

Die intelligente Datenbank - Jenseits von Relationen

Karin Patenge
Oracle Deutschland B.V. & Co. KG
Potsdam

Schlüsselworte

Datenbank, Objekt-relational, Oracle Extensibility Architecture, Objekt, Methoden, Multimedia, Spatial, DICOM¹, XML² DB

Einleitung

Die Zeit, in welcher wir heute leben, wird nicht selten als das Informationszeitalter bezeichnet.

Wir Menschen leben also in einer Informationsgesellschaft und nehmen tagtäglich daran teil, Informationen zu produzieren, zu konsumieren, auszutauschen, zu transformieren oder auch zu vernichten. Hilfestellung erhalten wir dabei durch entsprechende Informations- und Kommunikationstechnologien.

Informationen, insbesondere qualitativ hochwertige und konsistente, stellen den Schlüssel zum Erfolg einer jeden Unternehmung bzw. eines jeden Unternehmens dar. Es gilt, die richtigen Informationen zum richtigen Zeitpunkt in der richtigen Menge bedarfsgerecht und performant vorzuhalten, so dass sie für die richtigen Geschäftsprozesse genutzt werden können. Nur ein Bruchteil dieser Informationen lassen sich dabei in das Korsett von Relationen und Attributen pressen, welche seit Jahrzehnten die bevorzugte Technologie sind, Informationen strukturiert abzubilden und vorzuhalten. Dieses Korsett hat die Oracle Datenbank vor langer Zeit abgestreift.

Von relational zu objekt-relational

Mit der Version 8i zum Ende der 90er Jahre wurden erstmals objekt-relationale Erweiterungen, also objektorientierte Datentypen und Methoden, in den Kern der Datenbank integriert, um komplexe Datenstrukturen abzubilden. Was angefangen hat mit den Data Cartridges interMedia Text, Spatial und Visual Information Retrieval wurde seitdem beständig erweitert. Weitere Objekttypen sind hinzugekommen wie XMLTYPE, SDO_GEORASTER oder ORDDICOM ebenso wie Methoden für die Verarbeitung oder typgerechte Indexierung von Objekten.

So ist die Oracle Datenbank schon lange in der Lage, nicht mehr nur relationale Daten (über Datentypen wie NUMBER oder VARCHAR abgebildet) sondern darüberhinaus komplex strukturierte Daten effizient zu speichern, zu verarbeiten oder anzufragen.

Für derart komplexe Informationen sind eigene Datentypen definiert ebenso wie die für die Analyse und Anfragebearbeitung dieser Informationen notwendigen spezifischen Operatoren (Methoden). Die Datenbank „weiß“ somit um die Art, Struktur und Bedeutung dieser Information und wie diese mit anderen in Beziehung stehen. Die Grundlagen der Verarbeitung von Daten, wie sie aus relationaler

¹ Digital Imaging and Communication in Medicine

² Extensible Markup Language

Datenbanksystemen für relationale Daten gefordert sind (u.a. ACID³), gelten dabei auch für objekt-relational gespeicherte Daten. Der gemeinsame effiziente Zugriff sowohl auf relationale als auch komplexe Daten (Objekte) erfolgt dabei über einheitliche Schnittstellen wie die Standard-Abfragesprache SQL⁴, PL/SQL und Java.

Auf diese Weise müssen für Anwendungsgebiete mit besonderen Anforderungen, wie z.B. Geoinformationssysteme, nicht mehr gesonderte Datenhaltungssysteme installiert werden, die sich häufig nur schwer in eine immer mehr standardisierte IT-Landschaft integrieren. Die Oracle Datenbank übernimmt die Aufgabe, sämtliche Unternehmensinformationen integriert konsistent und auswertbar vorzuhalten. Mehr noch. Durch das Vorhandensein entsprechender objektspezifischer Methoden der Verarbeitung und Analyse bringt die Oracle Datenbank schon viel an Funktionalität mit, die sonst von spezialisierten Anwendungen bereitgestellt wird. Wesentlicher Vorteil ist dabei, dass die Verarbeitung und Analyse nah an den Daten, nämlich in der Datenbank selbst passiert. Das „Herausziehen“ der Daten aus der Datenbank, deren Verarbeitung oder Analyse von externen *Engines* und das Zurückschreiben in die Datenbank (oder ein Dateisystem) entfällt somit.

Darüberhinaus erlaubt die Entscheidung, Bilder, Dokumente und dergleichen in der Oracle Datenbank zu speichern auch, die gleichen Maßnahmen zur Hochverfügbarkeit oder Sicherheit der Informationen anzuwenden wie bei relationalen Daten auch. - und Sicherheitsaspekten zu tun,

Die Oracle Extensibility Architecture

Was im vorangegangenen Text konzeptionell recht einfach beschrieben wurde, hat seine Grundlage in der sogenannten Oracle Extensibility Architecture. Diese stellt ein erweiterbares Typsystem bereit ebenso wie eine entsprechende Serververarbeitungsumgebung, spezialisierte Indizes, nutzerdefinierte Operatoren und Aggregate, abstrakte Tabellen, einen erweiterten Optimizer und natürlich auch die notwendigen Basisservices wie das Speichermanagement oder die Fehlerbenachrichtigung, um hier nur zwei dieser Services anzuführen.

Die Vielfalt der Informationen - Oracle Multimedia, XML DB & Co.

Gerade die immer rasantere Verbreitung der Nutzung des Internets zielen auf die Bereitstellung und den Austausch von Inhalten wie Bildern, Videos, Audiostreams, Karten, Dokumenten oder DNA Sequenzen, um nur einige Beispiel anzuführen. Für all diese Inhalte oder Informationen gibt es bereits Entsprechungen in Bezug auf Datentyp, Indexierung(en) und Methoden in der Oracle Datenbank.

Die nachfolgenden Komponenten stellen das technologische Grundgerüst für das Informations- und Wissensmanagement in der Oracle Datenbank:

- Mit Oracle Text werden unstrukturierte Informationen wie Textdokumente in der Datenbank verwaltet. Das beinhaltet eine leistungsfähige Volltextrecherche mit Mitteln der Standardabfragesprache SQL.
- Oracle Multimedia erweitert die Möglichkeiten von Oracle Text im Hinblick auf den Umgang mit unstrukturierten, aber nicht-textuellen sondern binären Informationen wie Bildern, Audios oder Videos.
- Oracle XML DB kommt zum Einsatz, wenn es um die Verwaltung, Verarbeitung und den Zugriff auf XML Dokumente geht. Sie bietet drei verschiedene Speichervarianten für XML, ein integriertes XML DB Repository mit Protokollservern für HTTP(s), WebDAV und FTP sowie die Unterstützung von Standards wie SQL/XML, XQuery, XPath, XML Schema.

³ Atomicity, Consistency, Isolation und Durability

⁴ Structured Query Language

- Oracle Spatial als substantielle Erweiterung der Oracle Locator Funktionalität kümmert sich um das Management räumlicher Informationen in Form von z.B. 2- oder 3-dimensionalen Vektordaten, Rasterdaten oder Netzwerkdatenmodellen. Über das darin integrierte Netzwerkdatenmodell ermöglicht die Spatial RDF Engine, dass auch „Wissen“ in Form von Ontologien (Semantische Netze) abgebildet werden kann und dieses ebenso in Auswertungen einbezogen wird.

Oracle Multimedia hat vier Objekt-Datentypen implementiert, die ihrerseits den gemeinsamen Objekt-Datentyp ORDSOURCE enthalten. ORDIMAGE wird für Bildinformationen, ORDAUDIO für Audio-Informationen und ORDVIDEO für Video-Informationen verwendet. Lassen sich diese multimedialen Inhalte organisatorisch nicht trennen voneinander (alles muss rein und zwar in eine Tabellenspalte), dann kann der Datentyp ORDDOC dafür herangezogen werden. Darüberhinaus steht für Bilder, welche im Rahmen von bildgebenden Verfahren der Medizin erzeugt werden, der Datentyp ORDDICOM zur Verfügung. Zu den Objekten werden spezifische Metadaten vorgehalten, wie z.B. Höhe, Breite und Farbattribute von Bildern, die Länge von Video- oder Audiosequenzen oder im Falle von medizinischen Bildern ganze XML Dokumente mit zusätzlichen Informationen wie Geräte- oder Patientenname. Methoden zum Umgang mit diesen Datentypen sind zum Beispiel das Drehen, Skalieren, Spiegeln von Bildern, die Erstellung von Vorschaubildern (sogenannter *Thumbnails*), das Anonymisieren von personenbezogenen Daten in den DICOM-Metadaten (siehe Listing 1) bzw. generell das Extrahieren von Informationen aus den Metadaten wie zum Beispiel die Koordinaten (Länge- und Breitengrad) des Aufnahmeortes.

```
SELECT
  column_value
FROM
  TABLE (ORDIMAGE.GETMETADATA (GET_IMAGE_BLOB (), 'ALL')); -- Alternativ 'EXIF'
```

Listing 1: Metadatenextraktion aus ORDIMAGE

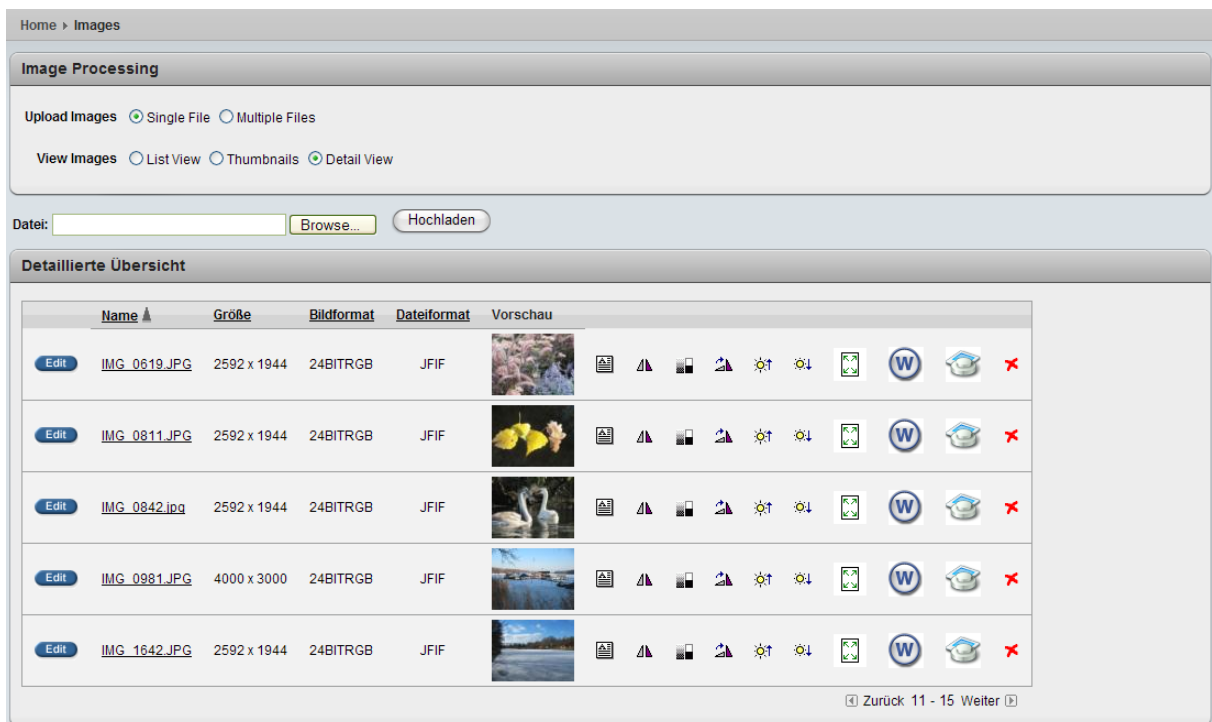


Abb. 1: Einfache APEX-Bildarchivanwendung mit symbolisierten typischen Bildfunktionen

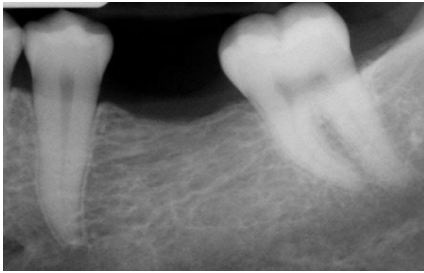


Abb. 2: DICOM-Bild

```
<?xml version="1.0"?>
<DICOM_OBJECT xmlns="http://xmlns.oracle.com/ord/dicom/metadata
<OTHER_BYTE tag="00020001" definer="DICOM" name="File Meta I

</OTHER_BYTE>
<UNIQUE_ID tag="00020002" definer="DICOM" name="Media Storag
<UNIQUE_ID tag="00020003" definer="DICOM" name="Media Storag
<UNIQUE_ID tag="00020010" definer="DICOM" name="Transfer Syn
<UNIQUE_ID tag="00020012" definer="DICOM" name="Implementat
<SHORT_STRING tag="00020013" definer="DICOM" name="Impleme
<APPLICATION_ENTITY tag="00020016" definer="DICOM" name="S
<CODE_STRING tag="00080008" definer="DICOM" name="Image Ty
<CODE_STRING tag="00080008" definer="DICOM" name="Image Ty
```

Abb. 3: DICOM Metadaten

XML hat sich zu einem universellen Speicherungs- und Austauschformat für strukturierte und semi-strukturierte Daten entwickelt, welches plattform- und herstellerunabhängig ist. Für die Speicherung, Verwaltung und Verarbeitung von XML Dokumenten wurde die Datenbank um die Funktionalität XML DB erweitert. Oracle XML DB stellt den Datentyp XMLTYPE und spezielle Funktionen zum Erzeugen, Extrahieren und Indizieren von XML Dokumenten bereit. So ermöglicht die XML DB, dass aus Datenbanktabellen XML erzeugt werden kann und umgekehrt mittels SQL auf XML Dokumente bzw. Teile dieser zugegriffen werden kann (siehe Listing 2). Jedes XML-Dokument läßt sich dabei als Instanz eines Objektes auffassen. Typische XMLTYPE-Funktionen, die neben weiteren zur Verfügung stehen, sind:

- ExtractValue, welches den Wert eines Knotens anhand eines XPath-Ausdrucks zurückgibt,
- ExistsNode, dass überprüft, ob ein bestimmter Knoten in einem XML-Dokument vorhanden ist oder
- Transform zur Umwandlung eines XML Dokuments in ein anderes Format (z.B. HTML⁵) unter Verwendung einer XSLT⁶-Datei.

```
SELECT
  XMLElement (
    "Department",
    XMLAttributes(d.department_id AS "DepartmentId"),
    XMLForest(d.department_name AS "Name"),
    XMLElement (
      "Location",
      XMLForest (
        street_address AS "Address",
        city AS "City",
        state_province AS "State",
        postal_code AS "Zip",
        country_name AS "Country")),
    XMLElement("EmployeeList", (
      SELECT
        XMLAgg (
          XMLElement (
            "Employee",
```

⁵ Hypertext Markup Language

⁶ Extensible Stylesheet Language Transformation

```

XMLAttributes (e.employee_id AS "employeeNumber" ),
XMLForest(
  e.first_name AS "First Name",
  e.last_name AS "Last Name",
  e.email AS "Email Address",
  e.phone_number AS "Phone Number",
  e.hire_date AS "Start Date",
  j.job_title AS "Job Title",
  e.salary AS "Salary",
  m.first_name || ' ' || m.last_name AS "Manager"),
XMLElement("Commission", e.commission_pct)))
FROM
  hr.employees e,
  hr.employees m,
  hr.jobs j
WHERE
  e.department_id = d.department_id
  AND j.job_id = e.job_id
  AND m.employee_id = e.manager_id))) .extract('/*') AS XML
FROM
  hr.departments d,
  hr.countries c,
  hr.locations l
WHERE
  d.location_id = l.location_id
  AND l.country_id = c.country_id;

```

Listing 2: Extrahieren relationaler Daten aus Schema HR und Darstellung als XML

Über den Datentyp SDO_GEOMETRY können ortbezogene 2D, 3D oder 4D Vektordaten standardkonform in der Datenbank vorgehalten werden. Mit Hilfe dieser Informationen lassen sich dann z.B. thematische Karten generieren.

```

with mittelpunkt as (
  select
    sdo_geom.sdo_centroid(
      geometry,          -- Geometriespalte von GEO_STAATEN
      1                  -- Toleranz: 1m
    ) geometry
  from de_admin_areal
  where feature_name = 'GERMANY'
)
select
  -- Extraktion der Koordinaten aus der "Punkt"-Geometrie
  e.geometry.sdo_point.x laengengrad, -- Längengrad
  e.geometry.sdo_point.y breitengrad, -- Breitengrad
  e.geometry.sdo_srid "CS"           -- Koordinatensystem
from mittelpunkt e;

```

Listing 3: Bestimmen des Mittelpunktes einer Geometrie (hier Deutschland) und Ausgabe der Koordinaten

Zusammenfassung

Die Oracle Datenbank hat die heutigen Herausforderung, vielfältige Informationen in beliebig großen Mengen zu integrieren und Unternehmen verfügbar zu machen, angenommen und stellt umfangreiche Funktionalität dafür bereit.

Links weiterführende Dokumentation

Oracle Dokumentation: Object-Relational Developer's Guide

http://download.oracle.com/docs/cd/E11882_01/appdev.112/e11822.pdf

Oracle Text Overview

<http://www.oracle.com/technetwork/database/enterprise-edition/index-098492.html>

Oracle Multimedia Overview

<http://www.oracle.com/technetwork/database/multimedia/overview/index.html>

Oracle XML DB Overview

<http://www.oracle.com/technetwork/database/features/xmldb/index-093086.html>

Introduction to Semantic Technologies

http://download.oracle.com/docs/cd/B28359_01/appdev.111/b28397/sdo_rdf_concepts.htm#CIHECGI

[I](#)

Einfache Bildbearbeitung mit Oracle Application Express

<http://www.oracle.com/global/de/community/tipps/ordimage/index.html>

Bilder manipulieren mit Oracle Application Express

<http://www.oracle.com/global/de/community/tipps/textintoimage/index.html>

Oracle Spatial Blog

<http://oracle-spatial.blogspot.com/>

Kontaktadresse:

Karin Patenge

Oracle Deutschland B.V. & Co. KG

Schiffbauergasse 14

D-14467 Potsdam

Telefon: +49 (0) 331-200 7214

Fax: +49 (0) 331-200 7559

E-Mail karin.patenge@oracle.com

Internet: www.oracle.com