Information Integration Software Solutions
March 2003

DB2 Information Management Software

Creating a flexible infrastructure for integrating information

By Holly A. Hayes
IBM Software

Table of Contents

3 ***Executive summary***

3 ***Integration challenges***

4 ***An Internet generation***

5 ***Information explosion and diversity***

5 ***Technological challenges to integrating information***

6 ***A changing environment***

6 ***IBM business integration***

9 ***Single view of data***

10 ***Flexible access***

10 ***Rich features***

12 ***Comprehensive business integration***

14 ***Introducing IBM DB2 Information Integrator***

16 ***DB2 Information Integrator: a server for federated data and replication***

19 ***DB2 Information Integrator for Content: federated access across diverse content***

20 ***Current information management offerings***

23 ***Summary***

Executive summary

In today's e-business on demand environment, integrating information across and beyond the enterprise is a competitive mandate. Initiatives such as customer relationship management, supply chain management and business intelligence are based on successfully integrating information from multiple data sources, both structured and unstructured. Because of organizational or operational constraints, these data sources do not generally lend themselves to being fully replicated or consolidated within a single database. Hence, there is an increased demand for federated access to distributed sources.

Information integration technology plays a crucial role in meeting this demand. Information integration technology enables integrated, real-time access to traditional and emerging data sources, transforms information to meet the needs of business analysts and manages data placement for performance, currency and availability. This, in turn, leads to fast, constant and easy access for customer-facing e-business solutions.

This white paper is intended for IT decision makers and influencers who are evaluating integration technologies. The paper describes IBM's business integration solution, the most comprehensive in the industry. Additionally, it addresses the various integration challenges that businesses face today—including user interaction, application connectivity, process integration, building for integration and information integration. The paper also specifically profiles ways in which the DB2® family of information management products from IBM addresses information integration needs.

Integration challenges

Integrating information for an enterprise is a daunting job. IT executives must deal with pressure from the top to deliver results quickly, meet end-user expectations for ease of access to highly diverse data sources, plan around technology limitations and adapt to a constantly changing environment.

In today's competitive environment, businesses want to:

- Reach a broader customer base over the Internet while integrating new Web applications with existing core business processes
- Enhance the value of portals with information from critical e-business systems, improving the productivity of business users
- Speed product delivery by integrating order processing with internal application areas, such as manufacturing and shipping, as well as with suppliers and other trading partners
- Capitalize on emerging opportunities more quickly by correlating information across competitive analysis, analyst research, sales information and customer demographics
- Integrate all customer information across the enterprise with purchased demographics data to personalize interactions for improved customer loyalty and revenue generation
- Streamline information flow between people, processes and applications to minimize unnecessary work and delay.

To meet these needs, businesses are focusing their IT initiatives on integration. According to a 2002 survey by CIO Magazine, corporate information officers list integration as their top strategic priority.[1]

An Internet generation

The Internet has revolutionized customer expectations for service and information access. Through search engines like Google, Lycos, Yahoo! and others, people have access to a collection of information on virtually any topic. Response times are typically dictated more by connection speed than by any other factor.

Businesses have attracted customer attention and loyalty by bringing together complete views of client portfolios, adding additional value by incorporating interesting and relevant content. The bottom line is that customers expect fast, online access to a holistic view of their portfolios with broad-based, value-added content.

[1] Lorraine Cosgrove Ware, "IT Spending," *CIO Magazine* (March 1, 2002): *http://cio.com/archive/030102/spending.html*

Information explosion and diversity

There is no shortage of content. Digitized information is growing rapidly, seemingly beyond the ability of businesses to manage and leverage it. Industry analysts at the School of Information Management and Systems at the University of California, Berkeley, predict that more data will be generated from 2001 through 2003 than in all of recorded history. According to the Berkeley researchers, "The world produces between 1 and 2 exabytes [1 and 2 billion gigabytes] of unique information per year, which is roughly 250 megabytes for every man, woman and child on earth."[2]

The growth in information combined with the diversity of information sources further complicates the retrieval of useful information. Businesses must access not only traditional application sources such as relational databases, but also Extensible Markup Language (XML) documents, text documents, scanned images, video clips, news feeds, Web content, e-mail, analytical cubes and special-purpose stores (both internal and external). Because of organizational or operational constraints, information from diverse and distributed data sources does not generally lend itself to being fully replicated or consolidated into a single database. Yet hidden information can be uncovered, opportunities more readily recognized and customers better served when information is correlated. Giga Information Group estimates that at least 30 percent of all new e-business applications face the problem of integrating multiple data sources.[3]

Providing coherent access to diverse data is a significant obstacle for most businesses. Not only are technology solutions limited, but there is also the potentially larger problem of gaining corporate agreement to a common taxonomy.

Technological challenges to integrating information

Technology has struggled to keep pace with integration requirements. Businesses cobble together integration solutions to meet urgent needs, only to realize later that the solution lacks scalability, availability and flexibility.

[2] P. Lyman, H. Varian, J. Dunn, A. Strygin, K. Swearingen, "How Much Information?", University of California, Berkeley, October 2000, *http://sims.berkeley.edu/research/projects/how-much-info/*

[3] Giga Information Group, Emerging Internet Data Integration Solutions, November 2000

Technology vendors in a myriad of markets—such as enterprise application integration, data warehousing, enterprise content management, portals and application servers—have begun to shift their focus toward the overall integration problem. This makes it more difficult for a customer to choose the best technology to meet the business needs. Moreover, the niche orientation of point products often makes it difficult to leverage the benefits of one implementation in a subsequent project. Customers may find themselves integrating the integration solutions.

A changing environment

The integration job is never finished. IT environments are in a constant state of flux. New applications are coming online. Release-level changes to packaged applications can cause a ripple effect throughout the infrastructure. There is always the next new tool or technology to try. Investments must be made with an eye to the future. Thus, organizations are emerging within businesses to focus on an integration architecture. Whether called Information Management, Integration Services or Data Architecture, specialized departments within companies are tackling the issue of integrating the business, and defining an integration architecture and infrastructure that will provide the foundation for their future.

IBM business integration

Recognizing the need to bring structure and clarity to the market, IBM has introduced a framework for comprehensive business integration. Such a framework is fundamental to becoming an on demand e-business. IBM defines an on demand e-business as one which is highly responsive to customers, partners and employees; focused on its core competencies and backed by a resilient IT infrastructure. Such a business also offers variable cost structures for financial agility. IBM's framework for business integration is also key to meeting the criteria for an on demand operating environment, which has four essential characteristics: it is integrated, open, virtualized and autonomic.

Fundamentally, integration revolves around people, processes, applications and information. Different integration technologies are necessary for different classes of integration problems. For example, online customer orders must

be enabled through an application, not a database application programming interface (API). Business rules embedded in application programming logic protect the database from inappropriate use. Alternatively, the application that responds with a projected delivery date might well access correlated information across manufacturing and shipping databases, and depend on the data management system to handle the complexity of joins and mask differences among the data sources. As in this example, the best solution often utilizes several technologies, which emphasizes the need for moving easily from one technology to another.

While competitors may only provide niche integration, IBM can deliver comprehensive integration with offerings that work seamlessly together. IBM has had more than 30 years' experience in building and evolving the base technologies for middleware, and enabling these technologies to work together in thousands of different business environments. As presented in Figure 1, IBM has identified 5 integration approaches based on an open services infrastructure that can be used together or separately to address these issues.

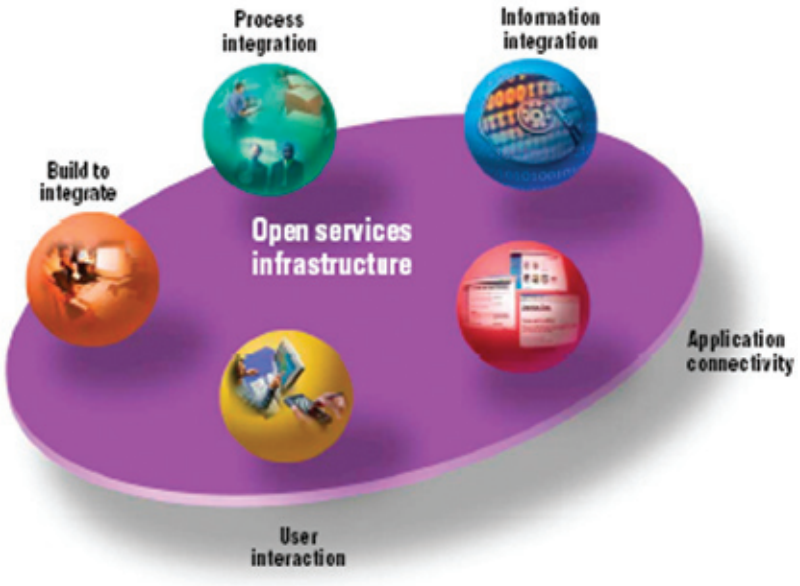

Figure 1. IBM provides an open services infrastructure for business integration.

These approaches are:

- Information integration, which enables the integration of diverse forms of business information across and beyond the enterprise. Instead of sequentially accessing individual information sources, information integration enables coherent search, access, replication, transformation and analysis over a unified view of information assets to meet business needs.
- Application connectivity, which links applications to share and leverage information. Business assets are efficiently connected to allow information on disparate systems to flow across the enterprise.
- Process integration, which changes how businesses automate processes. Process integration takes application connectivity to the next level by allowing a business to change how it operates through the modeling, automation and monitoring of processes across different groups of people and heterogeneous systems, inside and outside the enterprise.
- User interaction, which is about creating a single, interactive user experience across applications and devices. By presenting a single, tailored user interface, making it available through virtually any device, enabling full transactional support and integrating with multiple business systems, businesses can enjoy higher customer loyalty, more effective collaboration and new opportunities.
- Build to integrate, which focuses on building and deploying new integration-ready applications that leverage Web services and existing assets. Instead of traditional silos, new solutions need to be implemented in a manner that enables them to be immediately integrated with existing software assets.

IBM expands opportunity and productivity in a heterogeneous business world through an open services infrastructure that delivers the foundation for open, on demand e-business platforms. Based on industry standards and on a commitment to interoperability, IBM offers the best approach to address today's needs—and to help enable flexibility for the future—with its WebSphere® software platform and DB2 Information Management family.

Single view of data

IBM envisions an information integration infrastructure that will give the application layer a unified view of the data it must access, regardless of differences in data format, data location and access interfaces. The evolution of data management software is not simply about managing single-instance data stores, but about providing value-added integration across all forms of data, dynamically managing data placement to match availability, currency and performance requirements, and providing autonomic features that continue to reduce the burden on IT staffs for managing complex data architectures. To that end, IBM has consolidated a project, code-named Xperanto, to address customer requirements for integrating structured, semi-structured and unstructured data. Based on ongoing research investments and proven data management technologies in areas such as relational data, XML, content management, federation, search and replication, IBM is developing the integrated infrastructure shown in Figure 2.

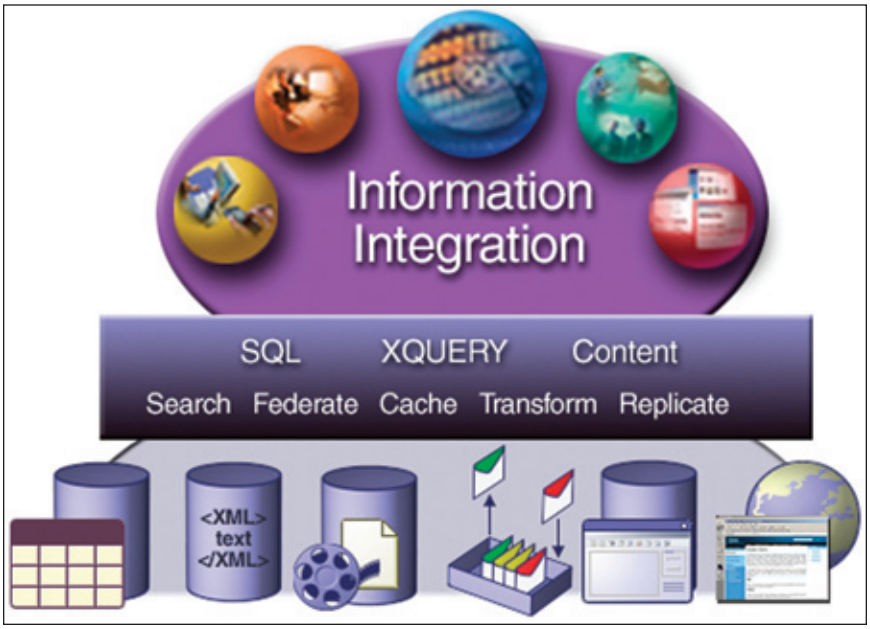

Figure 2. The IBM infrastructure provides flexible access based on a range of programming models, a rich set of integration features and interoperability with IBM's overall business integration framework.

Flexible access

IBM's vision is to provide flexible access to an information integration infrastructure through industry-standard interfaces. Client access may be through ODBC, JBDC, Web services, native client or asynchronous client interfaces. Supported query languages will include:

- Structured query language (SQL), the industry's most mature and powerful query language with widespread adoption in the marketplace
- XQuery, the emerging standard for access to XML data, currently being standardized through the World Wide Web Consortium (W3C)
- IBM DB2 Content Manager object-oriented application programming interfaces that support the content management lifecycle, including rich text and image query.

Regardless of the client access and query language used, the application should be able to access all of the data connected through the integration server. This combination of client access and query language flexibility makes it possible for existing development and analytical tools to take immediate advantage of the broader data access and integration features provided by the integration server. It also allows the infrastructure to plug into service-oriented architectures using Web services, provide asynchronous clients for easy integration with workflows or scheduling long-running queries, and protect your investment in current and new application infrastructures.

Rich features

The IBM information integration infrastructure will enable integration of diverse, distributed and real-time data as if it were a single source, no matter where it all resides. The key features of the infrastructure include the ability to federate, search, cache, transform and replicate disparate data:

Federate. IBM provides industry-leading federation over diverse data sources. Federation is the concept that a collection of data sources can be viewed and manipulated as if they were a single source, while retaining their autonomy and integrity. The resources may be uniform or diverse, collocated or distributed, depending on the implementation. IBM's federation engine provides:

- *Transparency, which helps mask from the user the differences, idiosyncrasies and implementations of the underlying data sources, to make the set of federated sources appear like a single system.*
- *Heterogeneity, which implies the ability to federate highly diverse types of data, including structured data (e.g., relational databases), semi-structured data (e.g. XML documents) and unstructured data (e.g. free-form text).*
- *Extensibility, such that the federation can be extended to almost any data source. Specifically, the extensibility has been designed to minimize the effort required to integrate a new source, yet offer the flexibility to provide the necessary information to optimize query access.*
- *Rich functionality, which includes the functions available through the supported query languages, compensation for missing functions in the backend data sources, plus the ability to surface source-specific capabilities into the query language seamlessly.*
- *Autonomy for data sources such that data sources can be federated with little or no impact on existing applications or systems.*
- *Performance characteristics that make federated query a real-world option. With more than 25 years of research and development and patented optimization technology, IBM can demonstrate the performance that is required to make federation a viable option.*

Search. IBM's infrastructure will provide advanced search and query capabilities including the ability to crawl the Web, index documents, federate search results from multiple search engines, categorize and summarize text documents for intelligent access, and understand semantics. In 2002, IBM formed the IBM Search and Text Analysis Institute to unify and accelerate IBM's research and deployment of advanced search and mining functions through an integrated architecture. The results will fuel IBM's information integration platform as well as other IBM offerings.

Cache. IBM's information integration infrastructure will support placing and managing data at multiple points in the data hierarchy to improve performance. Beyond simple caching, this is policy-based data placement and management. A range of caching strategies is required to provide adequate performance, currency and availability characteristics to requesting applications.

Transform. The infrastructure must provide rich transformation features to facilitate analysis, interchange or presentation.

Replicate. Replication is required as a fundamental characteristic of an information integration infrastructure. It complements the distributed access features, enabling management of centralized data stores, and provides the necessary infrastructure for efficiently managing data caches.

Comprehensive business integration

As referenced earlier, information integration is just part of an overall business integration infrastructure. To support a business's evolving requirements, this infrastructure is complemented by robust data and content stores, as well as additional integration technologies, and is based on industry standards.

Storing the data: IBM today offers leading relational database management systems and content management systems, and is leading the industry in delivering integrated XML support. Beyond current capabilities, which are based on the relational data model, the XML store must completely embody and exploit the XML data model. A built-in XML registry is planned to provide easy management of the vast number of XML artifacts, such as XML schema documents, document type definitions (DTDs) and Web services description documents that can be generated as the quantity and variety of XML data increases.

Leveraging complementary integration technologies: Key to developing a corporate integration infrastructure is the ability to easily leverage appropriate integration technologies, together or separately. IBM continues to focus on integration across the IBM Software Group portfolio—and most importantly with the WebSphere business integration portfolio—to deliver a comprehensive business integration infrastructure.

The WebSphere software platform provides an open, integrated Java™ technology development environment for database and federated database applications. At the same time, IBM is investing in efforts to work well and integrate with Microsoft VisualStudio. In short, IBM offers the broadest, most complete integrated solution in the industry.

Support for industry standards: IBM continues to be at the forefront of industry standards development and adoption, facilitating broad interoperability among vendor tools.

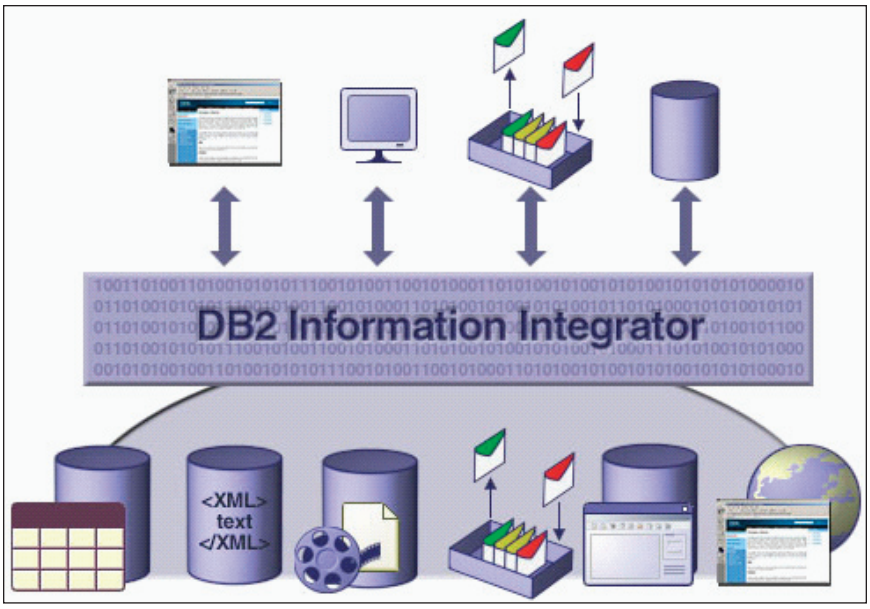

Figure 3. The DB2 Information Integrator products provide integrated access to diverse, distributed and real-time data as if it came from a single data source.

Introducing IBM DB2 Information Integrator

IBM recently announced the beta for the IBM DB2 Information Integrator software shown in Figure 3. This software provides the foundation for a strategic information integration framework. Such a framework helps customers to access, manipulate and integrate diverse, distributed and real-time data. The portfolio consists of:

- IBM DB2 Information Integrator, V8.1, a new product based on DB2 information management technology
- IBM DB2 Information Integrator for Content, V8.2, formerly IBM Enterprise Information Portal.

Each of these products enables customers to abstract a common data model across diverse and distributed data and content sources, and to access and manipulate them as though they were a single source. Each backs a user community defined primarily by the data its members access and the development community they support. This product set supports the predominantly read-access scenarios common to enterprisewide reporting, knowledge management, business intelligence, portal infrastructures and customer relationship management.

With the DB2 Information Integrator family of products you can:

- *Choose the data access strategy to match the business value.* Consolidating data local to the application makes application development easier and provides better data access performance and availability. However, it also introduces the burden and cost of moving data from place to place, storing the data and managing its synchronization. DB2 Information Integrator V8.1 delivers centralized access by providing replication and caching to support performance and availability requirements. Alternatively, when there is wide diversity in the data accessed, it is impractical or too expensive to replicate the data, or when the data is owned outside the enterprise, a better approach may be to access this data in place.
- *Integrate data and content without moving the data or changing the platform.* The DB2 Information Integrator family lets you access diverse and distributed data no matter where it resides, as though it were located in a single source. It provides a broad range of data source access out-of-the-box, covering structured and unstructured data across and beyond the enterprise. With more information available more easily, you have the opportunity to gain greater return on existing information assets.

- *Make more progress, more quickly and at a lower cost.* With the DB2 Information Integrator products you can more easily and quickly develop the new generation of composite applications that require efficient integration of disparate data. Developers can choose whether to use an SQL or DB2 Content Manager programming model based on the product selection. And now, you have a practical way to integrate diverse relational data and combine it with unstructured data from content repositories, the Web, spreadsheets, etc. Thus, the DB2 Information Integrator offerings can help you speed project deployment, leverage existing skills over a broader range of projects and reduce ongoing maintenance costs.

DB2 Information Integrator: a server for federated data and replication

DB2 Information Integrator V8.1 is targeted at the application development community familiar with relational database application development. Applications that use SQL, or tools that generate SQL (e.g., integrated development environments, reporting and analytical tools, etc.), can now access and manipulate distributed and diverse data through a federated data server. This product is most appropriate for projects whose primary data sources are relational data augmented by other XML, Web or content sources.

DB2 Information Integrator V8.1 is based on the DB2 technology infrastructure, leveraging IBM's prior investments in products such as IBM DB2 DataJoiner®, IBM DB2 Relational Connect and IBM DiscoveryLink® (as well as subsequent product enhancements). DB2 Universal Database consists of a modern database architecture known worldwide for its scalability and extensibility. DB2 exploits parallelism within a single system and over clustered systems, and can support thousands of concurrent users and terabytes of data. Its object relational underpinnings provide an extensible framework for adding new data types with tailored functions. DB2 runs on numerous platforms worldwide and leads the industry in exploiting standards. Basing DB2 Information Integrator on the DB2 technology underpinnings results in a platform-independent, standards-based, high-performance, scalable, reliable, available and manageable information integration infrastructure.

DB2 Information Integrator includes the ability to federate, search, cache, transform and replicate data. As a federated data server it provides out-of-the box access to DB2 Universal Database, including DB2 Informix® Edition products, as well as databases from Microsoft, Oracle, Sybase and Teradata. In addition, it can also access semi-structured data from WebSphere MQ messages, XML documents, Web services, Microsoft Excel, flat files, ODBC or OLE DB sources, plus a variety of formats unique to the life science industry. Integrated support for IBM Lotus® Extended Search provides the solution's broad content access to a variety of content repositories, including DB2 Content Manager, as well as e-mail databases, third-party Internet search engines and LDAP directories. In addition, a developer's kit extends the federation capability to virtually any data source.

Search and query access is provided through a standard SQL API and combines the broad content access of Lotus Extended Search with the precision of a relational engine. The query may produce standard SQL answer sets or XML documents. The federation engine includes a cost-based optimizer that takes into account not only standard database statistics and indexes, but also network and server resources and the query power available in the data source engine. Two search approaches for text are available:

- The ability to create a global index for backend relational stores. Using this approach, text search semantics—such as fuzzy search, thesaurus support and search within sections—may be used within the query.
- A brokered search architecture that does not require the creation or maintenance of a central index to access content across multiple sources. The extended search engine translates each full text query into the native query language of the target data source.

The first release of federated caching provides administrator-managed caching of integrated views across relational database backends. The optimizer automatically routes queries to the cache to satisfy queries when appropriate.

A rich set of transformation features includes standard SQL functions, such as string manipulation, arithmetic calculations, statistical computations, online analytical processing functions and procedural logic. Type-specific features—such as the application of scoring algorithms, spatial analysis or chemical similarity searches—further enhance this already rich set of transformations. Extensible stylesheet language (XSL) translations facilitate document interchange and dynamic style matching to diverse display characteristics. User-defined functions enable customers to standardize virtually any function for any data type. In addition, the ability to access Web services as a built-in function means that any Web service, such as a currency conversion, can become an embedded transformation function.

DB2 Information Integrator V8.1 also includes a replication server for mixed relational databases. Customers can replicate data between IBM (DB2, including DB2 Informix Edition products), Microsoft, Oracle, Sybase and Teradata (target only) databases. Customers can flexibly configure a variety of topologies, latencies and consistency characteristics.

In a recent CRN article, "IBM: Xperanto Rollout to Start In Early 2003"[4], industry analysts lauded IBM's upcoming offering:

- IBM's big advantage will be its reliance on its own optimization logic "which is second to none," and its ability to handle structured and unstructured data. [Doug Laney, Vice President Application Delivery Strategies at Meta Group]
- Philip Russom, research director of Data Integration at Giga Information Group, agreed: "IBM will raise the bar on enterprise information integration. It has so much more to offer than small companies, and it has broadened the category to include unstructured data."

[4]For more information, please visit: *http://www.crn.com/sections/BreakingNews/breakingnews.asp?ArticleID=39187*

DB2 Information Integrator for Content: federated access across diverse content

DB2 Information Integrator for Content V8.2 is targeted at the content application developer who needs to search for and access text and non-text information across a wide range of content sources. Providing seamless reach into diverse data environments, DB2 Information Integrator for Content represents a renaming and repositioning of the Enterprise Information Portal offering.

DB2 Information Integrator for Content offers a rich set of integration features, such as connectors to diverse content sources, sophisticated information mining and advanced workflow. To speed implementation time of content integration projects, DB2 Information Integrator for Content provides access to a variety of data sources out-of-the-box—all of which can be federated into a single search. These connectors include access to the DB2 Content Manager family and other content repositories, Lotus databases, relational databases and wide-ranging content available with Lotus Extended Search.

Additionally, DB2 Information Integrator for Content includes a sophisticated information mining capability that uses Web crawling and text-mining algorithms to provide structure to unstructured content. The mining algorithms include the ability to identify the language in which a document is written, identify such features within documents as names, classify documents according to a defined taxonomy, group documents by category and summarize documents. By building additional knowledge about enterprisewide information, businesses can reap additional return from existing content assets.

Finally, DB2 Information Integrator for Content provides an advanced workflow application to enable businesses to increase productivity, reduce production times, and improve communication and collaboration. Using a graphical workflow builder, developers easily define workflow processes across the enterprise.

The DB2 Information Integrator portfolio addresses the need to access and integrate structured and unstructured data. Supporting today's popular programming models, each offering enables businesses to quickly capitalize on their existing skills and tools infrastructure in enterprise content management and SQL-related deployments. IBM has a compelling vision, a significant research and development investment, and a roadmap for the product family that represents the evolution of data management technology.

Current information management offerings

The DB2 Information Integrator offerings have evolved from a rich set of technologies that are available in the market today. These proven solutions include:

- DB2 Universal Database, which provides the foundation for storing, managing and federating data. DB2 includes a rich set of statistical and analytical functions and offers the best optimization technology in the industry, including distributed optimization and automated caching functions. DB2 provides traditional client access, as well as access from Web service clients. According to Bloor Research, "It is also worth noting a major difference in approach that differentiates IBM from Oracle. IBM takes the view that DB2 should support the ability to integrate and federate information rather than centralize it. This is in direct contrast with Oracle, which is pushing centralization."[5]
- DB2 Extenders™—including DB2 XML, Text, Net.Search, Audio, Visual, Image and Spatial Extenders—which provide data type-specific extensions to query, access, update and manage various data objects. For example, with DB2 Extenders, you can manipulate XML documents, query by image shape or color, or query based on proximity to a given location. Consider Satellite Records, which plans to use IBM DB2 XML Extender to process XML-based documents from distributors and store them in DB2. The company will thus be able to easily move vinyl record information it receives from distributors

[5] Bloor Research, "Databases: an evaluation & comparison," January 2002

directly to its own Web site. Steve Shapero, IT director at Satellite, estimates that DB2 XML Extender will ultimately save the company 40 percent in development time and costs by providing a framework for its extranet, reducing the amount of new code that the company would otherwise need to write.[6]

- DB2 DataPropagator™, a replication engine supporting point-in-time and near real-time replication with embedded transformation capabilities for populating data warehouses and data marts. IBM IMS® DataPropagator extends replication to IMS systems and IBM DB2 DataJoiner extends replication to Oracle, Sybase, Microsoft and Teradata databases. One IBM customer, s.Oliver, estimated that the lack of inventory visibility was costing the company approximately four percent of its potential sales. It uses DataPropagator to replicate order and inventory information between staging and production servers. According to Jose Monteagudo, CIO at s.Oliver, "Although technology changes rapidly, we cannot afford to chase every fad and fashion in e-business. The software and hardware infrastructure we've built with IBM has given us the availability, scalability and reliability we can count on to help ensure a fair return on our current and future e-business investments."[7]
- DB2 Relational Connect (available only with DB2 Universal Database, V7) provides the source-specific wrappers that extend the reach of the core federation engine to heterogeneous relational databases, including DB2 (as well as DB2 Informix Edition products), Oracle, Microsoft and Sybase. Based on years of research and proven experience, the DB2 optimizer provides unmatched high-speed access, not only to a single, local database, but to diverse and distributed databases. Automatic compensation built into DB2 helps mask the differences between the DB2 capabilities and the data source capabilities without any loss of function.

[6] Satellite Records success story at *http://www-3.ibm.com/software/success/cssdb.nsf/CS/NAVO-5C6TVN?OpenDocument&Site=dmmain*

[7] s.Oliver success story at *http://www-3.ibm.com/software/success/cssdb.nsf/CS/NAVO-4ZUQ2S?OpenDocument&Site=dmmain*

DB2 Relational Connect is a component of DiscoveryLink, a solution that enables integration and analysis of large quantities of diverse scientific data from a variety of life sciences domains. Extensions for the life sciences industry embedded in DiscoveryLink extend access to Excel spreadsheets, flat files and proprietary genomic and proteomic data stores. These extensions also demonstrate the flexibility that is needed in extending the technology to arbitrarily complex data stores. According to Peter Loupos, global head of Drug Innovation and Approval Information Systems at Aventis, "DiscoveryLink allows us to access and mine the physical data in a way never before possible, significantly speeding up the drug discovery and development process."[8]

- DB2 DataJoiner, IBM's first-generation federation engine, additionally delivers read and write capabilities, two-phase commit across heterogeneous relational sources and heterogeneous replication among DB2, Oracle, Sybase, Microsoft and Teradata. As Bloor research says of DB2 DataJoiner, "Also at the global level, DB2 DataJoiner has an optimization component. That is, it aims to optimize SQL according to whichever data source it is addressing. This is unusual. Most vendors tend to give their own products an edge when it comes to optimization. IBM appears to be taking a genuinely dispassionate approach in this respect, which we commend."[9]
- Enterprise Information Portal, IBM's second-generation federation engine primarily targeting unstructured data. Hadley Reynolds, research director at the Delphi Group, states, "…IBM's … [Enterprise Information Portal] will add important new capabilities to the functionality available to portal implementers." Referring to those new features, Reynolds adds, "The integration of unstructured and structured information sources is a huge challenge for virtually all business portals, and adding in-the-box APIs for structured data sources, richer unstructured content facilities, and full development platform and Web application support streamlines an already rich platform offering."[10]

[8] Aventis success story at *http://www-1.ibm.com/industries/healthcare/doc/content/casestudy/318234105.html*

[9] Bloor Research, "Databases: an evaluation & comparison," January 2002

[10] Sandra Haimila, "Managing knowledge at Heritage Mutual," *KMWorld Magazine* (June 26, 2001): *www.kmworld.com/resources/featurearticles/index.cfm?action=readfeature&feature_id=130*

- DB2 Warehouse Manager and other partner offerings round out IBM's information integration portfolio in the area of extract, transform and load technology. According to Sherilyn Jensen, data warehouse manager, Hillman Group, "DB2 and DB2 Warehouse Manager have proven to be a powerful combination in helping us get the most value from our data, enabling us to recoup our investment within two years."[11]

Summary

Businesses today need to integrate information to drive customer loyalty and satisfaction, improve operational efficiency, compete for online customers and trading partners, and identify and respond to emerging opportunities. In short, information integration provides a competitive advantage.

A complete integration architecture should include multiple technology approaches to address the range of integration challenges businesses face. IBM understands this, providing the most comprehensive business integration solution on the market. And IBM has heard and understands the requirements for integrating diverse data. D.H. Brown and Associates sums it up: "IBM's goal is to provide a comprehensive, scalable and open database platform and infrastructure for building the new generation of Web-based database applications, while allowing users to rely on their existing databases, including those from other vendors. DB2 is, perhaps, better positioned than any of its competitors to achieve its goal because its platforms, infrastructure and middleware are so comprehensive and able to seamlessly integrate DB2 with all heterogeneous databases."[12] Indeed, with its new DB2 Information Integrator portfolio, IBM continues to drive state-of-the-art innovation in technologies that enable businesses to take advantage of all of their information assets.

[11] Hillman Group success story at *http://www-3.ibm.com/software/success/cssdb.nsf/CS/NAVO-4X323B?OpenDocument&Site=default*

[12] D. H. Brown and Associates, "Database Trade-offs for IBM and Oracle: Availability, Scalability, and Performance," September 2001

For more information

Please contact your IBM marketing representative or an IBM Business Partner, or call 1-800 IBM CALL within the U.S. Also, visit our Web site at:
ibm.com/software/data/integration

IBM Corporation
Silicon Valley Laboratory
555 Bailey Avenue
San Jose, CA 95141
U.S.A.

Printed in the United States of America
03-03

Printed in the United States on recycled paper containing 10% recovered post-consumer fiber.

GC18-7560-00