

XQuery Saving Lives: Data Integration Using DataDirect XQuery®

Customer Profile

When the Volpe Center and the Transportation Information Project Support (TRIPS) contractors, lead by CSC, were requested by the National Highway Traffic Safety Administration (NHTSA) - part of the U.S. Department of Transportation - the goal was simple: provide an efficient way to deliver Car Crash Information in an open, public format following the US-federal standards and compliance rules.

The information, collected since 1997 under the Electronic Car Crash Data System, averages out data involving approximately 5,000 accidents every year. Data is collected in 24 sites around the country and consolidated in a large Oracle database (500 GigaBytes). NHTSA wanted to make the information available to several groups of people, including Department of Transportation researchers, insurance company investigators, vehicle manufacturers and the general public.

Previously, all data distribution had taken place using flat files and various proprietary formats; new federal regulations and technology evolution empowered the Volpe Center to move to the current open standard: XML.

Doing Things the Old-Fashioned Way

Initially, the engineers at the Volpe Center explored using a conventional approach — SQL queries driven by Java code. They determined that moving data from a database into XML using SQL and Java would require about 50 Java classes and 150 SQL statements for one study type. Obviously, a project of this magnitude would be complex to design and debug, and would require considerable maintenance of code.

Doing Things a Better Way

After careful consideration in November 2005 it was determined that a better option was to store the data in an XML repository. The Volpe Center decided to use DataDirect XQuery® to handle their XML and relational tasks. XQuery allowed them to move all the data directly into XML files validated against the specific Car Crash XML-schema design; all the XQueries for sub processing data features such as search and dataset construction, are performed against the XML files. The new architecture enabled by DataDirect XQuery requires just a single Java class and 7 XQueries. Each XQuery encapsulates the transformation rules clean of any other type of code, considerably reducing maintenance efforts, and simplifying code.

“Building the XML repository from the Oracle relational database and later executing the XQueries on the XML to traverse the data resulted in amazing performance,” reports Juan Alfonso, senior software engineer at Core Computer

“DataDirect XQuery is the one that worked; it is very stable and demonstrates excellent performance.”

Juan Carlos Alfonso,
Senior Software Engineer
Core Computer Group
www.corecomputergroup.com

Group, a TRIPS project contractor. "Moreover, a large extraction from 500MB of XML data building datasets intended for SAS analysis took significantly less time compared to using different methods. Simple XQueries with XPath expressions pointing to the XML repository locations performed easily using the DataDirect XQuery implementation."

Get the Information to the People



DataDirect and the TRIPS project team at the Volpe Center were instrumental in developing the National Highway Traffic Safety Administration's new system, which allows their customers to access important information quickly, efficiently, and reliably. DataDirect XQuery, from Core Computer Group's perspective, was instrumental in enabling the Volpe Center to efficiently deliver a powerful and reliable solution to their customer.

"DataDirect XQuery is the one that worked, and it worked with excellent performance, which we demonstrated with our project," Alfonso says in summing it up. "With DataDirect XQuery we were able to reduce code complexity, simplify code maintenance and improve performance and scalability."

"Additionally, the next release of our series of car crash studies will deploy a set of new features like dynamic search capabilities and custom dataset creation. The processing will be performed using a backend generic XQuery factory, based on DataDirect XQuery. The advantage to doing this process directly from the XML repository, is that it will allow us to isolate the final-public-deliverable data from the Oracle production database."

Goal

Real-time lookup and distribution of high-volume data from XML and a large (500 GB) database.

Why they selected DataDirect

By moving away from more conventional approaches (such as Java code and SQL query statements), and exploring different XQuery implementations, DataDirect XQuery was the most efficient solution for both the relational to xml and later XML to XML data integration.

Business benefits

National Highway Traffic Safety Administration was able to provide a way for their researchers and customers to access the Car Crash Information Collection database quickly and efficiently while saving development time and effort.

Technical benefits

By choosing DataDirect XQuery, the Volpe Center and the TRIPS contractors were able to move the data to XML files residing on the file system and to drastically simplify the query code. The architecture the Volpe Center developed using XQuery is powerful, efficient, stable, and easy to maintain.

"With DataDirect XQuery we were able to reduce code complexity, simplify code maintenance and improve performance and scalability."

Juan Carlos Alfonso

About DataDirect Technologies

DataDirect Technologies is the software industry's only comprehensive provider of software for connecting the world's most critical business applications to data and services, running on any platform, using proven and emerging standards. Developers worldwide depend on DataDirect® products to connect their applications to an unparalleled range of data sources using standards-based interfaces such as ODBC, JDBC and ADO.NET, XQuery and SOAP. More than 300 leading independent software vendors and thousands of enterprises rely on DataDirect Technologies to simplify and streamline data connectivity for distributed systems and to reduce the complexity of mainframe integration. DataDirect Technologies is an operating company of Progress Software Corporation (Nasdaq:PRGS). For more information, visit www.datadirect.com.

DataDirect and DataDirect XQuery are trademarks or registered trademarks of DataDirect Technologies Corp. in the U.S. and other countries. Java and all Java based marks are trademarks or registered trademarks of Sun Microsystems, Inc. in the U.S. and other countries. Any other trademarks contained herein are the property of their respective owners.