**XML-Datenbanksysteme (BaseX)**

**Leistung von XML-Datenbanksystemen**

XML-Datenbestände sind Bürger erster Klasse

* Datenmodell XDM, XML-Kodierung auf Nutzungsebene
* Datenstrukturen zum Speichern transparent
* Unterstützung von **XQuery** als Abfragesprache

Geeignet für große Datenbestände, effizientes Extrahieren und Aktualisieren von Information

* wichtig für Effizienz: Indexierung

Nutzungsverwaltung, Zugriffseinschränkungen / Security

Unterstützung von Nebenläufigkeit

Transaktionskonzept (ACID-Kriterien)

Batch Loading, Batch Export

Unterstützen **Zugriff über HTTP**

* **geeignet zur Realisierung von Webanwendungen auf Basis von XML-Technologie**

**Systeme**

eXist (open source)

BaseX (open source)

MarkLogic (kommerziell)

* Kommerzielles Produkt, Marketing als NoSQL-Lösung für Enterprises
* Alleinstellungsmerkmal: Performanz für sehr große Datenmengen

**BaseX**

Offene, frei verfügbare XML-Datenbank

* Organisiert XML-Daten
* Enthält XQuery-Prozessor
* Unterstützt W3C XQuery Update Facility und XQuery Fulltext Facility
* Verfügt über graphisches Front End für lokales System
* Verfügt über Administrations-Tool als Web-Anwendung
* Integrierbar in Entwicklungsumgebungen oXygen, Eclipse etc

**Welche Datenmodelle für XML gibt es als W3C-Standard?**

XDM (XML Data Model), XML Information Set

**Wie sind die Daten in diesen Datenmodellen strukturiert?**

XML Information Set: hat Informationseinheiten für die Komponenten eines XML-Dokuments, nämlich **Elemente, Attribute, Text …**

* Für jede dieser Komponenten legt das Information Set fest, was sie für Information enthalten
  + Elemente: Namen, Kindknoten, Attributknoten, Elternknoten
  + Attribut: Namen und Wert, Element an dem das Attribut steht
  + Text: Textinhalt, Elternknoten
* bildet XML-Dokument ab

Daten in dem XDM Datenmodell sind Sequenzen von Referenzen auf Informationseinheiten im Information Set: Folgen von Konstrukten in einem XML-Dokument (Elementknoten, Attributknoten…) **Beziehung der Datenmodelle**

Die Komponenten in einer Sequenz in XDM sind (Referenzen auf) Knoten im Datenmodell Information Set. XDM wurde eingeführt, damit Queries und andere Operationen gegenüber dem Datenmodell abgeschlossen werden können.

**Welche Funktion haben diese Datenmodelle? Warum sind sie wichtig?**

„Damit man weiß worüber man redet“ Was sind die relevanten Komponenten eines XML-Dokuments unabhängig von der Syntax (Abstraktion der Syntax)

**Was ist die Beziehung des Baummodells zu diesen Datenmodellen?**

Baummodell ist graphische Abbildung des Information Sets, Diagramm in Baumform

**XML-Kodierung vs XML Information Set**

Nennen Sie Aspekte der Syntax von XML-Dokumenten, von denen ein XML Information Set abstrahiert.

**Lösung**

* Sind Attributwerte mit „“ oder mit ‚‘ eingerahmt?
* Weitere syntaktische Details die nicht abgebildet werden
  + Leerelemente-Tag <hr/> <hr> </hr>
  + Whitespace innerhalb von Tags <name x=“7“/> oder <name x = „7“ />
* Referenzen auf Entitäten werden aufgelöst.
* XML-Deklaration <?xml version = „1.1“ encoding = „utf-8“?>
* Reihenfolge der Attribute
* DTDs

**Prinzipien des XQuery-Entwurfs**

* Closure (erreicht über Datenmodell XDM)
* Compositionality (funktionale Ausdruckssprache, keine Seiteneffekte, Schachtelbarkeit)
* Completeness (berechenbarkeits-universell; Gegensatz: transitive Hüllen)
* Generality (mit und ohne Schema, text-orientiert und daten-orientiert, diverse Quellen für XML-Daten)