Paper #3006

Decision Support in the 90's

George Ferguson

Hewlett-Packard
19490 Homestead Road
Cupertino, CA 90014
U.S.A.

What is Decision Support?

Decision support is the use of Information Technology to help *knowledge workers* (executives, managers, financial analysts, product managers, marketing analysts, etc.) find and analyze the information they need to make *faster* and *better* recommendations and decisions.

Decision support and transaction processing (either batch or on-line) together make up what is known as *information management.* Transaction processing, which is the use of computer technology with which companies are most familiar, involves the automation of clerical tasks. Examples of transaction processing include:

> an order processing clerk taking a customer order over the telephone and entering it into the computer system, or
>
> a data entry clerk in personnel entering into a computer system a manager-approved raise for an employee, or
>
> a retail bank customer withdrawing money from her savings account through an ATM

In contrast, decision support involves the automation of knowledge worker tasks. Examples of decision support include:

> a product manager for a detergent product trying to decide with the aid of the computer system whether a 25 cent coupon will increase sales volume sufficiently to make up for the lower price, or
>
> an assistant vice president at a bank trying to determine through the computer system the return on investment for a proposed new mutual fund account that will be marketed to retirees, or
>
> a high-level manager at a telecommunications company determining with the help of the computer system whether an additional 20% discounting of international calls during non-business hours would result in higher profitability than an expensive expansion of satellite communications capacity, or
>
> a hospital administrator trying to decide with the aid of the computer system which of two hip replacement procedures results in the highest recovery rates and the lowest number of days in the hospital.

Decision support has existed for the past 20 to 30 years in several different forms and is used to some extent by all major companies and many smaller companies and organizations.

Batch Reports

From the 1960s to the present, many companies have used batch reports for decision support. At first, these paper reports were generated by writing third-generation programs (typically in COBOL or RPG). Later the reports were produced with fourth-generation languages such as IBI Focus. Although widely used, there are many problems with this approach:

It can be hard to find the right information in a large stack of paper.

Requests for new information (or new ways of looking at the information) can take a long time for over-burdened IT departments to handle. Many IT departments have report backlogs ranging from a few weeks to over a year. In contrast, the information may only be valuable to the knowledge worker if it is available in minutes, hours, or days.

It is difficult to work with the data received using desktop analysis and presentation tools. In many cases, highly paid professionals have resorted to retyping originally computerized data from a paper report into their Lotus spreadsheet program.

Terminal-Based Decision Support

From the 1970s to the present many companies have also used terminal-based decision support systems (DSS) of varying levels of sophistication. These systems have also been developed with fourth-generation languages and executive information system (EIS) toolkits, such as the older product offerings from Information Resources Inc. (IRI) and Comshare.

Terminal-based decision support systems have more flexibility and better timeliness than the batch reports, but still suffer from inadequate flexibility, difficult user interfaces, poor integration with desktop tools, and often times poor integration of data sources. This last point means that data is only available on a variety of different computer systems that each have to be accessed separately and between which it is difficult to integrate information.

PC Data Access Tools and Slide-Show EIS

The 1980s saw two new developments in the decision support area—both of which continue to be widely used. PC data access tools, such as HP Information Access and many third-party offerings (Q+E, Lotus Datalens, etc.), finally gave end users direct access to corporate and departmental databases with good desktop tool integration. What the end users saw through these tools, however, was boggling to everyone except the most serious power users. The databases presented to the end users through these tools were usually filled with obscure and inconsistent codes, complex database schemes, and a scarcity of
the historical data needed to view trends over time. The users quickly discovered that access to the databases was not sufficient if the databases were not organized for end user usage.

At the same time, EIS vendors introduced slide show executive information systems with very easy graphical interfaces for high-level executives. The problem with these systems was that they were typically very inflexible and required a significant number of programmers/analysts to generate the graphical slides for very small numbers of executives. In some cases, six programmers would work overnight to produce EIS slides for six executives to view the next morning.

Decision Support in the 90's

The hottest area of decision support in the 1990's is data warehousing--a concept that includes a wide variety of decision support components. The following discussion will first examine (1) what is data warehousing, (2) the components of data warehousing, and (3) non-data warehouse approaches to decision support.

What is a Data Warehouse?

To solve the many problems with past generations of decision support systems, the idea of a *data warehouse* evolved. A data warehouse can be viewed as a decision support database that is maintained separately from an organization's operational databases. Some aspects of data warehousing have existed since the mid 1970s. The ideas were first crystallized and formalized, however, by Bill Inmon (now an executive vice president at Prism Solutions) in the mid 1980s in a set of articles and books, culminating in his 1992 book, Building the Data Warehouse.

A more rigorous definition, taken from Bill Inmon, is that a data warehouse is a subject oriented, integrated, time variant, non-volatile collection of data that is used primarily to aid organizational decision making. The following discussion describes these four key characteristics of data warehouses.

Subject Orientation

Operational databases, such as order processing and payroll databases, are organized around business processes or functional areas. These databases grew up out of the applications they served. The data was, thus, relative to the order processing application or the payroll application. Data on a particular subject, such as products or customers or employees, was maintained totally separately (and usually inconsistently) in a number of different databases. In contrast, a data warehouse is organized around subjects. This subject orientation presents the data in a much easier-to-understand format for end users and non-IT analysts.

	Operational Database	**Data Warehouse**
Function	Data processing:support of business operations	Support of decision making
Data	Process oriented; current values; highly detailed	Subject oriented; current and historical values; summarized and sometimes detailed
Usage	Structured, repetitive	Ad-hoc; some repetitive reports and structured applications
Processing	Data entry; batch or OLTP	End user initiated queries

Integrated

Integration of data within a warehouse is accomplished by making the data consistent in format, naming, and other aspects. Operational databases, for historic reasons, often have major inconsistencies in data representations. For example, a set of operational databases may represent "male" and "female" by "m" and "f", by "1" and "2", by "x" and "y". Often, the inconsistencies are more complex and subtle. For example, a seemingly simple concept like *customer* may easily represent several different ideas. In one operational database, *customer* may mean the major corporate entity making a purchase, for example General Electric. In another operational database, *customer* may refer to the business unit making a purchase, for example GE-Appliances or GE-Aircraft Engines. In a third operational database, *customer* may refer to the General Electric facility at 867 Lakeside Drive in Cincinnati Ohio, USA. In a data warehouse, on the other hand, data is always maintained in a consistent fashion, and multiple concepts are not represented by the same name.

Time Variant

Data warehouses are time variant in the sense that they maintain both historical and (nearly) current data. In certain cases, data warehouses may even store future data, such as forecasts or budgets. Operational databases, in contrast, contain only the most current, up-to-date data values. Furthermore, they generally maintain this information for no more than a year (and often much less). Data warehouses contain data that is generally loaded from the operational databases daily or weekly and which is then typically maintained for a period of 3 to 10 years. This is a major difference between the two types of environments.

Historical information is of high importance to decision makers, who often want to understand trends and relationships between data. For example, the product manager for a soft drink may want to see the relationship between coupon promotions and sales. This is information that is impossible to determine with an operational database.

Non-Volatile

Non-volatility, the final primary aspect of data warehouses, means that after the data warehouse is loaded, there are no changes, inserts, or deletes performed against the informational database except during a periodic batch update cycle. The data warehouse is, of course, first loaded with transformed data that originated in the operational databases. The data warehouse is subsequently reloaded or more likely appended on a periodic basis (usually nightly or weekly) with new transformed data from the operational databases.
Outside of this loading process, the data warehouse stays static. The few exceptions to this rule are covered later in the section on "Updating Data Warehouses". Due to non-volatility, the data warehouse can be heavily optimized for complex query processing. Non-volatility also avoids certain problems such as maintaining detail data in synch with pre-built summaries.

Why are data warehouses important?

Data warehouses are very important to nearly all large businesses because they help solve several major problems experienced by both IT and users for many years.

Decision makers in companies want and need information to help them make better, quicker decisions. This has become especially critical in today's global competitive environment in which fast access to the right information can provide the competitive edge to help companies cut costs and get products to market faster. Unfortunately, these information requests are being passed to backlogged IT departments that are being asked to downsize and keep costs under control. Even before recent downsizing efforts, many IT departments faced 3 month to 1 year backlogs of requests for custom reports and data extracts. Obviously, a major change needs to be made if business needs and management requirements are to be met. This is where a data warehouse can help.

A data warehouse allows informational processing to be off-loaded from the mainframe (where most operational databases still reside) onto much lower cost servers. Furthermore, a significant portion of user requests for information can be handled by the end users themselves (or non-IT analysts) through use of graphical data access and analysis tools that operate on data stored in the server. This off-loads IT departments, while providing much better service to the end users.

The question that is often asked at this point is why not give the users direct (albeit perhaps restricted) access to the data where it resides (in the operational databases) rather than introducing an additional system with redundant data storage. This is a very important issue that needs to be fully understood. Following are some of the major problems with the single system approach:

1) Inability to optimize performance for both operational and informational needs

Use of a single database for both operational and informational needs necessitates optimizing the database for one purpose or the other. This optimization can be significant since operational databases are generally optimized for simple updates, while informational databases should generally be optimized for complex queries.

2) Inconsistent data formats and semantics

As has already been pointed out, data formats and meanings in operational databases are often inconsistent since the databases have grown up over time around differing applications written by a variety of different developers with no consistent data model. The combination of these semantic inconsistencies with obscure and indecipherable codes make the operational databases an extremely hostile world to all but the most initiated non-IT analysts. In many cases, the semantic inconsistencies can make cross-subject-area analysis and broad cross-organization queries difficult or impossible even impossible for the skilled analysts.

3) Lack of historical data

The operational databases generally do not contain the historical data that is so badly needed by decision makers. Saving of this historical data in the data warehouse (whether in detailed or summary format) is not redundant storage since the data does not even exist in the operational environment.

For all of these reasons, it is usually important to separate operational and informational data. Exceptions to this rule are covered later in this paper in the section on "Direct Data Access".

What are the Components of Data Warehousing?

In order to understand the components of a data warehousing solution, the flow of data from operational systems through the data warehouse and down to business users will be examined. The components and flow are depicted using HP's OpenWarehouse framework (see diagram on the following page). The basic elements of the OpenWarehouse framework are widely accepted in the industry today--although they were very precedent-setting when first advocated by Hewlett-Packard in 1993.

Data begins its electronic life in the operational databases (typically on mainframes) where it has generally been entered by clerical personnel doing batch or on-line transaction processing. The data is extracted from the operational databases and then transformed (or cleaned up) into subject-oriented, consistent formats that will be more understandable to end users. This clean data is then transported by network connection to a server, where it is loaded into the data warehouse.

Data within the warehouse is pulled down to the desktop systems, by data access and reporting tools and similar decision support applications or by advanced analytical tools, such as On-Line Analytical Processing (OLAP) tools, statistical tools, geographic information systems, executive information systems, and data mining tools. It is in these desktop tools and applications that the data truly takes on meaning and thus becomes information that can be used to help analysts, managers, and executives in making business decisions.

The following pages show most of the primary software components in the OpenWarehouse Alliance Partners program. The lists focus on the horizontal software vendors and thus do not include, for example, some of the more important vertical players in HP's OpenWarehouse-based Risk Management program (for financial services) and Call Behavior Analysis programs (for telecommunications).

Having now looked at the basic OpenWarehouse framework at a high level, it would be helpful to go through it again at a lower level, discussing in a more detailed fashion both the processes and the relevant components.

Operational Systems

Operational databases running batch and on-line transaction processing (OLTP) have been the mainstay of the computer industry for decades. These databases are used for tracking orders, managing accounts payable, handling payroll, etc. In most Fortune 1000 companies, these databases are on expensive IBM or plug-compatible mainframes, although in some companies they may be on other mainframes (e.g. Bull, Siemens, or UNISYS), DEC VAXes, HP 3000s, or high-end commercial UNIX servers. As has been previously stated, databases on these systems are typically application oriented (rather than subject oriented), and data formats and semantics are frequently obscure and usually inconsistent.

In addition to using data originating in internal operational databases, some companies also purchase high value, industry specific *syndicated* data from external sources. Prominent examples include A.C. Nielsen and IRI for the consumer packaged goods industry, A.M. Best for the insurance industry, and IMS America for the pharmaceutical industry.

Extraction/Transformation Tools

The extraction/transformation process is that of extracting data from the operational databases, transforming (or cleaning up) the data, moving the data to the server on which the data warehouse is

located, and loading the data into the warehouse. The extraction/transformation process also typically should include the production of relevant warehouse metadata (this will be explained further below).

The extraction/transformation process can be handled by COBOL or 4GL programs written specially by a customer or integrator to handle a specific custom warehouse. In many cases, however, it is more appropriate to use a special extraction/transformation tool for
generating these programs. Reasons for using a specialized tool include the following:

1) Developing custom code to perform the extract, transformation, and load processes is typically a long, tedious process that greatly lengthens the time necessary to develop and deploy the data warehouse. This process may seem easy, but the realities of most corporations' operational databases normally make this process long and difficult. Companies who have implemented data warehouses frequently complain that the extraction/transformation development process was much more difficult than anticipated and took a great deal longer than planned. Going with a custom-programming approach can be a drag on scarce IT resources and cause frustration in management, who are looking for fast results.

2) Custom coding does not produce organized metadata (literally "data about data") that describes the relationship between operational data and the warehouse data. This is crucial to warehouse maintenance and quality verification.

Network transport is generally done through high speed file transfer mechanisms rather than the much slower database gateways. Extraction, transformation, transport, loading and indexing normally must all take place in a relatively narrow time window (generally a night or a weekend). Database gateways often do not have the necessary bandwidth to transport the large volumes of data involved and still leave adequate time for the other steps within the available time window.

Note that most extraction/transformation tools do not actually initiate or manage the file transfer process or handle the scheduling of jobs and processes. In complex warehouse environments, this scheduling and management can be quite complex. Specialized tools are beginning to appear to manage these processes (like Software AG's SourcePoint) or generalized scheduling software (such as Maestro or CA-Unicenter). UNIX CRON is not normally adequate for complex environments.

Extraction/Transformation Tools

Vendor	Product
Carleton Corporation	Passport
Evolutionary Technologies Inc.	Extract
Informatica	OpenBridge
Information Builders Inc.	EDA Copy Manager
Platinum Technology	InfoRefiner
Prism Solutions	Prism Warehouse Manager

Scrubbing Tools

Closely related to the extraction/transformation tools are the scrubbing tools. Basically these are very sophisticated transformation tools that can be used in a complementary fashion with the extraction capabilities of the extraction/transformation tools. For example, a scrubbing tool might be used to

recognize that the following are all the same person:

John Doe	123 Main Street
J. Doe	123 Main Street
John Doe	123 Main
J. P. Doe	123 Main St.

This is of particularly high value in environments where there is a high payback from having very clean data (particularly customer or mailing data), such as in financial services, telecommunications, and retail/mail order.

Scrubbing Tools

Vendor	Product
Apertus	Enterprise/Integrator
PostalSoft	PostalSoft
Vality	IPE

The Data Warehouse

As has been previously stated, the data warehouse is a decision support database that maintains a subject oriented, integrated, time variant, non-volatile collection of data. The data warehouse is normally maintained within a relational database management system (RDBMS) or a similar flexible data store (such as a SAS data file).

The primary criteria that should be used in selecting a warehouse data store are as follows:

load/index time
query response time
database size requirements/limitations
quality
ratio of raw data size to full database size (including indices, temp space, etc.)
parallel capabilities
price
company DBMS standardization policy

A final criteria that sometimes comes into play is whether the warehouse will be used as a pure warehouse (i.e. no or practically no inserts/changes/deletes allowed) or whether some OLTP will also be performed on it. Although the data warehouse model cautions against allowing any OLTP against the data warehouse, some businesses will choose to do this for their own purposes. If this is done, it can become an important selection criteria since some excellent decision support DBMS's do not support robust OLTP.

Data Warehouse

Vendor	Product
Computer Associates	CA-Ingres
Hewlett-Packard	Allbase/SQL
Informix	Informix, Informix XPS
Oracle	Oracle7, Oracle Parallel Server
Red Brick	Red Brick Warehouse

SAS Institute	SAS
Software AG	ADABAS
Sybase	SQL Server, MPP

Data Warehouse Management

In spite of the many benefits of data warehousing, early data warehouses have exposed a number of issues and difficulties that keep data warehouses from effectively delivering the many benefits they promise. In particular most attention to date has been given to getting the data warehouse up and running, while less attention has been given to managing the warehouse after it has been established. Some of these particular management issues include the following:

Difficulties in managing warehouse tuning, security, and metadata integration

Limited usage of the data warehouse (due to poor performance and difficulties in determining warehouse contents and meaning)

Non-integrated data warehouse islands

The primary market offering in this area is HP's Intelligent Warehouse. Intelligent Warehouse is open middleware (residing between the data warehouse and end user data tools) that aids in data warehouse management. Intelligent Warehouse uses a generalized set of metadata to map user queries to the optimal summary tables in the data warehouse (thus improving ad hoc query performance by one or more orders of magnitude). Intelligent Warehouse also logs all user queries so that an administrator, using a graphical interface, can determine what summary tables to populate and depopulate, what indices to generate, and how to avoid problem queries.

Intelligent Warehouse works with all popular relational DBMS's and most ODBC-compliant data access and OLAP tools. Through its generalized metadata, administrators can provide a virtual, business-like view of the warehouse through what ever combination of data access and/or OLAP tools are used with the warehouse. This is a huge advantage to administrators as most data warehouses will (and should) be accessed from a variety of tools based on user needs.

Intelligent Warehouse has strong benefits to simple warehouse environments encompassing a single non-partitioned data warehouse residing on a single server. It also includes a number of optional enterprise extensions that make it ideal for complex, distributed warehouse environments. For example, Intelligent Warehouse allows multiple homogeneous or heterogeneous warehouses to appear as a single warehouse to end users. It also allows queries to span multiple subject areas (e.g. orders and shipments), thus overcoming the normal outer-outer join restrictions of SQL. This allows the logical unification of multiple independently-developed data warehouses provided that they utilize common data element definitions (e.g. customer in one warehouse is the same as customer in the other warehouse).

Data Access and Reporting Tools

Data access and reporting tools are various desktop tools for graphically building SQL queries that operate against the data warehouse. In addition to allowing queries to be made, most of these tools have full report writers that allow the returned data to be easily formatted into reports.

A wide variety of data access and reporting tools exist on the market and can be used with the OpenWarehouse. These tools support a wide variety of desktop systems including MS Windows, Windows NT, Macintosh, UNIX/Motif, and OS/2.

In some cases, more than one tool will be used by a company based on differing end-user needs. HP consultants can advise companies on the most appropriate data access and reporting tools based on the company's specific business needs and requirements.

Data Access & Reporting Tools

Vendor	Product
Andyne Computing	GQL
Brio	BrioQuery
Business Objects	Business Objects
Cognos	Impromptu
Information Builders Inc.	Focus for Windows
Oracle	Discoverer2000
Platinum Technology	SQL*Assist, InfoReports
PowerSoft	InfoMaker
SAS Institute	SAS/Assist
Software AG	Esperant
Sterling Software	CLEAR:Access

OLAP and Executive Information Systems

An executive information system (EIS) is simply a very easy-to-use data access tool (whose use is not at all restricted to executives). Common characteristics in today's EIS's include drill-down, threshold reporting, and exception reporting. The term EIS is now losing favor in many parts of the world. In contrast, much market excitement is now being directed at On-Line Analytical Processing (OLAP) server and client tools. In some cases these are re-architected EIS tools and in other cases are totally new tools. OLAP servers and clients provide an easy-to-use multi-dimensional interface for easily drilling down from high-level to low-level summary data and for slicing and dicing data. They also typically provide support for complex derived functions that are common in financial reporting and consolidation. Some of these tools are based on a multi-dimensional database (MDD), while others impose a multi-dimensional model directly on relational data (Relational OLAP). The OLAP tools have powerful analytic capabilities and are well-suited for fast response time (especially on small data sets), complex calculations, and budgeting/forecasting what-if analysis.

OLAP and Executive Information SystemsVendor	Product
Andyne Computing	Pablo
Arbor	Essbase
Cognos	PowerPlay
Comshare	Commander OLAP
Holistic Systems	Holos
Information Advantage	Decision Suite
Informix	Metacube
Microstrategies	DSS/Agent
Oracle	Express
Pilot	LightShip
Planning Sciences	Gentium
Platinum Technology	InfoBeacon, Forest & Trees
SAS Institute	SAS/EIS, OLAP++
Speedware	Media

Data Mining

Data mining is the use of intelligent agents or computer assisted methods to find patterns and anomalies in data. These tools assist analysts in finding data relationships in a specified area. For example, they might help an analyst recognize that males below age 24 have very high automobile accident rates (a pattern) or identify individuals with more than 3 whiplash injuries in the past year (an anomaly that might indicate insurance fraud). A variety of data mining techniques exist including statistical analysis and neural networks. This very leading-edge area in decision support is now beginning to find its way into commercial applications--particularly in financial services and large retail. Many more companies and organizations will likely be adopting data mining over the next several years.

Data Mining Vendor	Product
DataMind	neurOagent
Information Discovery	IDIS
NeoVista	Decision Series

SAS Institute	SAS/Neuronets

4GL's, GUI Builders, and PC Databases

In certain situations, end users may prefer a specific, custom decision support application over the ad-hoc capabilities of a data access and reporting tool. This is especially likely to be the case for clerical workers, customer support representatives, and sales personnel. To satisfy the needs of these users, OpenWarehouse can work with a variety of different fourth generation languages (4GL's), graphical user interface (GUI) builders, and PC databases. These application development tools may be used by IT departments, by HP's Professional Services Organization, or by third party integrators to develop needed decision support applications.

4GL's, GUI Builders, and PC Databases

Vendor	Product
Information Builders	Focus
Lotus	Approach
Microsoft	Access, Visual Basic
MITI	SQR/Workbench
PowerSoft	PowerBuilder
SAS Institute	SAS/AF

Web/Internet Access

Many companies and organizations are now considering web browser access to their data warehouses for several reasons:

> To roll-out data warehouse access to large numbers of knowledge workers while minimizing client-software installation/maintenance costs
> To serve users with differing desktop hardware and operating systems
> To open data warehouse access (over the internet) to users external to the company or organization--particularly to business partners, suppliers, distributors, and major customers.

A variety of tools are now becoming available that provide such web-warehouse access. These tools are effectively web-browser versions of all of the end-user tool categories just covered: data access and reporting tools, OLAP and EIS tools, data mining tools, and 4GL's & GUI builders.

When considering web-warehouse access, crucial concerns should be the *management* of the large number of potential users and the *security* of an internal data warehouse when opened to external users. Products, such as HP Intelligent Warehouse and HP VirtualVault, should be considered to deal with these issues.

Enterprise Data Warehouses and Data Marts

The original concept of a data warehouse was that of a very large, centralized database serving an entire enterprise. As companies have implemented data warehouses over the past several years, this original concept has seen a couple of variations.

Large centralized companies have tended to implement large, centralized data warehouses, but have added a second set of smaller *data marts* based on geography or functional areas--each of which contains a subset of the data in the central warehouse. These data marts are established to provide higher performance and flexibility than that provided by the central warehouse.

In contrast, large decentralized companies have tended to implement many data warehouses or data marts around separate business units or functional areas--without deploying any central warehouse. Some of these companies are now beginning to experiment with unifying these multiple warehouses into a single logical warehouse through middleware, such as HP's Intelligent Warehouse. This approach is more complex than the centralized approach, but it better fits the needs of highly decentralized companies in which construction of a
huge centralized warehouse is usually impractical.

Non-Tangible Aspects of Data Warehousing

Too many people believe that a data warehouse is simply a hardware box and a relational DBMS. A data warehouse is much more than just hardware and software. To be fully effective and thus help organizations take full advantage of their data in organizational decision making, data warehouses will usually involve major reorganization of data, changes in processes, and often even changes in organizational roles.

Data warehouse success is by no means a foregone conclusion, and companies planning to implement a data warehouse should seriously examine options for supplementing internal IT capabilities with experienced data warehousing consultants unless internal IT staff already have experience in this area. Note that the skills and procedures for setting up a successful transaction processing system are *very different* from those involved in establishing a successful data warehouse.

Non Data-Warehouse Approaches to Decision Support

Having examined the various components of decision support solutions based on a data warehouse, it is now important to also examine non data-warehouse approaches to decision support. Given the stated benefits of data warehousing, why should non data-warehouse solutions be considered? Two reasons are evident:

1) Cost - Data warehouses have been shown by industry analyst IDC to have excellent return on investment--401% mean ROI in a three-year period (Steve Graham, *The Foundations of Wisdom: A Study of the Financial Impact of Data Warehousing*, (Toronto: International Data Corporation, 1996), p. 5). The same IDC report, however, showed that the mean data warehousing cost for that same three-year period exceeded $2 million. Data warehouses are not inexpensive and some companies will find the costs prohibitive. Alternate decision support approaches may have lower returns on investment, but they also are typically much less expensive.

2) Problem Applicability - A data warehouse is ideally suited to a certain class of problems--typically those involving broad and high-level organizational decision making. For various operational needs, such as customer contact applications and other applications focusing on time-critical process-state and detail data, a data warehouse may be overkill and even ill-suited to meet organizational requirements.

Note that the following non data-warehouse approaches utilize some of the same software components as the data warehousing solutions, but are either directly based on operational systems or utilize a hybrid integrated data store, such as an operational data store.

Direct data access

Decision makers, sales people, customer service representatives, and clerical personnel will at times need access to detailed or time critical information contained in operational databases. A variety of sophisticated database gateways can support these needs. Particularly noteworthy are the following:

Information Builders Inc.	- EDA/SQL
Oracle	- SQL*Net
Sybase	- OmniCONNECT and Direct Connect gateways

These gateways provide connectivity between a very wide variety of mainframe/minicomputer/server databases and either custom applications or data access and reporting tools. They are most appropriate for access to detail transaction data (i.e. small amounts of data) where minimal or no integration and transformation of data is required.

Application-Specific Decision Support

Many specific and integrated operational applications have decision support components. These DSS components typically operate against the operational database. They produce a variety of reports and sometimes include more sophisticated OLAP-type analytic tools. They have the significant benefits that they require little set-up and configuration (assuming that you are already purchasing or using the application in question) and are significantly less expensive than a full data warehouse. They have the major disadvantages that they may impact the performance of the operational database and their tools often lack the flexibility of a fully open data warehouse. They are also application-oriented, rather than subject-oriented and are thus limited to the data domain of the specific application. If a company needs to access and analyze data from multiple applications, a separate data warehouse will still need to be established.

Operational data stores

An operational data store (ODS) is a concept that is very similar to a data warehouse. Like a data warehouse, an ODS is a subject oriented, integrated collection of data that is maintained separately from the operational database systems. Unlike a data warehouse, however, an ODS does not contain historical information , but instead contains only current or nearly current information that is extracted and updated from the operational databases. Furthermore, because they are primarily used by operational workers and clerical personnel, operational data stores focus on transaction-level detail data and typically do not include summaries. Operational data stores are generally used for customer service applications or occasionally for clerical reporting needs. They are also sometimes used for large-scale transformations in complex warehouse environments in which the magnitude of data integration and transformation makes it impractical to do the integration/transformation work on either the operational systems or the data warehouse system. In contrast to an ODS, a data warehouse is generally used to aid management decision making and is used by higher-level decision makers.

The extraction/transformation and scrubbing components of the OpenWarehouse framework can be used for either true data warehouses or operational data stores. Since operational data stores are typically accessed through custom applications for repetitive queries and updates, ad hoc data access tools and analytical tools (such as OLAP and EIS) are not normally used.

Summary

Decision support systems are crucial for companies to compete in the global competitive market place of the 1990's. Recent years have seen both the emergence of both major new software tools and fully re-architected approaches to aid companies and organizations in effectively harnessing the power of information.