

# DATABASE

## TRENDS AND APPLICATIONS

Solutions for the Information Project Team • [www.dbta.com](http://www.dbta.com)

Volume 21, Number 8 • August 2007

## XML Powers Advanced Integration

By Joe McKendrick

It was well over a decade ago that XML was first introduced as a lingua franca that could bring together even the most disparate data environments. While it has become a fairly ubiquitous part of the enterprise landscape, has it lived up to its promises? A mixed picture emerges. Many companies are leveraging the standardization that is possible through XML-based interfaces to link up and better integrate with partners' systems. But challenges still remain, and adoption still touches relatively few applications.

In one sense, XML is helping to transform the industry as we know it, well beyond its original predictions. Nowhere is this more apparent than in the rise of software as a service, in which data and calls to applications can be seamlessly exchanged across corporate boundaries. "We wouldn't as a company be able to have the success that we have, nor deliver the kind of results that we have for our customers, without XML," said Adam Gross, vice president of developer marketing for Salesforce.com. "XML powers the advanced integration capabilities for our largest sophisticated enterprise accounts, like Dell, Cisco, and Merrill Lynch."

Indeed, whether it's for the external enterprise, or inside-the-firewall deployments, there is still no shortage of enthusiasm over the potential benefits XML can deliver. One company that serves the publishing industry, Mark Logic, reports that all of its customers are either already using XML or are interested in adopting it as a new standard for managing content. "Companies are increasingly becoming

excited about what XML will enable them to do with their content," John Kreisa, director of product management at Mark Logic, told DBTA. "Organizations are realizing the importance of storing the information about their data along with the data itself, and that this is the key to enabling content re-use and repurposing."

Companies are increasingly becoming excited about what XML will enable them to do with their content.

### Standardized Integration

For many companies concerned with data and application integration, there simply has been no better way than through XML. "XML provides a standards-based way for data to interact between applications, databases, and even legacy operational systems," Scott Gidley, co-founder and chief technology officer for DataFlux (a subsidiary of SAS), told DBTA. "From a vendor perspective, we can focus on development of data quality and data management services instead of developing APIs for specific platforms, applications, and programming languages. From a customer perspective, companies have a standardized method of application integration that doesn't change from vendor to vendor."

Salesforce's Gross observes that his company now handles 50 percent of all of its transactions with XML. "All of the actual requests into our servers and

our data center, are now XML and SOAP, as opposed to HTTP and HTML Web browser requests. We actually do more XML than we do non-XML transactions," he said. As recently as four years ago, all transactions were HTML, Gross said.

"One of the things that people frequently have questions about is integrating with our service," Gross explained. "How do you integrate with an application that lives out in somebody else's data center? For a lot of our customers, integration is a key requirement. XML and Web services and SOAP have allowed us to do that."

### Adoption Rates

However, while adoption of XML is wide, it is not deep. Recent data from Evans Data, for example, finds that 61 percent of applications at 400 surveyed companies use XML in at least some of their applications. However, only three percent report that XML is now supported by the majority of their applications.

For example, Bart Grantham, director of software development for LogicWorks, told DBTA that his company currently uses "XML for very little of our internal development work, generally restricted to AJAX for client-side work." Grantham pointed out, however, that his company's commercial products use XML.

Kreisa agrees that the adoption is still relatively low, noting that XML "is still a relatively new format." He added, however, that "with the adoption of XQuery as a standard query language for XML, organizations will now have the ability to build applications and more fully leverage their XML content.

Consequently, we see more and more organizations creating their content directly into an XML format.”

What are some of the issues impacting XML? Grantham stated that the markup language adds “a great deal of overhead to the processing of data as well as complexity for software development. Not all data structures map cleanly onto trees, and many data stores do not need human readability or editability.”

“One of the biggest challenges with XML, besides separating the hyperbole from the reality, is syntax fragility,” Grantham noted. “Solid libraries do much to alleviate this problem, but as a data format there is much that can go wrong in syntax and parsing of XML. XML’s strength is in being human-readable, and leveraging the industry’s familiarity with SGML-derived languages. It is also quite flexible. But, in my opinion, it should not be a first choice for a machine-to-machine data format, due to processing and memory overhead tradeoffs.”

There are situations where data is better left as it is, and not converted to XML.

### Performance Issues

Appliance solutions on the market, such as IBM Datapower, are designed to offload XML processing overhead from servers and onto dedicated hardware that reside on the network.

However, not everyone believes XML will drag down performance. Mark Logic’s Kreisla, for one, believes hardware capacity will keep up with XML-based workloads. “With the continued reduction in storage costs we rarely see this as a consideration regarding converting existing content to XML,” he

explained. “Organizations should look for a content server or content base that has the scalability and performance to be able to handle the high volume of XML content.”

The industry has been responding with new approaches, including the introducing of Binary XML by the World Wide Web Consortium (W3C), which delivers XML capabilities as object code, rather than as more human-readable - but more verbose - text.

The most pronounced issue with XML, as cited by the Evans Data research, included the ability to write XML schemas and document-type definitions (DTDs), which are the building blocks of XML-based documents. One out of four enterprise developers say this is an issue, along with one out of five who feel that XML syntaxes create performance overhead, especially at the server level for XML parsing.

Issues with semantics also hamper XML adoption as well. “XML has helped standardize the ‘syntax’ for sharing data between systems, but has not addressed the far more important issue of semantics - what the data means,” Cliff Longman, chief technology officer of Kalido, told DBTA.

“It is as though XML has allowed a phone connection between two people, but if one speaks English and the other speaks Chinese, the phone connection does not help much. It is semantic interoperability on top of XML that has become the new battleground for standardization at the enterprise level,” Longman said.

“Metadata really is one of the dirtier aspects of information integration,” Michael Curry, director product strategy and management, IBM Information Platform and Solutions, told DBTA. “For example, a business might refer to customer information in one database with the phrase ‘customer ID,’ and put the same information under the phrase ‘customer account number’ in another database. This adds to the confusion.”

### Other Challenges

Other issues dampen XML adoption as well. Xiong Wang, associate professor in the Department of Computer Science at California State University, Fullerton, notes that the industry still suffers from “a shortage of mature XML authoring tools, a lack of stable standards, and numerous customizations required to make software solutions work.”

XML is not a good fit for structured data formats as well, Wang stated. “With structured data XML induces too much unnecessary redundancy. RDBMS on the other hand is a perfect fit in such data,” he said. In addition, Wang continued, “lack of efficient storage structures is another challenge with XML.”

There are situations where data is better left as it is, and not converted to XML. “It’s not necessary to use XML in situations where the data has a regular structure and fits neatly into rows and columns,” Kriesa pointed out. “While it is possible to use XML for this type of content, you’re probably not fully leveraging the strengths of the XML standard.” Clearly XML has to be thought of as an option that must be applied appropriately.

In addition, the RDBMS vendors are increasingly addressing XML integration. For example, IBM’s recently released DB2 9 (“Viper 2”) is considered as a “hybrid data server to serve data from both pure relational and pure XML structures,” Bernie Spang, director of IBM data servers, told DBTA. IBM’s intention is to “lower development time and cost savings that makes ‘XML as data’ cost-effective for the first time.”

DB2 accomplishes this functionality by storing XML data in a hierarchical structure that naturally reflects the structure of XML, which allows DB2 to efficiently manage this data and eliminate much of the complex and time-consuming parsing required for XML,” Spang said.