B2B Enterprise Technology Story

Web Summary:

Profitable Knowledge from Data

by Leon A. Enriquez

Reading Time:
28 minutes

Reader Benefit:
- ♦Understanding knowledge access as compared to data mining;
- ♦Knowing the difference between correlated "patterns" in the data from random data analysis;
- ♦Pattern query language (PQL) is the latest approach in finding the logical connections in massive pile of historical data.

Jargons proliferate in the database jungle. These arise from the business transactions and the logical connections needed to keep track, manage and correlate data – and then, refine the data to impact decision making. Discover more as we follow the route where data becomes strategic information.

Most companies now collect and refine massive quantities of data in their data warehouses.

Data mining and knowledge access have a variety of applications in many industries. They improve an organisation's effectiveness, efficiency and value by enhancing the knowledge the organisation possesses.

Data mining is the automatic extraction of patterns of information from historical data to enable companies to focus on the most important aspects of their business. Data mining yields valuable information – and reveals to the company what they did not know, about what they never thought of asking!

With knowledge access, almost all the relevant trends or "patterns" in the data are found beforehand, and stored for use by business users such as marketing analysts, bank branch managers, store managers, etc. Business users get the interesting patterns of change every week or month or can query the pattern base at will.

Such companies realise that to succeed in a fast paced world, business users need to be able to get information on demand. And often, they are pleasantly surprised by unexpected but useful information. The computer should do this job itself because there is never enough time to think of all the important questions.

Not surprisingly, many of the more successful companies now view information as one of their most valuable assets. Thus, a company needs to make full use of valuable information derived from their data assets. By exploring the database itself and extracting invaluable nuggets of information, it can provide a boost to the business.

Profitable Knowledge from Data

by Leon A. Enriquez

Jargons proliferate in the database jungle. These arise from the business transactions and the logical connections needed to keep track, manage and correlate data, and then, refine the data to impact decision making. Discover more as we follow the route where data becomes strategic information.

Computers have been used to capture details of business transactions such as retail sales, banking and credit card records, telecommunications, etc. From these transactional systems, the data contain details of the key trends that impact various aspects of each business. These include products that sell together, sources of profits, etc. This data is gathered over time and stored in a separate database, usually a data warehouse.

In contrast, operational data deals with daily activities whereas the warehouse data is historical in nature and is used to gain insights into the business trends. The analysis of historical data gathered renders a fresh perspective and insight that can be used to improve business decisions.

Data mining is the automatic extraction of patterns of information from historical data to enable companies to focus on the most important aspects of their business. Data mining yields valuable information – and reveals the company what they did not know, about what they never thought of asking! Most companies now collect and refine massive quantities of data in their data warehouses.

Such companies realise that to succeed in a fast paced world, business users need to be able to get information on demand. And often, they are pleasantly surprised by unexpected but useful information. The computer should do this job itself because there is never enough time to think of all the important questions.

By exploring the database itself and extracting invaluable nuggets of information, it can provide a boost to the business. Most companies now view information as one of their most valuable assets, and as such, data mining enables a company to make full use of valuable information derived from their data assets.

Decision Support

Decision support is a term referring to the use of information as a strategic corporate asset, enabling companies to utilise their databases to make better decisions. Decision support systems (DSS) have traditionally relied on three types of analyses:

- *Query and Reporting:* This is where a user asks a question, e.g., “What were the sales for a specific product?”
- *Online Analytical Processing (OLAP):* This amounts to the processing of queries along multiple dimensions, e.g., state, month, etc. For instance: “Categorise sales by month, by brand, by outlet.”
- *Data Mining:* This provides influence factors and relationships in data, e.g., “What impacts sales in Singapore for a given product.”

With statistics and reports, just summaries of data were available to business users. And, the data could only be obtained by request from an analyst, e.g., sales summaries per quarter. With data warehouses, some query and reporting could be performed by business users on their own, e.g., product and store performance reports, etc.

With OLAP, multi-dimensional summary questions could be addressed by business users, e.g., the total of sales by product, by channel, by month. With data mining, analysts and a sophisticated subset of business users could gain insight into the influence factors and trends in data. But often, significant analysis was needed before key questions could be answered.

With knowledge access, almost all the relevant trends or “patterns” in the data are found beforehand, and stored for use by business users such as marketing analysts, bank branch managers, store managers, etc. Business users get the interesting patterns of change every week or month or can query the pattern base at will.

Because large databases often provide too much data, approaches based on Query and OLAP usually encounter a problem. Too much data leads to “information myopia” where a user fails to derive the real value for decision-making. To solve this problem, data mining has the ability to search through the data on its own initiative, thereby discovering key patterns without human intervention.

Although the three approaches mentioned are useful, they share a common trait. The user has to perform analysis to gain knowledge. Obviously, a more dynamic approach is to empower business users with refined information, where data analysis is performed beforehand and the user just looks up the pre-mined knowledge on demand.

Strategic Information

To distil information from a database, we obviously need to analyse it at some time. The key question is: “When?” Does the analysis take place at the time the user needs the knowledge, or is it done beforehand, with the knowledge ready to access? Traditionally, data mining analyses were performed upon user request.

Thus, there are two distinct paradigms for empowering users with strategic knowledge (or refined data/information) as follows:

- *Data Analysis:* Where users operate on data to discover information. This paradigm relies on the “analysis on-demand” approach, i.e., when a user wants knowledge, analysis is performed.
- *Knowledge Access:* Where the analysis is automatically done beforehand, refined patterns are pre-generated and users just get knowledge when needed, i.e., the “knowledge on demand” approach – which changes the game in favour of the business user. The user can just reach for refined knowledge when needed, without the need for analysis.

Obviously, this approach to acquire strategic knowledge provides a multitude of benefits to the business user, namely:

1. *Easy-to-use, yet powerful:* Business users without technical know-how can access knowledge without training – they just click a graphic user interface from within a web-browser. And, the knowledge access approach is more powerful because multiple types of powerful patterns are automatically merged to answer serious questions. With the analysis paradigm, business users inevitably rely on simple models and can not deal with complex situations on their own.
2. *Condensed information:* Because of disk space limitations, many organisations only store 12 or 24 months' worth of historical data. However, because knowledge is so much more compact than data, the pattern-base is only a fraction of the size of the database, allowing many years' worth of patterns to be stored with ease, even when the data is no longer available.
3. *Fast response and overall efficiency:* When a user requests knowledge, no analysis is needed, and follow-up questions are answered quickly without delay. Data mining on a very large database may take time, but pattern look-up is fast. And, because patterns are not re-computed each time for each user, the overall system efficiency is much higher. Computations take place once, and users access the refined knowledge again and again with ease.
4. *Accuracy and quality:* Because sampling and extract files are avoided, the discovered patterns correspond to the entire database and have high accuracy, resulting in better decisions. And, because patterns are stored in a single repository, all users get similar answers, rather than relying on fragmented analyses. This is in contrast to the data analysis paradigm where different users may draw different conclusions from the same data.
5. *Up-to-date knowledge:* Because the pattern-base is incrementally updated, recent patterns are always available. With the data analysis paradigm, there is usually not enough time to continuously analyse new data, and often users are forced to rely on out-of-date analyses.

The knowledge access paradigm is an innovative idea with a multitude of business and technical benefits that reinforce each other. As a tool, knowledge access will change the way you access your corporate knowledge.

Industry Applications

Data mining and knowledge access have a variety of applications in many industries. They improve an organisation's effectiveness, efficiency and value by enhancing the knowledge the organisation possesses. Industry examples include the following:

1. *Customer prospecting and segmentation:*

In industries such as banking, credit cards and insurance, data mining helps break the market into meaningful clusters and segments. This helps identify customer groups that not only have a higher likelihood to respond to offers, but will provide better profits in terms of higher volumes of product usage.

And, by analysing results of direct mail campaigns higher-value responses, e.g., in profit dollars, etc. as well as higher response-rates are obtained. This information helps marketing managers and business executives better focus promotional activities and design new marketing campaigns.

2. *Customer relationship management:*
Here, data mining finds patterns of product usage and helps determine consumer behaviour. It helps you understand what causes customer attrition and improves customer retention. It discovers cross-product utilisation behaviour, improving the management of channels such as bank branches, ATMs, etc.

These patterns can re-shape the thinking of a customer loyalty department. And, by combining trend analysis with wallet-share and affinity results, “right time marketing” is implemented based on a customer life cycle model.

3. *Retail sector:*
For example, in supermarkets, department stores, etc., data mining helps identify how products sell together in specific stores and/or regions through affinity and market basket analysis. Data mining also helps you understand profit patterns per linear foot of shelf space based on store layouts, product combination offerings, etc., and determines when promotions work best and which item layout combinations are the most effective.

Data mining also helps identify products which are traffic builders, and those which are profit makers. The impact of advertising and in-store promotions is identified and inventory control issues and carrying costs are analysed.

4. *Consumer packaged goods companies:*
Analyse one manufacturer’s products in many store chains, not several manufacturers’ products in one chain. Here, market-share analysis by store-chain and demographics, and understanding where and when promotions work and when they don’t helps focus advertising dollars.

And, by analysing historical performance from the chain level, down to the region and store levels, the factors influencing the effectiveness of advertising are identified, contributing to profitability.

5. *Telecommunications:*
In this industry, changes in regulation complicate the dynamic nature of the local and cellular telephone markets. Segmentation and prospecting strategies based on data mining help organisations change strategies to keep up with a changing world.

Data mining identifies the patterns of change in the market, allowing the marketing department to better focus on customers with a high acceptance ratio and longer usage lifecycles. The key characteristics shaping each segment of the market are automatically identified, both based on customer attributes and in terms of geographic clusters.

In network capacity utilisation, planning for telecommunication markets that expand, and the ability to service customers is highly affected by the capacity planning that has taken place long before.

Data mining provides an understanding of the underlying patterns and structures of service usage by customer groups. This insight allows capacity planners to optimise the investments in network facilities to better serve customers, while avoiding costly over-expansions – having enough capacity to deal with growing markets just at the right time.

Trend Management?

The concept of data management is often taken for granted simply because it has been around for a long time, and has most notably been identified with databases (DB). From the early concepts, further refinements in managing such a database resulted in database management systems (DBMS), and the process of correlating and synchronising the various elements of the data collection have led to relational database management systems (RDBMS).

Thus, to understand data management in a meaningful way, we need to look at the two practical approaches to capitalising on the database repository, which are as follows:

- Management of operational data by systems that process transactions, which occur often on a real-time basis; and
- Management and analysis of historical data by decision support systems (DSS) that provide business users with insight.

According to Information Discovery, Inc. (IDI), it is clear today that business users do not want massive volumes of data, but are interested in the trends or “patterns” buried within data. Such patterns need to be accessed, manipulated and managed, in the same fashion as data elements need to be managed. Thus, pattern management system (PMS) pioneered by IDI deals with patterns just as a data management system deals with data.

PMS requires distinct repositories and query languages just as languages have been developed for data management. In pattern-oriented systems, patterns are treated as first-class elements of languages and repositories. Note that the traditional tools meant for data management such as SQL (structured query language) are not designed to deal with patterns.

Notice also that pattern management is not any of the following, namely, knowledge management (KM), data mining (DM), or the construction of a knowledge-based system.

For instance, the knowledge that is based on patterns known to humans is often a mere fraction of the patterns that a database implicitly contains.

Note also that data mining is a process that precedes pattern management; here pattern discovery then feeds a pattern repository that will be managed thereafter. Put another way, pattern management deals with patterns of knowledge – after they have been discovered by data mining. One of the software tools developed specially for this is the PQL (Pattern Query Language) which is a unique innovation from IDI.

As data warehouses grow, the need for pattern management becomes a critical aspect of data management. Why? Simply because it frees the business user from the burden of dealing with data. Here, the business user deals with patterns as the basic tokens of the information system thus going beyond the rudimentary facilities provided by data management systems.

Data versus Patterns

Consider the difference. Data is raw whereas patterns are refined. A pattern expresses relationships between data items but not the data. There are several classes of patterns, namely:

- *Influence patterns* – which are often reflecting probabilities or likelihood, as well as affinity patterns that deal with associations e.g., market-basket patterns;
- *Comparative patterns* – that point out differences among data sets. Each pattern class has specific rules of inference for the manipulation of patterns.

Let's illustrate the case using a simple analogy. Consider data as grapes; patterns of knowledge as wine; thus, data mining is similar to the wine making process. Therefore, a data repository is a storage facility for grapes – and a pattern repository somewhat akin to a wine cellar. In this case, the data mining tools are like wine making equipment. Although users can make their own wine by getting grapes from the warehouse, this requires time and know-how to turn the grapes into wine. Quite naturally, most business users prefer bottled wine. With pattern management, the data mining process still takes place transparently such that the business user never needs to know about it at all.

Data Analysis vs. Pattern Management

Pattern management follows data analysis, and is hence distinct from it. Data mining analyses are frequently performed upon user request.

The pattern management approach is dynamic, and help users by pre-mining refined knowledge. Consequently, there are two distinct mindset shifts for empowering users with actionable knowledge as follows:

- With data analysis, users operate on data to discover information. This scenario relies on the analysis-on-demand approach, i.e., when a user wants knowledge, analysis is performed – but not ahead of the demand.
- With pattern management, the analysis is automatically done beforehand, and refined patterns are pre-generated, i.e., users just get actionable knowledge when needed, on demand.

Here we can see the compelling case for pattern management. To the business user, a multitude of benefits can be obtained using the pattern management:

1. Business users without technical know-how can access knowledge without training, i.e., they just click a graphic user interface from within a Web browser.
2. When a user requests knowledge, no analysis is needed, and follow-up questions are answered quickly, without delay.
3. Data mining on a very large database may take time, but pattern look-up is fast.
4. Because patterns are not re-computed each time for each user, the overall system efficiency is much higher. Computations take place once, and users access the refined knowledge with ease.
5. Because sampling and extract files are avoided, the discovered patterns correspond to the entire database and have high accuracy, resulting in better decisions.
6. Moreover, because patterns are stored in a single repository, all users get similar answers, rather than relying on fragmented analyses. Contrast this to the data analysis paradigm where different users may draw different conclusions from the same data.
7. Pattern management avoids the probability of 100 users getting 100 different answers from the same data, because now corporate knowledge is truly centralised.

Pattern Management Components

To deal with patterns, we need to collect, store, manipulate, access and visualise them. Thus, we need repositories, query languages such as PQL, and systems to deal with refined patterns rather than raw data. Each of these has an equivalent in the data management world.

The concept of a data warehouse was championed in the 1980's as a repository for corporate data elements. The idea was to create a central storage facility where everyone in the corporation could go and get "data" on demand, whenever they needed it. And, the central repository would help increase corporate data quality and consistency because everyone obtained data from a single source. This idea has now achieved widespread global acceptance with many companies owning data warehouses.

Similarly a pattern repository can be used to hold historical patterns as opposed to historical data. With such a pattern repository, almost all the relevant patterns in the data can be found beforehand, and stored for use by business users, e.g., marketing analysts, bank branch managers, store managers, etc. Actionable knowledge becomes more accessible as business users get the interesting patterns of change every week or month, and can query the patterns at their convenience.

For example, patterns can be represented as a set of "pattern-tables" within a traditional relational database. This solves several potential issues regarding user access rights, security control, multi-user access, etc.

Obviously, we need a language to access and query the contents of pattern repository. Structured Query Language (SQL) may be considered the obvious first choice. Note that data mining was not a major issue when SQL was designed over 30 years ago. SQL was specifically designed to access data stored in databases. Obviously, it’s a daunting task to access these patterns using SQL. To do a better job, pattern-oriented languages are needed to access the pattern repository storing various types of exact and inexact patterns.

Patterns cannot be conveniently queried in a direct way using a relational query language. Because some patterns not easily stored in a simple tabular format, just looking up influence factors in pattern-tables may bring you incorrect results. We need a “pattern-kernel” in the software that consistently manages and merges patterns.

Note that pattern query uses the “pattern algebra” whereas SQL relies on the relational algebra. Thus, pattern query process should use SQL as part of its operation, i.e., where pattern queries are decomposed into a set of related SQL queries, after which the results are re-combined. However, business users just click on a graphic user interface to retrieve patterns on the intranet. They can begin immediately to access knowledge without lengthy training sessions or analytical know-how.

With pattern visualisation, the user still performs analysis, e.g. visualises affinity patterns – the results delivered for the same level of computational effort are much better because the user now analyses refined knowledge and not raw data. Consider now that 100 different analysts will no longer get 100 different answers from the same data because there is a central knowledge repository.

For instance, a natural way of delivering pattern-based information to users on the Web is a document organised as a collection of information of different types, e.g., text, data, graphs, etc. Such an “explainable document” looks like any other web page at first, but does much more by allowing users to dynamically obtain explanations that clarify, justify and substantiate the patterns presented within the document.

Box Story 1:

Database Management

The need to track large quantities of data in your company can become an overwhelming experience. No wonder then, that many users are frustrated by this ever-escalating dilemma. Are you one of the many who are suffering from data distress?

Perhaps you’re currently using a paper filing system, text documents or a spreadsheet to help you keep track of your critical information. If you’re searching for a more flexible data management system, a database might be just the solution that can help you do your job more effectively and efficiently.

You may hazard the question: What is a database? Quite simply, it is an organised collection of data. A database management system (DBMS) such as Microsoft Access, Oracle or SQL Server provides you with the software tools that you need to organise data in a flexible manner. It includes facilities to add, modify or delete data from the database, ask questions (also known as queries) about the data stored in the database and produce reports summarising selected contents.

Fields We need databases because we need a way to structure a big mess of data. The very simplest database is concerned with only one category or "field" of data, e.g., a list of customer names. But by itself, a list of names is not very useful. What we truly need is more information to relate to these customer names. For instance, the total bill on the last invoice. With this in mind, we now have a simple database with two fields which are related to each another, i.e., customer names, and total bill on the last invoice.

To help us with our simple illustration, we will use Microsoft Access software as a typical example. Going further our discussion, let's first examine just three of the major components that most database users will encounter. These three components are: Tables, Queries, and Forms.

Tables Tables comprise the fundamental building blocks of any database. Are you familiar with spreadsheets? You will find database tables similar to spreadsheets. To help you get a better understanding, just think of a table as a spreadsheet-style listing of information, in any particular order that you like. (See Table 1)

Customer Name	**Last Invoice Amount ($)**	**Product Description & Brand**
Jess S	450.00	Card Scanner
Anne S	4,250.00	Laptop PC, Sony
Elen S	2,400.00	Desktop PC, HP
Kelly T	1,700.00	Digital Camera, Canon
Bhad M	1,850.00	iMac PC, Apple
Allen E	2,950.00	Laptop PC, IBM

Table 1: Customer Name, Invoice Amount and Product Description

Query: From the simple example in Table 1, we can see how a table allows us to create a framework for storing information in a database. However, storing information is not the reason for the creation of a database as it would be a futile exercise. Here we need to introduce a new element that will enable us to retrieve particular information fields using a logical selection method or process as well. This process of looking for useful information is referred to as a query.

In the more sophisticated database software programs, there is a specific feature designed to enhance the meaningful extraction of the desired information presented in a meaningful way. This functionality is known as SQL or structured query language.

For instance, if you want to recall the information stored in a table, the software allows you to open the table and scroll through the records contained within it. The real power of a database lies in its capabilities to answer more complex requests or queries. For example, Microsoft Access queries provide the capability to combine data from multiple tables, and place specific conditions on the data or information retrieved.

A simple example
Let’s imagine a real-life situation. Your company sells a vast variety of computer products. And your boss requires a simple method to create a list of those products that are currently selling well with a gross margin of 25 percent or more. This will enable him to ensure that the inventories of such products are sufficient to fulfill all customer orders.

Under normal circumstances, this is what happens to fulfill this simple task:
Step 1: You retrieve the product information table;
Step 2: You spend a large amount of time sorting through data;
Step 3: You perform calculations manually;
Step 4: You now have to compile and tabulate the data in a new form; and
Step 5: You repeat the sequence for as many times as you need before you get the final statistics.

Did you notice that this process is a labourious and very time-consuming and exercise, plus the fact that it is somewhat unproductive to carry out this fact-finding process manually. Thus, the obvious solution is presented by relying upon the built-in software ingenuity of a database software program.

Here, the power of a query enables you to simply request that the database software only return those records that meet the “gross margin of 25 percent or more” condition. Alternatively, you can instruct the database to only list the name, unit price and stock quantity of those items that are currently selling well with your stipulated condition.

Thus far, we have learned the basic concepts behind organising the information in a database, and the process of retrieving information from a database. In order to facilitate data extraction, we need to set up the mechanisms to place information into the tables in the first place.

Here, the database software program provides two primary mechanisms to achieve this objective. The first method is to simply call-up the table in a window, and then adding the pertinent information in the correct field or column, or row. This is quite similar to adding information to a spreadsheet.

Form: Nowadays, most database software come with GUI and user-friendly forms interface that help the user to enter information in a graphical form and then, have that information transparently passed to the database. Also, there are often form wizards to help out in simplifying the task of data inputting.

Up to this point, you've learned the basic concepts that comprise the core of a database software.

Reports: Reports provide the capability to quickly produce attractively formatted summaries of the data contained one or more tables and/or queries. For instance, through the use of wizards, database users can create reports in literally a matter of minutes.

Consider the simple illustration. Suppose your company wishes to produce a catalogue to share product information with current and prospective clients. Previously, we learned that this sort of information could be retrieved from our database through the judicious use of queries.

However, recall that this information was presented in a tabular form, that is, not exactly the most attractive marketing material. Therefore, reports allow the inclusion of graphics as well as attractive formatting and pagination.

Features: One of the powerful features that greatly enhance the usability and flexibility of a database platform from which we can derive reports is the SQL or structured query language functionality. The capability of the database software to tightly integrate SQL functionality means that you can retrieve, manipulate and work with the data stored in your company's database environment. Just use the SQL utility to extract the information that you need from the database, and then tabulate it in the way you require it in terms of look-and-feel.

If you browse through any article involving databases, you will frequently come across the term "relational database". This is clearly the type of database that we discussed in the example mentioned above, for the simple reason that the data fields can correlate to each other in a logical way. Thus, relational databases are commonplace because they provide useful correlation between different data fields, as well as different tables, and so forth.

For instance, the most likely database that you will usually encounter and want to query from is a web page. This is normally executed using the Search feature where you key-in the subject or topic that you're interested about. It follows quite naturally that all of the major vendors' database software packages, e.g., Microsoft Access, SQL, Oracle, Informix, etc. are geared towards creating and managing relational databases. The management process of such databases is referred to as relational database management system (RDBMS).

Editorial Thoughtscapes
Words Impact Business Success
www.et-writer.com

About the Author

Leon A. Enriquez is managing editor and business manager of Editorial Thoughtscapes – a professional writing firm. Leon can be reached at leonenriquez@et-writer.com.

<ENDS>