ACADEMIA DE STUDII ECONOMICE - Bucureşti
Academy Of Economic Studies - Bucharest

FACULTY OF BUSINESS ADMINISTRATION
(Facultatea de Administrare a Afacerilor cu predare în limbi străine)

Informatics for Business Administration

-

Business Intelligence (BI)

By: Professor Vasile AVRAM, PhD

- suport de curs destinat studenţilor de la sectia engleză - master aprofundat-
(lecture notes for 1st year students of english division – Master in Business Administration)
- anul I - Zi -

Bucureşti 2008

COPYRIGHT© 2008-2011
All rights reserved to the author Vasile AVRAM.

Chapter 7: Business Intelligence (BI)

7.1 Definition

Keys defines Business Intelligence in [JK-2006] as "Business intelligence (BI) is a set of methodologies and technologies for gathering, storing, analyzing, and providing access to data to help users make better business decisions. The goal of BI is to transform data into information and ultimately into knowledge and wisdom. BI is usually implemented to gain sustainable competitive advantage and is a valuable core competence in some instances."

BI is defined by Gartner Group as "An interactive process for exploring and analyzing structured, domain-specific information (often stored in data warehouses) to discern business trends or patterns, thereby deriving insights and drawing conclusions. The business intelligence process includes communicating findings and effecting change. Domains include customers, suppliers, products, services and competitors." (http://www.gartner.com/6_help/glossary/GlossaryB.jsp)

Figure 7.1 shows the evolution of using computers technologies considering the main purpose of that technology.

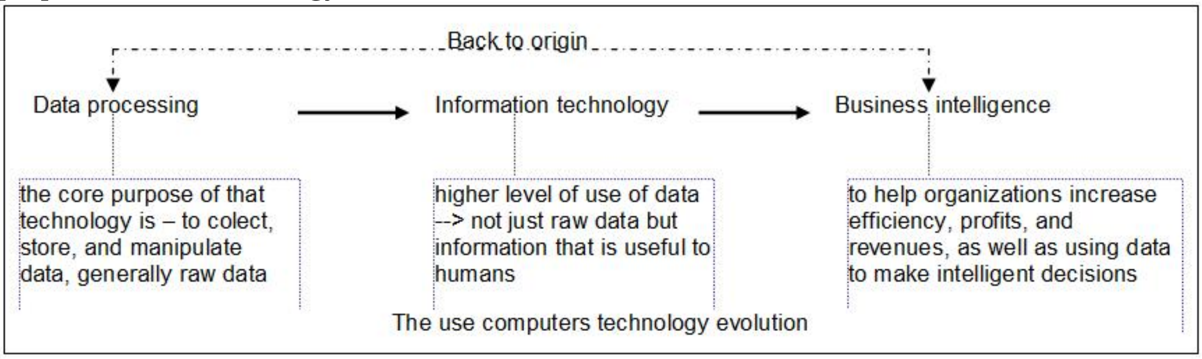

Figure 7.1 The evolution of using computers technologies

The first term used was those of Data Processing whose purpose was to collect, store, and manipulate data viewed as raw data. This evolves into the term Information Technologies (IT) that uses structured data and produces and manipulates information. That information and structured data are stored generally in data bases, data warehowses (DW). The term IT evolves into those of Business Intelligence (BI) that puts back to origins and gives access to most knowledge workers to DW/BI system. BI is gained by the timely display of data to business analysts enabling them to make the best possible business decisions.

The part of information technology that focuses on reporting and analysis currently goes by the name business intelligence (BI).

Data are the materialization, the representation of information or more simply a set of unconnected facts (facts, images, figures, sounds etc) refeared as raw data.

The information is equivalent to knowledge and has to do with the semantic aspect of the meaning of data so that it is data associated with meaning (*What?, Who?, When?, Where?*) and relates to description, definition and perspective. It is data organized, summarized, formatted, and interpreted data.

> Information in a work system can potentially take a variety of forms including numbers, text, sounds, pictures, video etc, and they can be created, modified or deleted with the system or other information can be simply received from other systems.

The information is not exactly knowledge. In fact knowledge is information associated with strategy or process (*How?*) and comprises strategy, practice, methods or approach flow. It requires intuition, contextual reasoning and evaluation, ideas, and heuristics. Wisdom embodies principle, insight, moral or archetype (*Why?*) and is biult from experience, instincts, and judgement.

Figure 7.2 shows the relationships between data, information, knowledge, and wisdom considering the context independence and understanding.

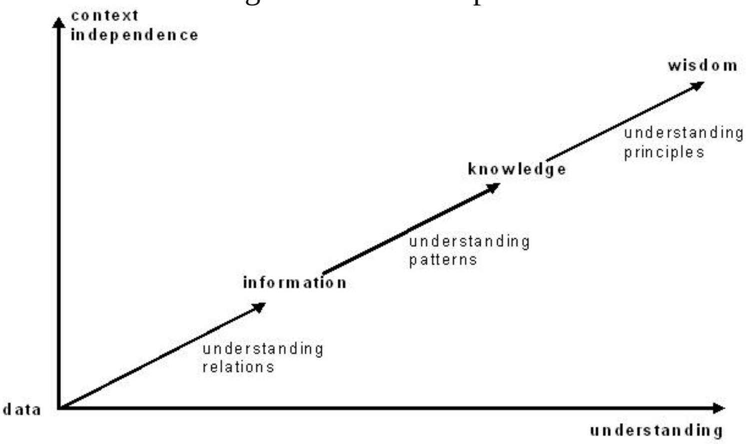

Figure 7.2 The relationships between data, information, knowledge, and wisdom considering the context independence and understanding

(Source: Neil Fleming, Coping with revolution: Will the Internet change learning? Lincoln University, Canterbury, New Zealand)

7.2 Types of BI applications

BI applications fill a critical gap in meeting the organization's range of data access needs. Figure 7.3 shows the range of reporting needs, together with the BI application types, from strategic to operational. At the top end of the information access spectrum, some users will build their own queries against the data warehouse from scratch; this is what we mean by *ad hoc access*. These business analysts, or power users, are usually charged with creating more complex analyses in an effort to address a pressing problem or opportunity. At the other end of the spectrum, operational BI occurs at the tactical level; each execution is usually low value, has short-term impact, and is predefined. Operational BI is usually leveraged over all transactions in an operational system, or across a large set of users, like all the agents in a call center. The gap between these two endpoints is wide. A large percentage of the DW/BI system user base, perhaps as high as 80 to 90 percent, will fall in the middle or toward the lower end of this spectrum. These users will need prebuilt, parameterized reports. As BI matures, organizations are finding even more ways to meet the information needs of these groups in the middle and at the lower end.

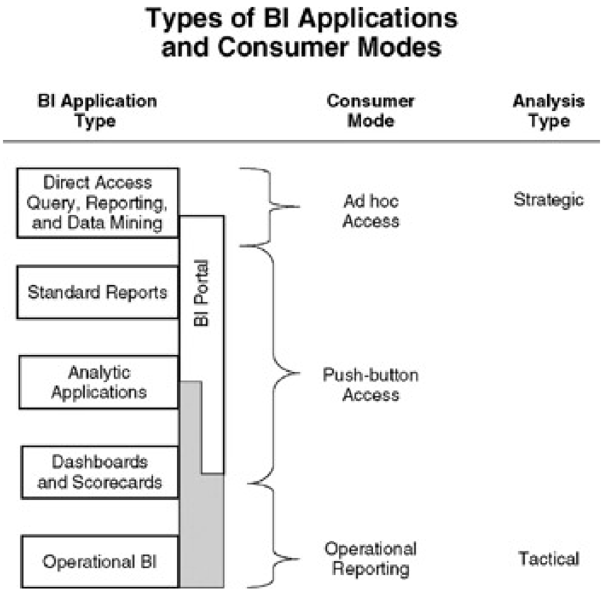

Figure 7.3 Types of BI applications

Figure 7.3 break down BI applications into several major subsets:

- Direct access query and reporting tools;
- Data mining;
- Standard reports;
- Analytic applications;
- Dashboards and scorecards;
- Operational BI and closed loop applications.

Some of the categories of tools and techniques that commonly cover the BI types outlined in figure 7.3 include:

1. Online Analytical Processing (OLAP), sometimes simply called *analytics* (based on dimensional analysis and the hypercube or “cube”);
2. Scorecarding, dashboarding, and data visualization;
3. Data warehouses (DW);
4. Data mining (DM);
5. Document warehouses;
6. Text mining;
7. Executive Information Systems (EIS);
8. Decision Support Systems (DSS).

7.2.1 Online Analytical Processing (OLAP)

OLAP takes a snapshot of a relational database and restructures it into dimensional data. An OLAP structure created from the operational data is called an *OLAP cube*. The cube is created from a star schema of tables. At the centre is the fact table, which lists the core facts that make up the query. Numerous dimension tables are linked to the fact tables. These tables indicate how the aggregations of relational data can be analyzed. The number of possible aggregations is determined by every possible manner in which the original data can be hierarchically linked.

All SAP OLAP-oriented applications do have one or more of these star schemes, which are called “InfoCubes” as sugested in figure 7.4.

The core of InfoCubes are “facts,” such as sales orders, financial postings, or work orders. Dimensions like time, place, locations, and regions surround facts. In order to complete the reporting, you can add extra information to the cube in terms of master data, texts, or hierarchies. Each item has its own database table. The InfoCubes consists of one or more fact tables, dependent on the level of compression. Each row (a fact) in this table points

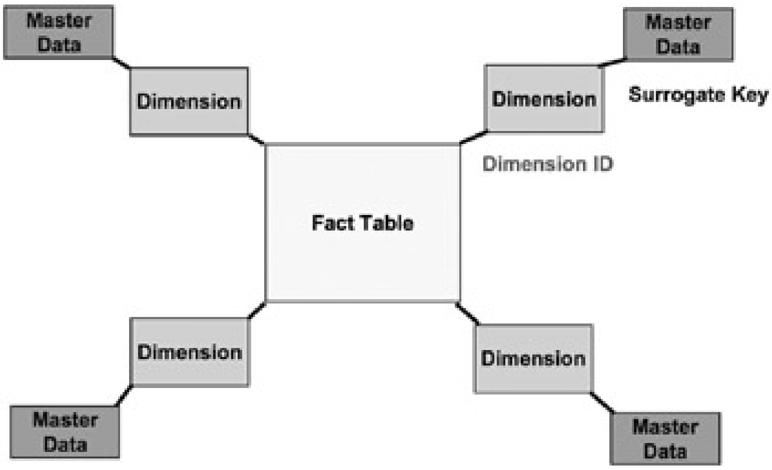

Figure 7.4 Technical layout of an SAP InfoCube (©Copyright SAP AG)

to one or more (up to a maximum of 16, a restriction of almost common databases engines) dimension tables. A row in the dimension table can point to one or more master data records. Figure 7.5 shows an example of a cube recording selling transactions, products and customers.

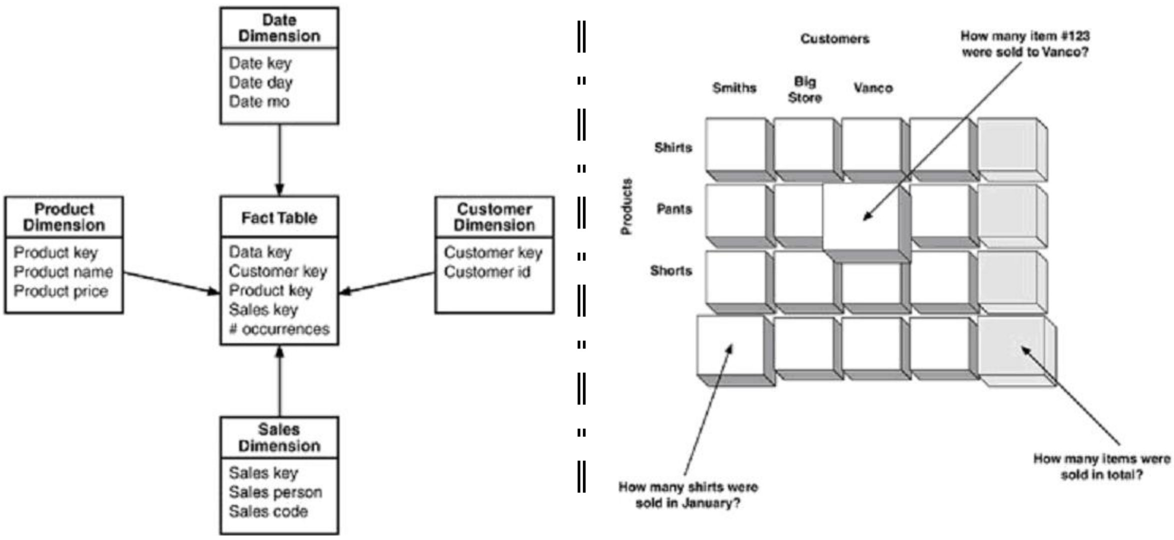

Figure 7.5 An example of an OLAP cube

The major advantage of such a data model is the query performance you can obtain from such a structure—especially when queries are based on one of the dimensions. Looking at differences between the two models, there are not really any consequences for the way upgrades are performed on such an application.

7.2.2 Scorecarding, dashboarding, and data visualization

Balanced Scorecards. In 1992, Drs. David Norton and Robert Kaplan developed a strategic management tool based on measurements of financial status, customer feedback and other outcomes, and internal process flows that illustrate the state of the business and expose areas where improvement might be desirable and named it the Balanced Scorecard. A Balanced Scorecard incorporates a feedback loop around business process outputs and the outcome of the business strategies. This double-loop feedback provides a comparison to financials

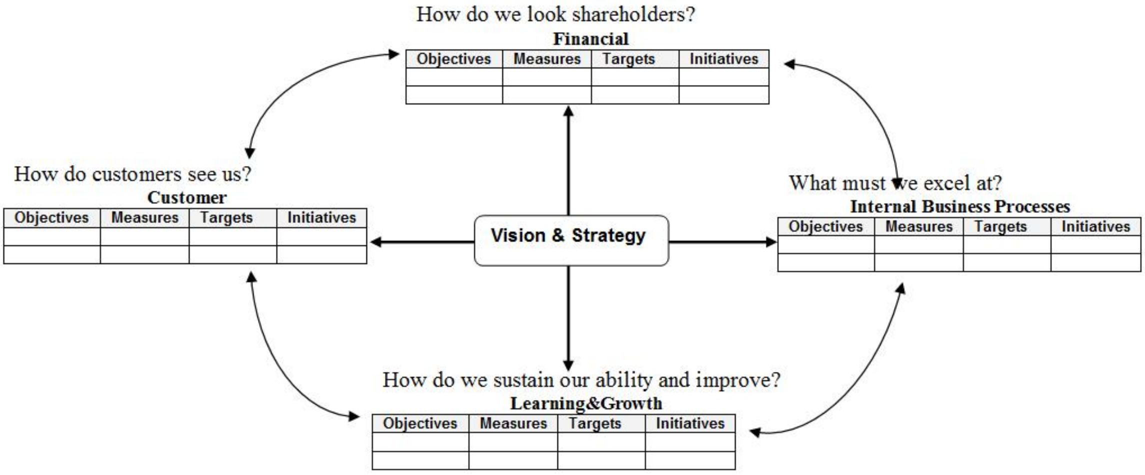

Figure 7.6 Balanced scorecard perspectives

resulting in a more balanced approach to business management.

In the balanced scorecard approach, management must first determine the organization's goals and strategy. These are used to define metrics that measure the organization's performance against its strategy from four perspectives: internal business processes, learning and growth, customer, and financial. These measures, also known as key performance indicators (KPIs) in other approaches, are then made available via the scorecard where they are tracked and trended. Everyone in the organization who has the ability to impact one or more of these measures must be able to monitor their performance. Typical metrics viewed show present status of an organization, provide diagnostic feedback and trends in performance over time, indicate which metrics are critical, and provide input for forecasting. Figure 7.6 shows an example of perspectives used for knowledge management. Each one of these perspectives defines a set of objectives, measures, targets, and initiatives to achieve the goals of that perspective.

Dashboards provide the interface tools that support status reporting and alerts across multiple data sources at a high level, and also allow drilldown to more detailed data. The alerts can be generated by comparing the actual number, like sales, to a fixed number, a percentage, or, in many cases, to a separate target or goal number. A dashboard is usually aimed at a specific user type, such as a sales manager. The sales management dashboard might include a sales by region vs. plan report with a drilldown to district and sales rep and highlights for those entities below 90 percent of plan. It might also have pipeline information including deal counts by sales cycle stage and projections of dollar sales compared to plan. Applications based on dashboard tools are particularly appealing to management because they offer the possibility of quickly identifying a problem anywhere in the business and drilling down into the detail to identify its causes.

Stephen Few defines the dashboard in [SF-2004] as “A dashboard is a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance.”

The caracteristics required by dashboards to do their job effectively can be resumed as:

> [Stephen Few; SF-2004] Dashboards have small, concise, clear, and intuitive display mechanisms. Display mechanisms that clearly state their message without taking up much space are required, so that the entire collection of information will fit into the limited real estate of a single screen. If something that looks like a fuel gauge, traffic signal, or thermometer fits this requirement best for a particular piece of information, that's what you should use, but if something else works better, you should use that instead. Insisting on sexy displays similar to those found in a car when other mechanisms would work better is counterproductive.

Dashboards are customized. The information on a dashboard must be tailored specifically to the requirements of a given person, group, or function; otherwise, it won't serve its purpose.

Dashboards can be categorized in several ways. The following table lists several variables that can be used to structure dashboard taxonomies, along with potential values for each.

Table 7.1

Variable	Values
Role	Strategic Analytical Operational
Type of data	Quantitative Non-quantitative
Data domain	Sales Finance Marketing Manufacturing Human Resources
Type of measures	Balanced Scorecard (for example, KPIs) Six Sigma Non-performance
Span of data	Enterprise-wide Departmental Individual
Update frequency	Monthly Weekly Daily Hourly Real time or near real time
Interactivity	Static display Interactive display (drill-down, filters, etc.)
Mechanisms of display	Primarily graphical Primarily text Integration of graphics and text
Portal functionality	Conduit to additional data No portal functionality

Data visualization is the attempt to display mined information in new ways. For instance, a map was once a “flat” (2D) piece of paper or parchment. As time went on, the map was overlaid onto a sphere, and we had our first globe. Data visualization hopes to take these concepts into the 21st century. From a BI perspective, data visualization, called *information visualization* in this context, uses the computer to convert data into picture form. The most basic visualization is that of turning transaction data and summary information into charts and graphs. Visualization is used in computer-aided design (CAD) to render screen images into 3D models that can be viewed from all angles and can also be animated.

Dashboards and visualization are cognitive tools that improve your "span of control" over a lot of business data. These tools help people visually identify trends, patterns and anomalies, reason about what they see and help guide them toward effective decisions. As such, these tools need to leverage people's visual capabilities. With the prevalence of scorecards, dashboards and other visualization tools now widely available for business users to review their data, the issue of visual information design is more important than ever.

7.2.3 Data warehouse

A *data warehouse* is a database designed to support decision making in an organization or enterprise. It is refreshed, or batch-updated, and can contain massive amounts of data. When the database is organized for one department or function, it is often called a *data mart* rather than a data warehouse. The data in a data warehouse is typically historical and static in nature. Data marts also contain numerous summary levels. It is structured to support a variety of elaborate analytical queries on large amounts of data that can require extensive searching.

A data warehouse is a record of an enterprise’s past transactional and operational information, stored in a database designed for efficient data analysis and reporting (especially OLAP). Two basic ideas guide the creation of a data warehouse:

1. *Integration* of data from distributed and differently structured databases, which facilitates a global overview and comprehensive analysis in the data warehouse. Periodically, one imports data from enterprise resource planning (ERP) systems and other related business software systems into the data warehouse for further processing.
2. *Separation* of data used in daily operations from data used in the data warehouse for purposes of reporting, decision support, analysis, and controlling.

Enterprise data warehouses were initially proposed as a single repository of all historic data for a company. In many companies, this approach became bogged down in the politics of creating a single data model and gaining corporate-wide approval. Inevitable delays resulted and little or no business value was delivered in a timely fashion. Where only enterprise data warehouses were created during this time, the schema design was typically third normal form enabling flexible update strategies but not providing the ease of use many business analysts expected.

By using a single data model, users can easily see across the boundaries sometimes imposed by separate functional areas of a business and executives can look back over limited historical data (since data is not typically retained for long periods in an operational system) and project into future scenarios for the entire business of the organization. And, of course, a single data model makes it much easier for Oracle applications to deliver a standardized set of reports and metrics for business analysts and executives.

The alternative exercised in many companies was the building of data marts as departmental initiatives. These marts could be more quickly deployed when driven by local business needs, but they were often designed and deployed without any inter-departmental coordination. Questions that crossed two departments could be asked of separated data marts and two different answers would often be obtained. Data marts were typically deployed with star schema enabling business users to more easily submit ad hoc queries and perform analyses.

A solution that emerged in the late 1990s to address the downside of both of these deployment models was the incremental building of an enterprise data warehouse at the same time dependent data marts were built using consistent data definitions. The marts were often fed from the enterprise data warehouse that became the single source of truth. Figure 7.7 is a representation of this commonly deployed business intelligence topology where an operational data store is also present as a gathering point of transactional data.

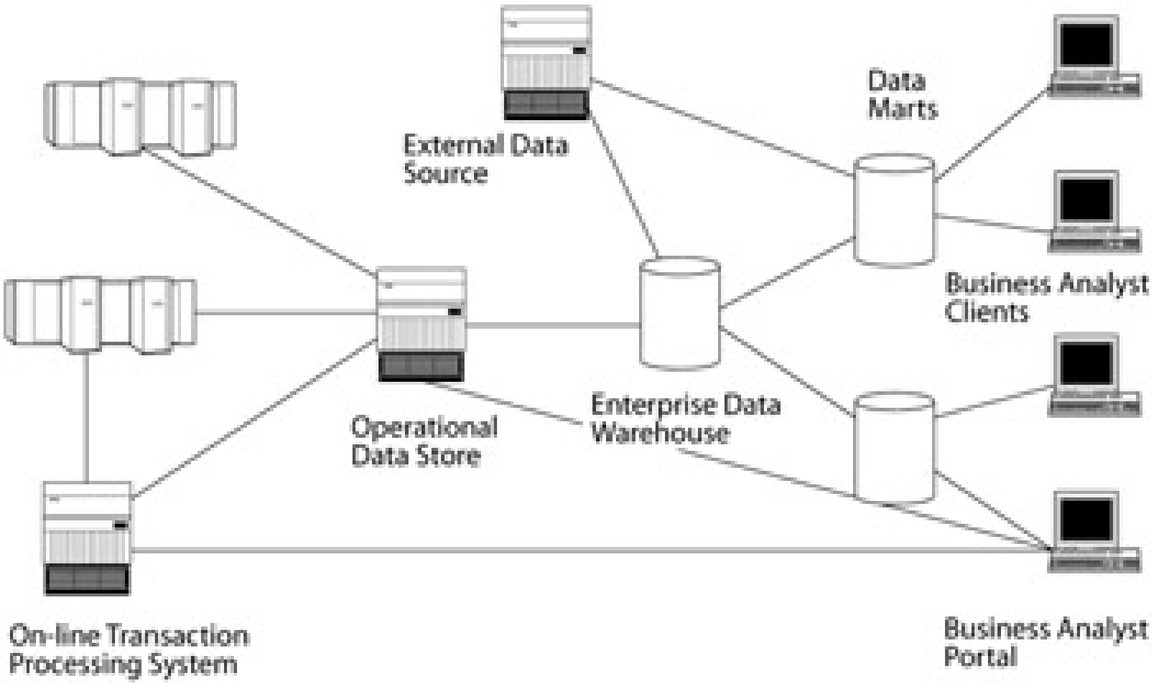

Figure 7.7 Topology that includes an enterprise data warehouse and marts (Source: Oracle Data Warehousing and Business Intelligence Solutions by Robert Stackowiak, Joseph Rayman and Rick Greenwald John Wiley & Sons © 2007)

Data warehouses are typically deployed to isolate the potential performance impact of complex unplanned queries operating on large amounts of data, or to create different data designs to efficiently service these queries, or to store a consolidated cleansed source of data apart from the different sources. By their nature, classic data warehouses often have somewhat older data than transactional systems, with data loads ranging from every minute to hourly to daily to weekly.

The warehouse is no special technology in itself. The *data warehouse database* is a relational data structure that is optimized for distribution. It collects and stores integrated sets of historical, nonvolatile data from multiple operational systems and feeds them to one or more data marts. It becomes the one source of truth for all shared data.

The easiest way to conceptually view a data mart is that a mart needs to be an extension of the data warehouse. Data is integrated as it enters the data warehouse from multiple legacy sources. Data marts then derive their data from the central data warehouse source. The theory is that no matter how many data marts are created, all the data is drawn from the one and only version of the truth, which is the data contained in the warehouse. Distribution of the data from the warehouse to the mart provides the opportunity to build new summaries to fit a particular department's need. The data marts contain subject-specific information supporting the requirements of the end users in individual business units. Data marts can provide rapid response to end-user requests if most queries are directed to precomputed, aggregated data stored in the data mart.

Today, many companies are leveraging the advanced data warehousing features in databases such as Oracle to consolidate the data mart schema into the same database as the enterprise data warehouse. The resulting hybrid schema is the third normal form for the detailed data and star schema or OLAP cubes for the summary-level data. Oracle added performance features such as extensive parallelism, static bitmap indexes, advanced star join techniques, materialized views and embedded analytic functions, multi-dimensional cubes (OLAP), and data mining. As the databases have grown in size and complexity, Oracle developed more extensive management and self-management capabilities including Enterprise Manager's specific features for data warehousing management, automatic degree of parallelism, memory allocation at query time, a database resource manager, table compression, the Automatic Database Diagnostics Monitor (ADDM), and the Automatic Storage Manager (ASM). Of course, as data warehouses obtain their data from other sources, Oracle added and optimized many data movement capabilities within the database including SQL*Loader for direct path loading, extraction, transformation, and loading (ETL) extensions to SQL, transportable tablespaces, and Streams for advanced queuing and replication.

7.2.4 Data mining

Data mining can help us not only in knowledge discovery (i.e., the identification of new phenomena) but also in enhancing our understanding of known phenomena. One of the key steps in data mining is pattern recognition, namely, the discovery and characterization of patterns in image and other high-dimensional data. A pattern is defined as an arrangement or an ordering in which some organization of underlying structure can be said to exist. Patterns in data are identified using measurable features or attributes that have been extracted from the data.

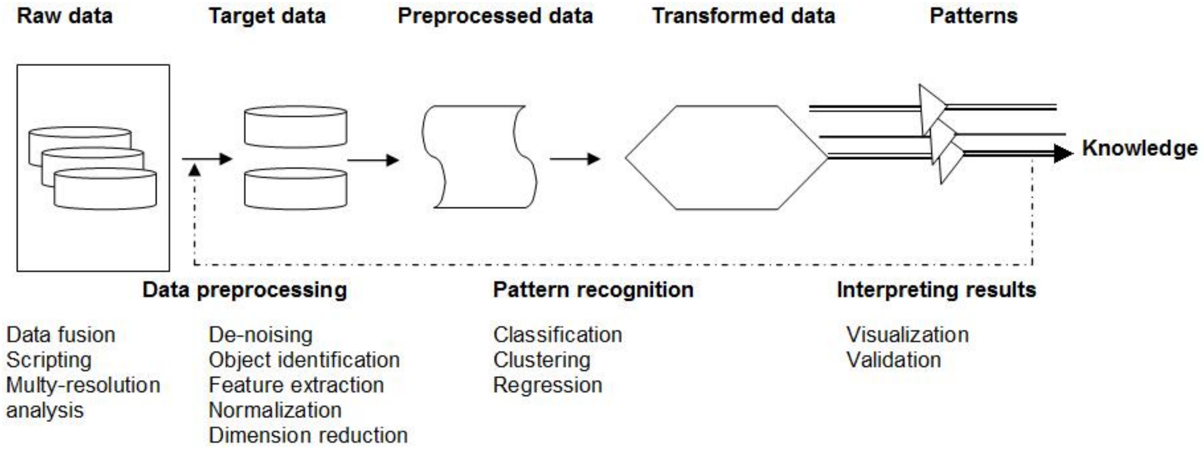

Figure 7.8 The process of data mining

"Data mining is a process of data exploration with the intent to find patterns or relationships that can be made useful to the organization. Data mining takes advantage of a range of technologies and techniques for exploration and execution. From a business perspective, data mining helps you understand and predict behavior, identify relationships, or group items, such as customers or products, into coherent sets. The resulting models can take the form of rules or equations that you apply to new customers, products, or transactions to better understand how you should respond to them. These models can then be accessed from analytic applications or operational BI applications to provide guidance, alerts, or indicators to the user.” [RK et al.-2008]

Data mining is an interactive and iterative process involving data preprocessing, search for patterns, knowledge evaluation, and possible refinement of the process based on input from domain experts or feedback from one of the steps as sugested in figure 7.8.

The preprocessing of the data is a time-consuming but critical first step in the data-mining process. It is often domain and application dependent; however, several techniques developed in the context of one application or domain can be applied to other applications and domains as well. The pattern-recognition step is usually independent of the domain or application.

Data mining starts with the raw data, which usually takes the form of simulation data, observed signals, or images. These data are preprocessed using various techniques, such as sampling, multiresolution analysis, denoising, feature extraction, and normalization.

Sampling is a widely accepted technique to reduce the size of the dataset and make it easier to handle. However, in some cases, such as when looking for something that appears infrequently in the set, sampling may not be viable. Another technique to reduce the size of the dataset is multiresolution analysis. Data at a fine resolution can be “coarsened,” which shrinks the dataset by removing some of the detail and extracts relevant features from the raw dataset. In credit card fraud, for instance, an important feature might be the location where a card is used. Thus, if a credit card is suddenly used in a country in which it has never been used before, fraudulence seems likely. Thus, the key to effective data mining is reducing the number of features used to mine data, so that only the features best at discriminating among the data items are retained.

Once the data is preprocessed or “transformed,” pattern-recognition software is used to look for patterns. *Patterns* are defined as an ordering that contains some underlying structure. The results are processed back into a format familiar to the experts, who then can examine and interpret the results.

To be truly useful, data-mining techniques must be scalable. In other words, when the problem increases in size, we do not want the mining time to increase proportionally. Making the end-to-end process scalable can be very challenging, because it is not just a matter of scaling each step but of scaling the process as a whole.

Large-scale data mining is a field very much in its infancy, making it a source of several open research problems. To extend data-mining techniques to large-scale data, several barriers must be overcome. The extraction of key features from large, multidimensional, complex data is a critical issue that must be addressed first, prior to the application of the pattern-recognition algorithms. The features extracted must be relevant to the problem, insensitive to small changes in the data, and invariant to scaling, rotation, and translation. In addition, we need to select discriminating features through appropriate dimension reduction techniques. The pattern-recognition step poses several challenges as well. For example, is it possible to modify existing algorithms or design new ones that are scalable, robust, accurate, and interpretable? Further, can these algorithms be applied effectively and efficiently to complex, multidimensional data? And, is it possible to implement these algorithms efficiently on large-scale multiprocessor systems so that a scientist can interactively explore and analyze the data?

The six steps of effective data mining are:

1. *Business understanding:* Understanding the specific business problem’s objectives and goals
2. *Data understanding:* Initial data collection and familiarization, identifying data quality problems, and identifying data subsets

3. *Data preparation:* Selecting and clearing final set of data
4. *Modeling:* Selecting and applying modeling techniques
5. *Evaluation:* Determining whether the result meets the business problem's specifications
6. *Deployment:* Presenting the data in a meaningful way so that an end user can use it

7.3 Analytic Cycle for Business Intelligence

Business analysis follows a common process from monitoring activity to identifying a problem or opportunity and determining an action to take, and finally back to monitoring the results of that action. The analytic cycle breaks the process into five distinct stages as shown in figure 7.9.

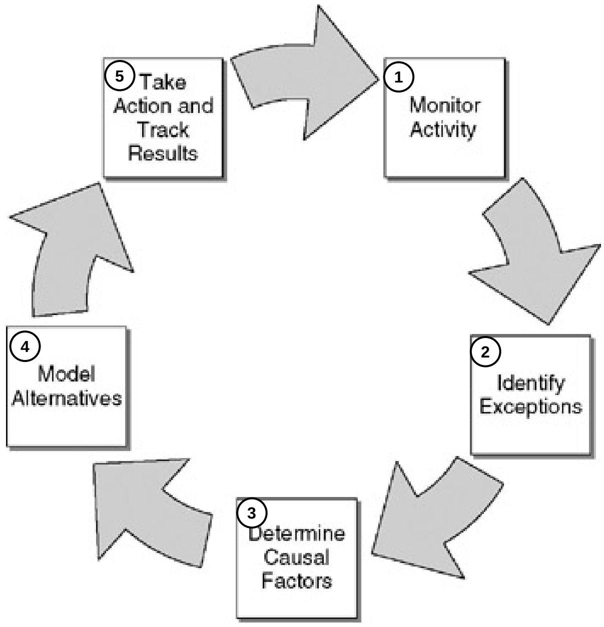

Figure 7.9 Analytic cycle for BI analysis

Stage 1: Monitor Activity - Users work with standard reports to examine current results versus previous periods or plan in order to provide a report card on the state of the business. BI application requirements in the monitor activity stage focus on the presentation layer and include technologies such as *dashboards, portals, and scorecards*.

Stage 2: Identify Exceptions - Focuses on the identification of "what's the matter?" or "where are the problems?" (identifies the exceptions to normal performance, the opportunities etc). The identify exceptions stage requires additional capabilities, such as distribution servers that distribute alerts to users' devices of choice based upon exception triggers and visualization tools to view the data in different more creative ways, including trend lines, spark lines, geographical maps, or clusters.

Stage 3: Determine Causal Factors - Understand the "why" or root causes of the identified exceptions. Identifying reliable relationships and interactions between variables that drive exceptional performance is the key. Successfully supporting users' efforts in this stage will require your DW/BI system architecture to include additional software including statistical tools and/or data mining algorithms, such as association, sequencing, classification, and segmentation, to quantify cause-and-effect. This step frequently requires new data to investigate causal factors.

Stage 4: Model Alternatives - Develop models for evaluating decision alternatives based on cause-and-effect relationships. The system must have the ability to perform what-if analysis and simulations on a range of potential decisions. Ther data warehouse architecture may also need to accommodate additional technologies in the model alternatives stage, including statistical tools and data mining algorithms for model evaluation, such as sensitivity analysis, Monte Carlo simulations, and goal seeking optimizations.

Stage 5: Take Action and Track Results - Places additional demands on data warehouse architecture such as access to operational systems via APIs or SOAP calls. The system must record the results of specific business decisions and determine which decisions worked and which ones didn't. Technologies applicable in this area include distribution services that enable users to respond with recommended actions from email, PDAs, pagers, or cell phones, not just deliver alerts. The results of the action should be captured and analyzed in order to continuously fine tune the analysis process, business rules, and analytic models. This brings you right back to the monitor activity stage to start the cycle all over again.

The environment needs to proactively "guide" users through the analysis of a business situation, ultimately helping them make an insightful and thoughtful decision. The goals are to:

- Proactively guide business users beyond basic reporting.
- Identify and understand exceptional performance situations.
- Capture decision-making "best practices" for each exceptional performance situation.
- Share the resulting "best practices" or intellectual capital across the organization.
- Understanding the analytic cycle helps ensure that you provide a complete solution to each analytic opportunity identified in the business requirements definition process.

References

Books

1. [RK et al.-2008] The Data Warehouse Lifecycle Toolkit, Second Edition by Ralph Kimball et al. John Wiley & Sons © 2008
2. [WLVS-2007] Enterprise Architecture and Integration: Methods, Implementation and Technologies , by Wing Lam and Venky Shankararaman (eds), IGI Publishing © 2007 (364 pages) ISBN:9781591408871
3. [VS-2008] Intelligent Information Technologies: Concepts, Methodologies, Tools, and Applications, by Vijayan Sugumaran (ed), IGI Publishing © 2008 (2469 pages) ISBN:9781599049410
4. [RSJRRG-07] Oracle Data Warehousing and Business Intelligence Solutions by Robert Stackowiak, Joseph Rayman and Rick Greenwald John Wiley & Sons © 2007
5. [JK-2006] Knowledge Management, Business Intelligence, and Content Management: The IT Practitioner’s Guide by Jessica Keyes Auerbach Publications © 2006 (500 pages) ISBN:9780849393853
6. [DWJW-04] SAP NetWeaver For Dummies, by Dan Woods and Jeffrey Word, John Wiley & Sons © 2004 (414 pages)
7. [Sens-08] Upgrading SAP, by M. Sens, Infinity Science Press © 2008 (344 pages), ISBN:9781934015155
8. [SF-2004] Information Dashboard Design, by Stephen Few, O'Reilly, 2004, ISBN: 0-596-10016-7

On line resources

1. www.oracle.com
2. www.sap.com
3. www.ibm.com
4. www.sas.com
5. www.gartner.com