

EMC Documentum xDB

High-performance native XML database optimized for storing and querying large volumes of XML content

The Big Picture

- Ideal for content-oriented applications like dynamic publishing and content warehousing
- More flexible than RDBMS for unstructured content and for aggregating diverse data sets
- Scalable, high-performance architecture with minimal overhead
- High reliability
- Simplified administration
- XQuery: Powerful search, retrieval, linking, and updates
- Comprehensive, highly productive application development
- Designed for integration and interoperability
- Built-in content management features

Introduction

Many organizations view XML as a means of improving access to information. Since XML can be used to describe any type of data, it offers a common format for representing both structured and unstructured data. This common format allows diverse sets of content to be integrated into the 360-degree view of corporate information that leads to better decisions, greater innovation, lower risk, and more efficient processes. In addition, a number of mature standards for manipulating XML content make XML ideal for unlocking information from the strictures of proprietary, application-defined formats.

Adoption trends also are pushing organizations to create and manage XML content. Many government agencies, for example, are prescribing XML as the format for regulated documents like financial filings and food and drug applications. And XML is now the default format in Microsoft Office and other office suites. These adoption trends, driven by the advantages of XML as a semantically-rich, application-neutral content format, have led to an explosion in the amount of XML content in the enterprise.

If your organization uses content-oriented applications, then it's not a question of if, but when you will use XML. Now is the time to begin thinking about how your organization is going to manage that XML content and make it available to users. To start, you need a scalable XML database platform that is optimized for storing, managing, and querying XML content.

Overview

Designed for content-oriented applications such as publishing, archiving, information mashups, regulatory filings, collaboration, and knowledge management, EMC® Documentum® xDB provides a scalable architecture and complete support for the XQuery language, allowing organizations to warehouse content in an application-neutral format, not dependent on a proprietary application for information retrieval.

Unlike relational databases, xDB allows content schemas to be easily modified to adapt to changing information requirements; supports queries against complex data structures that are not easily modeled in rows and columns; and even supports automatic versioning of content and schemas. And, with a powerful and extensible development and runtime toolset—as well as integration with the leading enterprise content management system—xDB is a powerful platform for even the most complex and demanding content-centric applications.

Common applications for xDB include the following.

Dynamic publishing

Traditional content delivery systems have been static, with the same content pushed to all users regardless of their needs. Frustration with finding relevant content can challenge customer loyalty and limits up-selling or cross-selling opportunities. And, in these traditional systems, content can easily grow stale, which can lead to compliance and liability issues. Attempts to build dynamic delivery sites on relational databases often lead to scalability and maintenance problems.

xDB has the scalability, performance, and functionality to power high-volume dynamic sites that deliver highly relevant, targeted content. Modeling your content as XML allows you to capture its richness and depth, while making it easy to adapt your content models to changing business requirements. The complete XQuery implementation of xDB can aggregate, filter, and assemble the content as needed to:

- Increase user satisfaction by delivering relevant, targeted content
- Avoid performance problems caused by the content transformation overhead that bogs down relational systems
- Increase developer productivity and drive down maintenance costs by leveraging standards-based tools and technologies
- Scale your sites to handle any volume of requests with a high-performance architecture
- Simplify site administration with built-in content management features, like versioning and metadata support
- Eliminate stale content by synchronizing automatically with Documentum

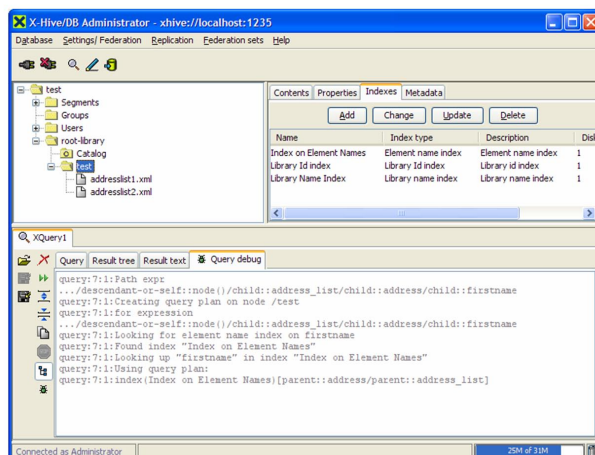
You can also take advantage of EMC Documentum Dynamic Delivery Services, a powerful and productive web-based platform that makes it easy and productive to build, maintain, and deploy content delivery applications on xDB.

Content warehousing

While there is a growing need for organizations to retain content for regulatory compliance, eDiscovery, and more flexible search and reporting, operational systems with high transaction volumes often need to offload historical records to reduce overhead. When data is offloaded, it is not available for query and analysis. When information needs to be available beyond the useful life of the application that created it, legacy applications are kept around so that the data can live on, but with high maintenance costs. Even when data is preserved, it resides in application-specific silos that restrict an organization's view of the information, limiting knowledge dissemination and innovation, hampering business process improvement, and increasing exposure to risk.

XML is a perfect format for aggregating information into a content warehouse to support archiving and decision support. It is application-neutral—you don't need a proprietary application to search and process it—and it's self-describing, which means that it retains its information value, or identity, even when the application that created it is not available. You can also easily modify the content definitions (schemas) to adapt to changing needs. A high-performance repository is required for efficient storage and access, as is a flexible query language for mining the content for its information. That's where xDB comes in. The product's scalable architecture and support for the XQuery language enables your organization to:

- Reduce risk and improve decision making by aggregating content from multiple applications in a single repository for query and analysis
- Eliminate the maintenance costs on outdated applications by archiving their data in a repository that can preserve application context



The xDB Admin Tool allows you to test and debug your queries.

Features

Scalable, high-performance architecture

Meet enterprise-level scalability and performance requirements with an architecture optimized for storing and querying large volumes of XML content, but requiring minimal storage and memory overhead.

- **Unlimited storage capacity:** Manage huge volumes of content with a flexible storage model.
- **Minimal storage overhead:** Minimize storage overhead with efficient space usage—space required for content is often less than or equal to original content size, including indexes.
- **Low memory requirements:** Reduce hardware costs with efficient memory usage and low memory requirements (starting at only 256 MB).
- **Extensive tuning options:** Fine-tune query performance with multiple indexing options. Index types include Value, Id attribute, Element name, Library name, Library id, Full text, and Path / Value. Indexes can be scoped to a specific collection or document, minimizing overhead and improving selectivity. Data partitioning enables highly selective pruning of query results and also optimizes space management.
- **High concurrency:** Avoid user lock-out scenarios where readers block writers, or vice versa. Read-only transactions allow readers to view a consistent snapshot of content without requiring read-locks that block updates.
- **Reduced processing overhead for compound operations:** Minimize processing overhead and network latency effects with multi-statement transactions that batch updates and queries.
- **Unlimited horizontal scaling:** Scale your applications across multiple servers. Query processing can be separated from back-end page servers and distributed with load-balancing across multiple nodes in a server farm. Query load can also be distributed across multiple page server replicas, with each query processor able to access one or more replicas of the page server.

High reliability with simplified administration

Ensure that your database is well-managed and protected from data loss and downtime.

- **High availability:** Ensure the reliability and availability of your content with support for full, incremental, and “hot” backups (with consistent state), plus integration with third-party backup utilities for efficient and convenient management. Replication support automatically synchronizes content changes across multiple page servers, with failover to a surviving replica if the primary server fails.
- **Robust transactions:** Protect your data with transaction control (commit/rollback) across multiple statements, with checkpoints for saving changes within a transaction and automatic deadlock detection.
- **Intuitive navigation:** Organize XML content into folder-like collections (called libraries) for easy browsing and management.
- **Powerful, but easy-to-use, administration:** Manage users, browse data, and test XQueries with a graphical administration console.
- **Embeddability:** Minimize the need for ongoing management when xDB is embedded in an application through xDB’s effortless administration, 100 percent Java architecture, silent installer, and very low resource requirements.
- **Easy copying between databases:** Export libraries or individual documents with all related metadata to a portable format that simplifies copying between databases.

Related Products

xDB anchors a family of complementary and related products from EMC. Products available separately include:

- **EMC Documentum XML Store:** Integrates xDB into EMC Documentum Content Server, the market leading enterprise content management repository. XML store leverages the power of XML for adding meaning and relationships to your enterprise content, while applying an extensive set of enterprise content and process services to your XML content—improving security, lifecycle management, and integration with other information assets.
- **EMC Documentum Dynamic Delivery Services:** A platform for building and deploying XML-based content delivery applications that enable user-driven, realtime assembly of content for viewing through a variety of channels including print, web, and mobile.
- **EMC Documentum XML DIFF engine:** A utility for calculating the differences between any two XML documents, or between two documents derived from a single source document (a “3-way diff”). The DIFF engine can identify deletions, insertions, moves, element and attribute name changes, and changes in comments and processing instructions. The comparisons are visually rendered in HTML or can be further processed (for example, to merge or accept/reject changes) using the Java API. The XML DIFF engine can be used with xDB, XML Store, or can operate standalone.

Powerful search, retrieval, linking, and updates

Enable powerful search and retrieval with the industry’s most complete implementation of the XQuery standard, and leverage common developer skills with extensive support for XML processing standards.

- **Fully standards based:** Future-proof your content and applications with support for key XML processing standards, including XPath, XLink, XPointer, DOM, and the most complete conformance with the W3C XQuery specification.
- **DOM processing for collections:** Process multiple documents simultaneously with extended DOM Level 3 functions; for example, modifying all documents within a library.
- **Full-text searches:** Search full-text in your content with full-text indexes and XQuery extensions.
- **Efficient XQuery update:** Perform updates on your XML content utilizing the W3C XQuery Update Facility candidate recommendation. Updates are applied directly to affected content fragments, without having to read and write entire documents.

Comprehensive application development

Rapidly build and deploy powerful content applications with an extensive set of standards-based development and runtime features. These features are ideal for content delivery applications that enable user-driven, realtime assembly of content.

- **Complete API:** Utilize a comprehensive Java API with methods for storing, querying, retrieving, transforming, and publishing XML data.
- **Built-In transformations:** Transform XML into other formats, such as HTML, PDF, and mobile formats, with built in XSLT and XSL-FO processors.
- **Rapid development:** Build, maintain, and deploy powerful, interactive XML applications with a complementary product, EMC Documentum Dynamic Delivery Services.

Designed for integration and interoperability

Comply with the latest industry and platform standards, and ensure that your XML applications are easily integrated with the rest of your IT infrastructure.

- **Supported standards:**

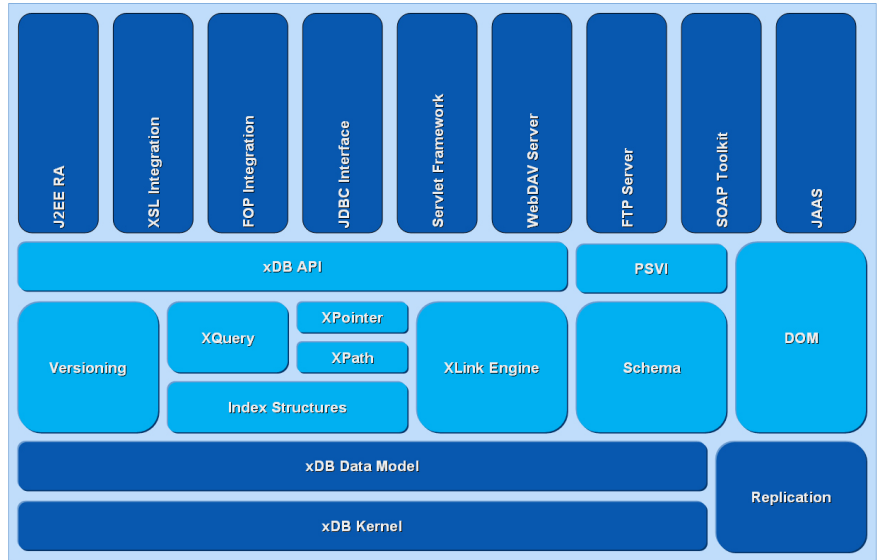
- | | |
|-------------------|-----------|
| o XML 1.0 and 1.1 | o DOM |
| o XQuery | o WebDAV |
| o XML Schema | o J2EE |
| o XPath | o JSP |
| o XPointer | o JAAS |
| o XLink | o SOAP |
| o XSL | o Unicode |

- **Supported platforms:**

Supports any Java platform (JDK 5 or higher), including:

- o Linux
- o Sun Solaris
- o HP-UX
- o Any Windows platform
- o Apple Macintosh
- o AIX

- **Content validation:** Ensure that XML content is valid against either DTDs or XML schemas. Maintain a catalog of DTDs and schemas, and also store well-formed (non-validated) XML.
- **WebDAV interface:** Access the database as a file repository from WebDAV clients.



The xDB standards-based architecture makes integration, interoperability, and application development a snap.

- **Database import:** Import content directly from any JDBC-compliant database, or via a SQL loader utility.
- **File import:** Load files directly from file systems with a high-speed loading utility.
- **Documentum integration:** Push content automatically from EMC Documentum Content Server through Documentum Site Caching Services (available separately).

Extended content management

Take advantage of powerful content management capabilities that go beyond conventional database storage and processing.

- **Manage non-XML content:** Store and manage any file-based content.
- **Versioning:** Maintain document history and allow documents to be restored to a previous version using built-in linear versioning with branches.
- **Metadata:** Track properties and characteristics for libraries and content with an extensible metadata model.
- **XML differencing:** Compare document versions with an XML differencing engine, available separately.



EMC Corporation
Hopkinton
Massachusetts
01748-9103
1-508-435-1000
In North America 1-866-464-7381
www.EMC.com

Take the next step

To learn more about EMC Documentum xDB, visit www.EMC.com or call 800.607.9546 (outside the U.S.: +1.925.600.5802).