Data Warehousing and Web Enablement: Opportunities, Issues, and Trends

Hamid R. Nemati
Department of Information Systems and Operations Management,
The University of North Carolina,
Greensboro, NC 27014
Tel: (336) 334-5666
Fax: (336) 334-4083
nemati@uncg.edu

Sherrie Cannoy
Department of Information Systems and Operations Management,
The University of North Carolina,
Greensboro, NC 27014
Tel: (336) 334-5666
Fax: (336) 334-4083

Robert Delk
Department of Information Systems and Operations Management,
The University of North Carolina,
Greensboro, NC 27014
Tel: (336) 334-5666
Fax: (336) 334-4083
Data Warehousing and Web Enablement: Opportunities, Issues, and Trends

Abstract

The Web is becoming the biggest contributor towards the growth of business intelligence and data warehousing. It is anticipated that 80% of all decision support queries will be across the Web in the near future. Using Internet/Intranet technologies to implement data warehouses is significantly reducing the cost associated deployment. Security and administration are centralized along with the elimination of the need to maintain client code on the end-user’s computer. New users can be granted access more quickly and application updates are automatic and contribute to the reduction in deployment and maintenance costs. The compelling advantages in using the Web for access are magnified even further in a data warehouse (Berson). In this paper, we discuss opportunities and issues relating to making an enterprise-wide datawarehouse available on the web and discuss a number of important trends and technologies.

Keywords:
Data Warehouse, Internet, Interanet, World Wide Web, Java, XML,
The Information Asset

Information and the ability to share information are two vital assets of any corporation. Business battles in the next decade will be won and lost according to how well companies know their customers and how fast they respond to customer needs (Hibbard). Focusing on the customer will require the monitoring of customer data minute by minute.

Internet technologies can help companies leverage their information assets by making it easy to publish and make this information accessible to more users. The following lists some of the reasons that are pushing organizations to look at Web technologies to publish their information assets on the Internet/Intranet:

- Companies looking to capitalize on their data
- Lower costs for database maintenance and administration
- Organizations are interested in reducing client software requirements
- Users want access to data and they want the access now!

Information is increasingly becoming the ultimate weapon that corporations can use to gain a competitive advantage over their competitors. The value of data warehousing is maximized when the right information gets into the hands of those individuals who need it, where they need it, and when they need it the most (Berson). Getting the information in the right hands is critical to gaining the maximum return on investment with your data warehouse. The delivery of information over the Internet offers the potential of reduced application development costs along with the ability to reach more users. Intranets enable companies to improve communication and collaboration among employees, vendors, and customers, thereby increasing productivity and providing significant savings in time and money (Kumar).

In order to adapt to the changing demands of the marketplace, the major database vendors are introducing new data warehousing technologies. The top six database vendors, in terms of market share--Computer Associates, IBM, Informix Software, Microsoft, Oracle, and Sybase--are turning out at a rapid clip technologies designed to help users publish their data to the Web, as well as to use the Web as a way to query their databases more efficiently (Foley). The following lists examples of technologies that the major database vendors are implementing or are planning to implement:

- Internet database connectivity
- Extraction and dissection tools (ex. OLAP and Data Mining)
- Database development and web-authoring tools
- Support for enhanced data types (ex. graphics, multimedia, etc.)
- Scalability
- Added support for data warehouses and data marts
- Transaction processing
- Handle objects using predefined business rules
An Introduction to Data Warehousing

For today’s businesses to remain competitive it is important for these companies to learn how to massage the mountains of available raw data into information. Once this data is transformed into information a company then can use this information to create a competitive advantage. Data warehousing encompasses the technologies that can help corporations manage this raw data and identify trends, reduce costs, retain customers and obtain new customers. The following list some of the business problems that data warehousing can help with:

- **Customer Retention** - minimize customer turnover or determine whether it is cheaper to acquire new customers or keep existing customers
- **Sales** - aid sales force by making product recommendations to customers
- **Customer Service** - predict what a customer may need and offer these services to them
- **Marketing** - aids in targeting promotions to the appropriate audience
- **Risk Assessment** - determine whether a customer is a high or low risk based upon information such as address, age, or educational background
- **Fraud Detection** - identify fraudulent practices by individuals that are using different names

A data warehouse can be defined as follows:

*A data warehouse is a subject-oriented, integrated, time-variant, non-volatile collection of data in support of management decisions.*

-- W. H. Inmon (Father of data Warehousing)

But a data warehouse is more than just a collection of data. A data warehouse encompasses all of the technologies used to both store and access the data. These technologies include the database management system (DBMS), query tools, report tools, data transformation tools and network infrastructure necessary to support the technology. Additionally, a data warehouse has the following attributes:

- The database is separate from on-line transaction processing (OLTP) systems
- The database is typically large and combines information from multiple applications
- The database contains aggregate and summary information
- The database is typically not normalized (contains redundant data)
- The business analyst must be able to drill-down from the various summarization levels to potentially the lowest grain of information (i.e. the actual transaction)
- The database supports direct access by the business analyst
- The database is read intensive and is updated on a periodic bases
- The database contains both current and historical information
- Tools used by the business analysts are easy to use
- The data within the database is consistent in that a report for a specific time period generates the same results even when initiated at different times
As previously defined, a data warehouse is separate from your existing OLTP systems. The reasons for this separation are outlined as follows:

- **Performance** – Data within a data warehouse are organized to support the subject-oriented view of the user rather than the performance requirements of OLTP systems; minimize the impact that decision support queries will have on the OLTP system; OLTP systems are tuned for each individual transaction where a data warehouse is tuned for complex queries.
- **Data Sources** – The data warehouse consolidate the various OLTP data sources into one data source.
- **Aggregation and Summarization** – A data warehouse will typically contain aggregate and summarization information in order to speed access by the end-user.
- **Data Quality** – Data quality of information flowing into the data warehouse is guaranteed. The data transformation process acts upon the information flowing into the data warehouse by validating the data, cleaning the data and properly aggregating the data.

Data Warehousing has become an integral part of today's organizations. Taking data stored in on-line transaction processing (OLTP) systems, aggregating and summarizing the data, and then presenting this information to business users, allows corporations to achieve greater productivity and allows business analysts to make more informed decisions. Because of these benefits, more and more organizations are looking to implement data warehouse technologies to achieve a competitive advantage.

As a result of successful data warehouse implementations, IS departments are being pressured to increase the amount of information that is available in the data warehouse, increase the number of users that can access the information, and provide access to the information instantly. The following lists some of the reasons that are pushing organizations to expand their of current data warehouse systems:

- Competitive pressures
- More and more data warehouse implementations are successful
- Decrease in hardware and software costs
- Increase in hardware and software performance
- Internal pressures to reduce costs while improving productivity

**Advantages of Web-Enabled Data Warehousing**

Enabling your data warehouse for access by either the Internet or your corporate Intranet is more than just the latest craze. Web style access has several advantages and in some cases makes more sense than the current client/server paradigm. The following lists some advantages of Web-enabling your data warehouse implementation:

- Make information available to a wider audience (ex. internal and external customers)
- Make information more accessible (ex. browser based access)
• Increase the use of the data warehouse
• Information can be accessed remotely
• Companies can sell subscriptions to the information to generate revenue
• Enhance customer service by aiding customers in helping themselves
• Increase in the sharing of information within the organization
• Lower client workstation requirements
• Browser supplies a common user interface across the organization
• Lower training cost because user interface is familiar
• Server-centric licensing model is typically cheaper than per-seat licensing model
• Lower client workstation support costs
• Cheaper to communicate over the Internet than implementing a private network
• Centralized system administration
• Outsource network support (ISP)

10. Browsers are universal - they look the same for intranets, extranets or the Web.
9. Internet access is cheaper and easier than remote dial up.
8. Those Mac fanatics in advertising can keep their smiley-faced machines.
7. Desktop operating systems don't usually provide 3270 or 5250 clients, but they do have a IP stacks (the first battle in the legacy link war).
6. You can dump those proprietary client/server apps that were supposed to save the world.
5. Web-to-host integration has lots of security options; you can ID'em at the Web page and at the database gateway, and encrypt everything for transmissions besides.
4. Your business partners are clamoring for a Web-based peek at order status or other electronic commerce applications.
3. Your competitors already offer their business partners Web access to order status.
2. It's a way to make those twenty-something Web wizards acknowledge the mainframe's contribution.
1. It's not just the latest buzz, it's a trend

Top 10 Reasons you should Webify your legacy data (InformationWeek, 1998).

Web Browsers

Every end-user workstation that is being installed today includes a Web browser. Web browsers are free and provide uniform access to the Internet and corporate Intranets. In addition, Web browsers by their nature contain functions that aid in information retrieval and processing. Web browsers incorporate as standard features many attributes that are especially useful in analytical applications, including local caching of pages, which drastically improves the response time for repetitive analysis; viewing of partial results as the page is loaded, which is particularly useful for speed-of-thought analysis; and asynchronous processing, data compression, and data encryption-features that may not be included in client/server OLAP, query, or reporting tools (Berson).

Using a web browser to access data does eliminate the need for additional software to be installed, configured, or maintained on the client machine. As a result, corporations can reduce the costs associated with software installation and maintenance. In addition, the Web browser
allows for corporations to reach the home market without requiring additional software installations. The effort associated with software installations and upgrades has been placed in the hands of the Web browser vendors and the end-user.

**Easy Access to Information**

The Internet and data warehousing are both technologies that are intended to offer easy access to information. The Internet supplies an infrastructure where more users can access corporate information. In addition, the Internet can increase the number of sources of information to include both internal and external data.

For many organizations the Internet eliminates many of the infrastructure issues of implementing a data warehouse. Teresa Wingfield of Giga Information Group, states "The Web is having a tremendous impact on data warehousing because it removes many of the geographical, technical, and organizational barriers to information delivery" (Weil). By relying on the inexpensive access offered by Internet service providers, corporations do not have to implement their own network architecture. In addition, the Internet offers connectivity to external companies and users that would previously not be possible or would be cost prohibitive.

Today's end-users expect up-to-date data that is delivered over the Web. Smart organizations are deploying data over the Web and letting end-users choose from a variety of tools -- avoiding a one-tool-fits-all mentality (Watterson). In addition, the Internet supplies easy access to data from anywhere. This allows for users to access the information they need while traveling or even from home.

**Extend Access to More Users**

Corporations are trying to move the decision-making process down the organizational ladder so that those closest to the problem can solve them. As a result, the decision-making is shifting from a few highly trained business analysts to more of the general population. This decision making shift is being created by competitive pressures that are forcing organizations differentiate their products, deliver these products to market faster, and deliver these products to the right customers. In order to meet these competitive pressures, information has to be delivered to this new group of users in a timely fashion and in a form that supports the appropriate level of analysis for this user group.

In addition to competitive pressures, organizations are under pressure to decrease costs. Cost reduction is inline with data warehouse initiatives that are intended to empower end users without requiring the assistance of IS resources. CEO of Sybase Mitchell Kertzman stated that both customers and companies can profit by exploiting the Web as a vehicle for service and products and for building brand loyalty (Hammond).

**The Impact of the Internet**

The Web is becoming the biggest contributor towards the growth of business intelligence and data warehousing. It is anticipated that 80% of all decision support queries will be across the
Web in the near future. The techniques for data retrieval are wide ranging but, increasingly, seem to be converging on the Internet and corporate Intranets (Foley).

Using Internet/Intranet technologies to implement data warehouses is significantly reducing the cost associated deployment. Security and administration are centralized along with the elimination of the need to maintain client code on the end-user's computer. New users can be granted access more quickly and application updates are automatic and contribute to the reduction in deployment and maintenance costs. The compelling advantages in using the Web for access are magnified even further in a data warehouse (Berson).

The Web is very efficient at processing static requests. The connection between a Web browser and a Web server is considered stateless since there is not an ongoing connection between the two. For data analysis, this stateless connection is inefficient. In addition, the majority of the processing is performed on the server rather than the workstation. To overcome the inefficiencies of the stateless connection, it is necessary to establish a direct connection between the Web clients and the DBMS. In addition, it is important to support this communications without requiring each query to reestablish a connection or login to the database. Once the data is retrieved the client has to have the ability to manipulate this information without requiring additional trips to the server.

The delivery of information on the Web is still evolving. Initially the Web was used for the delivery of static reports that were stored in an HTML format and accessed on demand by the end-user. While static report delivery using the Web was an improvement over paper based distributions; these static reports did not offer much in interactive data manipulation and analysis. The Web then evolved into a more interactive data collection environment by allowing users to initiate queries against enterprise data. Queries are initiated using an HTML form that in turn initiates a server side script such as Common Gateway Interface (CGI) script. The CGI script will initiate the appropriate SQL query against the database, format the response into HTML content ad present the information to the user. The current evolution of Web processing include the creation of Java or ActiveX clients that give the user similar functionality that can be found in existing client/server applications. These Web-enabled clients give the user the ability to perform enhanced analysis on the information.

Centralized Administration and Management

Even outside the data warehouse market there is a significant desire of corporations to centralize servers and support of applications. The promised benefit of client/server applications has placed several servers into the corporate network that are difficult to manage. Web-based application development has created an environment that allows for corporations to re-centralize these servers and ease the management headache. The Web allows companies to store and manage both data and applications on servers that can be centrally managed, maintained, and updated, thus eliminating problems with software and data currency (Berson).

Centralization into a few large servers also enables corporations to provide increased scalability for the changing application needs of users. It is cheaper to upgrade processors and hard drives on a few centralized servers than to upgrade several decentralized servers. At the
same time, this centralization is also forcing corporations to look into the scalability of the applications that they deliver. Strategic applications need to rapidly scale up to the amount of data they manage, the number of users they support, and the types of functionality they include (Berson).

From a cost perspective Internet/Intranet delivery of data warehouse applications makes sense. Web-centric data warehouses make data warehousing activities less expensive and easier to manage (Weil). Information can be made available to a wider audience but at a lower cost. Web technology combined with data marts make data warehousing technology more available to smaller organizations. Inexpensive data deployment to a wide audience is one of Webhousing's biggest advantages (Weil).

The Web allows for the reduction or even the elimination of custom report requests. By creating HTML or Java based clients, Web-based data warehouse projects seek to empower new categories of database users while lowering the cost of database administration and maintenance (Greenberg). Corporate Intranets can provide information which is always current, accessible from any platform, cost effective, rich in content, and easy to use (Kumar).

**Architecture for a Warehouse on the Web**

An Intranet refers to a network that is accessible only by the organization’s employees or others with authorization. A firewall surrounding the Intranet prevents unauthorized access. An extranet allows outsiders access via username and password to the network. In either case, there are specific pieces of architecture and tools that are included in the Warehouse on the Web solution. The Web itself is a universal, cross platform solution.

Some type of network-based architecture must be utilized so that contents are not duplicated. Handling common content requires common service management such as distributed object computing, which can access objects on the network. Distributed object computing can be administered through Common Object Request Broker Architecture (CORBA) or Distributed Component Object Model (DCOM), which is integrated into Microsoft operating systems (Web-based BI).

As stated earlier, script-based solutions such as Common Gateway Interface (CGI) can be utilized, but they do not allow for “stateful sessions” between the client and server. However, script-based solutions are only as effective as the Web server on which the run. As a result, scalability and performance may be limited. Another alternative is that of Plug-Ins which move processing to the desktop via the Web browser. They, too, have limitations with standardization because they are browser-specific, platform specific, and must be individually installed (Web-based BI).

Internal clients using a Web browser will access corporate information through a Web server. The Web server can then access the desired information directly from the database server using some sort of database connectivity like ODBC or JDBC. In cases where information is desired from a legacy system the Web server may use a CGI script that communicates with the application server that in turn accesses legacy information. Once the desired information is
retrieved this information is then presented to the client workstation for processing. With an appropriate client technology such as Java or ActiveX, the client workstation can then manipulate the information as desired. If no client technology exists then the Web server has to perform the desired manipulation and then present the results to the user in the form of an HTML page.

External clients will use a Web browser to access the Web server from the Internet through a corporate firewall. The firewall prevents unauthorized access to the internal corporate network by outsiders. Once authorized, the processing is the same as for internal clients. Keep in mind that the connection from the Internet maybe significantly slower than access from the internal network. The connection speed is design factor that must be considered when building your web-enabled data warehouse architecture.

The following diagram illustrates a possible architecture of a web-enabled data warehouse implementation:

![Web-enabled Data Warehouse Architecture](image)

**Figure 1 – Web-enabled Data Warehouse Architecture**

**Major Trends**

Analyst project that 60% of data warehouse projects over the next year will incorporate Web-enabled technology (King). In addition, The Data Warehouse Institute (TDWI) predicted that the market for data warehousing hardware, software and services would grow from $16.9 billion in 1996 to $40.5 billion in 2001 (Ubois). Industry data warehouse experts also believe that the Web is critical to a successful data warehouse implementation. Ralph Kimball believes that data warehouse results should be published everywhere, preferably over the Internet (Kimball).
While vendors have developed Web technologies for data warehouses, companies have been hesitant to adopt such remedies for enterprise-wide deployment. This is due to reasons such as Web technologies that lack architecture for standardized solutions or applications that duplicate content. The lack of architectural standardization is due to “...trade-offs between the extensive development time required to integrate Web solutions with desktop applications and the time-to-market pressures most vendors feel with emerging technologies” (Web-Based BI). Businesses succumb to pressures of justifying the expenditures of costly technologies. This is ironic since the reason for Web-enabled data warehouses is to save money by offering technology on an enterprise-wide basis. The area of webhousing is fairly new, though, and vendors are vying for market share by developing new products that will combat these problems.

Several major trends are driving the database vendors to enhance their data warehouse product offerings. Many of the data warehouse enhancements make vendor product offerings easier to implement in a Web environment. The following lists some of these trends:

- Internet database connectivity options
- Server side OLAP processing
- Data access from many data sources
- Internet enabled middleware and applications
- Enabling databases to work closely with Web servers
- Extensible Markup Language (XML) support

Technologies such as Java are intended to overcome some of the processing inefficiencies of the Web. Java allows for data to be retrieved from the source systems and manipulated without requiring a second trip back to the server. The data is manipulated on the client workstation. On the other hand, Java is considered slow and buggy. As a result, the savings achieved by implementing your application over the Web can be lost because the quality of the applications declines versus their client/server counterparts.

Regardless of the inherent limitation that still exist with implementing web-based applications, there are many technological advances that are occurring in the marketplace that address current issues while at the same time enhance Web access. This section touches on some of these trends and the technologies that are emerging in the business arena associated with the accessing of information over the Internet/Intranet.

**Hardware and Software**

Successful data warehouse implementations will allow users to find the information they need when they need. As users gain more familiarity with the data warehouse technologies, their use of these applications will grow and their information needs will become more complex. As usage patterns change the scalability of the data warehouse will become more important. Some of the scalability issues can be addressed with additional hardware and software.

Currently hardware performance is increasing while at the same time prices for new hardware is decreasing. Today’s processors are getting faster and faster and as a result, hardware vendors are generating faster single and multiprocessor systems. In addition to processor speeds,
storage capacities are also increasing with an associated decrease in price. The increase in storage capacities is allowing for more and more data to be captured and stored.

To support the increased appetite for information, software prices are decreasing and at the same time their analytical capabilities are increasing. For example, Microsoft is bundling OLAP services and metadata management with their latest SQL Server release but is not significantly increasing the cost of DBMS.

As a result of the current hardware and software trends, companies are better able to implement more comprehensive and scalable data warehouse systems. Hardware and software capabilities and costs are becoming less of a hindrance to data warehouse implementations.

**Data Brokers**

Everyone realizes that there is an enormous amount of information on the Web. The problem is finding the right information that can aid your decision-making. In response to this issue data brokers may begin to emerge on the Internet. These brokers would gather information from surveys and from other companies and then supply query-based access over the Internet to organizations. This technology would allow you to make your own interpretations of the facts rather than depending upon research organizations. Companies would be willing to supply data to these data brokers to ensure that their name shows up beside their competitors when queries are performed.

Data brokers will implement the technologies such as Java, XML, JDBC, and CGI that will merge bring the information from several data sources together for presentation to the end-user. You can view the data broker as the Web server that integrates all of the data sources that you desire.

The data brokers would offer subscription plans to businesses that are interested in the information that the brokers have to offer. Companies would then have the ability to query the data reserves and make their own interpretations of the information. The actual data could still reside at the source corporation rather than at the data broker. The data broker would link to the various source corporations over the Internet and unify the information into one delivery mechanism.

**Intelligent Agents**

Intelligent agent solutions in a data warehouse environment maximizes your warehouse investment and enable business intelligence software solutions in ways never before attainable (Bragger). As data warehouse implementations grow the time it takes to analyze the information also grows. This growth of information can have a negative impact on the business analyst by overwhelming them with information ("information overload"). As analysis becomes more advanced, users might lose sight of what they wanted in the first place (Bragger).

Intelligent agents are a technology that enables a business analyst to delegate tasks to computer programs whose job is to monitor the business surroundings on behalf of the user and
Intelligent agents can also be turned lose on your data warehouse in a data mining mode and return information to the user. The key is that the acts independent of any other processing.

Intelligent agents have their greatest potential on the Web. Consider the fact that the most important information that you will need is external to your company rather than within your existing data warehouse. In addition, this information is not stored in a database but is stored as bits of data that are embedded in a Web page. Not just one Web page but thousands of Web pages across the Internet. Obviously a business analyst does not have the time to search through the mountains of information that is available on the Internet. But a business analyst could inform an intelligent agent of the type of information that is desired and then turn that agent loose to discover this information and report back its findings.

Intelligent agent technology has been mentioned for use with XML. XML is a standard mechanism for representing data within a Web page. Intelligent agents could leverage this standard to produce dramatic results in their information gathering power.

Support for Unstructured Data

The ability to store and manipulate rich, complex data types via user-defined or vendor provided functions will likely result in the clear business need to manage and analyze this complex data as an extension of a data warehouse (Berson). The following outlines some of these complex data types:

- Image
- Voice
- Animation
- Multimedia
- Full-motion Video
- Spatial Data
- HTML Pages
- Business Rules
- Etc.

Support for complex data types is a real business need. Database management systems need to become extensible to support the storage of voice, image, video, text, and spatial information in their proprietary formats. IBM’s approach to nontraditional data is to provide links to it through the database. DB2 Universal Database Extenders are utilized for complex types of data. They can be manipulated though an SQL query. Oracle is taking a unique approach for the storing of complex data types with their technology iFS. Internet File System (iFS) allows for users to drag and drop data from Windows-based apps into the database, where it can be indexed, managed and queried (Foley). As a result a query could be initiated that spans e-mail, text files, and your data warehouse. Proprietary file formats such as Word and Excel can be imported into the database where they can be subsequently viewed using a Web browser.
In addition to supporting the storage of complex data types, future database management systems should allow for the creation of user-defined data types and should support add-on modules that can be used to increase the functionality that currently exists within the database engine through a well-defined API. IBM, Informix, and Oracle are currently taking this approach. IBM offers Relational Extenders, Informix offers DataBlades, and Oracle offers Data Cartridges. The logic is that if the support you need is not currently in the DBMS, add your own support using add-on modules or purchase add-on modules from a third party vendor.

**Extensible Markup Language (XML)**

As more and more people begin to use the Web as a medium for delivering information, the limitations of the current markup language HTML will become more apparent. HTML is not extensible and does not provide structure or data checking. The following outlines these limitations (Bosak):

- **Extensibility** – HTML does not allow users to specify their own tags or attributes in order to parameter or otherwise semantically qualify their data.
- **Structure** – HTML does not support the specification of deep structures needed to represent database schemas or object-oriented hierarchies.
- **Validation** – HTML does not support the kind of language specification that allows consuming applications to check data for structural validity on importation.

XML is an emerging standard for representing data that can enable applications to more easily exchange data. As a result, XML is moving from a document technology to a solution for data integration challenges. XML is seen as a method to meet a business need by giving access to diverse data sources and support the manipulation of them many times on the desktop without a trip back to the database (Karpinski). Since it is an open standard, XML can operate across disparate platforms and applications. The cross platform capability combined with support from vendors like IBM and Oracle is making XML the technology of choice for doing business on the Web. The following outlines some applications and benefits of XML technology:

- XML is useful in allowing Web clients to communicate with many heterogeneous databases.
- XML is useful in distributing the processing load from the server to the client.
- XML allows for different views of the same data to be presented to different users.
- XML can work with intelligent agents to tailor information discovery and delivery to users.

Extensible Markup Language (XML) is a tag language that supports a way to create data on the Web in a format that allows for information to be shared easier. XML is a simplification of the Standard Generalized Markup Language (SGML) and is considered a Web definition technology. Currently, XML is a formal recommendation of the World Wide Web Consortium (W3C) and is intended to make the Web a more versatile environment. In addition, XML has been designed to be powerful, easy to teach and learn, and easy to implement.
XML is similar to HTML in that both use tags that are embedded within Web pages. But where HTML describes how information is to be displayed, XML describes the contents of the Web page. XML allows for its pages to be processed as input into a program or displayed using a XML compatible browser. Another difference between XML and HTML is in the tag language that is used by both. HTML has a fixed set of tags that are valid for use in Web pages. On the other hand, XML tags are self-defining and there is no limit to the number of tags that can be created. This is what is makes XML extensible. XML allows developers to create their own tags to extend the functionality that is not found in HTML. The following outlines how XML differs from HTML (Bosak):

- Information providers can define new tag and attribute names at will.
- Document structures can be nested to any level of complexity.
- Any XML document can contain an optional description of its grammar for use by applications that need to perform structural validation.

One area that XML can be used is in conjunction with intelligent agents. For example, industries could standardize their XML tags across a product line. Car manufacturers could publish automobile information on their Web site using XML tags that indicate make, model, color, and engine size. As a result, a consumer could initiate an intelligent agent against the various car manufacturer’s Web sites that retrieves information on cars that meet the consumer’s criteria. The intelligent agent would use the XML tags to locate the attributes of the car and then retrieve documents that contain the appropriate attributes.

XML allows for a standard way to express the increasingly complex structured data found on Web sites and Intranets (Seltzer). In other words, XML is being established as a standard way of defining all data interchange. Another view of XML is as an EDI (Electronic Data Interchange) technology for the Internet. XML is a messaging format that can have the following uses:

- Messaging mechanism between clients and servers
- A mechanism for enabling and scaling distributed computing environments
- A format for structured documents
- A language upon which other markup languages can be built

Various vendors have announced support for XML. These vendors include database vendors IBM and Oracle. Oracle has announced support for XML in their products Oracle8i, Application Server 4.0, and Internet development tools. Oracle officials say that XML will play a role in their data warehouse products and application development tools (Walsh). Oracle8i will include a XML parser for processing XML documents and iFS will automate rendering of data from XML and the database. The XML support within Oracle8i and iFS will allow for the support of non-relational data files. Oracle's XML support will be made up of three components (Walsh):

- XML parser to process the XML document
- Internet File System (iFS) for parsing and rendering the XML to the database
- interMedia to provide better search tools for processing XML documents
IBM will support XML in their WebSphere Application Server which, when integrated with IBM's DB2 Universal Database, is a complete package for designing, developing, debugging, and deploying e-business applications.

Like any new technology, XML does have critics. Some industry analysts note the performance issues associated with XML parsers that render XML information to and from relational database management systems. As hardware performance increases this may be less of an issue. Secondly, other analysts are concerned that since developers have a mechanism to define their own tags that the widespread use of XML will be hindered since there is no control over these new tags. The proliferation of new data tags without uniform meaning may not offer a lot of value to the business analyst. A tag indicating sales in one document may not have the same meaning as sales in another document.

The Gartner Group estimates that the publishing of XML documents will exceed HTML documents by the first half of 2000 (Karpinski). If widely accepted by vendors, XML could lead to more vendor neutrality. As a result database management systems that store information in XML would allow data moved from one database management system to another without a transformation step.

The Internet is full of data, and XML is going to allow you to get at that data and sort it into more legible information (Dalton). For more information on XML refer to the web page http://www.w3.org/XML.

**XML Metadata Interchange (XMI)**

An extension of XML is the XML Metadata Interchange (XMI) format. XMI is intended to allow users of different development tools and environments to share information over the Internet. XMI format standard is a subset of XML and will become the cornerstone of an open interchange model for exchanging programming data over Web-based networks (Gonsalves).

XMI should allow for the exchanging of data between various tools, repositories, and applications that should enhance collaboration over the Web among developers. XMI should define a standard for storing and sharing objects that will in turn increase teamwork and productivity of developers.

**Java**

One of the biggest factors that has to be overcome when attempting to deliver information over the Web is how to deliver the necessary application code that will be used to access and manipulate the information. One of the technologies can help to overcome the application code distribution to end-users is Java.

Java is a programming language that was introduced by Sun Microsystems and was originally designed for use in consumer electronic devices and to be platform independent. Java has since been adapted for the distributed computing environment that includes the Internet.
Java programs can run on client workstations and servers. You will typically find Java programs as small applications that are embedded within Web pages. As a result, Java has the capabilities of making Web pages more interactive.

Java is represented in a machine independent bytecode that is interpreted by the client workstation. As a result of this machine independent representation, Java applications require a runtime environment to execute. This runtime environment is known as a Java Virtual Machine (JVM). The major Web browser vendors such as Microsoft and Netscape have incorporated a JVM into their products to support Java applications. In addition, an optional just-in-time (JIT) compiler may be embedded within the JVM to convert the bytecode into an executable in order to improve execution performance. One of the drawbacks of Java applications since its introduction has been its execution speed. The bytecode interpretation slows down the execution process. A JIT can speed up processing at execution time but with a tradeoff of the time it takes to be compiled.

When a Web page is loaded and it contains a reference to a Java application, the bytecode for the application is automatically downloaded to the client workstation and executed. This automatic download process allows for Java applications to be maintained on a central server. The client workstation will receive any application modifications automatically the next time the Web page is accessed.

The Java programming language bears a strong resemblance to C++ but is considered easier to use. Java improves on memory management and garbage collection that are considered weak points of C++. Overall, Java applications have the following characteristics:

- Java programs are compiled into a machine independent bytecode
- A Java Virtual Machine (JVM) on the client workstation interprets the byte an runs the application
- Java applications do not have the ability to reference data external to the application or other known objects. For example, a Java application does not have the ability to interact with files on your local workstation.
- Java is object-oriented and supports methods and inheritance
- Compared to C++ Java is easier to learn

As long as the client workstation has support for a JVM, either within the operating system or within the Web browser, the client workstation can execute Java applications. The appeal is that a developer can write an application once and have this application run on any computer platform. As a result, Java is considered a strategic language for Web development. For more information on Java, refer to the Web page: [http://java.sun.com](http://java.sun.com).

Within the spectrum of implementing a data warehouse over the Web, Java can be a tool that gives the end-user the same look and feel of today’s client/server applications without the software deployment headaches. In addition, Java is intended to be platform independent and should allow for information delivery to a wider audience without concerns about their client workstation’s operating environment. From a development standpoint Java offers the
opportunity to develop data warehouse applications that can run on many operating environments.

Vendors have responded to the potential benefits that Java support may offer by integrating Java Virtual Machines (JVM) within their database engines. Oracle will incorporate a JVM and support for Java stored procedures in their Oracle8i database. Oracle states that their Java support is the first database resident JVM. Java applications will be able to run inside the Oracle8i environment. The JVM will allow for developers to extend the functionality of Oracle8i by allowing for the development of components that can run inside the database engine. In addition, Oracle8i will allow developers to use Java for database programming rather than requiring the use of Oracle's PL-SQL language. The Java support within Oracle8i will enable Java applications to query the Oracle databases and support the deployment of these applications on any kind of client.

**Java Database Connectivity (JDBC)**

Database connectivity is a must for allowing end-users to access corporate information over the Web. The complication with database connectivity is the installation of the appropriate database drivers on the end-user’s workstation. JDBC has been developed to overcome this driver implementation problem.

![Java Database Connectivity (JDBC)](image)

**Figure 2 - Application Connectivity using JDBC**

Java Database Connectivity (JDBC) is an object-oriented application-programming interface that allows programs written in Java to connect to SQL compliant database management systems (DBMS). JDBC represents the industry standard for database-independent connectivity between Java applications and database management systems. JDBC makes it possible to write a single database application that can run on different platforms and interact
with different DBMSs (*JDBC*). JDBC further enhances the “write once, run anywhere” position of Java applications by supporting access to enterprise-level data.

JDBC provided call-level API for SQL-based database access (*Java*). JDBC allows for the encoding of SQL requests that are then passed on to the DBMS where the request is processed. The results are returned through the same interface. As a result JDBC provides three things:

- Connection to a database
- Encoding and transmission of SQL statements
- Receive and process the results

The JDBC API is implemented via a driver manager that can support multiple drivers connecting to different databases (*Overview*). The following diagram illustrates the JDBC architecture:

![Figure 3 - Application Connectivity using JDBC-ODBC](image)

The left side of the diagram illustrates a direct database connection where the JDBC call is converted into the protocol used by the DBMS. This option allows network applications to connect directly to the DBMS server. The right side of the diagram represents connection to a database using middleware. The JDBC call is translated into a protocol understood by the middleware server. The middleware server performs the actual connection to the DBMS server.
JDBC also supports database connectivity through ODBC using a bridge program. This connectivity to ODBC is often referred to as a JDBC-ODBC bridge. The JDBC-ODBC bridge allows for backward compatibility with DBMS systems that do not support JDBC directly. The following illustrates the JDBC-ODBC bridge architecture:

The left side of the diagram illustrates connectivity using the proprietary ODBC driver for the DBMS. The ODBC driver must be installed on each of the client workstations using this mechanism. The right side of the diagram illustrates connectivity using a proprietary Java driver that will connect with the DBMS server. This option also requires client code to be installed.

For both architectures, communications between the application and the DBMS occurs through the JDBC API and the JDBC Driver Manager. The following lists the various classifications of JDBC drivers:

- JDBC-ODBC Bridge
- Native-API Partly-Java Driver
- Net-Protocol All-Java Driver
- Native-Protocol All-Java Driver

You can refer to the Web page http://java.sun.com/products/jdbc/overview.html for detailed descriptions of these classifications.

Using JDBC for database connectivity offers the following advantages:

- Supports access to existing DBMS *
- Supports access to enterprise Metadata
- A pure JDBC require does not require any client installation (downloaded automatically)
- Remote database connections are made using a URL like reference
- Write once, run anywhere – JDBC is part of the Java Platform

Connection to remote database systems occurs using a URL like reference. For example:

Example 1: jdbc://host.domain.com:400/databasefile
Example 2: jdbc:odbc://host.domain.com:400/databasefile

Example 1 references an application connection to a DBMS using the JDBC driver. The ODBC reference in this second example is called a sub-protocol reference and is used to inform the JDBC Driver Manager how to access the database. In Example 2, the DBMS will be accessed using a JDBC-ODBC bridge.

JDBC also support the addition of vendor specific SQL extensions. Like the ODBC specification, JDBC supports the embedding of vendor specific SQL statements using keyword/parameter syntax. But unlike ODBC, JDBC is not language independent. JDBC is
designed specifically for Java applications. JDBC support is downloaded automatically as part of a Java application.

JDBC is an integral component for opening up database access over the Web. Combine Java with JDBC access and all the tools exist to replace the functionality of existing client/server tools. In addition, since JDBC is platform independent there isn’t the need for acquiring and installing ODBC drivers for each client workstation. Sybase offers jConnect, which is a JDBC driver that can be used to connect to Sybase data sources. The JDBC driver is completely written in Java and is fully compliant. jConnect will enable Java applications to easily connect to Sybase data sources.

For additional information on the vendors that endorse and are building JDBC-based products refer to the Web page http://java.sun.com/products/jdbc/jdbc.vendors.html.

**Very Large Databases (VLDB)**

The Internet offers the potential for a data explosion by enabling corporations to acquire information from outside resources rather than relying on the information produced from internal systems. As a result there is a potential the space requirements of data warehouses to grow previously unimaginable sizes. The goal for database management systems is to support these larger data warehouse systems without sacrificing performance. In response to the growing size of corporate databases the vendors are implementing Very Large Database (VLDB).

Database vendors will also have to increase their support for very large databases (VLDBs). As the amount of information increases, database management systems will not only need to support gigabytes of information but also maintain responsive query speeds. In addition, database management systems will also need to allow for backup and recovery of these large databases. With SQL Server 7.0, Microsoft has added the capabilities of managing up to one terabyte of data. In addition, Oracle8i from Oracle offers significant improvements in very large database support, including increased scalability, robust data partitioning, and enhanced availability.

The Web is making more and more information available to corporations. Combine this increase in information with the space associated with newer data types such as voice and video and the space requirements of databases will increase beyond what corporations may be used to. Databases grow steadily in size, and the support for very large databases (VLDBs) presents new technical challenges (Berson).

When implementing a Web-enabled data warehouse consideration has to be given to the DBMS and its ability to handle both current information and information needs. As corporations pull more and more information from external sources VLDB support will be a must.

**Issues in Implementing a Warehouse on the Web**
The following lists some of the issues associated with implementing data warehouse technologies over the Web:

- Cost
- Performance
- Scalability
- Security
- Total Solution versus Multi-Vendor
- Systems and Network Management Tools

**Cost**

The white paper developed by Hummingbird Communications “Web-based Business Intelligence” discusses the Lowest Cost of Ownership Imperative and its relationship with creating a business intelligence solution on the web. Included are factors affecting the cost of owning a web-based business intelligence solution such as:

- End-user training: both formal and casual
- Installation and troubleshooting
- Help Desk support
- Security, setup, integration, and administration
- Standards enforcement
- Content management

They did not, however, discuss the process of developing a plan to implement the web-based solution. The time and effort involved in gathering information and building a plan for implementation should also be included as part of the overall cost.

**Performance**

Enabling access to a data warehouse on the Web requires analysis of data and disk requirements, query complexity, desired response times, concurrent users, backup and restore cycles, data refresh rate, and amount of data redundancy desired. Users expect a short response time, regardless of how many users are on the system and what types of data (multimedia vs. text) are being accessed. With the emergence of multimedia data, performance issues will be raised regarding retrieval response time.

**Scalability**

 Corporations will be wise to undertake their initial Web data warehouse implementation by starting with a small subset of data. Starting small will allow for the corporation to gain confidence in the new technologies and security systems associated with implementing a web-based data warehouse. For the most part this initial implementation is confidence booster to show that these Internet-based technologies can work to deliver information from the legacy
systems to Web-based clients. But at the same time an initial implementation allows for the evaluation of the scalability of the system. Do additional servers need to be implemented? Does information need to be partitioned across many servers? Can the network infrastructure handle the anticipated volume? Since each corporate environment is different there is not a cookie-cutter implementation that can be used for all corporations. Architectures need to be tailored to fit the corporation’s specific needs. The following offers some guidelines to consider when implementing a data warehouse over the Web:

- Keep scalability in mind throughout the entire data warehouse design process
- Minimize the size of transmissions as much as possible
- Gear most of the processing to occur on the server
- Make sure that your server can grow as more users are added or as functionality increases
- Address any weak links to maintain availability
- Give data warehouse security equal focus

Security

It is easy to be enamoured with the idea of making corporate information available over the Web. This love for the Web makes it easy to forget the inherent security issues associated with making information available on the Internet. The loss of data warehouse data to hostile parties can have extremely serious legal, financial, and competitive impacts on an organization \((\text{Kimball, 449})\). Security, accessibility, and user acceptance are of paramount importance when developing Web-based data warehouse. Without them the best conceived projects are destined to fail \((\text{Meyers})\). The data warehouse manager is ultimately responsible for ensuring that information is made available to the appropriate users while at the same time protecting this corporate information from unauthorized access and modification.

Data Warehouse architects must balance the two competing forces of accessibility and unauthorized access to corporate information when developing a web-enabled data warehouse. This balancing must occur for both internal Intranet applications and those that are delivered over the Internet. Consider that fact that current employees perform nearly half of all hacking damage to corporate information \((\text{Meyers})\). This stresses that fact that just because you publish information only available via the corporate Intranet does not mean that less attention can be spent on securing the information from unauthorized access. Surprisingly, human error and human omission and not computer hackers typically cause information corruption. In 1997, the Aberdeen Group offered the following breakdown on the causes of “compromised information” \((\text{Kimball, 458})\):

- 35% - human error
- 25% - human omission
- 15% - disgruntled employees
- 10% - external people
- 7% - fire
- 5% - flood
• 3% - other natural disasters

The bottom 15% can be addressed using sound contingency planning while the top 60% can be addresses to a certain extent with proper controls and change management.

Making corporate information available over the Internet is inherently a risky proposition. The Internet was designed to connect everyone together and not to protect privacy (Kimball, 468). The underlying technologies of the Internet are based upon supporting this universal connectivity. In fact, anyone with a connection to the Internet backbone has the ability of “spoofing” the IP source address of packets that flow over the network. This spoofing creates the illusion that the packets are originating from a desired client but in fact could be originating from a hacker with access to this backbone connection. It is key that the data warehouse architect understands the inherent security risk of using the Internet so that these risks can be mitigated using the appropriate combination of technologies.

The nature of today’s Ethernet-based personal computers also generates risk on protected corporate Intranets. The nature of Ethernet networks creates an environment where every computer on the LAN has the ability to see all packets that flow on the network. A packet sniffer can be placed on the Ethernet segment an decode all the traffic that flows on the network and gain access to potentially sensitive information.

There is no one technology that can protect your corporate information from unauthorized access. Several technologies have to be combined in order to control access to the information stored in your web-enabled data warehouse. The following outlines many of these technologies:

**Push and Subscription Technologies** – Push and subscription technologies are mechanisms that delivers information to various users rather than having the user pull the information. Since information that is delivered is controlled centrally these technologies are effective in controlling access to the information in the data warehouse (Meyers).

**Packet filtering routers (firewalls)** – Firewalls are a combination of hardware and software that can be used to control access to the data warehouse both on the Internet and on the corporate Intranet. Firewalls provide the following useful functions (Kimball):

- Reject connection attempts form unknown hosts
- Reject attempts from the outside that are spoofing as insiders
- Isolating traffic to come from a single machine or a small group of machines such as database servers
- Prevent packet sniffing on the private network from machines on the public network

**Bastion/Proxy server** – The bastion/proxy server processes all communications that originate from the outside and routes the communications to the appropriate servers that are located on the corporate network. In addition, bastion/proxy servers are also used for communications that originate within the organization that are intended for the Internet.
The proxy application has the ability to determine whether the communications are valid and safe prior to routing the traffic. The proxy component offers the following functions (Kimball):

- Monitor and audit Internet/Intranet usage and access patterns
- Restrict access to prohibited sites Internet/Intranet sites
- Cache Web content for more efficient access

**Encryption** – Another technique that can be used to protect your corporate information is to use data encryption. Encryption is a key component in protecting data transmissions that occur between the clients and servers. Strong encryption processes allow for secure communications over the Internet.

**Private key encryption** – Private key encryption is a process where data is scrambled (encrypted) using a secret code (key) that is only known by the individual. This secret code is also required to unscramble the data. If this secret code is shared with another trusted individual then a mechanism for secure communications between the two can be established. The most widely accepted private key encryption mechanism is the Data Encryption Standard (DES).

**Public key encryption** – Public key encryption combines a private key and a public key mechanism to scramble or unscramble the data. Public keys are known to everyone but private keys are only known to the individual. Information encrypted using an individual’s public key can only be decrypted using the individuals private key. This mechanism supports the secure delivery of information to an individual. On the other hand, information encrypted using an individual’s private key can be decrypted using an individuals public key. Though this is not inherently secure this does offer assurances to the receiver that the message is genuine and originated from the desired individual.

**Virus Detection** – Computer viruses are as much a threat to corporate information as hackers. All servers and workstations that are used to support and access the data warehouse should have an appropriate level of virus protection.

**Monitoring/Auditing** – Unauthorized access to corporate information can go undetected if the access to the data warehouse is not monitored or audited. A combination of Web site analyzers and database analyzers can be used to monitor: who visited the Web site, for how long, and what path was followed (Meyers).

**Directory Server/Lightweight Directory Access Protocol (LDAP)** – LDAP is a standard mechanism for accessing a directory server which serves as a central repository that controls access to network resources such as servers, printers, host systems, etc. The directory server offers the following (Kimball):

- Single point of access, administration, and auditing
- Single point of authorization
Secure Sockets Layer (SSL) – SSL is an industry standard protocol that supports secure communications of the Internet using primarily private key encryption. The major Internet browser vendors support SSL within their products. During initial communications both the server and the client agree on a private encryption key under which communications will be secured. To maintain secure communications the key is changed frequently.

Virtual Private Networks (VPN) – VPN is technology that creates secure communications path on a public network between two points. This technology is frequently used for Extranets and to support secure communications to internal networks over the Internet.

In Ralph Kimball’s book “The Data Warehouse Toolkit”, it is mentioned that the goal should be to protect the contents of the information that flows over the network. This protection helps for both Internet and Intranet communications. This protection is accomplished using connection-based encryption of the information that flows between the client and the server. Secure communications between parties can be accomplished using public key encryption using the following technique:

1. User A encrypts a message using their private key
2. User A then encrypts the message again using the public key of User B
3. The message can now be transmitted from User A to User B
4. User B decrypts the message using their private key
5. User B then decrypts the message again using the public key of User A
6. The message can now be read by User B

Over all security is a major component of building a data warehouse, especially if the information is to be made available over the Internet. Security issue must be understood by the data warehouse architect and addressed appropriately. Data warehouse managers must have a fairly good understanding of security issues, so they can hire and supervise security experts who are dedicated to the data warehouse effort (Kimball, 450).

Scalability

Technologies are emerging that require data warehouses to be scalable on the Web. Data warehouses are growing larger, and the next generation of Very Large Databases (VLDB) of over 1 terabyte will increase hardware and software requirements through increased disk space and processors for parallel processing. Accessing complex data types such as image, voice, animation, and video will require a way of integration with current data warehouses into data mining techniques. Vendors are beginning to realize the importance and inevitability of complex data types and have developed products to make such data more available.

Total-solution versus Multi-vendor Approach
The total-solution approach utilizes one major vendor such as IBM, Oracle, or Sybase as the source of support throughout the design and implementation of a data warehouse. This idea is also known as the best-of-breed approach since the products and services are highly integrated, consistent, and easily managed. The opposite solution would be to use multiple vendors for databases and tools, resulting in a system with products specific to the warehouse needs. However, this requires more effort, time, risk, less power in licensing issues, and dealing with more vendors when problems arise. Choosing a vendor or multiple vendors who are enabling their products for the Web and future enhancements in this area is of vital importance.

**Systems and Network Management Tools**

When searching for vendors to implement a warehouse on the Web, it is of essence to choose one that also provides complete tools for warehouse management. Some vendors which will be discussed later are including tape and storage management, performance monitoring, network configuration, and allocation of processor and bandwidth resources as part of the management tool kit. Many vendors are providing remote access of servers from one centralized point as a way of managing databases. These tools are critical for low cost management.

**Conclusion**

Competition will drive organizations to make their information available over the Web. Corporations need to deliver information further down the corporate ladder to the individuals that are in the best position to make day-to-day decisions for the organization. Using the corporate Intranet and the Internet as the communications vehicle will aid in the facilitation of this information delivery.

Whether companies will be successful delivering information over the Web will be a function of how well they can combine the various Web technologies such as Java, JDBC, XML, security systems and hardware that are necessary to successfully deliver Web content. Technologies such as Java and XML are ever evolving and improving. These improvements will aid corporations in delivering their information over the Web. Hardware and communications capacities are increasing and are making the Intranet/Internet a more viable solution for information delivery.

Database vendors realize the potential of the Web and are gearing their products to better support the Web environment. Improved data transformation technologies allow for the integration of legacy information. Improved datatype support will allow these database management systems to incorporate the information typically found on the Internet such as graphics and multimedia support. Database vendors are increasing performance and the ability to handle enormous amounts of data. If your Web data warehouse project is a success you can expect that the amount of information that the data warehouse will need to support will grow significantly. The addition of JDBC support will allow Java-based applications to connect to the corporate databases over the Intranet/Internet and make Java an excellent candidate for a data warehouse front-end. In addition, pure JDBC support will allow for information delivery to many clients the overhead associated with installing the appropriate database drivers on each client.
As previously stated security concerns need to me mitigated in a Web environment. Information is among one of the most important assets of the corporation. Preventing unauthorized access to this information is a critical task that needs to be accomplished. The ultimate goal of securing the information provided by the data warehouse is the data warehouse manager. Their role is to ensure that adequate time and technologies are applied to securing data. Technologies such as firewalls, proxy servers, LDAP, and VPNs can be used to secure the information. The goal is to implement the right combination of technologies that will secure information on both the corporate Intranet and the Internet.

Many of the technologies that will be considered for a Webhouse implementation are new and are still evolving. Technologies such as XML and JDBC are relatively recent advances that can aid in the implementation of a data warehouse over the Web. As vendors incorporate support for these new technologies into their products, corporations will begin to experiment with their potential in their environment. As experience grows so will advancements. The advantages of data warehousing on the Web are expansive. It is a less expensive way of delivering information to many remote users allowing for time-critical decisions. Many of the geographic, technical, and organizational barriers to information delivery would be removed via web access. Training would be minimal and inexpensive, resulting in fewer requests to information technology staff. A user could drill-down into data and create reports quickly and easily. This new paradigm will enable businesses to gain sustainable competitive advantage by utilizing real-time data to make enterprise-wide decisions.
Bibliography


“SQL Server 7.0: Reviewers Guide”, Microsoft Corporation, November 1998


Angus, Jeff and Tyo, Jay, “Accessible Decision Support”, InformationWeek, September 21, 1998


Bethoney, Herb, “Handhelds Gain Oracle8 Power”, PC Week, November 2, 1998


Colkin, Eileen, “Informix, Sybase Improve in 3Q”, InformationWeek, October 26, 1998


Cornetto, John, “IBM, Sybase latest to put RDBMS on Linux”

Cornetto, Jon, “Microsoft enters the enterprise”, InfoWorld, November 16, 1998


Dalton, Gregory, “XML Examined”, InformationWeek, November 9, 1998

Dovey, Tom and Sweat, Jeff, “Mobility Matters”, InformationWeek, November 23, 1998
Davis, Beth, “Data Warehouses Grow”, InformationWeek, November 2, 1998


Dyck, Timothy, “SQL Makes InRoads”, PC Week, November 23, 1998


Gonsalves, Antone, “XMI to Enable Component Sharing”, PC Week, November 2, 1998


Hammond, Mark, “Sybase's Kertzman: Use the Web to reach (and keep) customers”, PCWeek, April 29, 1998


Hibbard, Justin, “Maximum Value from Customer Data”, InformationWeek, September 14, 1998


Johnston, Stuart J., “Big Servers or Big clients?”, InformationWeek, November 23, 1998

Johnston, Stuart, “SQL 7.0: More power”, InformationWeek, November 16, 1998


Kimball, Ralph, “Coping with the Brave New Requirements”, Intelligent Enterprise, November 1998


Maknuri, Sukan, “Internet Commerce”, Intelligent Enterprise, December 1998


Nicolai, James, “Ellison goes for speed with Oracle8i at Comdex”, ComputerWorld, November 11, 1998

Nicolai, James, “Oracle pitches ‘Internet platform’”, Computerworld, November 10, 1998
Niccolai, James, “Oracle warehouse plan challenges IBM, Microsoft”, Computerworld, November 4, 1998


Reichard, Kevin, “Believe the Hype”, Intelligent Enterprise, December 1998

Ritter, David, “Déjà vu All Over Again”, Intelligent Enterprise, November 1998


Sanchez-Klein, Jana, “SQL Server bows at Comdex”, Computerworld, November 11, 1998

Schwartz, Jeffrey, “Data Warehouse Landscape to Change”, InternetWeek, November 9, 1998

Schwartz, Jeffrey, “DB2 Connects External Files Through Links”, InternetWeek, October 5, 1998


Sweat, Jeff, “Sybase Links Data Sources”, InformationWeek, 1998

Swoyer, Stephen, “Microsoft Bets the Farm on SQL Server 7.0”, MIDRANGE Systems, 1998

Ubois, Jeff, “TDWI Meeting Focuses on Growth of Data Warehousing”, 1998

Walsh, Jeff, “XML announcements pop up everywhere”, InfoWorld, November 16, 1998

Walsh, Jeff, “XML Explodes Onto the Scene”, PC Week, November 23, 1998

Waltner, Charles, “Remote Data Ties”, InformationWeek, November 2, 1999


