together, multidimensional analysis and data mining can yield greater analytical value than either method alone.

Introducing DB2 OLAP Miner

According to research conducted by the IBM Almaden Research Center, DB2 OLAP Miner (OLAP Miner)—an opportunity-discovery feature of DB2 OLAP Server—introduces a new technology that marries discovery-driven analysis with multidimensional analysis. Applying data mining algorithms to OLAP cubes, OLAP Miner discovers interesting values through an exhaustive search of the multidimensional data. Values that the algorithm marks as interesting are then presented to the analyst to be used as a starting point for further hypothesis-driven exploration.

By automating analysis that results in new insights, OLAP Miner can support a fast path to actionable information. It also adds another dimension to OLAP analysis, enabling IBM to offer—in addition to high-speed, ad hoc data exploration of OLAP cubes—exhaustive, machine-driven, algorithm-based data mining to discover unexpected values.
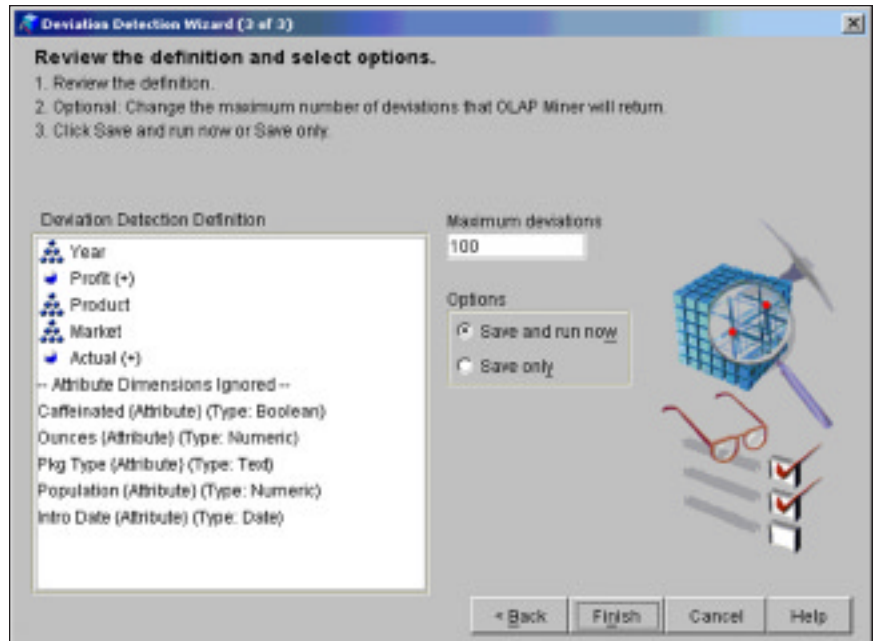

Highlights

Running algorithms against
OLAP cubes

Graphical interface to guide users.

OLAP Miner was developed as a set of components that work together with DB2 OLAP Server. Since OLAP Miner runs algorithms against OLAP cubes, the starting point for any deviation detection run is an OLAP cube that has been loaded and calculated. The OLAP Miner client component is a graphical user interface that guides the user through the set up of a deviation detection definition. This involves selecting:

• A source cube and the number of deviations to report
• Individual members and member combinations from dimensions to fully define the subcube to be mined
• Individual deviations to view in broader contexts.

In the example below, we want to find the unexpected values in actual sales results. The subset of the OLAP cube to be analyzed is the slice identified by *Sales* and *Actual*. Sales is a member of the Measures dimension. Actual is a member of the Scenario dimension.

The Deviation Detection Wizard retrieves data deviations.

Highlights

Determining data deviations

OLAP Miner first extracts the OLAP data corresponding to Sales and Actual. It then applies the deviation detection mining algorithm to the extracted cells. Once the algorithm has traversed all required cells, a list of deviations is written to a report. The report can be viewed using the OLAP Miner Deviation Viewer or a spreadsheet application. From the sample deviation detection definition above, we get the results shown below. The values in the Magnitude column indicate how much the deviant value differs from other surrounding values. The larger the magnitude, the more significant the deviation.


OLAP Miner compiles reports of data deviations and presents the information in its Deviation Viewer.

Highlights

Putting deviant values into context

Individual deviant cells can be further investigated by using the Deviation Viewer or the regular DB2 OLAP Server spreadsheet interface. This allows the analyst to see the deviant value in context. For example, a spreadsheet chart was built from the first deviation shown on the list on Page 4. This shows Sales for Televisions in various markets in October, November and December. The deviation report indicated that OLAP Miner expected Sales to be higher for Boston in December. This value is highlighted in red in the report. The chart clearly shows why Boston television sales would be expected to be higher.
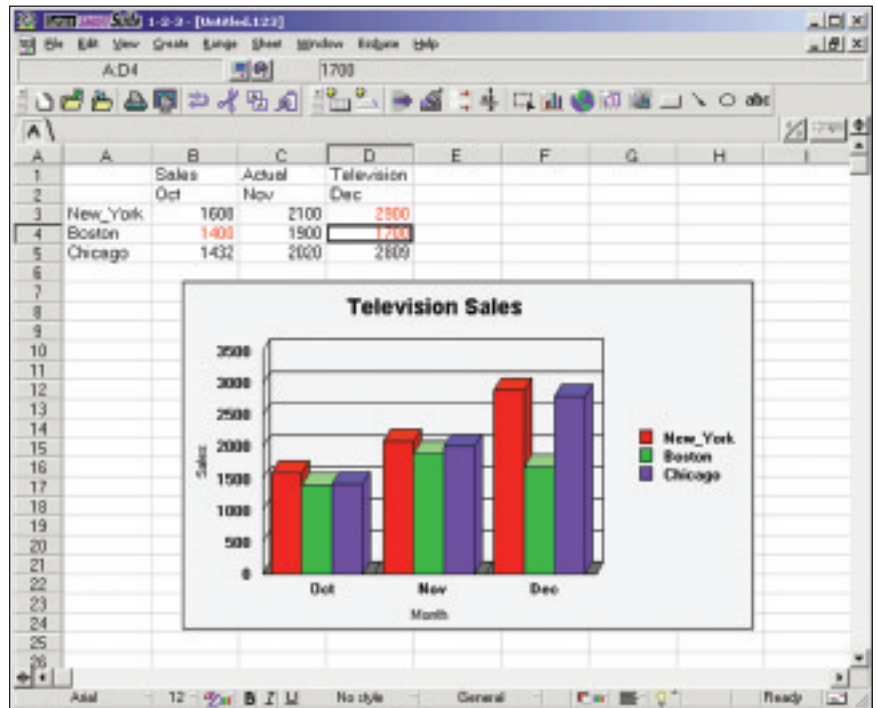
A spreadsheet application such as Lotus® 1-2-3® is another option for viewing OLAP Miner reports.

Highlights

Understanding OLAP Miner

ING Antai speeds claims payments.

How does OLAP Miner work?

The value of a cell in a cube is considered deviant if it is significantly different from the expected value, which is calculated by the data mining algorithm. OLAP Miner uses a statistical model that computes the expected value of the cell in the context of its position within the subcube, taking into account data along the dimensions in which the cell belongs. OLAP Miner maintains multidimensional context to discover anomalies that fall outside the expected values.

For each cell, an expected value is calculated. The absolute difference between the actual and expected values is termed the residual of the model. Any cell where the residual is relatively large is considered a deviation. A statistically valid definition of relatively large is required, so the residual values are scaled based on the standard deviation associated with the residual. If the standardized residual is higher than a defined threshold, the value of the cell is considered a deviation.

The data mining algorithm has been optimized so that it can be run against large OLAP cubes. A form of multiplicative model-fitting is used to calculate the expected values. This is reduced to addition by taking logs. Coefficients are calculated based on these logs. Standard deviations are calculated using all the residuals associated with the cell in question.

Case in point: ING Antai

Established in 1988, ING Antai of Taiwan (formerly Aetna Life of Taiwan) is the country's first foreign life insurance organization. Its 20,000 agents sell individual and group life insurance, health, travel and accident insurance as well as financial products such as mutual funds.

Highlights

Gaining insights into customer behavior

Uncovering new business opportunities.

Committed to strengthening customer loyalty by providing the best service possible, ING Antai relies on business intelligence software from IBM to gather insights into customer behavior. By analyzing data associated with customer accounts and interactions, the company can better gauge product performance, market products and services to targeted audiences, and even combat insurance fraud.

Now, ING Antai has taken its business intelligence solution to the next level, by participating in the beta program for OLAP Miner. A no-charge, value-added feature that can be installed from DB2 OLAP Server, Version 7.1 FixPak, OLAP Miner enables ING Antai to quickly discover significant fluctuations in areas such as sales numbers and budget costs, so the company can respond in a timely manner to swiftly changing conditions.

"DB2 OLAP Miner adds an extremely useful and easy-to-use tool to our business intelligence environment," notes Austin Wang, section chief, ING Antai. "It is dramatically cutting our costs and enabling us to pay claims much faster—a competitive advantage that improves our customer retention and acquisition rates for enhanced, long-term profitability. For example, with OLAP Miner, it takes just two or three days to analyze fraud claims data, a process that used to take two to three weeks. In addition, by automating the scanning of OLAP cubes, OLAP Miner enables us to quickly gain a rich understanding of our customers so that we can cross- and up-sell from our multiple lines of business. This allows ING Antai to increase revenues while providing better customer service."

IBM

*Paradigma is an early adopter of
IBM data mining technologies*

Paradigma extends its IBM business intelligence platform

Paradigma is an IT consultancy specializing in strategic analysis, decision
support and business intelligence. Since its beginnings, Paradigma has relied
on IBM DB2 Universal Database™ for its performance, availability and
scalability in support of various client applications.

An early adopter of IBM's data mining technologies, Paradigma has
successfully completed projects in telecommunications and retail. Deploying
OLAP Miner, Paradigma has been able to provide a valuable contribution
to the multifaceted analysis of European market data. OLAP Miner has
enabled the consultant to identify opportunities in market segments that
would otherwise have remained hidden.

*Elytics enhances enterprise
analytics platform.*

Elytics incorporates OLAP Miner into its scalable enterprise analytics platform

Founded in Cambridge, Massachusetts in 1999, Elytics, Inc. is a privately
held developer of business intelligence solutions. Its software solution enables
e-businesses to quickly gain an understanding of Web site visitor behavior, as
well as other patterns and opportunities.

"Our mission at Elytics is to provide the most advanced set of ROI metrics
for the e-channel available anywhere in the market through a secure, scalable
enterprise analytics platform," says Elytics CEO Jon Christensen. "The proven
track record of DB2 in the areas of high performance, rapid response and
24x7 availability make it a natural choice for Elytics' database platform.
Elytics Analysis Suite (EAS) is the OLAP product of choice for e-channel ROI
measurement and CRM analysis.

Highlights

OLAP Miner provides cost-effective analytical tool

Driving analysis with data mining discoveries.

"With IBM DB2 OLAP Miner, our customers can easily deploy EAS and customer data marts to leverage key business intelligence features in DB2, including high-speed load, federated data access and sophisticated mining to discover new trends and directions in customer buying segments. With OLAP Miner, we know our customers have an affordable, leading-edge technology they can use to easily gain additional insight and a financial data mart that enables them to discover new trends in sales and customer behavior patterns to make critical business decisions."

Summary
OLAP Miner for DB2 OLAP Server brings together IBM data mining and OLAP technologies. Moving beyond the simple application of existing data mining technologies to OLAP cubes, the deviation detection algorithm is designed and optimized to apply statistical models and data mining techniques to OLAP cubes to find interesting values.

Business analysts can then use the results of the data mining analysis to focus on the interesting values in the context of the source OLAP cube. The usual slice, dice and drill operations can be used to explore the data, but the analysis is driven by the discoveries made by the data mining algorithm.

In this way, OLAP Miner offers a combination of exhaustive, machine-driven statistical analysis and discovery-driven human analysis. As a result, potential opportunities and problems will not be hidden in the data, and the analyst can quickly find important, actionable information.

IBM ®