

Information Integration and XML in IBM's DB2

Pat Selinger

IBM Silicon Valley Lab
555 Bailey Ave.
San Jose, CA
USA
pgs@us.ibm.com

Customer Challenges in Information Integration

In today's e-business environment the demand to integrate information across and beyond the enterprise is a competitive mandate. Bottom-line initiatives such as customer relationship management, supply chain management, and business intelligence are predicated on successfully and rapidly integrating information from diverse data sources. Customers want to:

- Reach a broader customer base over the Internet while integrating new Web applications with existing core business processes.
- Speed product delivery by integrating order processing with manufacturing availability and shipping capability.
- Capitalize on emerging opportunities more quickly by correlating information across competitive analysis, analyst research, sales information, and customer demographics.
- Integrate all customer information cross-enterprise with purchased demographics to personalize all customer interactions for improved customer loyalty and revenue.
- Streamline information flow between people, processes, and applications to eliminate unnecessary work and delay.

To accomplish these, businesses are focusing their IT initiatives on integration. According to a recent survey by Morgan Stanley, CIO's list integration as their top strategic priority.

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the VLDB copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Very Large Data Base Endowment. To copy otherwise, or to republish, requires a fee and/or special permission from the Endowment

**Proceedings of the 28th VLDB Conference,
Hong Kong, China, 2002**

IBM Data Management Solutions

Due to organizational or operational constraints, the diverse data sources an enterprise uses do not generally lend themselves to being fully replicated or completely consolidated under a single database -- hence the increased demand for data interchange and for federated access to distributed sources. IBM has ongoing work in information integration technology that 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 leading to fast, constant, and easy access for customer e-business solutions.

IBM's Information Integration infrastructure today supports SQL, a mature, powerful query language, plus a number of SQL extensions in support of XML. SQL and SQL extensions for specific datatypes (such as XML or image) support access to any type of data object: structured, semi-structured, and unstructured.

IBM will also support XQuery, the emerging standard for access to XML data. XML is a World Wide Web Consortium (W3C) standard that packages data along with its metadata, making XML self-describing. XML provides several advantages for customers:

- It enables business to business communication, because XML makes it easy to send structured data across the web so that nothing gets lost in translation, masking differences in end-point infrastructures.
- It enables information to be rendered appropriately for a variety of devices, because XML separates the document content from the document presentation.
- It enables smart searches, because it provides a context for search arguments, e.g. the name Chip rather than the food chip.

IBM's information integration capabilities are aimed at making it easy to access any information from any client. Thus, in addition to traditional database clients, IBM supports access from Web service clients. Any SQL statement, stored procedure, or XML request can be accessed from a Web service.

DB2 Capabilities

It is our strong conviction that customers will be best served by an integrated database engine, supporting both SQL and XML, with a feature set that includes the following:

- **Relational and XML Store:** the ability to store and retrieve structured relational table, semi-structured XML documents, and unstructured content such as byte streams, scanned images, etc.
- **Search:** advanced search and query capabilities. The query processor provides industry leading cost-based optimization for high speed query response
- **Federation:** support for single request query over multiple diverse data sources. Federation is the concept that a collection of resources can be viewed and manipulated as if they were a single resource while at the same time retaining their autonomy and integrity. The resources may be uniform or diverse, co-located or distributed depending on the implementation.

This session's presentation will focus on the XML support that has been and is being added to the DB2 product family as part of its information integration work.