

Data Management? Two Perspectives



Sandeepan Banerjee

DBTA: XML is hot. What needs to be done in the development cycle?

Banerjee: The major foundations are in place for XML. On the standards side, the big pieces, except for XML query, are out. The adoption of the XML data model, which is called XML schema, last year was a big achievement. Other standards to transform XML, to bind XML to Java and to invoke XML remotely are already in place. We have also seen XML support gradually permeate the enterprise software stack. The Oracle database, for example, supports XML schema. The mid-tier supports XML access from Java. Browsers can display XML, and so on. Various industry consortia are describing industry-specific XML-based document exchange and processing for legal, financial, government, trade and other applications. There are also cross industry efforts.

DBTA: How are these standards being brought to bear today?

Banerjee: We are seeing mainstream customers deploying XML solutions. So in my view, XML is coming along nicely, especially in the area of exchanging data and instructions.

DBTA: What challenges lie ahead?

Banerjee: The focus of XML is shifting to the content space. In this space, a big wave is beginning to emerge using XML to manage enterprise content and multi-channel publishing of that content. People are trying to take all the unstructured assets they have and see if marking them up in XML unlocks their value in ways that have not been done before.

DBTA: What needs to be done to make that happen?

What does XML mean to the relational database vendors and users? We talked to Sandeepan Banerjee, the director of product management for SQL, Text and XML at Oracle Corp. The short answer — the RDBMS will become a data integration platform upon which companies will manage their structured, unstructured, and semi-structured data.

Banerjee: It means work on XML repositories, XML-based content authoring and content composition, XML text-based search and so on. This is really the work in progress for XML.

DBTA: What remains undone?

Banerjee: For the exchange of data and instructions, XQuery is the major piece that is undone. All the other standards are going to go through incremental revisions. But the versions that are out are pretty solid. And customers are deploying real applications.

DBTA: Has the fear that incompatible flavors of XML would emerge dissipated?

Banerjee: There is always the issue of if everything is standard, how do vendors differentiate themselves. I think that XML has done a good job of balancing the core, common pieces, which are standard, adding enough extensibility so that vendor-specific increments can be made. XML had done a good job in the standards we have seen so far in harmonizing the general and the specific.

DBTA: So this is a standards effort that is really on track.

Banerjee: By and large, I think so. We are satisfied.

DBIA: For database vendors, XQuery is going to be significant, isn't it?

Banerjee: XQuery certainly has a lot of interest in the market. We think that the proposed standard in the W3C is targeted toward querying XML data and will be important in the future, when people have a lot of XML data. We are on the XQuery committee. We have made some of our earlier implementations downloadable from our developer site. But we think XQuery will first be useful in the mid-tier. In the database tier, SQL will continue to rule because it has been optimized for multi-user, concurrent, volume access to structured data. What you will see in the database tier is some gradual SQL and XML interoperability. The bigger story will be in the middle tier, where there is limited or no query today. That is where XQuery will debut.

DBTA: Will XML be a functionality absorbed into relational databases? Some people argue that because XML is hierarchical, it is not the best fit with relational technology. What will XML mean to relational database vendors and what will it mean for new alternatives in data management?

Banerjee: First let me take the assertion

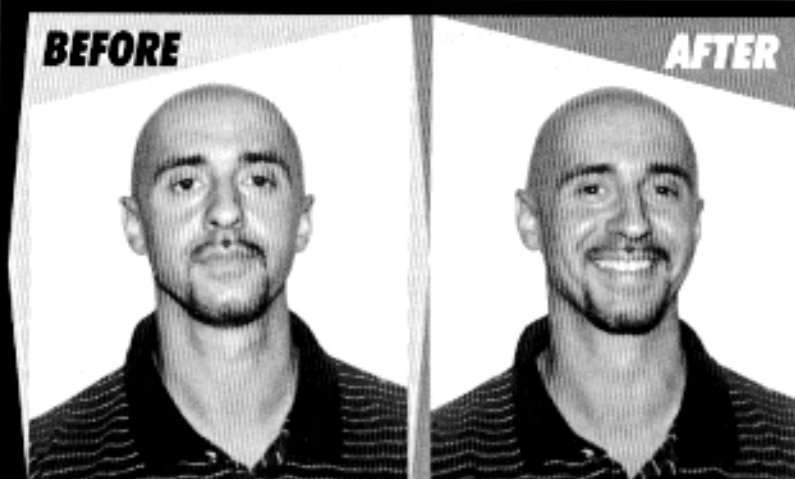
that XML does not map well into a relational database. We have solved this problem at Oracle. Seven years ago, with Oracle 8.0, we started to enhance the SQL data model to allow for nesting and hierarchical structures. There was a 1999 version of SQL that provided for nesting tables within tables and so on. With 8.0, 8i and now with 9i, we added more and more of the SQL 99 mechanisms. We can deal with hierarchical data inside the database. We call this the extended relational or SQL 99 model. To this, we have added specialized indexes performing very efficient directory-path-like traversals for XML documents.

DBTA: And the result has been?

Banerjee: We can completely represent the XML schema data model within Oracle 9i release 2, using our extended relational mechanisms. The leading banking software platform, a partner of ours, recently benchmarked 22 million XML financial transactions per day using this infrastructure. So it is efficient. These are loads we can churn out with relative ease.

DBTA: Some people have suggested that with a standard data interchange, it will be easier to add all

Banerjee Q & A page 28



BEFORE **AFTER**

Attitude Adjustment

Three out of Four PICK developers experience immediate attitude improvement after using **AccuTerm K2** for their GUI projects.

Check out our new upgraded features! @ www.asent.com

8041 Foothill Blvd., Sunland, CA 91040
818-951-1891 FAX 818-951-3606

AccuSoft
ENTERPRISES

Banerjee Q & A

Continued from page 23

kinds of non-conventional systems to the information infrastructure.

Banerjee: Exchanging data does not eliminate the need to optimize access to individual data stores, nor to make the XML world interoperate with the SQL world and the Java world. XML will make it easy to interchange data between systems, as they all support XML. And the database will address XML. And this will be especially valuable for unifying access to structured and unstructured content. You will see the mainstream relational databases will become more prominent as data integration platforms. They will all address documents and messages in addition to tables.

DBTA: What will XML mean to the Oracle family of products?

Banerjee: XML is fundamentally an interesting new set of standards for data management. XML schemas are the first standards the world has had that can unify structured and unstructured data. In our strategic vision, we want to be the best software platform for managing XML, both the best implementation of open standards and the best platform for managing all of your data.

DBTA: How do your products reflect that vision?

Banerjee: In Oracle 9i Release 2, we released Oracle XML DB, the name for all the XML features integrated into the standard Oracle server. This release is a big realization of our vision.

DBTA: Were you ahead of the curve with those features?

Banerjee: In the exchange and integration space, I think we hit the nail on the

head. A number of customers are using this. One of the largest PC makers has used this to re-implement their entire order processing system. There is a lot of variability in each order, which would be hard to model using traditional relational technology. These XML orders come in; they are parsed and placed in different order entry systems and move through their life cycle. They are all XML documents moving back and forth. This is well supported using Oracle. Content management will be the focus of forthcoming XML work. We have a number of features that lead the market.

Want more answers from Sandeepan Banerjee about the XML standards process, native XML database vendors, and content management? Visit the In depth section of www.dbta.com.