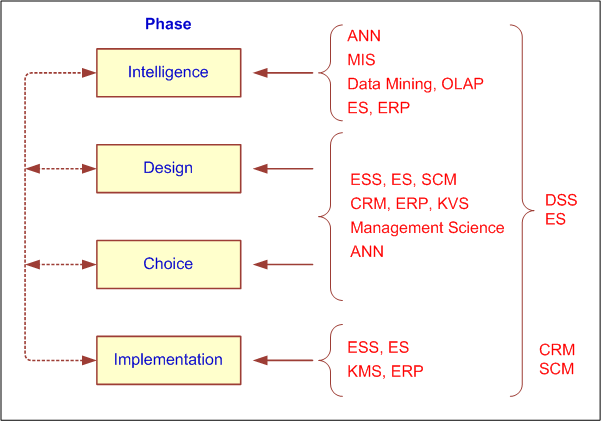
**How Decisions are Supported?**



Here we relate specific technologies to the decisionmaking process. Databases, data marts, and especially data warehouses are important technologies in supporting all phases of decision making. They provide the data that drive decision making.

**Support for the Intelligence Phase**

The primary requirement of decision support for the intelligence phase is the ability to scan external and internal information sources for opportunities and problems and to interpret what the scanning discovers. Web tools and sources are extremely useful for environmental scanning. Web browsers provide useful front ends for a variety of tools, from OLAP to data mining and data warehouses. Data sources can be internal or external. Internal sources may be accessible through a corporate intranet. External sources are many and varied.

Decision support/BI technologies can be very helpful. For example, a data warehouse can support the intelligence phase by continuously monitoring both internal and external information, looking for early signs of problems and opportunities through a Web-based enterprise information portal (also called a dashboard). Similarly, (automatic) data (and Web) mining (which may include expert systems [ES], CRM, genetic algorithms, neural networks, and other analytics systems) and (manual) OLAP also support the intelligence phase by identifying relationships among activities and other factors.

Geographic information systems (GIS) can be utilized either as stand-alone systems or integrated with

these systems so that a decision maker can determine opportunities and problems in a spatial sense. These relationships can be exploited for competitive advantage (e.g., CRM identifies classes of customers to approach with specific products and services). A KMS can be used to identify similar past situations and how they were handled. GSS can be used to share information and for brainstorming.

Another aspect of identifying internal problems and capabilities involves monitoring the current status of operations. When something goes wrong, it can be identified quickly and the problem can be solved. Tools such as business activity monitoring (BAM), business process management (BPM), and product life-cycle management (PLM) provide such capability to decision makers. Both routine and ad hoc reports can aid in the intelligence phase. For example, regular reports can be designed to assist in the problem-finding activity by comparing expectations with current and projected performance. Web-based OLAP tools are excellent at this task. So are visualization tools and electronic document management systems.

Expert systems (ES), in contrast, can render advice regarding the nature of a problem, its classification, its seriousness, and the like. ES can advise on the suitability of a solution approach and the likelihood of successfully solving the problem. One of the primary areas of ES success is interpreting information and diagnosing problems. This capability can be exploited in the intelligence phase. Even intelligent agents can be used to identify opportunities.

Much of the information used in seeking new opportunities is qualitative, or soft. This indicates a high level of unstructuredness in the problems, thus making DSS quite useful in the intelligence phase.

**Support for the Design Phase**

The design phase involves generating alternative courses of action, discussing the criteria for choices and their relative importance, and forecasting the future consequences of using various alternatives. Several of these activities can use standard models provided by a DSS (e.g., financial and forecasting models, available as applets). Alternatives for structured problems can be generated through the use of either standard or special models.

However, the generation of alternatives for complex problems requires expertise that can be provided only by a human, brainstorming software, or an ES. OLAP and data mining software are quite useful in identifying relationships that can be used in models. Most DSS have quantitative analysis capabilities, and an internal ES can assist with qualitative methods as well as with the expertise required in selecting quantitative analysis and forecasting models. A KMS should certainly be consulted to determine whether such a problem has been encountered before or whether there are experts on hand who can provide quick understanding and answers. CRM systems, revenue management systems, ERP, and SCM systems software are useful in that they provide models of business processes that can test assumptions and scenarios. If a problem requires brainstorming to help identify important issues and options, a GSS may prove helpful. Tools that provide cognitive mapping can also help. Several Web-based tools that provide decision support, mainly in the design phase, by providing models and reporting of alternative results. Each of their cases has saved millions of dollars annually by utilizing these tools. Such DSS are helping engineers in product design as well as decision makers solving business problems.

**Support for the Choice Phase**

In addition to providing models that rapidly identify a best or good-enough alternative, a DSS can support the choice phase through what-if and goal-seeking analyses. Different scenarios can be tested for the selected option to reinforce the final decision. Again, a KMS helps identify similar past experiences; CRM, ERP, and SCM systems are used to test the impacts of decisions in establishing their value, leading to an intelligent choice. An ES can be used to assess the desirability of certain solutions as well as to recommend an appropriate solution. If a group makes a decision, a GSS can provide support to lead to consensus.

**Support for the Implementation Phase**

This is where “making the decision happen” occurs. The DSS benefits provided during implementation may be as important as or even more important than those in the earlier phases. DSS can be used in implementation activities such as decision communication, explanation, and justification.

Implementation-phase DSS benefits are partly due to the vividness and detail of analyses and reports. For example, one chief executive officer (CEO) gives employees and external parties not only the aggregate financial goals and cash needs for the near term, but also the calculations, intermediate results, and statistics used in determining the aggregate figures. In addition to communicating the financial goals unambiguously, the CEO signals other messages. Employees know that the CEO has thought through the assumptions behind the financial goals and is serious about their importance and attainability. Bankers and directors are shown that the CEO was personally involved in analyzing cash needs and is aware of and responsible for the implications of the financing requests prepared by the finance department. Each of these messages improves decision implementation in some way.

Reporting systems and other tools variously labeled as BAM, BPM, KMS, EIS, ERP, CRM, and SCM are all useful in tracking how well an implementation is working. GSS is useful for a team to collaborate in establishing implementation effectiveness. For example, a decision might be made to get rid of unprofitable customers. An effective CRM can identify classes of customers to get rid of, identify the impact of doing so, and then verify that it really worked that way. All phases of the decision-making process can be supported by improved communication through collaborative computing via GSS and KMS. Computerized systems can facilitate communication by helping people explain and justify their suggestions and opinions.

Management is a process by which organizational goals are achieved by using resources. The resources are considered inputs, and attainment of goals is viewed as the output of the process. The degree of success of the organization and the manager is often measured by the ratio of outputs to inputs. This ratio is an indication of the organization’s *productivity*, which is a reflection of the *organizational and managerial performance*.

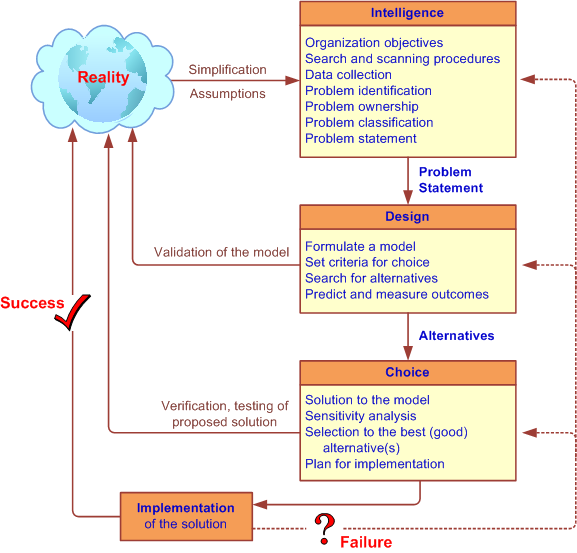
The level of productivity or the success of management depends on the performance of managerial functions, such as planning, organizing, directing, and controlling. To perform their functions, managers engage in a continuous process of making decisions. Making a decision means selecting the best alternative from two or more solutions.

The process of decision making:

* Defining the problem (a decision situation that may deal with some difficulty or with an opportunity)
* Constructing a model that describes the real-world problem
* Identifying possible solutions to the modeled problem and evaluating the solutions
* Comparing, choosing, and recommending a potential solution to the problem

Decision making is difficult, because:

* Technology, information systems, advanced search engines, and globalization result in more and more alternatives from which to choose
* Government regulations and the need for compliance, political instability and terrorism, competition, and changing consumer demands produce more uncertainty, making it more difficult to predict consequences and the future
* Other factors are the need to make rapid decisions, the frequent and unpredictable changes that make trial-and-error learning difficult, and the potential costs of making mistakes



**Phases of the Decision-Making Process**

There is a continuous flow of activity from intelligence to design to choice (see the bold lines in Figure ), but at any phase, there may be a return to a previous phase (feedback). Modeling is an essential part of this process. The seemingly chaotic nature of following a haphazard path from problem discovery to solution via decision making can be explained by these feedback loops.

**Decision Making: The Intelligence Phase**

Intelligence in decision making involves scanning the environment, either intermittently or continuously. It includes several activities aimed at identifying problem situations or opportunities. It may also include monitoring the results of the implementation phase of a decision-making process.

**Problem (or Opportunity) Identification**

The intelligence phase of decision-making begins by identifying organizational goals and evaluating if they are being met, often triggered by dissatisfaction with the status quo. This dissatisfaction arises from a gap between desired and actual outcomes. In this phase, decision-makers examine if a genuine problem exists by identifying symptoms, assessing their extent, and defining the problem clearly. Often, what appears to be a problem (e.g., high costs) may only be a symptom (e.g., poor inventory management). By analyzing productivity and gathering real data, organizations can clarify if an issue exists and differentiate between underlying problems and symptomatic indicators.

1. **Problem classification** is the conceptualization of a problem in an attempt to place it in a definable category, possibly leading to a standard solution approach. An important approach classifies problems according to the degree of structuredness evident in them. This ranges from totally structured (i.e., programmed) to totally unstructured (i.e., unprogrammed).
2. **Problem Decomposition**

Decomposition simplifies complex problems by dividing them into manageable subproblems, often revealing structure within seemingly unstructured issues. This approach aids in understanding, communication, and collaboration among decision-makers. In DSS development, decomposition enhances problem structure and supports the analytical hierarchy process, leading to more effective, informed decisions.

1. P**roblem Ownership**

In the intelligence phase, establishing problem ownership ensures responsibility for addressing an issue. If no ownership is assigned, the problem remains unaddressed or unidentified. Someone must step forward to take responsibility or designate ownership.The assignment of authority to solve the problem is called **problem ownership**. For example, a manager may feel that he or she has a problem because interest rates are too high. Because interest rate levels are determined at the national and international levels, and most managers can do nothing about them, high interest rates are the problem of the government, not a problem for a specific company to solve.

The intelligence phase ends with a formal problem statement.

**Decision Making: The Design Phase**

The decision-making process begins with the intelligence phase, where the problem is identified and ownership is assigned. In the design phase, a simplified model of the system is built, assumptions are made, and variables are related. This model aids in evaluating and identifying alternative solutions.The design phase involves finding or developing and analyzing possible courses of action. These include understanding the problem and testing solutions for feasibility. A model of the decision-making problem is constructed, tested, and validated.Modeling combines science and art. Scientifically, it uses established model classes to fit situations; artistically, it requires creativity to choose suitable assumptions and integrate model features for valid solutions. Models include decision variables representing alternatives, result variables defining objectives (like profit), and uncontrollable variables describing the environment. Modeling establishes mathematical or symbolic relationships among these variables to guide decision-making.

**Decision Making:Selection of a Principle of Choice**

A **principle of choice** is a criterion that describes the acceptability of a solution approach. The choice phase in decision-making involves selecting a specific course of action based on evaluating alternatives. It includes searching for, assessing, and recommending a solution, though refining options may overlap with the design phase. Solutions are judged for viability and profitability. Notably, solving the model provides a recommended action, but solving the actual problem requires implementing this recommendation effectively. Search approaches involve analytical methods, algorithms, heuristics, or logical trial-and-error.Each alternative should be evaluated carefully, especially if it has multiple goals that need balancing. Sensitivity analysis tests the stability of an alternative by assessing if small parameter changes affect outcomes. What-if analysis explores major shifts in parameters, while goal seeking identifies decision variable values to meet specific objectives effectively.

**Decision Making: the Implementation Phase**

The implementation phase involves applying a recommended solution, not necessarily a computer system. Key issues include managing resistance to change, top management support, and user training—crucial for successful information system-supported decisions. Previous initiatives like business process reengineering (BPR) and knowledge management faced challenges due to change management issues. Recognizing change management’s complexity, the importance of project management has risen, with certifications growing as companies value structured approaches in implementation success.Implementation should include data collection and analysis to refine future decisions. Analytics in feedback processes ensures valid data for problem identification, especially vital in public policy decisions.

**Data Integration and the Extraction, Transformation, and Load (ETL) Process**

**Data integration** comprises three major processes that, when correctly implemented, permit data to be accessed and made accessible to an array of ETL and analysis tools and the data warehousing environment: data access (i.e., the ability to access and extract data from any data source), data federation (i.e., the integration of business views across multiple data stores), and change capture (based on the identification, capture, and delivery of the changes made to enterprise data sources).

A major purpose of a data warehouse is to integrate data from multiple systems. Various integration technologies enable data and metadata integration:

• Enterprise application integration (EAI)

• Service-oriented architecture (SOA)

• Enterprise information integration (EII)

• Extraction, transformation, and load (ETL)

**Enterprise application integration (EAI)** provides a vehicle for pushing data from source systems into the data warehouse. It involves integrating application functionality and is focused on sharing functionality (rather than data) across systems, thereby enabling flexibility and reuse. Traditionally, EAI solutions have focused on enabling application reuse at the application programming interface (API) level. Recently, EAI is accomplished by using SOA coarse-grained services (a collection of business processes or functions) that are well defined and ocumented. Using Web services is a specialized way of implementing an SOA. EAI can be used to facilitate data acquisition directly into a near–real-time data warehouse or to deliver decisions to the OLTP systems. There are many different approaches to and tools for EAI implementation. **Enterprise information integration (EII)** is an evolving tool space that promises real-time data integration from a variety of sources, such as relational databases, Web services, and multidimensional databases. It is a mechanism for pulling data from source systems to satisfy a request for information. EII tools use predefined metadata to populate views that make integrated data appear relational to end users. XML may be the most important aspect of EII because XML allows data to be tagged either at creation time or later. These tags can be extended and modified to accommodate almost any area of knowledge.Physical data integration has conventionally been the main mechanism for creating an integrated view with data warehouses and data marts.With the advent of EII tools new virtual data integration patterns are feasible.

**Extraction, Transformation, and Load**



At the heart of the technical side of the data warehousing process is **extraction, transformation, and load (ETL)**. ETL technologies, which have existed for some time, are instrumental in the process and use of data warehouses. The ETL process is an integral component in any data-centric project. IT managers are often faced with challenges because the ETL process typically consumes 70 percent of the time in a data-centric project. The ETL process consists of extraction (i.e., reading data from one or more databases), transformation (i.e., converting the extracted data from its previous form into the form in which it needs to be so that it can be placed into a data warehouse or simply another database), and load (i.e., putting the data into the data warehouse). Transformation occurs by using rules or lookup tables or by combining the data with other data. The three database functions are integrated into one tool to pull data out of one or more databases and place them into another, consolidated database or a data warehouse. ETL tools also transport data between sources and targets, document how data elements (e.g., metadata) change as they move between source and target, exchange metadata with other applications as needed, and administer all runtime processes and operations (e.g., scheduling, error management, audit logs, statistics). ETL is extremely important for data integration as well as for data warehousing. The purpose of the ETL process is to load the warehouse with integrated and cleansed data. The data used in ETL

processes can come from any source: a mainframe application, an ERP pplication, a CRM tool, a flat file, an Excel spreadsheet, or even a message queue. The process of migrating data to a data warehouse involves the extraction of data

from all relevant sources. Data sources may consist of files extracted from OLTP databases, spreadsheets, personal databases (e.g., Microsoft Access), or external files. Typically, all the input files are written to a set of staging tables, which are designed to facilitate the load process. A data warehouse contains numerous business rules that define such things as how the data will be used, summarization rules, standardization of encoded attributes, and calculation rules. Any data quality issues pertaining to the source files need to be corrected before the data are loaded into the data warehouse. One of the benefits of a well-designed data warehouse is that these rules can be stored in a metadata repository and applied to the data warehouse centrally. This differs from an OLTP approach, which typically has data and business rules scattered throughout the system. The process of loading data into a data warehouse can be performed either through data transformation tools that provide a GUI to aid in the development and maintenance of business rules or through more traditional methods, such as developing programs or utilities to load the data warehouse, using programming languages such as PL/SQL, C++, Java, or .NET Framework languages. This decision is not easy for organizations. Several issues affect

whether an organization will purchase data transformation tools or build the transformation process itself:

• Data transformation tools are expensive.

• Data transformation tools may have a long learning curve.

• It is difficult to measure how the IT organization is doing until it has learned to use

the data transformation tools.

In the long run, a transformation-tool approach should simplify the maintenance of an organization’s data warehouse. Transformation tools can also be effective in detecting and scrubbing (i.e., removing any anomalies in the data). OLAP and data mining tools rely on how well the data are transformed.

The following are some of the important criteria in selecting an ETL tool:

• Ability to read from and write to an unlimited number of data source architectures

• Automatic capturing and delivery of metadata

• A history of conforming to open standards

• An easy-to-use interface for the developer and the functional user

Performing extensive ETL may be a sign of poorly managed data and a

fundamental lack of a coherent data management strategy. Karacsony (2006) indicated that there is a direct correlation between the extent of redundant data and the number of ETL processes. When data are managed correctly as an enterprise asset, ETL efforts are significantly reduced, and redundant data are completely eliminated. This leads to huge savings in maintenance and greater efficiency in new development while also improving data quality. Poorly designed ETL processes are costly to maintain, change, and update. Consequently, it is crucial to make the proper choices in terms of the technology and tools to use for developing and maintaining the ETL process.

A number of packaged ETL tools are available. Database vendors currently offer ETL capabilities that both enhance and compete with independent ETL tools. SAS acknowledges the importance of data quality and offers the industry’s first fully integrated solution that merges ETL and data quality to transform data into strategic valuable assets. Other ETL software providers include Microsoft, Oracle, IBM, Informatica, Embarcadero, and Tibco.

Simple Taxonomy for Data Mining Tasks.



A Simple Taxonomy for Data Mining Tasks.

**Classification**, or supervised induction, is perhaps the most common of all data mining tasks. The objective of classification is to analyze the historical data stored in a database and automatically generate a model that can predict future behavior. This induced model consists of generalizations over the records of a training data set, which help distinguish predefined classes. The hope is that the model can then be used to predict the classes of other unclassified records and, more importantly, to accurately predict actual future events. Common classification tools include neural networks and decision trees (from machine learning), logistic regression and discriminant analysis (from traditional statistics), and emerging tools such as rough sets, support vector machines, and genetic algorithms.

Statistics-based classification techniques (e.g., logistic regression and discriminant analysis) have received their share of criticism—that they make unrealistic assumptions about the data, such as independence and normality—which limit their use in classification-type data mining projects.

Neural networks involve the development of mathematical structures (somewhat resembling the biological neural networks in the human brain) that have the capability to learn from past experiences presented in the form of well-structured data sets. They tend to be more effective when the number of variables involved is rather large and the relationships among them are complex and imprecise. Neural networks have disadvantages as well as advantages. For example, it is usually very difficult to provide a good rationale for the predictions made by a neural network. Also, neural networks tend to need considerable training. Unfortunately, the time needed for training tends to increase exponentially as the volume of data increases, and, in general, neural networks cannot be trained on very large databases. These and other factors have limited the applicability of neural networks in data-rich domains.

Decision trees classify data into a finite number of classes based on the values of the input variables. Decision trees are essentially a hierarchy of if-then statements and are thus significantly faster than neural networks. They are most appropriate for categorical and interval data. Therefore, incorporating continuous variables into a decision tree framework requires *discretization,* that is, converting continuous valued numerical variables to ranges and categories.

A related category of classification tools is rule induction. Unlike with a decision tree, with rule induction the if-then statements are induced from the training data directly, and they need not be hierarchical in nature. Other, more recent techniques such as SVM, rough sets, and genetic algorithms are gradually finding their way into the arsenal of classification algorithms.

**Clustering** partitions a collection of things (e.g., objects, events, etc., presented in a structured data set) into segments (or natural groupings) whose members share similar characteristics. Unlike classification, in clustering the class labels are unknown. As the selected algorithm goes through the dataset,identifying the commonalities of things based on their characteristics, the clusters are established. Because the clusters are determined using a heuristic-type algorithm, and because different algorithms may end up with different sets of clusters for the same data set, before the results of clustering techniques are put to actual use it may be necessary for an expert to interpret,and potentially modify, the suggested clusters. After reasonable clusters have been identified, they can be used to classify and interpret new data.

Not surprisingly, clustering techniques include optimization. The goal of clustering is to create groups so that the members within each group have maximum similarity and the members across groups have minimum similarity. The most commonly used clustering techniques include *k*-means (from statistics) and self-organizing maps (from machine learning), which is a unique neural network architecture developed by Kohonen. .

Firms often effectively use their data mining systems to perform market segmentation with cluster analysis. Cluster analysis is a means of identifying classes of items so that items in a cluster have more in common with each other than with items in other clusters. It can be used in segmenting customers and directing appropriate marketing products to the segments at the right time in the right format at the right price. Cluster analysis is also used to identify natural groupings of events or objects so that a common set of characteristics of these groups can be identified to describe them.

**Associations**, or *association rule learning in data mining,* is a popular and well-researched technique for discovering interesting relationships among variables in large databases. Thanks to automated data-gathering technologies such as bar code scanners, the use of association rules for discovering regularities among products in large-scale transactions recorded by point-of-sale systems in supermarkets has become a common knowledge-discovery task in the retail industry. In the context of the retail industry, association rule mining is often called *market-basket analysis*.

Two commonly used derivatives of association rule mining are **link analysis** and **sequence mining**. With link analysis, the linkage among many objects of interest is discovered automatically, such as the link between Web pages and referential relationships among groups of academic publication authors. With sequence mining, relationships are examined in terms of their order of occurrence to identify associations over time. Algorithms used in association rule mining include the popular Apriori (where frequent itemsets are identified) and FP-Growth, OneR, ZeroR, and Eclat.

Data mining.

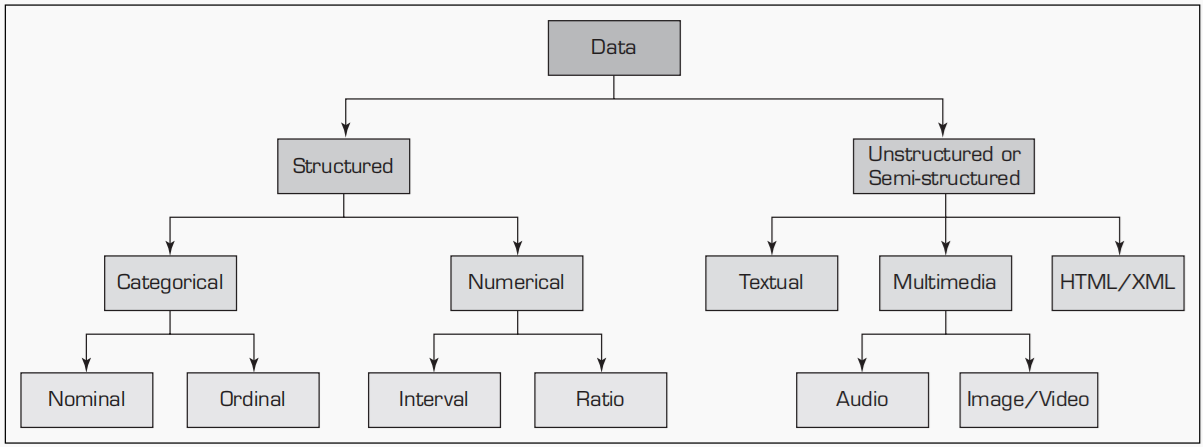


Data mining is a process that uses statistical, mathematical, and artificial intelligence techniques to extract and identify useful information and subsequent knowledge (or patterns) from large sets of data.

The following are the major characteristics and objectives of data mining:

* Data are often buried deep within very large databases, which sometimes contain data from several years. In many cases, the data are cleansed and consolidated into a data warehouse. Data may be presented in a variety of formats.
* The data mining environment is usually a client/server architecture or a Web-based information systems architecture.
* Sophisticated new tools, including advanced visualization tools, help to remove the information ore buried in corporate files or archival public records. Finding it involves massaging and synchronizing the data to get the right results. Cuttingedge data miners are also exploring the usefulness of soft data (i.e., unstructured text stored in such places as Lotus Notes databases, text files on the Internet, or enterprise-wide intranets).
* The miner is often an end user, empowered by data drills and other power query tools to ask ad hoc questions and obtain answers quickly, with little or no programming skill.
* Striking it rich often involves finding an unexpected result and requires end users to think creatively throughout the process, including the interpretation of the findings.
* Data mining tools are readily combined with spreadsheets and other software development tools. Thus, the mined data can be analyzed and deployed quickly and easily.
* • Because of the large amounts of data and massive search efforts, it is sometimes necessary to use parallel processing for data mining.

A Simple Taxonomy of Data



**Categorical data** represent the labels of multiple classes used to divide a variable into specific groups. Examples of categorical variables include race, sex, age group, and educational level. Although the latter two variables may also be considered in a numerical manner by using exact values for age and highest grade completed, it is often more informative to categorize such variables into a relatively small number of ordered classes. The categorical data may also be called *discrete data,* implying that it represents a finite number of values with no continuum between them. Even if the values used for the

categorical (or discrete) variables are numeric, these numbers are nothing more than symbols and do not imply the possibility of calculating fractional values.

• **Nominal data** contain measurements of simple codes assigned to objects as labels, which are not measurements. For example, the variable *marital status* can be generally categorized as (1) single, (2) married, and (3) divorced. Nominal data can be represented with binomial values having two possible values (e.g., yes/no, true/false, good/bad), or multinomial values having three or more possible values (e.g., brown/green/blue, white/ black/Latino/Asian, single/married/divorced).

**Ordinal data** contain codes assigned to objects or events as labels that also represent the rank order among them. For example, the variable *credit score* can be generally categorized as (1) low, (2) medium, or (3) high. Similar ordered relationships can be seen in variables such as age group (i.e., child, young, middle-aged, elderly) and educational level (i.e., high school, college, graduate school). Some data mining algorithms, such as *ordinal multiple logistic regression,* take into account this additional rank-order information to build a better classification model.

• **Numeric data** represent the numeric values of specific variables. Examples of numerically valued variables include age, number of children, total household income (in U.S. dollars), travel distance (in miles), and temperature (in Fahrenheit degrees). Numeric values representing a variable can be integer (taking only whole numbers) or real (taking also the fractional number). The numeric data may also be called *continuous data,* implying that the variable contains continuous measures on a specific scale that allows insertion of interim values. Unlike a discrete variable, which represents finite, countable data, a

continuous variable represents scalable measurements, and it is possible for the data to contain an infinite number of fractional values.

• **Interval data** are variables that can be measured on interval scales. A common example of interval scale measurement is temperature on the Celsius scale. In this particular scale, the unit of measurement is 1/100 of the difference between the melting temperature and the boiling temperature of water in atmospheric pressure; that is, there is not an absolute zero value.

• **Ratio data** include measurement variables commonly found in the physical sciences and engineering. Mass, length, time, plane angle, energy, and electric charge are examples of physical measures that are ratio scales. The scale type takes its name from the fact that measurement is the estimation of the ratio between a magnitude of a continuous quantity and a unit magnitude of the same kind. Informally, the distinguishing feature of a ratio scale is the possession of a nonarbitrary zero value. For example, the Kelvin temperature scale has a nonarbitrary zero point of absolute zero, which is equal to –273.15 degrees Celsius. This zero point is nonarbitrary, because the particles that comprise matter at this temperature have zero kinetic energy.

Other data types, including textual, spatial, imagery, and voice, need to be converted into some form of categorical or numeric representation before they can be processed by data mining algorithms. Data can also be classified as static or dynamic (i.e., temporal or time-series). Some data mining methods and algorithms are very selective about the type of data that they can handle. Providing them with incompatible data types may lead to incorrect models or (more often) halt the model development process. For example, some data mining methods need all of the variables (both input as well as output) represented as numerically valued variables (e.g., neural networks, support vector machines, logistic regression). The nominal or ordinal variables are converted into numeric representations using some type of *1-of-N* pseudo variables (e.g., a categorical variable with three unique values can be transformed into three pseudo variables with binary values—1 or 0). Because this process may increase the number of variables, one should be cautious about the effect of such representations, especially for the categorical variables that have large numbers of unique values.

Similarly, some data mining methods, such as ID3 (a classic decision tree algorithm) and rough sets (a relatively new rule induction algorithm), need all of the variables represented as categorically valued variables. Early versions of these methods required the user to discretizenumeric variables into categorical representations before they could be processed by the algorithm. The good news is that most implementations of these algorithms in widely available software tools accept a mix of numeric and nominal variables and internally make the necessary conversions before processing the data.

DSS:

A **Decision Support System (DSS)** is an interactive computer-based system that assists individuals or organizations in making informed decisions by analyzing large volumes of data, providing insights, and helping predict outcomes. DSS combines data, sophisticated analytical models, and user-friendly interfaces, allowing decision-makers to evaluate different scenarios and make data-driven choices. These systems are widely used in fields like healthcare, business, finance, and engineering to improve decision quality, increase efficiency, and reduce uncertainty.

**How Decisions are Supported?**

Here we relate specific technologies to the decisionmaking process. Databases, data marts, and especially data warehouses are important technologies in supporting all phases of decision making. They provide the data that drive decision making.

**Support for the Intelligence Phase**

The primary requirement of decision support for the intelligence phase is the ability to scan external and internal information sources for opportunities and problems and to interpret what the scanning discovers. Web tools and sources are extremely useful for environmental scanning. Web browsers provide useful front ends for a variety of tools, from OLAP to data mining and data warehouses. Data sources can be internal or external. Internal sources may be accessible through a corporate intranet. External sources are many and varied.

Decision support/BI technologies can be very helpful. For example, a data warehouse can support the intelligence phase by continuously monitoring both internal and external information, looking for early signs of problems and opportunities through a Web-based enterprise information portal (also called a dashboard). Similarly, (automatic) data (and Web) mining (which may include expert systems [ES], CRM, genetic algorithms, neural networks, and other analytics systems) and (manual) OLAP also support the intelligence phase by identifying relationships among activities and other factors.

Geographic information systems (GIS) can be utilized either as stand-alone systems or integrated with

these systems so that a decision maker can determine opportunities and problems in a spatial sense. These relationships can be exploited for competitive advantage (e.g., CRM identifies classes of customers to approach with specific products and services). A KMS can be used to identify similar past situations and how they were handled. GSS can be used to share information and for brainstorming.

Another aspect of identifying internal problems and capabilities involves monitoring the current status of operations. When something goes wrong, it can be identified quickly and the problem can be solved. Tools such as business activity monitoring (BAM), business process management (BPM), and product life-cycle management (PLM) provide such capability to decision makers. Both routine and ad hoc reports can aid in the intelligence phase. For example, regular reports can be designed to assist in the problem-finding activity by comparing expectations with current and projected performance. Web-based OLAP tools are excellent at this task. So are visualization tools and electronic document management systems.

Expert systems (ES), in contrast, can render advice regarding the nature of a problem, its classification, its seriousness, and the like. ES can advise on the suitability of a solution approach and the likelihood of successfully solving the problem. One of the primary areas of ES success is interpreting information and diagnosing problems. This capability can be exploited in the intelligence phase. Even intelligent agents can be used to identify opportunities.

Much of the information used in seeking new opportunities is qualitative, or soft. This indicates a high level of unstructuredness in the problems, thus making DSS quite useful in the intelligence phase.

**Support for the Design Phase**

The design phase involves generating alternative courses of action, discussing the criteria for choices and their relative importance, and forecasting the future consequences of using various alternatives. Several of these activities can use standard models provided by a DSS (e.g., financial and forecasting models, available as applets). Alternatives for structured problems can be generated through the use of either standard or special models.

However, the generation of alternatives for complex problems requires expertise that can be provided only by a human, brainstorming software, or an ES. OLAP and data mining software are quite useful in identifying relationships that can be used in models. Most DSS have quantitative analysis capabilities, and an internal ES can assist with qualitative methods as well as with the expertise required in selecting quantitative analysis and forecasting models. A KMS should certainly be consulted to determine whether such a problem has been encountered before or whether there are experts on hand who can provide quick understanding and answers. CRM systems, revenue management systems, ERP, and SCM systems software are useful in that they provide models of business processes that can test assumptions and scenarios. If a problem requires brainstorming to help identify important issues and options, a GSS may prove helpful. Tools that provide cognitive mapping can also help. Several Web-based tools that provide decision support, mainly in the design phase, by providing models and reporting of alternative results. Each of their cases has saved millions of dollars annually by utilizing these tools. Such DSS are helping engineers in product design as well as decision makers solving business problems.

**Support for the Choice Phase**

In addition to providing models that rapidly identify a best or good-enough alternative, a DSS can support the choice phase through what-if and goal-seeking analyses. Different scenarios can be tested for the selected option to reinforce the final decision. Again, a KMS helps identify similar past experiences; CRM, ERP, and SCM systems are used to test the impacts of decisions in establishing their value, leading to an intelligent choice. An ES can be used to assess the desirability of certain solutions as well as to recommend an appropriate solution. If a group makes a decision, a GSS can provide support to lead to consensus.

**Support for the Implementation Phase**

This is where “making the decision happen” occurs. The DSS benefits provided during implementation may be as important as or even more important than those in the earlier phases. DSS can be used in implementation activities such as decision communication, explanation, and justification.

Implementation-phase DSS benefits are partly due to the vividness and detail of analyses and reports. For example, one chief executive officer (CEO) gives employees and external parties not only the aggregate financial goals and cash needs for the near term, but also the calculations, intermediate results, and statistics used in determining the aggregate figures. In addition to communicating the financial goals unambiguously, the CEO signals other messages. Employees know that the CEO has thought through the assumptions behind the financial goals and is serious about their importance and attainability. Bankers and directors are shown that the CEO was personally involved in analyzing cash needs and is aware of and responsible for the implications of the financing requests prepared by the finance department. Each of these messages improves decision implementation in some way.

Reporting systems and other tools variously labeled as BAM, BPM, KMS, EIS, ERP, CRM, and SCM are all useful in tracking how well an implementation is working. GSS is useful for a team to collaborate in establishing implementation effectiveness. For example, a decision might be made to get rid of unprofitable customers. An effective CRM can identify classes of customers to get rid of, identify the impact of doing so, and then verify that it really worked that way. All phases of the decision-making process can be supported by improved communication through collaborative computing via GSS and KMS. Computerized systems can facilitate communication by helping people explain and justify their suggestions and opinions.