


XML-Dokumente und die Oracle-Datenbank



Carsten Czarski


Business Unit Database
Oracle Deutschland GmbH

ORACLE

 IT-Symposium 2004
- Experten im Dialog -


Warum XML in der Datenbank ...?

ORACLE

 IT-Symposium 2004
- Experten im Dialog -

... und nicht einfach im Dateisystem?


ORACLE

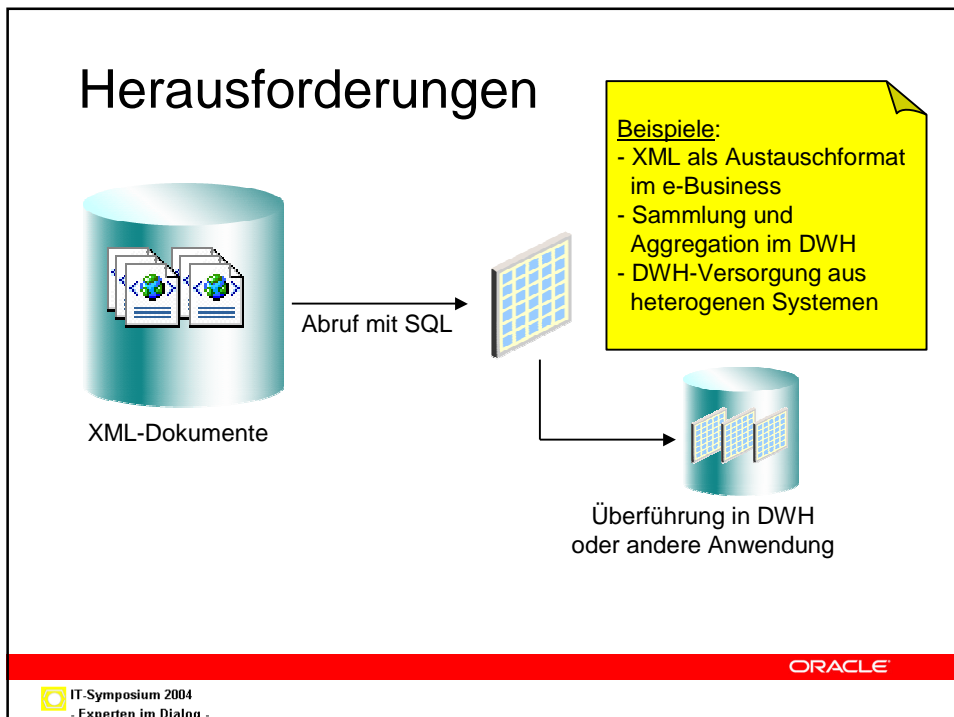
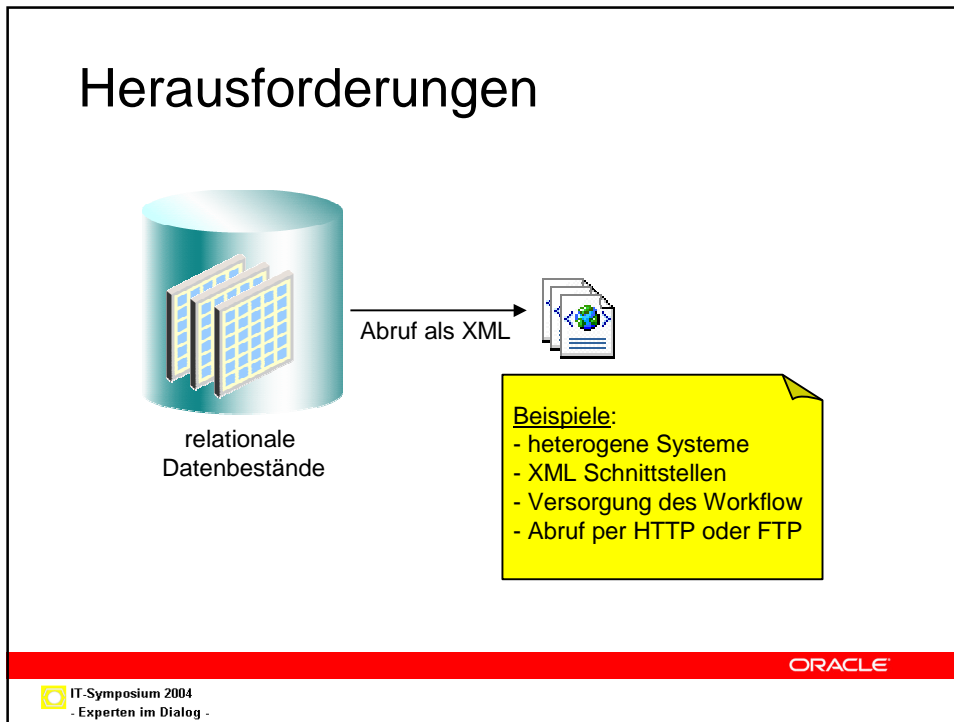
 IT-Symposium 2004
- Experten im Dialog -

Warum XML in der Datenbank und nicht im Dateisystem?

- Relationale Daten bereits vorhanden
 - Kombination mit XML Dokumenten
 - Ausgabe als XML-Dokument
- Erweiterung des XML-Datenmodells
 - Zusätzliche Datenbankvalidierung
 - Transaktionskonzept
 - Datennahe Sicherheitskonzepte
 - Datennahe Logik
 - ...

ORACLE

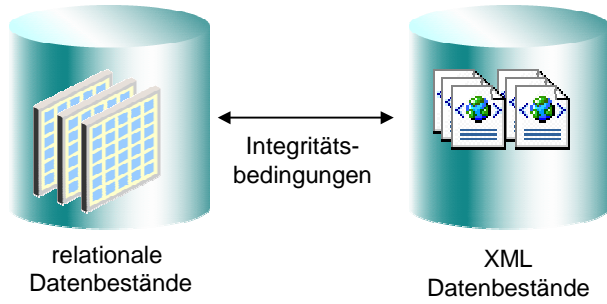
 IT-Symposium 2004
- Experten im Dialog -



Herausforderungen

Beispiele:

- XML Dokumente nativ speichern (jur. Gründe)
- Validitätsprüfung gegen Stammdaten in einer relationalen Datenbank
- Validierung im laufenden Workflow



Herausforderungen Referentielle Integrität in XML

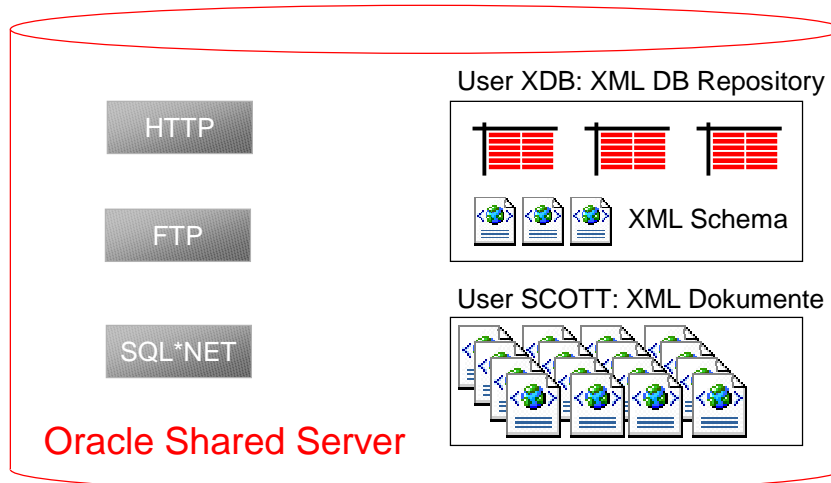
```

<?xml version="1.0" encoding="utf-8" ?>
<news:news xmlns:news="http://www.oracle.com/aktis/nachrichten.xml" type="Unternehmensne"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://
http://www.oracle.com/aktis/nachrichten.xml"/>
  <title xml:lang="de" id="831299" name="IBM" reuters="IBM" branche="Hard und Software" />
  <datum:2002-02-11T10:09:50 />
  <source:Finanzen.net />
  <contact:art="email" />
  <case:IBM />
  <address:bm@pubserve.com />
  <contact:art="email" />
  <case:Finanzen.net />
  <address:raaz.chiefredakteur@finanz.net />
  <contact: />
  <title:IBM: CEO erholt in 2002 Bonuszahlungen von 4,5 Mio. Dollar />
  <body>
    <p>Wie Samuel J. Palmisano, CEO der International Business Machines Corp., am gest.
    Borsenaufsicht SEC mitteilte, hat er in 2002 in seinem neuen Amt 45 Prozent mehr v.
    er als President und COO krieg mal.</p>
    <p>Palmisano, der im Marz 2002 nach dem Rucktritt von Louis V. Gerstner Jr. zum CEO
    Computerherstellers ernannt wurde, erholt demnach ein Grundgehalt von 1,4 Mio. $
    von 4,5 Mio. Dollar.</p>
    <p>Im 2001 belief sich das Basisgehalt von Palmisano auf lediglich 1,1 Mio. Dollar und
    Mio. Dollar.</p>
    <p>Gerstner, der sein Amt als Chairman im Januar niederlegte, erholt in 2002 ein Gru
    Bonuszahlungen in Hohe von 1,2 Mio. Dollar nach 2 Mio. Dollar bzw. 8 Mio. Dollar in 2
    <p>Die Aktie von IBM schloss gestern an der NYSE bei 74,76 Dollar (+0,62 Prozent).</p>
  </body>
</news:news>

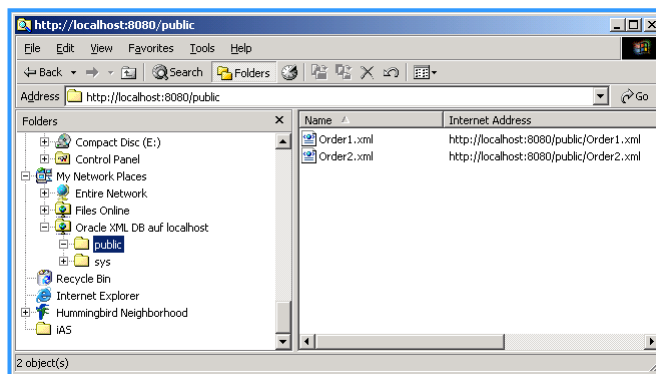
```

- XML-Datenmodell
- referentielle Integritat
 - ID
 - IDREF
 - IDREFS
- Dokumentbezogen

Oracle XML DB: Architektur



WebDAV



FTP

```

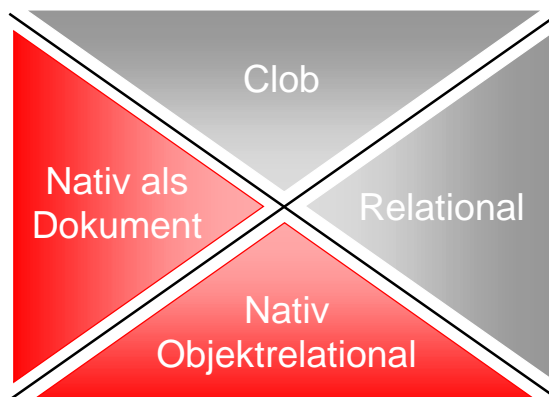
ftp> open localhost 9021
Connected to cczarski-de.de.oracle.com.
220 cczarski-de FTP Server (Oracle XML DB/Oracle9i Enterprise Edition
Release 9.2.0.4.0 - Production) ready.
ftp> user SCOTT
331 pass required for SCOTT
Password:
230 SCOTT logged in
ftp> dir
200 PORT Command successful
150 ASCII Data Connection
drw-r--r--  2 SYS      oracle      0 FEB 27 14:01 home
drw-r--r--  2 SYS      oracle      0 APR 07 09:17 public
drw-r--r--  2 SYS      oracle      0 MAR 23 14:55 sys
-rw-r--r--  1 SYS      oracle      0 DEC 11 17:49 xdbconfig.xml
226 ASCII Transfer Complete
ftp: 250 bytes received in 0,07Seconds 3,57kbytes/sec.
ftp> cd /public
250 CWD Command successful
ftp>

```

ORACLE

IT-Symposium 2004
- Experten im Dialog -

Unterschiedliche Speicherarten



Die Speicherart richtet sich nach den Anforderungen der Anwendung:

- Konkurrierender Zugriff
- Vokabularänderung
- Ladepformance
- Abfrage-Performance
- Redundanzen
- XML Schema Support

ORACLE

IT-Symposium 2004
- Experten im Dialog -

Dokumentorientiertes XML

- Dokument im Vordergrund
- Dokument-Transformationen
- Dokument als kleinste Einheit
 - Lesen und Schreiben als Ganzes
 - Dokumentteile nur selten
 - Skalare Daten nur selten
 - Häufig Volltextsuche
- XML als medienneutrales Speicherformat


ORACLE

 IT-Symposium 2004
- Experten im Dialog -

Datenorientiertes XML

- Daten im Vordergrund
- Dokumentkontext nicht unbedingt wichtig
- Häufig strukturierte Suche
 - größer-als und kleiner-als Abfragen
- XML als Austauschformat
 - Standardisierung der Werkzeuge
 - Integration in die IT-Landschaft

ORACLE

 IT-Symposium 2004
- Experten im Dialog -

Mischformen

- Beide Konzepte in einem Dokument
- Beispiele:
 - Zeitungsartikel mit strukturierten Zusatzinformationen
 - Autor, Zeitung, Erscheinungsdatum
 - ...
 - Kreditanträge
 - Strukturierte Kunden- und Kreditdaten
 - Unstrukturierte Begründungen

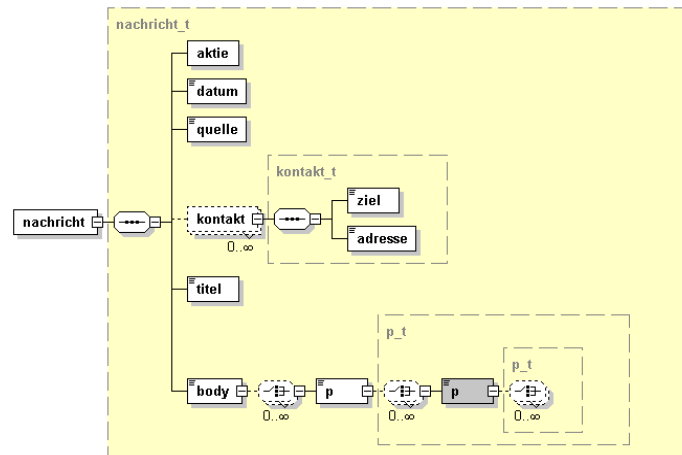
Beispiel: Börsennachrichten

```

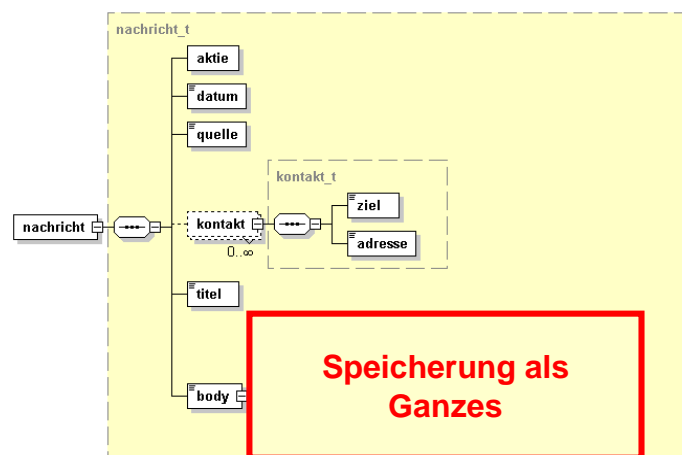
<?xml version="1.0" encoding="utf-8" ?>
- <nachricht xmlns="http://www.oracle.com/aktie/nachrichten.xsd" typ="Unternehmensmeldung"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.oracle.com/aktie/nachrichten.xsd
  http://www.oracle.com/aktie/nachrichten.xsd">
  <aktie wkn="888811" name="Oracle Corporation" reuters="ORCL" branche="Software" />
  <datum>2003-02-18T10:23:00</datum>
  <quelle>Finanzen.net</quelle>
- <kontakt art="email">
  <ziel>Oracle</ziel>
  <adresse>investor_us@oracle.com</adresse>
  </kontakt>
- <kontakt art="email">
  <ziel>Finanzen.net</ziel>
  <adresse>fred.redakteur@finanzen.net</adresse>
  </kontakt>
  <titel>Oracle kooperiert mit Nokia und Alcatel</titel>
- <body>
  <p>>Der Softwarekonzern Oracle meldete am Dienstag eine Kooperation mit den
  Telekommunikationsausrüstern Nokia (Finnland) und Alcatel (Frankreich).</p>
  <p>>Auf diese Weise will Oracle die Verbreitung eigener Software-Lösungen vorantreiben. Bislang
  hält der Softwarekonzern Microsoft das Geschäft rund um die Anwendungsgebiete der "Outlook
  Office Software" fest in der Hand. Microsoft hat bereits seit einiger Zeit damit begonnen,
  Kooperationen im Mobilfunkbereich zu schließen.</p>
  <p>>Im Zuge der Zusammenarbeit mit Nokia und Alcatel wird der finnische Konzern die Email-
  Software, das Voicemail-Modul sowie das Kalender- und File-Management-System auf seinen
  Terminals installieren. Mit Alcatel ist eine verstärkte Konzentration auf den
  Geschäftskundenbereich geplant.</p>
  <p>>Die Aktie von Oracle schloss gestern an der NASDAQ mit einem Gewinn von 1,39 Prozent bei
  11,70 Dollar. Das Papier von Nokia verliert in Amsterdam derzeit 2,03 Prozent auf 13,01 Euro. Der
  Anteilsschein von Alcatel gibt in Paris aktuell um 1,35 Prozent auf 7,30 Euro nach.</p>
  </body>
</nachricht>

```


Beispiel: Börsennachrichten Beschreibung als XML Schema

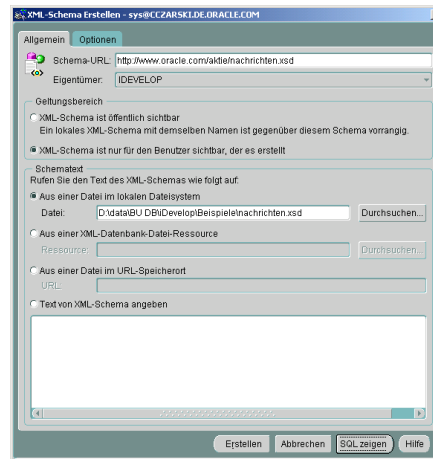


Beispiel: Börsennachrichten Beschreibung als XML Schema



Beispiel Börsennachricht Registrierung des XML Schemas

- Objektrelationale Speicherung
- Enterprise Manager oder PL/SQL
- Standard XML Schema (W3C)
- Feinsteuerung mit Annotations




Laden

- Unabhängig von der Speicherungsform
- FTP und WebDAV
 - XML Schema muss in XML-Dokument referenziert sein
- SQL und SQL*Loader

Schemavalidierung

- "Leichte" Schemaprüfung
 - Beim Einfügen
 - Strukturprüfung
 - Performancegründe
- Vollständige Schemavalidierung
 - Aktivierbar
 - RDBMS-Trigger

ORACLE


 IT-Symposium 2004
- Experten im Dialog -

Beispiel Börsennachrichten Zugriff mit HTTP

- Zugriff mit URL
 - Protokoll
 - Hostname und TCP/IP Port
 - Datenbank-Kürzel
 - Datenbankschema
 - Tabelle
 - Optional: XPath Einschränkung

[http://localhost:9080/oradb/<User-ID>/NR_NACHRICHT_TAB/ROW/nachricht/kontakt\[ziel='Finanzen.net'\]](http://localhost:9080/oradb/<User-ID>/NR_NACHRICHT_TAB/ROW/nachricht/kontakt[ziel='Finanzen.net'])

ORACLE

 IT-Symposium 2004
- Experten im Dialog -

Zugriff mit SQL

Gib mir von allen Oracle-Nachrichten
das Datum und den Titel!

```

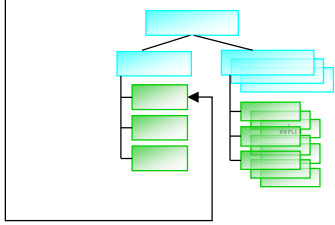
select
  extractvalue
    (value(e), '/nachricht/datum') as datum
  ,extractvalue
    (value(e), '/nachricht/aktie/@name') as ag
  ,extractvalue
    (value(e), '/nachricht/titel') as titel
from nr_nachricht_tab e
where existsnode
(
  value(e), '/nachricht/aktie[@wkn="871460"]'
) = 1
    
```

Ohne „Query Rewrite“

```

SELECT
  extractvalue ...
FROM nr_nachricht_tab e
    
```

DOM-Baum

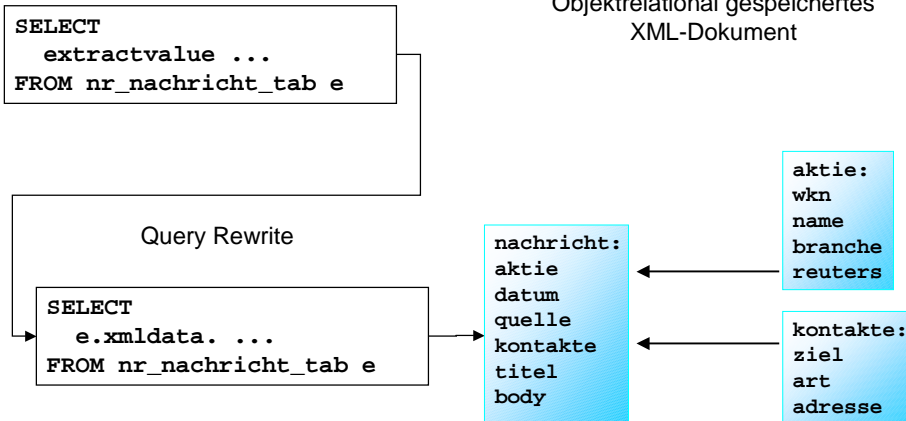


XML-Dokument

```

<?xml version="1.0" encoding="utf-8" ?>
<nachricht xmlns="http://www.oracle.com/aktie/nachrichten.xsd" type="Unternehmensne
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://
http://www.oracle.com/aktie/nachrichten.xsd">
<aktie wkn="871460" name="IBM" nae="IBM" branche="Hard und Software" />
<datum="2002-02-11T13:09:00"/datum>
<quelle="Finanzen.net"/quelle>
<kontakt art="email">
<ziel="IBM"/ziel>
<adresse="IBM@equiserve.com"/adresse>
</kontakt>
<kontakt art="email">
<ziel="Finanzen.net"/ziel>
<adresse="franz.chefredakteur@finanzen.net"/adresse>
</kontakt>
<titel="IBM: CEO erhielt in 2002 Bonuszahlungen von 4,5 Mio. Dollar"/titel>
<body>
<p>Wie Samuel J. Palmisano, CEO der International Business Machines Corp., am gest
Börsenaufsicht SEC mitteilte, hat er in 2002 in seinem neuen Amt 4,5 Prozent mehr v
er als Präsident und COO tätig war.</p>
<p>Palmisano, der im März 2002 nach dem Rücktritt von Louis V. Gerstner Jr. zum CEO
Computerherstellers ernannt wurde, erhielt demnach ein Grundgehalt von 1,4 Mio. £
von 4,5 Mio. Dollar.</p>
<p>In 2001 belief sich das Basisgehalt von Palmisano auf lediglich 1,1 Mio. Dollar und
Mio. Dollar.</p>
<p>Gerstner, der sein Amt als Chairman im Januar niederlegte, erhielt in 2002 ein Gru
Bonuszahlungen in Höhe von 1,5 Mio. Dollar nach 2 Mio. Dollar bzw. 8 Mio. Dollar in 2
<p>Die Aktie von IBM schloss gestern an der NYSE bei 75,70 Dollar (-2,82 Prozent).</p>
</body>
</nachricht>
    
```

Mit „Query Rewrite“



Piecewise Update

- XML Dokumentknoten gezielt ändern
- kein Ersetzen des Gesamtdokumentes
- nur auf SQL-Ebene
 - FTP und WebDAV sind dateiorientiert
- SQL-Funktion **updateXML()**
- Durch Query Rewrite direkte Änderung der Objektrelationalen Tabellenspalten

Piecewise Update

Ändere den Firmennamen bei den Nachrichten mit der WKN 871460 auf "Oracle Corporation"

```
update nr_nachricht_tab e
set value(e) = updatexml
(
  value(e),
  '/nachricht/aktie/@name'
  , 'Oracle Corporation'
)
where existsnode
(
  value(e)
  , '/nachricht/aktie[@wkn="871460"]'
) = 1
```

ORACLE

Volltextsuche mit Oracle TEXT

- Volltextindex auf jedes XML-Dokument möglich
- Oracle TEXT Index
 - Wildcard Suche (links und rechts)
 - Phrasensuche
 - Unschärfe Suche (Fuzzy)
 - Wortstamm
 - Stopwort-Listen
 - Thesaurus-Unterstützung

ORACLE

Volltextsuche mit Oracle TEXT

- Volltextrecherche ...

```
select score(1), value(e)
from nr_nachricht_tab e
where
  CONTAINS
  (
    value(e)
  , '
    (Software and Kooperation)
    INPATH
    (/nachricht/body)
  '
  ,1
  )>0
```

ORACLE

Stylesheet-Transformationen

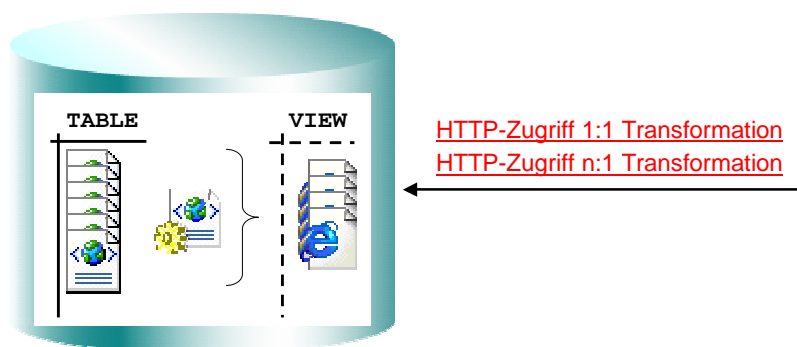
- 1:1-Transformation ...
- n:1 Transformation ...

```
select
  xmltransform
  (
    value(e)
  , xdburitype
  (
    '/public/idevelop/nachricht_html.xsl'
  ).getxml()
  )
from nr_nachricht_tab e
```

ORACLE

Stylesheet-Transformationen

- View-Definition



ORACLE

IT-Symposium 2004
- Experten im Dialog -

Beispiel Börsennachrichten Kontakt-View

```

create or replace view nr_kontakte
as
select
  extractvalue
    (value(e), '/nachricht/aktie/@wkn') as wkn
,extractvalue
    (value(e), '/nachricht/aktie/@name') as gesellschaft
,extractvalue
    (value(k), '/kontakt/@art') as kontaktart
,extractvalue
    (value(k), '/kontakt/ziel') as kontaktziel
,extractvalue
    (value(k), '/kontakt/adresse') as adresse
from
  nr_nachricht_tab e
,table(xmlsequence(extract(value(e), '/nachricht/kontakt'))) k

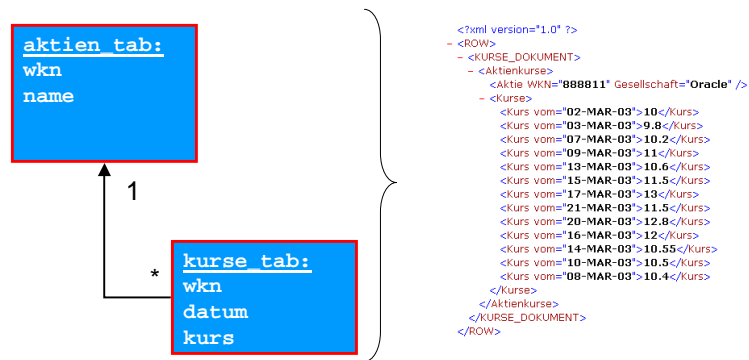
```

ORACLE

IT-Symposium 2004
- Experten im Dialog -

Beispiel Aktienkurse

Relationales Tabellenschema



Beispiel Aktienkurse

XML Views

- Erstellen der View

```

create or replace view aktienkurse_xml as
select
 XMLElement(
    "Aktienkurse",
    XMLElement("Aktie",
      XMLAttributes(
        wkn as "WKN", ak.name as "Gesellschaft")),
    XMLElement("Kurse", XMLAgg(
      XMLElement(
        "Kurs",XMLAttributes(
          k.datum as "vom",
          k.kurs))))
  )
as kurse_dokument
from aktien_tab ak join kurse_tab k using (wkn)
group by wkn, ak.name

```

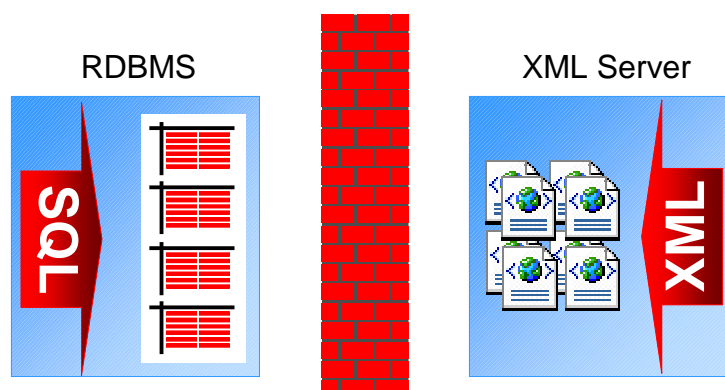
XML Views

- Nutzbar wie XMLType Tabellen
- Stylesheet-Transformationen
- Zugriff durch HTTP
- einfacher Austausch relational vorliegender Daten

ORACLE

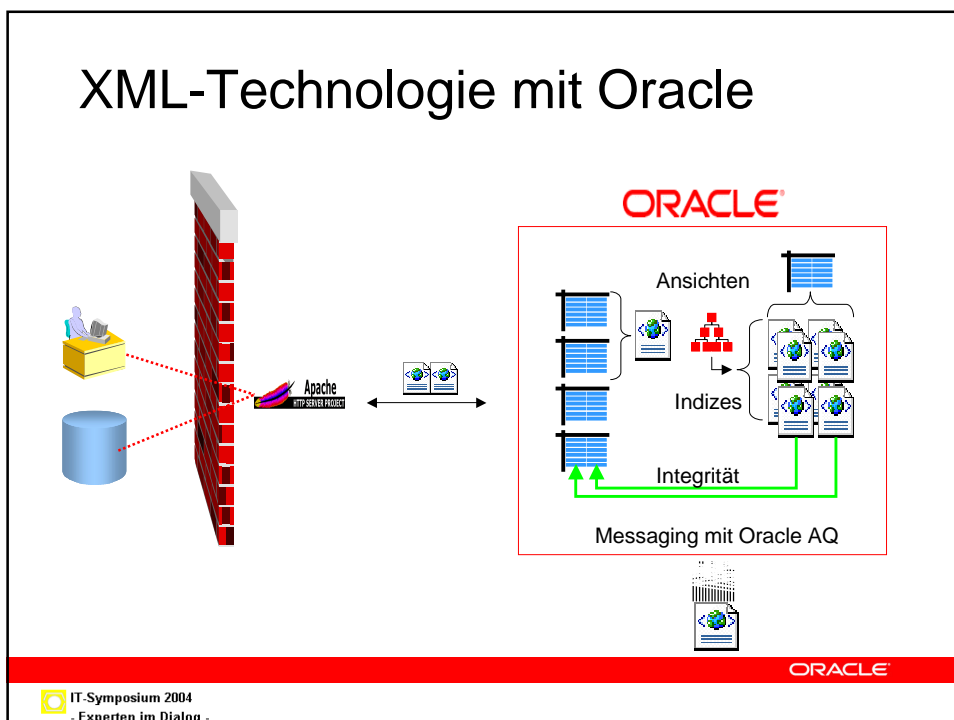
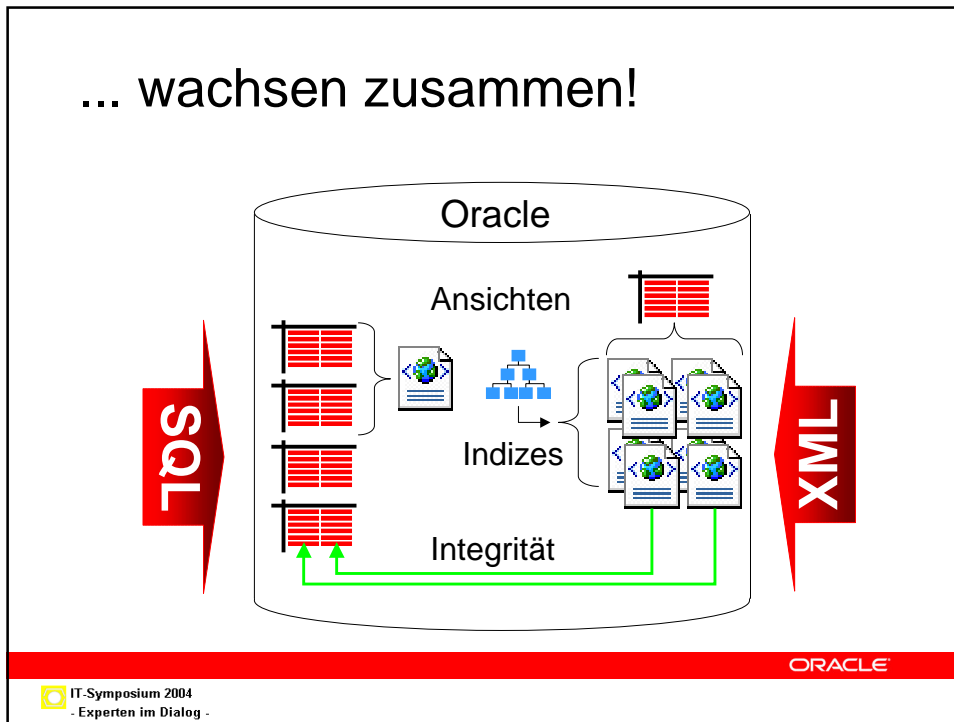
IT-Symposium 2004
- Experten im Dialog -

Zusammenfassung: Oracle XML DB: zwei Welten ...



ORACLE

IT-Symposium 2004
- Experten im Dialog -



Weitere Informationen

- Oracle Technology Network
 - <http://otn.oracle.com/tech/xml/xmlldb/content.html>
 - <http://otn.oracle.com/products/aq/index.html>
- Dokumentation
 - Oracle XML database Developers Guide
http://download-west.oracle.com/docs/cd/B10501_01/appdev.920/a96620/toc.htm ↗
 - Oracle Application Developer's Guide - Advanced Queuing
http://download-west.oracle.com/docs/cd/B10501_01/appdev.920/a96587/toc.htm ↗

ORACLE