

DOAG

Deutsche ORACLE-Anwendergruppe e.V.

News

Der lange Arm von Big Data



Best Practice

Mobil integrierte Geschäftsprozesse mit ADF Mobile

Im Interview

Roland M. Dürre,
Vorsitzender des Vorstands der Interface AG



Oracle Hidden Secrets

Netzwerk-Verschlüsselung ohne jeden Aufwand

Trends & Entwicklungen, Erfahrungsaustausch, Tipps & Tricks

- *OWB to ODI Transition*
- *Innovative Datenmodellierung für das DWH*
- *Big Data wird erwachsen...*
- *BI & DWH in the Cloud*

Technologie, Methodik und fachliche Hintergründe



FRÜHBUCHER
BIS 28. FEBRUAR 2014

DOAG 2014 **Business Intelligence**

3. April 2014, Hilton München City Hotel

Community Konferenz für das BI- und DWH-Umfeld

Sponsoren:





Robert Szilinski
Leiter der Development
Community

Liebe Mitglieder der DOAG Deutsche ORACLE-Anwendergruppe,
liebe Leserinnen und Leser,

aktuell geht man davon aus, dass sich das weltweite Datenvolumen alle zwei Jahre verdoppelt. Der Großteil der neu hinzukommenden Daten ist unstrukturiert, was vor allem am Siegeszug der Smartphones, dem Internet der Dinge und am Wachstum der sozialen Netzwerke liegt. Aber auch in den Unternehmen fallen inzwischen gigantische Datenmengen an und der Hunger nach immer mehr aktuellen Informationen und Analysen, die den Umsatz steigern sollen, wird permanent größer. Nicht zuletzt nutzen auch die Geheimdienste alle Mittel, um sämtliche verfügbaren Daten zu kombinieren und für sich auszuwerten.

Haben Sie sich schon mal gefragt, wie Google es schafft, auf Ihre Fragen innerhalb von wenigen Millisekunden Antworten zu finden? Wie es sein kann, dass Amazon fast in Echtzeit Produkte im Rahmen Ihres Einkaufs empfiehlt? Die Antwort ist natürlich „Big Data“ und ein anschauliches Beispiel aus dem 19. Jahrhundert verdeutlicht meiner Meinung nach sehr schön, welche Bedeutung die Auswertung gigantischer Datenmengen in kürzester Zeit für uns alle hat. 1880 dauerte die Volkszählung von 50 Millionen Einwohnern in den USA rund 7 Jahre – per Hand. Der Einsatz von Lochkarten stellte zehn Jahre später die Auswertung bereits innerhalb von sechs Wochen bereit. In ähnlichen Dimensionen bewegen wir uns heute, wenn wir das Potenzial von Big Data in unseren immer schnelleren Alltag übertragen. Mit In-Memory-Datenbanken, Column-Stores und Map-Reduce-Konzepten stehen technologische Treiber zur Verfügung.

Das Zusammenspiel von strukturierten mit unstrukturierten Daten (NoSQL) ermöglicht es Unternehmen mittlerweile, in Echtzeit Geschäftsanalysen durchzuführen und auch bisher nicht wahrgenommene Beziehungen zwischen den Daten aufzudecken. Entsprechend groß sind das Potenzial, das Big Data mit sich bringt, sowie die Auswirkungen, die es auf die klassische Oracle-Welt haben wird.

In diesem Sinne wünsche ich viel Spaß beim Lesen dieser Ausgabe.

Ihr R. Szilinski

ORACLE Platinum
Partner

HUNKLER
GmbH & Co. KG



„ Sicher, schnell, kompakt verpackt:
Das Datenbanksystem
für Ihre Höchstleistung “

Aktuelle Infos und Aktionen
zu Oracle unter
www.oracle-on-oracle.de

LIZENZBERATUNG &
-VERTRIEB



HOCHVERFÜGBAR-
KEITSLÖSUNGEN &
PERFORMANCE
TUNING



DATA WAREHOUSING &
BUSINESS
INTELLIGENCE
LÖSUNGEN



ORACLE
APPLIANCES



Oracle Database Appliance: perfekt abgestimmt – bis zum letzten Treiber

Server, Storage, Software, Netzwerk: Das ist die Oracle Database Appliance (ODA). Das Datenbank-Komplettsystem ist vorkonfiguriert und bis ins Detail perfekt abgestimmt. Keine aufwändige Verkabelung, keine fehleranfällige Anpassung. Dafür maximale Zuverlässigkeit und Ausfallsicherheit – sofort betriebsbereit!

ODA heißt: Pay as you grow. Sie schalten erstmal nur die Leistung frei, die Sie brauchen, und kaufen die entsprechenden Lizenzen. Dann wächst die ODA mit Ihren Anforderungen mit: Sie erwerben einfach weitere Lizenzen und aktivieren die zusätzliche Power per Knopfdruck – installieren müssen Sie dafür nichts!

Oracle Database Appliance: Das clevere System aus einer Hand. Fragen Sie uns!

Hauptsitz Karlsruhe
Bannwaldallee 32, 76185 Karlsruhe
Tel. 0721-490 16-0, Fax 0721-490 16-29

Geschäftsstelle Bodensee
Fritz-Reichle-Ring 6a, 78315 Radolfzell
Tel. 07732-939 14-00, Fax 07732-939 14-04

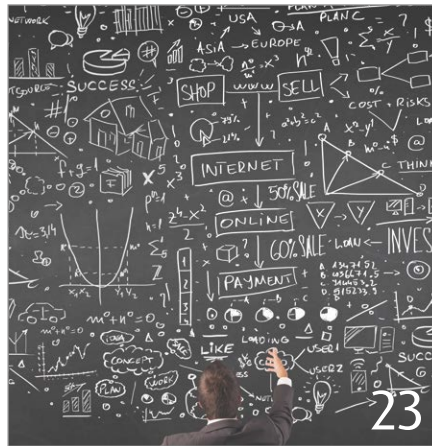
info@hunkler.de, www.hunkler.de

ODA AUF EINEN BLICK

- Komplettlösung: Datenbanksoftware, Server, Storage, Netzwerk in einer Box
- Vollständig vorkonfiguriert und perfekt abgestimmt
- Auf Knopfdruck betriebsbereit
- Wächst mit: für mehr Leistung einfach weitere Lizenzen erwerben und aktivieren



Interview mit Roland M. Dürre, Vorsitzender des Vorstands der Interface AG, Seite 8



Das Berufsbild „Data Scientist“ und dessen Aufgaben, Seite 23



Mobil integrierte Geschäftsprozesse mit ADF Mobile, Seite 39

Einleitung

- 3 Editorial
- 5 Spotlight
- 6 Die DOAG in der Presse
- 8 „Da findet eine Evolution statt, die kaum vorhersehbar ist...“
Interview mit Roland M. Dürre, Vorsitzender des Vorstands der Interface AG

Big Data

- 10 SQL und Hadoop – eine gemeinsame Zukunft
Jean-Pierre Dijcks
- 15 Data-Mining in sozialen Online-Netzwerken
Bianca Böckelmann
- 19 Customer Analytics: Neues Buzzword oder effektives Steuerungsinstrument?
Marco Poloni
- 23 Das Berufsbild „Data Scientist“
Angi Voss
- 25 Ereignissen auf der Spur: Big-Data-Analysen in der Datenbank
Alfred Schlaucher

Security

- 34 Der DBA mit einem Fuß im Gefängnis?
Oliver Pyka

Buchrezensionen

- 60 Oracle Application Express in der Praxis
gelesen von Niels de Bruijn
- 61 SQL Thinking
gelesen von Andreas Schmidt

Best Practice

- 35 Eine Einführung in Business Transaction Management
Arne Brüning
- 39 Mobil integrierte Geschäftsprozesse mit ADF Mobile
Enno Schulte
- 44 Web-Services in der Oracle-Datenbank
Detlev Schütte
- 50 Eine Alternative zu Database Change Notification bei Massendaten-Änderungen
Michael Griesser und Thies Rubarth
- 55 Quasi-Modale Dialoge in Apex
Frank Weyher

Tipps & Tricks

- 57 Oracle Hidden Secrets: Netzwerk-Verschlüsselung ohne jeden Aufwand
Heinz-Wilhelm Fabry
- 59 Tipps und Tricks aus Gerds Fundgrube Heute: Zeituntersuchungen im Mikrosekundenbereich
Gerd Volberg

DOAG intern

- 61 Neue Mitglieder
- 62 #DOAG 2013 - Twitterstrecke zur Anwenderkonferenz
- 64 DOAG-Botschafter 2013
- 64 Aus dem Verein
- 65 Impressum
- 66 Termine
- 66 Inserentenverzeichnis



Montag, 4. November 2013

Cornel Albert aus der DOAG Geschäftsstelle klärt mit dem Catering-Unternehmen Lehrrieder die letzten Details für den Bewirtungsservice auf der DOAG 2013 Konferenz + Ausstellung in Nürnberg. Am Community-Abend wird es wieder verschiedene Themenbuffets mit Spezialitäten aus allen Teilen der Welt geben.

Montag, 18. November 2013

Das Team der DOAG-Geschäftsstelle ist mit dem Aufbau der DOAG 2013 Konferenz + Ausstellung beschäftigt. Abends bei der Abnahme durch die Konferenzleitung ist alles für die große Veranstaltung vorbereitet.

Dienstag, 19. November 2013

Die DOAG 2013 Konferenz + Ausstellung verzeichnet gleich am ersten Tag eine Rekordbeteiligung. Noch nie zuvor waren bei einer DOAG-Anwenderkonferenz 1.700 Besucher gleichzeitig anwesend. Damit ist die Teilnehmerzahl im Vergleich zum Vorjahr um zehn Prozent gestiegen.

Mittwoch, 20. November 2013

Im Anschluss an seine interessante Keynote zum Thema „Anstiftung zum Querdenken“ diskutiert der Buchautor Dr. Peter Kreuz mit Prof. Dieter Kempf, Präsident der BITKOM und Vorstandsvorsitzender der DATEV eG; Reiner Ernst, Sales Director ORACLE Deutschland B.V. & Co. KG; Roland Dürre, Vorstandsvorsitzender der Interface AG und DOAG-Vorstand Urban Lankes über unterschiedliche Aspekte des Querdenkens.

Donnerstag, 21. November 2013

Die DOAG führt eine Online-Befragung bei rund 200 Unternehmen durch, die in Deutschland ihr Data Warehouse auf Basis der Oracle-Database-Infrastruktur betreiben. Sie belegt, dass die Anwender den Warehouse Builder auch über das Release 1 der Datenbank 12c hinaus behalten möchten. Oracle muss hier dringend handeln, immerhin sind fast drei Viertel der Befragten bereit, alternative Anbieter zu eruieren.

Freitag, 22. November 2013

Die Mitgliederversammlung des Interessenverbands der Java User Groups e.V. (iJUG) bestätigt den Vorstandsvorsitzenden Fried Saacke sowie seine beiden Stellvertreter Tobias Frech und Stefan Koospal für weitere zwei Jahre in ihrem Amt.

Dienstag, 3. Dezember 2013

Die Business Solutions Community diskutiert das Konzept der DOAG 2014 Applications und startet eine Umfrage, um die Veranstaltung noch attraktiver für die Teilnehmer zu machen.

Freitag, 6. Dezember 2013

Die Vorstandssitzung der DOAG stellt auf Basis der Beschlüsse der Delegiertenversammlung die Weichen für die Aktivitäten im Jahr 2014.

Montag, 9. Dezember 2013

Fried Saacke, DOAG-Vorstand und Geschäftsführer, bespricht mit Josef Opitz, Ansprechpartner der DOAG für Hochschulen, die Zukunft des Programms für Studierende.

Mittwoch, 1. Januar 2014

Die DOAG wünscht allen Mitgliedern und Interessenten ein gesundes und erfolgreiches neues Jahr!



Die DOAG in der Presse

Die nachfolgenden Ausschnitte reflektieren die Einschätzung der Fach- und Wirtschaftspresse zu bestimmten Themen über Oracle; die Veröffentlichungen geben nicht die Meinung der DOAG wieder und sind auch nicht im Vorfeld mit der DOAG abgestimmt. Lediglich die Zitate einzelner DOAG-Vorstände geben die Meinung der DOAG wieder.

Infomarkt vom 1. November 2013

An der Konferenz „Applications“ der Deutschen Oracle Anwendergruppe (DOAG) nahmen vom 9. bis 11. Oktober 2013 rund 280 Personen teil. Dabei drehte sich vieles um die Zukunftsperspektiven von Oracle im deutschsprachigen Raum.

Nach den Untersuchungen von Professor Norbert Gronau von der Universität Potsdam hat Oracle verglichen mit SAP und Microsoft deutlichen Nachholbedarf in puncto Sichtbarkeit auf dem Markt für Business Applications. Es fehlten vor allem lokale Marketingstrategien sowie Referenzen über alle Produktlinien hinweg. Dietmar Neugebauer, Vorstandsvorsitzender der DOAG, sagte: „Es kommen große Herausforderungen sowohl auf die Kunden als auch auf den Hersteller zu. Der Einsatz der Fusion Applications spielt insbesondere auf dem deutschsprachigen Markt derzeit noch keine bedeutende Rolle.“

silicon.de vom 21. November 2013

Deutsche Oracle-Anwender nutzen für Ihre ETL-Prozesse zum überwiegenden Teil den Oracle Warehouse Builder. Jetzt will Oracle mit einem neuen Datenbank-Release OWB nicht mehr zertifizieren.

Die Deutsche Oracle Anwendergruppe DOAG meldet sich mit den Ergebnissen einer Befragung zu 'Oracle Warehouse Builder' zu Wort. Die DOAG

reagiert damit auf die Pläne Oracles, mit dem Release 1 der Datenbank 12c, OWB nicht mehr zu zertifizieren.

Rund 200 Nutzer, die ein Data Warehouse auf Basis der Oracle-Datenbank-Infrastruktur betreiben, nahmen an der Umfrage teil. 95 Prozent, die Extract, Transform, Load (ETL) auf Oracle betreiben, setzen dabei auf OWB. Mit ETL werden Daten aus verschiedenen Datenquellen in ein einziges Data Warehouse zusammengefasst.

Bei den Anwendern stehen häufig hohe Investitionen in Schulung, Aufbau von Bibliotheken und andere Bereiche an. Wie die DOAG mitteilt, sind die Basic-ETL-Komponenten des OWB Teil der kostenpflichtigen Datenbank-Editionen und sind somit kostenfrei nutzbar, wenn ein Anwender eine Datenbank lizenziert hat. Laut der DOAG-Umfrage nutzen 87 Prozent der Befragten diese Basic-ETL-Komponenten.

Oracle will laut dem Statement of Direction zum OWB und Oracle Data Integrator (ODI) den OWB nach dem Release 1 der Datenbank 12c nicht mehr zertifizieren. Stattdessen sollen Unternehmen künftig ODI einsetzen. Dafür ist jedoch ein Umschulungsaufwand erforderlich, den rund 80 Prozent der Befragten mit „hoch“ oder „sehr hoch“ einschätzen.

Darüber hinaus gehen 89 Prozent der Unternehmen von einem „hohen“ oder „sehr hohen“ Migrationsaufwand aus. Zudem ist die Oracle Data Integrator Enterprise Edition ein kostenpflichtiges Produkt, das zusätzlich zur Datenbank lizenziert werden muss.

„Mit dieser Umfrage können wir nun Oracle mit konkreten Zahlen belegen, dass die Anwender den Warehouse Builder auch über das Release 1 der Datenbank 12c hinaus behalten möchten“, so Dr. Dietmar Neugebauer, Vorstandsvorsitzender der DOAG.

„Die Ergebnisse der Umfrage zeigen, dass wir die richtige Zielgruppe adressiert haben, und dass das aktuelle Statement of Direction von Oracle für diese Zielgruppe einen hohen Leidensdruck bedeutet“, ergänzt Christian Weinberger, Themenverantwortlicher für BI & Data Warehouse zusammen. Oracle müsse hier dringend handeln, denn knapp 75 Prozent der Befragten hätten laut Weinberger auch erklärt in diesem Fall auch die Produkte anderer Hersteller evaluieren zu wollen. Daher sieht die DOAG die Möglichkeit einer Migrationswelle zu anderen Herstellern.

Dennoch will man auch auf Seiten der DOAG dem Hersteller entgegenkommen: „Eine kostenfreie Light-Version des ODI im Rahmen der Datenbank-Lizenz, die die Funktionalitäten der bisherigen Basic-ETL-Komponenten enthält, könnte zumindest eine längerfristige Alternative für die Benutzer des OWB sein“, schlägt Michael Paege, stellvertretender Vorstandsvorsitzender der DOAG und Leiter Competence Center Lizenzierung vor. Auf diese Weise könne Oracle bei den Anwendern die Investitionen schützen.

Immerhin scheint auch Oracle das Ansinnen der Anwender zumindest gehört zu haben. So signalisiert Günther Stürner, Vice President Server Technologies und Sales Consulting bei der ORACLE Deutschland B.V. & Co. KG, Verhandlungsbereitschaft, wenn er erklärt: „Wir werden das Problem gemeinsam mit der DOAG besprechen und eine gute Lösung für die betroffenen Anwender anstreben.“

datacenter-insider vom 26. November 2013

Die in der Deutschen Oracle Anwendergruppe (DOAG) organisierten Oracle-Nutzer zeigen ihrem Lieferanten die

rote Karte - die Lizenzpolitik passt nicht.

Nachdem kürzlich die deutschen SAP-Anwender in Nürnberg tagten, versammelte sich jetzt auch die Oracle-Gemeinde in der Lebkuchenstadt. Die Doag meldet ein Rekordergebnis, was die Teilnehmerzahlen angeht: Schon am ersten Tag kamen knapp 1.700 Besucher – so viel wie nie zuvor. Sie hatten unter anderem Ärger im Gepäck.

Das neue Datenbank-Release „12c“ stand naturgemäß im Mittelpunkt des Tagungsprogramms aber auch die angekündigte Verfügbarkeit der In-Memory-Datenbank. Damit versucht Oracle mit den Konkurrenten aus dem Hadoop-Lager gleichzuziehen.

Ein weiterer Schwerpunkt bildete die Vortragsreihe über die „Soft Skills“, die erstmals ins Programm aufgenommen wurde: Auch der Oracle-Technologie steht im Wandel und sieht sich mit Herausforderungen aus dem Business-Bereich konfrontiert. Project-Coaching und Wissenstransfer sind neue Aufgaben.

Dazu passte die Keynote des Bestseller-Autors Peter Kreuz, der seit Jahren zum „Querdenken“ anstiften will. Und auch die „Doag Unconference“ geht unübliche Wege. Diese Plattform legt Wert auf Interaktion und Diskussion der Teilnehmer, die selbst Themen vorschlagen konnten.

Um handfeste Probleme mit ETL (Extract, Transform, Load) ging es bei einer Online-Befragung, die die Doag kürzlich durchgeführt hat und deren Ergebnisse jetzt präsentiert wurden. Die rund 200 Befragten betreiben in Deutschland ihr Data Warehouse mit Database-Infrastruktur aus dem Haus Oracle. Rund 95 Prozent der Befragten, die dazu ETL-Tools nutzen, setzen den „Oracle Warehouse Builder“ (OWB) ein.

Die Basis-ETL-Komponenten von OWB sind in der Lizenz der kostenpflichtigen Datenbank-Edition enthalten, schlagen also nicht extra zu Buche, wenn die Datenbank lizenziert ist. Das soll sich jetzt ändern: Oracle will laut einem „Statement of Direction zu OWB und Oracle Data Integrator (ODI)“ den OWB nach dem Release 1 der Datenbankversion 12c nicht mehr zertifizieren. Unternehmen sollen stattdessen ODI verwenden.

Das würde bedeuten, dass die Unter-

nehmen ziemlich in Schulung investieren müssten und zudem auch noch migrieren sollen: Laut Umfrage erwarten 80 Prozent der Befragten einen hohen oder sehr hohen Schulungsaufwand und knapp 90 Prozent fürchten hohen oder sehr hohen Migrationsaufwand. Und das ist noch nicht alles: Die Oracle Data Integrator Enterprise Edition ist kostenpflichtig und muss zusätzlich zur Datenbank lizenziert werden.

Christian Weinberger, bei der DOAG Themenverantwortlicher für BI & Data Warehouse, fasst die Umfrage-Ergebnisse so zusammen: „Oracle muss hier dringend handeln, immerhin sind fast drei Viertel der Befragten bereit, alternative Anbieter zu eruiieren“. Oracle hat bereits Verhandlungsbereitschaft signalisiert.

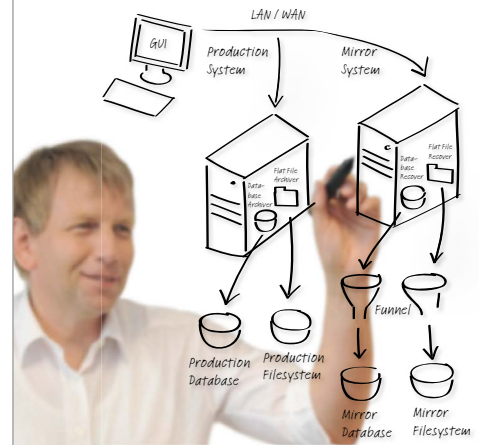
Überhaupt klagen die Oracle-Anwender über die undurchsichtige Lizenzpolitik des Herstellers. Oft werden Anwender dazu animiert, teure Lizenzen für Enterprise-Versionen zu kaufen obwohl oft schon die Standard-Version ausreicht. Wertvolle Tipps gibt es auf der DOAG-Homepage.

Java Spektrum vom 1. Dezember 2013

Am 25. und 26. März 2014 findet zum ersten Mal die JavaLand-Konferenz statt. Sie wird von der DOAG, der Deutschen Oracle-Anwendergruppe, gemeinsam mit dem Heise Zeitschriften Verlag präsentiert und richtet sich sowohl an Java-Einsteiger als auch an Java-Experten. Die Organisatoren haben die im iJUG e. V. vertretenen Java User Groups eingeladen, die Konferenz mitzugestalten, und ein Programmkomitee berufen, das auch aus JUG-Vertretern besteht. Ungeöhnliche Wege auch beim Veranstaltungsort: Der Vergnügungspark Phantasia-land in Brühl (in der Nähe von Köln) steht der Java-Community an beiden Tagen exklusiv zur Verfügung. Für zwei Tage sollen die Teilnehmer also in ihrem eigenen Java-Land „leben“ können - mit der Option, die Attraktionen des Parks mitzunutzen. Darüber hinaus sind etliche Community-getriebene Aktivitäten geplant, die über die „Features“ des Konferenzorts hinausgehen.

Weitere Pressestimmen lesen Sie unter <http://www.doag.org/presse/spiegel>

Libelle BusinessShadow®



Unabhängig bezüglich

- Fehlerursache
- Entfernung
- Hardware / Architektur
- Komplexer Systeme

Schnelle Arbeitsaufnahme

- Mit konsistenten Daten
- Auf Knopfdruck
- Automatisiert
- ...

Hans-Joachim Krüger
Chief Technology Officer
Libelle AG

Erfahren Sie mehr:
www.Libelle.com/business



ORACLE Gold Partner



Libelle

Libelle AG

Gewerbestr. 42 • 70565 Stuttgart, Germany
T +49 711 / 78335-0 • F +49 711 / 78335-148
www.Libelle.com • sales@libelle.com



Dr. Dietmar Neugebauer (rechts) beim Interview mit Roland M. Dürre, Vorsitzender des Vorstands der Interface AG

„Da findet eine Evolution statt, die kaum vorhersehbar ist...“

Die Informationstechnologie entwickelt sich schneller als je zuvor. Dr. Dietmar Neugebauer, Vorstandsvorsitzender der DOAG, und Wolfgang Taschner, Chefredakteur der DOAG News, sprachen darüber mit Roland M. Dürre, Vorsitzender des Vorstands der Interface AG.

Sie haben die Interface AG gegründet. Was ist deren Geschäftsmodell?

Dürre: Gestartet im Jahr 1984 als Produkt-Unternehmen, sind wir mit der professionellen Textverarbeitungs-Software HIT/CLOU innerhalb weniger Jahre zum europäischen Marktführer auf UNIX-Systemen aller Hersteller geworden. Um die Jahrtausend-Wende haben wir uns zum IT-basierten und Branchen-neutralen Dienstleistungs- und Beratungsunternehmen gewandelt. HIT war das UNIX-Textsystem, CLOU die ergänzende 4GL-Maschine für Textgenerierung. Schon in den ersten CLOU-Versionen gab es die Möglichkeit, SQL-Skripte einzubetten, sodass der Textautomat aus der zugewiesenen Datenbank qualifiziert lesen und bei Bedarf auch in sie schreiben konnte. In vielen Einsätzen war dies die Oracle-Datenbank.

Neben der Interface AG gibt es unter dem Dach der IF-Group noch weitere Unternehmen. Was machen diese?

Dürre: Unsere Tochter-Unternehmen sind spezialisiert auf Dienstleistungen im Umfeld besonderer Hersteller-Technologien wie Microsoft oder allgemeiner IT-Technologien wie Virtualisierung. Ein Unternehmen fällt aus der Rolle – es ist die IF-Localization, die Übersetzungen in alle Sprachen organisiert und interkulturelle Anpassungen durchführt.

In welchem Rahmen setzen Sie bei Ihren Kunden Produkte der Firma Oracle ein?

Dürre: Fast alle unsere Kunden setzen in den geschäftskritischen Prozessen die Oracle-Datenbanken ein. Diese Datenbanken sind die Basis für IT aller Art.

In welchem Bereich sehen Sie die Stärken der Oracle-Produkte, wo die Schwächen?

Dürre: Mein Eindruck ist, dass die Oracle-Produkte in der Regel immer sehr gut im Wettbewerb abschneiden. Bedauerlich ist aus meiner Sicht eher, dass es gerade im Middleware-Bereich keine Übersicht über die Vielfalt der

Produkte gibt. Da wäre noch viel Platz nach oben für Oracle drin.

Wie beurteilen Sie generell die Produkt-Strategie von Oracle?

Dürre: Es steht mir nicht zu, die Produkt-Strategie von Oracle zu bewerten. Zumal ich gar nicht in der Lage bin, die Vielfalt ganz zu überschauen. Vielleicht wäre da eine Produkt-Landkarte sinnvoll, die alle Produkte und ihre Möglichkeiten beschreibt.



Roland M. Dürre

Würden Sie Ihren Kunden empfehlen, ein Komplettsystem von der Hardware bis zu den Applikationen von einem einzigen Hersteller wie Oracle einzusetzen?

Dürre: Warum sollte ich nicht? Ich bin schließlich ein alter Fan von Sun-Systemen. Aber im Ernst, eine integrierte und abgestimmte Architektur hat viele Vorteile, die sich wahrscheinlich auch rechnen. Eine Abhängigkeit vom Hersteller besteht ja immer, ganz gleich ob der Anbieter Microsoft, IBM, SAP oder Oracle heißt.

Was hat Sie damals an Sun so beeindruckt?

Dürre: Wir waren ja eine UNIX-Firma und ich war mir damals sicher, dass Windows keine Chance im professionellen Bereich hat. Sun war damals neu auf dem Markt und in der grafischen Oberfläche der Sun-Work-

stations sah ich eine große Zukunft. Zudem waren mir die damaligen Sun-Mitarbeiter auf ihren Veranstaltungen sehr sympathisch.

Was ist Ihnen durch den Kopf gegangen, als Oracle Sun übernommen hat?

Dürre: Einerseits war ich entsetzt, andererseits hatte ich die Hoffnung, dass diese Akquisition eine große Chance ist. Die große Performance der Oracle Engineered Systems ist ein gutes Beispiel dafür.

Die InterFace AG setzt stark auf soziale Medien. Welche Erfahrungen haben Sie damit gemacht?

Dürre: Persönlich nur die besten. Allerdings ist es nicht immer einfach, die Menschen und auch zentrale Abteilungen zum Mitmachen zu bewegen. Witzigerweise erfahre ich oft aus Facebook, Google+ oder Twitter, was unsere Mitarbeiter gerade so machen, auch dienstlich.

Welche Vorteile sehen Sie im Einsatz der sozialen Medien im Unternehmen?

Dürre: Ich betrachte das heute als ein unbedingtes Muss. Das Management hat sich über die Jahre verändert, anstelle des hierarchisch aufgebauten Systems gibt es heute eine vernetzte Struktur. Bei größeren zusammenarbeitenden Gruppen sind deshalb moderne Hilfsmittel erforderlich, um beispielsweise die zahllosen Besprechungen zu vermeiden. Entscheidend ist jedoch, alle Teammitglieder von den Vorteilen der sozialen Medien zu überzeugen.

In welche Richtung wird sich die IT in den kommenden Jahren entwickeln?

Dürre: Wenn ich das wüsste, würde ich nicht hier sitzen, sondern auf einer schönen Yacht in der Karibik und mit einem Glas Champagner in der Hand. Seriös: Ich gehe davon aus, dass die IT ihren Weg, und dies sehr schnell, gehen wird. Da findet eine Evolution statt, die kaum vorhersehbar oder gar bestimmbar ist. Ich gehe aber davon aus, dass die großen Impulse nicht mehr aus Europa und auch nicht mehr so stark wie bisher aus den USA kommen werden, sondern aus Asien. In

der Pionierzeit von Unix war der große Nachteil für uns deutsche Anbieter, dass der deutsche Markt so viel kleiner als der Englisch-sprechende war. Heute bestimmen China und auch immer mehr Indien, wo es lang geht.



Zur Person: Roland M. Dürre

Roland M. Dürre hat 1977 sein Studium mit Diplom in Mathematik mit Nebenfach Informatik abgeschlossen. Von 1973 bis 1981 war er für die Siemens AG tätig, zuerst fast vier Jahre parallel zum Studium auch während der Semester als Werkstudent und später vier Jahre als fest angestellter Mitarbeiter in zwei verschiedenen Abteilungen. 1981 wechselte er zu Softlab und kam 1983 zur InterFace Computer GmbH.

Anfang 1984 gründete Roland M. Dürre gemeinsam mit Wolf Geldmacher, Claus M. Müller und Dr. Peter Schnupp die InterFace Connection Gesellschaft für Datenverarbeitung und Kommunikationssoftware mbH. 1999 wurde die InterFace Connection GmbH in eine AG umgewandelt, an der er die Mehrheit der Aktien hält.

Beruflich ist Roland M. Dürre als System-Programmierer gestartet, hat in ziemlich vielen Sprachen programmiert und mit unterschiedlichsten Betriebssystemen gearbeitet. Seine absolute Lieblingssprache war C. Der Schwerpunkt seiner Tätigkeiten hat sich im Laufe seines Lebens von der Programmierung zu unternehmerischen Tätigkeiten verändert. Bei der InterFace AG erfüllt er die Aufgabe des Vorstandsvorsitzenden.

Die InterFace AG

Die InterFace Connection hat Ende 1984 das Textsystem HIT und Anfang 1985 die 4GL für Produktionstext CLOU entwickelt und war damit für gut zehn Jahre in Deutschland Marktführer bei Produktionstext auf Sinix und verschiedenen UNIX-Systemen. Besonders stark wurde HIT/CLOU bei Bundesbehörden (wie Bundesanstalt für Arbeit) oder Landesbehörden (wie Polizei und Gerichte der Länder) und in Unternehmen mit starkem Dokumenten-Ausstoß wie Banken, Versicherungen, Handel etc. vertreten. Das System war aber auch in vielen Ländern Europas im breiten Einsatz, es gab auch eine kyrillische Version für Russland. Die InterFace Connection heißt heute InterFace AG und hat sich vom Produktunternehmen zu einem IT-Beratungs- und -Dienstleistungs-Unternehmen mit den Themen IT-Infrastruktur, IT-Organisation und IT-Applikationen gewandelt. Das Unternehmen entwickelt sich heute zu einer Unternehmensgruppe, intern „IF-Group“ genannt.

Was erwarten Sie dabei von einem IT-Unternehmen wie Oracle?

Dürre: Die Kunst wäre, die richtige Mischung und/oder Gewichtung zwischen „open“ und „proprietary“ zu schaffen. Die Märkte in der IT wandeln sich schnell und sind nicht mehr von einem Unternehmen allein beherrschbar. So sind technologische Offenheit und strategische Transparenz wünschenswert und bestimmt langfristig auch nützlich für IT-Unternehmen.

Wie sehen Sie den Stellenwert einer Anwendergruppe wie der DOAG?

Dürre: Ich habe ja eigene Siemens-Erfahrungen, als Mitarbeiter wie als Lie-

ferant. Bei Siemens gab es die Siemens Anwender-Vereinigung. Ich war immer beeindruckt, wie darüber sehr konstruktive Beziehungen gerade zwischen Großkunden und Lieferant möglich waren und welcher Nutzen für beide Seiten entstanden ist.

Welche Erfahrungen aus Ihrem langen Berufsleben können Sie an andere Unternehmer weitergeben?

Dürre: Vielleicht die wichtigste: Immer fähig sein, sich ein eigenes Urteil zu bilden. Und agil bleiben und dogmatische Entscheidungen und Handlungen um alles in der Welt verhindern. Ansonsten versuche ich, die Regeln für

modernes Management von Hans Ulrich, dem Erfinder des St.-Gallener-Management-Modells, zu beherrsigen. Diese sind: Die Unvorhersehbarkeit von Zukunft als Normalzustand zu akzeptieren, die Grenzen des Denkens weiter zu stecken, sich besser vom „sowohl als auch“ als vom „entweder oder“ leiten zu lassen, mehrdimensional zu denken, Selbstorganisation und Selbstlenkung als Gestaltungsmodell zu begreifen, Managen als sinngebende und sinnvermittelnde Funktion aufzufassen, sich auf das Wesentliche zu konzentrieren und Gruppendynamik auszunutzen. Allerdings stammen diese Regeln aus den 1980er Jahren.

SQL und Hadoop – eine gemeinsame Zukunft

Jean-Pierre Dijcks, Product Manager Big Data Oracle Corp.

Kein Zweifel, Big Data wird bleiben. In den letzten 18 Monaten avancierte die Thematik zu einem wichtigen Element in jeder Data-Warehouse-Strategie. Dennoch bleiben eine Menge Fragen über den optimalen Weg zur Realisierung einer Big-Data-Architektur und darüber, welche Tools die richtigen für diese Aufgabe sind. Ein interessantes Lösungsszenario ist die Integration von SQL und Hadoop. Der Artikel geht auf die Vor- und Nachteile einer solchen Integration ein und zeigt auf, wie solche Lösungen die Data-Warehouse, aber auch die Hadoop-Landschaft verändern.

Data Warehouse und Hadoop – brauchen wir beides? Die spontane Antwort ist ein klares „Ja“. Allerdings sollte zunächst immer die weitergehende Frage nach dem besten Weg zum Erreichen der jeweiligen Geschäftsziele gestellt werden. Zur Konkretisierung ein Beispiel: Ein erfolgreiches Unternehmen braucht eine klar definierte Geschäftsstrategie, um seine Geschäftsziele zu erreichen. Für einen Automobilhersteller mag dies lauten: „Produziere qualitativ hochwertige Autos, die ein angenehmes Fahrgefühl vermitteln und konstant die Erwartungen der Kunden erfüllen“. Doch wie kann das Unternehmen diese Strategie umsetzen? Wie lässt sich die positive Erfahrung hoher Qualität beim Kunden messen, wie kann man das Erfül-

len von Erwartungen prüfen oder ein positives Fahrerlebnis feststellen? Unternehmen brauchen dafür Daten aus der eigenen Organisation, aber auch von außerhalb – also fremde Daten.

Das Data Warehouse ist heute oft die Basis, auf der Unternehmen ihre Mission verfolgen: Es integriert Daten aus allen operativen Systemen sowie aus der Fertigung und dem Stammdaten-Management. Erfolgreiche Data-Warehouse- und Business-Intelligence-Anwendungen steuern heute Bedarfsplanung, Lieferanten-Management, Lagerhaltung und das Monitoring von Qualität und Kundenzufriedenheit. Allerdings reichen diese traditionellen Datenquellen und Managementverfahren heute immer weniger aus. Sich ausschließlich auf diese

gut strukturierten und über die Jahre gereiften Daten und Informationen zu verlassen, führt zu verpassten Chancen oder zum Nachsehen, wenn Wettbewerber solche Chancen nutzen.

Man denke beispielsweise an das Potenzial und die Auswirkungen dessen, was wir heute „Connected Cars“ nennen. In früheren Fahrzeuggenerationen wurden gerade mal die zurückgelegte Kilometer-Leistung, der Füllstand im Benzintank und der Öldruck gemessen. Es war eine „analoge Welt“. Messwerte außerhalb dieser beschränkten Möglichkeiten zu erhalten, war äußerst schwierig oder nur unter besonderen Laborbedingungen möglich. Heute verfügen moderne Fahrzeuge über weit mehr als hundert unterschiedliche Sensoren zur Messung von Zuständen.

Sie ermöglichen einen permanenten Check des Fahrzeugs: Konsistenz und Alter des Motor-Öls, Laufleistung, Geschwindigkeit, Temperatur, Drehzahl, Luftdruck im Reifen, Verwendung elektronischer Geräte durch den Fahrer, die Art der Beschleunigung, Art der Bremsvorgänge, Häufigkeit des Schaltens. Geo-Positionen mit GPS und lokale Wetterdaten bilden weitere Dimensionen möglicher Analysen. Darüber hinaus hinterlassen die Fahrer der Fahrzeuge durch Nutzung mobiler Geräte, Telefonieren, Interaktion über soziale Netzwerke, Nutzen von Diensten oder Teilnahme an Bonusprogrammen eine individuelle Spur persönlicher Verhaltensweisen und Merkmale.

Was bedeutet das für den Automobilhersteller, dem durch diese „Connected Cars“ sowieso schon ein riesiges Angebot an Daten zur Verfügung steht? Zunächst sind die Analyseverfahren nicht mehr nur beschränkt auf direkt im Fahrzeug entstehende Daten. Einblicke in reale Lebenssituationen von Kunden, Aufgreifen von Stimmungen und Verhalten im Umgang mit Mobilität beziehungsweise mit dem Fahrzeug sind möglich. Richtig umgesetzt liefert dies ganz neue Ideen und Impulse für die Entwickler neuer Features und Eigenschaften von Fahrzeugen. Zusätzlich kann man die so gewonnenen Erkenntnisse für das Aufstellen von Vorhersagemodellen und zur Steuerung von Bordcomputern nutzen. Fahrzeuge bleiben seltener liegen, weil schon vor Eintreten von Defekten durch vorgezogene Wartungsarbeiten reagiert werden kann.

Die technische Herausforderung des Sammelns und Integrierens von großen Volumen schwach strukturierter Messdaten der „Connected Cars“ ist mittlerweile durch die rasch entwickelten, neuen Big-Data-Techniken gelöst worden. Hadoop- und NoSQL-Datenhaltungen bieten heute eine ökonomisch sinnvolle Lösung.

Bei Big Data redet man oft von „Low Value“-Daten. Diagnose-Daten einzelner Fahrzeuge werden je nach Nutzung Tausende Male pro Tag eingesammelt. Wir sammeln alles, was kommt. Einzelne Messdaten gehen jedoch durch Messfehler, Funklöcher, Softwarefehler etc. verloren. Im Vergleich zur Masse



Abbildung 1: Die klassische Optimierung

der Daten fallen solche vereinzelt „Dirty Data“ nicht auf. Sie werden die Gesamt-Tendenz einer Analyse kaum oder gar nicht beeinflussen, die eine Menge von Millionen oder gar Milliarden Einzel-Informationen verwertet. Bei traditionellen „High-Value“-Daten wäre dies ein Drama. Man stelle sich vor, Kontobewegungen einer Bank würden verloren gehen oder bestellte Produkte kämen nicht bei Kunden an, weil der Bestelldatensatz fehlerhaft ist.

Analyse-Methoden bleiben unverändert

Die klassischen Analyse-Verfahren zur Optimierung von Geschäftsabläufen und Zielerreichung sind gemeinhin bekannt:

- Zunächst formuliert man ein Ziel wie die Verbesserung der Produktqualität
- Man sammelt alle zur Verfügung stehenden Daten, von denen man glaubt, dass man sie braucht und sucht mithilfe von „Predictive Analytics“ – also Data-Mining-artiger und heuristischer Zusammenhangesuche – nach Abhängigkeiten in diesen Daten
- Sind Abhängigkeiten gefunden, formuliert man diese als Regeln (sogenannte „Modelle“) und wendet sie zur Überprüfung auf neue Datenbestände an

- Sind die Modelle verifiziert, nutzt man sie zur Steuerung von operativen Systemen
- Zur Kontrolle setzt man ein Monitoring auf die Ergebnisse in den operativen Prozessen und misst den Erfolg über die Laufzeit der Anwendung

Das Ganze ist ein immer wiederkehrender Prozess (siehe Abbildung 1). Mit der Zeit verfeinert und verändert sich das Modell – je nach den bereitgestellten Daten und Randbedingungen. In einem Fahrzeug mag sich erst nach Jahren eine Fehlkonstruktion eines Bauteils herausstellen, wenn Erfahrungen unter bestimmten Fahrbedingungen und klimatischen Verhältnissen gesammelt wurden. Nicht nur das Bauteil, sondern auch das zugrunde liegende Modell wird geändert. Die Ingenieure berücksichtigen diese Erfahrungen bei Folgeentwicklungen neuer Bauteile. Klingt aufwändig und zeitintensiv. Da ist es nachvollziehbar, dass man solche Modell-Iterationen optimieren und beschleunigen will.

Hadoop oder relationale Datenbank

Das Hinzufügen von schwach ausgeprägten zusätzlichen Datenmengen ändert nichts an diesem Analyse-Prozess. Man zieht nur andersartige Daten mit heran und sie mögen auch anders verarbeitet werden. Am Ende geht es je-

doch immer darum, die Ergebnisse von Analysen in Modelle zu überführen, um Geschäftsprozesse zu optimieren. Die Frage, wo die technische Verarbeitung stattfindet – in Hadoop oder in einer relationalen Datenbank –, ist von einer Reihe von Faktoren abhängig:

- Entwicklungsreife von Werkzeugen
- Performance der Verarbeitung
- Security-Aspekte
- Fähigkeit, Daten schnell entgegen zu nehmen
- Wirtschaftlichkeit bei der Speicherung von „Low Value“-Daten
- Einfache ETL-Prozesse
- Vollständigkeit der Daten (Machen zu viele Lücken den Datenbestand sinnlos?)
- Unterschiedlichkeit der Datenstrukturen (Variety)
- Komplexität des Daten-Managements

Für einige Faktoren weist jede der beiden Plattformen Stärken auf, potenziell sind sicher alle Faktoren von beiden Plattformen lösbar. Hadoop ist „Schema-los“. Die Daten sind also zunächst ohne Struktur abgelegt. Diese Struktur wird erst durch die lesenden Programme erkennbar („Schema on Read“). Diese Flexibilität hat Vorteile:

- Das extrem schnelle Laden der Daten in einen Hadoop Data Store. Die Daten werden praktisch ohne nähe-

ren Struktur-Syntax-Check einfach „abgekippt“.

- Wenn sich die Formate der Fahrzeug-Diagnose-Sätze mit der Zeit ändern, dann hat das wenig Einfluss auf die ETL-Strecken. Die Änderungen müssen nur das Lesen berücksichtigen.
- Das Map-Reduce-Framework arbeitet massiv parallel. Die geringe Strukturierung der Daten kommt dieser hohen Parallelisierung sehr entgegen.
- Die überschaubaren Kosten für Storage und Rechenleistung für die zum Teil sehr hohen Datenmengen.

Relationale Systeme hingegen haben diese Stärken:

- Die bestehende Strukturierung ermöglicht „Schema on Write“, also Prüfung von Daten im Zuge der Speicherung.
- Es gibt eine hohe Zahl von gut entwickelten Werkzeugen für eine einfache, reibungslose und performante Verarbeitung beziehungsweise Datenanalyse.
- Anwender können ohne besondere Hürden mit einfachen Zugriffen auf einem standardisierten Datenmaterial ad hoc und multidimensional analysieren.
- Die Daten sind in der Regel qualitäts gesichert und die Analyse-Verfahren sind erprobt.

Die Herausforderung heute lautet also nicht „Können wir alles in Hadoop oder in einem Data Warehouse machen?“, sondern „Wie können beide Technologien so verzahnt werden, dass die spezifischen Vorteile der jeweiligen Technologien am besten die Unternehmensziele unterstützen?“ Tatsächlich ist die Koexistenz der Technologien nicht nur kurzfristig als Übergangszeit zu sehen. Die vornehmliche Speicherung bestimmter Daten auf der einen oder anderen Technologie wird zwar bleiben, aber die Kombination Data Warehouse/Hadoop wird ein fester Bestandteil künftiger IT-Infrastrukturen werden.

Gemischte Datenablage und Abfragesprache für Big Data

Werden Daten mit beiden Technologien gespeichert, stellen sich die Fragen „Wie sollen Daten in einer gemischten Umgebung, also aus Hadoop und aus dem Data Warehouse, gelesen und verarbeitet werden, ohne dass dies zu doppeltem beziehungsweise für Endbenutzer nicht zumutbarem Aufwand führt?“ und „Mit welcher Abfragesprache soll dies geschehen?“

Mit zunehmender Reife der Hadoop-Technologie und der wachsenden Akzeptanz in den Unternehmen nimmt auch der Nutzen einer integrierten Verwendung von Hadoop und Data Warehouse zu (siehe [Abbildung 2](#)). Die Hauptgründe sind:

- Reduzierung von Komplexität
- Schnellere Verfügbarkeit der zu analysierenden Daten (Wegfall von Datenbewegungen/ETL)
- Die Möglichkeit, Daten beliebig zu kombinieren beziehungsweise korrelierende Zusammenhänge zu analysieren

Das Adaptieren neuer Technologie erfolgt in den Unternehmen meist in mehreren Phasen. Auf dem Weg zur integrierten Verwendung von Hadoop und des Data Warehouse lassen sich drei Entwicklungsstadien erkennen. Wo die einzelnen Industrien heute stehen, lässt sich beantworten, indem man diese näher betrachtet.

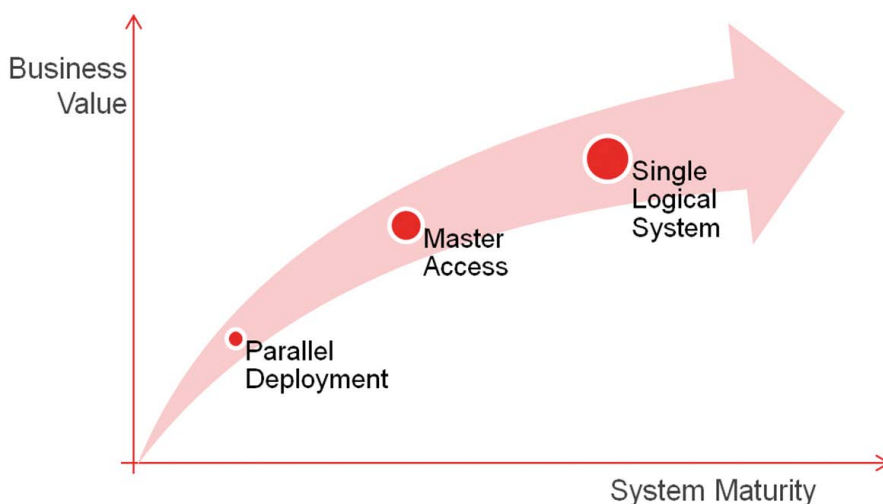


Abbildung 2: Reife-Phasen von Hadoop- und RDBMS-Implementierungen

Phase 1: Parallel Deployment

In der ersten Phase erfolgen Hadoop-Implementierungen parallel zu bestehenden RDBMS-Umgebungen. Die Hadoop-Umgebungen sind isoliert. In einer Art „Labor-Situation“ kann das Unternehmen den Umgang mit Hadoop erlernen und Best Practices mit allen Stärken und Schwächen von Hadoop erfahren. Aktuell ist eine große Anzahl von Unternehmen in dieser Phase. Es gibt Test-Umgebungen, in denen man Sinn und Zweck von Hadoop für das eigene Unternehmen erforscht. Das Lösen technischer Klippen und spezieller Use-Cases steht im Vordergrund.

Hadoop besitzt ein mächtiges Map-Reduce-Framework, das allerdings für die meisten Mitarbeiter – ob Fachabteilung oder IT – fremd ist. Folglich ist einer der ersten Schritte die Beschäftigung mit Hive oder der HiveQL-Abfragesprache. Hive bietet eine (RDBMS-) Datenbank-analoge Beschreibung von Hadoop-Daten (HDFS) in Tabellen, Spalten und sogar Partitionen. Auf relationalen Metadaten bietet HiveQL eine deklarative, SQL-artige Abfragesprache an, die die Abfragen im Hintergrund in Map-Reduce-Programme umwandelt. Der wichtigste Vorteil von Hive ist die Wiederverwendung der SQL-Skills durch SQL-erfahrene Benutzer, die damit leicht auf Daten in Hadoop zugreifen können. Der große Nachteil (verglichen mit einer traditionellen Datenbank) sind die schlechteren Antwortzeiten von Standard-Abfragen. HiveQL wird über Map-Reduce-Programme abgearbeitet, was Zeit kostet.

Erkennen Unternehmen einen Sinn in der Hadoop-Praxis, entsteht der Wunsch nach einer verstärkten integrierten Anwendung von Hadoop mit bestehenden Datenbank-basierten Anwendungen – abgesehen von spezifischen Anwendungsfällen, in denen ein isoliertes Hadoop-System einen Sinn ergibt. Der Übergang zur Phase 2 ist beschritten, jetzt wird SQL ein wichtiger Bestandteil der Big-Data-Debatte.

Phase 2: Hybrid Master Access

Hat sich ein Unternehmen für den Einsatz von Hadoop entschieden, wird überlegt, wie und in welcher Form

Endbenutzer auf diese neue Datenwelt zugreifen und Analysen durchführen. Man benötigt die Möglichkeit, auf beide Datenwelten gleichzeitig zuzugreifen, und erwartet eine Art konsolidierte Sicht, einen sogenannten „Master Data Point“. Dieser ist in drei Varianten vorstellbar:

1. Das zentrale Data Warehouse mit klassischen ETL-Prozessen. Hier betrachtet man das Hadoop-System lediglich als weitere Datenquelle.
2. Ein BI-Tool, das Zugriffe auf beide Welten erlaubt und die Konsolidierung im BI-Tool vornimmt.
3. Ein führendes Hadoop-System, das Daten auch aus einem Data Warehouse in sich aufnimmt.

Die meisten Unternehmen, die sich schon sehr früh mit Hadoop-Techniken beschäftigten, befinden sich mittlerweile in dieser zweiten Phase. Pragmatischerweise nutzen sie eine Mischform der drei genannten Varianten. Solche hybriden Lösungen bieten Zugriffe auf ein Subsystem an, während man gleichzeitig Daten über ETL austauscht – es wird vorsortiert und stellenweise aggregiert. Greift beispielsweise ein Teil der Benutzer über Hive auf Hadoop-Daten zu, liest ein großer Teil der übrigen Benutzer über ein BI-Tool Daten aus einem Data Warehouse mit voraggregierten Kennzahlen.

Diese pragmatische Hybrid-Lösung führt natürlich zu dem Nachteil der Komplexität von systemübergreifenden Zugriffen und Daten-Replikationen. Fährt man ein föderatives System, belässt also Daten an ihrer originären Stelle in Hadoop beziehungsweise im Data Warehouse, so heimst man sich „Latency“-Probleme ein: Typische HiveQL-Abfragen sind langsamer als SQL-Abfragen in einem Data Warehouse, ein Join von Daten in beiden Systemen fällt auf die Geschwindigkeit des langsamsten Systems zurück. Das Resultat sind unzufriedene BI-Benutzer.

Oft ist das Data Warehouse der primäre Zugriffspunkt für die Benutzer, weil hier bereits viele wohlsortierte und überprüfte Daten zu finden sind. Die Umgebung ist zudem sicher und erfüllt oft regulatorische Anforderun-

gen. Die Performance ist in der Regel besser als in Hadoop-Systemen.

Hat man jedoch die Masse der Rohdaten im Hadoop-System belassen und nur Extrakte beziehungsweise Aggregate in das Data Warehouse überführt, werden nicht alle Analysen und Anfragen über das Data Warehouse zu lösen sein. Tritt eine solche Situation ein, muss man Rohdaten in dem Hadoop-System nachladen, zusätzliche Filter bemühen und Daten erneut in das Data Warehouse überführen beziehungsweise aggregieren.

Zur Lösung solcher Probleme entstehen immer mehr technische Anwendungen. Auf der Hadoop-Seite werden SQL-Engines entwickelt, die gegenwärtig neu auf den Markt kommen. Diese ersetzen die Map-Reduce-Komponente und führen neue Optimizer zur Plan-Ausführung von Abfragen ein. Solche Anwendungen sind jedoch erst in ihrer frühen Entwicklungsphase und verfügen noch nicht über gewohnte Features von eingeführten RDBMS-Datenbanken. Sie schließen schon gar nicht die Lücke zwischen der Hadoop-Welt und einem Data Warehouse. Sie sind tauglich als Einzelzugriffs-Tools innerhalb der Hadoop-Welt.

Mittlerweile verfügen auch BI-Tools über einen direkten Hive-Zugriff und können Hadoop-Daten in ihrem Server-Cache für begrenzte Analysen vorhalten (wie Oracle Exalytics). Solche Lösungen sind heute schon stabiler und ausgereifter als SQL-Engines innerhalb von Hadoop. Die meisten der fortgeschrittenen Einsätze in den Unternehmen nutzen die hier dargestellten hybriden Lösungen und nehmen das Bewegen von Key-Daten beziehungsweise aggregierte Daten in Kauf.

Phase 3: Single Logical System

Die sinnvollste Lösung zur Eliminierung aller Nachteile der jeweiligen Plattformen ist ein integrierter Datenspeicher, der die spezifischen Vorteile von Hadoop auf der einen und eines relationalen Datenbanksystems auf der anderen Seite für Analysezwecke vereint. Nahezu keine aktuelle Implementierung erreicht gegenwärtig diese Phase. Die meisten Entwicklungen fokussieren sich auf eine der beiden Tech-

nologien, anstatt an einem einheitlichen logischen System zu arbeiten.

Das in der dritten Phase beschriebene, logisch integrierte System ist Voraussetzung für neue Möglichkeiten einer breiteren Anwenderschicht. Damit einher geht die Standardisierung der Zugriffssprache auf diesen neuen Datenspeichern. Es muss eine deklarative Sprache sein, die Zugriff auf beide Welten (Hadoop und RDBMS) erlaubt. Wie bereits ausgeführt, entwickelt sich SQL zu der bevorzugten Abfragesprache in der Hadoop-Welt, so wie sie es im Data Warehouse bereits seit vielen Jahren ist. Obwohl es momentan noch andere Abfragesprachen für Hadoop gibt, wird SQL die Abfragesprache der Wahl für integrierte Systeme werden.

Beim Einsatz von SQL haben alle bestehenden Werkzeuge (BI und sonstige Anwendungen) uneingeschränkten Zugang auch zu Hadoop-Daten, ohne dass man an dem bestehenden Ökosystem etwas ändern muss. Ein solches System muss jedoch Zugriffe quer über das Data Warehouse und Hadoop hinweg möglichst optimiert erlauben, ohne die Nachteile, die man heute bei Hive findet. Hadoop muss Hadoop bleiben, mit der Möglichkeit, schnell und massenweise Daten zu importieren und diese Daten flexibel und einfach auswerten zu können, bei gleichzeitiger Skalierbarkeit.

Abbildung 3 zeigt den Bauplan für eine solche Lösung. Sie umfasst ein Hadoop- und ein RDBMS-System für die jeweiligen beschriebenen spezifischen Anforderungen. Zwischen beiden Bestandteilen sollte ein High-Speed-Netzwerk liegen. Beispiele dafür sind InfiniBand oder die neueste Ethernet-Technologie (> 10GigE).

Die Zugriffe mittels SQL müssen über einheitliche Metadaten erfolgen, die sich über beide Systeme erstrecken. Hive und HCatalog stellen die Basis für den Hadoop-Anteil des Metasystems dar und sind mit dem Datenbank-Katalog verbunden, sodass ein einheitliches SQL-System alle Objekte sowohl im Hadoop als auch in der Data-Warehouse-Datenbank zusammenhängend abfragen kann.

Die Benutzer können ohne Änderung ihrer Umgebung oder ihrer Abfra-

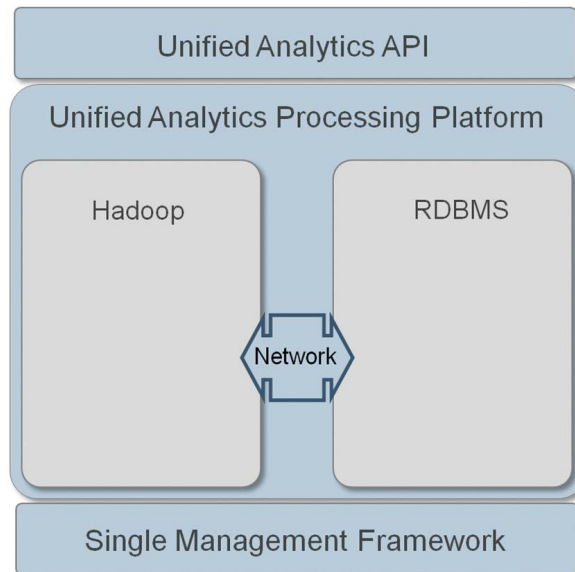


Abbildung 3: Vereinheitlichte Zugriffsplattform Hadoop/RDBMS

gesprache in beide Welten hinein abfragen und in einer einzigen Abfrage Daten aus beiden Welten zusammen in ihre Analyse miteinbeziehen. Die Performance der Abfragen ist gut, weil die Abarbeitung direkt auf Prozesse in der jeweiligen Schicht übertragen und dort erledigt wird. Es findet keine aufwändige Datenbewegung statt, außer in den Systemen selbst. Nur die Ergebnisse fließen an die Benutzer zurück.

Auch das föderative Konzept lässt sich sinnvoll umsetzen. BI-Server müssen jetzt nicht mehr alle Daten in ihren Cache laden, sondern können sich auf die wesentlichen Daten fokussieren. Die Hadoop-spezifische Verarbeitung erfolgt im Hintergrund durch die jeweiligen Werkzeuge innerhalb von Hadoop.

Wie gesagt, bewegen sich die technischen Entwicklungen im Augenblick auf dem Level der zweiten Phase. Hadoop, oft verstanden als Erweiterung des Data Warehouse, leistet bereits an einigen Stellen seinen Wertbeitrag in den Unternehmen. Anwender sollten sich mit den eingestellten Erfolgen jedoch nicht zufriedengeben, sondern ein Ziel-Szenario anstreben, in dem SQL als einfaches und standardisiertes Zugriffsmedium zu allen relevanten Daten vorhanden ist.

Fazit

Hadoop spielt in den kommenden Jahren eine immer bedeutsamere Rolle in

Data Warehouse-Architekturen und SQL wird sowohl innerhalb von Hadoop als auch als eine übergreifende Abfragesprache wichtiger denn je. Heutige Lösungen ziehen Daten entweder hoch in den BI-Layer oder sie kopieren Daten beziehungsweise Aggregate in eine Data-Warehouse-Umgebung. SQL-Lösungen innerhalb von Hadoop leiden heute noch an Performance-Limitierungen. Eine sinnvolle Lösung sind SQL-gestützte Brückenlösungen über beide Datenhaltungen hinweg. Diese Lösungen werden in den kommenden Jahren zum Hauptstandbein, wenn Unternehmen über alle ihre Daten hinweg Nutzen ziehen wollen.

Jean-Pierre Dijcks



Übersetzt und angepasst von
Alfred Schlaucher
alfred.schlaucher@oracle.com

Data-Mining in sozialen Online-Netzwerken

Bianca Böckelmann, Robotron Datenbank-Software GmbH

Soziale Online-Netzwerke wie Facebook bieten Nutzern zwar die Möglichkeit, Profil-Informationen durch Einstellungen vor Unbefugten zu verbergen, jedoch stellt sich die Frage nach einem ausreichenden Schutz der Privatsphäre. Inwieweit lassen sich zum Beispiel in solchen Netzwerken mittels Data-Mining durch Verknüpfung von öffentlichen Informationen der Freunde Rückschlüsse auf die privaten Daten eines Nutzers ziehen, ohne dass diese Daten explizit offengelegt wurden?

Im Rahmen einer Masterarbeit wurde diese Frage im Jahr 2012 an der Brandenburgischen Technischen Universität Cottbus in Kooperation mit der Robotron Datenbank-Software GmbH und der IHP GmbH untersucht. Ziel war es, die Risiken aufzuzeigen, die sich durch die Bekanntgabe von Freundschaften ergeben. Ist es also nötig, neben den eigenen Profil-Informationen auch die Freundschaftsbeziehungen privat zu halten, um die eigene Privatsphäre zu schützen?

Dieser Artikel gibt zunächst einen Überblick über die Schritte des zugrunde gelegten „Knowledge Discovery in Databases“-Prozesses (KDD). Nach einer kurzen Einführung in Oracle Data Mining wird der Prozess praxisnah am Beispiel des Forschungsthemas beschrieben. Hierbei kam der Oracle Data Miner (ODMr) mit Daten freiwilliger Facebook-Nutzer zum Einsatz. Neben der Anwendung der eigentlichen Data-Mining-Algorithmen für die Prognose beinhaltet der KDD-Prozess die vor-

herige Vorverarbeitung und Transformation der Nutzerdaten, die zeitlich einen Großteil des Gesamtprozesses einnehmen. Abschließend werden die Ergebnisse vorgestellt, die die Vermutungen bekräftigen, dass auch die Betrachtung der Freundes-Informationen wichtig ist, um seine Privatsphäre zu schützen.

Der KDD-Prozess

Knowledge Discovery in Databases (KDD) ist der (semi-)automatische [1], nicht triviale Prozess des Identifizierens von gültigen, bisher unbekanntem, potenziell nützlichen und verständlichen Mustern in Daten [2]. Der KDD Prozess besteht aus mehreren Phasen, die iterativ durchlaufen werden können (siehe **Abbildung 1**). Zu Beginn werden bei der Selektion das Ziel festgelegt sowie die Daten für die Wissensextraktion herangezogen, bezüglich ihrer Qualität beurteilt [1] und im nächsten Schritt für

das Erzielen von Integrität, Konsistenz und Vollständigkeit vorverarbeitet. Anschließend folgt die Transformation der vorverarbeiteten Daten in eine für den Analyse-Schritt geeignete Repräsentation. Dazu zählt unter anderem die Auswahl relevanter Attribute.

In der eigentlichen Data-Mining-Phase werden die Algorithmen für die Extraktion von Mustern angewendet. Abschließend werden die erzielten Ergebnisse interpretiert und hinsichtlich des festgelegten Ziels bewertet. Sind die gefundenen Modelle gut, kann das hieraus gewonnene Wissen beispielsweise gezielt zum Ableiten von Handlungsdirektiven eingesetzt werden.

Oracle Data Mining

Oracle bietet mit Oracle Data Mining (ODM) als Teil der Oracle-Advanced-Analytics-Option die Möglichkeit, Da-

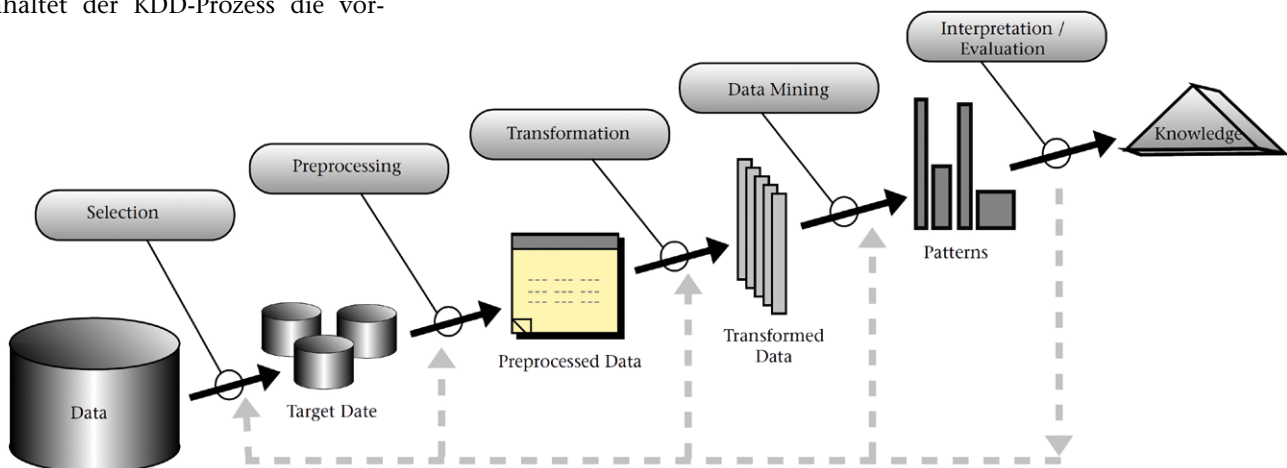


Abbildung 1: Schritte im KDD-Prozessmodell [3]

ta-Mining-Algorithmen direkt in der Oracle-Datenbank auszuführen. Für einen Zugriff auf die Quelldaten, die Modelle sowie die Analyse-Ergebnisse ist somit kein zusätzlicher Datentransport nötig. Die ODM-Funktionalitäten sind über PL/SQL, SQL und R im Zugriff [4]. Zudem steht mit dem Oracle Data Miner (ODMr) eine grafische Benutzeroberfläche bereit, die im Oracle SQL Developer eingebunden ist.

Mittels Workflows, die sich aus Knoten und Verbindungen zusammensetzen, lässt sich der gesamte KDD-Prozess abbilden. **Abbildung 2** zeigt einen einfachen Workflow. Neben einfachen Statistiken, die einen Überblick über die Daten erlauben, können die Daten geeignet gefiltert, aggregiert und transformiert werden. Hierfür stehen beim ODMr verschiedene Knoten im Workflow-Editor zur Auswahl.

Grundlegend unterscheidet ODM drei Arten von Daten: numerische, kategorische und unstrukturierte. Die letzte Kategorie ist bei Texten relevant und erst in ODM 12.1 verfügbar [5, 6]. Anhand dieser Differenzierung werden die Daten in verschiedenen Data-Mining-Algorithmen anders behandelt. Eine Zuordnung findet entweder automatisch über den Datentyp oder manuell statt.

Für das Lernen von Data-Mining-Modellen unterstützt ODM die Klassifikation, die Regression, die Assoziations-Analyse, das Clustering, die Anomalie-Erkennung sowie die Feature-Extraktion und die Attribut-Wichtigkeit. Tabelle 1 gibt einen Überblick über die in ODM 12.1 zur Verfügung stehenden Data-Mining-Funktionen und zugehörigen Algorithmen. Die in blauer Schriftfarbe aufgeführten Algorithmen sind in der Version 12.1 neu hinzugekommen.

Da für die Masterarbeit ODM 11.2 zum Einsatz kam, wird im Folgenden vorwiegend Bezug auf diese Version genommen. Für eine Klassifikation können beispielsweise der Entscheidungsbaum, Naive Bayes, Support Vector Machine (SVM) sowie das generalisierte lineare Modell (GLM) in der logistischen Regression eingesetzt werden [4].

Bei ODM wird für das Anwenden der Algorithmen vorausgesetzt, dass alle Daten in genau einer Tabelle oder Sicht vorliegen, der sogenannten „Fall-Tabelle“. Einstellungen zu den Algorithmen können entweder standardmäßig belassen oder individuell zugeschnitten auf den jeweiligen Anwendungsfall vorgenommen werden. Für eine Be-

wertung der gelernten Modelle stehen in Abhängigkeit vom Verfahren diverse Evaluierungsmaße und Visualisierungen bereit. Beispielsweise kommt bei der Klassifikation die Gesamtklassifikationsgenauigkeit zum Einsatz. Diese ist der Anteil der korrekt klassifizierten Objekte in der gesamten Menge.

Indirekte Bestimmung privater Informationen

In der Masterarbeit wurde untersucht, ob es in sozialen Online-Netzwerken mittels Data-Mining möglich ist, nur anhand der öffentlichen Informationen der Freunde auf die privaten Daten einer Person selbst zu schließen. Müssen also neben den eigenen Profil-Informationen auch die Freundschafts-Beziehungen vor der breiten Öffentlichkeit privat gehalten werden, um personenbezogene Daten vor dem Ausspähen durch Fremde zu schützen?

Zu Beginn wurde dafür die Facebook-Anwendung „Hidden Profile“ in PHP entwickelt, um knapp 700 freiwillige Nutzer zu gewinnen. Neben Profil-Informationen wie Geburtstag, Geschlecht, Wohn- und Heimatort, „Gefällt mir“-Angaben (unter anderem TV, Film, Musik), Informationen zur Arbeit und Bildung wurden auch

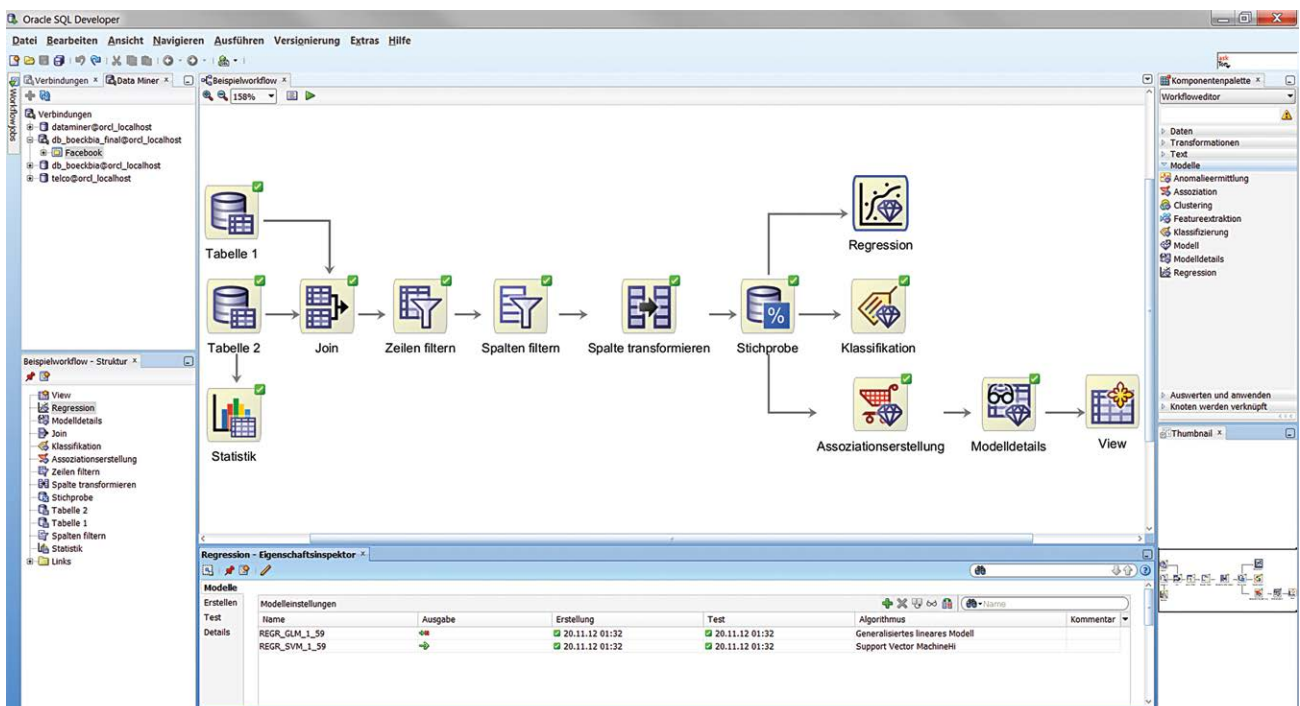


Abbildung 2: Benutzeroberfläche des Oracle Data Miner mit der Anzeige eines Beispiel-Workflows

Freundschafts-Beziehungen erfasst. Hierbei wurden nur diejenigen Freunde in der Datenbank gespeichert, die auch explizit der Anwendung zugestimmt hatten.

Im Schritt der Vorverarbeitung wurden zum einen fehlende Werte behandelt und zum anderen die Daten bereinigt. Während einige fehlende Daten wie der Wohn- und Heimatort anhand anderer im eigenen Profil zu findenden Informationen abgeleitet werden konnten, war dies für andere Attribute nicht möglich. Bei letzteren wurde der Median beziehungsweise der Modus unter allen Teilnehmern eingesetzt, um fehlende Werte zu behandeln.

Selten angegebene Attribute wurden jedoch von der weiteren Betrachtung ausgeschlossen, da zu viele künstliche Werte die Analyse verfälschen könnten. Zudem war eine Bereinigung der Daten notwendig, da in Facebook aufgrund von beispielsweise Rechtschreibfehlern und Mehrsprachigkeit viele unterschiedliche Bezeichnungen für das semantisch gleiche Objekt existieren.

Für die Erstellung eines hierfür notwendigen einheitlichen Vokabulars kamen DBpedia [8], Freebase [9] und andere frei verfügbare Datenquellen zum Einsatz. Die Profil-Informationen aus Facebook wurden automatisch je-

weils dem Wort aus den zuvor genannten Datenquellen zugeordnet, das die maximale Jaro-Winkler-Ähnlichkeit besitzt und einen gewissen Schwellenwert überschreitet. Die Jaro-Winkler-Ähnlichkeit ist als Funktion im Oracle-Datenbank-Package „UTL_MATCH“ verfügbar. Im Anschluss daran wurde eine manuelle Nachkorrektur aufgrund von Synonymen und Homonymen vorgenommen.

In der Transformationsphase wurde neben der Normalisierung von metrischen Attributen und Datentyp-Konvertierungen auch eine Kategorisierung vorgenommen, um durch die Reduzierung der Kardinalität von Attributen möglicherweise kompaktere Modelle zu erzielen. Hierfür wurden vorwiegend die zuvor erwähnten Datenquellen genutzt. Beispielsweise hat man eine Orts-Hierarchie aufgebaut sowie TV-Sendungen ihren Genres beziehungsweise das Alter mittels Quantil-Binning Altersgruppen zugeordnet.

Darüber hinaus wurden mit dem ODMr Ausreißer anhand von One-Class SVM identifiziert und entsprechend behandelt. Als Ausreißer gelten Nutzer mit seltenen, von der Mehrheit abweichenden Informationen. Beispielsweise hat man unrealistische Altersangaben im Bereich von 100 Jahren geglättet.

Ein weiterer wichtiger Punkt ist die probate Abbildung der Freundschaftsbeziehungen für die Prognose. Nur ein Teil der knapp 700 teilgenommenen Facebook-Nutzer konnte für die Vorhersage genutzt werden, da nur bei wenigen auch genügend Freunde der Anwendung zugestimmt hatten und somit deren Daten vorlagen. Während beispielsweise 241 Nutzer mindestens drei teilgenommene Freunde besaßen, waren es bei mindestens fünf Freunden nur noch 150 Facebook-Mitglieder.

Außerdem wurden solche Personen gänzlich von der Analyse ausgeschlossen, die mehr als 700 Leute in ihrer Facebook-Freundesliste pflegten. Dies liegt in der Vermutung begründet, dass neben engen Freundschaften auch viele Bekanntschaften enthalten sind, die eine Prognose wahrscheinlich verschlechtern würden. Zudem war es vorstellbar, neben den Personen aus der eigenen Freundesliste zusätzlich die indirekten Freunde einzubeziehen, also die Freunde der Freunde, um möglicherweise bessere Modelle zu erhalten.

Da die Anzahl der Freunde über verschiedene Personen hinweg variiert und ODM genau eine Fall-Tabelle voraussetzt, ist dazu eine Aggregation der Freundeswerte notwendig gewesen. Es wurde deshalb für jeden Nutzer ein Freundesvektor für jedes Attribut erstellt, der dem gewichteten relativen Vorkommen eines jeden Attributwerts unter seinen Freunden entspricht.

ODM 11.2 stellt für die Speicherung solcher verschachtelter Daten die Datentypen „DM_NESTED_NUMERICALS“ und „DM_NESTED_CATEGORICALS“ bereit [6], in ODM 12.1 stehen zusätzlich „DM_NESTED_BINARY_DOUBLES“ und „DM_NESTED_BINARY_FLOATS“ zur Verfügung [5]. Da die Anzahl der indirekten Freunde viel größer ist als die der direkten, wurde der relative Anteil für die zwei Freundespartitionen separat berechnet. Um den direkten Freunden eine größere Bedeutung zukommen zu lassen als den indirekten, wurde ihnen bei der anschließenden Summierung der Anteile ein größeres Gewicht zugewiesen.

Irrelevante und korrelierte Attribute verfälschen ein Modell. Aus diesem Grund wurden mit dem ODMr

Data-Mining-Funktionen	Data-Mining-Algorithmen
Klassifikation	<ul style="list-style-type: none"> • Entscheidungsbaum • Naive Bayes • Support Vector Machine (SVM) • Generalisiertes Lineares Modell (GLM) → <i>logistische Regression</i>
Regression	<ul style="list-style-type: none"> • GLM → lineare Regression • SVM
Assoziations-Analyse	<ul style="list-style-type: none"> • Apriori
Clustering	<ul style="list-style-type: none"> • k-Means • O-Cluster • <i>Erwartungs-Maximierung</i>
Anomalie-Erkennung	<ul style="list-style-type: none"> • One-Class SVM
Feature-Extraktion	<ul style="list-style-type: none"> • Nicht-negative Matrix-Faktorisierung • <i>Singulärwert-Zerlegung mit Hauptkomponentenanalyse</i>
Attribut-Wichtigkeit	<ul style="list-style-type: none"> • Minimum Description Length (MDL)

Tabelle 1: Überblick über die Data-Mining-Algorithmen in ODM 12.1; Algorithmen in blauer Schriftfarbe sind in ODM 12.1 neu hinzugekommen [7]

irrelevante Attribute für die Vorhersage bestimmter Merkmale automatisch mittels Minimum Description Length (MDL) gefiltert.

Data-Mining

Das Ziel der Analyse war eine Prognose der Attributwerte eines Facebook-Nutzers ausschließlich anhand der Angaben seiner Freunde. Der vorherzusagende Nutzer stellte somit eine Blackbox dar. Die Klassifikation und die Regression sind für eine solche Aufgabenstellung geeignet. Für eine exemplarische Prognose von Geschlecht und Wohnort kamen die Klassifikations-Algorithmen SVM und Naive Bayes zur Anwendung. Aufgrund einer binären Ziel-Variablen kam beim Geschlecht zusätzlich die logistische Regression zum Einsatz.

Da der Entscheidungsbaum erst in ODM 12.1 verschachtelte Daten verarbeiten kann [5, 6], wurde deshalb beim Freundesvektor ein repräsentativer Wert ausgewählt. Für die Vorhersage des metrischen Alters wurden bei der Regressions-Analyse die lineare Regression und SVM angewendet. Eine Evaluierung der verschiedenen Modelle war durch eine Aufteilung der Objektmenge in zwei Drittel Trainings- und ein Drittel Test-Menge gegeben.

Darüber hinaus existiert neben den klassischen Prognose-Verfahren auch die Möglichkeit, anhand von Assoziationsregeln Gemeinsamkeiten zwischen allen Nutzern, also nicht nur zwischen seinen Freunden, für die Prognose genau eines Merkmals einzusetzen. Das setzt voraus, dass der Nutzer selbst für ein Attribut mindestens einen Wert öffentlich hält, damit anhand von Regeln weitere private Angaben vorhergesagt werden können.

Die Assoziations-Analyse wurde exemplarisch für TV-Sendungen durchgeführt. Gehört beispielsweise „The Big Bang Theory“ zu den Lieblings-Fernsehsendungen eines Nutzers, so kann man anhand der gefundenen Regel „The Big Bang Theory“ → „How I Met Your Mother“ die nicht in seinem Profil öffentlich einsehbare Serie „How I Met Your Mother“ zu seinen favorisierten Sendungen zählen.

Unter der Annahme, dass alle Informationen der Freunde eines Nutzers

öffentlich zugänglich sind, konnte die Vermutung bestätigt werden, dass die Daten der Freunde gut geeignet sind, um mehr über einen Facebook-Nutzer zu erfahren. Beachtlich ist die gute Prognose des Alters anhand des Geburtstages der Freunde. Durchschnittlich wich das vorhergesagte Alter nur um 2,05 Jahre vom tatsächlichen ab. In sieben von zehn Fällen konnte der Wohnort korrekt anhand verschiedener Angaben der Freunde geschlussfolgert werden, das Geschlecht war in 65 Prozent aller Fälle richtig.

Neben der Untersuchung der prinzipiellen Ableitbarkeit wurden verschiedene Einflussfaktoren für die Güte der Ergebnisse analysiert. Es hat sich gezeigt, dass die Qualität der extrahierten Modelle unter anderem von der Wahl des Data-Mining-Algorithmus abhängt. Für die Vorhersage des Wohnorts war Naive Bayes am besten geeignet, beim Geschlecht war es hingegen SVM.

Der Entscheidungsbaum, der im Gegensatz zu anderen Modellen vom Menschen leichter interpretierbar ist, hatte im Allgemeinen schlechte Ergebnisse geliefert. Die lineare Regression war für das Ableiten des Alters besser geeignet als die SVM. Zudem stellte sich heraus, dass die zusätzliche Betrachtung indirekter Freunde zu keiner Verbesserung führt. Es war außerdem zu erkennen, dass die Ergebnisse beim Einbeziehen von mindestens fünf direkten Freunden besser waren als bei mindestens drei.

Es konnte jedoch keine allgemeine Aussage hieraus abgeleitet werden, da die Testmengen unterschiedlich groß waren. Die zuvor durchgeführte Bereinigung der Daten und die Einordnung in Kategorien führte nur beim Bestimmen des Wohnorts zu besseren Resultaten. Dieser Aufwand war bei der Vorhersage des Geschlechts und des Alters nicht nötig gewesen.

Es wurden außerdem Assoziationsregeln auf TV-Sendungen gelernt. Hierbei konnten 21 Regeln gefunden werden, die einen Support von mindestens 10 Prozent, eine Konfidenz von mindestens 40 Prozent und einen Lift größer als eins besitzen. Es konnten keine Regeln extrahiert werden, die sowohl hohe Support-, Konfidenz- als auch Liftwerte besitzen.

Fazit

Aus diesen Erkenntnissen lässt sich schlussfolgern, dass neben privaten Profil-Informationen auch die Freundschafts-Beziehungen vor der Öffentlichkeit verborgen werden sollten, um die eigene Privatsphäre zu schützen. Hierbei müssten jedoch beide Freunde ihre Privatsphäre-Einstellungen in Facebook ändern, um Fremden nicht die Möglichkeit zu geben, die Beziehung indirekt über die andere Person herausfinden zu können.

Literatur

- [1] M. Ester und J. Sander, Knowledge Discovery in Databases - Techniken und Anwendungen, Springer, 2000.
- [2] U. M. Fayyad, G. Piatetsky-Shapiro und P. Smyth, „From Data Mining to Knowledge Discovery: An Overview,“ in Advances in Knowledge Discovery and Data Mining, Menlo Park, CA, USA, American Association for Artificial Intelligence, 1996, pp. 1-34.
- [3] U. M. Fayyad, G. Piatetsky-Shapiro und P. Smyth, „From Data Mining to Knowledge Discovery in Databases,“ AI Magazine, Bd. 17, Nr. 3, pp. 37-54, 1996.
- [4] K. L. Taylor, „Oracle Data Mining Concepts, 11g Release 2 (11.2), E16808-06,“ Juli 2011. [Online]. Available: http://docs.oracle.com/cd/E11882_01/datamine.112/e16808.pdf.
- [5] K. L. Taylor, „Oracle Data Mining User's Guide, 12c Release 1 (12.1), E17693-13,“ Mai 2013. [Online]. Available: http://docs.oracle.com/cd/E16655_01/datamine.121/e17693.pdf.
- [6] K. L. Taylor, „Oracle Data Mining Application Developer's Guide, 11g Release 2 (11.2), E12218-07,“ Juli 2011. [Online]. Available: http://docs.oracle.com/cd/E11882_01/datamine.112/e12218.pdf.
- [7] K. L. Taylor, „Oracle Data Mining Concepts, 12c Release 1 (12.1), E17692-13,“ Mai 2013. [Online]. Available: http://docs.oracle.com/cd/E16655_01/datamine.121/e17692.pdf.
- [8] DBpedia. [Online]. Available: <http://dbpedia.org>.
- [9] Freebase. [Online]. Available: <http://www.freebase.com/>.

Bianca Böckelmann

bianca.boeckelmann@robotron.de



Customer Analytics: Neues Buzzword oder effektives Steuerungsinstrument?

Marco Poloni, IBM Global Business Services

Die Google-Suche nach „Customer Analytics“ liefert mehr als 60 Millionen Ergebnisse. Das Thema ist also nicht unbekannt und in der Fachliteratur sehr präsent. Alle Big Player der IT-Branche (Software und Beratung) haben das Thema auf der Tagesordnung und bieten neue Angebote/Produkte in diesem Bereich. Kunden-Analysen sind demnach nichts Neues – doch was ist heute anders als in der Vergangenheit?

Der Artikel stellt Möglichkeiten der Analyse vor, die vor wenigen Jahren noch undenkbar waren, weil sie nur Großkonzernen zur Verfügung standen oder weil es Datenquellen wie Social Media noch nicht gab, und die Integration der Ergebnisse (teilweise in Echtzeit) in die Geschäftsprozesse. Damit ist eine effektivere Steuerung der Front-Office-Prozesse und in Teilen der Produktentwicklung möglich. Fallstudien zeigen Anwendungen in verschiedenen Branchen, um mögliche Implementierungsbeispiele konkret darzustellen.

Die fachliche Herausforderung

Durch die wachsende Digitalisierung der Geschäftsprozesse und der Kommunikation ist es heute möglich, fast die komplette Kunden-Interaktion und sämtliche Kunden-Stammdaten für Analysezwecke verfügbar zu haben – wobei natürlich angenommen wird, dass alle Datenschutz-

Rahmenbedingungen erfüllt sind. Auf diese Art und Weise sind neue Datentöpfe entstanden, die auch durch das Internet ein enormes Wachstum erreichen (Social Media, Blogs, Foren etc.). Hinzu kommt die immer weiter steigende Erwartungshaltung der Kunden gegenüber Unternehmen in Bezug auf personalisierte und relevante Ansprache. Folgende wesentliche Aspekte ändern die Analyse im Vergleich zur Vergangenheit:

- Es ist wesentlich preiswerter, große Datenmengen zu speichern
- Analyse-Methoden und Tools für größere Datenmengen sind für viele Unternehmer bezahlbar geworden (auch als Open Source beziehungsweise Cloud-Anwendung verfügbar)

Beide Punkte sind unter dem Begriff „Big Data“ zusammengefasst. Darüber hinaus gilt:

- Kunden verwenden eine Vielzahl von Kommunikations- und Vertriebskanälen gleichzeitig
- Die Akzeptanz der Kundendaten und deren Analyse als „Asset“ für das Unternehmen ist auch in der Vorstandsetage vorhanden

Der CIO einer amerikanischen Versicherung formuliert diese Akzeptanz sehr passend als „Amazon Effect“. Seine Vorstandskollegen sind auch „smart consumers“, die mit Smartphone und iPads arbeiten und Vorschläge von Amazon erhalten, da sie dort schon verschiedene Produkte gekauft haben.

Die Wichtigkeit der Verwendung von Daten, um effizient und innovativ mit Kunden zu interagieren und Produkte zu entwickeln, wird als Wettbewerbsfaktor klar verstanden. Die Tatsache, dass die Technologie gleichzeitig preisgünstig und stabil geworden ist,

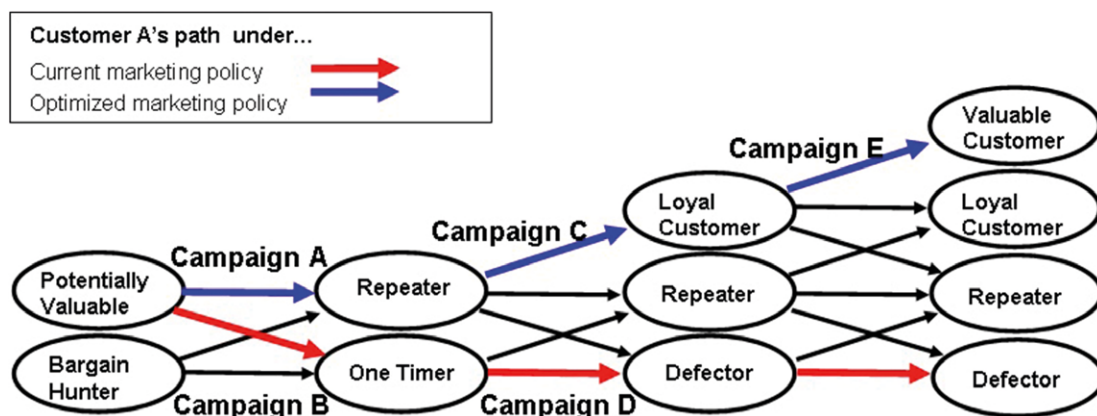


Abbildung 1: Customer-Life-Time-Value-Optimierung

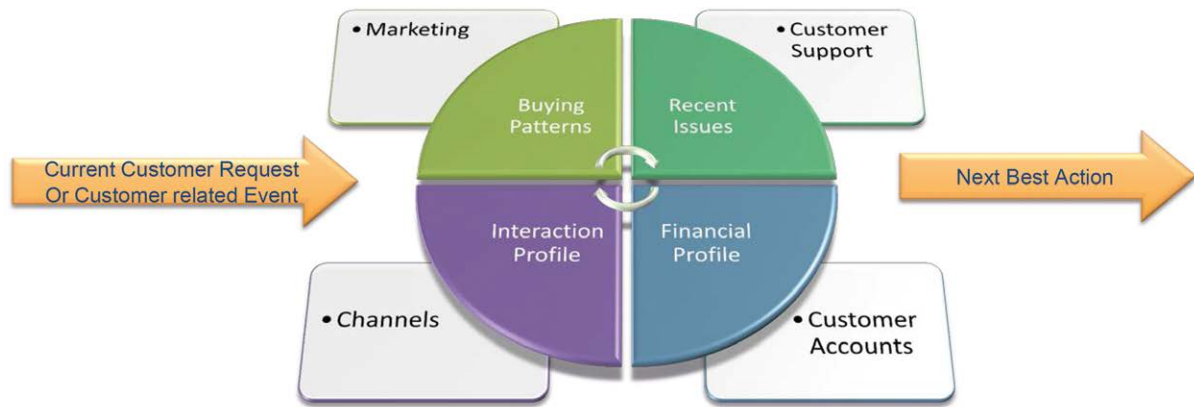


Abbildung 2: Kundensicht steuert Kunden-Interaktion

ist ein Argument mehr, um Business Cases für neue Systeme beziehungsweise Geschäftsprozesse zu entwickeln. Sogar sogenannte „Expertensysteme“ wie IBM Watson werden ernsthaft für sehr unterschiedliche Branchen und Geschäftsprozesse berücksichtigt.

Wenn man es gewohnt ist, mit seinem Smartphone über Sprache zu kommunizieren (Siri in iOS oder Skyvi in Android), ist die Vorstellung eines Expertensystems in einem Call Center oder in einer Beratungsstelle nicht mehr sehr weit. Die Herausforderung ist dann fachlicher Natur: Wie kann man mithilfe der vorhandenen Daten beziehungsweise Analyse-Ergebnisse seine Kundenbeziehungen optimieren und den Kundenwert steigern? Analyse-Methoden sind bereits vorhanden (etwa Markov-Deci-

sion-Prozesse gekoppelt mit Reinforcement Learning in dem IBM Next Best Action Optimizer), die durch Vorschläge für die bestfolgende Kunden-Interaktion (Next Best Action) eine Implementierung der Kundenwert-Optimierung unterstützen (siehe Abbildung 1). Es wird somit eine Kundenwert-Optimierungsstrategie entwickelt, die gleichzeitig folgende Bereiche berücksichtigt:

- Kunden-Historie
- Kunden-Profil
- Kunden-Segmentierung
- Unternehmensziele
- Ressourcen (wie Kanalkapazitäten)
- Geschäftsregeln

Eine zusätzliche Herausforderung ist die Entwicklung sogenannter „Data

Driven Products“ – also statt die Vermarktung der existierenden Produkte und die Kundenbeziehung durch die Datenanalyse zu optimieren, werden die Datenanalyse und die Erzeugung von Daten als Startpunkt für Produktentwicklung und Kunden-Interaktion genutzt. Einige innovative Ansätze sind im Markt schon vorhanden, beispielsweise eine Versicherungspolice über alle Geräte und Leitungen im Haushalt, gekoppelt mit einem Portal, über das der Kunde die Verwendungsdaten seiner Geräte eingibt, um somit Vorschläge für die optimale Wartung und Nutzung (speziell hinsichtlich Energiesparen) zu bekommen. Eine bessere Verwendung der Geräte reduziert die Risiken der Versicherung und gleichzeitig sind durch die Kun-

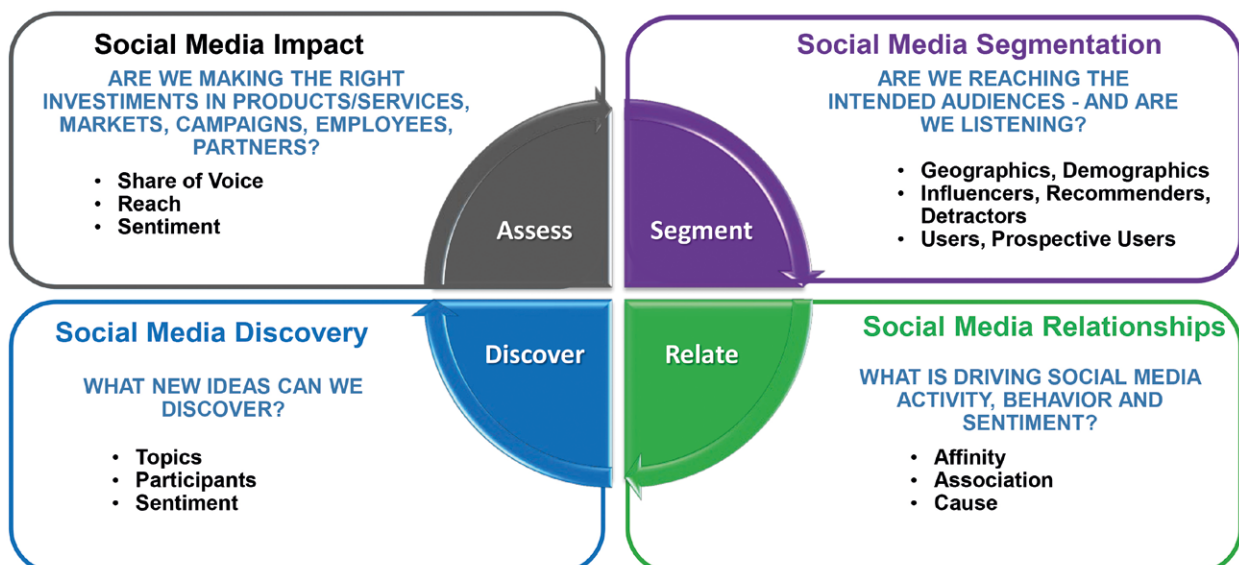


Abbildung 3: Die Bereiche der Social-Media-Analyse

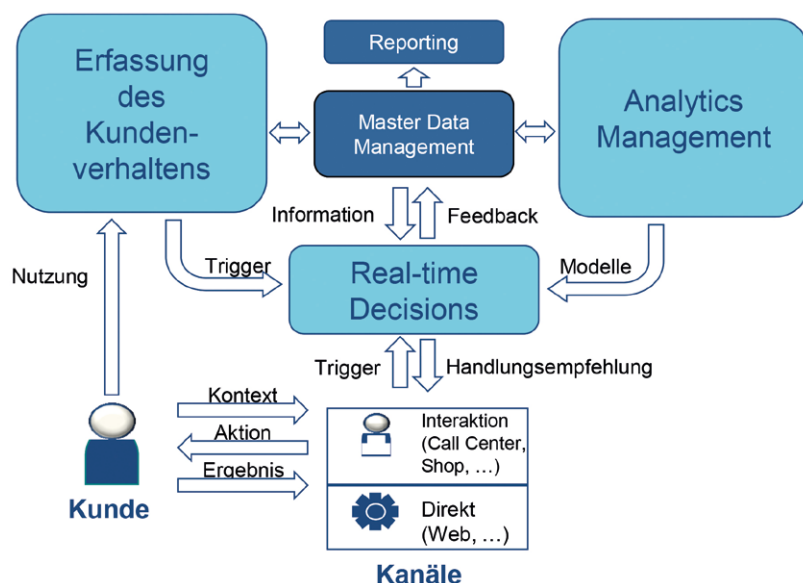


Abbildung 4: Logische Architektur

den-Analyse die Anpassung der Deckungen und die Änderungen an den Tarifen gesteuert.

In einem anderen Beispiel wird durch ein kleines Gerät im Auto die Autoversicherung nach Kilometern bezahlt und nicht mehr pauschal. Die gesammelten Daten stehen dem Versicherungsnehmer in Form eines Online Dashboard zur Verfügung, um Fahrverhalten und Autoverwendung zu analysieren – und für das Unternehmen eine genaue Steuerung neuer Tarife zu ermöglichen. Diese „Pay As You Go“-Modelle in der Kfz-Versicherung sind in Europa bei verschiedenen Versicherungen bereits im Test.

Datenquellen und Analyse-Möglichkeiten

Die Datenquellen sind sehr vielfältig geworden und umfassen auch Daten, die in der Vergangenheit nicht zur Verfügung standen, weil sie zu umfangreich oder digital nicht verfügbar waren (wie Textbausteine in Call Centern oder Daten aus TK-Anlagen). Es ist strategisch wichtig, durch die Integration der Datenquellen zu einer umfassenden Sicht des Kundenverhaltens zu kommen. In Echtzeit wird somit die Interaktion mit dem Kunden gesteuert und jede abgeschlossene Interaktion nicht nur aufgenom-

men, sondern auch als Erfahrung bei der nächsten Interaktion mit diesem Kunden und mit allen Kunden mit ähnlichem Profil berücksichtigt (siehe [Abbildung 2](#)).

Ein Hype-Thema ist die Analyse von Social-Media-Daten (Facebook, Twitter etc.). Aber welche Informationen kann ein Unternehmen tatsächlich aus Social-Media-Analysen gewinnen? Wie in [Abbildung 3](#) dargestellt, betreffen die Informationen hauptsächlich Unternehmens- beziehungsweise Produkt-Wahrnehmung. Diese Informationen sind für Produkt-Entwicklung/-Anpassung verwendbar, indirekt aber nur für Kampagnen-Management (wie Anpassungen der Kommunikation bezüglich Produkten/Dienstleistungen).

Eine interessante Entwicklung, die aber noch nicht sehr verbreitet ist, ist die Verwendung von „Facebook Applications“. Um eine Facebook-Anwendung zu nutzen, ist explizit ein Opt-in notwendig. Danach sind der Kunde und das Unternehmen „Freunde“ – d.h. das Unternehmen bekommt Zugriff auf die Kundendaten in Facebook und kann den Kunden auch direkt ansprechen.

Speziell für die neue Generation (die „Digital Natives“) eine Möglichkeit, einen zusätzlichen Kommunikationskanal zu öffnen.

Integration in die Geschäftsprozesse

Eine Schlüsselrolle in der Anwendung von Customer Analytics ist die Integration in die Front-Office-Prozesse – alle Kunden-Interaktionen, inbound oder outbound, sollen mit einer einheitlichen Kunden-Profilierung und Ansprachen-Alternativen bestückt werden. So werden die Kommunikation und das Angebot für den Kunden konsistent über alle Interaktionspunkte hinweg gehalten. Kanal-abhängige Abweichungen sind dann auf bewusste Marketing-Entscheidungen zurückzuführen und nicht auf organisatorisch bedingte Einschränkungen. [Abbildung 4](#) zeigt die ideale logische Architektur für eine vollständige Integration der Kunden-Analyse in die Unternehmensprozesse.

Die typische Herausforderung in einer kundenzentrierten Architektur ist der Aufbau einer Strategie, um die notwendigen Ergänzungen/Anpassungen vorzunehmen. Eine sorgfältige Planung anhand von Use Cases, Mehrwert für den Geschäftsprozess beziehungsweise Return On Investment und technische Komplexität sind wichtige Voraussetzungen, um die ersten Schritte zu priorisieren. [Abbildung 5](#) zeigt ein Beispiel aus der IBM-GBS-Vorgehensweise.

Die folgenden drei Fallstudien zeigen Beispiele von Customer Analytics in den drei unterschiedlichen Branchen Telekommunikation, Banken und Handel. Gemeinsame Basis ist das Bestreben, den Kunden mit der richtigen Botschaft zum richtigen Zeitpunkt zu erreichen, unter Berücksichtigung aller relevanten betriebswirtschaftlichen Kennzahlen (Kundenwert, Kanalkosten, Unternehmensziele etc.).

Fallstudie 1: Abwanderungsprävention für Prepaid-Kunden in der Telekommunikation

Das Unternehmen hatte folgende Herausforderungen:

- Die Kundenbasis ist hauptsächlich im Prepaid-Bereich
- Sehr kompetitives Umfeld
- Das Marketing-Budget sollte gezielt eingesetzt werden
- Der Kampagnen-Erfolg ist sehr schwierig zu ermitteln

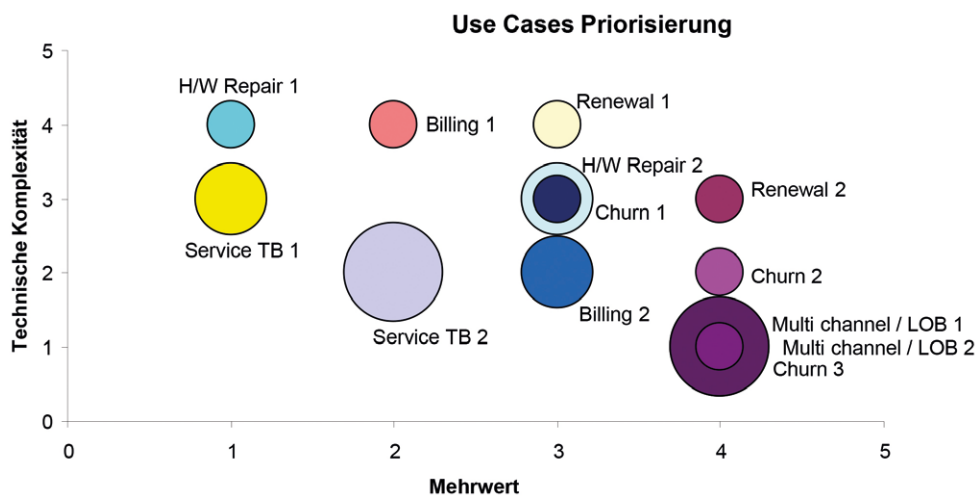


Abbildung 5: Priorisierung der Use Cases für die Implementierung

- Die Kampagnen für Abwanderungsprävention beziehungsweise für Cross-/Up-Selling sollten sehr zeitnah stattfinden, um die Angebote zum richtigen Zeitpunkt wirken zu lassen (typischerweise per SMS)

Der Lösungsansatz:

- Near-Real-Time-Analyse der CDR-Daten, um besondere Nutzungsprofile zu entdecken und den laufenden Kampagnen zuzuordnen
- Das Kampagnen-Ergebnis klar messen und diese Daten als Basis für die Predictive-Analyse-Methoden verwenden

Aus den ersten Ergebnissen mit Predictive Analytics wird erreicht:

- Eine Steigerung der Response Rate bei Kampagnen von etwa 25 auf 50 Prozent
- Eine Abwanderungsreduktion von 15 bis 20 Prozent
- CDRs werden innerhalb von 30 Sekunden analysiert, vorher brauchte man einen Tag dafür

Fallstudie 2: Next Best Action in Banking

Die Bank braucht eine Lösung für folgende Aktionen:

- Kundenverhalten auf Einzelkunden-Ebene analysieren
- Schnell auf Veränderungen reagieren

- Das beste Angebot identifizieren
- Die optimale Nutzung der Kommunikationskanäle (Kapazität, Kosten etc.) gewährleisten

Der Lösungsansatz:

- Eine „Event Detection Engine“ erkennt relevante Änderungen im Kundenverhalten
- Predictive-Analytics-Modelle schlagen eine Handlungsweise vor, basierend auf dem Kundenprofil und vorherigen Interaktionen mit diesem und ähnlichen Kunden
- Vorhandene Kommunikationskanäle werden optimal eingesetzt, um die Handlung durchzuführen

Das Ergebnis:

- Conversion Rates für Sales Leads mehr als 30 Prozent
- Reduzierte Kosten durch gezielte Kunden-Ansprache
- 25 Prozent durchschnittliche Steigerung beim Produktumsatz
- 10 Prozent weniger Abwanderung
- Erhöhte Mitarbeiter-Zufriedenheit durch effiziente Geschäftsprozess-Gestaltung

Fallstudie 3: Kunden-Analysen zur Steuerung der Angebote und Kommunikationskanäle im Retail

Das Problem:

- Optimierung der Marketing-Aktionen auf Einzelkundenbasis über alle Kanäle hinweg
- Bessere Segmentierung basierend auf der gesamten Kundenkontakt-Historie
- Optimierung der Aktionen hinsichtlich des Kundenwerts (Customer Life Time Value)

Der Lösungsansatz:

- Kampagnen werden als Aktionen dargestellt
- Analyse und Berücksichtigung der Responses auf Einzelkunden-Ebene
- Analyse-Erkenntnisse werden verwendet, um den optimalen Kampagnen-Mix für einen Einzelkunden vorherzusagen

Das Ergebnis:

- Proof of Concept zeigt eine Umsatzsteigerung von 20 bis 30 Prozent
- Durchlaufzeit für eine neue Aktion ist um drei Wochen reduziert

Fazit

Customer Analytics ist ein strategisches Instrument, um die Interaktion mit den Kunden effektiv zu steuern. Ein integrierter Kreislauf von Informationen ermöglicht nicht nur eine optimale Steigerung des Kundenwerts, sondern auch eine effiziente Anpassung und Entwicklung von Produkten und Dienstleistungen. Das Unternehmen kann durch die Analyse nicht nur die bestehenden Produkte optimal vermarkten und die Kunden-Zufriedenheit erhöhen, sondern neue Produkte aufgrund der gewonnenen Erkenntnisse entwerfen.

Marco Poloni
marco.poloni@de.ibm.com





Das Berufsbild „Data Scientist“

Angi Voss, Fraunhofer IAIS

Dieser Artikel stellt den Data Scientist vor und gibt einen Überblick über dessen Aufgaben, Kompetenzen, Werkzeuge sowie Aus- und Weiterbildungsmöglichkeiten.

Parallel zum neuen Technologiefeld „Big Data“ entwickelt sich derzeit das Berufsbild der Data Scientists. Diese machen „Big Data“ zu „Smart Data“. Mit Erfindungsreichtum halten sie Ausschau nach verschiedensten Datenquellen, führen sie zusammen, explorieren die Daten und entwickeln Ideen, um sie gewinnbringend zu nutzen. Dazu benötigen sie einschlägiges Fachwissen –

zum Beispiel in Handel, Gesundheitswesen, Finanzwirtschaft, Energiewirtschaft oder Produktion und Logistik. Data Scientists müssen ihr Unternehmen gut kennen, um richtig einschätzen zu können, wo datenbasierte Erkenntnisse zu mehr Effizienz im operativen Betrieb, zu besseren Plänen, besseren Produkten oder gar neuen Dienstleistungen und Geschäftsmodellen führen können.

Mit den identifizierten Datenquellen – seien sie noch so ungewöhnlich, vielfältig, umfangreich oder schnelllebig – müssen Data Scientists effizient umgehen können. Hier sind IT-Kenntnisse im skalierbaren Daten-Management und in der Parallelverarbeitung gefragt. Cloud Computing, Stream-Processing, MapReduce, SQL- und NoSQL-Datenbanken sind Technologi-



Specialized Oracle Database



Datenbanken mit iQ



- IT-Consulting
- Schulungen
- APEX und PL/SQL
- Telefon-/Remotesupport
- Lizenzen

www.munisoft.de · info@munisoft.de

en, die sie einschätzen und einsetzen können.

Zur Entwicklung einer Anwendung gehört die Analyse der verfügbar gewordenen Daten. Data Scientists explorieren die Daten und ihre Qualität und konzipieren die Verarbeitungsschritte, die in der Anwendung später vollautomatisch durchgeführt werden. Dazu gehören Daten-Bereinigung, -Vorverarbeitung, -Fusion und -Anreicherung. Mit den vorhandenen Daten trainieren und evaluieren sie prädiktive statistische Modelle.

In der Anwendung erlauben sie, in den neuen Daten Muster zu erkennen, bestimmte Arten von Informationen zu extrahieren, Entscheidungen zu treffen und Prognosen aufzustellen. Hierzu beherrschen Data Scientists klassische statistische Verfahren und Methoden aus dem Data Mining und dem maschinellen Lernen. Sie nutzen SQL und Skriptsprachen wie R oder Python. Darüber hinaus verlangen digitalisierte Daten, Bilder und Videos nach speziellen Methoden; Textdaten erfordern zum Beispiel eine mehrstufige Vorverarbeitung, natürliche Sprachverarbeitung und Modellierungstechniken für hochdimensionale Merkmalsräume.

Bei der Analyse und Interpretation unüberschaubarer Datenmengen spielt die interaktive Visualisierung eine wichtige Rolle. Gleich am Anfang hilft dem Data Scientist die visuelle Exploration der Daten, später das visuelle Debugging bei der Beurteilung und Verbesserung der Modelle und schließlich das visuelle Reporting bei der Kommunikation der Ergebnisse.

Neben Informationsgewinn und Skalierbarkeit sind bei der Bewertung einer Big-Data-Anwendungsidee Datensicherheit und Datenschutz wichtige Faktoren. Für individualisierte Kundenansprache, Produkte und Dienstleistungen finden sich bereits heute viele Anwendungsbeispiele aus den USA, aber längst nicht alle lassen sich mit dem deutschem Datenschutz in Einklang bringen. Hierzu wird künftig in Deutschland eine Reihe von Forschungsaktivitäten gefördert, über deren Ergebnisse sich Data Scientists auf dem Laufenden halten sollten. Zunehmend etablieren sich auch Datenanbieter, die proprietäre oder offene Daten

bereinigen und gesetzeskonform aggregieren. Solche Plattformen und Dienste sollten Data Scientists kennen, da die Kombination von externen und internen Daten besondere Potenziale bietet.

Data Science ist Teamarbeit und macht Spaß

Kaum jemand wird sich in allen Gebieten gleich gut auskennen, sodass Data Scientists in der Regel in Teams arbeiten, die interdisziplinär zusammengestellt sein können. Hier finden sich neben Informatikern, Betriebswirten, Statistikern und Mathematikern auch Physiker und Sozialwissenschaftler sowie Wissenschaftler aus weiteren Anwendungsdisziplinen.

Der Harvard Business Review sprach im Jahr 2012 vom „Sexiest Job of the 21st Century“ [1]. Es fallen Begriffe wie Künstler, Hacker-Mentalität, Daten-Jiu-jitsu, Kreativität, Unternehmertum und holistische Herangehensweise. Interdisziplinäre Data-Scientist-Teams bilden hoch attraktive Arbeitsumgebungen.

Das Schulungsangebot wächst

Eine Studie von McKinsey [2] hat bereits im Jahr 2011 eine große Nachfrage nach Data Scientists in den USA prognostiziert. Als Reaktion darauf entstand eine Vielzahl von Studienangeboten im Bereich der Datenanalyse. In Europa rechnet man mit einem Bedarf von 300.000 in den nächsten Jahren [3].

Zwar gibt es in Deutschland noch keine speziellen Studiengänge für Data Scientists, jedoch inzwischen eine Reihe von Seminaren an den Hochschulen und kostenpflichtige Fortbildungsangeboten für Berufstätige. Die Lehrinhalte umfassen inhaltlich meist drei Bereiche:

- *Infrastruktur und Datenmanagement*
Typische Inhalte im Bereich „Infrastruktur und Datenmanagement“ sind Cloud Computing, verteilte Systeme und Parallelisierung, MapReduce und das Hadoop-Ecosystem, Stream- und Event-Processing, approximative Algorithmen, Graphenverarbeitung, Datenintegration, -ex- und -import.
- *Analytik*
Im Bereich „Analytik“ geht es um Daten-Aufbereitung und -Qualität,

Statistik, fortgeschrittene Analytik, Modellierung, Data Mining und maschinelles Lernen, R, Mahout, Bezug zu BI, Textanalyse und Visualisierung.

- *Anwendung*

Themen im anwendungsspezifischen Teil der Seminare sind Einsatzmöglichkeiten, Fallbeispiele, Praxisberichte, Wirtschaftlichkeit, Geschäftsmodelle, Datensicherheit, -governance und -schutz sowie Vorgehensweisen im Unternehmen.

Das Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS bietet seit Anfang 2013 Schulungen zu Big Data Architecture, Basic Analytics, Text Analytics und Visual Analytics an [4]. Das Angebot wird im Jahr 2014 im Rahmen der Fraunhofer-Initiative „Big Data“ (siehe www.big-data.fraunhofer.de) erweitert.

Der Auf- und Ausbau von Qualifizierungsprogrammen für Data Scientists wird in den nächsten Jahren öffentlich gefördert, zum Beispiel als Teil eines Fünf-Punkte-Plans für Big Data in Frankreich, im Horizon-2020-Programm der Europäischen Kommission und voraussichtlich auch durch die Big-Data-Kompetenzzentren, die das Bundesministerium für Bildung und Forschung in Deutschland ausgeschrieben hat.

Quellen:

- [1] <http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>
- [2] www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation
- [3] www.journaldunet.com/solutions/emploi-rh/plan-numerique-big-data-0713.shtml
- [4] www.iais.fraunhofer.de/data-scientist.html

Angi Voss

angelika.voss@iais.fraunhofer.de



Ereignissen auf der Spur: Big-Data-Analysen in der Datenbank

Alfred Schlaucher, ORACLE Deutschland B.V. & Co. KG

Viele Analysen im Big-Data-Umfeld suchen in den Unternehmensdaten nach Zusammenhängen, die durch einfaches Aggregieren nicht zu finden sind. Weit verbreitet sind sogenannte „Pattern-Analysen“. Die einfachsten Analysen dieser Art verwenden das Suchen von Strings.

Viele Map-Reduce-Programme der ersten Stunde suchen und zählen Strings in Unternehmensdaten oder zugekauften Daten. Man entwickelte eine Liste von bewerteten Begriffen und zählte die Häufigkeit der Verwendung solcher Begriffe in Blogs, Mails oder Presseartikeln beispielsweise in Verbindung mit den Namen der eigenen Produkte. Diese „Häufigkeiten“-Analyse-Art berücksichtigt jedoch nicht die Beziehungen oder Abhängigkeiten, die zwischen den einzelnen Treffern (Vorkommnissen, Ereignissen oder nennen wir sie hier schon mal „Events“) stattfinden:

- Gibt es Abhängigkeiten zwischen verschiedenen Entwicklungen?
- Welche Rolle spielt die Zeit bei der Abfolge einzelner Events?
- Treten unterschiedliche Events zeitlich zusammenhängend auf?

Zum besseren Verständnis sind solche Fragen auf folgende konkrete Beispiele übertragen:

- Ist am Verlauf von Aktienkurs-Änderungen ein wiederkehrendes Muster

zu erkennen, mit der Möglichkeit, Prognosen zu treffen?

- Kann man anhand der Abfolge von Seitenaufrufen eines Web-Shop-Benutzers erkennen, warum bestimmte Abfolgen zum Kaufabschluss führen und andere nicht?
- Kann man anhand von Sensordaten einer Fließband-Maschine zu Stromverbrauch, Drehzahl, Temperatur etc. ein typisches Muster im zeitlichen Verlauf erkennen, um Ausfälle vorhersagen zu können?

Alle diese Analysen haben folgende Gemeinsamkeiten:

- Es werden mehrere Merkmale erhoben und in Bezug zueinander gesetzt
- Es wird eine Serie der Merkmale gemessen
- Meist orientiert sich diese Serie an einem zeitlichen Verlauf

Ein erstes Beispiel

Zum allgemeinen Verständnis stellt das Beispiel den Verlauf eines Aktienkurses vor. Der Wert einer Aktie wird täglich festgestellt. **Tabelle 1** zeigt die

einzelnen Datensätze. Beim Betrachten des zeitlichen Verlaufs kann man eine Entwicklung des Aktienkurses feststellen (siehe **Abbildung 1**).

Die Entwicklung ist von längeren Aufwärtsphasen und mehr oder weniger abrupten oder steilen Abwärtsbewegungen geprägt. In der unsortierten Ansammlung der massenhaften Datensätze einer Tabelle kann man diese Entwicklung nur schwer erkennen.

Das Beispiel ist überschaubar. Man stelle sich aber nicht 30 Sätze pro Monat und Aktie vor, sondern Sätze bezo-

SYMBOL	ZEIT	PREIS
ORCL	10.12.13	32
ORCL	11.12.13	32
ORCL	12.12.13	32
ORCL	13.12.13	33
ORCL	19.12.13	34
ORCL	09.12.13	31
ORCL	08.12.13	30
ORCL	30.12.13	35

Tabelle 1: Beispiel Aktienkurse

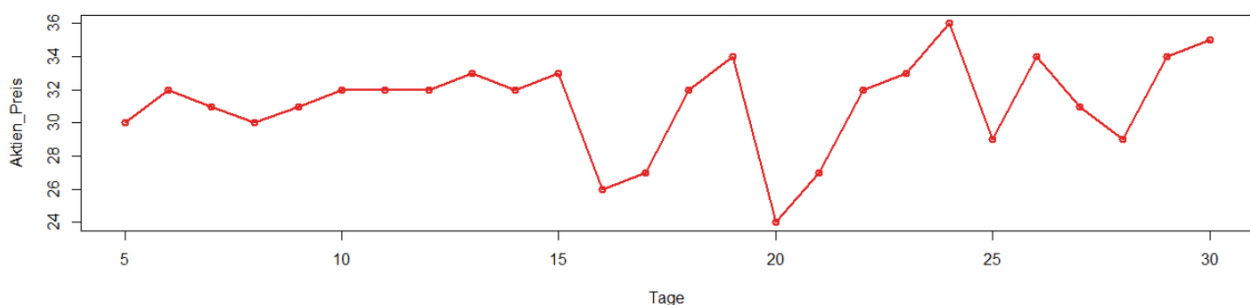


Abbildung 1: Der Aktienkurs aus der Beispiel-Tabelle über die Zeit visualisiert

```

SELECT * FROM Aktien
MATCH_RECOGNIZE (
  PARTITION BY SYMBOL ORDER BY Zeit
  MEASURES
    match_number()           AS Pattern_Nr,
    STRT.zeit                 AS Aufsetz_zeit,
    first(DOWN.zeit)         AS Start_Down_Zeit,
    LAST(DOWN.zeit)          AS Ende_Down_Zeit,
    LAST(UP.zeit)            AS Ende_Up_zeit,
    count(Down.zeit)         AS anzahl_downs,
    count(Up.zeit)           AS anzahl_up
  one ROW PER MATCH
  AFTER MATCH SKIP TO LAST UP
  PATTERN (STRT DOWN+ UP+)
  DEFINE
    DOWN AS DOWN.preis < PREV(DOWN.preis),
    UP   AS UP.preis   > PREV(UP.preis)
) MR
ORDER BY MR.symbol, MR.Aufsetz_zeit
/

```

Listing 1

gen auf 24 Stunden mal 365 Tage mal 10.000 Kurse. Wer will in diesen knapp 100 Millionen Sätzen Muster einer Abfolge finden? Dazu ein Hinweis: Es geht hier nicht um klassisches Aggregieren – also die Betrachtung einzelner Sätze und das Aufsummieren von Werten. Hier dreht es sich um Satz-übergreifende Analysen, das Identifizieren von unterschiedlich umfangreichen Gruppen in einer Menge verstreuter Sätze und das Analysieren der Sätze bezogen auf einen gruppierenden Zusammenhang. Das tun zum Beispiel Finanzunternehmen, bezogen auf Aktienkurse. Sie haben dafür recht komplexe Routinen entwickelt, die solche Analysen meist außerhalb der Datenhaltung sehr Zeit- und Ressourcen-aufwändig erledigen.

SQL-Pattern-Matching

Das SQL-Pattern-Matching in der Oracle-Datenbank 12c macht diese Art der Ana-

lyse jetzt wesentlich einfacher und vor allem ohne aufwändigen Datentransport innerhalb der Datenbank möglich. Die Analyse ist in SQL als mächtige Funktion „MATCH_RECOGNIZE“ eingebaut, man muss also kein zusätzliches Tool mit zusätzlichen Kosten und zusätzlichem Know-how einsetzen. Die Abfrage in Listing 1 analysiert die bereits vorgestellten Aktien-Datensätze. Tabelle 2 zeigt das Ergebnis.

Die letzten beiden Spalten der Ausgabe verraten bereits ein erstes Analyse-Ergebnis: Über alle erkannten Muster hinweg gibt es mehr Aufwärts- als Abwärtsbewegungen – über die Größe, also die Qualität der Schritte, wird allerdings noch nichts gesagt.

Wie das Ganze funktioniert

Der erste Schritt ist das Sortieren der Daten: Weil das tägliche Wechselspiel der Aktienkurse im Fokus steht, müs-

sen die Tabellensätze für die Verarbeitung in einen zeitlichen Verlauf gebracht werden, und weil der Verlauf bezogen auf eine Aktiensorte interessiert, sind die Sätze nach Aktiensorten zu sortieren. Dies erreicht man über die Klausel: „PARTITION BY Symbol ORDER BY Zeit“.

Im zweiten Schritt muss definiert werden, was ein Muster (Pattern) ist: Fallende und steigende Kurse haben das Muster eines „V“. Die Definition dieses Musters lautet also:

- *Phase des Fallens (down)*
Finde den Zeitpunkt, an dem ein Kurs fällt
- *Phase des Steigens (up)*
Finde den Zeitpunkt, an dem der Kurs steigt

Die grafische Darstellung in Abbildung 1 zeigt dieses „V“ genau sechs Mal.

SYMBOL	PATTERN_NR	AUFSETZ_ZEIT	START_DOWN_ZEIT	ENDE_DOWN_ZEIT	ENDE_UP_ZEIT	ANZAHL_DOWNS	ANZAHL_UP
ORCL	1	06.12.13	07.12.13	08.12.13	10.12.13	2	2
ORCL	2	13.12.13	14.12.13	14.12.13	15.12.13	1	1
ORCL	3	15.12.13	16.12.13	16.12.13	19.12.13	1	3
ORCL	4	19.12.13	20.12.13	20.12.13	24.12.13	1	4
ORCL	5	24.12.13	25.12.13	25.12.13	26.12.13	1	1
ORCL	6	26.12.13	27.12.13	27.12.13	30.12.13	1	2

Tabelle 2

```
PATTERN (STRT DOWN+ UP+)
DEFINE
  DOWN AS DOWN.preis < PREV(DOWN.preis),
  UP AS UP.preis > PREV(UP.preis)
```

Listing 2

findet das System auch andere Muster. **Abbildung 2** stellt den grafischen Verlauf der geänderten Beispieldaten dar. Die Ausgabe zeigt Muster mit zwei zeitlich eingrenzenden Aufwärtsphasen, die durch eine DOWN-Phase eingeschlossen sind (siehe **Tabelle 3**).

```
PATTERN (STRT DOWN+ UP+ UP+ DOWN+)
DEFINE
  DOWN AS DOWN.preis < PREV(DOWN.preis),
  UP AS UP.preis > PREV(UP.preis)
) MR
ORDER BY MR.symbol, MR.Aufsetz_zeit
/
```

Listing 3

Im dritten Schritt gibt die „MEASURE“-Klausel das sortierte Datensatz-Material aus. Zur Verfügung steht jeder Spaltenwert eines jeden Satzes in den einzelnen Pattern-Sets, beispielsweise der letzte Wert der „DOWN“-Phase mit „LAST(DOWN.ZEIT)“. Möglich sind hier „LAST“, „FIRST“, „MAXIMUM“ und „MINIMUM“, aber auch „AVERAGE“ und andere statistische Funktionen, die auf ein Set von Werten anwendbar sind.

In der Syntax der Funktion „MATCH_RECOGNIZE“ wird ein Muster über die „PATTERN“-Klausel und die zugehörige „DEFINE“-Unterklausele festgelegt (siehe **Listing 2**).

Es signalisiert, dass die „PATTERN“-Klausel auch „Regular Expressions“ versteht. Dies verdeutlicht, wie flexibel diese Art der Musterbeschreibung sein kann, denn „Regular Expressions“ können beliebige Kombinationen von Ereignissen beschreiben.

Die „DEFINE“-Subklausele definiert sogenannte „Pattern-Variablen“. In diesem Beispiel sind das „DOWN“ und „UP“. Diesen Variablen stehen für einen oder mehrere Datensätze, die bestimmte Bedingungen erfüllen („DOWN“-Set und „UP“-Set). In diesem Beispiel lautet die Bedingung für „DOWN“, dass der Aktienpreis von zwei aufeinanderfolgenden Sätzen sinkt. Über die Funktion „PREV(...)“ kann man auf den Vorläufersatz innerhalb des „DOWN-Sets“ zugreifen, da man Zugriff auf alle Felder aller zum Set gehörenden Sätze hat.

In dem Beispiel gibt es noch die dritte Variable „STRT“, die nicht über „DEFINE“ beschrieben ist. Als undefinierte Variable referenziert sie jeden Satz. Das bedeutet, dass das darauf folgende Muster „DOWN+“/„UP+“ an beliebigen Stellen in der Tabelle vorkommen kann. Ein Satz des „STRT“-Sets markiert den Beginnsatz eines neuen Musters.

Die „MEASURE“-Ausgabe liefert nur einen Satz pro gefundenem „V“-Muster, also sechs für das erste Beispiel und zwei für das zweite. Das wird durch die Angaben „ONE ROW PER MATCH“ erreicht. Diese Art der Ausgabe ist sinnvoll, wenn die Muster-Eigenschaften als Ganzes angesprochen sind, also etwa die Anzahl der Sätze in der „DOWN-Phase“ durch „COUNT(DOWN.ZEIT)“. Hier ergibt auch die Funktion „MATCH_NUMBER()“ einen Sinn. Sie zählt die einzelnen Pattern-Treffer mit einer laufende Nummer durch.

Die „PATTERN“-Klausel enthält die Kombination dieser Variablen: „UP“ folgt nach „DOWN“. Da es mehrere „DOWN“- oder „UP“-Phasen nacheinander geben könnte, steht hinter den Variablen-Namen das „+“-Zeichen.

In **Listing 3** sind die Beispieldaten und die Abfrage leicht verändert. Über die veränderte „DEFINE“-Klausel

Das Gegenstück dazu ist „ALL ROWS PER MATCH“ – hier wird jeder Einzelwert geliefert. Weil man die Sätze optisch in der Ausgabe nicht mehr so

SYMBOL	PATTERN_NR	AUFSETZ_ZEIT	ANZAHL_DOWNS	ANZAHL_UP
ORCL	1	06.12.13	2	3
ORCL	2	16.12.13	2	4

Tabelle 3

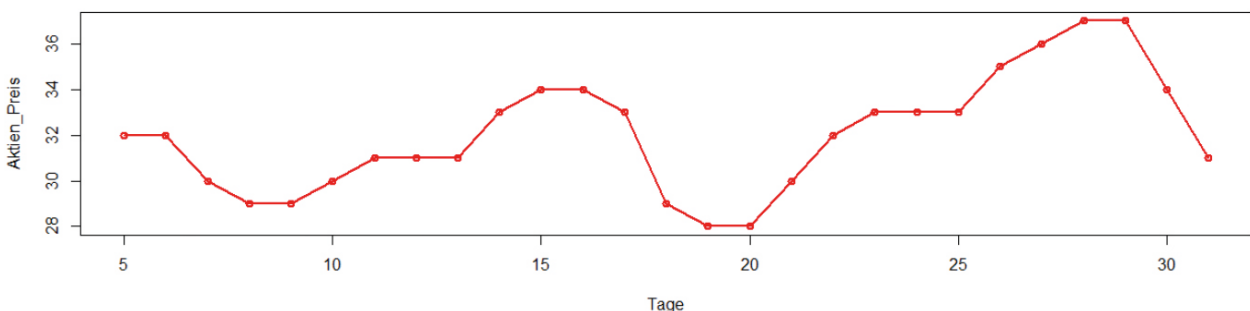


Abbildung 2: Der grafische Verlauf der geänderten Beispieldaten

leicht den einzelnen Phasen zuordnen kann, sollte man die Ausgabesätze mit der Funktion „CLASSIFIER()“ markieren. Sie liefert pro Ausgabesatz den zugehörigen Variablen-Namen, das ist der String „DOWN“ oder „UP“. Zur Ergebnismenge zählen alle Sätze, die einer „DOWN“- beziehungsweise „UP“-Phase zuzuordnen sind.

Nicht enthalten sind die Sätze mit gleichbleibendem Kursverlauf. Eine solche Ausgabesatzmenge kann selbst wieder Input für weitere Analysen sein. So könnte man beispielsweise den Output über eine View-Definition an eine R-Analyse oder eine grafische Aufbereitung weiterreichen.

Ein zweites Beispiel mit aufwändiger Ausgabe

Das nächste Beispiel erweitert den „MEASURE“-Bereich noch etwas. Es werden zusätzliche Ausgaben für den Aktienwert zu Beginn und am Ende der jeweiligen Fall- und Steig-Phasen und die durchschnittliche Fall- und Steig-Stärke pro Tag ausgegeben (siehe Listing 4). Die Möglichkeiten sind hier unbegrenzt, selbst Funktionen oder „CASE“-Ausdrücke sind möglich. Die Tabellen 4 und 5 zeigen die entsprechende Ausgabe.

```
SELECT *
FROM Aktien
MATCH_RECOGNIZE (
  PARTITION BY SYMBOL
  ORDER BY Zeit
  MEASURES
    match_number()           AS Pattern_Nr,
    STRT.zeit                AS Aufsetz_zeit,
    STRT.PREIS              AS Start_Preis,
    LAST(DOWN.PREIS)        AS Tief_Preis,
    first(DOWN.zeit)        AS Start_Down_Zeit,
    LAST(DOWN.zeit)         AS Ende_Down_Zeit,
    LAST(UP.zeit)           AS Ende_Up_zeit,
    LAST(up.preis)          AS Ende_Up_Preis,
    count(DOWN.zeit)        AS anzahl_downs,
    count(UP.zeit)          AS anzahl_up,
  case
    when count(DOWN.Preis) = 1 then
      STRT.Preis - LAST(DOWN.preis)
    when count(DOWN.Preis) > 1 then
      (STRT.Preis - LAST(DOWN.preis))/count(DOWN.PREIS)
  end durchschn_Fallwerte ,
  case
    when count(UP.Preis) = 1 then
      LAST(up.preis) - LAST(DOWN.PREIS)
    when count(UP.Preis) > 1 then
      round((LAST(up.preis) - LAST(DOWN.PREIS))/count(UP.PREIS),1)
  end durchschn_Steigwerte
  one ROW PER MATCH
  AFTER MATCH SKIP TO LAST UP
  PATTERN (STRT DOWN+ UP+)
  DEFINE
    DOWN AS DOWN.preis < PREV(DOWN.preis),
    UP   AS UP.preis  > PREV(UP.preis)
) MR
ORDER BY MR.symbol, MR.Aufsetz_zeit
/
```

Listing 4

SYMBOL	PATTERN_NR	AUFSETZ_ZEIT	START_PREIS	TIEF_PREIS	START_DOWN_ZEIT	ENDE_DOWN_ZEIT	~
ORCL	1	06.12.13	32	30	07.12.13	08.12.13	~
ORCL	2	13.12.13	33	32	14.12.13	14.12.13	~
ORCL	3	15.12.13	33	26	16.12.13	16.12.13	~
ORCL	4	19.12.13	34	24	20.12.13	20.12.13	~
ORCL	5	24.12.13	36	29	25.12.13	25.12.13	~
ORCL	6	26.12.13	34	31	27.12.13	27.12.13	~

Tabelle 4

~	ENDE_UP_ZEIT	ENDE_UP_PREIS	ANZAHL_DOWNS	ANZAHL_UP	DURCHSCHN_FALLWERTE	DURCHSCHN_STEIGWERTE
~	10.12.13	32	2	2	1	1
~	15.12.13	33	1	1	1	1
~	19.12.13	34	1	3	7	2,7
~	24.12.13	36	1	4	10	3
~	26.12.13	34	1	1	7	5
~	30.12.13	35	1	2	3	2

Tabelle 5

Ein drittes Beispiel mit komplexer Mustersuche

Das nächste Beispiel sucht in dem Datenstrom der sich ständig ändernden Aktienwerte nach einem plötzlichen

Fall von mehr als 8 Prozent (definierter Schwellenwert, siehe Listing 5 mit der Tabelle 6 als Ausgabe). Die Beispieldaten-Menge ist die gleiche wie im ersten Beispiel.

Ein viertes Beispiel

Das nächste Beispiel beobachtet die Menge der Aktien, die innerhalb einer bestimmten Zeit verkauft werden. Die Regel lautet: „Suche alle Aktienverkaufs-Events innerhalb einer Stunde, die einen zeitlichen Abstand von weniger als 60 Minuten haben und ein Volumen von mehr als 30.000 Stück ausmachen.“ Wenn solche Muster erkannt werden, kann ein Täuschungsverdacht bestehen (siehe Listing 6).

Der Wert „0.042“ ist der 24ste Teil von 1. Die Subtraktion der beiden Datumswerte liefert das Ergebnis bezogen auf ganze Tage zurück, wobei die Einheit ein Tag ist. Damit entspricht „0.042“ einem Wert von 60 Minuten. Die Alternative wäre „((A.Zeit - FIRST (A.Zeit))*24 < 1)“. Tabelle 7 zeigt die Ausgabe und Tabelle 8 die Daten für dieses Beispiel.

```
SELECT *
FROM Aktien
MATCH_RECOGNIZE (
  PARTITION BY SYMBOL
  ORDER BY Zeit
  MEASURES
    match_number()          as Pattern_Nr,
    A.Zeit                  as Ausgangs_Zeit,
    A.Preis                 as Ausgangs_Preis,
    B.Preis                 as Gefallener_Preis,
    count(C.*)+1           as Anzahl_Tage
  one ROW PER MATCH
  AFTER MATCH SKIP PAST LAST ROW
  PATTERN (A B C*)
  DEFINE
    B as (B.preis - A.preis) / A.preis < -0.08,
    C as C.preis < A.preis)
```

Listing 5

SYMBOL	PATTERN_NR	AUSGANGS_ZEIT	AUSGANGS_PREIS	GEFALLENER_PREIS	ANZAHL_TAGE
ORCL	1	15.12.13	33	26	3
ORCL	2	19.12.13	34	24	4
ORCL	3	24.12.13	36	29	5

Tabelle 6

```
SELECT * FROM Aktien_handel
MATCH_RECOGNIZE (
  PARTITION BY SYMBOL
  ORDER BY Zeit
  MEASURES
    match_number()          as Pattern_Nr,
    FIRST (A.Zeit)         as in_hour_of_trade,
    SUM (A.VOLUMEN)       as sum_of_large_volumes
  one ROW PER MATCH
  AFTER MATCH SKIP PAST LAST ROW
  PATTERN (A B* A B* )
  DEFINE
    A as ((A.volumen > 30000 ) and
          ((A.Zeit - FIRST (A.Zeit)) < 0.042 )),
    B as ((B.volumen <= 30000 ) and
          ((B.Zeit - FIRST (B.Zeit)) < 0.042))
  )
```

Listing 6

SYMBOL	PATTERN_NR	IN_HOUR	SUM_OF_LARGE_VOLUMES
ORCL	1	23.12.13	64300

Tabelle 7

Zeit	PREIS	SYMBOL	VOLUMEN
20.12.13	24	ORCL	1233
21.12.13	27	ORCL	453
23.12.13	32	ORCL	33300
23.12.13	36	ORCL	453
23.12.13	29	ORCL	31000
23.12.13	33	ORCL	4432
26.12.13	34	ORCL	422
27.12.13	31	ORCL	44
30.12.13	35	ORCL	453

Tabelle 8

Ein letztes Beispiel aus dem Bereich der Sensordaten

Einsatzfelder des „Pattern Matching“ gibt es in allen Branchen. Hier noch ein Sensordaten-Beispiel aus der Fertigungsindustrie. Die Maschinen einer Fertigungsanlage werden über eine Vielzahl von Sensoren überwacht. Etwa alle 20 Sekunden entsteht ein neuer Datensatz. Tabelle 9 zeigt einen Ausschnitt.

Für die gemessenen Maschinen gelten Maximalwerte für Temperatur (200°), Drehzahl (1000 Upm) und Stromleistungsaufnahme (1300 Watt). Während des Betriebs überschreiten die Maschinen diese Werte immer wieder, was zu

SATZNR	MESSZEITPUNKT	STUECKNR	DREHZAHL	WATT	TEMPERATUR
499994	30.04.2008 11:04:55	5000	1031	1094	139
499995	30.04.2008 11:05:16	5000	991	1315	159
499996	30.04.2008 11:05:37	5000	1030	1284	205
499997	30.04.2008 11:05:58	5000	882	1182	186
499998	30.04.2008 11:06:18	5000	818	1406	204

Tabelle 9

frühzeitigem Verschleiß und häufiger Wartung führt. In einer Abfrage sollen Zeiten für Konstellationen der drei Kennwerte gefunden werden, in denen alle Werte nacheinander die Höchstmarke überschreiten (siehe Listing 7).

Zunächst sind alle Sätze nach „MaschinenNr“ und „StueckNr“ gruppiert,

also nach dem Werkstück, das eine bestimmte Maschine gerade produziert. Pro Werkstück und Maschine fallen etwa 100 Mess-Sätze an. Sie sind nach der Mess-Zeit und einer künstlichen Satznummer sortiert.

Die „DEFINE“-Klausel sucht für die jeweiligen Kennzahlen („DREHZAHL“,

„WATT“ und „TEMPERATUR“) durch einen Vergleich des aktuellen Werts mit seinem Vorgänger nach einem Anstieg der Werte. Zudem betrachtet man nur die Sätze, in denen eine Überschreitung der Maximalwerte vorliegt (etwa „DREHZAHL > 1000“).

Die „PATTERN“-Klausel bringt die Vorkommnisse „A“, „B“ und „C“ in eine zeitliche Reihenfolge.

Man betrachte also zunächst die Drehzahl, in der zeitlichen Folge die Leistungsaufnahme und dann die Temperatur. Dieses Beispiel deckt nicht ganz die Realität ab, da Drehzahl und Leistungsaufnahme zeitlich wahrscheinlich synchron stattfinden, während der Temperaturanstieg nachhinkt, da eine Maschine nur dann schneller läuft, wenn auch mehr Strom fließt. Die Folge ist ein zeitlich versetzter Temperaturanstieg, der sich über mehrere Sekunden hinziehen kann. An diesem Beispiel ist gut sichtbar, wie mit wenigen Handgriffen die Logik der Analyse schnell und leicht änderbar ist. Auch das Einstellen der Überschreitung der Grenzwerte ist leicht machbar (siehe Listing 8 und Tabelle 10 mit einem Ausschnitt aus der Liste der gefundenen „PATTERN“).

Dieses Beispiel kann immer noch verbessert werden. Es kann sein, dass eine Steigerung der Leistungsaufnahme nicht unbedingt zu einer Erhöhung der Drehzahl führen muss. Im Gegenteil: Wird die Maschine durch einen unvorhersehbaren Vorgang an der Fertigungsstraße stark gebremst oder kommt sie sogar zum Stillstand, steigt der Stromverbrauch enorm, was in kurzer Zeit zur Schädigung der Maschine führt. Listing 9 zeigt das Muster, das eine solche Situation finden müsste.

```
SELECT * FROM Sensor_daten
MATCH_RECOGNIZE (
  PARTITION BY MaschinenNr, StueckNr
  ORDER BY MessZeit, satznr
  MEASURES
    match_number()          as Pattern_Nr,
    FIRST (A.satznr)        as first_Satznr,
    -- satznr                as satznr,
    LAST (A.satznr)         as Last_satznr,
    count(A.satznr)         as count_satznr,
    FIRST (A.drehzahl)      as first_drehzahl,
    LAST (A.drehzahl)       as LAST_drehzahl,
    -- LAST (A.Watt)         as last_A_Watt,
    -- FIRST (b.Watt)        as First_watt,
    LAST (b.Watt)           as LAST_watt,
    LAST (c.temperatur)     as last_temperatur,
    -- c.messzeit           as c_messzeit,
    FIRST (A.messzeit)      as Messzeit
  one ROW PER MATCH
  AFTER MATCH SKIP PAST LAST ROW
  PATTERN (A+ B C)
  DEFINE
    A as (A.drehzahl > PREV(A.DREHZAHL) and DREHZAHL > 1000) ,
    B as (B.watt > PREV(B.WATT) and B.watt > 1280),
    C as (C.Temperatur > PREV(C.Temperatur) and C.temperatur > 245)
)
```

Listing 7

```
PATTERN (A+ C)
DEFINE
  A as (A.drehzahl > PREV(A.DREHZAHL) and
        DREHZAHL > 1020) and
        (A.watt > PREV(A.WATT)) and
        (A.Watt > 1500) ,
  C as (C.Temperatur > PREV(C.Temperatur) and C.temperatur > 245)
```

Listing 8

STUECKNR	PATTERN_NR	LAST_SATZNR	LAST_DREHZAH	LAST_WATT	LAST_TEMPERATUR	MESSZEIT
8	1	720	1040	1547	252	01.01.2008 04:09:52
46	1	4531	1021	1530	250	02.01.2008 02:12:26
46	2	4534	1048	1548	247	02.01.2008 02:13:29
63	1	6220	1033	1524	248	02.01.2008 11:58:35
69	1	6882	1031	1514	252	02.01.2008 15:48:20
105	1	10426	1040	1527	247	03.01.2008 12:18:14
118	1	11728	1035	1528	249	03.01.2008 19:50:05
145	1	14414	1047	1543	248	04.01.2008 11:22:14
156	1	15504	1037	1535	248	04.01.2008 17:40:30
160	1	15910	1032	1527	255	04.01.2008 20:01:24

Tabelle 10

```

PATTERN      ( C A )
DEFINE
  C as (C.Temperatur > PREV(C.Temperatur) ,           -- einfacher Temperaturanstieg
        (A.watt > PREV(A.WATT)) and (A.Watt > 1300) ,   -- unterer Grenzwert / Anstieg Strom
  A as (A.drehzahl <= PREV(A.DREHZAHL)                -- gleiche oder sinkende Drehzahl

```

Listing 9

Vorstufe für zusätzliche Statistiken und Data Mining

Die Ergebnismenge des Beispiels zuvor kann je nach Festlegung der Grenzwerte mehrere Tausend Zeilen umfassen. Es liegt deshalb nahe, diese Menge einer weiteren Analyse zu unterziehen. So könnte man zum Beispiel herausfinden, ob Drehzahl und Temperatur miteinander korrelieren. Dafür eignet sich besonders gut die Sprache „R“, weil damit Analysen in der Datenbank möglich sind, ohne die Pattern-Ergebnisdaten zu bewegen.

Eine Korrelation zwischen allen drei Kennwerten der gefundenen Muster, also „DREHZAHL“, „LEISTUNGS-AUFNAHME“ und „TEMPERATUR“, könnte mit einer Routine leicht durchgeführt werden (siehe Abbildung 3). Als Vorarbeit in der Datenbank fasst eine View-Definition die Ergebniszeilen zusammen (siehe Listing 10). Listing 11 zeigt das Oracle-R-Enterprise-Skript. Es ist hier abgedruckt, um aufzuzeigen, wie schnell und einfach man zu Analyse-Ergebnissen gelangt.

Skalierung

Sensordaten werden oft als Beispiel für Big-Data-Anwendungen angeführt. Dieses Beispiel zeigt jedoch, wie gut

solche Daten in einer klassischen Datenbank aufgehoben sind. Obwohl in einem Abstand von 20 Sekunden immer ein neuer Satz pro Maschine entsteht, kommt man bei einer praxisrelevanten Testreihe kaum über eine Million Sätze. Um eine für Testzwecke höhere Datenmenge von 5 Millionen Sätzen zu erzeugen, wurden für das oben ausgeführte Beispiel über fünf Jahre Daten gesammelt.

Die hier vorgestellten „MATCH_RECOGNIZE“-Funktionen benötigen auf einem Standard-Laptop gerade mal 35 Sekunden Bearbeitungszeit. Durch den Parallelisierungsgrad von zwei (mehr ist auf einem solchen Laptop nicht spürbar), richtig aktualisierte Statistiken und Caching sinkt die Ausführzeit auf unter fünf Sekunden. Am Ausführungsplan der Datenbank erkennt man, dass die Bearbeitung kom-

```

CREATE VIEW SENSOR_PATTERN as
  SELECT Drehzahl, Watt, Temperatur FROM ( Select des vorigen
  Beispiels)

```

Listing 10

```

ore.connect (connect-Daten zum DB-Schema)]
ore.doEval(
  function(param) {
    library(ORE)
    ore.connect(-> DB-Schema)
    ore.sync(DB-Schema)
    ore.attach(SENSOR_PATTERN)
    cor.matrix <- cbind(Drehzahl,Watt,Temperatur)
    rcorr(cor.matrix)
  });

```

Listing 11

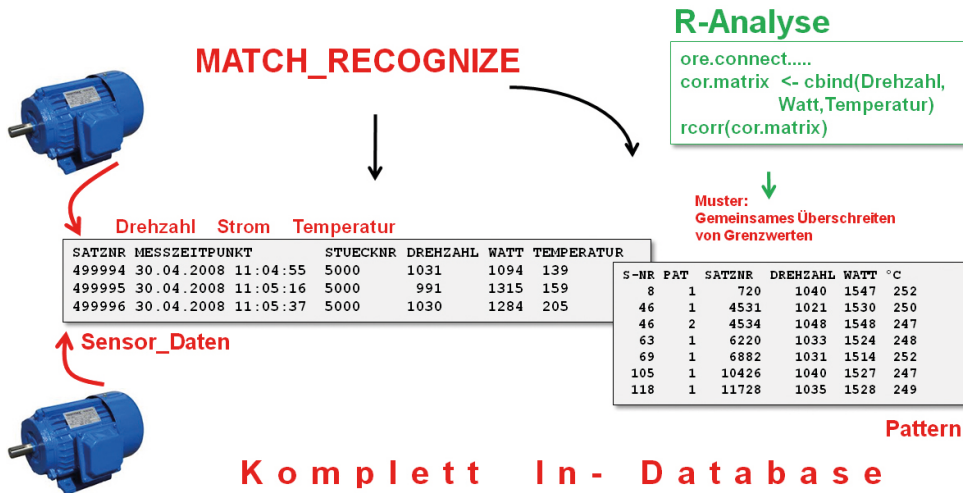


Abbildung 3: Einsatz-Szenario „Pattern-Suche“ – als Input die Korrelations-Analyse mit R

plett im Kern der Datenbank (letztlich im Hauptspeicher) abläuft (siehe Abbildung 4). Natürlich wird die Komplexität der Anwendung in der Praxis steigen. Dann verwendet man allerdings auch keinen Laptop, sondern einen größeren Datenbank-Server mit einer echten Parallelisierungs-Option. Damit würde die Pattern-Suche auf unter eine Sekunde Laufzeit sinken und wäre auch für Realtime-Szenarien einsetztauglich.

Streaming und Event-Measurement

Klassische Kennzahlen-Analysen betrachten in der Regel abgelaufene Berichtszeiträume. Sie liefern Tages-, Monats-, Quartals- oder Jahres-Ergebnisse und beantworten Fragen über das Wie, also wie etwas gelaufen ist. Um künftige Geschäftsentwicklungen vorherzusagen oder sogar durch frühzeitiges Erkennen neuer Entwicklungen planen zu können, reichen solche Kennzahlen meist nicht aus. Hierzu ist nach dem Warum zu fragen, also warum etwas geschehen ist.

Über das vorgestellte Pattern-Matching lassen sich Ereignisse und Zusammenhänge zwischen den Ereignissen aufdecken. Hierzu ist regelmäßiges

Beobachten der relevanten Geschäfts-transaktionen nötig (Event-Monitoring, siehe Abbildung 5). Die relevanten Transaktionsdaten werden dazu permanent den Analysen zugeführt (Streaming). Beim Eintreten eines nicht geplanten Ereignisses kann man spontan und rechtzeitig reagieren.

Pattern-Matching ist ideal für diese Verfahren. Allerdings ist „Streaming“ nicht gleich „Streaming“. Es ist nicht damit getan, Daten fließen zu lassen und sich diese Daten beim Vorbeifließen anzuschauen. Drei wichtige Aspekte sind zu berücksichtigen:

1. Es muss Regeln geben. Man kann nur etwas überprüfen, wenn man es mit etwas anderem vergleicht, von dem es dann abweicht oder nicht. Beim vorgestellten Feature sind das die Klauseln „PATTERN“ und „DEFINE“.
2. Es muss einen messbaren Bereich geben. Eine Veränderung (ein Ereignis) muss also auch als solches erkennbar sein. In den genannten Beispielen war das beispielsweise die „PREV“-Funktion innerhalb der „DEFINE“-Klausel. Man könnte sagen, Streaming bedeutet immer:

„Vergleiche den vorhergehenden Wert mit einem nachfolgenden“. Wenn man sich jedoch die möglichen Regeln anschaut, dann ist nicht immer klar, was genau ein Folge-beziehungswise ein Vorgängerwert darstellt. Bei dem Aktien-Beispiel könnte beispielsweise ein Wert nach einer „DOWN“-Phase über mehrere Tage hinweg (mehrere Messungen hinweg) gleich bleiben, ohne in eine „UP“-Phase zu wechseln. Die Ausdehnung des zu messenden Bereichs (die Menge der Sätze) ist also von anderen Faktoren abhängig.

3. Ein historisches Gedächtnis in Form eines Data Warehouse ist erforderlich. Durch das Wissen über das Vergangene lassen sich Verhaltensregeln und Planung modifizieren.

Zumindest der zweite Punkt führt dazu, dass das hier betrachtete Streaming doch keinen kontinuierlichen Datenstrom darstellt, sondern eher eine Kette von gruppierten Sätzen, die zusammenhängend zu analysieren sind. Das „PARTITION BY“ des „PATTERN MATCHING“ gibt bereits eine solche Gruppierung vor. Für ein kontinuierliches Event-Monitoring muss es darüber hinaus eine Ablaufumgebung geben, die das Pattern-Matching im Kern periodisch aufruft. Diese Datenbank-Techniken unterstützen das Streaming:

- Einsatz des „MATCH RECOGNIZE“ im Select einer Materialized View. Man kann das Refresh der Materialized View auf „ON COMMIT“ setzen, sodass bei Änderung der Ausgangs-Tabelle immer wieder aufs Neue Pattern gesucht werden. Diese Variante ist jedoch nur sinnvoll, wenn sich die Änderungen in der Ausgangs-Tabelle in Grenzen halten.

Id	Operation	Name	Rows	Bytes	Cost (%CPU)	Time
0	SELECT STATEMENT		4999K	166M	34419 (1)	00:00:02
1	VIEW		4999K	166M	34419 (1)	00:00:02
2	MATCH RECOGNIZE SORT DETERMINISTIC FINITE AUTOMATON		4999K	81M	34419 (1)	00:00:02
3	TABLE ACCESS FULL	SENSOR_DATEN	4999K	81M	6911 (1)	00:00:01

Abbildung 4: Ausführungsplan der „MATCH_RECOGNIZE“-Verarbeitung auf fünf Millionen Sätzen

ten. Die Alternative ist ein regelmäßig aufgerufener expliziter Refresh.

- Die Verwendung des „MATCH RECOGNIZE“ in einer „Table Function“. Man erreicht damit eine zusätzliche Parallelität von Pattern-Suche und Weiterverarbeitung bereits gefundener Pattern.

Fazit

Das neue Pattern-Matching der Oracle-12c-Datenbank ist ein genialer Griff in die Möglichkeiten des klassischen SQL. Einerseits löst man aktuelle Aufgabenstellungen, wie sie zum Beispiel im Rahmen des Big Data vorkommen. Andererseits minimiert man die Aufwände auf das Nötigste:

- Die Daten bleiben in der Datenbank
- Datenbank-Ressourcen mit Know-how und auch Infrastruktur werden wiederverwendet
- Komplexe Analysen über definierbare Datenbereiche hinweg sind ohne Programmierung leicht machbar

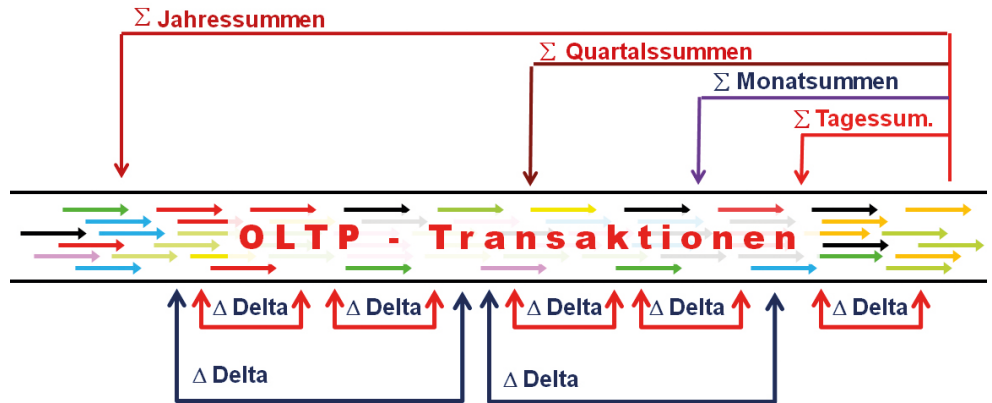


Abbildung 5: Event-Monitoring im Gegensatz zum klassischen Aggregieren vergangener Transaktionen

Alfred Schlaucher
alfred.schlaucher@oracle.com



Hinweis:

Die Beispieldaten zu diesem Artikel können unter „http://www.oracle-dwh.de/downloads/AutoIndex-2.2.4/index.php?dir=Artikel/&file=BSP_Daten_Match_Recognize.zip“ heruntergeladen werden.

PROLICENSE®
OPTIMIZING SOFTWARE ASSETS
Kompetent – Unabhängig – Erfolgsbasiert

SO RICHTIG UNTERLIZENSIERT?

Sprechen Sie mit uns!

Wir sind nur unseren Mandanten verpflichtet.

- > **Compliance sichern**
- > **Audit vermeiden**
- > **Kosten senken**

ProLicense GmbH

Friedrichstraße 191 | 10117 Berlin
Tel: +49 (0)30 60 98 19 230 | www.prolicense.com

Der DBA mit einem Fuß im Gefängnis?

Oliver Pyka, Mitglied in der Datenbank Community und Regionaler Repräsentant von Würzburg

Wir leben in einer Welt voller Vorschriften – aber muss man sie alle kennen? „Ja“, lautet die Antwort, wenn es um die Arbeit geht, da ansonsten sowohl ein Angestellter als auch ein externer Dienstleister Probleme bekommen kann.

Es geht um den Datenschutz. Genauer gesagt um das Bundesdatenschutzgesetz (BDSG), das in der Vergangenheit stark novelliert, um nicht zu sagen verschärft wurde. Dieser Artikel bietet keine juristische Betrachtungsweise, sondern eine praktische. Was muss also ein DBA machen, wenn ihm etwas auffällt? „Bitte kopieren Sie mal die Produktion in die Entwicklung“ – welcher DBA kennt diesen Satz nicht? Wie oft hat er das auch schon durchgeführt, ohne weitere Fragen zu stellen, eben weil es so ist oder er in dem Glauben handelt, er brauche nichts in Frage zu stellen, wenn er eine Anweisung bekommt – erst Recht, wenn diese von einem Teamleiter oder Vorgesetzten beziehungsweise Auftraggeber kommt. Doch genau das ist der Punkt: Er müsste sich fragen, ob in dem zu kopierenden System BDSG-relevante Daten enthalten sind oder nicht; falls solche etwa in der Datenbank in Form von Adressen von Privatkunden enthalten sind, dürfen diese Daten ausdrücklich nicht in ein Entwicklungs-System übernommen werden.

„Das bleibt doch bei uns im Haus“ hört man hier oft als Begründung dafür, dass es doch geht. Das ist aber eben nicht der Fall. Es geht darum, dass Entwickler nicht auf Echt-Daten zugreifen können dürfen, wenn dort schützenswerte Daten zu finden sind. Anders ist dies bei unternehmerischen Daten, sogar bei Betriebsgeheimnissen, denn diese werden nicht durch das BDSG abgedeckt.

Wie sollte ein DBA nun reagieren? Einfach ist es natürlich, wenn er erst gar nichts hinterfragt, denn dann weiß er nicht, ob ein Verstoß gegen das BDSG vorliegt und kann sich auch keiner Schuld bewusst sein. Mit dieser Einstellung sollte er sich jedoch fragen, ob er das gleiche Verhalten auch von einem Kollegen möchte, der mit einer Datenbank, in der seine persönlichen Daten gespeichert sind, ebenso verfährt und somit die Daten auch für andere als für die relevante Zielgruppe ersichtlich sind.

Im Security Panel auf der DOAG 2013 Konferenz in Nürnberg ist auch die Frage aufgekommen, ob man sich bei seinem Arbeitgeber nicht unbeliebt macht oder als externer DBA um den Job fürchten muss, wenn man so einen Missstand anspricht beziehungsweise aufdeckt. Die Fragestellung ist in der Praxis oft eine andere. Ein Anruf beim Datenschutzbeauftragten des Unternehmens (zu finden im globalen Adressbuch genau mit dieser Bezeichnung) kann Wunder bewirken. So hat sich auch der Autor als externer DBA bei einem großen Unternehmen verhalten. Ihm sind dadurch keine Nachteile entstanden und ausgesprochene Drohungen, dass er wohl mit keiner Verlängerung rechnen könne, wenn er die Arbeit verweigere, sind ins Leere gelaufen und die Daten wurden anonymisiert.

An dieser Stelle sei noch kurz auf die Oracle-Optionen hingewiesen, die für diesen Zweck die Anonymisierung sowohl beim Auslesen der Datenbank

als auch beim Kopieren in Akzeptanzbeziehungsweise Entwicklungs-Systeme anbietet. Zudem gibt es auch andere (kostenfreie) Möglichkeiten. Dabei ist allerdings auch zu beachten, dass es nicht mehr reicht, nur die Adress-Tabelle wegzulassen und zu meinen, dass durch die Kunden-Nummer keine Zuordnung mehr möglich sei. Genau dieses in der Vergangenheit oft genutzte Verfahren gestattet das BDSG nicht.

Die DOAG hat aus dem Security Panel einen Wunsch mitgenommen und wird eine Checkliste für den DBA erarbeiten, die auf der DOAG 2014 Datenbank vorgestellt wird. So wird es dem DBA leichter fallen, sich richtig zu verhalten. Zum Schluss ein Zitat aus dem Freundeskreis des Autors: „Man kann über Herrn Snowden denken, wie man möchte. Aber ab und an etwas mehr eigene Zivilcourage würde gut tun“.

Oliver Pyka
oliver.pyka@doag.org



Eine Einführung in Business Transaction Management

Arne Brüning, ORACLE Deutschland B.V. & Co. KG

Die Gattung des Business Transaction Management (BTM) ist heute noch relativ selten in freier Wildbahn zu beobachten. Liegt dies daran, dass es sich hierbei lediglich um einen neuen Hype der Software-Hersteller handelt? Oder ist diese Software-Gattung einfach noch nicht hinreichend bekannt?

Dieser Artikel klärt zunächst, was sich hinter BTM verbirgt. Anschließend wird Oracle BTM (OBTM) als eine mögliche Implementierung vorgestellt. Dabei kommen nützliche Auswahlkriterien zur Sprache, um zu entscheiden, ob BTM in der eigenen IT sinnvoll ist.

Möglichkeiten, komplexe Systemlandschaften zu überwachen, gibt es bereits so viele, dass jedes Unternehmen sich genau überlegen muss, welche davon für den eigenen Betrieb zweckmäßig sind. Überwachung nur um der Überwachung willen hilft niemandem und verursacht nur unnötige Kosten. Warum benötigt man daher nun noch Business Transaction Management?

Egal, ob im eigenen Unternehmen „SOA“ ein Unwort ist oder ein grundlegendes Architektur-Paradigma für die eigene IT-Strategie beschreibt, im Rückblick über die letzten zehn bis zwanzig Jahre hat sich die IT-Landschaft in den meisten Unternehmen grundlegend geändert. Monolithische Anwendungen sind eindeutig auf dem Rückzug. Einzelne Anwendungen verlieren an Bedeutung, systemübergreifende Geschäftsprozesse treten allmählich an ihre Stelle – ob explizit ausgeprägt oder nicht. Daher ist es nur naheliegend, dass sich auch die System-Überwachung diesen geänderten Rahmenbedingungen anpassen muss.

Früher gelangte ein Auftrag möglicherweise über einen einzigen Kanal in die Auftragsverarbeitung. Wenn bei der nächsten Stapelverarbeitung kein Versand-Dokument erstellt wurde, war relativ klar, wo nach dem Status des vermissten Auftrags zu suchen war. Bei dem gleichen Unternehmen sind heute vermutlich mehrere Systeme an der Auf-

tragsbearbeitung beteiligt, etwa weil für bestimmte Teilaufgaben Spezialsysteme beschafft wurden oder auch weil ein komplettes Unternehmen mitsamt seiner eigenen Software übernommen wurde. Somit ist die Suche nach dem System, in dem gerade ein bestimmter Auftrag festhängt, aufwändiger geworden.

Umgekehrt ist es oftmals mühselig, bei gemeinsam benutzten Services herauszufinden, welche Systeme und, vor allen Dingen, welche Prozesse diese nutzen. Wenn ein gemeinsam genutzter Dienst zur Abfrage von Kundendaten zeitweise nicht zur Verfügung steht, ist dann der Auftrag „12345“ von Kunde „A“ oder die Support-Anfrage „67890“ von Kunde „B“ betroffen? Um diese und ähnliche Fragen zu klären, tritt BTM an.

Im Gegensatz zum herkömmlichen Monitoring, das traditionell die Systeme im Fokus hat, steht bei Business Transaction Management (auch Business Transaction Monitoring oder Application Transaction Profiling), wie der Name schon sagt, die Geschäftstransaktion im Vordergrund. Das hat einige Implikationen, allen voran muss das Monitoring-System zunächst einmal die Geschäfts-Transaktionen kennen. Dann müssen diesen die beteiligten Systeme und Dienste zugewiesen werden können. Zur Laufzeit muss dann das BTM-System die Nachrichtenflüsse den Transaktionen zuordnen können.

Hierbei ist ein hoher Grad an Automation wichtig. Man kann sich leicht vorstellen, dass das manuelle Anlegen aller Geschäfts-Transaktionen inklusive aller beteiligten Systeme, Services und insbesondere deren Abhängigkeiten eine Sisyphus-Arbeit wäre. Daher ist es

hilfreich, wenn das System bereits alle beteiligten Komponenten selbstständig erkennt und der Anwender diese möglichst einfach den Geschäfts-Transaktionen zuordnen kann. Voraussetzung dafür ist natürlich, dass das BTM-System eine möglichst große Anzahl der im Unternehmen eingesetzten Systeme monitoren kann. Ein Abgleich der Liste der vom BTM-Anbieter zertifizierten Systeme mit der eigenen System-Landschaft ist unerlässlich.

Das BTM-System sollte auch die Möglichkeit haben, die für den Anwender relevanten Informationen herauszufiltern und Antworten auf typische Fragestellungen zu liefern. Auf die Frage „Wie ist der Status von Bestellung 123456?“ erwartet man eine schnelle und einfache Antwort, ohne dass zunächst eine Menge von SOAP-Nachrichten im Klartext analysiert werden muss.

Ebenso sollte ein BTM-System ein Alerting auf Transaktions-Ebene bieten. Rein technisches Monitoring kann beispielsweise die Antwortzeiten eines Services überprüfen und bei Problemen einen Administrator benachrichtigen. Ob dieser Dienst aber aktuell überhaupt in einer Geschäfts-Transaktion benutzt wird, in der die Störung zu Problemen führen wird, lässt sich darüber jedoch nicht beantworten. Gerade bei komplexeren Prozessen entscheidet sich der Prozessfluss erst zur Laufzeit. Daher ist es wichtig, eine Störung in Bezug zu den Geschäftsprozessen zu setzen.

Zum Beispiel könnte bei einem Order-Management-System ein spezieller Bezahlendienst für einen Werktag offline gehen. Solange kein Auftragsprozess diese Bezahloption benötigt, ist das überhaupt nicht kritisch. Selbst wenn

ein Auftragsprozess versucht, diesen anzurufen, muss das noch kein Problem darstellen, weil es sich möglicherweise um einen Artikel mit ohnehin mehrwöchiger Lieferfrist handelt. Wenn aber ein Kunde mit genau dieser Zahlungsart per Overnight-Express bestellt, sollte das Alerting greifen. In diesem Fall könnte der Kunde informiert und ihm die Möglichkeit gegeben werden, auf eine andere Zahlungsart zu wechseln, bei der das Problem nicht auftritt. Durch ein rein technisches Monitoring der Services wäre dies nicht ohne Weiteres herauszufinden.

Ein anderer wichtiger Aspekt ist die Mandantenfähigkeit eines Systems. Ein Beispiel ist hier ein Unternehmen mit mehreren Abteilungen, die zwar eigene Fachanwendungen besitzen, jedoch aufgrund der bereits im Unternehmen gelebten Wiederverwendung auf gemeinsam genutzte Services zugreifen. Ein übliches SLA-Monitoring berücksichtigt zum Beispiel Durchschnittswerte und Ausreißer; für die jeweilige Abteilung ist es hingegen von Bedeutung, wie die Antwortzeiten für ihre eigenen Prozesse waren. So nützt es einer Abteilung wenig, wenn die Antwortzeiten eines Dienstes im Mittel sehr gut sind, gerade bei den eigenen Anfragen aber deutlich schlechter als die vereinbarten Zeiten lagen.

Umgekehrt informiert der Betrieb die Fachabteilung zum Beispiel, wenn es durch Ausfall eines Dienstes zu Störungen kam. Interessant ist hierbei allerdings nicht, welche Prozesse und Abteilungen potenziell von einer Störung betroffen sein könnten, sondern nur, welche es tatsächlich auch waren. Stellt sich heraus, dass durch die Störung nur die Prozesse eines einzigen Bereichs betroffen waren, brauchen die weiteren Bereiche gar nicht damit behelligt zu werden. Ein mandantenfähiges BTM-System sollte also auch in Störungsfällen helfen, zielgerichteter zu informieren.

Oracle Business Transaction Management

Ein Produkt, das die genannten Anforderungen umsetzt, ist das Oracle Business Transaction Management (OBTM). Es kam durch die Übernahme von AmberPoint zu Oracle und ist heute Bestandteil des Oracle SOA

Management Pack Enterprise Edition. Technisch gesehen ist OBTM eine JEE-Anwendung und läuft auf Oracle WebLogic Servern. Dabei ist es möglich, das Deployment flexibel an das eigene Lastverhalten anzupassen. OBTM besteht zunächst aus den drei zentralen Servern (siehe Abbildung 1):

- *btmMain*
Hauptprogramm, Web-UI
- *btmPerformanceServer*
Überwachung der Services
- *btmTransactionServer*
Überwachung der Transaktionen

Die zentralen OBTM-Server halten ihre Informationen in einer Oracle-Datenbank, der BTM-Database.

Die „Observer“ sind Bibliotheken, die mit den überwachten JEE-Containern geladen werden, die Informationen darüber einsammeln und an das OBTM melden. Zwischen den Observern und den zentralen Servern sitzen die „Monitor Services“. Diese sammeln die Daten der Observer ein und schreiben sie in die BTM-Database. Aus Skalierungsgründen

ist hier auch eine Monitor-Gruppe aus mehreren Servern möglich. Diese mehrschichtige Architektur kann flexibel auf Performance-Engpässe reagieren.

Um den Einstieg zu erleichtern, gibt es für ein erstes Testen des OBTM von Oracle eine leichtgewichtige Variante, die sogenannte „Demo-Installation“. Hier sind die zentralen Server zusammen mit dem Monitor-Service in einem Archiv gebündelt. Damit ist die Installation kinderleicht: WLS installieren und Domain aufsetzen, OBTM-Demo-Installation deployen, Konfigurationsassistenten durchklicken und fertig. Sogar eine Embedded-Datenbank wird mitinstalliert und fertig konfiguriert. Dabei weist der Setup-Assistent ausdrücklich darauf hin, dass diese Variante nur zu Testzwecken gedacht ist. Falls man allerdings genau das möchte, bekommt man in weniger als einer halben Stunde eine lauffähige Umgebung.

In der Produktions-Variante ist das Setup etwas aufwändiger: Zunächst müssen Oracle-Datenbank und WebLogic Server installiert, dann die Oracle-Datenbank und die WebLogic Domain

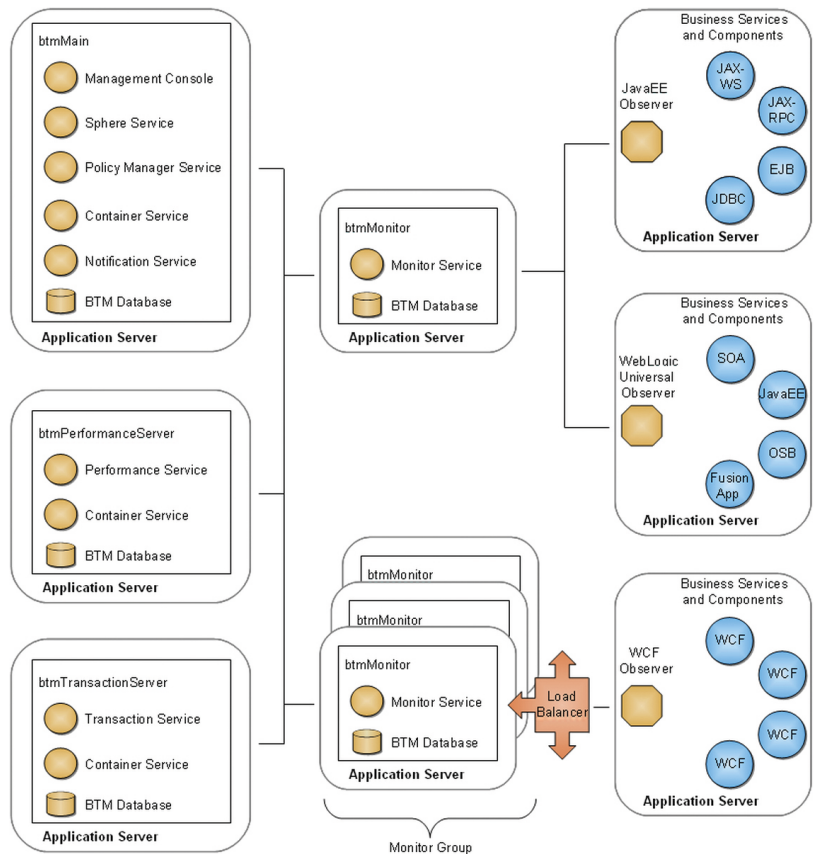


Abbildung 1: OBTM-Architektur-Diagramm

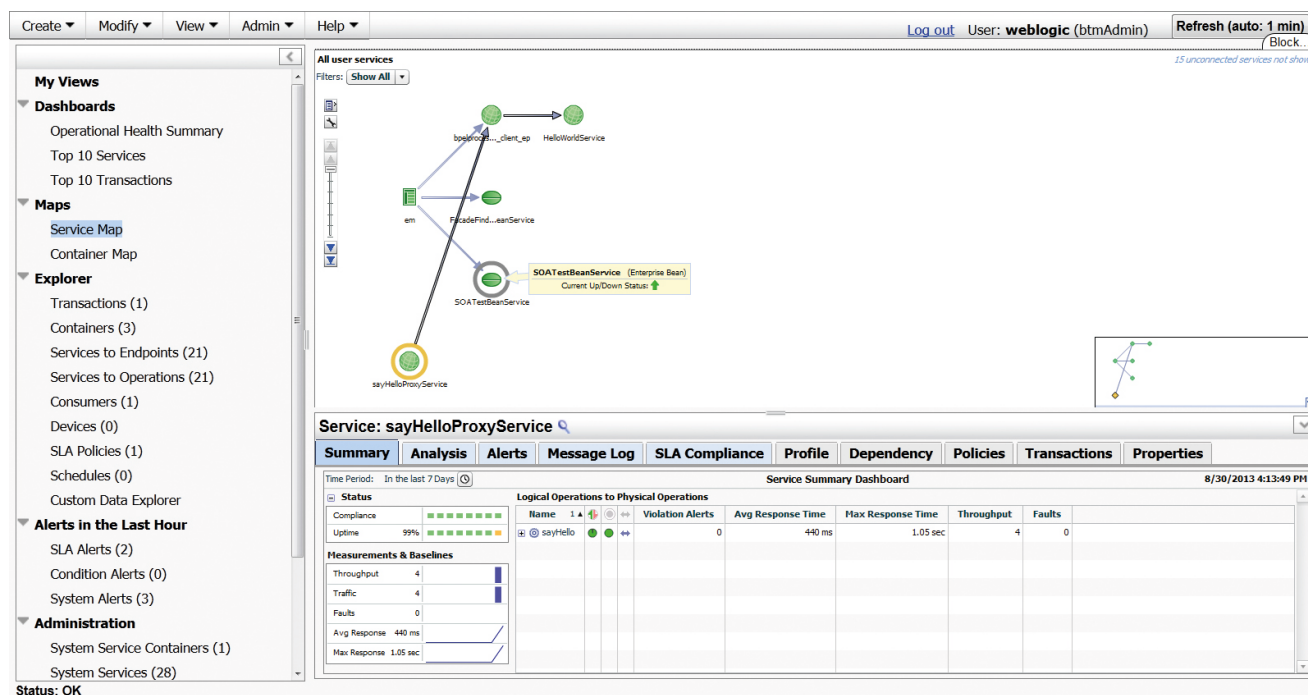


Abbildung 2: Service-Map im OBTM

aufgesetzt werden. Dort hinein werden dann die OBTM-Server installiert und alle Komponenten für die Zusammenarbeit konfiguriert. Je nach Bedarf wird gegebenenfalls noch eine Monitor-Gruppe mit einem Load-Balancer aufgesetzt. Während das für einen Produktivbetrieb unter Last absolut sinnvoll ist, mag das manch einen, der die Software zunächst nur ausprobieren möchte, eher abschrecken. Genau deshalb gibt es die erwähnte Demo-Installation.

Nachdem die OBTM-Server laufen, muss Last auf die beobachteten Systeme gebracht werden. OBTM erkennt die Infrastruktur durch Beobachten der Nachrichten, die die Systeme untereinander austauschen. Hierzu müssen allerdings noch die Observer installiert sein, die nicht nur für WebLogic Server zur Verfügung stehen, sondern auch für JBoss, IBM WebSphere und Microsoft-IIS. Sind diese installiert, kann man sich entspannt zurücklehnen und beobachten, wie OBTM die Abhängigkeiten identifiziert. Die beiliegende Web-UI aktualisiert sich standardmäßig automatisch, sodass man zusehen kann, wie die Komponenten gefunden werden. Als Erstes wird es die beteiligten Application-Server erkennen und ihren Zustand darstellen, das könnte man jedoch natürlich

auch mit klassischem System-Management erledigen.

Als Nächstes wird OBTM eine Service-Map aufbauen, um die Abhängigkeiten unter den Services zu visualisieren (siehe Abbildung 2). Dies erfolgt vollautomatisch, indem OBTM den Nachrichtenverkehr der überwachten Server verfolgt. Der immer noch entspannt zurückgelehnte Administrator kann jetzt dabei zusehen, wie OBTM die Abhängigkeiten erkennt und daraus die Service-Map erstellt. Bereits hier gibt es ein gutes Indiz, ob sich eine weitergehende Beschäftigung mit OBTM lohnt. Kann man die Informationen über diese dargestellten Abhängigkeiten auch aus der bestehenden Dokumentation und dem bestehenden System-Management ebenso einfach gewinnen oder gehört das vielleicht zu den Dingen, die man immer erledigen wollte, wenn dafür einmal Zeit ist? Falls Letzteres der Fall ist, ergibt es möglicherweise Sinn, sich weiter mit der BTM-Materie zu beschäftigen. Diese Darstellung kann bereits eine Impact-Analyse unterstützen. Nach kurzer Zeit ist bekannt, von welchen Systemen ein Service aufgerufen wird. Diese Information könnte man zwar auch aus einem gut gepflegten Enterprise-Repository bekommen, allerdings deckt das Auto-Discovery des Oracle-Enterprise-

Repository nur das Vorhandensein von Services ab, die Abhängigkeiten müssen aber manuell gepflegt werden. Daher ergänzen sich Enterprise-Repository und OBTM an dieser Stelle perfekt.

Die Service-Map ist jedoch nur eine Basisfunktion von OBTM. Darauf aufbauend lassen sich die namensgebenden Geschäfts-Transaktionen definieren. Hierzu werden der Einstiegspunkt der Transaktion ausgewählt und die abhängigen Services aus der Service-Map automatisch übernommen. Sofern nicht alle an der Transaktion beteiligten Services das gleiche Datenmodell nutzen, ist es wahrscheinlich, dass der gleiche Parameter bei jedem Service einen anderen Namen hat. So heißt die Auftragsnummer beispielsweise mal „ANR“, mal „AuftrNr“ oder noch anders. Will man jetzt herausfinden, warum ein bestimmter Auftrag irgendwo im System hängt, so müsste man eine Suche über alle beteiligten Services starten. OBTM bietet hierfür die Möglichkeit, die Parameter über alle Aufrufe mit standardisierten Namen zu versehen, sodass beispielsweise global nach „Auftragsnummer = 123456“ gesucht werden kann. Dabei ist es hilfreich, dass OBTM auch die Inhalte bereits gelaufener Transaktionen in einem konfigurierbaren Zeitfenster speichert, damit auch in den bereits durch-

gelaufenen Transaktionen gesucht werden kann (siehe Abbildung 3).

Über die reine Verfolgung von Services hinaus erfasst OBTM zusätzliche Metriken, sodass hierauf auch ein entsprechendes SLA-Monitoring und Alerting betrieben werden können. In Kombination mit fachlichen Inhalten lassen sich daraus interessante Alerts generieren. So kann OBTM benachrichtigen, wenn die Bestellung eines Premium-Kunden länger als zwei Tage für die Bearbeitung benötigt. Auch lässt sich das Monitoring nach Mandanten unterscheiden, was vor allen Dingen interessant ist, wenn gemeinsam genutzte Komponenten zwischen unterschiedlichen Mandanten geteilt werden.

Die Integration vom OBTM ist natürlich mit weiteren Oracle-Produkten optimal, aber es werden auch Produkte anderer Anbieter unterstützt. Hervorzuheben ist beispielsweise die Interaktion mit Oracle JVM Diagnostics. Im Fehlerfall wird dann nicht nur angezeigt, dass ein Service hängt. Es lässt sich dann direkt aus OBTM auch gleich die JVM-Diagnose für den betreffenden Service starten.

Fazit

Um die eingangs aufgeworfene Frage „Wem nützt BTM?“ zu beantworten: Es hängt sehr von der jeweiligen Infrastruktur des betreffenden Unternehmens ab.

Einige der aufgezeigten Problemstellungen lassen sich auch mit dem bestehenden Infrastruktur-Management klären. Ein gut gepflegtes Enterprise-Repository, das sich mit einem zentralen Enterprise-Service-Bus abgleicht, wird einen guten Teil der Antworten liefern können. Ebenso ist ein Service-Bus ein guter Ansatzpunkt für ein SLA-Monitoring. Auch ein fachliches Alerting lässt sich, abhängig vom eingesetzten Produkt, mehr oder weniger einfach umsetzen. Hat man diese Instrumente etabliert, so ist es zwar interessant, sich per BTM eine Service-Map zu erstellen, um die bestehenden Informationen zu überprüfen. In diesem Fall muss man sich allerdings die Frage stellen, ob die gewonnenen Informationen wirklich die zusätzliche Komplexität wert sind und im Zweifelsfall auf eine zusätzliche BTM-Implementierung verzichten.

Gleiches gilt, wenn die Anwendungs-Landschaft stark standardisiert oder komplett eigenentwickelt ist. Hier ist der zusätzliche Nutzen möglicherweise eher gering. Läuft beispielsweise die komplette Auftragsabwicklung in einem SAP-System, so würde man den Status eines Auftrags mit Bordmitteln abfragen und kein externes BTM-System befragen. Ebenso könnten bei einer Eigenentwicklung die Funktionalitäten für die fachlichen Abfragen von vornherein mit implementiert werden.

Anders sieht es hingegen beim Betrieb einer eher heterogenen Landschaft aus. Bei einer Mischung von Produkten verschiedener Anbieter sowie Eigen-Entwicklungen, die typischerweise auch aus verschiedenen Generationen der Software-Entwicklung stammen, sind die Instrumente zur Beantwortung der oben aufgeworfenen Fragen oft weniger ausgeprägt. Gerade in Unternehmen mit einer heterogenen, gewachsenen IT-Landschaft, möglicherweise im Kombination mit ein paar Jahren Kosteneinsparungen, ist der Innovationsdruck sehr hoch. Oft kommt dann die IT mit der Dokumentation der bestehenden Systeme nicht hinterher, während auf der anderen Seite die Fachabteilungen ständig neue Funktionalitäten benötigen. Auch bringen Firmen-Übernahmen oder neue Geschäftsfelder regelmäßig neue Herausforderungen für die IT mit sich.

Die schöne, heile Welt, in der die Anwendungs-Landschaft komplett aus einer integrierten Suite eines einzelnen Herstellers besteht, trifft man heute genauso selten an wie eine komplett eigenentwickelte Anwendungs-Landschaft, die auch noch durchgängig einem bestehenden Konzept folgt. Die meisten heutigen IT-Landschaften bestehen aus heterogenen Systemen, die aber aus fachlicher Sicht durchgängige Prozesse abbilden müssen. Die BTM-Systeme treten an, um diesem Umstand Rechnung zu tragen und den Überblick in heterogenen Umgebungen zu bewahren. Daher ist es durchaus sinnvoll, sich näher mit BTM zu beschäftigen.

Arne Brüning
arne.bruening@oracle.com

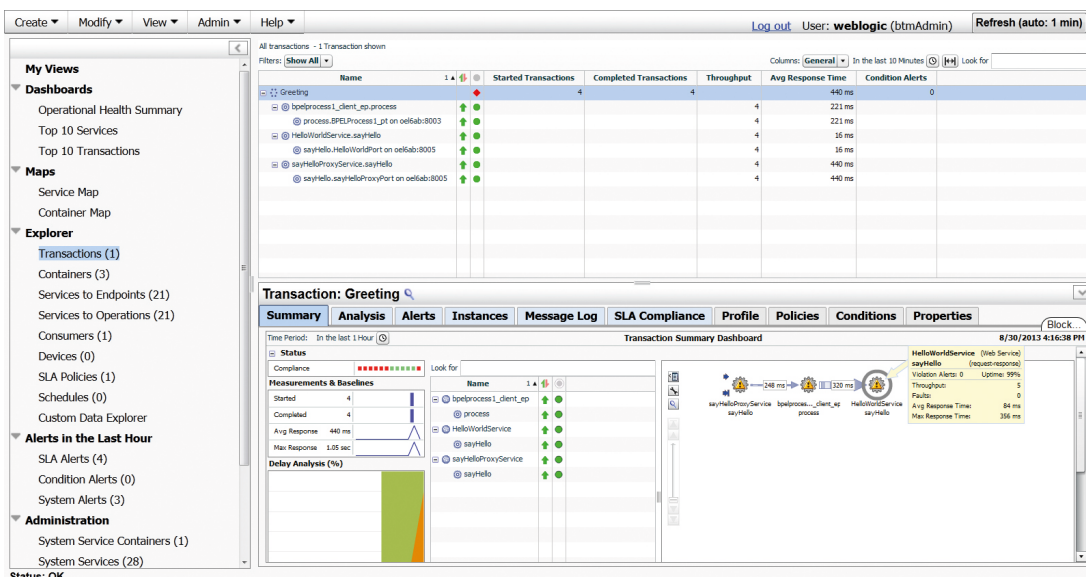


Abbildung 3: Darstellung einer OBTM-Transaktion

Mobil integrierte Geschäftsprozesse mit ADF Mobile

Enno Schulte, OPITZ CONSULTING GmbH

Die Integration mobiler Geräte in die Geschäftsprozesse stellt für viele Unternehmen eine Herausforderung dar, die sie angesichts der Möglichkeiten, die mobile Technologien bieten, erfolgreich meistern möchten.

Der Artikel zeigt die Hürden und die Chancen der mobilen Integration und stellt Umsetzungsmöglichkeiten sowie einen Ansatz zur Identifikation geeigneter Prozesse vor. Im Zentrum stehen die technologischen Anforderungen, die bei der Umsetzung von Bedeutung sind, und die spannende Frage, ob das Oracle Application Development Framework Mobile (ADF) ein geeignetes Werkzeug für die Realisierung mobiler Unternehmensanwendungen ist.

Mobile Technologien

In Zeiten, in denen die Internetnutzung und der mobile Datenverkehr gesamtgesellschaftlich zunehmen, wächst auch stetig der Bedarf an mobiler Unternehmenssoftware. Immer mehr Unternehmen binden heute mobile Technologien in den Arbeitsalltag ein, indem sie die Anbindung privater Smartphones an die eigene Infrastruktur erlauben. Dieser Sachverhalt wird häufig als „Bring-Your-Own-Device“-Bewegung bezeichnet.

Zurzeit setzen Firmen in diesem Umfeld überwiegend Anwendungen zur Anbindung an die Bürokommunikations-Software ein: Mitarbeiter lesen E-Mails oder erstellen Termine im Kalender der Groupware. Dank der mobil eingebundenen Mitarbeiter können viele Abläufe im Unternehmen beschleunigt werden. Mit dieser Form der Integration lässt sich jedoch nur eine geringe Teilmenge der möglichen Vorteile für die Geschäftsprozesse erreichen. Viele Optionen, die sich durch die neuen Möglichkeiten der Geräte in einen Prozess integrieren lassen, bleiben bei der schlichten Anbin-

dung einer bestehenden Anwendung ungenutzt.

Viele Abteilungen haben mobile Potenziale innerhalb ihrer Prozesse entdeckt und arbeiten inzwischen an geeigneten Lösungen, um ihre Abläufe zu mobilisieren [1]. So entsteht oftmals ein sogenannter „Flickenteppich“ mobiler Anwendungen, der sich aus verschiedensten Technologien mit unterschiedlichen Anbindungen an das Unternehmen zusammensetzt. Dieses Anwendungskonglomerat ist aufgrund der unterschiedlichen Technologien und der zerstreuten Entwicklung schlecht aufeinander abgestimmt. Der Autor bezeichnet die hier zu erwartende IT-Landschaft mittlerweile als „Wartungshölle“. Doch wie kann man ein solches Worst-Case-Szenario verhindern? Der Industrie mangelt es an ganzheitlichen Lösungen, um ihre Geschäftsprozesse durch die Integration mobiler Endgeräte zu unterstützen. Die geringe Geschwindigkeit, mit der mobile Lösungen eingeführt werden, ist häufig auf Bedenken zu Sicherheit, Reife und Weiterentwicklung der aktuell verfügbaren Technologien zurückzuführen [1]. Zudem ist vielen Firmen nicht bewusst, wie groß die Auswirkungen mobiler Unternehmensanwendungen auf ihre Geschäftsabläufe sein können. Stattdessen stehen Ängste im Vordergrund, die das ökonomische Risiko, die Komplexität der Entwicklung und die Frage, zu welchem Zeitpunkt welche Prozesse mobilisiert werden sollen, betreffen.

Das Potenzial der Mobilität

Mobil integrierte Geschäftsprozesse profitieren vor allem von der höheren

Erreichbarkeit und der Ortsunabhängigkeit der partizipierenden Mitarbeiter. Inzwischen besitzen mobile Endgeräte eine Vielzahl von sensorischen Möglichkeiten wie beispielsweise die Ortung der eigenen Position oder das Erfassen von audiovisuellen Medien. Mithilfe dieser Sensoren lassen sich Medienbrüche verhindern.

Die geringe Darstellungsfläche auf Smartphones zwingt Entwickler dazu, neue Oberflächen zu erstellen. Die Anzahl der Funktionen, die der Nutzer auf einen Blick zu sehen bekommt, wird damit verringert. Viele Benutzer empfinden diese Reduzierung als Arbeitserleichterung und ziehen diese Oberfläche der eines Desktop-PCs vor, bei dem sich die Suche nach denselben Funktionen aufgrund komplexer Menüstrukturen beschwerlicher gestaltet.

Geschäftsprozesse können also von der höheren Erreichbarkeit und der sensorischen Erfassbarkeit mobiler Systeme profitieren. Prof. Rahul C. Basole von der Georgia Tech in Atlanta [1] subsumiert den Nutzen für Unternehmen auf drei Bereiche: Effektivität, Effizienz und Bequemlichkeit. Vorteile wie Kontext-Sensitivität, Allgegenwärtigkeit oder Telemetrie-Funktionen resultieren letztendlich in einer effizienteren, effektiveren und bequemerem Organisation. Mobile Informationssysteme können, sofern sie in den richtigen Bereichen eingesetzt und den richtigen Mitarbeitern zu Verfügung gestellt werden, zu einem agileren, adaptiveren und kosteneffizienteren Unternehmen führen. Zeitkritische Abläufe werden beschleunigt und profitieren von ortsbezogenen Informationen, direkt am

„Point-Of-Action“. Durch den allgegenwärtigen Zugang zu qualitativ besseren Informationen zur richtigen Zeit und am richtigen Ort sind Mitarbeiter in der Lage, bessere Entscheidungen zu treffen (siehe **Abbildung 1**).

Um eine umfassende Lösung für die Integration mobiler Geräte in Geschäftsprozessen zu finden, empfiehlt es sich, zwei unterschiedliche Perspektiven auf das Thema einzunehmen: Aus dem strategischen Blickwinkel sind Prozesse mit den größten Nutzungspotenzialen zu identifizieren und eventuell an mobile Szenarien anzupassen. Aus technischer Perspektive muss in erster Linie eine geeignete Technologie ausgewählt werden, um die Anforderungen einer mobilen Unternehmenssoftware zu erfüllen.

Aspekte des mobilen Business

Im weiteren Verlauf kommen Aspekte der Mobilität in Unternehmen und die daraus resultierenden Anforderungen an mobile Unternehmensanwendung zur Sprache. Anschließend wird untersucht, ob ADF Mobile eine geeignete Technologie ist, um diesen Anforderungen zu entsprechen.

Im ersten Schritt müssen die Prozesse dargelegt werden, die die größten Mobilisierungspotenziale aufweisen. Danach untersucht das zuständige Team die Abläufe im Hinblick auf ihren wirtschaftlichen Mehrwert bei einer Mobilisierung. Beispielsweise mithilfe des Mobile Process Landscaping [5] können Architekten entsprechende Prozesse lokalisieren, um danach mit dem Mobility-M-Modell [3] den finanziellen Erfolg der Mobilisierung festzustellen. Dazu wird die Theorie der informationellen Mehrwerte nach Kuhlen [4] auf die Anbindung mobiler Anwendungssysteme bezogen.

Bezüglich der Beschreibung dieser Methoden sei hier nur zusammenfassend erwähnt, dass es wichtig ist, ein besonderes Augenmerk auf die Geschäftsprozesse zu richten, die zeitkritisch sind, die an unterschiedlichen Orten durchgeführt werden oder die viele Medienbrüche beinhalten. Im weiteren Verlauf unterscheidet man zwischen der einfachen Anbindung an ein mobiles System und der Anpassung des Prozesses.

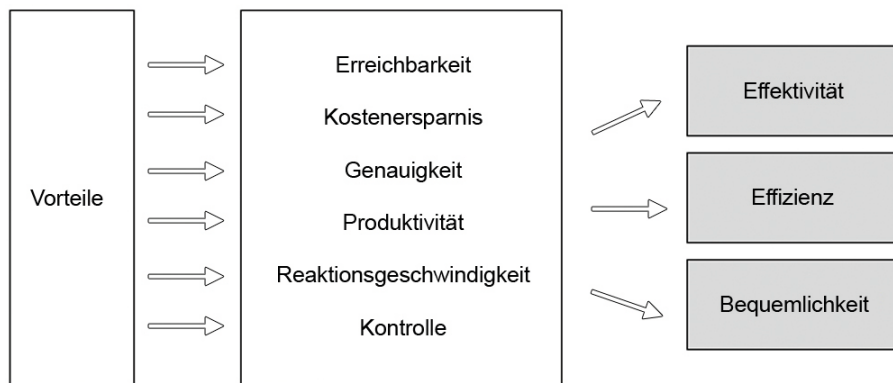


Abbildung 1: Potenziale mobiler Anwendungen (Quelle: Basole (1))

Mobile Business Process Improvement

Macht die Organisation ihre bestehenden Prozesse lediglich durch mobile Endgeräte zugänglich, so spricht man vom Mobile Business Process Improvement. Der dominierende Vorteil dieser Methode ist die einfache Effizienzsteigerung durch eine Beschleunigung der Durchlaufzeiten. Dabei profitiert der Prozess nicht in dem Ausmaß, wie es möglich wäre, wenn die volle Bandbreite der technologischen Funktionen in den Prozess mit einfließen würde. In diesem Zusammenhang wären beispielsweise die Anreicherung durch kontextbasierte Informationen oder die Qualitätssteigerung durch sensorische Werte zu nennen.

Angesichts der reinen Anbindung mobiler Mitarbeiter an die bestehende Prozesskette ist dies die günstigste Form einer Mobilisierung. Jedoch geht bei dieser Variante viel Potenzial verloren und der Wettbewerbsvorteil wird entsprechend geringer ausfallen, als es bei einem Mobile Business Process Reengineering der Fall wäre.

Mobile Business Process Reengineering

Im Gegensatz zum Mobile Business Process Improvement zielt das Mobile Business Process Reengineering auf eine Neugestaltung des gesamten Geschäftsprozesses ab, unter Berücksichtigung der neuen technologischen Möglichkeiten. Um den größtmöglichen Nutzen zu erreichen, betrachtet der Architekt nicht nur einzelne Prozesse. Stattdessen möchte er die Prozess-Landschaft des Unternehmens grundlegend überarbeiten. Auf diese

Weise nutzen Organisationen die vollen Potenziale mobil integrierter Geschäftsprozesse.

Formen der Mobilität

Bevor es an die Auswahl einer geeigneten Technologie beziehungsweise Architektur zur Integration mobiler Lösungen geht, muss die Organisation analysieren, welche Formen der Mobilität in welchen Prozessen vorhanden sind. Fragen wie „Welche Mitarbeiter sind involviert?“ und „Wie mobil sind die Tätigkeiten?“ führen hier weiter.

Auf diese Weise lassen sich vier Gruppen von mobilen Mitarbeitern in Unternehmen differenzieren. Die erste Gruppierung ist lediglich innerhalb des Firmengeländes mobil und bewegt sich damit innerhalb der Strukturen, die vom Unternehmen frei beeinflusst werden. Eine weitere Gruppe sind die Mitarbeiter, die ihren Arbeitsauftrag nicht nur auf dem Firmengelände erledigen und regelmäßig die Firmengrenzen verlassen. Angestellte, die ihr komplettes operatives Geschäft mobil durchführen, stellen eine weitere Gruppe mobiler Mitarbeiter.

Die letzte Gruppe bilden Entscheidungsträger. Selbst wenn ein Entscheidungsträger seine gesamte Arbeitszeit auf dem Firmengelände verbringt und somit eigentlich auch Bestandteil der ersten Gruppe wäre, so bekleidet er in vielen Prozessen eine verantwortliche Position und blockiert die Abläufe, wenn beispielsweise eine Genehmigung benötigt wird. Deshalb sind diese Mitarbeiter einer eigenen Gruppe zugeordnet.

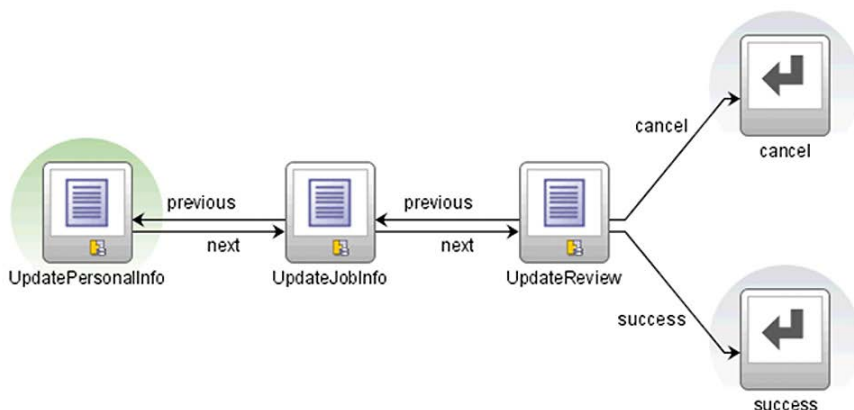


Abbildung 2: ADF Mobile Task Flow

Technologische Anforderungen

Hat die Organisation die Prozesse lokalisiert und gegebenenfalls angepasst, kann die IT mit der technischen Realisierung beginnen. Neben den Anforderungen, die der jeweilige Prozess definiert, stellt die Entwicklung für mobile Geräte aufgrund der neuen Geräteklasse zusätzliche Anforderungen.

Der geringe Formfaktor hat zur Folge, dass nahezu alle Ressourcen auf den Geräten nur in beschränkten Größen vorhanden sind. Obwohl Prozessorgeschwindigkeit und Arbeitsspeicher auf den Spitzengeräten inzwischen einen Grad erreicht haben, der durchaus mit kleinen Desktop-PCs mithalten kann, so muss man sich doch an den Durchschnittswerten des Marktes orientieren und eventuell Geräte unterstützen, die nur geringe Rechenleistung bereitstellen.

Die geringe Größe der mobilen Geräte erfordert zudem ein Umdenken bei der Entwicklung von Oberflächen. Zum einen ist der Platz für die Darstellung sehr eingeschränkt und erfordert Kreativität bei der Strukturierung der Oberfläche, zum anderen sind neue Bedienkonzepte durch Wischgesten und die Eingabe mit einem oder mehreren Fingern hinzugekommen.

Bei den üblichen Infrastrukturen, in denen Unternehmensanwendungen betrieben werden, bricht die Datenverbindung nahezu niemals ab. Mobile Geräte hingegen arbeiten autark und müssen sich permanent wandelnden Umweltbedingungen fügen. So kann die verfügbare Bandbreite zwischen 177.6 KB/s (GPRS) und 1 GB/s

(802.11ac) schwanken und es muss jederzeit mit Verbindungsabbrüchen gerechnet werden. Deshalb ist es wichtig, dass die eingesetzte Technologie in der Lage ist, die benötigten Daten vorzuhalten. Damit müssen auch kritische Unternehmensinformationen auf dem Gerät gespeichert sein, die im Falle eines Diebstahls oder Verlusts leicht in falsche Hände geraten können.

Aus Sicherheitsgründen sollten Anwender also so wenige Daten wie möglich auf den Geräten speichern und diese stattdessen auf den besser gesicherten Unternehmensservern belassen. Gleichzeitig sollten aber auch so viele Informationen wie nötig auf den Geräten verbleiben, um das Weiterarbeiten in Offline-Szenarien zu gewährleisten. Diesen Widerspruch bezeichnet man als das „mobile Dilemma“.

Eine geeignete Technologie für die Entwicklung von mobilen Unternehmensanwendungen muss sich also bei geringer Bandbreite an die Unternehmensdienste anbinden lassen, die abgerufenen Informationen persistieren können und diese zugleich verschlüsseln. Zusätzlich sollten die Entwickler die Oberflächen auf die mobilen Geräte abstimmen und den Zugriff auf Sensoren ermöglichen, die in den Geräten vorhanden sind. Außerdem muss die Anwendung auch ohne Datenverbindung betrieben werden können.

Diese allgemeinen Anforderungen decken wohl einen Großteil der möglichen Anwendungsfälle in Prozessen ab. Je nach Mobilitätsgrad können die Anforderungen auch wesentlich geringer ausfallen. Beschränkt sich der Kreis

der Anwender beispielsweise auf Mitarbeiter, die lediglich innerhalb der Unternehmensgrenzen mobil sind, so können die Voraussetzungen der Infrastruktur so gestaltet werden, dass persistente Informationen auf dem Gerät so gut wie ausgeschlossen sind.

Oracle ADF Mobile

Das auf der Oracle OpenWorld 2012 erstmals vorgestellte Application Development Framework (ADF) Mobile ist Oracles Antwort auf die Zukunft mobiler Unternehmensanwendungen. Anwendungen für Android- und iOS-Betriebssysteme lassen sich mit dem Framework denkbar einfach erstellen. Die Entwicklung der Anwendungen ist simpel und orientiert sich am deklarativen Entwicklungsstil, den viele Entwickler schon aus ADF gewohnt sind.

Im Kern setzt das Framework auf Java und macht damit aus jedem Software-Entwickler mit Java-Kenntnissen automatisch einen geeigneten Kandidaten für die Arbeit mit ADF Mobile. Der Einsatz bestehender Java-Ressourcen und die Unterstützung der zwei gängigsten mobilen Betriebssysteme ohne Anpassungen am Quellcode machen das Framework aus betriebswirtschaftlicher Perspektive besonders attraktiv. Zudem ist das Framework imstande, die zuvor genannten Anforderungen an eine mobile Unternehmensanwendung zu erfüllen. Die Anwendungen laufen nativ auf den Geräten und lassen sich auch ohne eine Datenverbindung ausführen. Dabei erzeugt der Entwickler die Oberfläche in HTML und fügt mithilfe von JavaScript dynamisches Verhalten hinzu (siehe Abbildung 2).

Die Entwicklung der Oberfläche erfolgt analog zu der von ADF-Anwendungen. Der Kontrollfluss wird in Task Flows definiert, die sich aus Steuerelementen und Seiten zusammensetzen. Diese Seiten sind in XML definiert und bestehen aus vordefinierten Komponenten, die anhand ihrer Attribute konfiguriert werden. Als Dateierweiterung nutzen sie die Abkürzung „*.amx“, weshalb sie auch AMX-Seiten heißen (siehe Abbildung 3).

Im Framework stehen bereits „Cascading Style Sheets“ zur Verfügung,

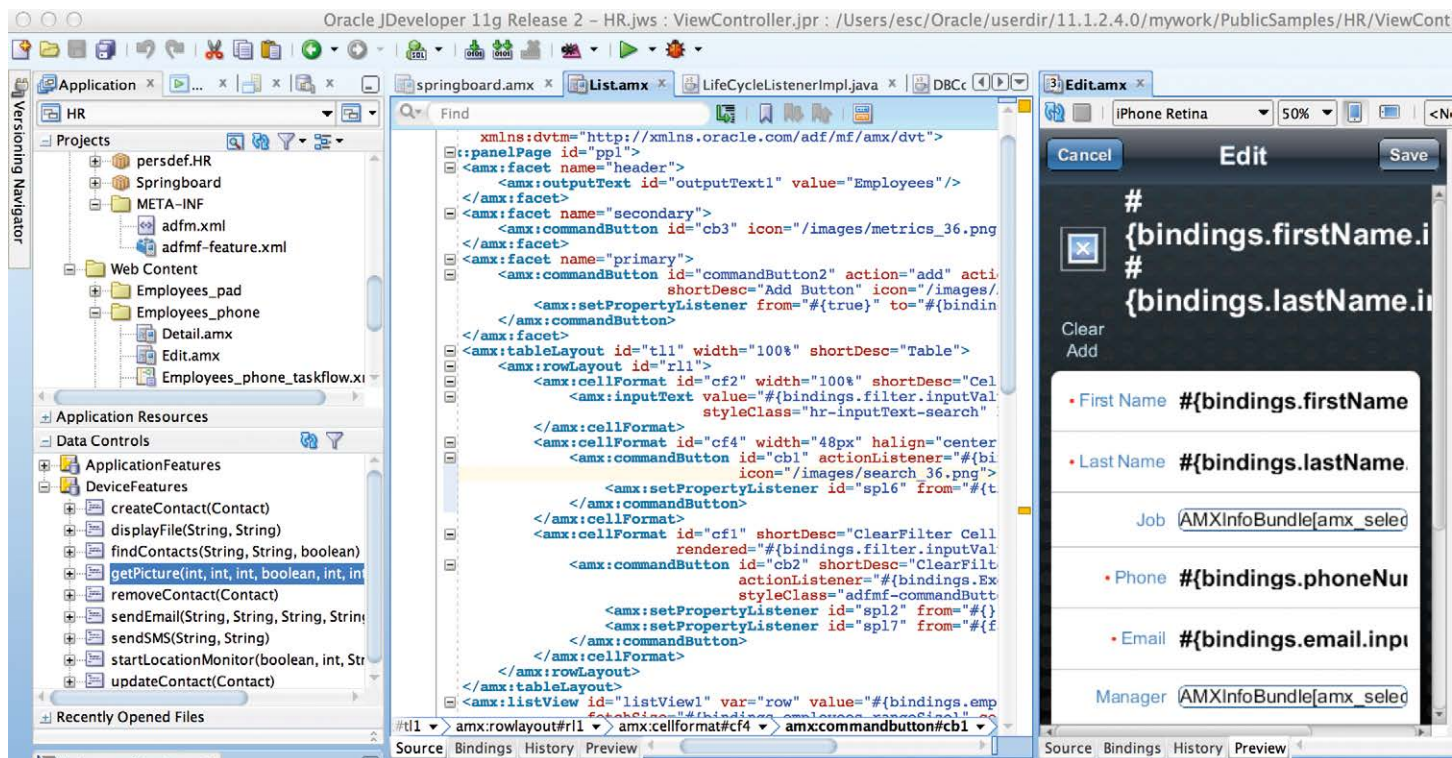


Abbildung 3: ADF-Mobile-AMX-Seite

die das „Look&Feel“ an die jeweilige mobile Plattform anpassen. Eigene Änderungen können anschließend noch hinzugefügt werden, sodass der Developer mit wenigen Schritten eine Oberfläche erstellen kann, die nicht vermuten lässt, dass es sich im Hintergrund um eine Website handelt.

Zur Anbindung an die Infrastruktur des Unternehmens bietet ADF Mobile Service-Adapter für SOAP- und RESTful-Webservices. Obwohl RESTful-Webservices aufgrund ihres geringeren Overheads für mobile Szenarien besser geeignet sind, ist die Integration von SOAP-Webservices etwas einfacher und besser etabliert. Man kann allerdings davon ausgehen, dass dies in kommenden Versionen angeglichen wird.

Zu Recht sind Bedenken bezüglich der Wartbarkeit solcher Anwendungen angebracht. Was ist, wenn sich die URL eines Webservice ändert? Wie schnell kann man gewährleisten, dass alle Anwender ein Update erhalten? Für dieses Problem bietet ADF Mobile einen Konfigurations-Service, der beim Start der App überprüft, ob eine neue Konfigurationsdatei (XML) unter einer zuvor definierten URL vorhanden ist. Sofern vorhanden, wird diese geladen und verwendet.

Adressiert die hinterlegte URL einen Bus wie den Oracle Service Bus (OSB) oder ein Gateway wie das Oracle Enterprise Gateway (OEG), so können die Services sowohl durch den Einsatz der Konfigurationsdatei als auch durch den Austausch hinter dem zentralen Einstiegspunkt geändert werden. Die dadurch erzielte doppelte Entkopplung verspricht ein hohes Maß an Flexibilität.

Zur Persistierung der Daten bietet ADF Mobile eine integrierte SQLite-Datenbank. Will man diese verwenden, so müssen die SQL-Skripte zur Installation des Datenbank-Schemas in der Anwendung enthalten sein und beim erstmaligen Start der Anwendung einmal ausgeführt werden. Der Zugriff auf die SQLite-Datenbank erfolgt via JDBC, wobei immer nur eine schreibende und viele lesende Verbindungen erlaubt sind.

Leider beinhaltet ADF Mobile kein Konzept für das objektorientierte Mapping und befördert den Entwickler damit gefühlte zwanzig Jahre in die Vergangenheit. Deshalb hat die ADF Enterprise Methodology Group (EMG) ein Tool namens „ADF EMG Mobile Accelerator“ entwickelt. Es unterstützt den Entwickler dabei, die erforderli-

chen Strukturen und Klassen mithilfe von Wizards zu erzeugen. Leider ist das Tool noch nicht offiziell veröffentlicht. Auf der DOAG 2013 Konferenz hat Oracle bestätigt, dass man selbst an einer ähnlichen Unterstützung für Entwickler arbeitet und diese direkt in das Framework integrieren wird.

In puncto Sicherheit kann die SQLite-Datenbank verschlüsselt und damit die Risiken beim Verlust eines Geräts verringert werden. Entwickler müssen an dieser Stelle jedoch ein wenig aufpassen. Wird die SQLite-Datenbankdatei nicht ordnungsgemäß entschlüsselt und anschließend erneut mit einem anderen Passwort chiffriert, kann man nie wieder auf die Daten zugreifen. Oracle verweist im Entwickler-Handbuch ausdrücklich auf diesen Sachverhalt.

Bevor die Sicherheitsmechanismen von ADF Mobile weiter erläutert werden, kommen zunächst das sogenannte „Springboard“ und die Definition von Features zur Sprache: Eine Anwendung in ADF Mobile besteht aus einem oder mehreren Features. Ein Feature besteht üblicherweise aus einem Task Flow oder einer einzelnen Seite. Dabei kann eine Seite verschiedene Ausprägungen haben. Im einfachsten Fall handelt es sich

lediglich um den Zugriff auf eine externe Seite über ihre URL, was nichts anderes ist als der Aufruf der Seite im Browser des mobilen Gerätes.

Ähnlich, aber offline verfügbar, ist der Aufruf einer HTML-Seite, die lokal innerhalb der Anwendung vorgehalten wird. Die letzte Möglichkeit ist der Verweis auf eine AMX-Seite. Ein Feature steht für die Definition eines Prozesses innerhalb der App. Um Features aufrufen zu können, bietet ADF Mobile ein übliches Pattern für Oberflächen in mobilen Anwendungen. Falls so konfiguriert, startet die Anwendung mit einem Springboard, auf dem alle verfügbaren Features übersichtlich dargestellt werden. Wem dieses Standardlayout missfällt, der kann mit wenig Aufwand ein eigenes Springboard definieren (siehe Abbildungen 4 und 5).

Hinsichtlich Authentifizierung und Autorisierung hat ADF Mobile bereits Vorkehrungen getroffen. Für einzelne Features kann eine Anmeldung erforderlich sein. Die Implementierung ist sehr einfach gehalten. In einer App lassen sich Verbindungen zu URLs mit einer HTTP Basic Authentication hinterlegen. Um ein spezielles Feature zu nutzen, muss der Anwender nun die erfolgreiche Verbindung zu der vorher hinterlegten Adresse durch die Eingabe des korrekten Passworts ermöglichen. Soll das Feature im Offline-Szenario verfügbar sein, stehen die Anmelde-daten in einem Keystore zur Verfügung. Andernfalls bräuchte man eine Datenverbindung, um das Feature zu nutzen.

Zugriff auf Geräte-Funktionen

Der Zugriff auf Funktionen wie GPS, Kamera oder die gespeicherten Kontakte lässt sich mit einem geringen bis gar keinem Programmieraufwand realisieren. Bei der Entwicklung einer ADF Mobile App stehen dem Developer Data Controls zur Verfügung. Diese bilden eine einheitliche Schnittstelle zur Nutzung der Gerätefunktionalitäten. Beispielsweise existiert eine Methode namens „sendSMS(String,String)“ im Device Feature „Data Control“. Der Aufruf dieser Methode mit der Nummer und dem Text in den Übergabe-Parametern führt zum Aufruf der Standard-Anwendungen für Nachrichten auf dem jeweiligen Gerät. Die Werte für Nummer und Text sind entsprechend mit den Parametern gefüllt.

Das direkte Senden von SMS ohne eine Interaktion mit dem Benutzer ist jedoch nicht möglich. Diese Restriktion seitens der mobilen Plattformen soll den Benutzer vor Schadsoftware schützen. Andernfalls bestünde die Gefahr, dass eine solche Software zum Beispiel unkontrollierbare Mengen von SMS an Bezahl-dienste schickt.

Zur Realisierung des Zugriffs wird im Hintergrund Apache Cordova (auch bekannt als „PhoneGap“) eingesetzt. Dabei handelt es sich um eine JavaScript-Bibliothek, die einen einheitlichen Zugriff auf Gerätefunktionen ermöglicht, um plattformübergreifende hybride Apps zu entwickeln (siehe Abbildung 6).

Fazit

Die Frage nach der Eignung von Oracle ADF für die Entwicklung von mobilen Unternehmenslösungen lässt sich nach Ansicht des Autors weitestgehend positiv beantworten. ADF Mobile bringt viele erforderliche Fähigkeiten mit, um mobile Anwendungen für Unternehmen zu realisieren. Besonders interessant sind die Unterstützung von Java und die Möglichkeit, eine Anwendung ohne viele Anpassungen auf verschiedenen Plattformen bereitzustellen. Diese Aspekte verschaffen dem Framework angesichts der großen Gemeinde von Java-Entwicklern einen wirtschaftlichen Vorteil. Der deklarative Entwicklungsstil mit der hohen Anzahl von Wizards beschleunigt zudem die Einarbeitungszeit. So erzielen auch Anfänger im mobilen Bereich sehr schnell erste Ergebnisse.

Bezüglich der im Vorfeld vorgestellten technologischen Anforderungen für die Entwicklung mobiler Unternehmensanwendungen erfüllt ADF Mobile nahezu alle Kriterien. Lediglich die Unterstützung von älteren Geräten (> 1,5 Jahre) lässt zu wünschen übrig. Ein angenehmes Arbeiten mit einer in ADF Mobile entwickelten App ist aus Erfahrungen des Autors erst ab einem iPhone 4s, eher noch iPhone 5 möglich. Aber hier spielt die Zeit für das Framework, da die Geräte immer leistungsfähiger werden.

Die Anbindung an die Unternehmens-Infrastruktur mit Webservices ist der konsequente Schritt hinsichtlich



Abbildung 4: ADF Mobile App mit Standard Springboard



Abbildung 5: ADF Mobile App mit Custom Springboard

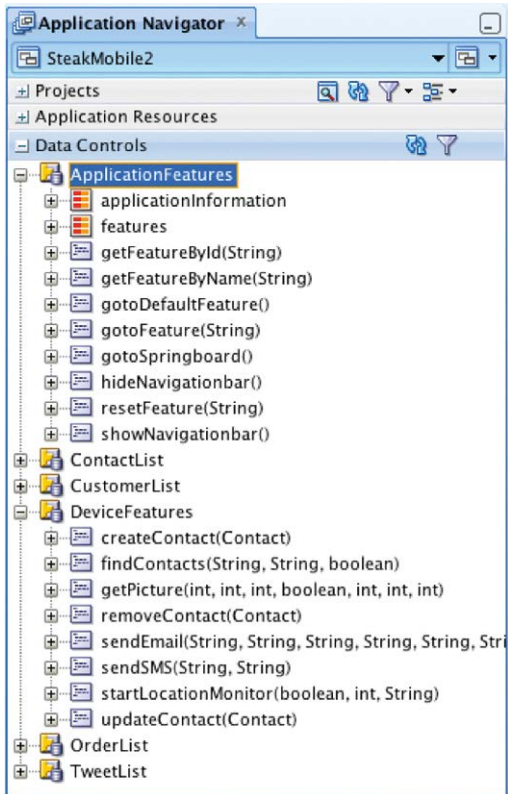


Abbildung 6: ADF Mobile Data Controls

einer mobilen Integration in Oracles Fusion Middleware. Unternehmen, die

bereits SOA im Einsatz haben, profitieren an dieser Stelle von den bei ihnen vorhandenen Voraussetzungen für die Anbindung von mobilen Endgeräten. Der Autor betont an dieser Stelle die Wichtigkeit, wartbare Architekturen zu schaffen und nur über eine zentrale Infrastruktur auf die Unternehmensdienste zuzugreifen (etwa mittels OSB oder OEG).

In der aktuellen Version hat das Framework zwar noch mit kleineren Kinderkrankheiten zu kämpfen, etwa mit mangelnden Lösungen für die Synchronisation von offline erfassten Daten. Allerdings könnte bereits im kommenden Jahr eine neue Version erscheinen, die dann dank einer schlanken Java 8 VM und neuer Features wohl auch die letzten Kritiker überzeugen wird.

Literatur

- [1] R. C. Basole: Mobilizing the enterprise: A conceptual model of transformational value and enterprise readiness, 26th ASEM National Conference Proceedings, 2005
- [2] A. Pousttchi, und K. Gump: The Mobility-M-Framework for Application of Mobile Technology in Business Pro-

cesses, 35. GI-Jahrestagung Informatik, 2005

- [3] A. Pousttchi und K. Turowski: Mobile Commerce, Springer, 2003
- [4] R. Kuhlen, R. Hammwöhner und J. Herget: Informationsmarkt: Chancen und Risiken der Kommerzialisierung von Wissen, Universitätsverlag, 1995
- [5] U. Wellen: Process Landscaping – Eine Methode zur Modellierung und Analyse verteilter Softwareprozesse, Dissertation, 2003

Enno Schulte

enno.schulte@opitz-consulting.com



Web-Services in der Oracle-Datenbank

Detlev Schütte, ORBIT Gesellschaft für Applikations- und Informationssysteme mbH

Dieser Artikel beschreibt Web-Services auf der Oracle-Datenbank, die auf der Funktionalität „XML DB“ basieren. Zielgruppe sind Entwickler und Administratoren. Einer allgemeinen Beschreibung folgt eine systematische Abhandlung der Schritte zum Einrichten eines funktionsfähigen Web-Service. Sicherheitsaspekte sind ebenfalls berücksichtigt.

Web-Services sind ein gängiger Weg des Informationsaustauschs über das Internet (B2B, B2C) und spielen daher in der heutigen IT-Welt eine bedeutende Rolle. Im Wesentlichen laufen folgende Prozesse ab:

- Der Konsument ruft beim Anbieter eine Information ab, indem er an dessen URL eine Anfrage sendet.

- Der Web-Service des Anbieters bearbeitet die Anfrage des Konsumenten und sendet die Antwort zurück.

Üblicherweise geschieht das alles mittels einer XML-Anfrage unter Nutzung des Netzwerkprotokolls SOAP (Simple Object Access Protocol). Seit der Datenbank-Version 10g bietet die neuen Funktionalität „XML DB“ die Möglichkeit,

Web-Services mit Datenbank-Mitteln anzubieten und zu konsumieren. Oracle XML DB unterstützt SOAP in der Version 1.1. Mit einem solchen Service lassen sich SQL- und XQuery-Abfragen erstellen, und eine Darstellung der Resultate im XML-Format ist ebenfalls möglich. Zu letzterem Zweck können PL/SQL-Prozeduren genutzt werden, deren In- und Out-Parameterliste das Format

der ein- beziehungsweise auslaufenden Dokumente festlegen. Die zur Parametrisierung des Web-Service benötigten WSDL-Dokumente (WSDL = Webservice Description Language) werden automatisch durch die Funktionalität „XML DB“ der Oracle-Datenbank generiert.

Oracle-XML-DB-Web-Services

Aus Sicherheitsgründen sind Web-Services in der Datenbank normalerweise nicht aktiviert. Für eine Nutzung muss man sie extra konfigurieren und aktivieren. Es wird dringend empfohlen, hierzu einen eigenen User namens „<<WS_USER>>“ ohne DBA-Privilegien anzulegen.

Beim Austausch sensibler Informationen ist die Nutzung eines Web-Service über HTTPS angeraten. Falls die Kommunikationspartner einander vertrauen, kann die Abfrage von Benutzername und Passwort entfallen. Dies ist bei Job-gesteuerten Abfragen immer der Fall, da hier eine manuelle Authentifizierung nicht praktikabel ist.

Um einen vertrauensvollen Zugang zu ermöglichen, müssen die beteiligten Partner Sicherheitszertifikate austauschen und an bestimmter Stelle ablegen. Ablageort für Sicherheitszertifikate bei XML DB ist das sogenannte „Wallet“. Diese Datei liegt im Verzeichnis „ORACLE_HOME/admin/<<Name der Datenbank Instanz>>“ des Datenbank-Servers; ihr Ort kann individuellen Wünschen angepasst werden. Der Pfad ist der Datenbank bekannt zu machen; in den Code-Beispielen ist er mit „<<Pfad für Oracle-Wallet>>“ gekennzeichnet.

Falls ein Oracle-Wallet noch nicht existiert, muss es erstellt werden. Dies kann entweder durch eine Client- oder eine Server-seitige GUI geschehen oder auf der Kommandozeile des Datenbank-servers. Voraussetzung ist ein X-fähiges Terminal, es muss also die grafische Version von UNIX vorhanden sein. Das Dienstprogramm zum Aufruf der GUI lautet „owm“. Es ist im Verzeichnis „ORACLE_HOME/bin“ zu finden.

Im Folgenden wird das Verfahren über die Kommandozeile beschrieben. Zur Ausführung der Kommandos ist eine Anmeldung als User „ORACLE“ auf dem Datenbank-Server erforderlich. Das Dienstprogramm zum Anle-

gen, Konfigurieren und Anzeigen eines Wallet lautet „orapki“ und ist im Verzeichnis „ORACLE_HOME/bin“. Es ist in mehrfacher Weise über eine Reihe von Parametern einsetzbar. Insbesondere wird ein Passwort erwartet, das in den Code-Beispielen als „<WALLET_Passwort>“ gekennzeichnet wird. **Listing 1** zeigt das Kommando zum Anlegen eines leeren Wallet.

Wie bereits erwähnt, lassen sich im Wallet Zertifikate der Kommunikationspartner ablegen. Bei Bedarf können sie zusammengesetzt sein, so können sie beispielsweise aus dem Zertifikat einer autorisierten, ausstellenden Organisation (Certification Authority, kurz CA) und einem User-Zertifikat bestehen (siehe **Listing 2**).

Zu Testzwecken lässt sich auch ein eigenes, selbst signiertes Zertifikat erstellen, das man dem Kommunikationspartner zur Verfügung stellen kann.

Seine Gültigkeitsdauer lässt sich begrenzen. Für seine Erstellung ist ein „Distinguished Name“ (DN) erforderlich. Dieser enthält identifizierende Angaben wie den Namen des Zertifikats, der Organisation, der Organisationseinheit, des Standorts sowie die Namen beliebiger geographischer Angaben. Die Angaben sind durch bestimmte Kürzel gekennzeichnet. Ein DN könnte beispielsweise so aussehen: „CN = Name_des_Zertifikats, O = Name_der_Organisation, OU = Name_der_Organisationseinheit, L = Stadt, ST = Bundesland, C = DE“ (siehe **Listing 3**). Der Inhalt eines Wallet lässt sich mit dem in **Listing 4** gezeigten Kommando darstellen.

Konfiguration eines Servlet

Der User „SYS“ muss die Servlet-Konfiguration erstellen. Dies geschieht mit der Konsolen-Anwendung „SqlPlus“.

```
$ORACLE_HOME/bin/orapki wallet create -wallet <Pfad für Oracle-Wallet> -pwd <WALLET_Passwort> -auto_login
```

Listing 1: Wallet anlegen

```
# User-Zertifikat importieren (Teil 1 von 2 der Zertifikats-Kette)
$ORACLE_HOME/bin/orapki wallet add -wallet <Pfad für Oracle-Wallet>
-trusted_cert -cert <User-Zertifikat> -pwd <WALLET_Passwort>
# CA-Zertifikat importieren (Teil 2 von 2 der Zertifikats-Kette)
$ORACLE_HOME/bin/orapki wallet add -wallet <Pfad für Oracle-Wallet>
-trusted_cert -cert <CA-Zertifikat> -pwd <WALLET_Passwort>
```

Listing 2: Zertifikate ins Wallet importieren

```
# zu Testzwecken eigenes, selbst signiertes Zertifikat erstellen
# Gültigkeitsdauer des Zertifikats: 3650 Tage (10 Jahre)
$ORACLE_HOME/bin/orapki wallet add -wallet <Pfad für Oracle-Wallet>
-keysize 1024 -dn '<distinguished Name>' -self_signed -pwd <WALLET_Passwort>
-validity 3650
# Eigenes, selbst signiertes Zertifikat exportieren in Datei
# Hinweis: die Datei kann dem Kommunikationspartner für Testzwecke
zur Verfügung gestellt werden
$ORACLE_HOME/bin/orapki wallet export -wallet <Pfad für Oracle-Wallet>
-dn '<distinguished Name>' -cert <Dateiname für Zertifikat> -pwd
<WALLET_Passwort>
```

Listing 3: Selbst signiertes Zertifikat erstellen und in Datei exportieren

```
$ORACLE_HOME/bin/orapki wallet display -wallet <Pfad für Oracle-Wallet>
-pwd <WALLET_Passwort>
```

Listing 4: Inhalt eines Wallet anzeigen

```

-- Servlet und Servlet-Mapping erstellen
DECLARE
  SERVLET_NAME VARCHAR2(32) := 'orawsv';

BEGIN
  DBMS_XDB.addServlet(NAME      => SERVLET_NAME,
                     LANGUAGE => 'C',
                     DISPNAME => 'Oracle Query Web Service',
                     DESCRIPT => 'Servlet for issuing queries as a Web Service',
                     SCHEMA   => 'XDB');
  DBMS_XDB.addServletSecRole(SERVNAME => SERVLET_NAME,
                             ROLENAM  => 'XDB_WEBSERVICES',
                             ROLELINK => 'XDB_WEBSERVICES');
  DBMS_XDB.addServletMapping(PATTERN => '/orawsv/*',
                             NAME     => SERVLET_NAME);

END;
/

-- Ergebniskontrolle-1: Check Servlet
SELECT XMLSerialize(content
  XMLQuery('declare namespace ns="http://xmlns.oracle.com/xdb/xdbconfig.xsd"; (: :)
    $r//ns:servlet[ns:servlet-name="orawsv"]'
    passing DBMS_XDB.CFG_GET() AS "r" returning content)
  AS CLOB ) servlet_config
FROM dual
;

-- Ergebniskontrolle-2: Check Webservice Mapping "orawsv"
select extractValue(value(x),'/servlet-mapping/servlet-name', 'xmlns="http://xmlns.oracle.com/xdb/xdbcon-
fig.xsd"') "WebServiceMapping"
  from table(xmlsequence(extract(xdburitype('/xdbconfig.xml').getXML(),'xdbconfig/sysconfig/protocolcon-
fig/httpconfig/webappconfig/servletconfig/servlet-mappings/servlet-mapping'))) x
;

```

Listing 5: Konfiguration Web-Service – Servlet erstellen

Der Name des Servlet ist in diesem Beispiel „orawsv“ (siehe Listing 5).

Falls keine andere Konfiguration vorgenommen wurde, läuft die Kommunikation bei der Oracle-Funktionalität „XML DB“ unverschlüsselt über den Standard-Port 8080. Aus Sicherheitsgründen ist aber eine verschlüsselte Kommunikation vorzuziehen, beispielsweise über den Port 8893 (siehe Listing 6).

Oracle-Listener

Der Oracle-Listener-Prozess muss über die konfigurierten Kommunikations-Ports informiert werden. Das geschieht zum einen innerhalb der Datenbank (Listener Endpoint) und zum anderen in den Dateien „listener.ora“ und „sqlnet.ora“. In der Datenbank wird der Listener Endpoint mit dem in Listing 7 gezeigten SQL-Kommando gesetzt.

In den Dateien „listener.ora“ und „sqlnet.ora“ (standardmäßig im Verzeichnis „ORACLE_HOME/network/admin“ zu finden) muss die Netzwerk-Konfiguration angepasst werden. Die

Datei „listener.ora“ ist wie folgt zu ergänzen (siehe Listing 8). Darüber hinaus ist in der Datei „sqlnet.ora“ folgender Eintrag zu ergänzen (siehe Listing 9).

Konfiguration der Zugangskontrolle

Aus Sicherheitsgründen empfiehlt es sich, die Nutzung eines Web-Service nur bestimmten Nutzern zu gestatten, die in der Zugangskontroll-Liste (engl.: Access Control List, kurz ACL) geführt werden. In der Datenbank 10g war die Liste beschränkt auf den User „PUBLIC“. Somit stand ein Web-Service allen Nutzern zur Verfügung. Mit der

Version 11g wurde dieses Konzept folgendermaßen verfeinert:

- Es können einzelne User in die Zugangsliste aufgenommen werden.
- Die ACL kann auf bestimmte Hosts eingeschränkt werden. Möglich ist die Angabe eines Host-Namens, einer Domäne, einer IP-Adresse oder eines Sub-Netzes. Beim Host-Namen ist die Groß- und Kleinschreibung zu beachten. Bei IP-Adressen und Domänen sind Wildcards möglich.
- Es gibt eine Beschränkung auf bestimmte Port-Bereiche.

```

-- SET XDB Port (HTTP)
exec DBMS_XDB.setHTTPPort(8893);

-- Ergebniskontrolle
select dbms_xdb.getHttpPort() as "HTTP-Port"
  from dual
;

```

Listing 6: Konfiguration Web-Service – Kommunikation über Port 8893

```

-- Listener ENDPOINT setzen
exec dbms_xdb.setListenerEndPoint(1,null,8080,1);
exec dbms_xdb.setListenerEndPoint(2,null,8893,2);

-- Ergebniskontrolle:
declare
  a1 varchar(2);
  b1 number;
  c1 number;
  a2 varchar(2);
  b2 number;
  c2 number;

BEGIN
  dbms_xdb.getlistenerendpoint(1, a1, b1, c1);
  dbms_xdb.getlistenerendpoint(2, a2, b2, c2);

  dbms_output.put_line(a1 || ' ' || b1 || ' ' || c1);
  dbms_output.put_line(a2 || ' ' || b2 || ' ' || c2);
END;
/

-- Check WS Ports
select extractValue(value(x),'httpconfig/http-protocol', 'xmlns="http://xmlns.oracle.com/xdb/xdbconfig.xsd"')
"HTTP Protocol"
, extractValue(value(x),'httpconfig/http-port',
'xmlns="http://xmlns.oracle.com/xdb/xdbconfig.xsd"') "HTTP Port#"
, extractValue(value(x),'httpconfig/http2-protocol', 'xmlns="http://xmlns.oracle.com/xdb/xdbconfig.xsd"')
"HTTPS Protocol"
, extractValue(value(x),'httpconfig/http2-port', 'xmlns="http://xmlns.oracle.com/xdb/xdbconfig.xsd"') "HT-
TPS Port#"
from
table(xmlsequence(extract(xdburitype('/xdbconfig.xml').getXML(),'/xdbconfig/sysconfig/protocolconfig/httpcon-
fig'))) x
;
```

Listing 7: Konfiguration Web-Service – Listener Endpoint

Listing 10 zeigt die gespeicherte Prozedur zum Anlegen einer ACL.

Sonstige Schritte

Wie bereits erwähnt wurde, muss bei Benutzung eines Wallet dessen Pfad in der ACL der Datenbank bekannt sein. Zur Bekanntgabe dient das in **Listing 11** gezeigte Kommando. Sobald dies erledigt ist, muss die XDB-Konfiguration aktualisiert werden (siehe **Listing 12**).

Aus Sicherheitsgründen sollte der Anbieter eines Web-Service nicht der User „SYS“ sein und über keine DBA-Privilegien verfügen. Vielmehr sollte ein geeigneter Datenbank-User ohne jegliche DBA-Privilegien angelegt werden. Sowohl Anbieter als auch Nutzer sind mit den in **Listing 13** gezeigten Privilegien auszustatten.

Sind die erforderlichen Privilegien zugewiesen, so steht dem User die Nutzung eines Web-Service über HTTPS zur Verfügung. Er hat dann den glei-

```

WALLET_LOCATION=
  (SOURCE=
    (METHOD=FILE)
    (METHOD_DATA=
      (DIRECTORY=<Pfad für Oracle-Wallet>)
    )
  )

SSL_CLIENT_AUTHENTICATION=FALSE

(DESCRIPTION=
  (ADDRESS = (PROTOCOL=tcp) (HOST=<hostname>)(PORT=8080))
  (Presentation=HTTP)(Session=RAW)
)
(DESCRIPTION=
  (ADDRESS = (PROTOCOL=tcps) (HOST=<hostname>)(PORT=8893))
  (Presentation=HTTPS)(Session=RAW)
)
```

Listing 8: Netzwerk-Konfiguration „listener.ora“

chen Zugang zu seinen Objekten wie bei einer manuellen Anmeldung, allerdings sind die PUBLIC-Objekte für den Web-Service gesperrt. Sollte der Web-Service des Users auch Zugang

zu PUBLIC-Objekten benötigen, so ist dies ab Version 11g gesondert zu veranlassen (siehe **Listing 13**). Falls nötig, kann zusätzlich ein Zugang über HTTP gewährt werden, siehe ebendort.

XML-DB-Web-Services nutzen

Zur Nutzung eines Web-Service mit Datenbank-Mitteln eignen sich PL/SQL-Prozeduren oder Packages. Damit kann der Konsument des Service automatisch einen SOAP-Request erstellen und an den Anbieter senden. Auch die SQL-Injektion wird mit Prozeduren beziehungsweise Packages ausgeschlossen.

```

WALLET_LOCATION=
  (SOURCE=
    (METHOD=FILE)
    (METHOD_DATA=
      (DIRECTORY=<Pfad für Oracle-Wallet>)
    )
  )

SSL_CLIENT_AUTHENTICATION=FALSE
SQLNET.AUTHENTICATION_SERVICES = (TCPS,NTS, BEQ) # add TCPS

```

Listing 9: Netzwerk-Konfiguration „sqlnet.ora“

```

-- ACL oraws.xml anlegen
begin
  if (dbms_db_version.ver_le_10_2) then
    dbms_output.put_line('Oracle-Version kleiner / gleich 10.2');
    dbms_output.put_line('keine ACL Konfiguration noetig');
  else
    dbms_output.put_line('ACL oraws.xml anlegen');
    dbms_network_acl_admin.create_acl ( acl          => 'oraws.xml'
                                       , description => 'ACL fuer Ora-WebService'
                                       , principal   => '<WS_USER>' -- Schema der Anwendung
                                       , is_grant    => true
                                       , privilege   => 'connect');

    dbms_output.put      ('- add privilege ...');
    dbms_network_acl_admin.add_privilege ( acl          => 'oraws.xml'
                                          , principal   => '<WS_USER>' -- Schema der Anwendung
                                          , is_grant    => true
                                          , privilege   => 'resolve');

    dbms_output.put_line(' OK');
    dbms_output.put      ('- assign acl ...');
    dbms_network_acl_admin.assign_acl ( acl          => 'oraws.xml'
                                       , host        => '<URL des
Datenbankservers>' -- z. B. ,10.1.10.*'
                                       , lower_port  => 8080
                                       , upper_port  => 8893);

    dbms_output.put_line(' OK');

    dbms_output.put_line('ACL oraws.xml fertig');
  end if;
end;
/

-- Ergebniskontrolle
SELECT dna.host
      , dna.lower_port
      , dna.upper_port
      , dna.acl
--      , dna.aclid
FROM DBA_NETWORK_ACLS dna;

-- ACL-Privileges checken
select dnac.accl
--      , dnac.aclid
      , dnac.principal
      , dnac.privilege
      , dnac.is_grant
--      , dnac.invert
      , dnac.start_date
      , dnac.end_date
from DBA_NETWORK_ACL_PRIVILEGES dnac;
select dwa.wallet_path
      , dwa.acl
--      , dwa.aclid
from dba_wallet_acls dwa;

```

Listing 10: Konfiguration Web-Service – ACL anlegen


```
-- ACL: Wallet-Lokation bekanntgeben
BEGIN
  DBMS_NETWORK_ACL_ADMIN.ASSIGN_WALLET_ACL ( acl => 'oraws.xml'
                                           , wallet_path => 'file:
<Pfad für Oracle-Wallet>' -- Wallet Path des Servers
                                           );
END;
/
```

Listing 11: Bekanntgabe des Wallet-Speicherorts

```
-- refresh XDB-Configuration ...
exec DBMS_XDB.CFG_REFRESH;
```

Listing 12: Auffrischung der XDB-Konfiguration

```
-- Zugang über HTTPS
GRANT XDB_WEBSERVICES TO <WS_USER>;
-- Zugang zu PUBLIC-Objekten
GRANT XDB_WEBSERVICES_WITH_PUBLIC TO <WS_USER>;
-- Zugang über HTTP (nicht nur HTTPS)
GRANT XDB_WEBSERVICES_OVER_HTTP TO <WS_USER>;
```

Listing 13: Zuweisung der Privilegien

Für die Anfrage wird ein XML-Dokument erstellt, das als SOAP-Request an den Anbieter gesendet werden kann.

Sein Format muss der WSDL (siehe oben) des Anbieters entsprechen. Anschließend wird eine Verbindung zum

Anbieter aufgebaut. Für vertrauensvolle Verbindungen muss der Speicherort des Wallet bekannt sein. Falls der Anbieter die Anfrage fehlerfrei verarbeiten konnte, sendet er eine Antwort zurück, die von der Prozedur abgeholt wird. Als Beispiel dient das in [Listing 14](#) gezeigte Package. Es steht aufgrund seiner Länge unter www.doag.org/go/doagnews/webservice zum Download bereit.



Detlev Schütte
detlev.schuette@orbit.de



DBMonipro: Die leistungsstarke, **kostenlose** Monitoring Lösung für **ALLE** Oracle Datenbank Editionen.

DBMonipro läuft via Nagios und enthält zahlreiche vorgefertigte Datenbank Checks. Neue Checks können in wenigen Minuten ohne Programmierkenntnisse erstellt werden.

Kostenloser Download: <http://www.dbmonipro.at>



Die Oracle Experten

Lizenz
Kosten

Praxis Seminar

Oracle Software Lizenzierung

✓ Klare Antworten

✓ Fehler vermeiden ✓ Kosten optimieren

✓ Zahlreiche Praxisbeispiele

Bei diesem eintägige Seminar lernen Sie alle Stolpersteine eindeutig zu erkennen und zu vermeiden, sowie Einsparungspotentiale zu identifizieren und umzusetzen.



Ing. Klaus-Michael Hatzinger
CEO DBConcepts GesmbH, Oracle DB Lizenzierungs Experte

25. März Düsseldorf, 26. März Frankfurt
27. März Stuttgart, 01. April Wien

Shortlink zur Veranstaltung: <http://bit.ly/praxis-seminar>

Eine Alternative zu Database Change Notification bei Massendaten-Änderungen

Michael Griesser und Thies Rubarth, Freiberufler, sowie Nis Nagel und Andriy Terletsyy, Berenberg Bank

Dieser Artikel zeigt eine Alternative zu Oracles „Database Change Notification“ beziehungsweise „Continuous Query Notification“ im Bereich der Massendaten-Änderung, wobei die multilingualen User-Clients mit unterschiedlichen territorialen Einstellungen in einem festgelegten Intervall über DML-Ereignisse aus der Datenbank informiert werden.

Bei Berenberg gibt es für viele Anwendungen die Anforderung, Datensatz-Änderungen (DML-Operationen) automatisch in der GUI aller aktiven Anwender anzuzeigen. Ein ständiges Neu-Abfragen der Daten durch die GUI würde jedoch die gesamte Applikation unnötig belasten und eine viel zu große Datenmenge mehrfach bearbeiten. Stattdessen soll die Datenbank proaktiv in den Aktualisierungsprozess für die GUI einbezogen und Daten-Spam aus der Quelle vermieden werden. Bisher gab es dafür eine Trigger-basierte Lösung, die mit zunehmendem Datenvolumen und bei hochfrequenten Änderungen an ihre Grenzen stieß. Ihre wesentlichen Nachteile lagen darin, dass jede einzelne DML-Änderung an die Clients gesendet wurde und dass die Ausführung der Trigger die Performance der Businessprozesse in der Datenbank negativ beeinflusste.

Anforderungen, Analyse und Entscheidung

Heutzutage ist es besonders beim Wertpapierhandel üblich, dass ein Or-

der-Datensatz innerhalb eines kurzen Zeitabschnitts mehrfach vom System geändert wird. Auch ist der Durchsatz an Order-Einstellungen in einem Handelssystem sehr hoch. Grund dafür ist eine hohe Anzahl maschinell gehandelter Orders an den Börsen (algorithmisches Trading).

Die neue Lösung soll deshalb in einem zeitlich festgelegten Intervall einen Snapshot der geänderten Datensätze an den Client senden und in unabhängigen, von der Datenverarbeitung getrennten Prozessen laufen (siehe [Abbildung 1](#)). Eine weitere Anforderung an das System ist die unterschiedliche Aufbereitung der Nachrichten, abhängig von den Sprach- und Ländereinstellungen der angemeldeten Benutzer. Mit diesen Anforderungen und der Tatsache, dass Oracle in ihrer Spezifikation angibt, dass das Datenbank-Feature „Database Change Notification“ nur bei geringen Änderungen effizient ist und ansonsten die gesamte Query aktualisiert wird, fiel

die Entscheidung zugunsten einer Eigen-Entwicklung.

Die Architektur

[Abbildung 2](#) zeigt die System-Architektur der Client Notification für Datensatz-Änderungen. Das System kann in fünf Komponenten unterteilt werden:

- Tabellen mit Row-Dependencies
- Consumer-Verwaltung mit Queue-Überwachung
- SCN-Capture-Jobs
- Message Enqueue
- Nachrichten-Empfang auf GUI-Ebene

Es muss zunächst definiert werden, für welche Tabellen die Client Notification umzusetzen ist. Die Tabelle ist dann mit „ROWDEPENDENCIES“ anzulegen oder mit „DBMS_REDEFINITION“ zu ändern. Damit speichert die Oracle-Datenbank für jeden Datensatz eine eigene „ROW_SCN“ (System Change Number). Diese ist eine Pseudo-Column des Datensatzes/Blocks

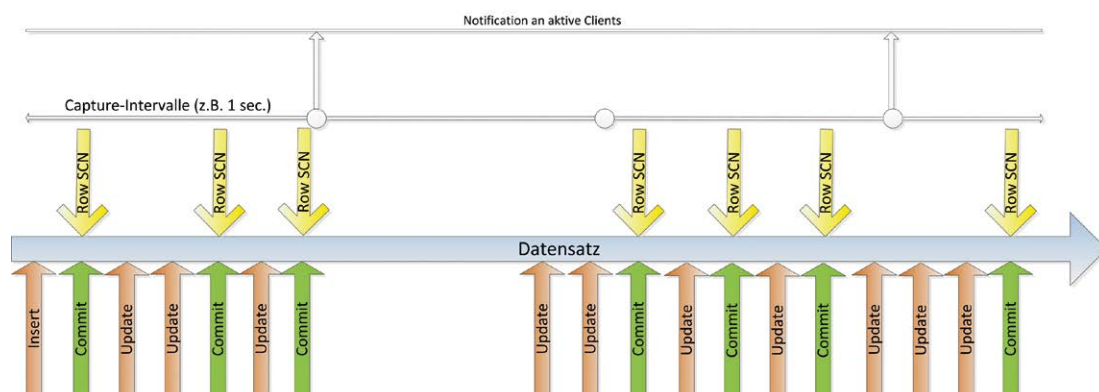


Abbildung 1: Row-SCN-Snapshots bei Massendaten-Änderungen

und beschreibt eine Art fortlaufenden „Transaktionszähler“.

Der GUI-Client (Consumer) meldet sich an das Notification-System über ein selbstgeschriebenes PL/SQL-API in der Datenbank an. Zusätzlich registriert sich jede Masken-Instanz im System als sogenannter „Client-Listener“. Die Consumer-Verwaltung stellt ein Mapping zwischen den Masken-Definitionen und den entsprechenden Business-Tabellen/Objekten her und definiert damit, welche Datenbank-Tabellen aktiv von den SCN-Capture-Jobs erfasst werden müssen. Ein zusätzlicher Job in der Consumer-Verwaltung übernimmt Monitoring-Aufgaben. Er erkennt inaktive Consumer (ausgesetzte Heartbeats) und meldet diese ab. Ein weiterer Job überwacht die Queues und entfernt Expired Messages, die in die Exception-Queue verschoben wurden.

Über Scheduler-Jobs werden die in das System konfigurierten Tabellen mithilfe der Pseudo-Spalte „ORA_ROWSCN“ auf DML-Ereignisse überwacht. Sobald sich Client-Listener für Änderungen interessieren, wird in dem jeweilig eingestellten Zeitintervall (abhängig vom Business-Umfeld) die Ergebnismenge der jüngsten DML-Änderung („ORA_ROWSCN BETWEEN <Letzter SCN> AND DBMS_FLASHBACK.GET_SYSTEM_CHANGE_NUMBER()“) gesammelt (siehe Listing 1).

Aus der Liste wird je eine Text-Message erzeugt, gruppiert nach den jeweiligen Sprach- und sonstigen territorialen Einstellungen der User, und an die entsprechende Recipient-Liste gesendet (Siehe Listing 2).

Als Schnittstelle zwischen Datenbank und GUI dient eine Multi-Subscriber-Queue mit einem JMS-Payload-Type (siehe Listing 3). Je nach Konfiguration können diese Messages „buffered“ oder „persistent“ verschickt werden, wobei hier auch im laufenden Betrieb eingegriffen und dieses ad hoc umkonfiguriert werden kann. Die Message wird im JSON-Format erzeugt (siehe Listing 4).

Um bei „buffered Messages“ das Problem einer zu hohen Enqueue-Rate zu umgehen, muss auf die Exception „ORA-25307 Enqueue-Rate zu hoch, Fluss-Steuerung aktiviert“ reagiert werden. Hier schaltet das System im laufen-

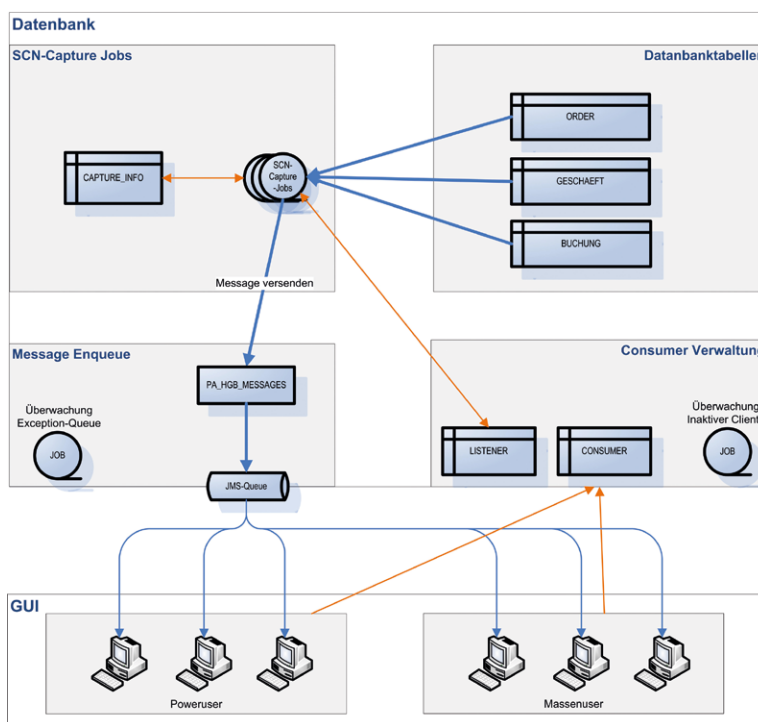


Abbildung 2: Architektur der Client Notification

```
CREATE TYPE TABLE_NUMBER AS TABLE OF NUMBER;

CREATE OR REPLACE PROCEDURE PR_CAPTURE
IS
  scn_start    NUMBER;
  scn_end      NUMBER;
  tab_order    TABLE_NUMBER := TABLE_NUMBER();
BEGIN

  -- Letzte SCN auslesen
  SELECT i.scn_last
  INTO scn_start
  FROM CAPTURE_INFO i
  WHERE i.TABLE_NAME = 'ORDER';

  scn_end      := DBMS_FLASHBACK.GET_SYSTEM_CHANGE_NUMBER();

  -- prüfen ob es angemeldete listener gibt
  IF PA_HGB_MESSAGES.LISTENER_EXISTS(IN_TABLE_NAME => 'ORDER') THEN
    -- capturen der tabelle
    SELECT o.ORDER_ID
    BULK COLLECT INTO tab_order
    FROM ORDER o
    WHERE o.ora_rowscn BETWEEN scn_start+1 AND scn_end;

    IF tab_order.COUNT() > 0 THEN
      PA_HGB_MESSAGES.VERSENDEN(
        IN_TABLE_IDS => tab_order
      , IN_CLIENT_TABLE_NAME => 'ORDER');
    END IF;
  END IF;

END;

-- Letzte SCN merken
UPDATE CAPTURE_INFO i
SET i.scn_last = scn_end
WHERE i.TABLE_NAME = 'ORDER';
COMMIT;

END;
```

Listing 1

den Betrieb automatisch auf „persistent Messages“ um und wechselt anschließend wieder auf „buffered Messages“.

Implementierung in Java

Um aus Java heraus auf Oracle Advanced Queuing zugreifen zu können, bietet Oracle zwei unterschiedliche APIs an. Zum einen ein proprietäres API, das direkt mit den Queuing-Funktionen arbeitet, und zum anderen ein JMS-API, über das zum größten Teil mit Standard-Java-Methoden auf die Queues zugegriffen werden kann.

Da die Oracle-Datenbank keinen JNDI-Context anbietet, muss das JMS-Topic über Oracle-Methoden erstellt werden. Zusätzlich wird das Oracle-spezifische Konstrukt „TopicReceiver“ verwendet, das benötigt wird, um auf die „Multi-Subscriber-Queues“ zuzugreifen, da JMS ein solches Konzept nicht vorsieht.

In **Abbildung 3** sind die in der Java-Anwendung notwendigen Schritte zum Empfangen von JMS-Nachrichten dargestellt. Die Schritte 1 plus 2 und 5 können über einen Anwendungsserver laufen, sodass die DB-Verbindung in den Clients auf das Empfangen von JMS-Nachrichten beschränkt werden kann.

Die Klasse aus **Listing 5** setzt voraus, dass die Anmeldung bereits erfolgt ist und ein Listener für eine Tabelle angemeldet wurde. Zunächst muss eine Topic-Connection zur Datenbank aufgebaut werden. Mit dem Namen des Consumers kann dann auf das Topic zugegriffen werden.

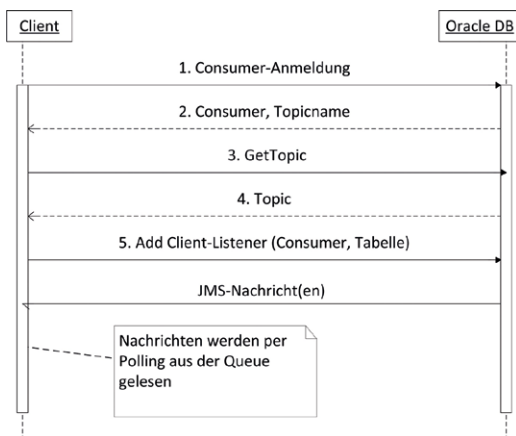


Abbildung 3: Sequenz-Diagramm für das Einrichten des Nachrichten-Empfangs

```

t_recipients SYS.DBMS_AQ.AQ$_RECIPIENT_LIST_T;
...
SELECT sys.AQ$_AGENT(c.NAME, NULL, NULL)
BULK COLLECT INTO t_recipients
FROM CONSUMER c
WHERE ...;

```

Listing 2

```

BEGIN
-- QueueTable anlegen
SYS.DBMS_AQADM.CREATE_QUEUE_TABLE(
    QUEUE_TABLE => 'AQ_JMS_HGB_TAB'
    ,QUEUE_PAYLOAD_TYPE => 'SYS.AQ$_JMS_TEXT_MESSAGE'
    ,COMMENT => 'QueueTable für HGBs'
    ,MULTIPLE_CONSUMERS => TRUE
    ,SORT_LIST => 'priority,enq_time'
);
-- Queue anlegen
SYS.DBMS_AQADM.CREATE_QUEUE(QUEUE_NAME => 'AQ_JMS_HGB_WPS'
    ,COMMENT => 'Queue für HGBs'
    ,QUEUE_TABLE => 'AQ_JMS_HGB_TAB'
    ,QUEUE_TYPE => SYS.DBMS_AQADM.NORMAL_QUEUE
    ,MAX_RETRIES => 2
);
-- Queue starten
SYS.DBMS_AQADM.START_QUEUE(QUEUE_NAME => 'AQ_JMS_HGB_WPS');
END;

```

Listing 3

```

... -- Bevorzugtes Messagingverfahren auswerten
IF IN_BUFFERED_PERSISTANT = ,BUFFERED'
THEN
-- Prüfung der Länge
-- Wenn > max VARCHAR, dann kann die Message nicht als
-- Buffered verschickt werden
IF IN_OBJECT.text_len <= 32767
THEN
    enq_opt.VISIBILITY := DBMS_AQ.IMMEDIATE;
    enq_opt.DELIVERY_MODE := DBMS_AQ.BUFFERED;
END IF;
END IF;
-- Nachrichten für explizite Recipients?
IF IN_RECIPIENTS.COUNT() <> 0
THEN
    msg_prop.recipient_list := IN_RECIPIENTS;
END IF;

DBMS_AQ.Enqueue(QUEUE_NAME => IN_QUEUE_NAME
    ,ENQUEUE_OPTIONS => enq_opt
    ,MESSAGE_PROPERTIES => msg_prop
    ,PAYLOAD => IN_OBJECT
    ,MSGID => msg_handle
);
...

```

Listing 4

Am einfachsten können die Nachrichten empfangen werden, indem der Topic-Receiver einen Message-Listener zugewiesen bekommt, der dann

für jede Nachricht aufgerufen wird. **Listing 6** zeigt ein Beispiel für einen Message-Listener, der die eingehenden Nachrichten auf der Konsole ausgibt.

Intern wird der Nachrichtenempfang durch das Oracle-JMS-API mittels Polling realisiert. Um die Dequeue-Frequenz einzustellen, können in der Java-VM die Parameter „oracle.jms.minSleepTime“ und „oracle.jms.maxSleepTime“ gesetzt werden. So kann beispielsweise für bestimmte Power-User eine höhere Dequeue-Frequenz als für die große Menge an Benutzern („Massen-User“) eingestellt werden.

Leider bietet das Oracle-JMS-API diese Funktionalität nur für persistente Nachrichten. Wenn Buffered Messages verwendet werden sollen, muss das Verfahren selbst umgesetzt werden. In **Listing 7** ist ein einfacher Polling-Mechanismus implementiert.

Um Buffered Messages empfangen zu können, muss die Methode „AQjmsConsumer.bufferReceive“ verwendet werden, die sowohl persistente als auch nicht-persistente Nachrichten aus dem Topic lesen kann. Die Methode gibt es in unterschiedlichen Varianten, die sich im Wesentlichen dadurch unterscheiden, dass sie die Nachrichten „blockierend“ oder „nicht-blockierend“ aus der Queue lesen. Welche der Methoden zum Tragen kommt, sollte davon abhängig gemacht werden, wie viele Clients an einem Topic lauschen. Wenn „nicht-blockierend“ gelesen wird, kommt es zu Latenzzeiten, da zwischen den Lesezugriffen Pausen eingehalten werden müssen, um den Client und die Datenbank nicht zu überlasten. Das blockierende Lesen erlaubt einen schnelleren Empfang der Nachrichten, bindet jedoch aktive Sessions in der Datenbank. Um die Zahl der Sessions in der Datenbank zu minimieren, kann ein JMS-Server zwischen der Datenbank und den Clients geschaltet werden, der als Message Broker fungiert.

Fazit

Im Bereich von hochfrequenter Masendaten-Änderung sollte in Bezug auf die Business-Anforderungen überlegt werden, ob alle DML-Änderungen visualisiert werden sollen oder ob es ausreicht, in bestimmten zeitlichen Intervallen einen Snapshot an die GUI zu senden. Buffered Messages haben einen klaren Performance-Vorteil und sind ideal für Systeme, bei denen die Ausfallsicherheit nicht im Vordergrund

```
import java.io.*;
import javax.jms.*;
import oracle.jms.*;

public class TestAqTopics {

    public static void main(String[] args) throws Exception {
        String jdbcUrl = ...;
        String userName = ...;
        String password = ...;
        String topicOwnerName = ...;
        String topicName = ...;
        String consumerName = ...;
        MessageListener msgListener = new TestAqMessageListener();
        TestAqThread thread = null;
        boolean persistentMessagesOnly = true;

        // Topic Connection aufbauen
        TopicConnectionFactory tcf = AQjmsFactory
            .getTopicConnectionFactory(jdbcUrl, null);
        TopicConnection con = tcf.createTopicConnection(userName,
            password);
        con.start();

        // Topic abonieren
        AQjmsSession session = (AQjmsSession) con
            .createTopicSession(false, Session.CLIENT_ACKNOWLEDGE);
        Topic topic = session.getTopic(topicOwnerName, topicName);
        TopicReceiver receiver = session.createTopicReceiver(
            topic, consumerName, "");
        if (persistentMessagesOnly) {
            // Wenn mit persistenten Nachrichten gearbeitet wird,
            // können wir den MessageListener direkt setzen
            receiver.setMessageListener(msgListener);
        } else {
            // Für nicht-persistente Nachrichten, muss das Polling
            // selbst implementiert werden
            AQjmsConsumer consumer = (AQjmsConsumer) receiver;
            thread = new TestAqThread(consumer, msgListener);
            thread.start();
        }
        String cmd = "";
        BufferedReader in = new BufferedReader(
            new InputStreamReader(System.in));

        while (!"QUIT".equals(cmd)) {
            cmd = in.readLine();
        }
        if (thread != null) {
            thread.interrupt();
            thread.join();
        }
        con.close();
    }
}
```

Listing 5

steht. Bei deren Einsatz sind die jeweils aktuellen Restriktionen des Datenbank-Release zu bewerten. Zurzeit werden folgende Oracle-Streams-Advanced-Queuing-Features für Buffered Messages nicht unterstützt:

- Message retention
- Message delay

- Transaction grouping
- Array enqueue
- Array dequeue
- Message export and import
- Posting for subscriber notification
- Messaging Gateway

Als Nachrichtenformat haben sich die Autoren für JSON entschieden, da die-

```
import javax.jms.*;

public class TestAqMessageListener implements MessageListener {
    @Override
    public void onMessage(Message message) {
        try {
            TextMessage tm = (TextMessage) message;
            String text = tm.getText();
            System.out.println("Message received: " + text);
        } catch (JMSEException e) {
            e.printStackTrace();
        }
    }
}
```

Listing 6

```
import javax.jms.*;
import oracle.jms.*;

public class TestAqThread extends Thread{

    private static long WAIT_TIME = 500;
    private AQjmsConsumer receiver;
    private MessageListener listener;

    public TestAqThread(AQjmsConsumer receiver, MessageListener listener){
        this.receiver = receiver;
        this.listener = listener;
    }

    public void run(){
        while(!isInterrupted()){
            try {
                Message msg = receiver.bufferReceive(1);
                if (msg != null) {
                    listener.onMessage(msg);
                }
            } catch (Exception e) {
                e.printStackTrace();
            }
            try {
                sleep(WAIT_TIME);
            } catch (InterruptedException e) {
                break;
            }
        }
    }
}
```

Listing 7

ses Format gegenüber XML beim Erstellen und Parsen schneller und die Nachrichtengröße deutlich kompakter ist. Als Payload-Type bietet JMS Vorteile bei der Performance und der Flexibilität bezüglich Nachrichtengröße und Zusatzattributen im Vergleich zu XML und RAW.

Ein Nachteil der hier vorgestellten Implementierung ist noch, dass bei

sehr großen Tabellen und der Tatsache, dass eine Indizierung der „ORA_ROWSCN“ in der aktuellen Datenbank-Version nicht möglich ist, der Snapshot etwas länger dauert. Ein Enhancement-Request wurde jedoch schon von Oracle angenommen.

Das JMS-API von Oracle wirkt an einigen Stellen noch nicht ausgereift.

Es fehlt die Möglichkeit, Buffered Messages mit einem Message-Listener zu empfangen und beim nicht-blockierenden Lesen von Buffered Messages kommt es dazu, dass Nachrichten verloren gehen. Als Workaround kann blockierend gelesen werden, mit einem Timeout von einer Millisekunde.

Andriy Terletskyy
andriy.terletskyy@berenberg.de



Nis Nagel
nis.nagel@berenberg.de



Michael Griesser
michael.griesser.ext@berenberg.de



Thies Rubarth
thies.rubarth.ext@berenberg.de



Quasi-modale Dialoge in Apex

Frank Weyher, ORBIT Gesellschaft für Applikations- und Informationssysteme mbH

Dieser Artikel richtet sich in erster Linie an Apex-Entwickler, die modale Fenster, wie sie in traditionellen GUIs gang und gäbe sind, realisieren möchten.

Im Rahmen einer größeren Apex-Applikationslandschaft bieten sich für beliebige Geschäftsobjekte applikationsübergreifend Funktionen an. Diese Funktionen, wie zum Beispiel eine Notiz- oder Dokumentenanhangs-Funktion, sollten durch einen Klick auf ein Icon erreichbar sein. Seit einiger Zeit unterstützt Apex Plug-ins, um den Entwickler von der Codierung einer eigenen Komponente zur Simulation modaler Dialoge zu befreien. Eine kurze Recherche führte zu dem Plug-in „SkillBuilders Modal Page“ (siehe <http://skillbuilders.com/oracle-apex/Apex-Consulting-Training.cfm?category=apex-plugin-ins>), das diese Anforderungen unterstützt. Der Autor verwendet das Plug-in in der Version 2.0.0 in einer Apex-4.2-Umgebung.

Notwendige Schritte

Die Konfiguration und Verwendung des Plug-ins ist einfach und erfordert einige im Folgenden beschriebene Maßnahmen. Das Plug-in wird über das Apex-Import-Utility des Application Builder importiert. Die Konfigurationseinstellungen in den „Shared Components → Component Settings“ sind eine Frage des persönlichen Geschmacks beziehungsweise des Corporate Designs. Allerdings

brauchte das CSS ein wenig Pflege, siehe Abschnitt „Was noch zu tun bleibt“.

Der nächste Schritt besteht darin, eine Dynamic Action (DA) zu erstellen. Diese soll das Klick-Ereignis auf das Icon verarbeiten. Es darf sich dabei sowohl in einem Bericht als Berichtsspalte als auch in einem Formular befinden. Die „Global Page“ ist der ideale Ort, um diese DA zu definieren, weil sie dann auf jeder Seite zur Verfügung steht. **Tabelle 1** zeigt die nötigen Einstellungen (nicht genannte Eigenschaften bleiben bei den Standardwerten). „md“ steht in diesem Fall für „modaler Dialog“. **Tabelle 2** zeigt die Eigenschaften der (True-)Action.

Auch bei der Definition von Buttons für Formulare kann man sich doppelte Arbeit ersparen. Dazu legt man sich ein eigenes Template für den Button an, um die nötigen Einstellungen nur einmal definieren zu müssen. Der Autor hat es als Anchor-Element mit einem Bild de-

finiert (siehe **Listing 1**). **Tabelle 3** zeigt die Bedeutung der einzelnen Attribute.

Der Aufruf des quasi-modalen Dialogs kann nun in Berichten anhand eines Column-Links und in Formularen durch einen Button erfolgen. Wichtig ist, dass in den Link-, beziehungsweise Button-Attributen bestimmte Eigenschaften gesetzt sind (siehe **Tabelle 4**).

Apex besteht darauf, dass die URL einen Link enthält. Er wird jedoch nicht benötigt, deshalb wird dort „javascript:void(0)“ eingetragen. In der Definition der DA wurde ein dynamischer „Event Scope“ mit einem bestimmten Namen festgelegt, der nun hier erforderlich ist. Im Header der Berichtsregion wird „<div id=“region_with_md_link“...>“ eingefügt.

Ist der „Display Point“ des Buttons „Above ...“ oder „Below Region“, kann das „div“-Tag im Regions-Header des Formulars definiert werden. Andernfalls bleibt nur der Seiten-Header.

```
<a href="#LINK#" style="padding:10px" onclick="return false;" #BUTTON_ATTRIBUTES# id="#BUTTON_ID#" class="md-link">

</a>
```

Listing 1

Attribut	Wert	Bemerkung
Event	Click	
Selection Type	jQuery Selector	
jQuery Selector	.md-link	
Event Scope	Dynamic	Bitte unbedingt den Punkt vor „md-link“ beachten. Er bedeutet, dass sich der jQuery-Selector auf das class-Attribut bezieht. Dieser Selector sorgt also dafür, dass alle Seitenelemente mit dem class-Attribut „md-link“ empfindlich für die DA sind.
Static Container	#region_with_md_link	Damit in Reports beim Weiterblättern die Linkspalte auch auf den Folgeseiten des Reports funktioniert, muss der Event-Scope den Wert „Dynamic“ haben. Der hier verwendete Name taucht dann wieder als „id“ im Regions-Header des Berichts auf.

Tabelle 1: Parameter der Dynamic Action

Attribut	Wert	Bemerkung
Fire On Page Load	No	
Dialog Title	Mein Pop-Up	
URL Location	Attribute of Triggering Element	Da kontextabhängige Informationen an die Seite im modalen Dialog transportiert werden sollen, kann es nicht „Statically Defined“, sondern muss „Attribute of Triggering Element“ sein.
Attribute Name	href_md	Dass hier nicht „href“ verwendet wird, sondern das selbst definierte Attribut „href_md“, liegt an der Art und Weise, wie Apex-Buttons die URLs-Aufrufe behandelt. Sie werden durch einen JavaScript-Aufruf maskiert, was dazu führt, dass der Aufruf des Plug-ins nicht funktioniert. Dadurch kann der URL nicht direkt über die Apex-Eigenschaften definiert werden, sondern muss direkt in den Link-Attributen erfolgen.
Autoclose On Element Selector	div#success-message	Standard – Die Dialoge werden nicht automatisch geschlossen, sondern müssen aktiv durch den Benutzer beendet werden.
Dialog Height/Width Mode	Auto	

Tabelle 2: Parameter der True Action der Dynamic Action

Attribut	Wert	Bemerkung
onclick	"return false;"	Damit der Klick auf den Button/Icon durch die DA behandelt wird.
class	md-link	
href	"#LINK#"	Wird nicht verwendet.
Style	"padding:10px"	Beliebig. In diesem Fall sorgt es für einen 10-Pixel-Abstand zu den anderen Buttons.

Tabelle 3: Button-Template-Attribute

Attribut	Wert	Bemerkung
class	md-link	Der Name der Klasse „md-link“ muss mit dem jQuery-Selektor aus dem Abschnitt „Definition einer Dynamic Action zum Aufruf des Plug-ins“ übereinstimmen.
onclick	"return false"	Dieser „onclick“-Handler sorgt dafür, dass die DA den Click behandelt.
title	"Mein Pop-Up"	Der Titel produziert einen Tooltip, der natürlich beliebig ist.
href_md	"?p=&APP_ID. :10001 :&SESSION. :DEFAULT :&DEBUG. :<Parameter>"	Das Attribut „href_md“ muss mit dem Attributnamen aus der Definition der DA übereinstimmen und ist die URL, die beim Klick auf den Link aufgerufen wird, in diesem Fall die Applikationsseite „10001“. Sie übergibt als „REQUEST“ den Kontext „DEFAULT“. Die weiteren Parameter sind natürlich applikationsspezifisch und enthalten typischerweise einen Objekttyp und die Id des Objekts.

Tabelle 4: Link-Eigenschaften

Die Popup-Fenster zeigen herkömmliche Apex-Seiten auf der Grundlage des Popup-Templates. Dieses bietet die typische reduzierte Optik. Es wird auch durch einige Standard-Themes bereitgestellt; gegebenenfalls muss man es noch an seine Bedürfnisse anpassen.

Für die Parameter-Übergabe an Popup-Seiten hat man alle auch sonst in Apex verfügbaren Möglichkeiten. Deshalb sind sie eine elegante Methode, um eigene Auswahldialoge zu implementieren oder auch zentrale Funktionen bereitzustellen.

Zu beachten ist, dass diese Seiten nicht zu unterschiedliche Größen besitzen sollten, besonders dann, wenn man mit horizontalen Listen in den Dialogen arbeitet, die die Registeroptik nachempfinden. Dies führt zu einem ruhigeren Bild beim Wechsel der Seiten.

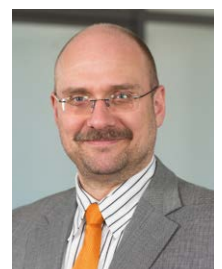
Was noch zu tun bleibt

Zurzeit muss jede nutzende Applikation innerhalb der Applikationslandschaft die DA selber definieren. Ein weiteres Ziel ist – um den Aufwand in den nutzenden Applikationen möglichst gering zu halten –, dass sie nur einen Link aufrufen und die zentrale Apex-Applikation sich um den Rest kümmert. Leider verhält sich die zentrale DA, die etwa auf „Page Load“ als auslösendes Ereignis reagiert, nicht wie gewünscht.

Das Plug-in bietet ein sogenanntes „Custom Theme“ an, das in den Component-Settings des Plug-ins eingestellt ist. Die dafür nötige „css“-Datei darf dazu nicht bei den Plug-in-Dateien liegen. Sie muss, damit sie vom Plug-in gefunden wird, im Bereich der Static-Files abgelegt werden. Zurzeit gibt es allerdings noch

Schwierigkeiten beim Laden der zusätzlichen Ressourcen (wie zum Beispiel Bilder). Der Ersetzungsmechanismus für „#WORKSPACE_IMAGES#“ funktioniert nicht wie erwartet. Der Autor hat sich damit beholfen, ein existierendes Theme gemäß seinen Wünschen anzupassen.

Frank Weyher
frank.weyher@orbit.de



Oracle Hidden Secrets: Netzwerk-Verschlüsselung ohne jeden Aufwand

Heinz-Wilhelm Fabry, ORACLE Deutschland B.V. & Co. KG

Die Konfiguration der Verschlüsselung von Daten aus Oracle-Datenbanken bei der Übertragung über Netzwerk-Verbindungen ist ein Kinderspiel. Bis Mitte des Jahres 2013 war die Netzwerk-Verschlüsselung allerdings Teil der lizenz- und kostenpflichtigen Advanced-Security-Option (ASO) und wurde deshalb nicht von allen Anwendern genutzt. Inzwischen ist das Sicherheits-Feature aber kostenlos. Damit sind alle DBAs aufgefordert, die bisher keine Netzwerk-Verschlüsselung einsetzen, diese umgehend zu implementieren. Der Artikel beschreibt, wie das am einfachsten und in wenigen Minuten geht.

Mit Freigabe der Datenbank 12c, Release 1, wurde Ende Juni 2013 eine ganze Reihe von Sicherheits-Features, die bis dahin Bestandteil der Advanced-Security-Option waren, als Feature der Datenbank verfügbar – und zwar sowohl in der Enterprise Edition als auch in der Standard Edition sowie auch in den vorangegangenen Releases (wie 11g), sofern diese das technisch unterstützen. Damit sind die starken Authentifizierungsmethoden Kerberos, PKI und RADIUS sowie die Verschlüsselung des Netzwerks über SQL*Net in nativer Form und auch über SSL/TLS ohne zusätzliche Kosten verfügbar. Die Änderung der Lizenzbedingungen ist unter anderem unter „http://docs.oracle.com/cd/E16655_01/license.121/e17614/options.htm#DBLIC143“ im Handbuch „Oracle Database Licensing Information“ dokumentiert. Da sich

die Netzwerk-Verschlüsselung über SQL*Net innerhalb weniger Minuten aktivieren lässt, wird hier die Konfiguration dieser Verschlüsselungsvariante dargestellt.

Das gesamte Netzwerk oder nur Datenbank-Zugriffe verschlüsseln?

Die Verschlüsselung des gesamten Netzwerks implementiert eine Verschlüsselung des Datenstroms zwischen den Netzwerkkarten. Das verhindert, dass durch Zugriff auf Datenleitungen Informationen in falsche Hände geraten. Sobald die Daten jedoch von einer Netzwerkkarte innerhalb eines Rechners an eine Applikation – also zum Beispiel an eine Datenbank – weitergeleitet werden, sind diese Daten nicht mehr verschlüsselt. Jeder, der Zugriff auf den Rechner hat, weil er etwa als Storage- oder System-Administrator ar-

beitet oder weil er sich in das Betriebssystem eingehackt hat, kann alle über das Netzwerk gesendeten Daten lesen.

Die Verschlüsselung des Netzwerks über SQL*Net verhält sich dagegen anders. Hier endet die Verschlüsselung nicht mit den Netzwerkkarten, sondern erst in den beteiligten Oracle-Komponenten: Man muss sich also in der Datenbank oder Datenbank-Anwendung einloggen und zugriffsberechtigt sein, um Daten lesen zu können. Die Verschlüsselung des Datenstroms über SQL*Net bietet daher einen höheren Schutz als die Verschlüsselung des gesamten Netzwerks.

Natürlich kommt es auf die eigene Risikoeinschätzung an, welche Form der Netzwerk-Verschlüsselung eingesetzt wird. Allerdings muss man heutzutage davon ausgehen, dass die kriminelle Energie von Individuen,

organisiertem Verbrechen und unterschiedlicher Geheimdienste irgendeine Form der Netzwerk-Verschlüsselung zwingend notwendig macht.

Verschlüsselung implementieren

Alle zurzeit durch den Oracle-Support unterstützten Datenbank- und Client-Versionen übertragen Authentifizierungs-Informationen grundsätzlich verschlüsselt. Der Schlüssel wird dabei für jede Benutzersitzung/Session immer wieder neu erstellt. Dieser Aspekt der Sicherheit ist also abgedeckt. Nach dem Aufbau der Verbindung erfolgt die gesamte Kommunikation zwischen Datenbank und Anwendung oder Remote-Datenbank allerdings prinzipiell unverschlüsselt.

Die Verschlüsselung, die implementiert werden soll, macht keinen Unterschied zwischen SQL-Befehlen und Aufrufen prozeduraler Komponenten oder Daten, die von einem Client an die Datenbank – zum Beispiel bei einem „INSERT“ und „UPDATE“ – oder von der Datenbank an einen Client, etwa als Ergebnis eines „SELECT“, geschickt werden. Der gesamte Informationsstrom ist verschlüsselt.

Im einfachsten Fall erfolgt dies über einen einzigen Eintrag in der Datei „SQLNET.ORA“, und zwar über eine Angabe zu den Parametern „SQLNET.ENCRYPTION_CLIENT“ (auch eine Datenbank kann Client sein) und/oder „SQLNET.ENCRYPTION_SERVER“. Die Einstellung „SQLNET.ENCRYPTION_SERVER=REQUIRED“ auf der Datenbank-Seite führt dazu, dass ausschließlich eine verschlüsselte Kommunikation mit dieser Datenbank erlaubt ist. Jeder andere Kommunikationsversuch wird mit einer Fehlermeldung abgelehnt.

Das ganze System ist allerdings sehr flexibel, denn neben „REQUIRED“ (Verschlüsselung ist obligatorisch) stehen für beide Seiten (Server und Client) drei weitere Einstellungen zur Verfügung:

- **REJECTED**
Verschlüsselung wird grundsätzlich abgelehnt
- **ACCEPTED (Default)**
Verschlüsselung wird akzeptiert, wenn der Kommunikationspartner das möchte

- **REQUESTED**
Verschlüsselung wird gewünscht, aber nicht verlangt

Steht der Parameter für eine Datenbank beispielsweise auf „REQUESTED“, könnten die Einstellungen für die darauf zugreifenden Clients nach Wunsch unterschiedlich gesetzt sein. Während einzelne Clients über ihre „SQL*NET.ORA“-Dateien eine Verschlüsselung erzwingen („REQUIRED“), könnten andere – sofern das Sinn ergibt – darauf verzichten („REJECTED“).

Entsprechend den Einstellungen auf beiden Seiten ergibt sich somit, ob verschlüsselt wird oder nicht, beziehungsweise ob überhaupt eine Datenbank-Verbindung zustande kommt. Per Default gilt für beide Seiten der Wert „ACCEPTED“. Das hat zur Folge, dass beide Seiten für eine Verschlüsselung zur Verfügung stehen, sie diese aber nicht ausdrücklich wünschen. In der Praxis führt das dann dazu, dass keine Verschlüsselung stattfindet. Die **Abbildung** zeigt die möglichen Kombinationen und ihre Auswirkungen auf die Verschlüsselung beziehungsweise den Verbindungsaufbau zwischen den Kommunikationspartnern.

Neben der grundsätzlichen Festlegung, ob verschlüsselt wird oder nicht, können weitere Einstellungen sinnvoll sein. Zunächst lässt sich der Algorithmus der Verschlüsselung über die Parameter „SQLNET.ENCRYPTION_TYPES_SERVER/_CLIENT“ bestimmen. Je nach Version der Datenbank beziehungsweise des Oracle-Client stehen verschiedene Algorithmen zur Verfügung. Für

ältere Datenbank-Versionen sind die verfügbaren Algorithmen im Handbuch „Advanced Security Administrators Guide“ beschrieben. Für die Datenbank 12c erläutert das Handbuch „Oracle Database Security Guide“ die Algorithmen. Ist kein Algorithmus festgelegt, vereinbaren Client und Server bei der ersten Kontaktaufnahme („handshake“) diesen Algorithmus zufällig aus der Reihe der gemeinsam zur Verfügung stehenden Algorithmen. Weil die Algorithmen unterschiedlich sicher und unterschiedlich performant sind, empfiehlt sich eine explizite Angabe. Der Algorithmus der Wahl ist hier „AES128“.

Darüber hinaus sind Prüfsummen-Verfahren möglich, um sicherzustellen, dass Daten im Transfer nicht modifiziert werden. Auch dies wird über Parameter in der Datei „SQLNET.ORA“ konfiguriert. Diese könnte schließlich verschiedenste Einträge enthalten (**siehe Listing**).

Die „CLIENT“-Parameter werden angegeben, da eine Datenbank, wie bereits mehrfach erwähnt, auch als Client agieren kann. Die gleiche Datei könnte auch auf der Client-Seite zum Einsatz kommen, vorausgesetzt der Client unterstützt die angegebenen Algorithmen. Dass auch datenbankseitige Parameter spezifiziert sind, stört dabei nicht.

Hinweise

Die Wirksamkeit der Netzwerk-Verschlüsselung lässt sich ohne fremde Hilfsmittel überprüfen. Dazu wird in der Datei „SQLNET.ORA“ nur das Tracing aktiviert „trace_level_client = 16“.



Native Verschlüsselung einschalten

		Client			
		REJECTED	ACCEPTED	REQUESTED	REQUIRED
Server	REJECTED	aus	aus	aus	Keine V.*
	ACCEPTED	aus	aus**	ein	ein
	REQUESTED	aus	ein	ein	ein
	REQUIRED	Keine V.*	ein	ein	ein

* Keine Verbindung ** Default ist Accepted

Abbildung: Einstellungen von „SQLNET.ENCRYPTION_CLIENT/_SERVER“

Jede neue Datenbank-Sitzung wird ohne Neustart der Datenbank ab sofort eine Trace-Datei erzeugen. In dieser kann man dann bei nicht eingeschalteter Verschlüsselung Daten und Befehle im Klartext finden, während bei eingeschalteter Verschlüsselung nichts Sinnvolles mehr zu erkennen ist.

Die hier vorgestellte Vorgehensweise gilt für alle Client-Server-Verbindungen, die über OCI oder über Java „thick“- und „thin client“-Anwendungen arbeiten. Mögliche Ausnahmen könnten lediglich ältere „thin client“-Anwendungen betreffen. Dies muss man im Zweifelsfall testen.

Testen sollte man zur eigenen Beruhigung auch, ob das Einschalten der Netzwerk-Verschlüsselung Auswirkungen auf die Performance des Systems hat. Dies ist sehr unwahrscheinlich – jedenfalls berichten Kunden in den Veranstaltungen der Oracle Business Unit Database immer wieder, dass sie

```
sqlnet.encryption_server = required
sqlnet.encryption_types_server = AES128
sqlnet.encryption_client = required
sqlnet.encryption_types_client = AES128
sqlnet.crypto_checksum_server = required
sqlnet.crypto_checksum_types_server = SHA1
sqlnet.crypto_checksum_client = required
sqlnet.crypto_checksum_types_client = SHA512
```

Listing

keinerlei Performance-Unterschiede zwischen einem System erkennen, das verschlüsselt, und einem, das unverschlüsselt überträgt.

Fazit:

Netzwerk-Verschlüsselung mit Oracle-Datenbankmitteln ist leicht zu implementieren und kostet weder Geld noch Performance. Diese Möglichkeiten sollte man zu seiner eigenen Sicherheit nutzen.

Heinz-Wilhelm Fabry
heinz-wilhelm.fabry@oracle.com



Tipps und Tricks aus Gerds Fundgrube

Heute: Zeituntersuchungen im Mikrosekundenbereich

Autor: Gerd Volberg, OPITZ CONSULTING GmbH

Die normalen Date-Funktionen reichen nicht aus, um Zeiträume unter einer Sekunde zu messen. Solche Untersuchungen ermöglicht der Datentyp „Systemstamp“.

Das Beispiel in [Abbildung](#) zeigt, wie man diese Technik einsetzt. In der Routine „Zeitmessung“ wird die Zeit untersucht, die ein Test-Loop benötigt. Dazu speichert man zwei Zeiten, deren Delta die Dauer des Loops ergibt (siehe [Listing](#)).

Im Beispiel sieht man nun, dass der Loop genau 94,688 ms dauerte. Diese Genauigkeit ist nur mit „Systemstamp“ möglich.

```
Start 06.12.13 07:24:06,010864
Ende 06.12.13 07:24:06,105552
Delta +0000000000 00:00:00.094688000
```

Abbildung: Das Beispiel



```

PROCEDURE Zeitmessung IS
  v_timestamp_1 timestamp := systimestamp;
  v_timestamp_2 timestamp;
begin
  FOR I IN 1 .. 10000 LOOP
    message (I, no_acknowledge);
  END LOOP;
  v_timestamp_2 := systimestamp;
  :Control.TI_Text := 'Start ' || v_timestamp_1 || const.cr ||
    'Ende ' || v_timestamp_2 || const.cr ||
    'Delta ' || to_char (v_timestamp_2 - v_timestamp_1);
END;
    
```

Listing

Oracle Application Express in der Praxis

gelesen von Niels de Bruijn

Es ist schon eine Weile her, seitdem das letzte deutschsprachige Buch über Oracle Application Express (Apex) erschienen ist. Die Entwicklung des Produkts ist rasant, daher ist das Buch „Oracle Application Express in der Praxis“ von Ralf Beckmann zu begrüßen. Ansprechen wird dieses Buch besonders (Oracle)-Entwickler, die mit Apex 4.2.x beginnen. Anhand eines Bei-

spiel-Projekts wird dem Leser Schritt für Schritt gezeigt, wie man eine Anwendung mit Apex-4.2-Standard-Features entwickelt. Dabei wird sowohl die Entwicklung für den Desktop als auch die für mobile Endgeräte thematisiert. Übergreifende Themen wie Installation, Architektur, Mehrsprachigkeit, Debugging, Sicherheit, Plug-ins, Websheets und Deployment kommen ebenfalls zur Sprache.

Das Buch besticht durch seinen klaren Aufbau und die Lesefreundlichkeit. Entwickler, die Anforderungen umsetzen müssen, die über den Apex-Assistenten hinausgehen, werden viel Freude an diesem Buch haben. Man wird die Architektur und die Möglichkeiten von Apex 4.2 besser verstehen und dadurch letztendlich produktiver in der Umsetzung von Web-Applikationen auf Basis von Apex. Vorsicht ist dennoch geboten: Das Buch bietet kaum Best



Practices und beschränkt sich im Wesentlichen auf die Standard-Features von Apex 4.2. Man sollte daher zu Projektbeginn in eine Schulung und/oder Coaching durch Apex-Experten investieren. Schließlich entscheidet ein Entwickler selbst über den Erfolg von Apex im Unternehmen. **Tabelle** zeigt weiterführende Quellen, die man auf dem Weg zum Apex-Experten verfolgen sollte.

Titel:	Oracle Application Express in der Praxis
Autor:	Ralf Beckmann
Verlag:	Hanser
Umfang:	416 Seiten
Sprache:	Deutsch
Preis:	59,90 Euro (Online-Ausgabe 47,99 Euro)
ISBN:	978-3-446-43896-5

Deutschsprachige Apex Community mit vielen Tipps	https://blogs.oracle.com/apexcommunity_deutsch
Demo-Anwendung von Denes Kubicek	https://apex.oracle.com/pls/apex/f?p=31517:101
Open-Source-Lösungen, Videos, Demos, Präsentationen	https://apex.mt-ag.com
Apex Blog Aggregator (verfolgt über 160 Apex Blogs!)	http://www.odtug.com/apex
Apex Forum	http://forums.oracle.com
Oracle Learning Library	http://bit.ly/1fwirpo
Apex-Dokumentation	http://apex.oracle.com

SQL Thinking

gelesen von Andreas Schmidt

„Denken in SQL“ könnte man den Buchtitel übersetzen. Das klingt nicht so treffend, vielleicht wurde deshalb hier ein englischer Buchtitel gewählt. Das Buch hingegen ist komplett auf Deutsch geschrieben. Die Autoren führen den Leser schrittweise in die relationale Abfragesprache SQL ein, und zwar „von der kaufmännischen Fragestellung hin zum SQL-Statement“.

Das Ziel ist kein einfaches Unterfangen, schließlich kämpfen täglich erfahrene Programmierer und DBAs mit komplexen SQL-Statements. Hierzu stehen zum Download Beispieldaten als DMP-Files zur Verfügung, außerdem werden die bekannten Beispiele des Oracle-Users „SCOTT“ verwendet. Damit kann der Leser die SELECT-Statements selbst ausprobieren, um ein besseres Gefühl für die Sprache SQL zu erhalten. Darum geht es auch im Wesentlichen in diesem Buch: Ein Gefühl für typische SQL-Konstrukte zu erfahren. Das Buch ist besonders für Einsteiger geeignet, die sich die Zeit nehmen, mit den Testdaten alle Beispiele nachzuvollziehen. Anders ausgedrückt: Das Buch kann als Kursersatz für Autodidakten dienen, um einen effektiven Einstieg in Oracle SQL zu erhalten. Nicht mehr, aber auch nicht weniger.

Im Buch sind neben den elementaren Projektionen und Filtern auch „INNER“ und „OUTER JOIN“ sowie Subqueries

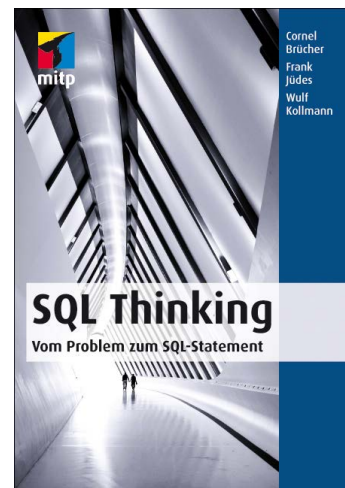
behandelt. Der „GROUP BY“ ist mit den Optionen „ROLLUP“ und „CUBE“ erläutert. Die rekursive Suche im Netzwerk mit dem „CONNECT BY“-Operator fehlt ebenfalls nicht. Darüber hinaus sind auch analytische Funktionen wie „LAG“ dargestellt. Damit geht dieses Werk über Umfang und Komplexität mancher Einführungsbücher hinaus.

Weitere Kapitel über Datenmodelle, DML-Befehle und Transaktionen sind eher „Abstecher“ und werden nur knapp skizziert. Dies trifft auch für die Auflistung der SQL-Funktionen im letzten Teil des Buches zu.

Fazit

Der Hinweis „Auch für Nichtprogrammierer“ auf der Rückseite bekräftigt, dass das Autorenteam auch nicht-technische Mitarbeiter ansprechen möchte, die selbstständig Anfragen an eine Oracle-Datenbank stellen wollen. Dies erklärt auch den Schwerpunkt auf den „SELECT“-Statements. Der humorvolle Stil im Buch entspricht einer lakonischen, aber durchaus realistischen Betrachtung der EDV-Realität. Zusätzliche Übungsaufgaben in jedem Kapitel wären eine sinnvolle Ergänzung gewesen.

Übrigens haben zwei der Autoren auch an dem Buch „Oracle Survival Guide“ mitgeschrieben, das bereits in der DOAG News Q3/2010 besprochen wurde. Das Buch wurde zwar bereits im



Jahr 2010 herausgegeben, ist aber in seiner Intention, SQL zunächst zu verstehen und in einem weiteren Schritt umzusetzen, weiterhin geeignet. Außerdem kam Ende 2012 ein umfassenderes Werk speziell für den MS SQL Server auf den Markt, der Titel lautet entsprechend „Microsoft SQL Thinking“.

Titel:	SQL Thinking – vom Problem zum SQL-Statement
Autoren:	Cornel Brücher, Frank Jüdes, Wulf Kollmann
Verlag:	mitp
Umfang:	362 Seiten
Sprache:	Deutsch
Preis:	29,95 Euro
ISBN:	978-3-8266-5505-0

Andreas Schmidt
andreas.schmidt98@freenet.de

Wir begrüßen unsere neuen Mitglieder

Persönliche Mitglieder

Benjamin Gampert
Theodor Ramminger
Janko Despotovic
Markus Böhmer
Jan Bossmann
Uwe Brusdeilins
Gerd Aiglstorfer
Marco Roth

Nikolay Kostadinov
Dennis Timotin
Gerhard Zischka
Frank Günther
Reto Kiefer
Andreas Ellerhoff
Henrik Nietschke
Simon Herfert
Christian Nietschke

Firmenmitglieder

Crayon Deutschland GmbH, Oliver Hartmann
pasolfora GmbH, Andreas Prusch
etomer GmbH, Jörg Stiller
baden IT GmbH, Gerhard Welte

#DOAG 2013

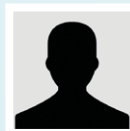
Eindrücke von der bislang größten DOAG-Anwenderkonferenz



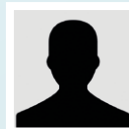
Peter Raganitsch @PeterRaganitsch
alle Papers für #DOAG2013 fertig bewertet. Viel und guter Content, freue mich schon auf die Konferenz
5. Juli 2013



Carsten Czarski @cczarski
Programm ist online. <http://www.doag.org/de/de/home/aktuelle-news/article/doag-2013-konferenz-ausstellung-zehn-streams-ueber-400-vortraege-und-jede-menge-aktivitaeten-am.html> ... Verspricht wieder mal eine spannende Konferenz - wie immer ...
5. August 2013



Jörg @joerg_whtvr
Das Programm zur #DOAG2013 ist online: <http://www.doag.org/de/events/konferenzen/doag-2013.html> ... Frühbucherrabatt gilt noch bis Ende September!
6. August 2013



Michel a.k.a. Emil @tbbol
Got my badge today
6. September 2013



Niels de Bruijn @nielsdb
Got my badge today
9. September 2013



BU DBTec @OracleBUDB
Statement Of Direction für APEX 5.0 - <http://www.oracle.com/technetwork/developer-tools/apex/application-express/apex-sod-087560.html> ... - #orclapex #apex5 - Vortrag zum Thema auf der #doag2013y
17. September 2013



DOAG e.V. @DOAGeV
#DOAG2013: 10 Streams, über 400 Vorträge und jede Menge Aktivitäten am Rande
1. Oktober 2013



Carsten Czarski @cczarski
See you in Nuremberg
7. Oktober 2013



edburns @edburns
RT @DOAGeV: Keynote: @edburns über Neuheiten in @jsf_spec auf #DOAG2013
17. Oktober 2013



Christian Antognini @ChrisAntognini
I'm looking forward for the OakTable stream!
18. Oktober 2013



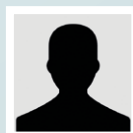
Bjoern Broehl @bbroehl
Letzte Vorbereitungen, morgen geht's los, wir freuen uns! pic.
twitter.com/n5DmBsvlKA
18. November 2013



DOAG e.V. SIG Java @sigjava
#DOAGDevZone: Hands-on - am Rande der Developer
Zone gibt's den Hack Me-Stand von Studenten der HM.
19. November 2013



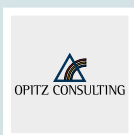
Christian Antognini @ChrisAntognini
Excellent keynote from Peter Kreuz
20. November 2013



Rolf Andreas Just @Handgemenge
RT @esentri: Um 16 Uhr verlosen wir ein iPad Air. #DOAG2013 Be-
sucher können heute noch teilnehmen <http://ow.ly/qOyUL>
20. November 2013



Svenja Schriever @Sv3nja86
Ask Apex: Es geht los! Und wieder gespannt wann die eine
zentrale Frage gestellt wird :-)) pic.twitter.com/e8puASCPqV
21. November 2013



OPITZ CONSULTING @OC_WIRE
Schon probiert? Gebrannte Edelnüsse in der Highland-
Atmosphäre unserer Denkfabrik.
21. November 2013



Christian Schwitalla @Christianschw67
#doag2013 eine fantastische Zeit, wie jedes Jahr.
Unglaubliche Themenvielfal, interessante Men-
schen - ein November-Highlight :-)) :-))
21. November 2013



Johannes Ahrends @streetkiter
The best conference ever
#DOAG2013 ist zu Ende
22. November 2013



Günther Stürner ist DOAG-Botschafter 2013

Als DOAG-Botschafter werden jährlich diejenigen Personen ausgezeichnet, die sich in der Arbeit der DOAG durch hohes Engagement und Kompetenz auszeichnen. Dabei wird die gesamte Mitwirkung der Person in der DOAG berücksichtigt: Aktivitäten auf Veranstaltungen und in elektronischen Netzwerken sowie das Schreiben von Beiträgen in den DOAG-Medien.

Ein Komitee der DOAG, bestehend aus fünf Vorständen und Führungskräften aus allen Communities, hat sich in diesem Jahr entschieden, den Preis „DOAG Botschafter 2013“ in der Kategorie „Technologie“ einer Person zuzusprechen, die der DOAG schon seit sehr langer Zeit eng verbunden ist: Günther Stürner, Vice President Server Technologies und Sales Consulting bei der ORACLE Deutschland B.V. & Co. KG, in dieser Funktion verantwortlich für die Bereiche „Datenbank“ und „Middleware“.

Günther Stürner hat von Beginn an bei allen DOAG-Konferenzen in irgendeiner Form mitgewirkt, meist selbst als Referent, der die neuesten Datenbank-Technologien immer anwendernah vermitteln konnte, oder als Motivator seiner Mitarbeiter, denen er die Wichtigkeit des Kunden-Feedbacks auf den DOAG-Konferenzen nahelegte. Er ist auch ein Mann der gesamten Community, der immer ein offenes Ohr für die Wünsche der DOAG hatte, es aber auch nicht scheute, sich

mit seiner eigenen Meinung gegenüber der DOAG zu positionieren. Aber auch wenn Oracle der DOAG nicht wohl gesonnen war, hat er die DOAG immer unterstützt, und bei vielen Wochenend-Veranstaltungen der DOAG seine Freizeit geopfert, um Rede und Antwort zu stehen.

Günther Stürner arbeitet bereits seit September 1985 für Oracle; sein beruflicher Werdegang begann als Sales Consultant Deutschland. Von 1987 bis 1993 widmete er sich dem Aufbau der Abteilung „Strategisch Technische Unterstützung“ (STU) und war anschließend sechs Jahre lang Leiter des Oracle SAP Competence Centers sowie des Oracle SAP Entwicklungszentrums. Er ist heute Vice President Server Technologies und Sales Consulting. Günther Stürner hat mehrere Fachbücher zur Oracle-Datenbank und zu SQL geschrieben sowie zahlreiche Fachartikel veröffentlicht.

Günther Stürner (auf dem Foto rechts) widmete die Anerkennung seinem gesamten Team und regte an, die Trophäe wie einen Wanderpokal der Reihe nach in allen Oracle-Geschäftsstellen in Deutschland herumzugeben.



*Dr. Dietmar Neugebauer
Vorstandsvorsitzender der DOAG*

DOAG-Vorstandssitzung am 6. und 7. Dezember 2013 in Berlin

Trotz des Sturmtiefs über Deutschland konnte der DOAG-Vorstand seine letzte Sitzung im Jahr 2013 planmäßig durchführen. Lediglich Michael Paege, Vorstandsmitglied aus dem Norden, musste wetterbedingt über Skype zugeschaltet werden.

Traditionell stellt zu Beginn der Sitzung ein Auszubildender aus der DOAG-Geschäftsstelle die Mitglieder-

und Veranstaltungsstatistik der DOAG für das abgelaufene Jahr vor. Diesmal übernahm Marius Fiedler diese Aufgabe für 2013. Eine wichtige Information war, dass innerhalb der letzten 24 Monate ein stetiger Rückgang der korporativen Mitglieder zu beobachten ist, bei einem gleichzeitigen Anstieg der persönlichen Mitglieder. Der Vorstand will die Gründe dafür herausfinden. Sehr positiv ist, dass die Anzahl der regionalen Veranstaltungen wieder nennenswert zugenommen hat. Hier nochmals vielen Dank an die regionalen Repräsentanten und deren Stellvertreter.

Im nächsten Sitzungspunkt ließ man nochmals die DOAG 2013 Konferenz + Ausstellung Revue passieren. Nicht nur im Vorstand, sondern auch durch das zahlreiche Feedback waren sich alle einig, dass dies die beste Konferenz der DOAG gewesen war. Dies ist Ansporn und zugleich eine große Herausforderung für die nächste Konferenz im Jahr 2014.

Im Anschluss gab Fried Saacke, DOAG-Vorstand und Geschäftsführer, einen kurzen Statusbericht zu JavaLand 2014. Das Programm steht seit Mitte November. Daneben ist ein umfangreiches Community-getriebenes Programm geplant (Hacker-Session, Hands-on-Sessions, offene Redaktion, Community-Café etc.). Hinzu kommt ein Partnerprogramm, das von Sponsoren getragen wird und als solches gekennzeichnet ist. Der Vorstand stimmte dem weiteren Vorgehen zu dieser Veranstaltung einstimmig zu.

Ein weiterer Schwerpunkt der Vorstandssitzung war die Vorbereitung der Delegiertenversammlung am 14. März 2014 in Düsseldorf. Neben dem Vorsitzenden werden auf dieser Versammlung alle verantwortlichen Community-Leiter und der Leiter der Querschnittsgruppen ihre Tätigkeitsberichte vorstellen. Finanzvorstand Urban Lankes präsentierte den aktuellen Forecast, der auch für das Jahr 2013 einen positiven Abschluss für den Verein erwarten lässt. Zudem stellte er den Entwurf des Budgetplans für 2014 vor. Nach einer ausführlichen Diskussion im Vorstand wurde dem Entwurf zugestimmt. Die-

ser wird der Delegiertenversammlung als Beschlussvorlage vorgelegt. Des Weiteren werden die geplanten Maßnahmen zu den Zielen 2013 bis 2016 auf der Versammlung vorgestellt und mit den Delegierten diskutiert.



Dr. Frank Schönthaler
Leiter der Business Solutions Community

Logistik-4.0-Konzepte und -Anwendungen auf der DOAG 2014 Logistik + IT

Zum Thema „Logistik 4.0“ veranstaltet die DOAG Business Solutions Community in Zusammenarbeit mit Oracle ihre diesjährige Community-Konferenz DOAG Logistik + IT. Sie findet am 7. Mai 2014 in der Logistikmetropole Dortmund statt. Der Veranstaltungsort – das Fraunhofer Institut für Materialfluss & Logistik IML – gilt mit seinem

einzigartigen Innovationsflair als die Industrie-4.0-Zukunftsschmiede.

Die Community-Konferenz behandelt Themen aus den Bereichen Beschaffungs-, Produktions-, Distributions-, Entsorgungs- und Ersatzteile-Logistik sowie „Logistik auf dem Weg zur Industrie 4.0“ und „Geschäftswissen für die Logistik“. Im Mittelpunkt stehen Innovationen in Form intelligenter Logistikprozesse und neuartiger IT-Systeme. Ziel der Konferenz ist es, innovative Technologien verständlich zu erklären und betriebswirtschaftliche Nutzenaspekte anhand von Berichten aus der Unternehmenspraxis aufzuzeigen. Dafür stehen hochkarätige Keynote-Speaker, erfahrene Anwender sowie renommierte Business- und Technologie-Experten. Vielfältige Networking-Elemente laden zur kritischen Diskussion und zur Vertiefung des vermittelten Wissens ein.

Mit dem Fraunhofer IML, dem LogistikCluster NRW und Oracle als Treiber von Prozess- und IT-Innovationen hat die Business Solutions Community starke Partner mit an Bord. Innovations-Atmosphäre garantieren die mit modernsten Technologien gespickten Räumlichkeiten des IML, die die Konferenzteilnehmer im Rahmen eines hoch informativen Rundgangs direkt begutachten können. Eine Fachausstellung und Aktiv-Workshops, in denen Logistik 4.0 live erlebt werden kann, runden das Angebot ab und machen die DOAG Logistik + IT zu einem echten Muss für Logistik- und SCM-Anwender.

Impressum

Herausgeber:

DOAG Deutsche ORACLE-
Anwendergruppe e.V.
Tempelhofer Weg 64, 12347 Berlin
Tel.: 0700 11 36 24 38
www.doag.org

Verlag:

DOAG Dienstleistungen GmbH
Fried Saacke, Geschäftsführer
info@doag-dienstleistungen.de

Chefredakteur (ViSdP):

Wolfgang Taschner, redaktion@doag.org

Redaktion:

Fried Saacke, Carmen Al-Youssef,
Mylène Diacquenod, Dr. Dietmar
Neugebauer, Dr. Frank Schönthaler,
Robert Szilinski

Titel, Gestaltung und Satz:

Alexander Kermas,
DOAG Dienstleistungen GmbH

Titelfoto: © Sergey Nivens / Fotolia.com

Foto S. 23: © ra2 studio / Fotolia.com

Foto S. 34: © AllebaziB / Fotolia.com

Anzeigen:

Simone Fischer, anzeigen@doag.org
DOAG Dienstleistungen GmbH
Mediadaten und Preise finden Sie
unter: www.doag.org/go/mediadaten

Druck:

Druckerei Rindt GmbH & Co. KG
www.rindt-druck.de



14.02.2014
DOAG Webinar: ORACLE Partitioning für Einsteiger
 office@doag.org

19.02.2014
Real World Performance Tour
 Christian Trieb
 christian.trieb@doag.org

20.02.2014
Regionaltreffen NRW (Forms Community)
 Andreas Stephan, Stefan Kinnen
 regio-nrw@doag.org

20.02.2014
Regionaltreffen Nürnberg/Franken
 André Sept
 regio-franken@doag.org



05.03.2014
Regionaltreffen Berlin/Brandenburg
 Michel Keemers
 regio-bb@doag.org

05.03.2014
Regionaltreffen Bremen
 Ralf Kölling
 regio-bremen@doag.org

06.03.2014
Regionaltreffen München/Südbayern
 Andreas Ströbel
 regio-muenchen@doag.org

10.03.2014
Regionaltreffen Osnabrück/Bielefeld/Münster
 Andreas Kother
 regio-osnabrueck@doag.org

14.03.2014
DOAG Delegiertenversammlung 2014
 office@doag.org

14.03.2014
DOAG Webinar: Wer hat die Daten gestohlen? Warum es der DBA nicht gewesen sein kann
 office@doag.org

18.03.2014
Regionaltreffen Jena/Thüringen
 Jörg Hildebrandt
 regio-thueringen@doag.org

20.03.2014
Regionaltreffen Karlsruhe
 Reiner Büniger
 regio-karlsruhe@doag.org

20.03.2014
Regionaltreffen Nürnberg/Franken
 André Sept, Martin Klier
 regio-franken@doag.org

20.03.2014
Regionaltreffen Stuttgart
 Jens-Uwe Petersen
 regio-stuttgart@doag.org

25./26.03.2014
JavaLand 2014 - die Konferenz der Java-Community
 office@doag.org

26.03.2014
SIG Security
 Franz Hüll, Tilo Metzger
 sig-security@doag.org



03.04.2014
DOAG 2014 Business Intelligence
 Christian Weinberger
 office@doag.org

08.04.2014
Regionaltreffen Hamburg/Nord
 Jan-Peter Timmermann
 regio-nord@doag.org

11.04.2014
DOAG Webinar: Proxy Authentication und Remote Login ohne sichtbare Passworte (Wallet)
 office@doag.org

14.04.2014
Regionaltreffen München/Südbayern
 Andreas Ströbel
 regio-muenchen@doag.org

15.04.2014
Regionaltreffen Rhein-Main
 Thomas Tretter
 regio-rhein-main@doag.org

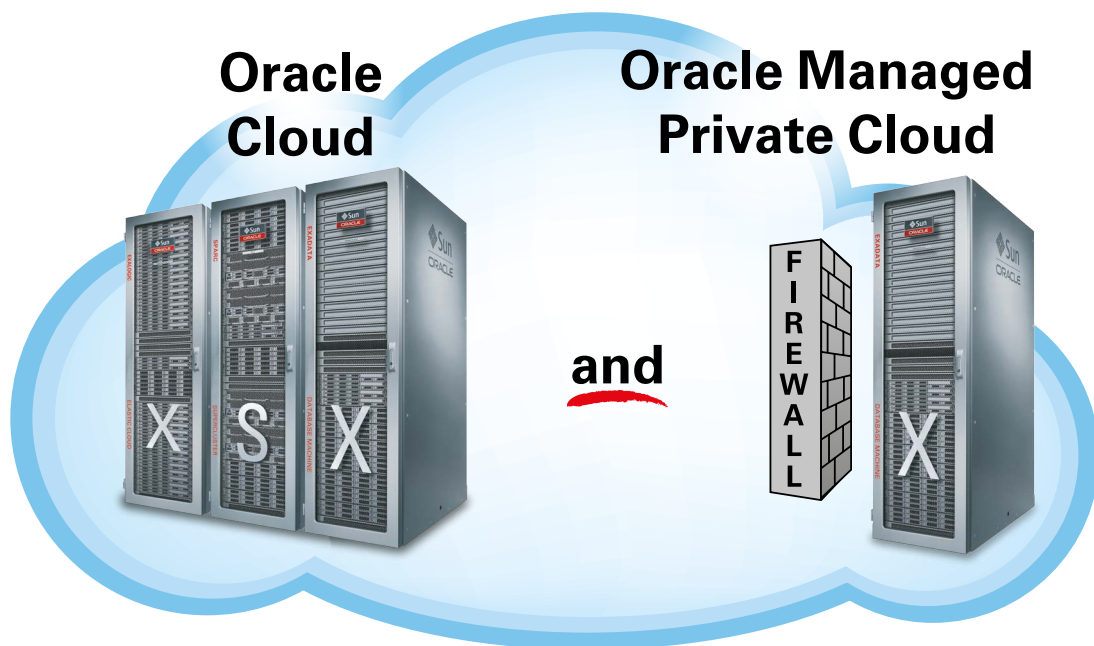
17.04.2014
Regionaltreffen Nürnberg/Franken
 André Sept, Martin Klier
 regio-franken@doag.org

Aktuelle Termine und weitere Informationen finden Sie unter www.doag.org/termine/calendar.php

Unsere Inserenten	
DBConcepts www.dbconcepts.at	S. 49
DOAG Deutsche ORACLE-Anwendergruppe e.V.	U 2
Hunkler GmbH & Co. KG www.hunkler.de	S. 3
Libelle AG www.libelle.com	S. 7
MuniQsoft GmbH www.muniqsoft.de	S. 23
ORACLE Deutschland B.V. & Co. KG www.oracle.com	U 3
ProLicense GmbH www.prolicense.com	S. 33
Trivadis GmbH www.trivadis.com	U 4

Oracle Cloud

Applications, Platform, Infrastructure



Our Data Center

Your Data Center

**Run some of your applications in the Oracle Cloud
and others in your Private Cloud. You Choose.**

ORACLE®

**oracle.com/cloud
or call 0800 1 81 01 11**

Schön, wenn man Zeit gewinnt.



■ Mehr Effizienz für Ihr Unternehmen. Aber auch mehr Zeit für Sie persönlich. Mit unseren IT-Lösungen erhalten Sie beides. Trivadis ist führend bei der IT-Beratung, der Systemintegration, dem Solution-Engineering und bei den IT-Services mit Fokussierung auf Oracle- und Microsoft-Technologien im D-A-CH-Raum. Unsere Leistungen erbringen wir auf den strategischen Geschäftsfeldern Business Intelligence, Application Development, Infrastructure-Engineering sowie Training und Betrieb. Sprechen Sie mit uns über Ihre Anforderungen.

Wir nehmen uns Zeit für Sie. www.trivadis.com | info@trivadis.com

ZÜRICH ■ BASEL ■ BERN ■ BRUGG ■ LAUSANNE ■ DÜSSELDORF ■ FRANKFURT A.M.
FREIBURG I.BR. ■ HAMBURG ■ MÜNCHEN ■ STUTTGART ■ WIEN

trivadis
makes IT easier. ■ ■ ■