zapthink
zapnote

Doc. ID: ZTZN-0117-1
Released: Dec. 25, 2001

ZAPTHINK ZAPNOTE™

XYZFIND
A NATIVE XML DATABASE FOCUSED ON SIMPLICITY

December, 2001

Analyst: Ronald Schmelzer

Abstract

XYZFind provides a general-purpose NXD storage and retrieval system for XML documents, regardless of their content or schema. XYZFind aims to solve the problem of search and query over identifiable "islands of data" where the content its too varied to accommodate with file system tools. XYZFind's utilizes a proprietary query language and has a proprietary, text-based storage architecture that utilizes a meta-schema encompassing an aggregate, schema-independent view of the data.

Native XML Databases

When a large volume of data is being created by applications originally in XML format, deconstructing the XML documents for storage in a relational or object-oriented database may not be the best approach. More likely, it would be best to simply insert the document intact into an XML-based "database" system that can then retrieve the document in whole or in parts using XML-based queries. This sort of functionality, namely the ability to interact with a data storage system by simply inserting XML documents intact and later retrieving them intact without exposing a relational or object-oriented database system to the user is what is known as the "Native XML Data Store".

Native XML data stores may themselves use RDBMS or OODBMS systems to store the XML content, but the important distinguishing factor is that these systems are not exposed to the end user. The end user interacts with the system as if it was storing XML documents natively, rather than working with an RDBMS system that happens to store XML documents in addition to relational fields. It's all in the interface.

Native XML Data stores (NXD) are capable of handling XML content in native XML form, rather than in some other representation. Most relational and object-oriented databases decompose XML into relational tables or objects, while NXD store XML in the original hierarchical form. The result is better suited for environments where XML is the main information structure being exchanged, and where the information environment needs to change as XML documents themselves change.

The XYZFind XML Data Store

XYZFind provides a general-purpose NXD storage and retrieval system for XML documents, regardless of their content or schema. XYZFind aims to solve the problem of search and query over identifiable "islands of data" where the content its too varied to accommodate with file system tools.

Work on the product began in July 1999, when they initially saw need for providing structured access to structured data without prior knowledge of the structure of that content. For example, an individual in an organization may need access to salary detail but not know where to get information and how to structure the query to get it. The XYZFind product of the same name provides a universal repository for XML data, and a universal method for providing search & query over the widely heterogeneous structured XML content.

XYZFind Server works in a manner unlike traditional database systems. Arbitrary, well-formed

XML documents can simply be inserted, and the system will be able to retrieve this information. This approach is meant to be as simple as using a file system. However, XYZFind adds a sophisticated and high-powered structured querying and keyword search environment one would expect from a relational database system. In common with other NXD systems, XYZFind's ability to accept arbitrary XML document that may or may not be validated against a DTD or Schema presents a major advantage over XML Extension products such as Microsoft SQL Server 2000 and Oracle 9i. This means that XYZFind users can insert any well-formed XML into the system and it becomes immediately available for query. Since no prior knowledge of XML document structures is required, nor is there a need to tweak database schemas every time a new type of XML document is inserted or an existing XML document's structure is modified.

XYZFind's architecture and storage technology is a proprietary, text-based format that utilizes a meta-schema encompassing an aggregate, schema-independent view of the data.

Support for Emergent XML Database Specifications

What differentiates XYZFind is its use of proprietary indexing and searching methods. XYZFind has produced a proprietary query language that makes optimal use of their indexing scheme, although they plan to announce support for XPath in the future. Their approach is to parse, decompose, and index documents for storage in a central database represented in an "intermediate" schema format. Keyword search and query access is provided over this common store, and documents, document ID's, and/or document fragments are returned from the database as results.

XYZFind doesn't yet support XQuery, XPath or other XML query standards as of yet. However, in response to customer requests, an extended subset of XPath functionality is planned for an intermediate release in the near future. XYZ Query Language (XYZQL) is the native query language used in the XYZFind product. XYZQL queries can be applied to any number of elements and attributes across one or more XML documents. It supports Boolean and wildcard operators that can be applied across multiple documents, elements, and attributes. Query results may be constrained based on the presence or absence of specific elements or attributes. XYZQL gives you focused control over returned values and documents, including the return of complete or partial documents, keyword and numeric-range search, individual elements or attributes from multiple documents, or lists of matching documents. Numeric-range search may likewise be applied, even when values contain mixed unit designations such as "12 inches" or "12 miles". XYZFind supports the return of document fragments as results, consolidated from matching documents. These results are returned in XML format so they can be integrated in XML-aware applications.

XYZFind is accessible via HTTP, a Java API, and a Web Services interface using SOAP. The product includes a Web-based console that demonstrates HTTP access, as well as a command-line utility and source code for a sample client application that uses the Java API. XYZFind can also be accessed as a Web Service (XYZQLService) via the Simple Object Access Protocol (SOAP) over HTTP. XYZQLService is defined in a WSDL file shipped with the product.

Performance and Footprint

While relational databases would require an XML document to be decomposed before it can be stored in schema-specific tables, XYZFind dynamically parses incoming documents, infers their structure, and automatically decomposes them into a single, "universal" data representation. All the associations between a document's elements, attributes, and content are retained to be able to return results to the user. Every aspect of inserted XML documents is indexed including text content, numbers, element and attribute names, comments, and document structures. All documents in the database then share this repository of decomposed information. While it is more processor-intensive to index all aspects of each

document, this approach optimizes query performance, and eliminates the need to constantly tune the database for new document types or extensions to existing schema.

In a relational database, queries may be heavily impacted by join performance, but XYZFind's database query time is predictable and relatively flat regardless of database size. However, practical limits to this indexing do exist. Given commonly employed single-CPU production server equipment, the system is currently able to handle about 2 Gigabytes of data for acceptable insert and query performance, with up to 10 Gigabytes practical for data that does not change very frequently. The XYZFind system does employ compression in the document update process, but also inserts additional information in the database necessary to support some query operators. Overall size of the database is between 75% and 200% of original content size, depending primarily on the character of the data.

Release History

XYZFind was originally released to beta in June 2000. The v1.0 release was launched in February 2001, and version 2 of the product, currently in beta, will be available in August 2001. A number of independent software vendors (ISVs) have expressed interest in using the XYZFind product to meet their document storage and retrieval needs. Mindwrap has leveraged XYZFind for their product catalog and content management needs. Also, Interwoven has integrated XYZFind Server into their TeamSite content management application, and will deploy it as a part of other products in the TeamSite suite. In fact, XYZFind is getting most of their initial interest from ISVs rather than end customers. However, they are getting interest from vendors who have major content requirements. Analog Devices have lots of chip and supplier data, and so they are using the product internally. Many academic libraries are employing XYZFind to provide storage of and query over metadata describing digital content in their collections as well.

XYZFind is licensed both as a standalone XML data store and as an embedded OEM component for redistribution. Their customer base is roughly split between these two licensing models.

Market and Competition

XYZFind faces competition on many fronts. Their initial competition is from the relational database vendors that offer XML Extensions to target this market. Of course, the differences in approach are tremendous, but it is an education effort to keep the customers aware of the differences. More direct challenges come from the XML-native database space. In this arena, Software AG's Tamino product is a major contender as well as similar offerings from Excelon, Neocore, IXIASoft, Infonyte, Ipedo, PyBiz, B-Bop, and others. They plan to compete with Tamino based on license price, the ability to accept arbitrary XML without need for DTDs or Schema, and the ability to return partial document results. IXIAsoft, another competitor in this arena, depends on third-party consulting for parsing returned documents and extracting matching results. There are also a number of minor players in the market that will shakeout soon.

Only those that have a real XML project and understand the benefits of a system like XYZFind will pursue this approach. As Kelvin Ginn of XYZFind says, "We're not going to put RDBMS companies out of business. Even Microsoft acknowledges the value of an arbitrary XML data store."

Key Conclusions & Recommendations

- Companies evaluating NXD projects should put XYZFind on their short-list if their needs are for a schema-independent data store.

- Automatic numerical and range queries without DTD or Schema requirement is a unique feature that may greatly help in situations where highly variable data needs to be searched or aggregated
- XYZFind's 10GByte practical limit for total XML documents in storage can be problematic for very large XML data stores.
- XYZFind doesn't yet fully support XPath or XQuery, which may be an issue, and is a competitive disadvantage to similar products on the market.
- Uses a proprietary storage, indexing, and query mechanism that has not yet been proven in highly scalable, robust environments. However, company is making major moves in that direction.

Profile: XYZFind (December, 2001)
Date Founded: July 1999
Funding: Privately-held, Venture-backed
XMLFund, Wilson, Sonsini, Goodrich, & Rosati, BEA Systems (NASD: BEAS), and Jonathon Lazarus
CEO: Tom Marvin
Employees: 20
Products:
- XYZFind Server

Address:
11200 Kirkland Ave, Suite 310
Kirkland, WA 98033
URL: www.xyzfind.com
Main Phone: 425.827.5838
Contacts:
Kelvin Ginn kginn@xyzfind.com

Related Research

- *XML Data Storage Technologies and Trends* Report (ZTR-ST101)
- *XML Data Storage* Multi-Client Study (ZTR-ST102)
- *Web Services Technologies and Trends* Report (ZT-WEBSRV)
- *B-Bop* ZapNote (ZTZN-0204)
- *Coherity* ZapNote (ZTZN-0144)
- *Excelon* ZapNote (ZTZN-0205)
- *Ipedo* ZapNote (ZTZN-0151)
- *NeoCore* ZapNote (ZTZN-0146)
- *Software AG Tamino* ZapNote (ZTZN-0116)
- *X-Hive* ZapNote (ZTZN-0200)
- *XAware* ZapNote (ZTZN-0154)
- *Xyleme* ZapNote (ZTZN-0326)

About ZapThink, LLC

ZapThink is an IT market intelligence firm that provides trusted advice and critical insight into XML, Web Services, and Service Orientation. We provide our target audience of IT vendors, service providers and end-users a clear roadmap for standards-based, loosely coupled distributed computing – a vision of IT meeting the needs of the agile business.

ZapThink's role is to help companies understand these IT products and services in the context of SOAs and the vision of Service Orientation. ZapThink provides market intelligence to IT vendors who offer XML and Web Services-based products to help them understand their competitive landscape and how to communicate their value proposition to their customers within the context of Service Orientation, and lay out their product roadmaps for the coming wave of Service Orientation. ZapThink also provides implementation intelligence to IT users who are seeking guidance and clarity into how to assemble the available products and services into a coherent roadmap to Service Orientation. Finally, ZapThink provides demand intelligence to IT vendors and service providers who must understand the needs of IT users as they follow the roadmap to Service Orientation.

ZapThink's senior analysts are widely regarded as the "go to analysts" for XML, Web Services, and SOAs by vendors, end-users, and the press. They are in great demand as speakers, and have presented at conferences and industry events around the world. They are among the most quoted industry analysts in the IT industry.

ZapThink was founded in October 2000 and is headquartered in Waltham, Massachusetts. Its customers include Global 1000 firms, public sector organizations around the world, and many emerging businesses. ZapThink Analysts have years of experience in IT as well as research and analysis. Its analysts have previously been with such firms as IDC and ChannelWave, and have sat on the working group committees for standards bodies such as RosettaNet, UDDI, CPExchange, ebXML, EIDX, and CompTIA.

Call, email, or visit the ZapThink Web site to learn more about how ZapThink can help you to better understand how XML and Web Services impact your business or organization.

ZAPTHINK CONTACT:
ZapThink, LLC
11 Willow Street
Suite 200
Waltham, MA 02453
Phone: +1 (781) 207 0203
Fax: +1 (786) 524 3186
info@zapthink.com