

Developing and Deploying XBRL-Based Applications in an Enterprise-Class Environment

Understanding the Challenges, Opportunities, and Benefits of XBRL Storage

UBMATRIX™

333 Twin Dolphin Drive
Redwood City, CA 94065
+1.650.264.4510
info@ubmatrix.com

Table of Contents

Executive Overview	3
Introduction	3
XBRL Storage, Query, and Retrieval Challenges	3
XBRL-DB Use Case Examples.....	5
Query and Retrieval of the Entire Instance Document	5
Query and Retrieval of Specific Fact Values for a Given Concept and Context.....	6
Retrieval of a Portion of an Instance Document	6
Supporting all Three Use Cases with UBmatrix XBRL-DB	6
Conclusion	7
Appendix A: Glossary of XBRL Terms.....	8

Executive Overview

This document is written for technical project managers and solutions architects considering the deployment of an XBRL-based application in an enterprise-class environment with a requirement to store, query, and retrieve XBRL documents in a scalable fashion. This document assumes the reader has a working knowledge of XBRL and is familiar with basic XBRL concepts, such as instance document, taxonomy, and linkbase. (For reference, a glossary of XBRL terms is included at the end of this document.) This document also assumes the reader has a working knowledge of SQL storage mechanisms.

After a review of this document, the reader should have a clear understanding of the issues surrounding the storage, query, and retrieval of XBRL documents. The paper will present the key benefits of using UBmatrix XBRL-DB, and review its main capabilities. The document will also describe common use cases for XBRL-DB.

Introduction

Today's enterprises need the ability to report, validate, and analyze business information efficiently, accurately, and in a scalable manner. Through the eXtensible Business Reporting Language (XBRL), financial institutions and regulators have established a standard language to describe reported data and business rules surrounding that data.

XBRL is an open standard based on XML that is currently supported by over 550 global organizations. It promises enterprises and government agencies a cost-effective method to deliver a 100% compliant communication process for both external and internal reporting. However, to deliver on this promise, XBRL documents must be able to be generated, stored, and queried seamlessly and effectively alongside other information within the enterprise.

The UBmatrix XBRL-DB represents the most powerful and effective method for managing, storing, and analyzing XBRL data. It fulfills the database storage requirements that provide the speed, scalability, and security required by organizations looking to implement an XBRL-based reporting system. The following sections of this paper will review the challenges related to storing, querying, and retrieving XBRL data, and introduce the capabilities of UBmatrix XBRL-DB.

XBRL Storage, Query, and Retrieval Challenges

Instance documents and taxonomies are at the core of XBRL. Any organization that is collecting XBRL documents (such as regulators, information hubs, or enterprises) will want to at a later date, query and retrieve information from these instance documents and/or taxonomies.

As we will see, to query and retrieve information from an XBRL instance document can require access to its associated taxonomy. This taxonomy contains critical metadata that express relationships among element definitions specific to XBRL, such as:

- **Presentation information** – Presentation information defines the organizational relationships (order) of elements using parent-child hierarchies.
- **Label information** – These are human-readable names for an element; each element has at least one, generally known as a “standard” label. Optional sets of labels can also exist; each is given a separate defining “role” or name.

- **Calculation information** – Additive relationships between monetary items expressed as parent-child hierarchies, which validate the instance document.

Most query and retrieval of XBRL information can take the form of one of three use cases:

- 1) **Document-based queries** – Query and retrieval of the **entire instance document** as filed (e.g., “Return the August 2007 10K filing for IBM”). To be complete, this also requires the retrieval of the associated taxonomy files.
- 2) **Fact-based queries** – Query and retrieval of **specific fact values** for a given concept and context (e.g., “Return the 2007 Total Revenue for SAP”).
- 3) **Taxonomy-based queries** – Query and retrieval of **fact values for a portion of an instance document**, referred to as “an extended link” (e.g., “Return the 2007 Balance Sheet for ORCL”). This retrieves facts organized by metadata contained in the taxonomy. Additionally one may query and retrieve **metadata stored in the taxonomy**.

Also note that all of these queries may cross multiple instance documents. For example: “give me the Income Statements for all companies that had total revenues greater than \$1 billion dollars in 2005.”

For cases 1 and 2, there are a variety of storage strategies currently available. For case 1, it would be storing the instance document and taxonomy as blobs (Binary Large Objects). Blobs are a collection of binary data stored as a single entity in a database management system – typically images, audio, or other multimedia objects. But this method has its limitations, since blob storage does not provide the ability to search for any particular content contained within the blob.

For case 2, a traditional SQL strategy would be appropriate where the concepts, context, and fact values are stored in a relational table (independent of the taxonomy). This would allow the querying and retrieval of facts such as “Total Revenue” across various time periods or entities.

For case 3, up until now, there has been no efficient storage strategy. With traditional (non-XBRL) storage mechanisms, the only approach is the same as case 1 – to store and retrieve the entire instance document and taxonomy as blobs. The problem with this approach is that it requires the overhead of reading into memory and resolving the entire taxonomy (i.e., the Discoverable Taxonomy Set¹), including extensions – and then reading into memory the entire instance document to retrieve just a portion of the instance document (e.g., “Income Statement for US-GAAP”). This introduces unacceptable overhead in terms of processing time and memory requirements, prohibiting the ability to scale.

Finally, it may be that different aspects of an XBRL-based application will require different types of query and retrieval. So the architect is faced with the dilemma of choosing one strategy over another, resulting in sub-optimal performance in the other use case, or keeping multiple copies of instance documents – creating a whole set of challenges relating to multiple sources of truth.

The key challenge for XBRL storage, query, and retrieval is to deliver a solution that provides enterprise-class efficiency and performance across **all three expected use cases** while preserving a single source of truth. SQL supports only document or fact-based queries that are independent of the taxonomy. In contrast, XBRL-DB supports **all** queries against the instance document, facts and taxonomy.

¹ The collection of all the schemas that are required to validate the syntax of an XBRL instance becomes part of that XBRL instance's Discoverable Taxonomy Set (DTS)

Query Type	Query Examples (using US GAAP)	XBRL-DB	SQL
Document-based	<ul style="list-style-type: none"> Give me all instance documents filed by Company X 	✓	✓
Fact-based	<ul style="list-style-type: none"> Give me Total Assets for Company X in for period ending 12-31-2007 Give me all companies that reported a value for Pension Liabilities for the period ending 12-31-2007 Return all periods where Company X's Gross Profit" exceeded \$Y 	✓	✓
Taxonomy-based (retrieve fact values)	<ul style="list-style-type: none"> Give me the Balance Sheet for Company X for the period ending 12-31-2007 with US-GAAP concepts Give me the Income Statement for Company X for the period ending 12-31-2007 with company extensions (such as custom labels or added concepts) 	✓	✗
Taxonomy-based (retrieve metadata)	<ul style="list-style-type: none"> Give me all companies that filed using "C&I" extension to US GAAP Give me all companies that have customized Property, Plant & Equipment Give me all customization of Property, Plant & Equipment by company X Give me all taxonomies & schema referenced by Company X 	✓	✗

XBRL-DB Use Case Examples

Query and Retrieval of the Entire Instance Document

```

<?xml version="1.0" encoding="US-ASCII" standalone="yes" ?>
<xbrli:xbrl xmlns:iso4217="http://www.xbrl.org/2003/iso4217"
xmlns:link="http://www.xbrl.org/2003/linkbase" xmlns:mda="http://xbrl.us/mda/2007-12-31"
xmlns:msft07="http://xbrl.microsoft.com/2007" xmlns:us-gaap="http://xbrl.us/us-gaap/2007-12-31"
xmlns:xbrldi="http://xbrl.org/2006/xbrldi" xmlns:xbrli="http://www.xbrl.org/2003/instance"
xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <link:schemaRef xlink:href="msft-20071231.xsd" xlink:type="simple" />
  <xbrli:context id="FY06e">
    <xbrli:entity>
      <xbrli:identifier scheme="http://www.sec.gov/CIK">0000789019</xbrli:identifier>
    </xbrli:entity>
    <xbrli:period>
      <xbrli:instant>2006-06-30</xbrli:instant>
    </xbrli:period>
  </xbrli:context>
  ....

```

Figure 1. A (partial) sample of an XBRL instance document.

Query and Retrieval of Specific Fact Values for a Given Concept and Context

ENTITY_SCHEME	CONTEXT_ID	PERIOD	ATTR_NAME	ATTR_VALUE	DECIMALS
http://www.ibm.com	IBM-2007-2Q-Instant	06/30/2007	TotalCurrentAssets	42293000000	-6
http://www.ibm.com	IBM-2006-Instant	12/31/2006	TotalCurrentAssets	44660000000	-6
http://www.ibm.com	IBM-2007-2Q-Instant	06/30/2007	Assets	102548000000	-6
http://www.ibm.com	IBM-2006-Instant	12/31/2006	Assets	103234000000	-6
http://www.ibm.com	IBM-2007-3Q-Instant	09/30/2007	TotalCurrentAssets	46588000000	-6
http://www.ibm.com	IBM-2006-Instant	12/31/2006	TotalCurrentAssets	44660000000	-6
http://www.ibm.com	IBM-2007-3Q-Instant	09/30/2007	Assets	108609000000	-6
http://www.ibm.com	IBM-2006-Instant	12/31/2006	Assets	103234000000	-6
http://www.ibm.com	IBM-2006-Instant	12/31/2006	Assets	103234000000	-6
http://www.ibm.com	IBM-2007-1Q-Instant	03/31/2007	TotalCurrentAssets	42446000000	-6
http://www.ibm.com	IBM-2006-Instant	12/31/2006	TotalCurrentAssets	44660000000	-6
http://www.ibm.com	IBM-2007-1Q-Instant	03/31/2007	Assets	101619000000	-6
http://www.ibm.com	IBM-2006-Instant	12/31/2006	Assets	103234000000	-6
http://www.ibm.com	IBM-2006-Instant	12/31/2006	TotalCurrentAssets	44660000000	-6
http://www.sec.gov	BalanceAsOf_30Jun2006_Unaudited	06/30/2006	Assets	27490100000	
http://www.sec.gov	BalanceAsOf_31Dec2006_Unaudited	12/31/2006	Assets	31612200000	

Figure 2. A sample result of a query of specific fact values for a given concept and context.

Retrieval of a Portion of an Instance Document

Statement Cash Flows (USD \$) (in Millions, except per share data)	3 Months Ended	
	Sep. 30, 2007	Sep. 30, 2006
Net Cash Flows Provided By/(Used In) Operating Activities	5,878	4,061
Proceeds from Issuance of Common Stock	646	385
Repurchase of Common Stock	2,930	7,683
Payment of Dividends - Common Stock	938	897
Stock-based Compensation - Excess Tax Benefit	69	40
Cash Provided by/(Used in) Financing Activities - Other	0	(20)
Net Cash Flows Provided By/(Used In) Financing Activities	(3,153)	(8,175)
Capital Additions	510	411
Acquisition of Businesses, Net of Cash Acquired	5,396	336
Purchase of Investments - Total	5,997	12,855
Proceeds from Sale of Available-for-Sale Securities	9,120	18,701
Net Cash Flows Provided By/(Used In) Investing Activities	(2,257)	6,461
Net Proceeds from Securities Lending	196	528
Increase/(Decrease) in Cash and Cash Equivalents	468	2,347
Effect of Exchange Rate on Cash and Cash Equivalents	58	15
Cash and Cash Equivalents - Beginning Balance	6,111	6,714
Cash and Cash Equivalents - Ending Balance	6,637	9,076

Figure 3. Sample XBRL Instance Document (after rendering).

Supporting all Three Use Cases with UBmatrix XBRL-DB

UBmatrix provides a single high performance storage mechanism which supports all three storage, query, and retrieval use cases. UBmatrix XBRL-DB™ is delivered on top of an enterprise-class relational database (either Oracle 11g), allowing it to benefit from capabilities of these scalable products.

Release 1.0 of XBRL-DB is available as an option to UBmatrix's Processing Engine. The key benefits of UBmatrix XBRL-DB include:

1) Supports all three storage, query, and retrieval use cases:

- Provides high performance query of instance documents, portions of instance documents (defined by extended links), and fact values from a single storage mechanism

2) Leverages information collected in XBRL documents:

- Enables both ad hoc and batch queries

3) Shortens application development times through high level APIs to navigate taxonomy and instance documents:

- No requirement to write taxonomy-specific code
- No requirement to create an object / relational table to map the XBRL instance document to relational stores
- Easy to create views of instance documents-based on the taxonomy (extended links)

4) Simplifies and improves data management over time:

- Provides a single source of truth
- Schemas / taxonomies can be updated easily without breaking the application or database
- Queries still work without having to migrate old instance documents to the new schema / taxonomy
- Maintains the fidelity and integrity of the original XBRL document as stored

5) Leverages the capabilities of industry standard relational database tools and services:

- Maximizes and enhances existing IT legacy systems, databases, business intelligence, data warehouse, ETL (Extract, Transform, and Load), EAI (Enterprise Application Integration), and enterprise applications
- Enables open and extensible data integration and business reporting
- Provides for backup and recovery
- Enables scalability

Conclusion

UBmatrix has been at the forefront of developing software that takes advantage of the XBRL standard to deliver robust and effective financial reporting systems to government agencies, market regulators, and enterprises alike. The UBmatrix XBRL Processing Engine (XPE) is recognized as one of the leading engines for generating, converting, validating, and consuming XBRL documents.

Release 1.0 of XBRL-DB will be distributed as an option to UBmatrix's Processing Engine. UBmatrix XBRL-DB provides a single high performance storage mechanism which **supports all three storage, query, and retrieval use cases**. It provides the ability to perform both fact-based and taxonomy-only queries. It leverage information collected in XBRL documents, shortens application development times, simplifies and improves data management, and leverages the capabilities of industry standard relational database tools and services.

For more information on XBRL-DB or other UBmatrix solutions, please visit www.ubmatrix.com. For inquiries, email info@ubmatrix.com.

Appendix A: Glossary¹ of XBRL Terms

abstract – Taxonomy element attribute used with the presentation linkbase hierarchy as a header to group items together. Abstract items cannot be used to tag data.

attributes – Defining aspects of XBRL elements such as name, balance, data type, abstract, etc.

authoritative reference – Information referring to specific accounting literature (pronouncements or standards) derived from various authoritative sources (SEC, FASB, AICPA, etc.) and used to help define an element.

balance – Attribute of a monetary item type; a balance must be the natural balance of the element—“credit” or “debit”—and, thus, indicates how related elements are to be summed, based on the weight attribute (-1 or +1).

calculation linkbase – Additive relationships between monetary items expressed as parent-child hierarchies. Each calculation child has a weight attribute (+1 or -1) based upon its natural balance of the parent and child items.

concept – Definition of a kind of fact that can be reported about the activities or nature of a business enterprise. Concepts may be equivalent to or composed of XBRL elements, which are normally the line items of a financial statement.

context – Entity and report-specific information (reporting period, segment information, etc.) required by XBRL that allows tagged data to be understood in relation to other information.

documentation label – Label role for capturing a human-readable description of a reporting concept.

dimensions – Structures that allow data to be organized and presented according to different criteria (such as in a pivot table). Dimensions represent each of the different aspects by which a fact may be characterized. For example, Total Sales Revenue for a period might be reported by business unit or division, by country, by product, etc. A dimension has only one effective domain.

domain – Set of members.

domain member – One of the possibilities within a set of members.

element – XBRL components (items, domain members, dimensions, etc.). A semantic or linguistic representation of a financial reporting concept and thus, often, the same as line items in a financial statement.

element type – Data types (monetary, string, share, decimal, etc.) that define the kind of data to be tagged with the element name.

extended link role – Highest level of a parent-child hierarchy used to categorize item relationships at the financial statement, schedule, or industry level.

extension taxonomy – A taxonomy that allows users to add to the base taxonomy in order to create new elements or customize taxonomy content (presentation, calculation, labels, etc.) without altering the original.

FRTA – Financial Reporting Taxonomies Architecture; guides the creation and use of taxonomies

¹ <http://xbrl.us/USGAAPreview/guide/Pages/appendixA.aspx>

by presenting a recommended design architecture as well as rules and conventions for taxonomies and instance documents.

fact – Usually a value or other information tagged by a taxonomy element or elements and contained in an instance document

GAAP – The first standard of reporting requires an auditor who has audited financial statements in accordance with generally accepted auditing standards to state in the auditor's report whether the statements are presented in conformity with generally accepted accounting principles. The phrase "generally accepted accounting principles" is a technical accounting term that encompasses the conventions, rules, and procedures necessary to define accepted accounting practice at a particular time. It includes not only broad guidelines of general application, but also detailed practices and procedures. Those conventions, rules, and procedures provide a standard by which to measure financial presentations. (AICPA, *Professional Standards*, vol. 1, AU sec. 411; AICPA, PCAOB *Standards and Related Rules*, AU sec 411)

hierarchy – Parent-child trees (presentation, calculation, etc.) used to express and navigate relationships.

instance document – XML file that contains business reporting information and represents a collection of financial facts and report-specific information using tags from one or more XBRL taxonomies.

item – Type of XBRL element used to represent financial reporting concepts and associated relationships for tagging values in instance documents.

label – Human-readable name for an element; each element has at least one, generally known as a "standard" label. Optional sets of labels can also exist; each is given a separate defining "role" or name.

linkbase – Part of a taxonomy used to define specific relationships and other data about elements. There are five standard linkbase types: Presentation, Calculation, Definition (Dimensions), Label, and Reference.

mapping – Process of determining which XBRL elements correspond to individual entries in a financial statement and which elements must be created by extension.

name – Assigns elements and their attributes a unique identifier in a taxonomy.

namespace – Universal resource identifier (URI) for qualifying elements used in XBRL instance documents to identify the taxonomy from which they originate.

parent-child hierarchy – Relationship between concepts or elements that indicates subordination of one to the other as represented in a print listing or financial statement presentation. XBRL taxonomy linkbases use parent-child hierarchies to model presentation, calculation, and definition (dimension) information.

period type – Attribute that determines if an element is reported as an instant or duration time period.

presentation linkbase – Defines the organizational relationships (order) of elements using parent-child hierarchies; it presents the taxonomy elements to users and allows them to navigate the content.

reference linkbase – Part of a taxonomy used to associate references to authoritative literature with elements.

root – First-level element in a parent-child tree (hierarchy); can appear only once in that tree.

scenario – Tag that allows for additional information to be included in the context of an instance document; this information encloses in particular the type of data reported (for example actual, budgeted, pro forma, etc.).

schema – Part of the taxonomy that defines XBRL elements and their attributes.

segment – Tag that allows additional information to be included in the context of an instance document; this information captures segment information such as an entity's business units or divisions and regions.

standard label – Human-readable text label for XBRL elements.

tag – Markup information that describes a unit of data in an instance document and encloses it in angle brackets (" $<>$ " and " $</>$ "). All values in an instance document are enclosed by tags that represent taxonomy elements in XBRL.

taxonomy – Electronic dictionary of business reporting concepts (elements) used to report business data. A taxonomy is composed of an XML schema (.xds file) and linkbase files directly referenced by that schema. The taxonomy schema files plus the linkbases define the concepts (elements) and relationships that form the basis of the taxonomy. The set of related schemas and linkbases altogether constitutes a taxonomy.

tree – Common name for parent-child hierarchies used to define relationships in an XBRL taxonomy.

tuple – Tuples must not be used in any version of the XBRL US GAAP Taxonomies. Tuples may be mentioned in software applications to ensure backward compatibility with previously created instance documents. The functionality previously addressed with tuples has been replaced with dimensions.

unit/measure – Part of an instance document context used to define the units in which numeric items have been measured (Dollars, shares, Euros, etc.).

validation – Process of checking instance documents to ensure that both taxonomies and instance documents correctly meet the rules of the XBRL specification.

weight – Calculation relationship attribute (-1 or +1) that works in conjunction with the balance of the parent and child monetary elements to determine the arithmetic summation relationship between them.

XBRL – Stands for **Extensible Business Reporting Language**; an XML-based standard for electronic communication of financial and business data.

XBRL specification – Detailed description of XML syntax, semantics, and structures, etc. that prescribe how XBRL is constructed. The current Specification 2.1 is used primarily by IT professionals in developing tools and software for XBRL applications.

XBRL US GAAP Taxonomies v1.0 – An electronic, machine-readable dictionary containing thousands of individual financial reporting concepts to which it has assigned human-readable labels, additive relationships, definitions, and applicable authoritative references. It can be used for building financial reporting instance documents for companies that conform to U.S. GAAP.

XML – Stands for Extensible Markup Language, which is used to describe and define data by allowing users to customize their own tags (in contrast to HTML where the tags are predefined). XBRL is an XML standard.