

Datenbanken und XML: Kommerzielle Anbieter (global players)

Rupert Späth

`info@rupert-spaeth.de`

Zusammenfassung

In dieser Arbeit wird untersucht, wie sich die wichtigsten vier Datenbanksysteme in den letzten Jahren bezüglich XML entwickelt haben und was sie aktuell bieten. Soweit verfügbar wird dabei auch ein kleiner Ausblick in die nahe Zukunft gegeben. Konkret wird untersucht, wie sie mit XML Daten umgehen (speichern) und welche Techniken sie zur Abfrage und Manipulation bieten.

1 Einleitung

XML hat sich in den letzten 10 Jahren zu einem der wichtigsten Datenaustauschformate entwickelt. Dementsprechend liegt es nahe, Datenbanksysteme zu entwickeln, die XML-Daten verarbeiten können. Prinzipiell können Datenbanksysteme in Systeme kommerzielle Anbieter (global players) und alternativer Anbieter unterteilt werden. Diese Arbeit untersucht, was die wichtigsten kommerziellen Datenbankanbieter im Bereich XML anbieten.

Die Datenbank-Systeme MySQL, Oracle Database, IBM DB2 UDB und MS SQL Server sind die führenden kommerziellen Datenbanksysteme [Joi08], [MySQL08h]. In Abbildung 1.1 wird die Entwicklung der Marktanteile grafisch dargestellt.

In dieser Arbeit werden keine Aussagen darüber gemacht, welches der vier Systeme im Bezug auf XML das bessere ist. Vielmehr werden die einzelnen Systeme neutral nacheinander dargestellt. Der Grund liegt darin, dass zu jedem der vorgestellten Systeme zahlreiche Texte gefunden werden kann, die genau erklären, warum das eine System besser als das andere ist. Tatsächlich zeigt sich häufig bei den Texten, die ein System eindeutig besser beurteilen, dass meist die entsprechende Firma dahinter steht. [Ora06]

Mit den Kapiteln 2 in dem eine kurze Einführung in den XML-Datentyp gegeben wird und dem Kapitel 3, das die Gemeinsamkeiten der dargestellten Datenbanksysteme aufzeigt, wird die Grundlage für die nachfolgenden Kapitel geschaffen. Diese beschäftigen sich jeweils mit einem Datenbanksystem. So wird den Systemen IBM DB2 UDB, Oracle Database, MS SQL Server und MySQL jeweils ein Kapitel (4-7) gewidmet. Kapitel 8 fasst die Ergebnisse nochmals zusammen.

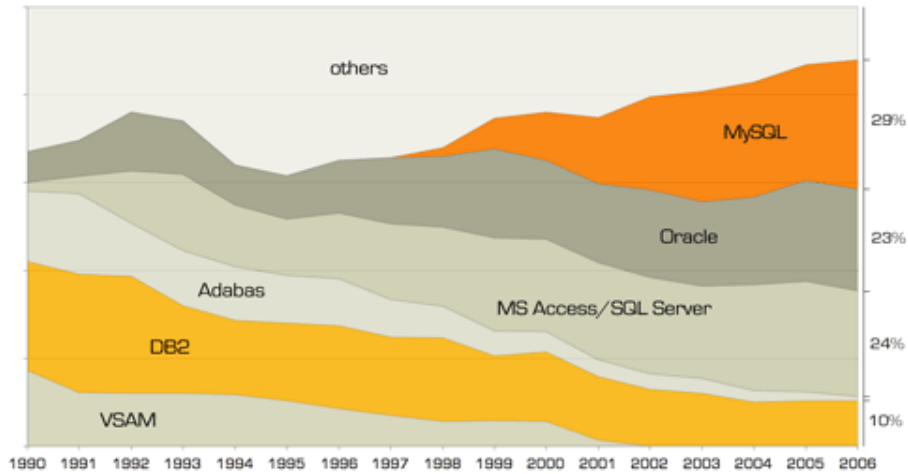


Abbildung 1.1: Marktanteil von Datenbanksystemen [MySQL08h]

2 Voraussetzung

Wichtige Voraussetzung für die vollständige Verarbeitung von XML in den Datenbanksystemen ist der XML-Datentyp der seit 2003 in ISO/IEC 9075-14 spezifiziert ist. [Rys05a] Auf diesen können entsprechende XML-Operationen (z.B. XQuery) ausgeführt werden. Dabei ist es für den Benutzer primär unerheblich, wie ein Feld vom Typ XML physikalisch gespeichert wird. Die einzelnen Vorteile des XML-Datentyps werden von Broseman (S. 24) detailliert aufgelistet [Bro04].

Da aber die unterschiedlichen Hersteller stark abweichende Konzepte verfolgen und diese erhebliche Auswirkungen auf die Leistungsfähigkeit haben, wird es bei den einzelnen Systemen dennoch betrachtet. Aufgrund der Tatsache, dass sich häufig noch alte Versionen im Einsatz befinden, werden nicht nur die aktuellen sondern auch die derzeit noch verbreiteten Systeme betrachtet. Dies ermöglicht gleichzeitig einen Einblick, wie sich die Systeme entwickelt haben. Schließlich soll jeweils noch ein kleiner Ausblick in die zukünftige Entwicklung gegeben werden, soweit darüber schon Informationen verfügbar sind.

Nach Klettke [Kle03] gibt es zahlreiche Methoden XML-Dokumente zu speichern. Sie unterscheidet drei verschiedene Klassen. Die erste speichert die XML-Dokumente in Character Large Objects (CLOB), die zweite speichert eine Graphstruktur der Dokumente und die dritte Klasse bildet die Struktur in ein relationales Schema ab. Diese Unterteilung wird in Abbildung 2.1 nochmals dargestellt.

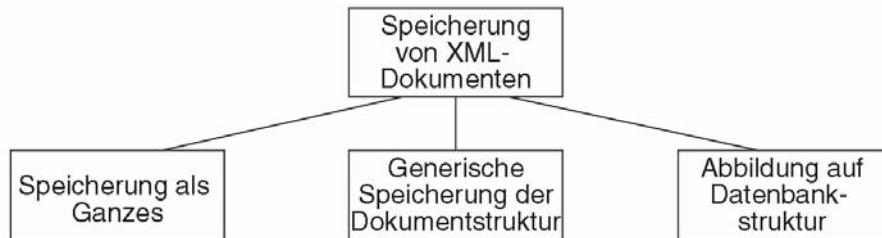


Abbildung 2.1: Klassifikation der Speicherung von XML-Dokumenten [Kle03]

3 Gemeinsamkeiten der vorgestellten Systeme

Bei den in dieser Arbeit betrachteten Datenbanksystemen (MySQL, Oracle Database, IBM DB2 UDB und MS SQL Server) handelt es sich um ursprünglich relationale Datenbanken, die um XML-Funktionalität erweitert wurden. Thomas Müller führt klar aus, dass diese Systeme auf absehbare Zeit nicht durch reine XML-Datenbanken in ihrer Vormachtstellung abgelöst werden, weil SQL dafür viel zu dominierend ist. [Mül04]

Von allen hier vorgestellten Datenbanksystem gibt es kostenlose Versionen (zumeist Express Edition genannt), die allerdings in der Regel im Funktionsumfang, bei den Administrationstools, und beim Support eingeschränkt sind. [IBM08b], [Mic08a], [MySQL08b], [Ora08a]. So werden bei unbeschränkten Vollversionen schnell Preise von über 20.000€ erreicht [Hei08b]. Ein Preisvergleich der Systeme würde den Rahmen dieser Arbeit sprengen, da es allein beim SQL Server 2005 verschiedene Editionen und Lizenzmodelle gibt [Mic08d].

Die hier vorgestellten Datenbanken haben alle einen relationalen Teil. Dies hat zum Vorteil, dass alle notwendigen Sicherungsmechanismen, z. B. Berechtigungskonzept und die Möglichkeit eine Datenbanksicherung (Backup) zu erstellen, gegeben sind. Weiterhin können alle diese Datenbanken Informationen, die logisch zu den XML-Daten gehören, aber nicht in XML-Form vorliegen, direkt speichern, ohne diese vorher in XML umzuwandeln.

Durch die Verwendung von XML und SQL in einem Datenbanksystem ergibt sich die Notwendigkeit der Typkonversion, da die XML-Datentypen nicht 100% kompatibel zu den SQL-Datentypen sind. [Mic08b]

4 IBM DB2 UDB

4.1 Grundsätzliches zur DB2

IBMs DB2 ist eine relationale Datenbank, die seit über 15 Jahren auf dem Markt verfügbar ist. [Jon03] Ihre Entwicklung geht aber bis in die 70er Jahre zurück. Das Besondere an der DB2 ist, dass sie nicht nur für Linux, Unix und Windows verfügbar ist, sondern auch für IBM-Betriebssysteme, wie sie in Großrechnern eingesetzt werden, erhältlich ist. Deswegen hat IBM der DB2 den Zusatznamen Universal Database

(UDB) gegeben. Die Unterstützung so vieler Betriebssysteme und eine so lange Geschichte führen allerdings dazu, dass die DB2 nicht immer ganz einfach zu administrieren ist. So ist zu beobachten, dass der Marktanteil der DB2 seit Jahren sinkt. Nach der Erfahrung des Autors ist sie heute hauptsächlich dort anzutreffen, wo sie schon lange eingesetzt wird.

4.2 Versionen 7-9

Bereits im Jahre 2001 gab es für Version 7 der DB2 durch den sogenannten XML-Extender XML-Unterstützung [Cheng2000]. Aber auch in der Version 8 wurde diese Funktionalität nicht in den Kern integriert. ([Bro04] S. 36).

Dem gegenüber gibt es nun in der aktuellen Version 9 (Codename Viper, verfügbar seit 2007) eine vollständige XML-Integration. [Sar06]

4.3 Der XML Extender

Der XML-Extender unterstützt DTDs. [Cheng00] Darüber hinaus kann er XML Dokumente wahlweise in relationale Tabellen und Spalten zerlegen oder als gesamtes Dokument in einem XML-Datentyp speichern. Weiterhin unterstützt er Indizes auf XML-Dokumente. Als Abfragemöglichkeiten bietet er Extensive Stylesheet Language Transformation (XSLT) und XPath. XPath-Ausdrücke können in die SQL-Syntax eingebunden werden. Allerdings ist man bei der Entwicklung noch stark davon ausgegangen, dass die XML-Dokumente in die relationale Struktur zerlegt werden und dementsprechend direkt abgefragt werden. Einen transparenten XML-Datentyp gibt es somit nicht. XML-Dokumente können mittels Stored Procedures oder einer Technologie namens Net.Data erzeugt werden, die inzwischen in IBM WebSphere aufgegangen ist. [IBM08a]

4.4 Native XML-Unterstützung in Version 9

In der Version 9 (Codename Viper) beschreitet IBM nun einen neuen Weg [Nic05]. Daten werden „nativ“ gespeichert. Dementsprechend nennt IBM dieses Feature pureXML [Ora06]. Gemeint ist damit das Speichern der Dokumente in Datenbank-Seiten in Baum-Strukturen, die dem XML-Datenmodell entsprechen. Ein relationales Mapping findet nicht mehr statt. Stattdessen wird dieses Konzept durch XML-Indizes, XQuery und SQL/XML unterstützt. Auch XML Schema ist verfügbar.

Wie in Abbildung 4.1 dargestellt besitzt die DB2 nun zwei gleichberechtigte Speicher-Mechanismen (Storage-Engines): Eine XML-Speichermechanismus und ein relationaler SQL-Mechanismus. [Sar06] Dabei ist der XML-Datentyp vollkommen transparent gegenüber der Anwendung (bzw. dem Benutzer).

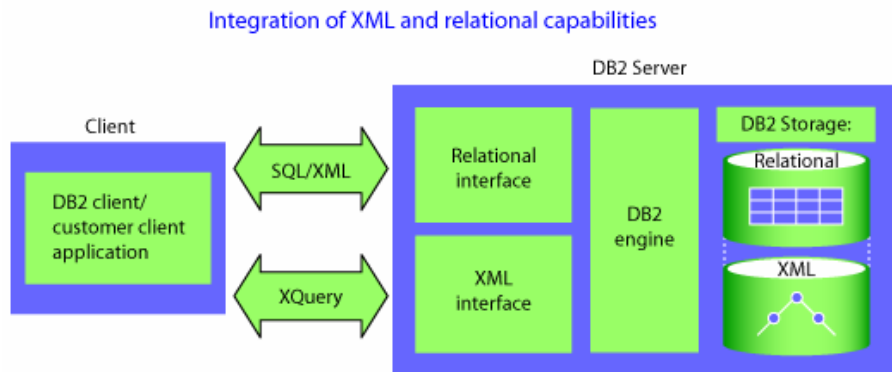


Abbildung 4.1: Architektur der DB2 Version 9 [Sar06]

In relationalen Tabellen können XML-Spalten wie gewöhnliche Spalten verwendet werden. Dabei können dort beliebige Daten(-strukturen) abgelegt werden so lange sie “well formed“ sind, also dem W3C Standard (World Wide Web Consortium) für XML entsprechen. [W3C06]

4.5 Besonderheiten der Implementierung

IBM stellt als besonderes Feature die kostenbasierte Abfrageoptimierung heraus. [IBM06a]. Dabei werden basierend auf Statistiken ein sogenanntes Kostenmodell erstellt, auf dessen Grundlage ein Ausführungsplan der Abfrage erstellt wird. Unter einem Kostenmodell versteht man eine Zeitschätzung für die einzelnen Teilaufgaben, die als Kosten bezeichnet werden. Dieses Vorgehen ist bei relationalen Datenbanken weit verbreitet.

Ab der Version 9.5 (Viper 2) können XML-Daten auch komprimiert gespeichert werden. Dies hat nicht nur den Vorteil, dass der Speicherbedarf reduziert wird, sondern auch, dass ein zweites neues Feature, die Speicherung von XML auf der selben Datenseite wie die relationalen Daten mehr an Bedeutung gewinnt. [Eat07] Üblicherweise werden relationale Daten in relativ kleinen Einheiten, sogenannten Pages (Datenseiten) gespeichert. Durch die Speicherung der XML-Daten auf derselben Datenseite kann ein Performance-Gewinn erzielt werden. Werden diese Daten komprimiert, steht dieser Performance-Gewinn auch noch für einige etwas größere XML-Dokumente zur Verfügung, die sonst an einer anderen Stelle gespeichert werden müssten, und somit beim Zugriff auf eine Tabellenzeile nur langsamer erreicht werden können.

4.6 Zusammenfassung

Mit der Version 9 der DB2 hat IBM eine leistungsstarke Datenbank vorgelegt, die laut IBM XML und SQL gleichberechtigt unterstützt. Dadurch, dass Strategien zur Abfrageoptimierung verwendet werden, mit denen IBM sehr lange Erfahrung hat, darf davon ausgegangen werden, dass die Implementierung sehr performant ist. Allerdings handelt es sich bei der XML-Implementierung um ein kostenpflichtiges Feature, dass selbst in der kostenlosen Express Edition bezahlt werden muss [Ora06].

5 Oracle Database

5.1 Grundsätzliches zur Oracle Database

Oracle Database ist das Datenbanksystem des gleichnamigen Softwareherstellers Oracle. Die Codebasis der Oracle Database geht auf Anfang der achtziger Jahre zurück. Das System war unter den hier aufgelisteten Datenbanksystemen das Erste, das einen XML-Datentyp zur Verfügung gestellt hat [Bro04]. Darüber hinaus wird dem System häufig die Marktführerschaft bescheinigt. [Pfe06]. All dies dürfte die Ursache dafür sein, dass in Datenbankkreisen das Oracle-Datenbanksystem häufig als das Maß der Dinge betrachtet wird.

5.2 Versionen 8-9

Bereits seit der Version 8 ist in der Oracle Database ein XML Datentyp vorhanden [Ora04]. Dieser ist bereits voll als XML-Datentyp integriert. In der Version 9 konnte dieser an XML Schemas gebunden werden. Außerdem wurde bereits der Standard SQL/XML unterstützt und es konnten XML Views erstellt werden.

Schwinn unterteilt XML-Dokumente in dokument- und datenzentrische Dokumente [Ora03a]. Dokumentenzentrisches Speichern wird mittels CLOB (Character Large Object) realisiert, während datenzentrisches Speichern mittels objektrelationalem Mapping realisiert wird. Dazu ist die Verwendung eines XML-Schemas notwendig. Damit wurde bereits das Fundament für die vollkommene Unterstützung von XQuery gelegt. Weiterhin wird das XML Schema aber auch zur Validierung des Dokuments während des Einfügens verwendet [Ora03b].

Beim sogenannten „shredding“ besteht leicht die Gefahr, dass das ursprüngliche Dokument seine Wiedergabetreue verliert, also das Dinge wie „whitespaces“ (Leerzeichen etc.) oder Prozessanweisungen verloren gehen [Ora08e]. Bei diesen Leerzeichen handelt es sich um zulässige Zeichen, die sich zwischen XML-Elementen befinden, auf die semantische Aussage des Dokuments keinen Einfluss haben, aber das Erscheinungsbild für den Betrachter verändern. Durch die Verwendung spezieller binärer Attribute wird der Verlust solcher Informationen verhindert.

5.3 XQuery

Die Implementierung von XQuery basiert seit der Version 9 auf einem hybriden Ansatz [Ora05]. Dabei wird versucht möglichst viele XQuery-Befehle in SQL zu übersetzen und nur dort, wo dies nicht möglich ist XQuery direkt „nativ“ auszuführen. So kann von der hohen Performance der relationalen Datenbankengine profitiert werden. Das Vorgehen wird in Abbildung 5.1 veranschaulicht.

Das Übersetzen von XML-Abfragen in SQL-Abfragen ist eine Besonderheit von Oracle Database und wird von Oracle als „XML Query Rewrite“ bezeichnet [Ora04]. Die SQL-Abfragen werden gegen die relationalen Tabellen ausgeführt, die vorher durch das in 5.2 genannte Mapping erstellt wurden.

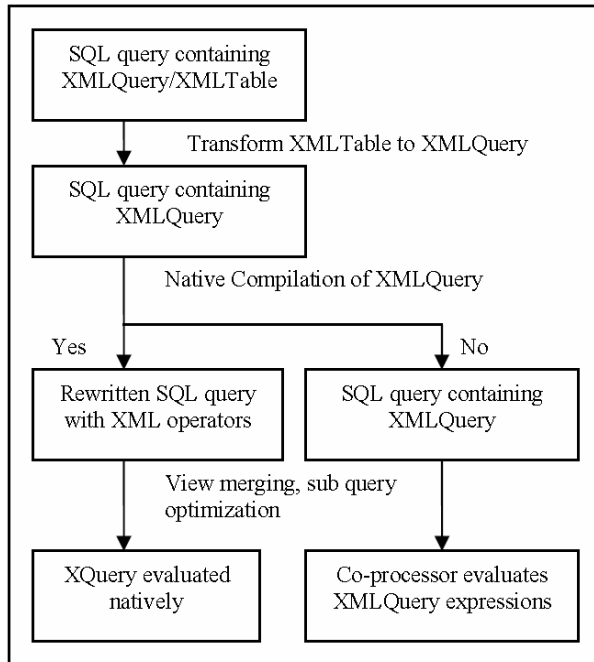


Abbildung 5.1: XQuery hybrid Strategie [Ora05]

5.4 RDF/OWL

Seit der Version 10.2 wird das Resource Description Framework (RDF) unterstützt [Ora08c]. In Version 11.1 wird nun zusätzlich eine Teilmenge der Web Ontology Language (OWL) unterstützt. Die Abbildung 5.2 zeigt die semantischen Fähigkeiten der Oracle Database Version 11.1

Um semantische Informationen speichern zu können besitzt die Oracle Database im Schema MDSYS Tabellen, in denen ein Tripple als eine Entität gespeichert wird [Ora08c]. Dabei kommen konventionelle relationale Techniken und kein XML zum Einsatz. Weitere Informationen zu semantischen Technologien in der Oracle Database finden sich bei [Ora08d].

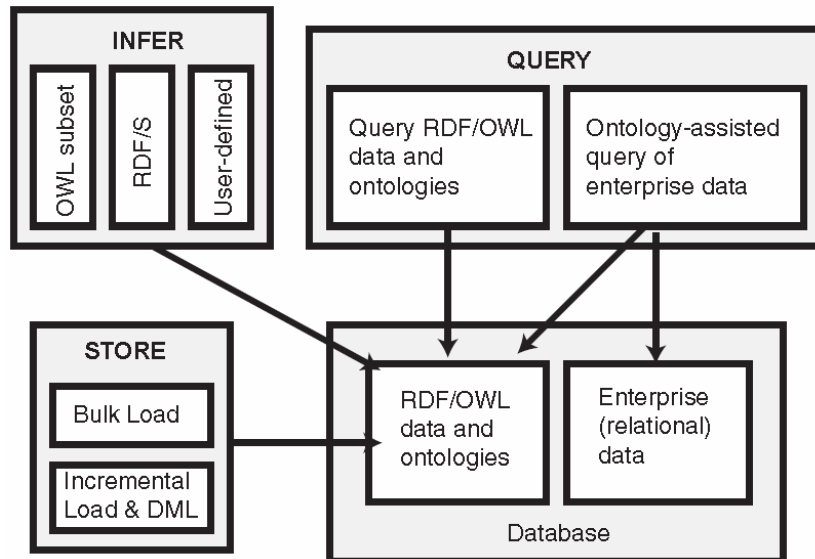


Abbildung 5.2: Semantische Fähigkeiten der Oracle Database [Ora08c]

5.5 Zusammenfassung

Oracle hat über mehrere Versionen hinweg systematisch die XML-Funktionalität der Oracle Database aufgebaut, erweitert und vervollständigt. Somit steht mit der Version 11 reichhaltige Funktionalität zur Verfügung. Im Bereich semantische Technologien ist sie zwar führend, aber die Funktionalität muss in der nächsten Version noch vervollständigt werden.

6 Microsoft SQL Server

6.1 Grundsätzliches zum SQL Server

Im Unterschied zu den anderen hier vorgestellten Datenbanksystemen ist der MS SQL Server nur für Windows verfügbar. Eine Installation mittels Wine (siehe Abkürzungsverzeichnis) auf einem Linux- oder UNIX-System hat bestenfalls experimentellen Charakter. Es konnten keine Anleitungen, Empfehlungen, Referenzinstallationen oder Erfahrungen mit entsprechenden Systemen gefunden werden. Selbst der Datenbankzugriff von einem Linux-Client auf einen SQL Server ist deutlich schwieriger, da es von Microsoft keinen entsprechenden Treiber gibt.

Der SQL Server besticht durch ausgereifte GUI-Tools, die bereits im Lieferumfang enthalten sind.

6.2 Die Versionen des SQL Servers

6.2.1 SQL Server 2000

Die XML-Unterstützung des SQL Servers 2000 war zunächst sehr rudimentär [Bro04]. So konnte z. B. die Ausgabe eines Select-Statements als XML formatiert werden. Ein XML-Datentyp war nicht verfügbar. Durch ServicePacks und Patches sowie durch die Integration in den IIS (Internet Information Services) wurde die XML-Unterstützung zumindest teilweise nachgerüstet. So konnte die Zeit überbrückt werden, bis mit dem SQL Server 2005 eine Datenbank mit einer belastbaren XML-Funktionalität zur Verfügung stand.

6.2.2 SQL Server 2005

Mit dem SQL Server 2005 steht erstmalig dem SQL Server ein XML-Datentyp zur Verfügung. Dieser wird intern als BLOB (Binary Large Object) gespeichert [Mic08b]. Beschränkungen bestehen hinsichtlich der Größe des XML-Dokuments (2 Gbyte) und der Hierarchietiefe (128). Die XML-Daten können an ein XML-Schema gebunden werden. Dies wird als „typed XML“ bezeichnet. Dann werden die Daten gegen dieses Schema validiert. Die Verwendung eines Schemas erhöht die Abfrageleistung. Dies dürfte daran liegen, dass der Abfrageoptimierer das Schema verwendet. XML-Daten die an kein XML-Schema gebunden sind werden als „untyped XML“ bezeichnet. Ansonsten verhält sich der XML-Datentyp in der Regel wie ein anderer SQL Datentyp. So kann dieser auch in der CLR (common language runtime) verwendet werden, sodass SQL-Funktionen in .net geschrieben werden können und Clientanwendungen diesen verwenden können.

Datenselektion und –manipulation ist mittels XQuery möglich. [Mic08b] Dazu stehen die fünf Methoden `query()`, `value()`, `exist()`, `nodes()` und für die Manipulation `modify()` zur Verfügung. Mit `modify()` können Teilbäume entfernt und hinzugefügt werden, sowie Skalare geändert werden. `Modify()` wird zusammen mit dem Update-Statement verwendet, die anderen Methoden sowohl mit den SQL Befehlen `select` als auch `update`.

XML-Daten haben einen sogenannten primären Index (basierend auf einem B⁺Baum [Mic04]), den gegebenenfalls mehrere sekundäre Indizes verwenden können [Mic08b]. Diese können von unterschiedlichem Typ (PATH, PROPERTY oder VALUE) sein. Weitere Informationen insbesondere zur Performance können [Mic04] entnommen werden.

Mit SQLXML steht ein Mapping-Mechanismus zwischen XML und dem relationalen Schema zur Verfügung.

6.2.3 SQL Server 2008

Nach [Hei08a] wird der SQL Server 2008 voraussichtlich im dritten Quartal 2008 verfügbar sein. Die offizielle Microsoftdokumentation „What’s New for XML in SQL Server 2008“ [Mic08c] trägt das Datum August 2008! Darin werden als wesentliche Verbesserungen in den Bereichen Schema-Validierung, XQuery Unterstützung und DML Einfügungen genannt. Konkret wird bei XQuery der `let`-Befehl genannt. Somit

müssen die Neuerungen dahingehend eingestuft werden, dass Funktionalität vervollständigt wird, die bereits in der Version 2005 zur Verfügung stand.

6.3 Besonderheiten der Implementierung

Als Besonderheit des SQL Servers 2005 ist zu nennen, dass nicht nur XML-Spalten in relationalen Tabellen erstellt werden können, sondern auch, dass XML-Tabellen erstellt werden können. [Rys05b] Damit kann ein Verhalten erzeugt werden, dass dem einer reinen XML-Datenbank entspricht.

6.4 Schlussfolgerung

Mit dem SQL Server 2008, der dieses Jahr voraussichtlich auf den Markt kommen wird, legt Microsoft ein Datenbanksystem vor, das im Bezug auf XML alles bietet, was man von einem Datenbanksystem erwarten darf. Allerdings sticht der SQL Server nicht durch besondere Features oder Implementierungen heraus.

7 Sun MySQL

7.1 Grundsätzliches zu MySQL

Wie aus der Grafik in der Einleitung entnommen werden kann, ist MySQL inzwischen die weitverbreiteste Datenbank. Im Jahre 2008 wurde MySQL AB von der Firma SUN Microsystems gekauft [Sun08], [MySQL08a]. MySQL wurde von Anfang an als kommerzielles Datenbanksystem entwickelt. Die Besonderheit an MySQL ist, dass es eine OpenSource-Datenbank ist. Das heißt nicht, dass das System von Freiwilligen entwickelt wird, aber dass das System „näher am Kunden“ entwickelt wird. Dieses Konzept dürfte der Grund sein, warum die MySQL inzwischen derartige Verbreitung erfahren hat. Insbesondere in der wichtigen OpenSource-Community spielt MySQL eine große Rolle, zum Beispiel im System LAMP.

Da die MySQL in vielen Bereichen im Funktionsumfang (noch) nicht mit den anderen Systemen mithalten kann (insbesondere im Bereich XML, wie nachfolgend gezeigt wird), häufig das Image der kostenlosen Open Source-Datenbank hat und jünger als die anderen Datenbank-Systeme ist, wird diese vielfach nicht in den direkten Vergleich zu den Systemen von Oracle, IBM und Microsoft gesetzt. [Bro04], [Rys05a], [Bau06] Dementsprechend haben die meisten Quellen in dem Literaturverzeichnis dieser Ausarbeitung, die Aussagen über mehrere Datenbanksysteme gemacht haben, MySQL nicht berücksichtigt.

Eine Besonderheit von MySQL ist, dass sie unterschiedliche Storage Engines unterstützt, unter anderem InnoDB (ein Produkt einer Oracle-Tochter) und IBMs DB2 [Ora08b], [IBM07a].

7.2 Versionen 5.0 bis 6.0

Noch im Jahre 2003 gab es keine Unterstützung für XML. [DuB03]. Dementsprechend werden nur die Versionen 5.0 und neuer betrachtet. Die aktuelle Version von MySQL ist 5.0. [MySQL08b] Diese wird demnächst von der Version 5.1 abgelöst, die

bereits als „Release Candidate“ zur Verfügung steht. Darüber hinaus befinden sich die Versionen 5.2 und 6.0 derzeit in Entwicklung, sind diese noch in dem sehr frühen, so genannten „Alpha“-Stadium. [MySQL08c]

In der Version 5.0 beschränkt sich die XML-Funktionalität auf die Ausgabe von XML mittels `mysql --xml`. [MySQL08d].

Mit der Version 5.1 sind erstmals XML-Funktionen verfügbar:

`EXTRACTVALUE (XML_document, XPath_string);`

`UPDATEXML (XML_document, XPath_string, new_value);`

Mit diesen Funktionen können Teile von XML-Dokumenten mittels XPath gelesen und geändert werden. Bei XML_document handelt es sich um einen Text-Datentyp.

Ein XML-Datentyp ist nicht verfügbar. [MySQL08e]

In der Version 6 können nun XML-Daten mittels `Load XML` auch eingelesen werden. [MySQL08f], [MySQL08g]

Weitergehende konkrete Pläne für einen XML-Datentyp oder XQuery existieren derzeit nicht, wenngleich bei MySQL erkannt wurde, dass die derzeitige XML-Funktionalität nicht ausreichend ist. [MySQL07a]

7.3 Erweiterungen für MySQL

Wie im vorigen Abschnitt gezeigt wurde, sind die XML-Fähigkeiten der MySQL nur gering. Nachfolgend werden nun noch zwei Entwicklungen kurz aufgezeigt, die die XML-Fähigkeiten der MySQL ergänzen sollen.

7.3.1 XQuery

Die Firma DataDirect bietet ein Tool, das mittels relationalem Mapping auf MySQL Datenbanken zugreift. [Dat08] Dafür haben sie einen eigenen XQuery-Prozessor entwickelt, der XML-Dokumente verarbeiten kann. Dieser Prozessor übersetzt XQuery in SQL und schickt dieses zur Datenbank

7.3.2 SPARQL für MySQL

Für MySQL gibt es eine SPARQL-Implementierung (SPARQL steht für SPARQL Protocol and RDF Query Language), die sich direkt in die Codebasis von MySQL einfügt [Pru08]. Allerdings bezeichnet der Autor Eric Prud'hommeaux (Mitverfasser der W3C Empfehlung zu SPARQL) seinen Ansatz immer noch als Experiment, obwohl es das Projekt schon über zwei Jahre gibt. So betrachtet der Autor selbst seine Implementierung noch nicht als vollständig

7.4 Schlussfolgerung

MySQL wird wie oben dargestellt auch in naher Zukunft (also bis einschließlich Version 6.0) nur rudimentäre XML-Unterstützung bieten können. Externe Tools und Patches können die Situation zwar etwas verbessern, aber nicht grundsätzlich ändern. Ein weitergehender Ausblick zu MySQL kann derzeit nicht gemacht werden, da abzu-

warten bleibt, was der neue Eigentümer von MySQL AB, die Firma SUN Microsystems mit MySQL vor hat. Zusammenfassung

8 Zusammenfassung und Ausblick

Wie gezeigt wurde, besitzen die neuesten Systeme von IBM, Microsoft und Oracle weitgehende XML-Funktionalität. Allein MySQL hinkt weit hinterher. Dies ist um so mehr verwunderlich, da gerade diese im Internet-Bereich (Stichwort LAMP) starke Verwendung findet und zur Zeit ihren Marktanteil ausbaut. Dies könnte ein Indiz darauf sein, dass XML-Funktionalität im Moment (noch) bei Datenbanken keine wesentliche Rolle spielt. Da aber alle genannten Anbieter an der XML-Funktionalität arbeiten, wird vermutlich von diesen erwartet, dass sich dies bald ändert. Dies dürfte wohl auch der Grund sein, warum reine XML-Datenbanken (z. B. Tamino von der Software AG) im Moment noch keine große Rolle spielen. Andererseits könnte auch die schlechte Abfrageleistung bisheriger Systeme dafür verantwortlich sein, dass XML in Datenbanken bisher kaum Verbreitung findet. So sind relationale Abfragen XML-Abfragen überlegen [Rys08]. Weiterhin sind relationale Datenbanken mit XML-Erweiterungen tendenziell performanter als reine XML-Datenbanken [Spe01]. Mit alternativen Ansätzen, mobilen Datenbanken und reinen XML-Datenbanken beschäftigt sich die Arbeit von Mark Mattern (Thema 10).

Abkürzungsverzeichnis

Abkürzung	Steht für	Ergänzende Erklärung
BLOB	Binary Large Object	SQL-Datentyp für (große) Binärdaten
CLOB	Character Large Object	SQL-Datentyp für große Texte
CLR	common language runtime	Virtuelle Maschine der .net Umgebung.
DB2 UDB		Datenbanksystem der Firma IBM
DML	Data Manipulation Language	Bei Datenbanksprachen wird häufig zwischen DML (Datenmanipulationssprache) und DDL (Datendefinitionssprache) unterscheiden.
DTD	Dokumenttypdefinition	Eine XML Schemasprache
GUI	Graphical User Interface	Grafische Benutzeroberfläche
IBM	International Business Machines	Ein amerikanischer Softwarehersteller
IIS	Internet Information Services	Webserver der Firma Microsoft
LAMP	Linux, Apache, MySQL, PHP	Eine weitverbreitete Kombination aus Betriebssystem, Webserver, Datenbank und Programmiersprache für Webanwendungen
MS	Microsoft	amerikanischer Softwarehersteller
OWL	Web Ontology Language	Beschreibungssprache für Ontologien
PHP	PHP: Hypertext Preprocessor	Scriptsprache für Webanwendungen
RDF	Resource Description Framework	Formale Sprache zur Bereitstellung von Metadaten im Web
RDFS	RDF Schema	Beschreibungssprache für Ontologien
SQL	Abgeleitet von SEQUEL	Eine Datenbanksprache
SPARQL	SPARQL Protocol and RDF Query Language	Protokoll und Abfragesprache für RDF
W3C	World Wide Web Consortium	Standardisierungsgremium des WWW
Wine	Wine is no Emulator	Eine Open Source Implementierung der Windows API
XML	Extensible Markup Language	Auszeichnungssprache, häufig zum Datenaustausch verwendet
XSLT	Extensible Stylesheet Language Transformation	Programmiersprache zur Transformation von XML
z. B.	Zum Beispiel	

Literaturverzeichnis

- [Bau06] Bauer, Cordula: XML in kommerziellen Datenbanksystemen, <http://www.inf.uni-konstanz.de/dbis/teaching/ss06/xmlsem/talks/commandout.pdf> - Abgerufen am 18.05.2008
- [Bro04] Matthias Brosemann 2004, XML in kommerziellen Datenbanksystemen, Studienarbeit, Uni Jena - Online nicht verfügbar
- [Cheng00] Josephine Cheng, Jane Xu, "XML and DB2" icde, p. 569, 16th International Conference on Data Engineering (ICDE'00), 2000, <http://www.cs.umd.edu/areas/db/dbchat/papers/xml-db2.pdf> - Abgerufen am 18.05.2008
- [Dat08] N. N., Using DataDirect XQuery with MySQL, <http://www.xquery.com/mysql> - Abgerufen am 18.05.2008
- [DuB03] Paul DuBois 2003, Using XML with MySQL, <http://www.kitebird.com/articles/mysql-xml.html> - Abgerufen am 18.05.2008
- [Eat07] Chris Eaton, Compression of XML in DB2 9.5, Ittoolbox Blogs, 2007, <http://blogs.ittoolbox.com/database/technology/archives/compression-of-xml-in-db2-95-20079> - Abgerufen am 18.05.2008
- [Hei08a] N. N., Microsoft SQL Server 2008 kommt später, heise online, 28.01.2008, <http://www.heise.de/newsticker/Microsoft-SQL-Server-2008-kommt-spaeter--/meldung/102574> - Abgerufen am 18.05.2008
- [Hei08b] N. N., Preisvergleich: Microsoft: SQL Server 2005 – Enterprise Edition, 1 CPU Lizenz X64, <http://www.heise.de/preisvergleich/a184121.html> - Abgerufen am 18.05.2008
- [IBM06a] Balmin et Al., Cost-based optimization in DB2 XML, IBM Systems Journal, vol 45, No 2, 2006, Seiten 299-319: <http://www.research.ibm.com/journal/sj/452/balmin.pdf> - Abgerufen am 18.05.2008
- [IBM07a] N. N., MySQL AB und IBM kündigen Open-Source-Datenbankunterstützung für die System i-Plattform an, Santa Clara, Kalifornien, USA - 25. April 2007, <http://www.ibm.com/news/de/de/2007/04/25.html> - Abgerufen am 18.05.2008
- [IBM08a] N. N., Net.data, <http://www-306.ibm.com/software/data/net.data/> - Abgerufen am 18.05.2008
- [IBM08b] N. N., DB2 Express-C, <http://www-306.ibm.com/software/data/db2/express/> - Abgerufen am 18.05.2008
- [Joi08] N. N., Skillmarkt - Datenbanken <http://www.joinvision.com/jv/x/n/t-TStatMarketDetail-statistic-db> - Abgerufen am 18.05.2008
- [Jon03] Jeff Jobs, 2003, The Big Picture: IBM DB2® Information Management Software and DB2® Universal Database™, <http://www3.software.ibm.com/ibmdl/pub/software/dw/dm/db2/0301jones/db2bigpicture.pdf> - Abgerufen am 18.05.2008
- [Kle03] Klettke, Meike und Meyer, Holger: Speicherung von XML-Dokumenten – eine Klassifikation Datenbankspektrum 5/2003,

- <http://www.dbis.prakinf.tu-ilmenau.de/papers/dbspektrum/dbs-05-40.pdf> - Abgerufen am 18.05.2008
- [Mic04] Pal et. Al., Indexing XML Data Stored in a Relational Database, 30th VLDB Conference, Toronto, Canada, 2004, <http://www.isys.ucl.ac.be/vldb04/eProceedings/contents/pdf/IND5P2.PDF> - Abgerufen am 18.05.2008
- [Mic08a] N. N., SQL Server Express - Easy to Use, <http://msdn.microsoft.com/en-us/express/aa718378.aspx> - Abgerufen am 18.05.2008
- [Mic08b] N. N., XML Support in Microsoft SQL Server 2005, <http://msdn.microsoft.com/en-us/library/ms345117.aspx> - Abgerufen am 18.05.2008
- [Mic08c] What's New for XML in SQL Server 2008 <http://download.microsoft.com/download/a/c/d/acd8e043-d69b-4f09-bc9e-4168b65aaa71/WhatsNewSQL2008XML.doc> - Abgerufen am 18.05.2008
- [Mic08d] N. N., Microsoft Product List, 2008, <http://www.microsoftvolumelicensing.com/userights/Downloader.aspx?DocumentId=1454> - Abgerufen am 18.05.2008
- [Mül04] Müller, Thomas: SQL-basierte Datenbankzugriffe und XML: Klassifizierung von Anwendungsprogrammen, Jena, 2004, <http://www.dbs.cs.uni-duesseldorf.de/gvd2004/papers/MuellerThomas.pdf> - Abgerufen am 18.05.2008
- [MySQL07a] Stephens, MySQL Forums :: XML :: Any plans for Xml data type?, 2007, <http://forums.mysql.com/read.php?44,155525,157883#msg-157883> - Abgerufen am 18.05.2008
- [MySQL08a] N. N., MySQL News Announcements (2008), <http://www.mysql.com/news-and-events/> - Abgerufen am 18.05.2008
- [MySQL08b] N., N. MySQL Downloads, <http://dev.mysql.com/downloads/> - Abgerufen am 18.05.2008
- [MySQL08c] N. N., MySQL 6.0 Reference Manual :: C MySQL Change History, <http://dev.mysql.com/doc/refman/6.0/en/news.html> - Abgerufen am 18.05.2008
- [MySQL08d] N. N., D.1.5. Release Notes for MySQL Community Server 5.0.41 (01 May 2007), <http://dev.mysql.com/doc/refman/5.0/en/releasenotes-cs-5-0-41.html> - Abgerufen am 18.05.2008
- [MySQL08e] Barkov, Gulutzan, MySQL 5.1's New XML Functions <http://dev.mysql.com/tech-resources/articles/mysql-5.1-xml.html> - Abgerufen am 18.05.2008
- [MySQL08f] Stephens, Using XML in MySQL 5.1 and 6.0 <http://dev.mysql.com/tech-resources/articles/xml-in-mysql5.1-6.0.html> - Abgerufen am 18.05.2008
- [MySQL08g] N. N., LOAD XML Syntax, <http://dev.mysql.com/doc/refman/6.0/en/load-xml.html> - Abgerufen am 18.05.2008

- [MySQL08h] N. N., MySQL :: Market Share, <http://www.mysql.com/why-mysql/marketshare/> - Abgerufen am 18.05.2008
- [Nic05] Nicola, van der Linden, 31st VLDB Conference, Trondheim, Norway, 2005, Native XML Support in DB2 Universal Database, <http://www.vldb2005.org/program/paper/thu/p1164-nicola.pdf> - Abgerufen am 18.05.2008 - Abgerufen am 18.05.2008
- [Ora03a] Ulrike Schwinn, XML in der Oracle Datenbank „relational and beyond“, in „Tagungsband der BTW 2003: Datenbanksysteme für Business, Technologie und Web“, 2003, <http://doesen0.informatik.uni-leipzig.de/proceedings/paper/IP6.pdf> - Abgerufen am 18.05.2008
- [Ora03b] Murthy, Banerjee, XML Schemas in Oracle XML DB, 29th VLDB Conference, Berlin, Deutschland, 2003, <http://www.oracle.com/technology/tech/xml/xmldb/Misc/vldbschemas4.pdf> - Abgerufen am 18.05.2008
- [Ora04] Krishnaprasad et. Al., Query Rewrite for XML in Oracle XML DB, 30th VLDB Conference, Toronto, Canada, 2004, <http://www.oracle.com/technology/tech/xml/xquery/pdf/vldb04-814-XMLRewrite-in-OracleXMLDB.pdf> - Abgerufen am 18.05.2008
- [Ora05] Liu, Krishnaprasad, Arora , Native XQuery Processing in Oracle XMLDB, Redwood Shores, CA 94065, U.S.A, 2005, http://www.oracle.com/technology/tech/xml/xquery/pdf/114_xquery_sigmod2005-final.pdf - Abgerufen am 18.05.2008
- [Ora06] Lee, Oracle XML DB vs. DB2 v9.1 pure XML, 2006, http://www.oracle.com/technology/tech/xml/xmldb/Current/xmldb_vs_viper_ext.pdf - Abgerufen am 18.05.2008
- [Ora08a] N. N., Oracle Database 10g Express Edition, <http://www.oracle.com/technology/products/database/xe/index.html> - Abgerufen am 18.05.2008
- [Ora08b] Harris et. al., InnoDB, 2007, <http://wiki.oracle.com/page/InnoDB> - Abgerufen am 18.05.2008
- [Ora08c] Murray et. al., 2007, Semantic Technologies Developer's Guide 11g Release 1 (11.1) http://download.oracle.com/docs/cd/B28359_01/appdev.111/b28397.pdf - Abgerufen am 18.05.2008
- [Ora08d] N.N., Semantic Technologies Center, http://www.oracle.com/technology/tech/semantic_technologies/index.html - Abgerufen am 18.05.2008
- [Ora08e] N. N., Getting Started with Oracle XML DB, http://www.filibeto.org/sun/lib/nonsun/oracle/10.2.0.1.0/B19306_01/appdev.102/b14259/xdm02rep.htm - Abgerufen am 18.05.2008
- [Pfe06] Pfele, Datenbanksysteme, Script zur Vorlesung, Heidelberg 2005, http://www-dbs.informatik.uni-heidelberg.de/teaching/ss2006/dbs/skript/dbs01_4pages.pdf - Abgerufen am 18.05.2008
- [Pru08] Prud'hommeaux, Eric, Notes on Adding SPARQL to MySQL, 2008, <http://www.w3.org/2005/05/22-SPARQL-MySQL> - Abgerufen am 18.05.2008

- [Rys05a] Rys, Chamberlin, Florescu, XML and Relational Database Management Systems: the Inside Story, SIGMOD 05, Baltimore, USA, 2005, <http://www.cs.uiuc.edu/class/fa05/cs591han/sigmodpods05/sigmod/pdf/p945-rys.pdf> - Abgerufen am 18.05.2008
- [Rys05b] Rys, Michael, XML and Relational Database Management Systems: Inside Microsoft® SQL Server™ 2005, SIGMOD 05, Baltimore, USA, 2005, <http://www.cs.uiuc.edu/class/fa05/cs591han/sigmodpods05/sigmod/pdf/p958-rys.pdf> - Abgerufen am 18.05.2008
- [Rys08] Rys Michael, Michael Rys on XML in SQL Server 2005 in Tim Anderson's ITWriting, <http://www.itwriting.com/sqlxml.php> und <http://www.itwriting.com/sqlxml2.php> - Abgerufen am 18.05.2008
- [Sar06] Saracco, Cynthia M., What's new in DB2 Viper, 2006, <http://www.ibm.com/developerworks/db2/library/techarticle/dm-0602saracco/> - Abgerufen am 18.05.2008
- [Spe01] Schindler, Specht, Schmatloch, Performanz von XML-Datenbanken, Ilmenau, 2001, http://www.old.netobjectdays.org/pdf/01/papers/ws-mik/specht_schindler_mik2001.pdf - Abgerufen am 18.05.2008
- [Sun08] N. N., Sun Microsystems Announces Agreement to Acquire MySQL, Developer of the World's Most Popular Open Source Database, Santa Clara, USA, 2008, <http://www.sun.com/aboutsun/pr/2008-01/sunflash.20080116.1.xml> - Abgerufen am 18.05.2008
- [W3C06] Bray et. al. Extensible Markup Language (XML) 1.0 (Fourth Edition), 2006, <http://www.w3.org/TR/xml/> - Abgerufen am 18.05.2008