

ZAPTHINK ZAPNOTE™

DATA DIRECT XQUERY 2.0 ENABLING XQUERY FOR VERY LARGE MESSAGES

Analyst: Jason Bloomberg

Abstract

As the quantity of XML in the enterprise dramatically increases, enterprise architects, application developers, and data integration specialists must deal simultaneously with the large number of XML-formatted messages on the network as well as such messages' increasingly large sizes. The XQuery approach to handling and processing XML is now recognized as one of the best approaches to dealing with the manipulation of information in XML format. DataDirect takes the power of XQuery to a new level in version 2.0 of the DataDirect XQuery tool, adding support for non-XML and non-relational data, XML streaming, and very large message support.

All Contents Copyright © 2006 ZapThink, LLC. All rights reserved. Reproduction of this publication in any form without prior written permission is forbidden. The information contained herein has been obtained from sources believed to be reliable. ZapThink disclaims all warranties as to the accuracy, completeness or adequacy of such information. ZapThink shall have no liability for errors, omissions or inadequacies in the information contained herein or for interpretations thereof. The reader assumes sole responsibility for the selection of these materials to achieve its intended results. The opinions expressed herein are subject to change without notice. All trademarks, service marks, and trade names are trademarked by their respective owners and ZapThink makes no claims to these names.



Data Integration, XML, and Very Large Messages

Companies are increasingly seeking to leverage XML, Web Services, and Service-Oriented Architecture (SOA) to gain significant business agility in the face of IT heterogeneity. XML provides a key to realizing these benefits through its powerful, interoperable, and metadata-based approach to representing information in environments of heterogeneity. However, these benefits don't come without tradeoffs in performance and efficiency. As the quantity of XML in the enterprise increases, and as Web-Services based SOA initiatives take hold that in turn depend on XML, the quantity and size of metadata-laden XML messages will only continue to increase over time.

Making problems worse is the fact that enterprises implementing XML and standards bodies expanding their reliance on an increasingly more verbose usage of XML-formatted documents continue to tax messaging systems with additional layers of metadata meant to abstract underlying infrastructure. Increasing numbers of messages with a corresponding increase in the size of those messages combine to create the challenge of *Very Large Messages* (VLM), as companies struggle to handle the increasing load of large volume of XML-formatted data efficiently.

Among the most important benefits of SOA is building more flexible approaches to data integration. Companies are increasingly realizing the application integration benefits of SOA, but they are also finding that simply getting systems to talk to each other is a small part of the challenge of making distributed computing in a heterogeneous infrastructure work. Indeed, the bigger challenge comes from trying to piece together data from these multiple systems in a way that makes them easy to understand and process. Traditionally, building such data integration logic is costly, time-intensive, complex, and error prone, because data integration specialists typically take a point-to-point approach, using a combination of brittle integration middleware, proprietary adapter interfaces, and human-intensive mapping. The net result is that businesses build one-time-use data integration projects that are difficult to manage and impossible to reuse or share across the organization, wasting a considerable portion of their IT budget in the process.

While XML aimed to solve the problem of interoperable data via a standards-based language for data integration, it doesn't provide the actual mechanism for integrating with endpoints nor converting different representations of XML data into different formats. The *XQuery* language was designed to support this kind of data integration, as well as native XML queries, and it is now gaining traction for both purposes. However, while XQuery offers significant power for data integration, developers still require an easy way to develop and manage XQuery-based data integration, especially as the size of XML messages increases. As a result, developers need tools to build the XQueries as well as to transform data from original data sources into an XML format, even when those data sources are streaming VLMs. The ability to abstract relational and other data formats into XML and the ability to aggregate both the inbound composite data from such sources is the key to success with XQuery.

Thank you for reading ZapThink research! ZapThink is an IT advisory and analysis firm that provides trusted advice and critical insight into the architectural and organizational changes brought about by the movement to XML, Web Services, and Service Orientation. We provide our three target audiences 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.

Earn rewards for reading ZapThink research! Visit www.zapthink.com/credit and enter the code **DDXQ2**. We'll reward you with ZapCredits that you can use to obtain free research, ZapGear, and more! For more information about ZapThink products and services, please call us at +1-781-207-0203, or drop us an email at info@zapthink.com.

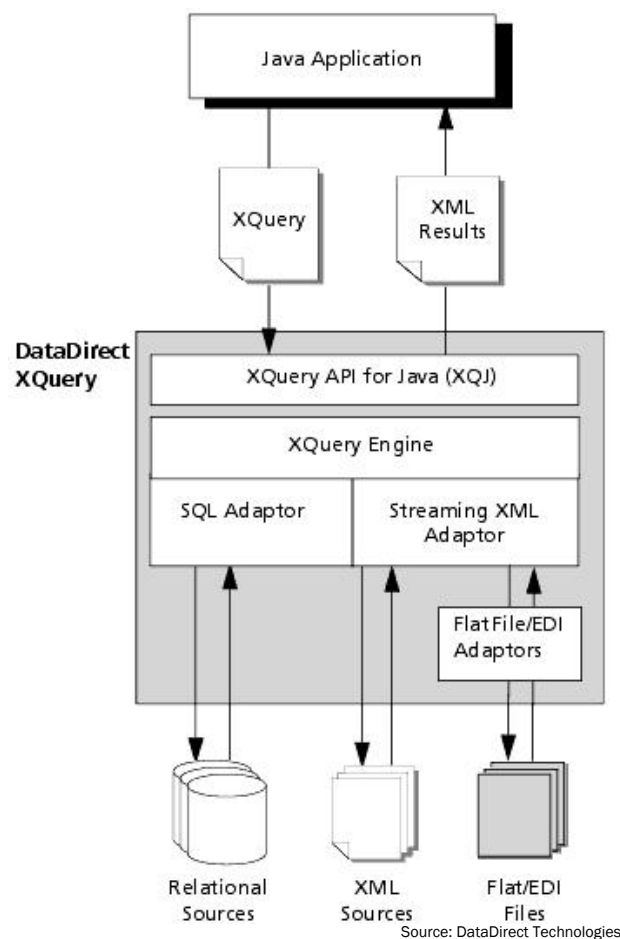


DataDirect XQuery 2.0: Optimized for Very Large Messages

To address these issues of large XML data volumes stressing existing infrastructure, DataDirect Technologies, a division of Progress Software, has recently released version 2.0 of *DataDirect XQuery*, their tool for database-independent processing and aggregating of XML, relational and legacy data formats such as EDI and CSV. DataDirect XQuery 2.0 is fundamentally an XQuery engine that includes performance enhancements for very large XML messages, as well as added functionality for optimal query performance and support for a larger number of legacy data formats. DataDirect XQuery 2.0 is particularly suited for environments of heterogeneous XML, relational and legacy data formats. In particular, the new streaming XML adapter added to version 2.0 reduces the amount of necessary memory for processing VLMS.

DataDirect XQuery provides a tool for data aggregation and integration across multiple applications and data sources. Like the earlier DataDirect XQuery 1.0, DataDirect XQuery 2.0 provides fast, reliable and scalable XQuery support for relational databases and runs on any Java platform. The tool also supports the *XQuery for Java API (XQJ)* specification, providing developers with the ability to deal with XML either as part of a set of heterogeneous data sources or as the targeted format for a data integration application. The DataDirect XQuery architecture is shown in the figure below:

DataDirect XQuery Architecture



DataDirect XQuery 2.0's streaming XML adapter enables DataDirect XQuery to handle VLMs of up to gigabytes in size. It also provides new user-configurable performance enhancements, providing a greater ability to choose optimizations for specific environments and to configure these performance enhancements directly. DataDirect XQuery 2.0 also provides extended database support and better integration with Stylus Studio XML Deployment Adapters, enabling Web developers to use DataDirect XQuery to query and integrate data sources that are neither XML or relational.

The ZapThink Take

As companies seek to increasingly make use of secure, reliable, process-driven, and loosely coupled SOA implementations, they will quickly become aware of the performance penalty that inefficient usage of the XML format imposes. In addition, the performance challenges imposed by very large XML messages might potentially threaten the viability of SOA implementations, and could precipitate an impending SOA performance crisis. It is imperative, therefore, for developers, integration specialists, technical architects, and anyone else who must deal directly with the XML flowing through the enterprise to have tools robust enough to work with such messages.

DataDirect XQuery 2.0 adds these VLM capabilities to an already robust set of features for the developer and data integration specialist. In fact, the combination of the flexibility and the power that XQuery 2.0 offers enables this application to be an *ad hoc* ETL (extract, transform, and load) tool. Data integration personnel are now able to accept a number of incoming XML messages of varying sizes, and use DataDirect XQuery to create a local summary file on the fly that they can then use as the basis for a variety of *ad hoc* queries. In other words, DataDirect XQuery 2.0 also addresses the XML reporting problem of how to provide business visibility into large quantities of XML data in a flexible manner worthy of SOA.

Product Profile

DataDirect XQuery 2.0	Availability: Now
-----------------------	-------------------

Overview:

DataDirect XQuery provides embeddable software capabilities using the XQuery and XQJ standards to enable Java applications to access both relational and XML data sources via a single query. DataDirect XQuery is standards-based and both database-independent and operating-system independent.

Value Propositions:

- *Performance and scalability enhancements* – includes performance optimizations that are now user configurable, allowing more control over specific optimizations.
- *Handling of very large XML messages* – XML streaming reduces the necessary amount of memory when processing large XML messages.
- *Integration with Stylus Studio XML Deployment Adapters* – These adapters provide the ability to use DataDirect XQuery to query and integrate data sources that are neither XML nor relational, including CSV and EDI files.

- *Updated XQuery Support* – DataDirect XQuery now supports the November, 2005 Candidate Recommendation XQuery specification and adds support for XQuery modules.

Key Differentiators:

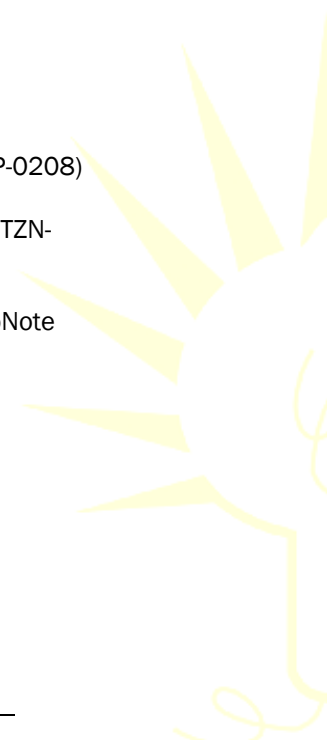
- *XQuery capabilities for data sources that are neither XML nor relational*
- *Handles very large XML documents without memory or scalability issues*
- *High performance in a variety of specific environments, supporting enterprise heterogeneity*

DataDirect Profile

Profile: DataDirect	March 2006
Funding:	Division of Progress Software
President:	Richard D. Reidy
Address:	14 Oak Park Drive Bedford, MA USA 01730
URL:	http://www.datadirect.com or http://www.stylusstudio.com
Phone:	800-876-3101
Contact:	Nancy.vodicka@datadirect.com

Related Research

- *Service Orientation Market Trends Report (ZTR-WS110)*
- *SOA Tools and Best Practices Report (ZTR-WS107)*
- *Business-to-Business Data Integration in a SOA World ZapForum Podcast (ZTP-0208)*
- *DataDirect Technologies: Simplifying Data Integration with XQuery ZapNote (ZTN-1189)*
- *DataDirect Technologies: Deploying XML Data Services with Stylus Studio ZapNote (ZTN-1181)*



About ZapThink, LLC

ZapThink is an IT advisory and analysis firm that provides trusted advice and critical insight into the architectural and organizational changes brought about by the movement to XML, Web Services, and Service Orientation. We provide our three target audiences 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 helps its customers in three ways: by helping companies understand IT products and services in the context of Service-Oriented Architecture (SOA) and the vision of Service Orientation, by providing guidance into emerging best practices for Web Services and SOA adoption, and by bringing together all our audiences into a network that provides business value and expertise to each member of the network.

ZapThink provides market intelligence to IT vendors and professional services firms that offer XML and Web Services-based products and services in order to help them understand their competitive landscape, plan their product roadmaps, and communicate their value proposition to their customers within the context of Service Orientation.

ZapThink provides guidance and expertise to professional services firms to help them grow and innovate their services as well as promote their capabilities to end-users and vendors looking to grow their businesses.

ZapThink also provides implementation intelligence to IT users who are seeking guidance and clarity into the best practices for planning and implementing SOA, including how to assemble the available products and services into a coherent plan.

ZapThink's senior analysts are widely regarded as the "go to analysts" for XML, Web Services, and SOA by vendors, end-users, and the press. Respected for their candid, insightful opinions, 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 Baltimore, Maryland. Its customers include Global 1000 firms and government organizations, as well as many emerging businesses. Its analysts have worked at such firms as IDC, marchFIRST, and ChannelWave, and have sat on the working group committees for standards bodies such as RosettaNet, UDDI, and ebXML.

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

ZAPTHINK CONTACT:

ZapThink, LLC
108 Woodlawn Rd
Baltimore, MD 21210
Phone: +1 (781) 207 0203
Fax: +1 (786) 524 3186
info@zapthink.com
www.zapthink.com