

# ORDIX<sup>®</sup> news

einfach. besser. informiert.



## Schöne neue vernetzte Welt

6 | Ein Blick auf das Internet der Dinge

14 | Interview zum Jubiläum mit Axel Röber, Benedikt Georgi und Christoph Lafeld

18 | Big Data - Informationen neu gelebt: Wie big ist Big Data?

26 | Java 8 - Die neue Version: Die Zeiten ändern sich

34 | Projektmanagement in der Praxis: Wer fragt, der führt!

Den Blick in die Zukunft gerichtet

# Wir wissen, wo der Weg hinführt!

25 Jahre zukunftsorientiertes Denken. Das bedeutet nicht nur stetige Qualifizierung sondern auch permanente persönliche Weiterentwicklung für unsere Mitarbeiter. Egal ob Sie als Einsteiger oder als Profi neue Perspektiven suchen - mit uns bringen Sie Ihre Karriere auf den richtigen Weg.

Den schnellsten Weg in unser Team finden Sie unter  
[www.ich-will-ins-ordix-team.de](http://www.ich-will-ins-ordix-team.de)



# 25.1

Paderborn, März 2015

Frauenquote, Mindestlohn (nicht der Lohn, sondern das drum herum macht es aus), Herdprämie, PKW-Maut und noch einige andere Forderungen, die so auf dem Tisch liegen, sind die neuesten Meisterstücke unserer Politiker. Selbst wenn uns als Firma heute nur wenige Punkte direkt betreffen (ich wäre froh, ich könnte die Gehälter meiner Mitarbeiter auf das Mindestlohn-Niveau absenken 😊), Regulierungswut und Bürokratiewahnsinn nehmen stetig zu.

Wenn ich 26 Jahre zurückblicke und an meine damaligen Vorstellungen und Gespräche vor der ORDIX Gründung denke, dann kann ich mit dem Wissen von heute nur sagen, dass es ORDIX in der jetzigen Form mit Sicherheit nicht gäbe. Die unglaubliche Vielfalt an Vorschriften, Gesetzen und ihren teilweise weder nachvollziehbaren noch logischen Auslegungen ist in Deutschland besonders groß und bei allem Respekt nicht immer förderlich.

Mich haben mittlerweile sieben Regierungen „begleitet“. Alle sind angetreten mit dem Ziel „weniger Bürokratie“ zu wagen bzw. Bürokratie abzubauen. Viele Phrasen der Politiker kann man so oder so einfach streichen. Es gibt aber kaum eine Aussage wie die von der Bürokratie, die kaum in den Mund genommen, genau das Gegenteil bedeutet.

Einmal in eine Suchmaschine<sup>1)</sup> eingegeben, können Sie jeden Politiker sofort überführen, da die Forderung zwar präsent ist, jedes neue umgesetzte Vorhaben aber zu neuen Vorschriften und zusätzlichem, nicht unerheblichem Aufwand in der Verwaltung führt. Dazu kommt mittlerweile auch noch der Wahn, alles und jedes Ding bis ins Detail zu überprüfen, ohne aber etwas Konkretes zu finden. Also wird eben nach Gutdünken ein Bericht geschrieben und schon haben Sie wieder eine Liste an Aufgaben vor sich, die mit ihrem normalen Geschäftsvorhaben nichts zu tun haben.

Bei Behörden, Ämtern und anderen Verwaltungen stehen Sie als Unternehmer ständig unter Generalverdacht bzw. der Anschuldigung, dass Sie jederzeit und bewusst gegen alle Vorschriften verstoßen. Bei jedem Mörder gilt die Unschuldsvermutung. Ein Unternehmer ist aber ohne Not ständig in der Pflicht nachzuweisen, dass er alles korrekt macht. Das ist nicht leicht, wenn man dann nach Jahren auch noch teilweise widersprüchliche Kommentierung von einigen Gesetzen zu lesen bekommt. Selbst hinzugezogene Rechtsanwälte verschaffen einem keine Rechtssicherheit.

Wenn mich also heute jemand um Rat fragen sollte, weil er/sie ein Unternehmen gründen will und Arbeitsplätze schaffen möchte, würde ich ihm/ihr eher davon abraten, dies in die Tat umzusetzen. Deutschland ist definitiv kein Gründerland und deshalb müssen wir uns nicht wundern, dass viele abwandern.

Gut, ich bin noch hier und werde auf meine alten Tage auch nicht mehr auswandern. Deshalb werden auch im ORDIX Jubiläumsjahr meine Mitarbeiter mit Artikeln in den ORDIX® news weiterhin für Ihre kostenlose Weiterbildung sorgen.

Mittlerweile schon mehrfach von mir erwähnt, sind zur Zeit die Softwarehersteller unterwegs, um die Lizenzierung der in Unternehmen eingesetzten Software zu prüfen. Unser Artikel zeigt Ihnen zumindest einige der Oracle-Lizenzierungsfallen.

Weiter geht es bei den Serien Oracle 12c und Java 8. Dem spannenden, hochaktuellen Thema Big Data widmen wir eine neue Reihe. In unserem Artikel zur Softwarearchitektur behandeln wir ebenfalls den einen oder anderen Aspekt des ständig größer werdenden Datenvolumens. Natürlich hat auch unser Data Warehouse Artikel mit Big Data zu tun.

Die Nichttechniker finden in dieser Ausgabe ebenfalls einige interessante Seiten: Im Artikel zum IT-Controlling beschäftigen wir uns mit einem Thema, welches für manchen technikverliebten Menschen eher der behindernde Teil des Projektgeschäftes ist. Sicherlich können aber auch Techniker aus unserem Artikel zur Interviewtechnik einiges gewinnbringend einsetzen.

In der Hoffnung, dass unser Interview mit meinen drei Vorstandskollegen aus Anlass des 25-jährigen Bestehens der ORDIX AG die Tipps des gerade erwähnten Artikels beherzigt, wünsche ich Ihnen für die nächsten Wochen und Monate bis zur nächsten ORDIX® news viel Erfolg und alles Gute



Wolfgang Kögler

<sup>1)</sup> Suchbegriff: „Regierungserklärung Abbau Bürokratie“





6

Internet der Dinge: Schöne neue vernetzte Welt

## Datenbanken

- 11..... **Datenbank-Performance im Data-Warehouse-Umfeld: Mit voller Kraft voraus**  
Die Geschwindigkeit der Datenbank entscheidet über die Leistung eines Data Warehouse. Viele Datenbankhersteller bieten hier eine Vielzahl von Funktionen um die Leistungsfähigkeit zu erhöhen. Wir zeigen Ihnen welche.
- 18..... **Neue Reihe: Big Data - Informationen neu gelebt: Wie big ist Big Data?**  
Viele der angebotenen Technologien versprechen eine unkomplizierte Verarbeitung Ihrer Datenmengen wie von Zauberhand. Einen ersten Einstieg in das Thema Big Data und welche Probleme beim Einsatz der einzelnen Technologien auftreten können, gibt dieser Artikel.
- 31..... **Neuerungen in der Oracle Database 12c (Teil VI): RMAN & Data Guard – Alles beim Alten?**  
Auch im Bereich RMAN und Data Guard gab es mit der neuen Version einige Erweiterungen. Sind diese sinnvoll? Wir sagen es Ihnen und zeigen Ihnen den Nutzen auf.
- 38..... **In-Memory-Option der Oracle Database 12c: Einfach und schnell mit dem In-Memory Column Store?**  
In-Memory ist eine neue, kostenpflichtige Option der Oracle Enterprise Edition 12c. Wir zeigen Ihnen wie einfach das Setup des neuen Speicherbereichs ist und welche Optionen es für den Aufbau gibt.
- 42..... **Oracle-Lizenzierung: Was ist neu unter VMware?**  
Ohne spezielles Lizenz-Know-how ist es nur schwer möglich, Oracle-Produkte „lizenzkonform“ einzusetzen. Dies gilt auch für den Einsatz unter VMware.



18

Wie big ist Big Data?

## Java / Java EE

- 26..... **Java 8 - Die neue Version (Teil IV): Die Zeiten ändern sich**  
Mit der neuen Java-Version hat auch eine neue Date/Time-API Einzug in den Standard gehalten. Welche Neuerungen diese API bringt und wie sich komplexe Datumsarithmetiken umsetzen lassen, wird in diesem Artikel beleuchtet.
- 36..... **Das Ende von „java.lang.OutOfMemoryError: PermGen space“: Auflösung der Permanent Generation**  
Wir stellen Ihnen die Auswirkungen des Wegfalls der Permanent Generation innerhalb der Java Hotspot JVM dar und zeigen, was zukünftig alles zu beachten ist.

## IT-Management

- 24..... **Strategisches IT-Controlling: Bestandteil einer nachhaltigen Unternehmensausrichtung**  
Das IT-Controlling wird zukünftig im Bereich der Unternehmensausrichtung eine immer größere Rolle spielen. Welche Arten des Controllings gibt es eigentlich? Wie sieht das Leitbild des Controllers aus?

## Projektmanagement

- 34..... **Projektmanagement in der Praxis: Wer fragt, der führt!**  
Für die Gesprächsform des Interviews gibt es im Bereich Projektmanagement sehr viele Einsatzmöglichkeiten. Wir geben Ihnen zahlreiche Tipps aus der Praxis.



26

Java 8: Die Zeiten ändern sich

## Softwarearchitektur

### 6 ..... Ein Blick auf das Internet der Dinge: Schöne neue vernetzte Welt

Wer kennt sie nicht, die Smartwatches, die den Puls messen oder die Elektrogeräte die über Smartphones ferngesteuert werden können? Aber was passiert mit den ganzen Daten, die hier im World Wide Web unterwegs sind?

## Aktuell

### 14 ..... Interview zum Jubiläum mit Axel Röber, Benedikt Georgi und Christoph Lafeld: „25 Jahre einfach. gut. beraten.“

### 17 ..... Larry Ratlos Rätsel

### 22 ..... Seminarübersicht: April bis Juli 2015

### 37 ..... ORDIX spendet - Hilfe statt Geschenke



38

In-Memory-Option der Oracle Database 12c

## Impressum

<b>Herausgeber:</b>	ORDIX AG Aktiengesellschaft für Softwareentwicklung, Beratung, Schulung und Systemintegration, Paderborn
<b>Redaktion/Layout:</b>	Jens Pothmann, Evelyn Kanne
<b>V.i.S.d.P.:</b>	Benedikt Georgi, Wolfgang Kögler
<b>Anschrift der Redaktion:</b>	ORDIX AG   Westermauer 12 - 16   33098 Paderborn Tel.: 05251 1063-0   Fax: 0180 1673490
<b>Auflage:</b>	7.000 Exemplare
<b>Druck:</b>	Druckerei Bösmann, Detmold
<b>Bildnachweis:</b>	© vectoropenstock.com   Computer functions circle icons © freepik.com   Technology background with planet © unplash.com   Juskteez   Stars © flickr   Nicholas A. Tonelli   Long Drive © fotolia.de   mindscanner   Word Cloud „Big Data“ © sxc.hu   pipp   notebook © freepik.com   Infographic chart © freebiesgallery.com   Asif Aleem   Vector Infographics © flickr   Jannis_V   Pocket Watch © flickr   MIKI Yoshihito   Display Casio © istockphoto.com   tiero   file cabinet and folder © psdgraphics.com   computer database psd icon © istockphoto.com   stokkete   Creative entrepreneur ... ideas © sxc.hu   coscurro   left or right © flickr   Esporte Clube Pelotas   Football © unplash.com   Kamil Lehmann   Road © shop.tipp-kick.de   Figur © istockphoto.com   Two Thumbs Up   sdominick
<b>Autoren:</b>	Andreas Flügge, Benedikt Georgi, Olaf Hein, Wolfgang Kögler, Christoph Lafeld, Sven Loer, Oliver Luser, Klaus Reimers, Thomas Rohde, Axel Röber, Michael Skowasch, Michael Spoden, Christian Wiesing
<b>Copyright:</b>	Die ORDIX® news erscheint 3 x jährlich. Alle Eigentums- und Nachdruckrechte, auch die der Übersetzung, der Vervielfältigung der Artikel oder von Teilen daraus, sind nur mit schriftlicher Zustimmung der ORDIX AG gestattet.
<b>Warenzeichen:</b>	Einige der aufgeführten Bezeichnungen sind eingetragene Warenzeichen ihrer jeweiligen Inhaber. ORDIX® ist eine registrierte Marke der ORDIX AG.
<b>Haftung:</b>	Eine Haftung für die Richtigkeit der Veröffentlichungen kann trotz sorgfältiger Prüfung durch die Redaktion vom Herausgeber nicht übernommen werden.

Sie können die Zusendung der ORDIX® news jederzeit ohne Angabe von Gründen schriftlich (z.B. Brief, Fax, E-Mail) abbestellen.



Ein Blick auf das Internet der Dinge

## Schöne neue vernetzte Welt

Das Internet der Dinge („Internet of Things“ – kurz IoT) ist zur Zeit in aller Munde und man stellt sich wie so oft die Frage, ob es sich dabei nur um eine weitere vorübergehende Modeerscheinung handelt oder ob wir uns damit in Zukunft ernsthaft auseinandersetzen müssen. Laut Gartner befinden wir uns gerade auf dem Höhepunkt des Hypes. Bei genauerem Hinsehen entpuppt sich das IoT als eine Folgeerscheinung, die durch die konsequente Weiterentwicklung bestimmter Technologien ermöglicht wird. Der folgende Artikel soll dies ein wenig genauer beleuchten und erklären, welche Ideen dem IoT zugrunde liegen und welche Folgen wir zu erwarten haben.

### Begriff

Der Begriff Internet of Things ist gar nicht so neu. Er tauchte bereits in den 90er Jahren im Zusammenhang mit Überlegungen zum Aufbau von RFID-Netzwerken auf. Schon damals stand die Idee im Mittelpunkt, Gegenstände elektronisch zu erfassen und miteinander zu vernetzen.

Im heutigen Sprachgebrauch verbindet man damit konkret die Vernetzung der alltäglichen Dinge unseres Lebens, insbesondere deren Integration in die bereits bestehenden Infrastrukturen, allen voran das Internet. Dies ist nur die konsequente Umsetzung und Weiterführung des Vorgehens, wie wir es in jüngster Vergangenheit bei Smartphones und Tablets beobachten konnten.

Das Szenario des „Ubiquitous Computing“, der allgegenwärtigen und computergestützten Informationsverarbeitung in unserem alltäglichen Leben, wird damit immer mehr Realität.

Die Zukunft des IoT sieht aber weit über das bloße Vernetzen von Dingen hinausgehende Fähigkeiten vor (siehe Abbildung 1). So sollen die Daten nicht nur ausgetauscht, sondern auch analysiert und mit einer Semantik versehen werden, die dann ihrerseits zur konkreten Entscheidungsfindung herangezogen werden kann und automatisierte Aktionen auslösen soll.

## Technologie

Ermöglicht wird das IoT im Wesentlichen durch technologische Fortschritte in den Bereichen, die auch bisher als die üblichen Verdächtigen galten. Die Miniaturisierung in Verbindung mit der Entwicklung von energiesparender Hardware und entsprechenden Funktechnologien wie z.B. Bluetooth Low Energy sind hier der entscheidende Treiber.

Hinzu kommen effiziente Produktionsprozesse, die sehr leistungsfähige Hardware extrem günstig herstellbar machen. So sind neben den eigentlichen rechnenden Einheiten wie z.B. Microcontroller auch relativ komplizierte Sensoren wie z.B. für GPS, Luftdruck, Neigung oder Beschleunigung sehr kostengünstig verfügbar (Beispiel siehe Abbildung 2, Preis ca. 4 Euro). Gerade dieser letzte Aspekt trägt dazu bei, dass sich in letzter Zeit eine florierende Crowdfunding- und Maker-Szene entwickelt hat, die mit ihren kreativen Ideen diese Entwicklung noch weiter beschleunigt.

## Anwendungsszenarien

Der jüngste Trend der Smartwatches und Fitness-Tracker ist vielleicht eins der bekannteren Beispiele für eine Ausprägung des IoT. Auf einer mit dem Smartphone vernetzten Armbanduhr werden Informationen zu E-Mails und sonstigen Ereignissen angezeigt. Fitnessarmbänder messen die Herzfrequenz und überwachen die Schlafphasen seines Trägers, um die Daten zu liefern, die andere Geräte zur Analyse und Visualisierung des Fitnesszustandes benötigen (siehe Abbildung 3).

Aber auch im eigenen Heim nimmt des IoT bereits Formen an. Elektrogeräte weit über die Entertainment-Komponenten hinaus, wie z.B. Kühlschränke und Waschmaschinen bekommen Netzwerkschnittstellen. Heizung, Licht und Rollläden lassen sich fernsteuern und eine Einbruchüberwachung mit Notruf ist ebenfalls keine Zukunftsvision mehr. Ähnliches gilt für die intelligenten Brandmelder, die untereinander vernetzt und mit einer Brandmeldezentrale verbunden sind. Welche Bedeutung diesem Thema zugemessen wird, kann man unter anderem daran erkennen, dass Google vor kurzem für über drei Milliarden Dollar Nest Labs, eine auf Thermostate und Brandmelder spezialisierte Firma, übernommen hat.

Auch des Deutschen liebstes Kind, das Automobil ist längst im Fokus des IoT. Die Integration von Informations- und Entertainmentsystemen schreitet immer weiter voran. Der Sensorik kommt im Automobil prinzipiell eine hohe Bedeutung zu, schließlich müssen Fahrerassistenzsysteme überwacht und gesteuert werden. Der Schritt zum autonom fahrenden Fahrzeug ist nicht mehr weit. Navigationssysteme liefern ebenfalls eine Vielzahl von Daten. Diese können neben der Wegbestimmung im Falle eines Unfalls z.B. zusammen mit den Airbag-Daten dazu genutzt werden, einen Notruf auszulösen und einen Rettungshubschrauber zum Unfallort zu leiten.



Abb. 1: Internet of Things (Quelle: Gartner, August 2014)

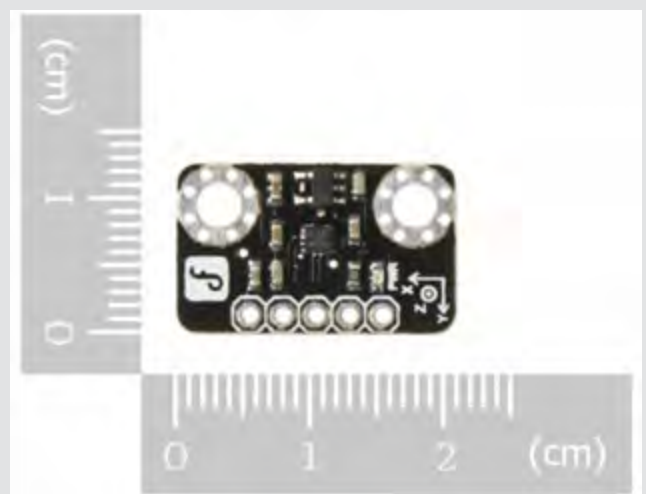


Abb. 2: 3-Achsen Beschleunigungssensor Bosch BMA220



Abb. 3: Samsung Gear Fit (Fitnessstracker mit Pulsmesser, Schrittzähler, Schlaf/Stress-Monitor und Bluetooth Low Energy-Schnittstelle) Quelle: [http://en.wikipedia.org/wiki/Samsung\\_Gear\\_Fit](http://en.wikipedia.org/wiki/Samsung_Gear_Fit)

## Potenzial

Aus den vorgestellten Anwendungsszenarien wird das enorme Potenzial des IoT deutlich. Es ist damit zu rechnen, dass durch die zunehmende Anzahl auch kleinsternetzter Geräte das weltweit anfallende Datenvolumen noch einmal deutlich steigen wird. Juniper Research rechnet einer Studie nach bis 2020 mit weltweit 50 Milliarden vernetzten Geräten jeglicher Art. Das stellt die IT, welche die anfallenden Daten speichern und verarbeiten soll vor große Herausforderungen und wird enorme Anstrengungen für die Weiterentwicklung z.B. in den Bereichen Mobile- und Cloud-Computing benötigen. Das haben auch viele Firmen erkannt und versuchen sich zurzeit entsprechend zu positionieren.

Einer der wichtigsten Aspekte dabei wird die Gewinnung und Kontrolle der anfallenden Daten sein. Denn nur der, der Zugriff auf die Daten besitzt, kann sie analysieren und daraus die richtigen Schlüsse ziehen. Hier wird es im IoT nicht anders sein, als heute bereits in den sozialen Netzwerken. Vereinfacht gesagt ist das Unternehmen im Vorteil, welches die Bestellung für die Nachlieferung bekommt, wenn der intelligente Kühlschrank meldet, dass die Milch aus ist.

## Risiken

An dieser Stelle werden auch die Risiken deutlich. Was passiert, wenn die Alarmanlage meines Hauses durch einen Hackerangriff von außen lahmgelegt wird? Bekomme ich noch eine Lebensversicherung, wenn die Daten meines Fitnesstrackers in die Hände der Versicherung gelangen und ein gesundheitliches Problem offenbaren? Möchte ich durch den für die Zukunft angedachten Fahrtschreiber (Blackbox für Unfallauswertung) in meinem Auto prinzipiell verfolgt oder ortbar sein?

Diese Fragen zeigen, dass wir es beim IoT nicht nur mit technischen Problemen zu tun haben werden. Das IoT wird mehr als alles bisher Dagewesene, tief in unser privates Leben eindringen und es beeinflussen. Wir werden nicht um die Klärung der damit verbundenen sozialen und rechtlichen Fragestellungen herumkommen. Dem Datenschutz wird zukünftig eine noch wichtigere Rolle als bislang zukommen müssen.

## Anforderungen und Randbedingungen

In der Kenntnis, dass die Vision des IoT natürlich weitreichendere Ziele als nur die Vernetzung hat, wollen wir uns aber zunächst auf diese beschränken, da die Vernetzung quasi den Grundstein für die Funktionsfähigkeit legt. Aus der technischen Sicht betrachtet ist dementsprechend zu klären, auf welcher Basis ein IoT aufgebaut sein könnte. Dazu stellt ein Softwarearchitekt in der Regel die Frage nach den Anforderungen (funktionale und nicht-funktionale), sowie nach den Randbedingungen und sonstigen Einflussfaktoren, um daraus eine passende Architektur zu entwickeln. Schauen wir uns im Folgenden also einige Anforderungen an:

Zunächst ist klar, dass wir es in der Problemdomäne mit verteilten Systemen (distributed systems) zu tun haben. Zur Verdeutlichung: Wir wollen eine große Anzahl an sehr unterschiedlichen (heterogenen) Systemen miteinander Daten austauschen lassen. Dabei handelt es sich fast ausschließlich um M2M-Kommunikation (machine to machine), also um ein klassisches Integrationszenario.

Bezüglich der zu vernetzenden Systeme oder Geräte können folgende Annahmen gemacht werden:

- Ein Gerät kann sowohl Datenproduzent als auch Datenkonsument sein.
- Die Geräte sind u.U. extrem beschränkt bzgl. ihrer Ressourcen (Rechenleistung, Speicher, ...), z.B. Sensoren.
- Die Netze, in denen sich die Geräte befinden, haben u.U. sehr beschränkte Bandbreiten (seriell, ISDN, mobile Verbindungen, ...).
- Aufgrund der begrenzten Stabilität der Netze sind die Geräte u.U. nicht ständig erreichbar (z.B. mobile Verbindungen).
- Das Netz soll skalierbar sein, d.h. zusätzliche Geräte sollen einfach hinzuzufügen sein.

## Architektur

Natürlich sind die Probleme aus dem Bereich der verteilten Systeme nicht grundsätzlich neu. Es gibt viele Parallelen zu Systemen, die in der Vergangenheit mit erprobten Lösungsmustern umgesetzt wurden. Man spricht übergreifend von nachrichtenbasierten Systemen (message oriented systems), denen gemein ist, dass sie die Kommunikation zwischen den Systemen auf den Austausch abstrakter Nachrichtenobjekte herunterbrechen.

Zwei Aspekte spielen dabei eine zentrale Rolle. Das ist zum einen der interne Status der einzelnen Systeme, über den andere Systeme Informationen benötigen. Zum anderen sind die Kommunikationsmuster von entscheidender Bedeutung, da sie die Topologie des Netzes mitbestimmen und Einfluss auf die konkreten Kommunikationsmechanismen und -aufwände haben.

Zur Lösung der Statusproblematik wird oft ein „Publish-Subscribe“-Muster angewandt. Dies ist mit einem Abonnement-Modell vergleichbar. Ein System A, das Statusinformationen von einem anderen System B benötigt, abonniert einfach dessen Status. System A registriert (subscribe) ein Interesse an den Statusänderungen von System B. Ändert sich in System B jetzt tatsächlich der Status, veröffentlicht (publish) System B diesen und alle Abonnenten (Subscriber) werden darüber benachrichtigt.

Ein Architekturpattern, das den zweiten Aspekt der Kommunikationsmuster adressiert und sehr häufig eingesetzt wird, ist der Broker (Vermittler). Er macht die Kommunikation für die beteiligten Systeme transparent und abstrahiert von den konkreten Implementierungen des Nachrichtenaustausches. Er sorgt dafür, dass nicht jedes System alle anderen kennen muss. Dadurch verhindert es ausufernde Kommunikationsaufwände und reduziert die Abhängig-



keiten zwischen den Systemen, was bei Punkt-zu-Punkt-Verbindungen häufig zu Problemen führt.

Der Verzicht auf eine synchrone Request-Response-Kommunikation führt in diesem Fall zu einer ereignisorientierten, asynchronen Kommunikation. Die einzelnen beteiligten Systeme kennen sich nicht direkt und wenden sich immer an den Broker (z.B. für die „Publish-Subscribe“-Aktionen, siehe Abbildung 4), während der Broker den Nachrichtenaustausch zwischen den Systemen vermittelt. Der Broker ist in Kenntnis aller beteiligten Systeme und bildet die zentrale Vermittlungsstelle, was ihn gleichzeitig zu einem potenziellen Problem (single point of failure) macht. Fällt er aus, bricht das Gesamtsystem zusammen. Um dies zu verhindern, werden Broker meist so implementiert, dass sie horizontal skalierbar werden. Dies bedeutet, dass sich in der Regel mehrere Instanzen eines Brokers in einem Gesamtsystem befinden, die sich gegenseitig synchronisieren. So ist gewährleistet, dass beim Ausfall eines Brokers eine andere Instanz dessen Aufgaben übernehmen kann.

### Beispiel MQTT

Das IoT kränkelt zurzeit noch an vielen verschiedenen, untereinander inkompatiblen Standards. Hier existiert geradezu ein Zoo von Protokollen, die unterschiedliche Aspekte adressieren. Mit MQTT (Message Queue Telemetry Transport) existiert allerdings ein Protokoll, das bereits 2013 bei der OASIS zur Standardisierung eingereicht wurde und dessen Verabschiedung noch für 2014 vorgesehen war. MQTT wurde bereits Ende der 90er Jahre von den Firmen IBM und Eurotech im Zusammenhang mit einem Projekt zur Überwachung einer Öl-Pipeline entwickelt. Es wurde unter einer freien Lizenz veröffentlicht und Eurotech hat seine API-Implementierung der Eclipse Foundation zur Verfügung gestellt, woraus das Eclipse-Paho-Projekt [1] entstanden ist. Auch für den dazugehörigen Broker existieren bereits Implementierungen - wie z.B. HiveMQ [2].

MQTT ist ein reines Transport-Protokoll und besitzt keinerlei Semantik, welche die speziellen Anwendungen im IoT unterstützt. Mit seinen spezifischen Mechanismen ist es aber extrem gut für die Bedingungen des IoT aufgestellt. Es implementiert die o.g. Muster und zeichnet sich weiter u.a. durch folgende Merkmale aus:

### Quality of Service (QoS)

MQTT setzt auf dem TCP-Protokoll auf und bietet von daher bereits recht zuverlässige Verbindungen. Das ist jedoch z.B. für naturgemäß instabile Mobilfunkverbindungen nicht ausreichend. Deshalb werden zusätzliche Garantien für die Nachrichtenzustellung definiert. So können die Teilnehmer festlegen, ob versendete Nachrichten höchstens einmal, genau einmal oder mindestens einmal zugestellt werden. Für den Fall von Verbindungsausfällen sorgt das Protokoll mit inhärenten Persistenzmechanismen dafür, dass keine Nachrichten verloren gehen können.

### Last will and testament / Retained message

Zusätzlich zu den QoS-Leveln können weitere Maßnahmen getroffen werden, die sich auf unerwartete Verbindungsabbrüche beziehen. So kann ein Client bei der Verbindung mit einem Broker eine Nachricht beim Broker hinterlegen, welche dieser versendet, sobald er bemerkt, dass die Verbindung zum Client abgebrochen ist (Last will and testament). Des Weiteren kann ein Datenproduzent (Publisher) beim Broker eine Nachricht hinterlegen, die ein neuer Datenkonsument (Subscriber) erhält, sobald beim Verbinden des Subscriber die Verbindung zum Publisher abgebrochen ist (Retained message).

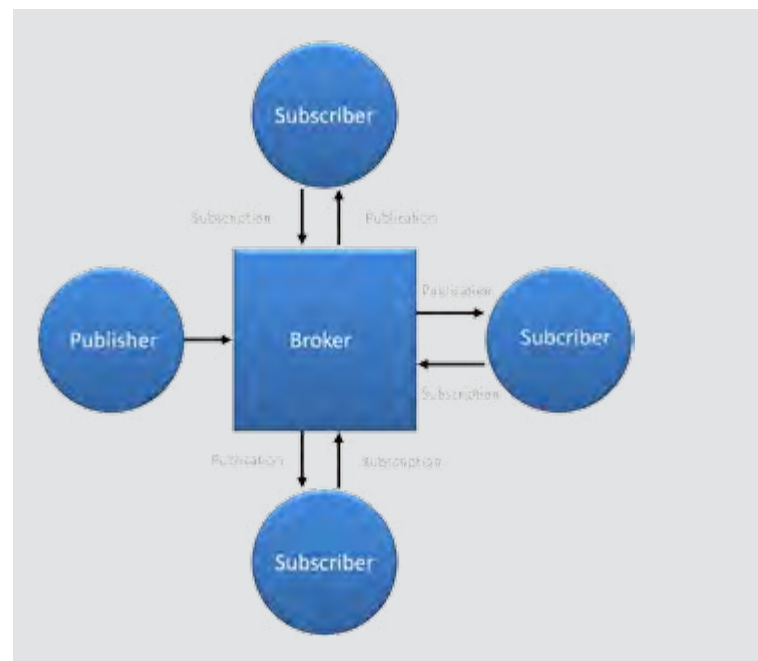


Abb. 4: Publish-Subscribe-Muster

### Links

[1] Webseite des Eclipse-Projektes „Paho“:  
<http://www.eclipse.org/paho/>

[2] Webseite von HiveMQ:  
<http://www.hivemq.com/>

[3] Webseite der Object Systems GmbH  
<http://www.objectsystems.de/>

### Quellen/Bildnachweis

[1] Gartner:  
<http://www.gartner.com/newsroom/id/2684616>

[2] Juniper Research  
<http://www.juniperresearch.com/>

© freepik.com | Technology background with planet  
© vectoropenstock.com | Computer functions circle icons

## Glossar

### Crowdfunding

Die Schwarmfinanzierung ist eine Art der Finanzierung. Mit dieser Methode der Geldbeschaffung lassen sich Projekte, Produkte, die Umsetzung von Geschäftsideen und vieles andere mit Eigenkapital, zumeist in Form von stillen Beteiligungen, versorgen. Eine so finanzierte Unternehmung und ihr Ablauf werden auch als eine Aktion bezeichnet. Ihre Kapitalgeber sind eine Vielzahl von Personen – in aller Regel bestehend aus Internetnutzern, da zum Crowdfunding meist im World Wide Web aufgerufen wird. (Quelle: Wikipedia)

### Maker

Maker ist die Bezeichnung für eine Subkultur, die man auch als Do-It-Yourself-Kultur mit dem Einsatz aktueller Technik beschreiben kann. Ziel vieler Anhänger ist es, mit eigenen Mitteln ein technisches Problem zu lösen, ohne den Einsatz kostspieliger Speziallösungen. Ein typisches Beispiel ist der Einsatz von 3D-Druckern, mit denen Ersatzteile für Geräte hergestellt werden. Für viele Mitglieder ist auch Wissensaneignung und Neugierde die Triebfeder.

### Microcontroller

Microcontroller sind Halbleiterchips, die im Gegensatz zu klassischen Prozessoren, neben der eigentlichen Recheneinheit auch komplexere Peripheriefunktionen wie z.B. Netzwerk, Digital-Analog-Wandler oder Signalverarbeitungsfunktionen auf einem einzelnen Chip beherbergen. Oft spricht man auch von einem „System on a Chip“ (SoC).

### MQTT

Message Queue Telemetry Transport (MQTT) ist ein offenes Nachrichtenprotokoll für Machine-to-Machine-Kommunikation (M2M), das die Übertragung von Telemetriedaten in Form von Nachrichten zwischen Geräten ermöglicht, trotz hoher Verzögerungen oder beschränkter Netzwerken.

### RFID

Die Radio Frequency Identification ist ein Verfahren zur Identifizierung von Gegenständen, die mit einem speziellen Etikett (Transponder) versehen sind. Mittels elektromagnetischer Wellen kann ein Sender berührungslos Informationen aus dem Transponder auslesen.

### TCP

Das Transmission Control Protocol ist ein Netzwerkprotokoll, das definiert, auf welche Art und Weise Daten zwischen Computern ausgetauscht werden sollen. Nahezu sämtliche aktuellen Betriebssysteme moderner Computer beherrschen TCP und nutzen es für den Datenaustausch mit anderen Rechnern.

abonniert werden. Für MQTT existieren APIs für die wichtigsten Sprachen wie C/C++, Java oder JavaScript. Das Protokoll wird bereits produktiv genutzt und skaliert auch bei mehreren tausend Geräten noch ausreichend gut. Ein weiterer Vorteil sind die geringen Ressourcenanforderungen, die MQTT an die jeweiligen Clients stellt.

## Fazit

Das IoT wird eine nochmalige Steigerung des Datenaufkommens bedeuten und für begleitende Paradigmen wie Mobile- und Cloud-Computing einen weiteren Schub bewirken. Technisch wird sich wahrscheinlich ein Trend verstärken, der schon seit einiger Zeit zu beobachten ist. Hier spielen leichtgewichtige Systeme, die sehr gut horizontal skalieren eine große Rolle. Dies bedeutet, dass wir uns zunehmend mit sehr reaktiven, asynchronen Programmiermodellen beschäftigen werden müssen. Der Hype um das „Internet of Things“ ist auf seinem Höhepunkt. Das bedeutet aber auch, dass es noch einige Zeit dauern wird, bis mit produktionsreifen standardisierten Produkten zu rechnen ist. Aber auch zum jetzigen Zeitpunkt existieren bereits Ansätze, mit denen Teilaspekte des IoT in ausreichender Qualität realisiert werden können.

Das IoT wird nicht aufzuhalten sein. Die IT wird noch weiter in alle Lebensbereiche vordringen, was nicht nur technische, sondern auch soziale und rechtliche Probleme verursachen wird. Spannend bleibt die Frage, wie wir damit umgehen werden.

Insbesondere die Frage nach der Nutzung und Verwertung der anfallenden Daten könnte uns dazu zwingen, unseren Umgang damit grundsätzlich zu überdenken. Es wird sich zeigen, ob evtl. Ideen wie die des amerikanischen Informatikers und diesjährigen Trägers des Friedenspreises des Deutschen Buchhandels Jaron Lanier hoffähig werden. Er plädiert dafür, Daten zu einem Handelsgut zu deklarieren, wonach sich private Nutzer für die Zurverfügungstellung ihrer Daten entlohnen lassen sollen. Dass sich diesbezüglich etwas bewegen muss, scheint unstrittig. Ob die Zeit bereits reif für alternative Ansätze wie die Laniers ist, wird sich allerdings erst noch zeigen müssen.

## Topic-Adressierung

Das Senden (publish) und Empfangen (subscribe) von Daten wird über sogenannte Topics geregelt. So könnte der Temperatursensor im Wohnzimmer eines Hauses seine Daten über das Topic /Haus/Wohnzimmer/Temperatur bereitstellen. Für die Küche wäre dies analog das Topic /Haus/Kueche/Temperatur. Wenn wir davon ausgehen, dass jeder Raum auch noch Sensoren für die Luftfeuchtigkeit und den Zustand des Lichtschalters bereitstellt, können alle Daten des Schlafzimmers über das Topic /Haus/Schlafzimmer/# abonniert werden. Neben dem Platzhalter „#“, der nur am Ende eines Topic stehen darf, existiert noch ein Platzhalter „+“. Mittels /Haus+/Licht können alle Lichtschaltersensoren des gesamten Hauses



Andreas Flügge  
(info@objectsystems.de)

Datenbank-Performance im Data-Warehouse-Umfeld

# Mit voller Kraft voraus

Die Leistung eines performanten Data Warehouse steht und fällt mit der Geschwindigkeit der darunterliegenden Datenbank. Datenbankhersteller bieten mittlerweile eine Vielzahl von Funktionalitäten zur Leistungsverbesserung. Welche Möglichkeiten die Datenbanken von Haus aus mitbringen, stellen wir in diesem Artikel vor.

## Einleitung

Die zugrundeliegende Datenbank ist die zentrale und wichtigste Komponente eines Data Warehouse (DWH). Für eine gute Antwortzeit bei den Reports und Analysen ist die Nutzung von bestimmten Datenbankfunktionen notwendig oder sogar unabdingbar.

## Ladetools

Neue Daten werden üblicherweise mit einer **Insert**-Anweisung in die Datenbanktabellen eingefügt. In einem Data Warehouse werden jedoch Massendaten eingefügt, beispielsweise die Buchungsdaten des letzten Tages. Die **Insert**-Anweisung hat hier den Nachteil, dass jede Transaktion protokolliert wird und in einem Logfile abgelegt wird. In einer Data-Warehouse-Umgebung werden aus diesem Grund so genannte Ladetools verwendet. Diese verfügen nicht über die üblichen Transaktionsprotokolle, haben allerdings den Nachteil, dass z.B. Trigger auf den Zieltabellen nicht verwendet werden dürfen und referentielle Integritäten nicht geprüft werden. Das sind jedoch alles Funktionalitäten, welche in der Regel in einem Data Warehouse nichts zu suchen haben. Vorteil ist die enorme Geschwindigkeit des Ladevorgangs, welcher um ein Vielfaches (bis zu einem Faktor 50 und mehr) schneller als der klassische **Insert** sein kann.

## Partitionierung

Typische Auswertungen im Data Warehouse vergleichen diverse Kennzahlen über verschiedene Zeiträume. In der Regel sind diese Reports noch eingeschränkt auf diverse weitere Dimensionen. Selten finden Analysen über den kompletten Datenbestand statt. Um nun beim Ermitteln der Daten nicht die komplette Faktentabelle per Index- oder Tablescan auslesen zu müssen, kann die Faktentabelle partitioniert werden. Meist bietet sich hierzu eine Partitionierung bzgl. der Zeit auf Monatsbasis an, aber auch andere Partitionierungskriterien sind durchaus denkbar. Durch die Einschränkung der Daten in der **where**-Bedingung auf diese Partitionierungsschlüssel kann der Datenbank-Optimizer schon beim Erstellen des Zugriffsplans viele

Partitionen vom Lesezugriff ausschließen und somit I/O-Zeit sparen. Weitere Partitionierungstechniken wie z.B. die Hash-Partitionierung sollten ebenfalls geprüft werden und können mit der klassischen Range-Partitionierung kombiniert werden.

## Komprimierung

In vielen Warehouse-Systemen stellt der I/O auf Platte den Flaschenhals dar. Durch historische Analysen müssen Massendaten analysiert werden, wo oftmals die Datenbank ohne einen Full Tablescan nicht auskommt. Das bedeutet, Millionen von Datensätzen müssen aufwändig von der Platte gelesen werden. Hier kann eine Komprimierung der Daten helfen. Viele Datenbankmanagementsysteme sind in der Lage, die Daten um bis zu 50 - 80% zu komprimieren. Das hat zur Folge, dass sich auch die teuren Lesevorgänge um einen ähnlichen Faktor verringern, da das Datenvolumen wesentlich kleiner ist und somit weniger zur Performance-Bremse wird.

## Datenbank-Caches

Durch das Cachen von Daten kann der I/O auf die Plattensysteme enorm reduziert werden. Im Data-Warehouse-Umfeld besteht jedoch durch das enorme Datenvolumen immer die Gefahr, dass wichtige Daten wieder aus dem Cache durch neue überschrieben werden. Insbesondere die Faktentabelle, auf welche praktisch bei fast jedem Report zugegriffen wird, fördert durch ihre Größe dieses Verhalten. Durch geschicktes Aufteilen der Tabellen auf mehrere Caches kann man dies jedoch reduzieren. In der Regel empfiehlt es sich für kleine zentrale Tabellen einen separaten Cache zu nutzen. Meist sind das die Dimensionstabellen, für welche in Oracle ein so genannter Keypcache bzw. in DB2 ein separater Bufferpool angelegt wird. Dadurch ist es möglich, für diese Tabellen eine Trefferquote des Cache von nahezu 100% zu erreichen und das gesamte I/O des Systems zu reduzieren.

## Glossar

### Data-Mart

Ein Data-Mart ist eine Kopie des Teildatenbestandes eines Data Warehouse, die für einen bestimmten Organisationsbereich oder eine bestimmte Anwendung/Analyse erstellt wird. Es kann auch als Teilansicht auf das Data Warehouse oder nicht-persistenten Zwischenspeicher verstanden werden.

### DBMS

Ein Datenbankmanagementsystem wird für die Informationsverwaltung eingesetzt. Ein solches System besteht aus Daten und Programmen und dient der Dateiverwaltung und -verarbeitung. Es kann als eine Art Magazin für alle Daten verstanden werden und ist verantwortlich für die Speicherung, Sicherheit, Integrität, Wiederherstellung und den Zugriff auf die Daten.

### DWH

Ein Data Warehouse ist eine Datenbank, in der Daten aus unterschiedlichen Quellen in einem einheitlichen Format zusammengefasst werden (Informationsintegration). Dadurch verbessert sich der Komfort beim Zugang zu diesen Daten.

### Faktentabelle

In der Regel ist die Faktentabelle die größte Tabelle eines Data Warehouse, in welcher nach dem multidimensionalen Datenmodell die Kennzahlen des Warehouse historisiert abgelegt werden.

### Hash-Partitionierung

Die Daten einer Tabelle werden anhand eines internen Hash-Algorithmus auf unterschiedliche Partitionen verteilt. Beim Selektieren können die Daten parallel auf den unterschiedlichen Partitionen gelesen werden.

### I/O

Mit Ein-/Ausgabe (englisch Input/Output, kurz I/O) bezeichnet man die Kommunikation/Interaktion eines Informationssystems mit seiner Außenwelt, z.B. seinen Benutzern oder anderen Informationssystemen und Aktionen zum Lesen und Schreiben von Daten. Eingaben sind Signale (z.B. Kommandozeilen) oder Daten, die das System empfängt, Ausgaben werden vom System gesendet.

### OLTP

Das Online-Transaction-Processing bezeichnet ein Benutzungsparadigma von Datenbanksystemen und Geschäftsanwendungen, bei dem die Verarbeitung von Transaktionen direkt und prompt, also ohne nennenswerte Zeitverzögerung, stattfindet.

### Range-Partitionierung

Die Daten einer Tabelle werden anhand bestimmter Wertebereiche (z.B. dem Monat) auf unterschiedliche Partitionen verteilt. Beim Selektieren der Daten mit einer entsprechenden Einschränkung muss somit nicht die komplette Tabelle gelesen werden.

### Referentielle Integrität

Unter der referentiellen Integrität (RI) versteht man die Bedingungen, die zur Sicherung der Datenintegrität bei Nutzung relationaler Datenbanken beitragen können. Nach der RI-Regel dürfen Datensätze (über ihre Fremdschlüssel) nur auf existierende Datensätze verweisen.

werden, empfiehlt es sich eine größere Block-/Seitengröße zu verwenden, da es performanter ist wenige große Blöcke bzw. Seiten zu lesen als viele kleine.

## Freiplatz in Tabellen

In einer OLTP-Umgebung ist das Ändern von Daten durch **Update**-Anweisungen nichts Ungewöhnliches. Damit die Wahrscheinlichkeit steigt, dass es beim Ändern von z.B. **varchar**-Feldern von kurzer String-Länge in eine längere String-Länge keine Performance-Einbußen gibt, wird in vielen DBMS ein gewisser Prozentsatz innerhalb eines Blockes freigehalten. Für OLTP-Systeme ist dies in der Regel auch eine gute und leistungssteigernde Funktionalität. In einem Data Warehouse allerdings werden – insbesondere auf die Faktentabelle – keine Änderungen vorgenommen. D.h. hier würde zum einen Platz verschwendet und andererseits unnötiger I/O durch das Lesen des freien Platzes innerhalb eines Blocks erzeugt. In der Datenbank eines Data Warehouse sollte daher dieser Wert für die Tabellen geprüft werden. In den meisten relationalen Datenbankmanagementsystemen nennt sich diese Option **PCTFREE**, welche beim Anlegen einer Tabelle auf 0 gesetzt werden kann.

## Materialised Views

Reports aggregieren die Kennzahlen in einem Data Warehouse oftmals nach verschiedenen Kriterien wie z.B. bestimmten Dimensionshierarchien. Für die Datenbank bedeutet dies, dass jedes Mal auf die detaillierten Basisdaten eine Summe, ein Durchschnitt, ein Maximalwert oder sonstige Aggregationen gebildet werden müssen. Was liegt also näher, als sich für häufig verwendete Aggregationen die Zwischenergebnisse in so genannten Materialised Views vorberechnen zu lassen? Wie der Name vermuten lässt, ist dies ähnlich einer herkömmlichen View allerdings werden die Daten nicht zur Laufzeit ermittelt, sondern schon vorab durch eine **Refresh**-Anweisung jedes Mal, nachdem die Datenversorgung abgeschlossen ist, physikalisch als Form einer Tabelle abgelegt.

Beim Ausführen der Queries ergeben sich dadurch enorme Performance-Vorteile, da die Daten nicht mehr berechnet werden müssen. Gegenüber der Reportingsoftware ist dies so transparent, dass weiterhin auf die darunterliegenden Basistabellen zugegriffen werden kann und die Datenbank automatisch erkennt, dass beim Selektieren der Daten die Materialised View verwendet werden kann. Materialised Views finden u.a. häufig als Data-Marts Verwendung.

## Block- und Seitengrößen

Die kleinste Einheit, in welcher Datenbankmanagementsysteme Daten lesen bzw. schreiben können, sind – je nach Datenbank – ein Block, eine Seite bzw. ein Extent. Da im DWH-Umfeld selten Einzelsätze gelesen und geschrieben werden, sondern im Allgemeinen Massendaten verarbeitet

## Parallelisierung

Die Datenbanken für ein Data Warehouse sind üblicherweise mit viel Hauptspeicher und CPU-Kernen ausgestattet. Durch die meist überschaubare Anzahl von Benutzern

und die Art der Queries, welche durch `Join`-Bedingungen verknüpft auf die Faktentabelle und diverse Dimensionstabellen zugreifen, bietet es sich an, die Parallelisierung für SQL-Anweisungen einzuschalten. Der Grad der Parallelisierung ist in Abhängigkeit zur Anzahl der CPUs und der gleichzeitig laufenden Analysen zu wählen. Ein zu hoher Parallelisierungsgrad birgt die Gefahr, dass wenige Queries die CPUs so stark auslasten, dass die Performance weiterer Queries darunter leiden würde. Ebenso steigt der Synchronisationsaufwand der einzelnen Prozesse.

## Fazit

Moderne Datenbankmanagementsysteme bieten eine Vielzahl von Funktionalitäten, um die Leistung eines Data-Warehouse-Systems zu steigern. So verwenden neue Datenbanktechnologien z.B. In-Memory-Techniken und das spaltenbasierte Speichern von Daten. Zu diesem Thema empfehlen wir Ihnen auch den Artikel „Einfach und schnell mit dem In-Memory Column Store?“ in dieser Ausgabe zu lesen.

Welche dieser Funktionen in einem konkreten Projekt sinnvoll sind, hängt von vielen Faktoren ab. Dies sind z.B. die verwendete Datenbank, das Datenvolumen, der SQL-typischen Query, der verwendeten Hardware und den Vorgaben bzgl. der gewünschten Antwortzeiten. Letztlich hängt es natürlich auch von den Lizenzkosten ab, die für einzelne Funktionalitäten ggf. anfallen können. Gerne unterstützen wir Sie beim Design oder der Performance-Optimierung Ihres Data Warehouse.

## Quellen

[1] Bauer, Andreas; Günzel Holger:  
„Data-Warehouse-Systeme: Architektur, Entwicklung, Anwendung“,  
4. überarbeitete und erweiterte Auflage, dpunkt.verlag GmbH, 2013



Michael Spoden  
(info@ordix.de)

## Seminarempfehlung: Data Warehouse Grundlagen

► Informationen/Online-Anmeldung: <http://training.ordix.de>

In diesem Seminar werden Sie mit den grundlegenden Konzepten und Komponenten eines Data Warehouse vertraut gemacht. Angefangen von der Anbindung von Quellsystemen, über die Befüllung des Data Warehouse mit ETL-Tools, bis hin zum Erstellen von Data-Marts werden alle Komponenten vorgestellt. Ebenso lernen Sie die verschiedenen Architekturansätze und Konzepte (z.B. Datenmodelle, Historisierung von Daten, etc.) innerhalb eines Data Warehouse kennen. Mehrere Livedemos vertiefen und veranschaulichen das Erlernete.

### Seminarinhalte

- Grundlagen DWH: Einführung und Begriffsdefinitionen
- Architektur eines Data Warehouse (Staging Area, Core Data Warehouse, Data-Mart, etc.)
- Datenmodelle (Star-Schema, Snowflake-Schema und Galaxie)
- Anbindung von Quellsystemen
- ETL-Überblick und Befüllung des Data Warehouse
- Historisierung von Daten
- Überblick Reporting
- Beispielhafte Datenbankoptimierungen für ein Data Warehouse
- Livedemos und Übungen

### Termine

27.05. - 29.05.2015 in Wiesbaden  
11.08. - 13.08.2015 in Wiesbaden  
20.10. - 22.10.2015 in Wiesbaden  
08.12. - 10.12.2015 in Wiesbaden

**Seminar-ID:** DB-DB-03  
**Dauer:** 3 Tage  
**Preis pro Teilnehmer:** 1.290,00 € (zzgl. MwSt.)  
**Frühbucherpreis:** 1.161,00 € (zzgl. MwSt.)



Buchen Sie gleich hier!



Im Gespräch mit unseren Vorständen Axel Röber, Christoph Lafeld und Benedikt Georgi

## „25 Jahre einfach. gut. beraten.“

Die ORDIX AG wird 25 Jahre alt. Dies hat die Redaktion der ORDIX® news zum Anlass genommen, in dieser Ausgabe drei der Vorstände zum Interview zu bitten. In der nächsten Ausgabe kommt der Vorstandsvorsitzende und Gründer Wolfgang Kögler zu Wort.

*Herr Röber, Herr Lafeld, Personalnummer 10 und 11, sagen Sie uns in wenigen Sätzen, was sich für Sie persönlich dahinter verbirgt.*

Axel Röber: Das sind zwei schöne binäre Zahlen. Bessere Personalnummern kann man sich als Informatiker gar nicht aussuchen. Meine Personalnummer 10 ist dann auch gleich noch die Basis unseres Dezimalsystems. Sie bildet auch die Summe der ersten vier Zahlen und wurde lt. Wikipedia schon früh zum Zeichen der Vollkommenheit und Vollendung.

Christoph Lafeld: Ich sehe das etwas leidenschaftsloser. Ich bin der 11. angestellte Mitarbeiter in der ORDIX Historie. Übrigens für meine Kollegen war das nicht immer ganz einfach: wir hatten eine Zeit lang Kurzwahlen, die mit 9 begonnen haben. Ich hatte die 911 und die Kollegen landeten meistens beim internationalen Notruf.

*Herr Georgi, Sie sind 2005 zur ORDIX AG gestoßen und haben mittlerweile auch schon Ihr 10-jähriges Jubiläum hinter sich. Was bewegt jemanden, von einem hochrangigen Managementposten bei einem weltweit tätigen Globalplayer zu einem kleinen, vom Unternehmer geführten Beratungsunternehmen zu wechseln und welche Erfahrungen haben Sie dabei in den letzten 10 Jahren gemacht?*

Benedikt Georgi: Ich möchte meine Zeit in einem Großunternehmen nicht missen. Ich hatte die Chance viele spannende Erfahrungen zu machen und herausfordernde Managementaufgaben zu übernehmen. Es gab dann aber einen Zeitpunkt in meinem Werdegang, an dem ich eine deutliche Veränderung wollte. Ich wollte wieder näher an den Kunden, näher an den Mitarbeitern und näher am Geschäft sein – und ich wollte weg von der Politik, die häufig in sehr großen Unternehmen herrscht. Diesen Wunsch

nach Veränderung konnte ich mit ORDIX perfekt umsetzen. Außerdem habe ich viel Neues gelernt, da ein Manager eines mittelständischen Unternehmens sich um viele kaufmännische und rechtliche Dinge im Detail kümmern muss, die in einem Großunternehmen durch spezialisierte Fachabteilungen übernommen werden.

*Sind Sie zufrieden mit dem, was Sie bisher erreichen konnten für sich persönlich als auch für Ihre (persönlichen) Ziele mit der ORDIX AG?*

Benedikt Georgi: Ich bin persönlich sehr zufrieden mit meiner bisherigen Zeit bei ORDIX. Bei der Bewertung bezüglich der von mir erreichten Ziele halte ich mich lieber zurück – das sollen andere beurteilen. Auf jeden Fall war es mein Ziel, das interne und externe Projektmanagement voranzubringen, den Erfolgsfaktor Kommunikation intern und extern zu stärken und den Vertrieb zu professionalisieren. Der Mensch steht bei uns im Mittelpunkt. Dafür habe ich mich immer engagiert.

*Herr Lafeld, Herr Röber, Sie waren jahrelang „normale“ Mitarbeiter der ORDIX AG, was sind Ihrer Meinung nach die wichtigsten Vorteile, die ORDIX als Arbeitgeber seinen Angestellten bietet?*

Axel Röber: Zu allererst bietet ORDIX seinen Mitarbeitern die Möglichkeit, sich permanent fachlich und persönlich weiter zu entwickeln – im Rahmen von Projekten, Seminaren und in der Zusammenarbeit in unserem internen Know-how-Netzwerk. Darüber hinaus ist der Umgang untereinander offen, unkompliziert und vertrauensvoll. Jeder kann jeden fragen. Hilfsbereitschaft ist für jeden bei ORDIX eine Selbstverständlichkeit. Das ist unsere Kultur!

Christoph Lafeld: Dem kann ich mich nur anschließen. Ich möchte aber ergänzen, dass für ein mittelständisches Unternehmen das Sozialpaket der ORDIX sehr gut – nahezu komfortabel – ist. Bei uns sind viele Dinge selbstverständlich geworden, die bei vergleichbaren, anderen Unternehmen eine Ausnahme sind bzw. gar nicht betrachtet werden.

*Herr Georgi, verglichen mit Großunternehmen wie Siemens oder ehemals Nixdorf, wo sehen Sie Unterschiede zur ORDIX als Arbeitgeber. Wo liegen die Vorteile für den Mitarbeiter, was machen „Große“ besser und wie kann ORDIX von diesem Wissen profitieren?*

Benedikt Georgi: Großunternehmen sind auch als Arbeitgeber sehr komplex – um nicht zu sagen kompliziert. Zur Illustration: Meine 16 Jahre in Großunternehmen machen 5 breite vollgepackte persönliche Ordner im Regal aus – meine 10 Jahre bei ORDIX nur 1 halbgefüllten schmalen Ordner. Da ich ein Fan von Komplexitätsreduzierung bin, finde ich, dass wir bei ORDIX gut unterwegs sind.

*Social Networking wird heute von vielen für unumgänglich gehalten, wo sind Sie selbst aktiv, welchen Nutzen sehen Sie für sich persönlich? Fordern und fördern Sie als Arbeitgeber die Präsenz der ORDIX AG auf diesen Plattformen?*

Axel Röber: Ich bin in den sozialen Netzen privat kaum aktiv, weil ich keinen Nutzen für mich sehe. Als Arbeitgeber und für mich beruflich sind die sozialen Medien mittlerweile eine völlig normale Angelegenheit, mit der man sich differenziert auseinander setzen muss. In erster Linie nutzen wir die sozialen Medien für die Ansprache von Bewerbern und für den fachlichen Austausch in Foren und Blogs.

Christoph Lafeld: Ich bin nur auf Xing eingetragen. Dieses Profil wird im Sinne der ORDIX genutzt. Privat bin ich gar nicht in sozialen Netzen unterwegs. Ich unterhalte mich mit der Familie und mit Freunden nach wie vor am liebsten persönlich. Die können dann selbst entscheiden, ob sie das Gesagte speichern oder nicht.

*Kommunikation findet heute weitestgehend elektronisch statt. Was erwarten Sie in Bezug auf die Erreichbarkeit per Mail etc. von Ihren Mitarbeitern? Gehen Sie dazu über wie Volkswagen oder andere Firmen, die berufliche Kommunikation zu Urlaubs- bzw. Nichtarbeitszeiten zu unterbinden?*

Axel Röber: Ich erwarte von meinen Mitarbeitern in der Regel nur, dass sie während der Arbeitszeiten erreichbar sind. Das Unterbinden der Kommunikation ist zwar ein sehr drastisches Mittel, aber im Einzelfall muss man darüber nachdenken, um bestimmte Mitarbeiter vor sich selbst zu schützen.

Christoph Lafeld: Grundsätzlich schließe ich mich Axel an. Allerdings gibt es Einzelfälle und Ausnahmen bei Mitarbeitern, denen man die Zusatzbelastung zutrauen kann und die auch damit einverstanden sind. Machen wir uns nichts vor: Wir sind ein kleines Unternehmen, das auch davon lebt, dass wir schnell reagieren. Da sind Ausnahmen in der o.g. Kommunikation schon mal notwendig.

*ORDIX ist bekannt dafür, dass jeder Mitarbeiter, egal wo er in der Hierarchie steht, intensiv in Kundenprojekte eingebunden ist. Wie direkt ist Ihr Kontakt zu Kunden? Was meinen Sie, was Ihre Kunden über Sie und Ihre Arbeit denken?*

Axel Röber: In unseren Kundenprojekten werde ich als Projektleiter oder technischer Ansprechpartner wahrgenommen. Meine Position bei ORDIX als Vorstand spielt dabei eine absolut untergeordnete Rolle.

Christoph Lafeld: Da bin ich sicher eine Ausnahme – auch im ORDIX Team. Ich bin zu 90% bei unserem Kunden vor

Ort und sehe mich aktuell eher wie ein Partner, der mit seinem Team beim Kunden agiert. Da kommt meine Vorstandstätigkeit gegenüber der Projektstätigkeit sicher etwas zu kurz. Was der Kunde über mich denkt, kann ich nicht sagen.

*Welche Kunden sind heute für ORDIX am wichtigsten und warum?*

Benedikt Georgi: Aus meiner Sicht sind für uns diejenigen Kunden am wichtigsten, die uns nicht so sehr als Lieferant sondern vielmehr als Partner sehen, mit dem sie vertrauensvoll und dauerhaft zusammen arbeiten und ihre Ziele erreichen wollen. Erfreulicherweise ist es uns gelungen, einige solcher wichtigen Kunden für eine langfristige Partnerschaft mit uns zu gewinnen.

*Tragen Sie mehr neue Technologien in die Firmen Ihrer Kunden oder beeinflusst eher die Umgebung bei Ihren Kunden die Weiterentwicklung bei ORDIX?*

Axel Röber: Das variiert natürlich in den unterschiedlichen Kunden- und Projektsituationen. In der Regel findet eine Symbiose statt, in der beide Seiten Technologien und Know-how einbringen.

Christoph Lafeld: Da sehe ich übrigens einen großen Vorteil in der ORDIX Strategie, Seminare zu geben. Wir bereiten für unseren Seminarbereich neue Technologien vor und müssen demzufolge einige Schritte weiter sein, als unsere Kunden. Natürlich gilt das nicht für alle Technologien, aber sicherlich für unsere Kernkompetenzen.

*Gibt es etwas, was Sie in der Zeit bei ORDIX heute anders sehen als zum damaligen Zeitpunkt?*

Axel Röber: Grundsätzlich hat sich meine Perspektive nicht verändert, allerdings hat ORDIX inzwischen eine Größenordnung angenommen, in der die Zusammenarbeit natürlich ganz anders organisiert werden muss als in den Gründungsjahren.

Christoph Lafeld: Wir sind ein mittelständisches Unternehmen geworden und sind nicht mehr die „große Familie“. Da hat sich schon einiges geändert.

*Wachstum ja oder nein, was meinen Sie und auf welches Wachstum ist ORDIX gut eingestellt?*

Axel Röber: Wachstum sollte nie ein Selbstzweck sein. ORDIX ist immer evolutionär und nachhaltig gewachsen – auch zu Zeiten der Internetblase. Diese Philosophie werden wir auch in Zukunft beibehalten.

Christoph Lafeld: Das stimmt. Jedoch müssen wir wachsen und meiner Meinung nach in Zukunft etwas schneller als in der Vergangenheit. Hier kommt uns jetzt unser Ausbil-

dungsbereich zu Gute, den wir von Beginn an aufgebaut haben und permanent weiterentwickeln und ausbauen.

*Ein wenig weg vom Beruflichen: Erlaubt es Ihnen Ihre Arbeit noch, Hobbys zu haben? Wenn ja welche haben Sie? Glauben Sie, dass Ihre Hobbys einen Einfluss auf Ihre Arbeit und Ihren Erfolg haben?*

Axel Röber: Ja, Hobbys sind für mich zwingend notwendig; sie ermöglichen mir den notwendigen Abstand zum Beruf zu gewinnen.

Benedikt Georgi: Mein Interesse für Kunst und Design war schon immer meine große Leidenschaft und war neben der Informatik durchaus auch eine mögliche Option für meinen beruflichen Werdegang. Ich habe dieses Interesse zu meinem Hobby gemacht und hole mir heute daraus Inspiration und Energie für meine Arbeit.

Christoph Lafeld: Ich habe immer gerne und viel Sport gemacht. Das ist für mich nach wie vor der beste Ausgleich und glücklicherweise teilen meine Familie und unser Hund diese Leidenschaft ;-).

*Und nun noch 2 Fragen kurz und knapp:*

*Was erzählen Sie Ihren Kindern (als diese noch jünger waren), was Sie bei ORDIX machen?*

Axel Röber: Ich mache irgendetwas mit Computern.

Christoph Lafeld: Dito. Irgendwann kam noch das Stichwort „Bank“ dazu.

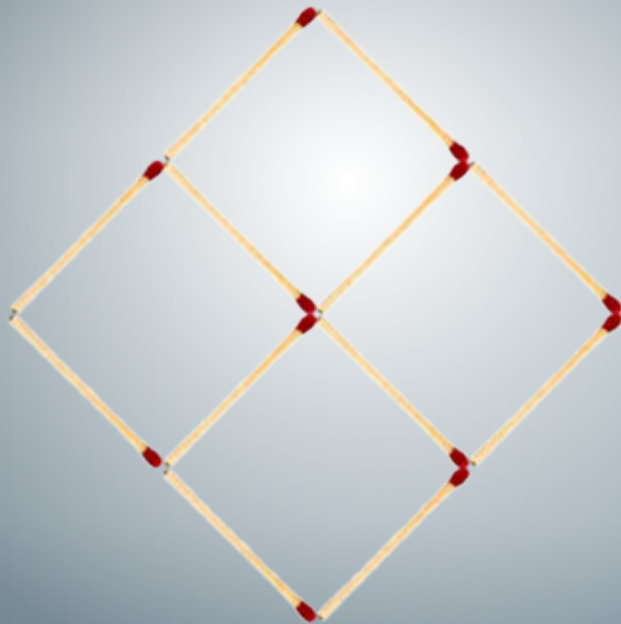
Benedikt Georgi: Mein jüngerer Sohn hat mich vor vielen Jahren mal bei ORDIX besucht und fand, dass am Computer „spielen“, mit dem Handy telefonieren und Kaffeetrinken keine richtige Arbeit sei.

*Wenn es nach Ihnen geht, gestalten Sie auch ORDIX 30?*

**Alle:** Aber sicher!

*Vielen Dank für das Interview.*





Rätsel

## Larry und die zehn Quadrate

Larry ist auf der Suche nach neuen Herausforderungen. Bei der Suche nach relevanten Jobs stößt Larry auf eine Stellenanzeige der ORDIX AG. Um nun beim bevorstehenden Vorstellungsgespräch zu punkten, möchte er die Lösung des abgebildeten Streichholz-Rätsels gleich präsentieren. Larry übt schon die ganze Zeit, aber er bekommt es einfach nicht hin.

### Bilde 10 Quadrate

Seit Tagen trägt Larry ständig 12 Streichhölzer mit sich herum und versucht mit diesen die folgende Aufgabenstellung zu lösen:

Die 12 Streichhölzer werden so aufgebaut, wie es in der obige Grafik gezeigt ist. Dann sollen vier Hölzer so umgelegt werden, dass genau zehn Quadrate entstehen.

### Können Sie Larry helfen?

Larry freut sich auf Ihren Lösungsvorschlag! Senden Sie Ihre Antwort bis zum **20. April 2015** an [kniffel@ordix.de](mailto:kniffel@ordix.de).

### Die Parole aus der letzten Ausgabe wurde gefunden!

Wir waren fast überwältigt von den vielen Antworten. Larry bedankt sich mit einer kleinen Aufmerksamkeit bei den drei schnellsten Einsendern:

Martin Eberhardt, Andreas Weller und Sabine Schiefer

### Lösung Larry Rätsel 3/2014:

Die richtige Antwort lautet 5.

Ausschlaggebend für die Parole ist die Anzahl der Buchstaben der vom Wächter genannten Zahl. Sagt der Wächter „12“ so besteht das Wort „Zwölf“ aus 5 Buchstaben.

Beispiele:

Sechzehn = 8 Buchstaben

Achtundzwanzig = 14 Buchstaben

**Bild:** © istockphoto.com | Computer Geek: Two Thumbs Up | sdominick

### Dieses Rätsel war keine Herausforderung für Sie?

Dann besuchen Sie doch unsere Seite:

[www.ich-will-ins-ordix-team.de](http://www.ich-will-ins-ordix-team.de)



Neue Reihe: Big Data - Informationen neu gelebt

## Wie big ist Big Data?

NoSQL, Hadoop, Map Reduce, HDFS, Storm, Hive, Spark,... Eine Vielzahl von Technologien wird im Umfeld von Big Data eingesetzt und wenn man manchen Veröffentlichungen Glauben schenken würde, dann werden durch sie alle unsere Probleme wie durch Zauberhand gelöst. Das ist in der Regel aber nicht der Fall! Durch den unüberlegten Einsatz können leicht Probleme gelöst werden, die es vorher gar nicht gab. Unsere neue Artikelreihe zum Thema Big Data startet mit einem ersten Überblick. In den folgenden Ausgaben werden die einzelne Technologien und Anwendungsfälle genauer beschrieben.

### Was ist Big Data überhaupt?

Wenn man nach einer Definition von Big Data sucht, dann stößt man schnell auf das 3V-Modell. Dieses geht auf ein Dokument von Doug Laney aus dem Jahr 2001 zurück [Quelle 1]. Getrieben durch E-Commerce, die fortschreitende Digitalisierung und die sozialen Netzwerke sind die Anforderungen an das Datenmanagement in den drei Dimensionen Volumen (Volume), Vielfalt (Variety) und Geschwindigkeit (Velocity) explodiert. Viele der dadurch entstehenden Probleme können durch den Einsatz von Big-Data-Systemen gelöst werden.

### Volumen - Wie big ist Big Data?

Viele Technologien sind aus der Notwendigkeit entstanden, riesige Datenmengen kostengünstig zu speichern und

zu verarbeiten. Beim benötigten Speicherplatz wird dabei oft in Terrabyte oder Petabyte gerechnet. Zum Beispiel hat Twitter aktuell 284 Millionen aktive Benutzer die 500 Millionen Tweets am Tag versenden. Zusammen mit Meta-informationen – wie Zeit, Geo-Lokation und Absender sowie den Log-Daten der beteiligten Systeme – werden jeden Tag mehr als 12 TB Daten erzeugt und gespeichert [Quelle 2].

### Vielfalt - Wie unstrukturiert ist Big Data?

Oft wird davon gesprochen, dass Big-Data-Systeme unstrukturierte Daten speichern und verarbeiten können. Das ist erst einmal richtig. Um die Informationen analysieren zu können muss dann aber wieder Struktur in die Daten

gebracht werden. Im Unterschied zu einer klassischen relationalen Datenbank wird die Struktur nicht beim Schreiben definiert sondern oft erst beim Lesen der Daten. Weiterhin können die Systeme sehr gut mit unterschiedlichen und sich verändernden Strukturen umgehen. So muss nicht jeder eingefügte Datensatz die gleichen Attribute besitzen. Ein typisches Beispiel für uneinheitliche Strukturen ist der Produktkatalog großer Online-Händler. Artikel aus verschiedenen Produktkategorien haben oft sehr unterschiedliche Eigenschaften. Die Speicherung und Analyse von Bild-, Video- und Audiodateien erhöht die Vielfalt der Daten nochmals beträchtlich.

### Geschwindigkeit - Wie schnell ist Big Data?

„Zeit ist Geld“ hieß es schon lange bevor über Big Data geredet wurde. Aber auch bei der Auswertung großer Datenmengen hat diese Redewendung ihre Berechtigung. Für ein soziales Netzwerk ist es zum Beispiel unerlässlich, dass Nachrichten schnell zugestellt werden. WhatsApp wäre nicht so erfolgreich, wenn die Zustellung einer Nachricht Minuten dauern würde. Ähnliches gilt für die Auswertung der Daten. Die klassische Analyse von Daten mittels einer Batch-Verarbeitung hat zwar noch lange nicht ausgedient, der Bedarf, große Datenmengen innerhalb weniger Millisekunden zu verarbeiten, steigt hingegen rasant an.

### Welche technischen Probleme werden gelöst?

Relationale Datenbanken skalieren vertikal in der Regel sehr gut. Mehr Hauptspeicher und CPUs sind oft ein geeignetes Mittel um die Performance zu verbessern. Allerdings sind der vertikalen Skalierbarkeit Grenzen gesetzt. Zum einen durch die maximale Erweiterbarkeit der verfügbaren Rechner und in der Praxis oft sehr viel früher durch den Preis sehr großer Rechner. Auch die horizontale Skalierbarkeit ist in der Regel sehr begrenzt, denn bei steigender Anzahl von Knoten steigt der Verwaltungsaufwand zur Durchführung von ACID-Transaktionen überproportional.

Viele Big Data Systeme sind von Haus aus verteilte Systeme. Sie skalieren sowohl vertikal als auch horizontal sehr gut. Während zum Beispiel ein Oracle RAC Cluster mit zwölf Knoten schon sehr groß ist, sind mehrere hundert Knoten in einem Big Data System nicht ungewöhnlich.

Die Systeme bestehen dabei oft aus vielen relativ günstigen Rechnern mit lokalen Festplatten. Auf große und teure Storage-Systeme wird vollständig verzichtet. Um die Verfügbarkeit der Systeme zu erhöhen und dem Verlust von Daten durch einen Defekt vorzubeugen werden alle Daten redundant im Cluster gespeichert. Dadurch kann auch die Verarbeitung der Daten beschleunigt werden, denn parallele Prozesse können auf mehrere Knoten verteilt werden. Ausfälle einzelner Knoten sind bei einem großen Cluster unvermeidlich. Diese werden aber akzeptiert und das System stellt beim Ausfall eines Knotens sicher, dass die Daten und die aktuell laufenden Prozesse im Cluster neu verteilt werden.

### Womit können die Probleme gelöst werden?

Es gibt eine riesige Auswahl an Technologien im Umfeld von Big Data. Während dieser Artikel einen ersten Überblick gibt, werden in den nächsten Teilen dieser Reihe einige Technologien genauer betrachtet. Den Anfang machen NoSQL-Datenbanken [1-4]. Diese gibt es für verschiedene Einsatzzwecke. Sie dienen in vielen großen Systemen als Ergänzung zu einer relationalen Datenbank, können diese aber unter bestimmten Voraussetzungen auch vollständig ersetzen. Ein kurzer Überblick folgt in den nächsten Abschnitten.

Eine weitere wichtige Komponente in vielen Big-Data-Systemen ist Apache Hadoop. Dabei handelt es sich nicht um ein einzelnes Produkt. Vielmehr ist es eine Sammlung von unterschiedlichen Open-Source-Projekten, die aufeinander aufbauen. Zentrale Komponenten sind dabei das verteilte Hadoop-Dateisystem (HDFS) und das darauf aufsetzende MapReduce-System zur parallelen Verarbeitung riesiger Datenmengen.

Als letzte Komponente werden die Stream-Processing-Systeme Storm und Spark Streaming kurz vorgestellt. Dabei handelt es sich um verteilte Systeme zur Verarbeitung von Datenströmen in Echtzeit.

### NoSQL-Datenbanken

Ursprünglich wurde der Begriff NoSQL für Datenbanken verwendet, die nicht relational waren und auch kein SQL unterstützt haben. Im Lauf der Zeit hat sich einiges geändert und SQL-fähige NoSQL-Datenbanken (zum Beispiel Apache Cassandra) entstanden. Heutzutage steht der Begriff für „Not only SQL“. NoSQL-Datenbanken können wie folgt kategorisiert werden (siehe Abbildung 1):

- Key-Value
- Spaltengruppenorientiert
- Dokumentenorientiert
- Graphen

Kategorie	Beispiele
Key-Value	Riak Amazon DynamoDB Oracle Berkeley DB
Spaltengruppenorientierte	Apache Cassandra Apache HBase Amazon SimpleDB
Dokumentenorientierte	CouchDB MongoDB
Graphen	Neo4J

Abb. 1: Verbreitete Vertreter der NoSQL-Datenbanken

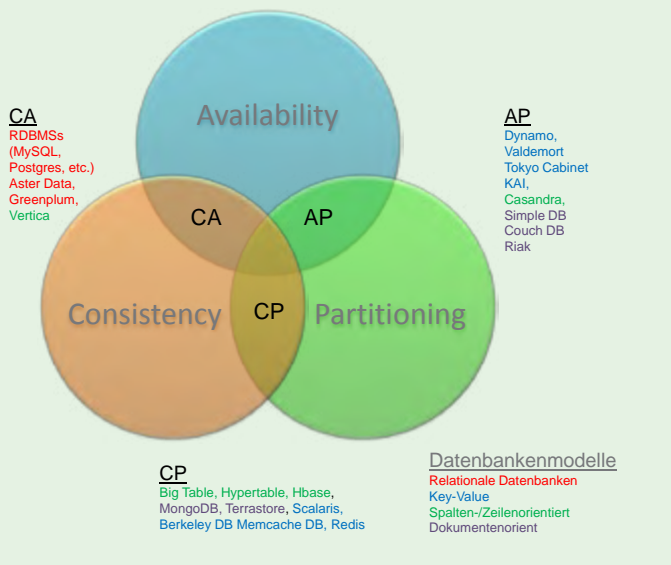


Abb. 2: CAP-Theorem

Komponente	Beispiele
Hadoop	Verteiltes Dateisystem HDFS und Dienste zum Ausführen von Map Reduce Jobs
Pig	Abfragesprache Pig-Latin und eine Laufzeitumgebung für die Ausführung von MapReduce Jobs
Hive	SQL-basierte Abfragesprache HiveQL
Sqoop	Werkzeug zum Datentransfer zwischen HDFS und einer relationalen Datenbank
Oozie	Workflow Scheduler
ZooKeeper	Verteilter Dienst zur Synchronisation und zur Verwaltung der Konfiguration im Cluster

Abb. 3: Hadoop-Distributionen

## Glossar

### ACID

ACID steht für die folgenden Eigenschaften eines Datenbanksystems: atomar (Atomicity), konsistent (Consistency), abgegrenzt (Isolation) und dauerhaft (Durability).

### CAP

Das CAP oder auch Brewers Theorem besagt, dass ein verteiltes System die Eigenschaften Konsistenz (Consistency), Verfügbarkeit (Availability) und Partitionstoleranz (Partition tolerance) nicht gleichzeitig garantieren kann.

### HDFS

Das Hadoop Distributed File System ist ein hochverfügbares Dateisystem, welches Daten verteilt auf mehreren Rechnern (Knoten) speichert.

### NoSQL

NoSQL steht für Not only SQL und bezeichnet Datenbanken, die nicht nur SQL unterstützen und die nicht relational sind.

## CAP-Theorem

Eine wichtige Eigenschaft von Big-Data-Systemen ist die Speicherung riesiger Datenmengen. Wie bereits erwähnt, wird dies durch die horizontale Verteilung der Daten in einem Cluster realisiert. Damit sind aber auch Einschränkungen verbunden. In einem verteilten System ist es laut dem CAP-Theorem nicht möglich die drei Eigenschaften Konsistenz (Consistency), Verfügbarkeit (Availability) und Partitionstoleranz gleichzeitig zu garantieren [Quelle 3]. Es muss immer auf eine dieser Eigenschaften verzichtet werden (siehe Abbildung 2). Aus diesem Grund sind verteilte Datenbanken auch nicht ACID-fähig.

Welche Eigenschaften unterstützt werden, ist von Datenbank zu Datenbank verschieden. Bei einigen Systemen hat der Entwickler sogar Einfluss auf die unterstützten Eigenschaften und kann sich entscheiden, ob zum Beispiel die Verfügbarkeit oder die Konsistenz garantiert werden soll. Nicht verteilte (No)SQL-Datenbanken (wie zum Beispiel Neo4J) haben diese Einschränkungen nicht und können somit auch ACID unterstützen.

## Polyglott Persistence

Vor allem die fehlende Unterstützung von ACID-Transaktionen führt dazu, dass eine NoSQL-Datenbank alleine oft nicht ausreicht, um ein Problem zu lösen. In einem E-Commerce-System kann der Produktkatalog zum Beispiel problemlos in einer dokumentenorientierten Datenbank verwaltet werden. Die Bestellungen und Bezahlvorgänge der Kunden würde man allerdings in einer relationalen Datenbank speichern. Für diese gleichzeitige Nutzung unterschiedlicher Datenbanken hat sich der Begriff „Polyglott Persistence“ durchgesetzt. Da eine zweite Datenbank zusätzliche Komplexität in ein System bringt, sollte die Entscheidung wohl überlegt sein. Sobald erhebliche Aufwände dadurch entstehen, dass man ein Problem mit der falschen Datenbank lösen will, sollte man nochmal in seinem Werkzeugkasten nach einem geeigneteren Werkzeug suchen.

## Apache Hadoop

Der Kern von Apache Hadoop besteht aus dem verteilten Hadoop Filesystem (HDFS) und dem MapReduce-System zur parallelen Verarbeitung der gespeicherten Daten. HDFS speichert alle Daten in relativ großen Blöcken ( $\geq 64$  MB) redundant im Cluster. MapReduce wiederum ist ein Programmiermodell für die Verarbeitung dieser Daten. Es abstrahiert vom physikalischen Lesen, Schreiben und Versenden der Daten im verteilten Hadoop-System. Weiterhin kümmert es sich automatisch um die Aufteilung von Jobs in einzelne Tasks und deren parallele Ausführung im Cluster.

## Hadoop-Distributionen

Das Hadoop-Ökosystem besteht nicht nur aus HDFS und MapReduce. Zusätzlich zu diesem Kern gibt es noch viele

weitere Open-Source-Projekte, in denen weitere Komponenten für Hadoop entwickelt werden. Einige davon sind in der Abbildung 3 aufgelistet. Allerdings werden diese Komponenten mehr oder weniger unabhängig voneinander weiterentwickelt. Um Probleme mit Inkompatibilitäten zu vermeiden oder wenigstens zu verringern empfiehlt es sich, auf eine vorgefertigte Hadoop-Distribution zurückzugreifen. Wie auch im Linux-Umfeld gibt es Hersteller, die aus den verschiedenen Komponenten ein vollständiges System zusammenstellen. Dies wird dann oft um Installations- und Management-Tools erweitert und zusammen mit Supportverträgen angeboten. Die bekanntesten Anbieter von Hadoop-Distributionen sind Cloudera, Hortonworks und MapR.

## Abfragesprachen

MapReduce-Jobs sind im Wesentlichen Programme, die in einer höheren Programmiersprache wie zum Beispiel Java implementiert sind und anschließend als Batch Job ausgeführt werden. Das ist auf der einen Seite extrem flexibel, da jeder Datentyp analysiert und verarbeitet werden kann. Für ad-hoc- oder Standardabfragen ist es aber eine große Hürde immer erst ein Programm schreiben zu müssen. Eine Abfragesprache, wie z.B. SQL, wäre hier sehr hilfreich.

Für dieses Problem gibt es verschiedene Lösungen. Die zwei bekanntesten sind Pig und Hive. Beide gehören standardmäßig zu den bekannten Hadoop-Distributionen. Während Pig die eigene Sprache PigLatin für Abfragen verwendet, werden Hive-Abfragen in SQL formuliert. In beiden Fällen wird im Hintergrund ein MapReduce-Job gestartet. Das ist für eine Batch-Verarbeitung in der Regel völlig in Ordnung. Für interaktive Abfragen, bei denen innerhalb von wenigen Sekunden eine Antwort erwartet wird, sind MapReduce-Jobs aber zu träge. Dieses Problem wird aktuell von mehreren Projekten angegangen. So gibt es die Stinger-Initiative, die sich um die Weiterentwicklung von Hive kümmert. Andere Projekte sind Impala von Cloudera und Spark SQL.

## Stream Processing

Auch für die Verarbeitung von Datenströmen in Echtzeit gibt es geeignete Lösungen. Genutzt werden diese zum Beispiel bei der automatisierten Analyse von Daten aus sozialen Netzwerken (z.B. Twitter Feeds). Ein weiterer typischer Anwendungsfall ist die Überwachung von Transaktionen zur Erkennung von Betrugsversuchen. Verbreitete Systeme sind Apache Storm und Apache Spark Streaming. Mit beiden kann sehr leicht ein horizontal skalierbares und hoch verfügbares System aufgebaut werden.

Im Gegensatz zu Hadoop oder NoSQL-Datenbanken können aber weder Storm noch Spark Streaming Daten speichern. Als Persistenzschicht wird in der Regel eine verteilte NoSQL-Datenbank verwendet. Einige der Hadoop-Distributionen enthalten bereits Storm und/oder Spark und bieten damit eine etwas leichtere Installation, Integration und Überwachung dieser Komponenten.

## Links

- [1] ORDIX® news Artikel 03/2012 „NoSQL vs. SQL - Hype oder echte Alternative? (Teil I): Was sind NoSQL-Datenbanken?“: <http://www.ordix.de/ordixnews/ordix-news-archiv/32012.html>
- [2] ORDIX® news Artikel 04/2012 „NoSQL vs. SQL - Hype oder echte Alternative? (Teil II): Oracle NoSQL“: <http://www.ordix.de/ordixnews/ordix-news-archiv/42012.html>
- [3] ORDIX® news Artikel 01/2013 „NoSQL vs. SQL - Hype oder echte Alternative? (Teil III): CouchDB - Time to Relax“: <http://www.ordix.de/ordixnews/ordix-news-archiv/12013.html>
- [4] ORDIX® news Artikel 02/2013 „NoSQL vs. SQL - Hype oder echte Alternative? (Teil IV): HBase – Spaltenorientiert“: <http://www.ordix.de/ordixnews/ordix-news-archiv/22013.html>
- [5] Seminarempfehlung:  
Big Data: Informationen neu gelebt (Seminar-ID: DB-BIG-01)  
<http://training.ordix.de>

## Quellen/Bildnachweis

- [1] Laney, Doug: „3-D Data Management: Controlling Data Volume, Velocity and Variety“, 2001  
<http://blogs.gartner.com/doug-laney/deja-vvvue-others-claiming-gartners-volume-velocity-variety-construct-for-big-data/>
- [2] Twitter:  
<https://about.twitter.com/company>  
<http://de.slideshare.net/kevinweill/analyzing-big-data-at-twitter-web-20-expo-nyc-sep-2010>
- [3] Wikipedia „CAP-Theorem“:  
<http://de.wikipedia.org/wiki/CAP-Theorem>
- © unplash.com | Juskteez | Stars
- © fotolia.de | mindscanner | Word Cloud "Big Data"
- © sxc.hu | pipp | notebook
- © freepik.com | Technology background with planet
- © freepik.com | Infographic chart

## Fazit

Wie bereits eingangs erwähnt, gibt dieser Artikel zunächst einen Überblick über das Thema Big Data und die eingesetzten Technologien. Für die ersten eigenen Gehversuche bietet es sich an, eine der verfügbaren Hadoop-Distributionen zu installieren oder aber eine fertige virtuelle Maschine zu verwenden. Beides ist kostenlos bei den großen Hadoop-Distributoren erhältlich. Für einen tiefergehenden Einblick werden in den nächsten Ausgaben der ORDIX® news weitere Artikel zu Big-Data-Technologien folgen.

Gerne begrüßen wir Sie auch zu unserem eintägigen Seminar „Big Data: Informationen neu gelebt“ [5] in dem wir Ihnen einen umfassenden Überblick über die verschiedenen Technologien und deren Einsatzmöglichkeiten geben. Oder aber unsere Experten kommen zu Ihnen, um ganz konkret über Ihre Anforderungen und deren Umsetzung zu sprechen.



Olaf Hein  
([info@ordix.de](mailto:info@ordix.de))

## Datenbanken

April - Juli 2015

DB-BIG-01	Big Data: Informationen neu gelebt	1 Tag	590,00 €	26.05.2015
DB-DB-03	Data Warehouse Grundlagen	3 Tage	1.290,00 €	27.05.2015
DB-ORA-01	Oracle SQL	5 Tage	1.890,00 €	15.06.2015
DB-ORA-01A	Oracle SQL Powerworkshop	3 Tage	1.290,00 €	27.07.2015
DB-ORA-02	Oracle Datenbankprogrammierung mit PL/SQL Grundlagen	5 Tage	1.890,00 €	22.06.2015
DB-ORA-34	Oracle Datenbankprogrammierung mit PL/SQL Aufbau	5 Tage	1.890,00 €	06.07.2015
DB-ORA-42	Oracle PL/SQL Tuning	3 Tage	1.290,00 €	20.04.2015   20.07.2015
DB-ORA-03	Oracle Datenbankadministration Grundlagen	5 Tage	1.990,00 €	20.04.2015   13.07.2015
DB-ORA-04	Oracle Datenbankadministration Aufbau	5 Tage	1.990,00 €	04.05.2015   27.07.2015
DB-ORA-32	Oracle Backup und Recovery mit RMAN	5 Tage	1.990,00 €	08.06.2015
DB-ORA-07	Oracle Tuning - Theorie und Interpretation von Reports	5 Tage	2.290,00 €	15.06.2015
DB-ORA-11	Oracle Troubleshooting Workshop	5 Tage	2.290,00 €	auf Anfrage
DB-ORA-08A	Oracle 12c RAC und Grid Infrastructure	5 Tage	2.290,00 €	22.06.2015
DB-ORA-33B	Oracle Security	5 Tage	1.890,00 €	15.06.2015
DB-ORA-31	Oracle Data Guard	5 Tage	1.990,00 €	22.06.2015
DB-ORA-35	Oracle Cloud Control	3 Tage	1.290,00 €	29.06.2015
DB-ORA-40	Oracle Capacity Planning	3 Tage	1.290,00 €	auf Anfrage
DB-ORA-46	Oracle APEX Anwendungsentwicklung Grundlagen	3 Tage	1.290,00 €	11.05.2015
DB-ORA-47	Oracle APEX Anwendungsentwicklung für Fortgeschrittene	3 Tage	1.290,00 €	26.05.2015
DB-ORA-48	Oracle Golden Gate	3 Tage	1.290,00 €	auf Anfrage
DB-ORA-49	Oracle 12c Neuheiten	5 Tage	2.090,00 €	20.04.2015   29.06.2015
DB-INF-01	IBM Informix SQL	5 Tage	1.790,00 €	13.07.2015
DB-INF-02	IBM Informix Administration	5 Tage	1.990,00 €	20.04.2015   27.07.2015
DB-INF-04	IBM Informix Backup und Recovery	3 Tage	1.290,00 €	auf Anfrage
DB-INF-03	IBM Informix Tuning und Monitoring	5 Tage	1.990,00 €	auf Anfrage
DB-DB2-01	IBM DB2 für Linux/Unix/Windows SQL Grundlagen	5 Tage	1.890,00 €	04.05.2015
DB-DB2-02	IBM DB2 für Linux/Unix/Windows Administration	5 Tage	1.990,00 €	13.04.2015
DB-DB2-05	IBM DB2 für Linux/Unix/Windows Monitoring und Tuning	3 Tage	1.290,00 €	11.05.2015
DB-MY-01	MySQL Administration	3 Tage	1.290,00 €	13.04.2015   29.06.2015
MS-SQL-01A	Querying SQL Server	5 Tage	1.990,00 €	20.04.2015   06.07.2015
MS-SQL-02A	Administering Microsoft SQL Server Databases	5 Tage	1.990,00 €	04.05.2015
MS-SQL-05A	Implementing a Data Warehouse with MS SQL Server	5 Tage	1.990,00 €	20.04.2015
MS-SQL-06	Developing Microsoft SQL Server Databases	5 Tage	1.990,00 €	15.06.2015
MS-SQL-07	Updating Your SQL Server Skills to Microsoft SQL Server 2014	5 Tage	1.990,00 €	08.06.2015
MS-SQL-08	Designing a Data Solution with Microsoft SQL Server	5 Tage	1.990,00 €	22.06.2015
MS-SQL-09	Implementing Data Models & Reports with Microsoft SQL Server	5 Tage	1.990,00 €	13.04.2015   13.07.2015
MS-SQL-10	Designing Self-Service Business Intelligence & Big Data Solutions	5 Tage	1.990,00 €	04.05.2015
MS-SQL-11	Microsoft SQL Server 2014 for Oracle DBAs	4 Tage	1.790,00 €	27.04.2015

## Entwicklung

April - Juli 2015

P-PHP-01	PHP Programmierung Grundlagen	5 Tage	1.690,00 €	15.06.2015
P-PERL-01	Perl Programmierung Grundlagen	5 Tage	1.690,00 €	22.06.2015
P-PERL-02	Perl Programmierung Aufbau	5 Tage	1.690,00 €	13.04.2015   29.06.2015
P-UNIX-01	Shell, Awk und Sed	5 Tage	1.690,00 €	08.06.2015
P-XML-01	Einführung in XML	3 Tage	1.190,00 €	06.07.2015

## Informationen und Anmeldung

Zentrale:  
**ORDIX AG**  
 Westernmauer 12 - 16  
 33098 Paderborn  
 Tel.: 05251 1063-0

Seminarzentrum:  
**ORDIX AG**  
 Kreuzberger Ring 13  
 65205 Wiesbaden  
 Tel.: 0611 77840-00

**Online-Anmeldung,  
 aktuelle Seminarinhalte  
 und Termine unter:**  
<http://training.ordix.de>



Unser Seminarstandort ist Wiesbaden.  
 Die hier angegebenen Termine entsprechen dem  
 ersten Seminartag.  
 Die Preise gelten pro Seminar pro Teilnehmer in Euro  
 zzgl. ges. MwSt., Inhouse-Preise auf Anfrage.

## Web- und Application Server

April - Juli 2015

INT-04	Apache HTTP Server Administration	3 Tage	1.190,00 €	08.06.2015
INT-07	Tomcat Konfiguration und Administration	3 Tage	1.290,00 €	29.06.2015
INT-08	WebSphere Application Server Installation und Administration	3 Tage	1.390,00 €	08.06.2015
INT-11_7	Administration und Konfiguration für JBoss 7	3 Tage	1.290,00 €	27.04.2015   06.07.2015
DB-ORA-50	Oracle Weblogic Administration Grundlagen	3 Tage	1.390,00 €	18.05.2015   20.07.2015

## Betriebssysteme und Monitoring

April - Juli 2015

BS-01	Unix/Linux Grundlagen für Einsteiger	5 Tage	1.690,00 €	20.04.2015   22.06.2015
BS-25	Unix Power-Seminar für Datenbank- und Applikationsbetreuer	5 Tage	1.890,00 €	13.04.2015   20.07.2015
BS-02	Linux Systemadministration	5 Tage	1.690,00 €	04.05.2015   13.07.2015
BS-09	Linux Hochverfügbarkeits-Cluster	5 Tage	1.890,00 €	13.04.2015
BS-03-11	Solaris 11 Systemadministration Grundlagen	5 Tage	1.990,00 €	27.07.2015
BS-04-11	Solaris 11 Systemadministration Aufbau	5 Tage	1.990,00 €	04.05.2015
BS-06-11	Solaris 11 für erfahrene Unix/Linux-Umsteiger	5 Tage	1.990,00 €	15.06.2015
BS-24	Solaris 11 Administration Neuheiten	3 Tage	1.290,00 €	08.06.2015
BS-18	Solaris Virtualisierung mit ZFS und Container (Zonen)	5 Tage	1.990,00 €	29.06.2015
BS-23	Solaris Virtualisierung mit LDOM	3 Tage	1.290,00 €	auf Anfrage
AIX-01	IBM AIX Systemadministration Grundlagen	5 Tage	1.990,00 €	22.06.2015
AIX-02	IBM AIX Installation, Backup und Recovery mit NIM	3 Tage	1.290,00 €	06.07.2015
SM-NAG-01	Systemüberwachung mit Nagios Grundlagen	3 Tage	1.190,00 €	auf Anfrage
SM-NAG-02	Systemüberwachung mit Nagios Aufbau	2 Tage	890,00 €	auf Anfrage

## Projekt-/IT-Management

April - Juli 2015

PM-01	IT-Projektmanagement - Methoden und Techniken	5 Tage	1.990,00 €	20.04.2015
PM-05	IT-Projektcontrolling	2 Tage	1.190,00 €	23.04.2015
PM-06	Projekte souverän führen - Systemisches Projektmanagement	4 Tage	1.850,00 €	04.05.2015
PM-08	Agiles Projektmanagement mit Scrum	2 Tage	1.190,00 €	08.06.2015
PM-08-Z	Scrum Praxis und Zertifizierung	1 Tag	590,00 €	10.06.2015
PM-10	IT-Controlling	3 Tage	1.650,00 €	11.05.2015
PM-14	IT-Anforderungsmanagement	2 Tage	1.190,00 €	27.04.2015
PM-16	Mit effektiver Kommunikation zum Projekterfolg	2 Tage	1.190,00 €	29.04.2015
PM-07	Krisenmanagement in Projekten	2 Tage	1.100,00 €	auf Anfrage
PM-11	Konfliktmanagement	2 Tage	1.100,00 €	auf Anfrage
PM-12	Stresskompetenz für IT-Führungskräfte	1 Tag	550,00 €	auf Anfrage
PM-13	Stresskompetenz für IT-Mitarbeiter	2 Tag	1.100,00 €	auf Anfrage
MGM-02	IT-Architekturen	3 Tage	1.650,00 €	06.07.2015
MGM-07	IT-Strategien effizient entwickeln	2 Tage	1.100,00 €	auf Anfrage
MGM-03	IT-Management	3 Tage	1.650,00 €	15.06.2015
MGM-05	IT-Risikomanagement	3 Tage	1.690,00 €	27.10.2014
MGM-04	Geschäftsprozessmanagement	3 Tage	1.650,00 €	22.06.2015

## Java / Java EE

April - Juli 2015

E-SWA-01	Softwarearchitekturen	5 Tage	1.890,00 €	08.06.2015
OO-01	Einführung in die Objektorientierte Programmierung und UML	3 Tage	1.190,00 €	26.05.2015
P-JAVA-01	Java Programmierung Grundlagen	5 Tage	1.690,00 €	20.07.2015
P-JAVA-03	Java Programmierung Aufbau	5 Tage	1.690,00 €	04.05.2015
P-JAVA-11	Java 8 Neuheiten	2 Tage	990,00 €	26.05.2015
P-JAVA-12	Java EE Powerworkshop	5 Tage	1.890,00 €	13.04.2015   20.07.2015
P-JEE-01	JEE für Entscheider	1 Tag	590,00 €	auf Anfrage
P-JEE-03A	JSP und Servlet Programmierung	5 Tage	1.590,00 €	auf Anfrage
P-JEE-05	Webanwendungen mit JavaServer Faces (JSF)	5 Tage	1.590,00 €	15.06.2015
INT-05	Java Web Services	3 Tage	1.190,00 €	11.05.2015
P-JEE-08	Java Performance Tuning	3 Tage	1.290,00 €	27.04.2015   13.07.2015

# Bestandteil einer nachhaltigen Unternehmensausrichtung

Das IT-Controlling hat das Ziel, die IT-Governance zu unterstützen. Die IT-Governance hat wiederum die optimale Umsetzung der IT-Strategie zum Ziel – zum einen, zur direkten Unterstützung der Einzelstrategien der einzelnen Geschäftsbereiche, zum anderen zur indirekten Unterstützung der Gesamtunternehmensstrategie. Die Unternehmensstrategie soll in der Regel den Erhalt bzw. den Ausbau der Marktanteile des Unternehmens sicherstellen und die Schaffung von Wettbewerbsvorteilen erreichen. Damit besteht die Hauptaufgabe des heutigen IT-Controllings, in der Unterstützung bei der Vorbereitung von Unternehmensentscheidungen.

## Bedeutung

Daraus ergibt sich, dass dem IT-Controlling, wie der gesamten Umsetzung der optimalen IT-Governance in vielen Unternehmen eine immer größer werdende Bedeutung zukommt. In vielen Branchen wird mittels der IT-Infrastruktur und den darauf aufbauenden Lösungen ein deutlicher Wettbewerbsvorteil gegenüber den Mitbewerbern erzielt.

Das IT-Controlling im Speziellen umfasst, neben dem vorgenannten Beratungsmandat für die Unternehmensführung, die Umsetzung des analytischen Planungsprozesses für die IT-Produkte des aktuellen IT-Leitungskataloges sowie die laufenden im Portfoliomanagement ausgewählten IT-Projekte, die IT-spezifische Prozesskostenverrechnung und das Berichtswesen unter steuerungsorientierter Zielrichtung.

## Arten des Controlling-Modells

Die Positionierung im Unternehmen und damit das Rollenspiel des IT-Controllings kann in unterschiedlichen Modellen erfolgen:

### Partnerschaftsmodell

Bei der Zusammenarbeit im Partnerschaftsmodell ist der Leiter des IT-Controllings direkt der Unternehmensführung unterstellt und damit auf der gleichen Hierarchieebene wie der CIO und der Leiter des Unternehmenscontrollings angesiedelt.

### CIO-Mitarbeiter-Modell

Bei diesem Modell ist der Leiter des IT-Controllings dem CIO untergeordnet. Gerade im Hinblick auf Meinungsverschiedenheiten kann diese disziplinarische Unterstellung zu Problemen führen.

## Controlling-Modell

Anders als bei den vorgenannten beiden Modellen ist im Controlling-Modell das IT-Controlling eine Teilaufgabe des Unternehmenscontrollings. Demzufolge berichtet der Leiter des IT-Controllings an den Leiter des Unternehmenscontrollings.

Wenn man sich auf das eigentliche Unternehmensziel, nämlich die nachhaltige Unternehmensausrichtung an den Anforderungen der Stakeholder des Unternehmens (wie z.B. Mitarbeiter, Führungskräfte, Aktionäre, Kunden, Lieferanten, Gesellschaft etc.) konzentriert und aus diesem Blickwinkel die vorangestellten Modelle der Zusammenarbeit des CIO mit dem Leiter des IT-Controlling betrachtet, dann empfehlen sich zwei Modelle: Das Partnerschaftsmodell und das Controlling-Modell.

Im Wesentlichen unterscheiden sich die beiden Modelle in der Form, dass sich beim Partnerschaftsmodell alle – der CIO, der Leiter des Unternehmenscontrollings sowie der Leiter des IT-Controllings – auf der gleichen Hierarchieebene begegnen. Dies wirkt zuerst optimal. Allerdings birgt es auch eine potenzielle Gefahr in der Gestalt, dass der Leiter des IT-Controllings für sein unternehmensweites Reporting eine autarke Datenbasis zum Unternehmenscontrolling aufbaut, die zusätzlich befüllt, gepflegt und gewartet werden muss.

Diese Gefahr ist bei dem Controlling-Modell, durch die hierarchische Unterstellung des IT-Controllings in dieser Form nicht gegeben. Wenn wir an dieser Stelle einen kleinen Vorgriff auf die nachfolgende Thematik der Darstellung des Leitbildes des IT-Controllers vorwegnehmen, dann ist dies eine der zentralen Aufgabenstellungen des IT-Controllers.

Um den Beratungsauftrag des Controllings in seiner Gesamtheit gegenüber der Unternehmensleitung optimal wahrnehmen zu können, empfiehlt sich das Controlling-Modell – zumindest in der Theorie. In der Praxis sollte



allerdings das gewählte Modell immer an der individuellen Unternehmensstruktur, -kultur und dem Organisationsaufbau des jeweiligen Unternehmens ausgerichtet sein.

Vor diesem Fazit wird nachfolgend das Leitbild des Controllers gemäß der Gesellschaft für Informatik e.V. dargestellt [Quelle 1].

### Leitbild des Controllers

Wenn wir das Leitbild des IT-Controllers analysieren und bewerten, dann zeigen sich auf dieser Granularität Beziehungspunkte der Einzelaufgabenstellungen im Hinblick auf die unterschiedlichen Unternehmensebenen, bestehend aus der normativen-, der strategischen- und der operativen Unternehmensebene.

Diese Darstellung veranschaulicht, dass der IT-Controller im Rahmen der eigenen Organisation (IT) neben dem jeweiligen Key Account Manager als zentraler Ansprechpartner des Geschäftsbereiches auch das entsprechende Senior Management in der Wahrnehmung seines Beratungsmandats in verschiedenen Teilaufgaben berät und unterstützt. Dies dient dazu, die vorgegebenen Ziele der IT-Governance zu realisieren bzw. bei auftretenden Abweichungen oder Problemen frühzeitig im Rahmen des beherrschten Reporting-Prozesses aufzuzeigen und dem Senior Management Alternativen zum ursprünglich festgelegten Vorgehen vorzulegen.

Deutlich wird auch, dass ähnlich dem Rollenbild des Key Account Managers auch der IT-Controller ein Mittler zwischen der fachlichen Welt der Geschäftsbereiche und der IT mit ihrer stärkeren technischen Fokussierung und dem damit verbundenen eigenen fachspezifischen Vokabular darstellt.

Damit die Implementierung eines unternehmensweiten Reportings die damit verbundene Transparenz auf beiden aufgeführten Seiten zur Analyse und Bewertung des aktuellen Status auf dem Weg zur jeweiligen Zielerreichung auch schafft, muss dem IT-Controller die Überwindung dieser fachspezifischen Hürden durch ein allgemeinverständliches Reporting glücken.

Der Grad des Erfolges, auf dem Weg dahin hängt zentral von der Wirksamkeit des unternehmensweiten Reportings und damit verbunden von der Qualität der erreichten Tiefe und Breite der Transparenz ab.

### Fazit

Nur das optimale Zusammenspiel zwischen der nachhaltigen Unternehmensstrategie und dem ganzheitlichen Controlling, abgestimmt auf die Qualitäts- und Risiko-managementstrategie sichern den dauerhaften Unternehmenserfolg (siehe Abbildung 1). Dieses Resümee lässt sich 1:1 auf das Zusammenspiel von IT-Controlling und IT-Governance übertragen.

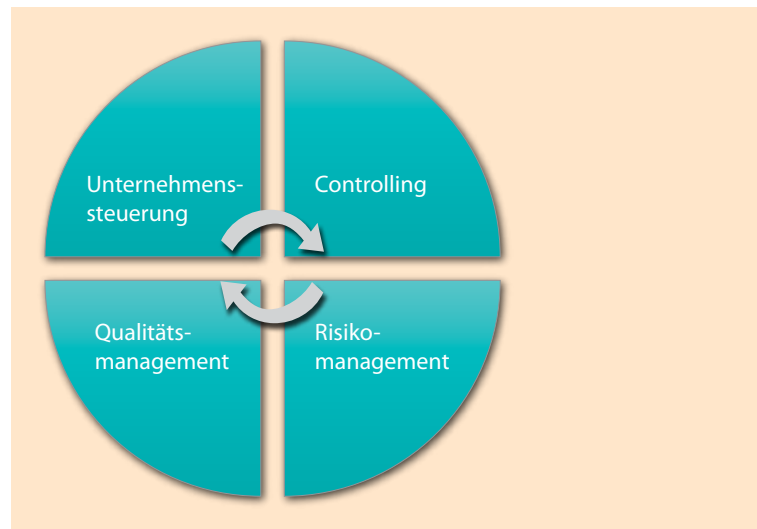


Abb. 1: Integriertes Managementsystem

### Links

[1] Internetseite der coniatos AG  
<http://www.coniatos.de>

[2] Seminarempfehlung: IT-Controlling (Seminar-ID: PM-10)  
<http://training.ordix.de>

### Quellen

[1] Leitbild des Controllers gemäß der Gesellschaft für Informatik e.V.:  
<https://www.gi.de/fileadmin/redaktion/empfehlungen/GI-Empfehlung-IT-Controllerin.pdf>

[2] Luser, Oliver: „Nachhaltigkeitsmanagement – Als Kernbestandteil der ganzheitlichen Unternehmensstrategie, E-Book, 1. Auflage, Oktober 2014  
<https://itunes.apple.com/de/book/nachhaltigkeitsmanagement/id929232256?mt=11>

[3] Helmke, Stefan, Uebel, Matthias: „Managementorientiertes IT-Controlling und IT-Governance“, SpringerGabler, 2013



Oliver Luser  
 (info@coniatos.de)



Java 8 - Die neue Version (Teil IV)

## Die Zeiten ändern sich

Endlich ändern sich die Zeiten, werden sicher viele sagen. Mit Java 8 hat auch die neue Date/Time-API Einzug in den Standard gehalten. Jeder der schon einmal das „Vergnügen“ hatte, mit den vorherigen Java-Versionen eine komplexere Datumsarithmetik zu entwickeln, wird schnell an die Grenzen bzw. an deren Eigenheiten gestoßen sein, die augenscheinlich nicht mehr in die moderne Java-Welt passen. Im folgenden Artikel, wollen wir die neue Date/Time-API einmal genauer betrachten und beleuchten, wie sie funktioniert und wie wir ihre Vorteile nutzen können.

### Warum ist eine neue API überhaupt notwendig?

Schnell kommt die Frage auf, warum wird überhaupt eine neue API benötigt? Die Antwort liegt auf der Hand, wenn man sich einmal die Nachteile der alten API vor Augen führt:

Fast jeder wird schon einmal bei dem Objekt `Calendar` über die Zählweise der Monate gestolpert sein. Es wurde wohl das alte Konzept zugrunde gelegt, dass es in der Entwicklung üblich ist, bei Iterationen mit 0 zu beginnen. An dieser Regel wurde auch bei den Monaten festgehalten.

Daher ist der Dezember „verständlicher Weise“ der Monat „11“ und der Januar dementsprechend der Monat „0“ eines Jahres. Über diese Logik könnte ggf. hinwegsehen werden, wenn man weiß, womit man es zu tun hat und mindestens einmal darauf hereingefallen ist.

Viel gravierender sind hingegen die Nachteile der Thread-Sicherheit. Das Objekt `Date` ist „mutable“ und ist demnach nicht Thread-sicher. Alle Zugriffe innerhalb von Threads müssten daher synchronisiert werden.

In Wirklichkeit ist das Objekt `Date` auch kein Datum, sondern ein Zeitstempel. Diesen kennen wir aus der Unix-Welt und startet entsprechend am 01. Januar 1970. Die Genauigkeit dieses Zeitstempels ist auf Millisekunden beschränkt. Eine Berechnung von beispielsweise Nanosekunden ist daher leider nicht möglich.

Es gibt noch eine Reihe weiterer Nachteile, die hier aufgelistet werden könnten, z.B. dass das Objekt `Date` keine Zeitzonen und auch keine Sommer-/Winterzeit kennt.

Wir wollen uns jedoch nicht nur mit den Nachteilen der alten Welt befassen, sondern uns lieber den Neuerungen widmen.

## Technischer Überblick

Die neue Date/Time-API wurde in Form des JSR 310 spezifiziert. Auch hier wurde das Rad nicht neu erfunden, sondern ein bewährtes Framework als Vorlage genommen. Große Teile der neuen API basieren auf der Bibliothek Joda-Time, welche bereits sehr früh als Referenz galt, wenn es in Java-Anwendungen um komplexe Datumsarithmetik ging. Sowohl die Bibliothek Joda-Time, als auch die neue Date/Time-API basieren auf der Designstudie „TimeAndMoney“ von Eric Evans (siehe Quelle [1]).

Alle Neuerungen der Date/Time-API befinden sich in den folgenden Packages:

- **java.time**  
Klassen für Datum, Uhrzeit, Datum und Uhrzeit kombiniert, Zeitzonen, Zeitpunkte, Dauer und Uhren
- **java.time.chrono**  
API für Kalendersysteme, die nicht dem Standard ISO 8601 entsprechen
- **java.time.format**  
Klassen zum Parsen und Formatieren von Daten und Uhrzeiten
- **java.time.temporal**  
Erweiterte API, die die Zusammenarbeit zwischen Datums- und Zeitklassen, Abfragen und Anpassungen an Daten und Uhrzeiten ermöglicht
- **java.time.zone**  
Klassen, die Zeitzonen, Differenzen zwischen Zeitzonen und Regeln der jeweiligen Zeitzonen unterstützen

Im Folgenden wollen wir uns ein paar ausgesuchte Aspekte der neuen API genauer anschauen. Natürlich ist es nicht möglich in diesem Artikel jedes Detail der neuen API aufzuzeigen, dennoch sollen Sie ein Gefühl dafür bekommen, wie die neue API einzusetzen und wie deren Mehrwert einzuschätzen ist.

### java.util.Date vs. java.time.Instant / Zeitpunkte

Das klassische `java.util.Date` ist genau genommen kein Datum, auch wenn der Name es vielleicht suggeriert. Es ist vielmehr ein definierter Zeitpunkt, der in Millisekunden ausgedrückt ist. In der neuen API gibt es hierfür die Klasse `java.time.Instant`. Auch sie repräsentiert einen bestimmten Zeitpunkt - mit dem Unterschied, dass die davon erzeugten Objekte unveränderlich und damit Thread-sicher sind und die Genauigkeit bis auf die Nanosekunde geht.

```
Date altesDateObjekt = new Date();

//Konvertierung von alt nach neu
Instant neuesInstantObjekt = altesDateObjekt.toInstant();

//Alternative
Instant neuesInstantObjekt = Instant.ofEpochMilli(altesDateObjekt.getTime());

//Konvertierung von neu nach alt
Date altesDateObjekt = Date.from(neuesInstantObjekt);
```

Abb. 1: Konvertierung von `java.time.Instant` nach `java.util.Date` und zurück

Ein Objekt vom Typ `java.util.Date` kann sehr einfach in ein Objekt vom Typ `java.time.Instant` konvertiert werden. Dafür wurde die alte Klasse `java.util.Date` um die Methode `getInstant` erweitert. Wie der Name schon verrät, wird hier ein neues `Instant`-Objekt auf Basis des `Date`-Objektes erzeugt und zurückgegeben. Wahlweise kann hierfür auch die Methode `ofEpochMilli` von der Klasse `Instant` verwendet werden. Hier muss lediglich der Wert übergeben werden, der von der Methode `getTime` des `Date`-Objektes zurückgeliefert wird. Für den umgekehrten Weg, wurde die Klasse `java.util.Date` um die Methode `from` erweitert.

Die Verwendung der Klasse wird in der Abbildung 1 verdeutlicht. In dem Code-Beispiel wird die einfache Verwendung und die Konvertierung der jeweiligen Klassen aufgezeigt.

### Daten und Uhrzeiten

Wird mehr als ein Zeitpunkt (`Instant`) benötigt, griff man bisher immer auf das Objekt `Calendar` zurück. Auf ein paar Nachteile dieser Klasse wurde bereits einleitend eingegangen. Bei der neuen Date/Time-API gibt es für die Datumsarithmetik nun eine Vielzahl von Klassen, die nur die Menge an Informationen beinhalten, welche auch wirklich benötigt werden. Hierzu zählen unter anderem die folgenden Klassen:

- **LocalDate**  
Datum ohne Zeitangabe
- **LocalTime**  
Uhrzeit ohne Datumsinformationen
- **LocalDateTime**  
Datum inklusive Uhrzeit

Es wird deutlich, dass für jedes Szenario eine eigene Klasse existiert. Das macht den Code in jedem Fall klarer und dadurch auch besser wartbar. Zuvor konnte anhand des Objektes nicht unterschieden werden, ob beispielsweise die Uhrzeit eine Rolle spielt oder nicht.

```

LocalDateTime now = LocalDateTime.now();
LocalDateTime lastDayOfFebruary = now.withDayOfMonth(1)
    .withMonth(Month.MARCH.getValue())
    .withHour(8)
    .withMinute(0)
    .withSecond(0)
    .minusDays(1);

System.out.println(lastDayOfFebruary.format(DateTimeFormatter
    .ofPattern("dd.MM.yyyy HH:mm")));
//AUSGABE: 28.02.2015 08:00

```

Abb. 2: Beispiel für die Verwendung der Fluent Interfaces: Setzen eines Datums auf 8:00 Uhr des letzten Tages im Februar

```

DateTimeFormatter format = DateTimeFormatter.ofPattern("dd.
MM.yyyy HH:mm");

// Abflug von Berlin am 15. Januar 2015 um 17:45 Uhr
LocalDateTime abflugDatum = LocalDateTime.of(2015, Month.
JANUARY, 15, 17, 45);
ZoneId zoneBerlin = ZoneId.of("Europe/Berlin");
ZonedDateTime abflug = ZonedDateTime.of(abflugDatum, zo-
neBerlin);

try {
    System.out.println("Abflug: " + abflug.format(format) + "
Uhr in" + zoneBerlin);
} catch (DateTimeException ex) {
    ex.printStackTrace();
}

// Der Flug dauert 16 hours und 50 minuten
ZoneId zoneTokyo = ZoneId.of("Asia/Tokyo");
ZonedDateTime ankunft = abflug.withZoneSameInstant(zone-
Tokyo).plusMinutes( (16*60+50) );

try {
    System.out.printf("Ankunft: " + ankunft.format(format)
+ " Uhr in" + zoneTokyo );
} catch (DateTimeException ex) {
    ex.printStackTrace();
}

//AUSGABE
//Abflug: 15.01.2015 17:45 Uhr in Europe/Berlin
//Ankunft: 16.01.2015 18:35 Uhr in Asia/Tokyo

```

Abb. 3: Beispiel für die Berechnung der Ankunftszeit eines Fluges von Berlin nach Tokio

Anstelle der bekannten Methoden `add` und `set` für die Datenmanipulation bei der Klasse `java.util.Calendar`, stehen nun die Methoden `plus`, `minus` und `with` zur Verfügung. Die Methoden `plus` und `minus` sind, wie der Name vermuten lässt, für die Addition bzw. Subtraktion von Daten und Uhrzeiten zuständig. Mit der Methode `with` werden einzelne Eigenschaften bzw. Informationen des Datums oder der Uhrzeit neu gesetzt. Zum Beispiel kann mit dem Aufruf `withDayOfMonth(1)` das zugehörige `LocalDate`- oder `LocalDateTime`-Objekt auf den ersten Tag des jeweiligen Monats gesetzt werden (siehe Abbildung 2).

Diese Methoden lassen sich beliebig kaskadieren. Das Konzept ist auch unter dem Begriff „Fluent Interfaces“ bekannt und wurde von Martin Fowler und Eric Evans entwickelt (siehe Quelle [2]).

## Zeitzone

Bei der Datumsarithmetik gibt es kaum eine schwierigere Aufgabe, als die Berechnung von Daten über verschiedene Zeitzone hinweg. Für diesen Anwendungsfall wurde die alte API so gut wie nie verwendet und das hat auch seinen guten Grund.

Die neue Date/Time-API versucht sich dem Problem anzunehmen und stellt für diesen Anwendungsfall u.a. drei weitere Klassen zur Verfügung:

- **ZoneId**  
Diese Klasse definiert eine Zeitzone als eindeutige ID und beinhaltet Regeln für die Konvertierung zwischen den Objekten `Instant` und `LocalDateTime`.
- **ZoneOffset**  
Die Klasse beinhaltet die Abweichung von einer Zeitzone zu der Zeitzone „Greenwich/UTC“.
- **ZonedDateTime**  
`DateTime`-Objekt mit zusätzlicher Information der zugehörigen Zeitzone

In der Abbildung 3 wird die örtliche Ankunftszeit eines Fluges von Berlin nach Tokio berechnet. Für eine Berechnung über die jeweiligen Zeitzone hinweg wird zunächst das Abflugdatum in Form eines Objektes `ZonedDateTime` benötigt. Um dieses Objekt zu erzeugen wird von der Klasse `ZonedDateTime` die Methode `of` aufgerufen. Dieser wird anschließend das zugehörige Datum in Form eines `LocalDateTime`-Objektes und die zugehörige Zeitzone in Form eines Objektes von Typ `ZoneId` übergeben. Im Anschluss daran, wird auf dem neuen Objekt mit der Methode `withZoneSameInstant` die Zeitzone von Tokio übergeben. Die Methode gibt eine Kopie des Abflugdatums zurück, jedoch auf Basis der Zeitzone von Tokio. Nun muss lediglich die Flugdauer mit der Methode `plusMinutes` addiert werden und man erhält das korrekte Ankunftsdatum in Tokio.

Dieses simple Beispiel zeigt sehr anschaulich, wie einfach und klar sich Berechnungen über verschiedene Zeitzone hinweg mit der neuen Date/Time-API realisieren lassen.

## Datumsarithmetik

Für Berechnungen, welche über die einfache Addition und Subtraktion von Daten und Uhrzeiten hinausgehen, wird das Package `java.time.temporal` zur Verfügung gestellt. Es bietet viele verschiedene Möglichkeiten komplexere Berechnungen auf Zeiten und Daten auszuführen.

Das Package liefert darüber hinaus weitere Schnittstellen, mit denen man eigene Data-Time-Klassen oder eigene Kalendersysteme implementieren kann.

Für die meisten Berechnungen kann der Enum-Datentyp `ChronoUnit` herangezogen werden. Sie besitzt viele verschiedene Konstanten für verschiedene Anwendungsfälle. Neben den üblichen Konstanten für Tage, Monate, Jahre, Stunden und Sekunden, werden auch Konstanten für spezielle Anwendungsfälle angeboten, wie z.B. Ära, Dekaden und Jahrhunderte. Diese Konstanten besitzen alle eine Vielzahl von Methoden, die sich gut für die Datumsarithmetik eignen. Stellt sich z.B. die Frage, wie lange ein Flug dauert, wenn der Flug um 02:00 Uhr in Berlin startet und am darauf folgenden Tag um 10:00 Uhr Ortszeit in Sydney landet, kann das mit der neuen Date/Time-API sehr einfach berechnet werden.

Dafür werden, wie im vorherigen Beispiel, einfach die passenden Objekte vom Typ `ZonedDateTime` erzeugt und diese Objekte anschließend mit der Methode `between` der Konstante `MINUTES` übergeben. Das `long`-Ergebnis repräsentiert die Minuten, die der Flug tatsächlich gedauert hat (siehe Abbildung 4). Das ist ein sehr einfaches Beispiel, dennoch wird deutlich, dass sich auch komplexere Berechnungen mit der neuen API einfach und sauber abbilden lassen.

## Parsen und Formatieren

Auf die bekannte Klasse `SimpleDateFormat` für das Parsen und Formatieren von Werten wird nun nicht mehr zurückgegriffen. Stattdessen wird die neue Klasse `DateTimeFormatter` verwendet.

Viele Klassen der Date/Time-API stellen dafür die Methoden `parse` und `format` zur Verfügung. Bei `parse` handelt es sich um eine statische Methode die dazu dient, ein Datum bzw. eine Uhrzeit in Form eines `string` in das jeweilige Objekt zu parsen. Neben diesem `string`-Parameter wird der Methode zusätzlich ein Objekt vom Typ `DateTimeFormatter` übergeben, in dem das Format des zu parsenden Datums definiert ist (siehe Abbildung 5).

Die Methode `format` unterstützt den umgekehrten Weg. Es wird lediglich ein Objekt vom Typ `DateTimeFormatter` übergeben und ein `String` in dem übergebenen Format zurückgegeben (siehe Abbildung 6).

Beim `DateTimeFormatter` können sowohl eigene Pattern definiert werden, als auch bereits vorhandene genutzt werden.

## Ist die Zeit reif?

Die Frage, ob die Zeit reif für eine neue Date/Time-API ist, kann nur mit einem klaren „Ja“ beantwortet werden. Der Ausdruck „überreif“ würde vielleicht noch ein wenig

```
// Abflug von Berlin am 01. Januar 2015 um 02:00 Uhr
LocalDateTime abflugDatum = LocalDateTime.of(2015, Month.
JANUARY, 01,02, 00);
ZoneId zoneBerlin = ZoneId.of("Europe/Berlin");
ZonedDateTime abflug = ZonedDateTime.of(abflugDatum, zo-
neBerlin);

// Abflug in Sydney am 0. Januar 2015 um 20:00 Uhr
LocalDateTime ankunftDatum = LocalDateTime.of(2015, Month.
JANUARY, 02,10, 00);
ZoneId zoneSydney = ZoneId.of("Australia/Sydney");
ZonedDateTime ankunft = ZonedDateTime.of(ankunftDatum, zo-
neSydney);

//Berechnung der Flugdauer in Minuten
long minutes = ChronoUnit.MINUTES.between(abflug, ankunft);
System.out.println("Der Flug dauert"+ (minutes/60) +" Stun-
den");

//AUSGABE:
//Der Flug dauert 22 Stunden
```

Abb. 4: Beispiel für die Berechnung der Flugdauer Berlin - Sydney

```
String datumAlsString ="02.12.2015 16:45";
DateTimeFormatter formatter = DateTimeFormatter.ofPattern("dd.
MM.yyyy HH:mm");
LocalDateTime dateTimeObjekt = LocalDateTime.parse(datumAlsString,
formatter);
```

Abb. 5: Konvertieren eines String-Objektes in ein Objekt vom Typ `LocalDateTime`

```
DateTimeFormatter formatter = DateTimeFormatter.ofPattern("dd.
MM.yyyy HH:mm");

LocalDateTime seminarStart = LocalDateTime.of(2015, Month.
FEBRUARY , 2, 10, 00);
System.out.println(seminarStart.format(formatter));
//AUSGABE: 02.02.2015 10:00
System.out.println(seminarStart.format(DateTimeFormatter.
ISO_DATE_TIME));
//AUSGABE: 2015-02-02T10:00:00
```

Abb. 6: Konvertieren eines Objektes vom Typ `LocalDateTime` in einen `String`

## Links

[1] Webseite der Java Community – Übersicht zu JSR 310:  
<https://jcp.org/en/jsr/detail?id=310>

[2] API von Joda-Time:  
<http://www.joda.org/joda-time/>

[3] Java Tutorial zur Date/Time-API von Oracle:  
<http://docs.oracle.com/javase/tutorial/datetime/TOC.html>

## Glossar

### API

Das Application Programmer Interface beschreibt eine Programmierschnittstelle, um die zugehörige Funktionalität nutzen zu können.

### Enum

Enum ist die Kurzform für Enumeration. Hierbei handelt es sich um eine spezielle Klasse in Java, die für Aufzählungen konzipiert wurde.

### ISO 8601

Dieser internationaler Standard beschreibt numerische Datumsformate und Zeitangaben.

### JSR

Ein Java Specification Request ist die Anforderung einer neuen Java-Spezifikation.

### Joda-Time

Joda-Time ist eine umfangreiche Java-Bibliothek für die Datumsarithmetik.

### Mutable

Mutable-Objekte in Java können von außen – meistens durch **set**-Methoden – geändert werden und sind demnach nicht Thread-sicher.

### Package

In Java werden Klassen in sogenannten „Packages“ organisiert. Die Package-Struktur entspricht im weitesten Sinn einer allgemeinen Verzeichnisstruktur.

besser passen, denn leider ist die neue Date/Time-API erst Bestandteil von Java 8. Das Problem daran ist, dass viele Frameworks noch länger mit der Einbindung der API warten werden, da sie häufig auch noch mit älteren Java-Versionen laufen und entsprechend abwärtskompatibel sein müssen.

Daher werden wir leider noch länger mit der alten API arbeiten müssen und sie noch einige Jahre in bestehenden Projekten vorfinden. Deshalb hätte es sicher nicht geschadet, wenn die neue API bereits in älteren Java-Versionen Einzug gehalten hätte.

Damit wir aber in unserem selbstgeschriebenen Code bereits die neuen Objekte einsetzen können, wurden viele Schnittstellen und Methoden bereitgestellt, die den Wechsel von der „alten“ in die „neue“ Welt so einfach wie möglich machen sollen.

## Fazit

Die neue Date/Time-API bietet viele Möglichkeiten, die sich zuvor nur schwer oder ausschließlich durch externe Bibliotheken, wie beispielsweise der Joda-API realisieren ließen.

Wir konnten in diesem Artikel nicht auf alle neuen Funktionalitäten eingehen. Dennoch ist deutlich geworden, wie die neue API funktioniert und welchen Mehrwert man bei der allgemeinen Datumsarithmetik durch sie erfährt. Sollten Sie noch weitergehende Fragen zu den Neuerungen in Java 8 haben, stehen wir Ihnen gerne zur Verfügung.

## Quellen/Bildnachweis

[1] Designstudie „TimeAndMoney“ von Eric Evans:  
<http://timeandmoney.domainlanguage.com/>

[2] Blogbeitrag von Martin Fowler zum Thema „Fluent Interfaces“:  
<http://www.martinfowler.com/bliki/FluentInterface.html>

© flickr | Jannis\_V | Pocket Watch  
© flickr | MIKI Yoshihito | Display Casio



Christian Wiesing  
([info@ordix.de](mailto:info@ordix.de))

Neuerungen in der Oracle Database 12c (Teil VI)

## RMAN & Data Guard – Alles beim Alten?

Im fünften Teil der Reihe zu den Neuerungen in der Oracle Database 12c [5] wurden die Neuerungen im Bereich Data Warehouse vorgestellt. In diesem Artikel geben wir Ihnen nun einen Überblick über die Erweiterungen im Bereich Backup und Recovery sowie in Data Guard.

### SQL-Unterstützung in RMAN

In Oracle 12c ist nun fast der gesamte Sprachumfang von SQL\*Plus im Recovery Manager (RMAN) integriert. Hierdurch entfällt die umständliche Syntax mit dem Präfix `sql` – sowie die aufwendige Quotierung von Hochkommata. Zudem erspart diese Neuerung dem Anwender in vielen Fällen den lästigen Wechsel zwischen RMAN und SQL\*PLUS.

### Neue Rolle SYSBACKUP

Für die Durchführung von DBA-Tätigkeiten steht in Oracle 12c die Rolle `SYSBACKUP` zur Verfügung. Diese Rolle ermöglicht es einem Backup-Administrator alle Tätigkeiten im Bereich Backup und Recovery durchzuführen ohne die Rolle `SYSDBA` zugewiesen zu haben. Da `SYSBACKUP`, ebenso wie `SYSDBA`, im Passwortfile eingetragen wird, ist es dem Benutzer möglich sich auch an heruntergefahrenen Datenbanken anzumelden.

### Table Recovery

Mit Oracle 12c ist es zudem endlich möglich, mit einem Befehl einzelne Tabellen aus einem Backup zurück zu holen. Die Durchführung erfolgt über den neuen Befehl `RECOVER TABLE`. Dieser Befehl führt in Oracle 12c alle Schritte aus, die in früheren Versionen einzeln durchgeführt werden mussten.

Für die Durchführung ist das Aufsetzen einer Auxiliary-Instanz notwendig und daher müssen entsprechende Ressourcen vorhanden sein. Der Aufbau der Auxiliary-Instanz wird über den Parameter `AUXILIARY DESTINATION` gesteuert. Dabei kann auf dreierlei Art vorgefahren werden:

- Backup der Tabelle(n) oder Partition(en) ins Filesystem als Basis für einen Datapump Input
- direktes Einspielen und Überschreiben in die Datenbank
- direktes Einspielen unter einem anderen Namen in der Zieldatenbank

### Cross Plattform Backup & Restore

Der Befehl `CONVERT DATABASE` erzeugt in Oracle 12c ab sofort Backup Sets und keine Image Copies mehr. Das Endian-Format muss beim Plattformwechsel nach wie vor identisch sein.

### Multi Section Image Copies

Mit Oracle 12c können nun auch Image Copies in einzelnen Teilschritten (Section Size) gesichert werden.

### Snapshots (Third Party)

Einige Snapshot-Technologien sind in der Lage, Snapshots einer Oracle 12c Datenbank zu erstellen ohne den Backup-Mode zu verwenden. Fragen Sie daher den Lieferanten Ihrer Snapshot-Technologie, unter welchen Voraussetzungen die Erstellung eines Snapshot ohne den Backup-Mode möglich ist. Diese optimierten Snapshots (ohne Backup-Mode) bieten zudem die Möglichkeit, mit dem Recovery direkt aufsetzen zu können.

### Duplicate Database

Beim Duplizieren von Oracle-Datenbanken mittels `RMAN` kann nun die Option `NOOPEN` verwendet werden. Hierbei wird die Klon-Datenbank nicht mit `OPEN RESETLOGS` geöffnet, sondern in der Mount-Phase belassen. Somit können vor dem Öffnen der Datenbank noch weitere Anpassungen vorgenommen werden.

Bei der Durchführung einer Active Database Duplication liefen in Oracle 11g die Prozesse auf dem Quellsystem und es wurden Image Copies übertragen (Push-Verfahren). In Oracle 12c laufen die Prozesse nun wie beim Duplizieren anhand vorhandener Backups auf dem Zielsystem und es werden nun Backup Sets verwendet (Pull-Verfahren). Hierdurch werden weniger Daten übertragen und es besteht die Möglichkeit, Multisection Backups und Compressed Backup Sets zu nutzen.

## Network Restore

Mit Oracle 12c ist es möglich, **RMAN** in Kombination mit Data Guard zu verwenden. So kann die primäre Datenbank über die Standby-Datenbank zurückgespielt werden. Weiterhin ist es möglich, die Standby-Datenbank über inkrementelle Backups der Produktion nach vorne zu fahren. Die Restore- und Recovery-Befehle werden hierfür um den Zusatz **FROM SERVICE <Quelle>** erweitert.

Schauen wir uns im Folgenden nun die Neuerungen in Data Guard genauer an:

## Einfacher Rollentausch

Der neue Befehl **ALTER DATABASE FAILOVER TO <standby>** stellt sicher, dass die Standby-Datenbank ohne Datenverlust zur Primary-Datenbank wird. Schlägt dies fehl, kann die Standby-Datenbank weiterhin über die alte Syntax aktiviert werden.

## Switchover verify

Neu ist in Oracle 12c die Option **VERIFY** beim manuellen Switchover ohne Data Guard Broker. Wird der Befehl ohne Fehler ausgeführt, ist es möglich, einen Switchover durchzuführen. Es wird also geprüft, ob alle Voraussetzungen für einen Rollenwechsel erfüllt sind bzw. ob es beim Rollenwechsel zu Fehlern kommen kann. Diese Möglichkeit ist auch in dem Broker eingebaut worden.

## Real-Time Apply per Default

In Oracle 11g musste beim Einschalten des Real-Time Apply der Zusatz **USING CURRENT LOGFILE** verwendet werden. In Oracle 12c ist dies nicht mehr nötig, da der Real-Time Apply nun der Default ist. Um das Default-Verhalten von Oracle 11g zu nutzen muss jetzt der Zusatz **USING ARCHIVED LOGFILE** verwendet werden.

## Neue Rolle SYSDBG

Neben der neuen Rolle **SYSBACKUP** wurde auch die Rolle **SYSDBG** mit Oracle 12c eingeführt. Diese Rolle ermöglicht es, Data Guard ohne die Rolle **SYSDBA** zu administrieren.

## Fast Sync

In Oracle 11gR2 musste der Maximum Availability Modus zwingend mit **SYNC + AFFIRM** betrieben werden. Mit Oracle 12c hat man nun die Möglichkeit **SYNC + NOAFFIRM** zu verwenden. Hierzu wird die Eigenschaft **LogXptMode** auf **FASTSYNC** eingestellt. Aufgrund des asynchronen I/O in die Standby Log Files erhält die Primary-Datenbank bei

**SYNC + NOAFFIRM** schneller eine Bestätigung als bei **SYNC + AFFIRM**. Dadurch kann sich die Performance der Primary verbessern. Andererseits wird bei dem Befehl **SYNC + NOAFFIRM** nicht gewährleistet, dass die Redo-Information auf der Standby-Seite physikalisch abgespeichert wurden. Dadurch sinkt die Sicherheit vor Datenverlust.

## Kaskadierende Standby-Datenbanken

Vor Oracle 12c gab es keine Möglichkeit, im Data Guard Broker den Transport zu einer kaskadierenden Datenbank zu administrieren. Hierfür gibt es nun die Eigenschaft **RedoRoutes**. Mit Hilfe dieser Eigenschaft können Regeln für den Redo-Transport definiert und überwacht werden.

## Active Data Guard

Die kostenpflichtige Option Active Data Guard wurde in Oracle 12c um einige Funktionalitäten erweitert. Diese werden im Folgenden vorgestellt:

### Erweiterte Sicherheit

Mit Hilfe der neuen View **V\$RO\_USER\_ACCOUNT** ist es in Oracle 12c möglich, fehlgeschlagene Logins auf der Standby-Seite eines Active Data Guard zu überwachen. Aus Security-Sicht ist dies eine sehr wertvolle Information.

### Temporäre Tabellen

Bisher konnten auf der Standby-Datenbank eines Active Data Guard keine DML-Statements auf temporären Tabellen ausgeführt werden. In Oracle 12c ist dies nun möglich. Um temporäre Tabellen auf der Standby-Datenbank nutzen zu können, müssen diese vorab auf der Primary-Datenbank angelegt werden.

### Sequenzen

Sequenzen können in Oracle 12c sowohl von der Primary- als auch von der Standby-Datenbank genutzt werden. Die Sequenz muss hierfür die Default-Optionen **CACHE** und **NOORDER** verwenden und auf der Primary-Datenbank angelegt werden.

### Session-Sequenzen

Session-Sequenzen können in Oracle 12c sowohl von der Primary- als auch von der Standby-Datenbank genutzt werden. Sie sind speziell für die Verwendung in Kombination mit temporären Tabellen gedacht.



## Far Sync

In Data-Guard-Projekten gibt es oft das Problem, dass ein synchroner Transport der Redo-Informationen aufgrund der schlechten Anbindung und einem hohen Transaktionsaufkommen auf der Primary-Datenbank zu langsam ist, um die Datenbank performant genug zu betreiben, d.h. die hohen Anforderungen an Datensicherheit und Performance sollen gleichzeitig erfüllt werden. Genau für ein solches Umfeld wurde Far Sync entwickelt.

Far Sync basiert auf einer abgespeckten Datenbankinstanz (Far Sync Instance), welche im Gegensatz zur Ziel-Standby-Datenbank gut an die Primary angebinden ist. Diese Instanz erhält die Redo-Information synchron von der Primary und leitet diese asynchron an die Standby-Datenbank weiter. Die Abbildung 1 veranschaulicht das Prinzip von Far Sync.

Mittels Far Sync kann ein Data Guard über eine große Entfernung bzw. über eine langsame Netzwerkverbindung gleichzeitig performant und hochverfügbar realisiert werden. Ziel hierbei ist es, maximale Datensicherheit zu gewährleisten ohne die Performance der Primary-Datenbank zu beeinträchtigen. Far Sync ist eine Funktion der kostenpflichtigen Option Active Data Guard, auch diese Instanz muss somit lizenziert werden.

## Fazit

Die Oracle Version 12c bietet im Bereich RMAN und Data Guard viele kleine und auch sinnvolle Neuerungen. Insbesondere die SQL-Unterstützung im RMAN und der neue Befehl `RECOVER TABLE` stellen nützliche Erweiterungen dar. Die meisten Erweiterungen bei Data Guard sind hingegen Bestandteil der lizenzpflichtigen Option Active Data Guard.



Sven Loer  
([info@ordix.de](mailto:info@ordix.de))

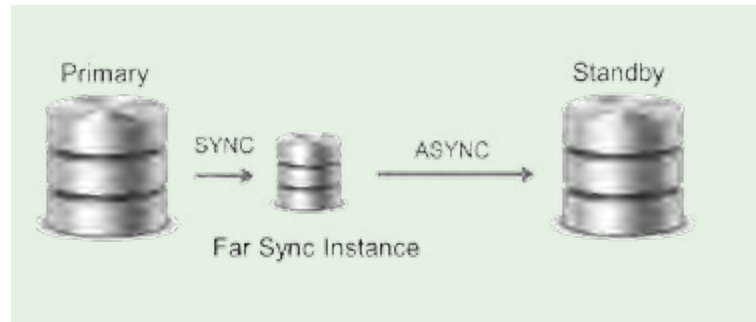


Abb. 1: Ablauf der Redo-Übertragung mittels der Far-Sync-Instanz

## Glossar

### Endian-Format

Das Endian-Format bestimmt die Interpretation von Byte-Reihenfolgen. Die bekanntesten Formate sind Big-Endian und Little-Endian.

### Image Copies

Mit RMAN können neben Backups auch 1:1-Kopien von Datenbankdateien erstellt werden. Diese heißen Image Copies.

## Links

- [1] ORDIX® news Artikel 3/2013  
„Neuerungen in der Oracle Database 12c (Teil I) - Oracle 12c - Eine für alle, alle in einer“:  
<http://www.ordix.de/ordixnews/ordix-news-archiv/3-2013.html>
- [2] ORDIX® news Artikel 4/2013  
„Neuerungen in der Oracle Database 12c (Teil II) - Multitenant-Architektur - eine Datenreise durch Raum und Zeit“:  
<http://www.ordix.de/ordixnews/ordix-news-archiv/4-2013.html>
- [3] ORDIX® news Artikel 1/2014  
„Neuerungen in der Oracle Database 12c (Teil III) - SQL-Neuerungen in Oracle 12c“:  
<http://www.ordix.de/ordixnews/ordix-news-archiv/12014.html>
- [4] ORDIX® news Artikel 2/2014  
„Neuerungen in der Oracle Database 12c (Teil IV) - Welche Veränderungen gibt es im Bereich Tuning?“:  
<http://www.ordix.de/ordixnews/ordix-news-archiv/2-2014.html>
- [5] ORDIX® news Artikel 3/2014  
„Neuerungen in der Oracle Database 12c (Teil V) – Erweiterungen im DWH-Umfeld“:  
<http://www.ordix.de/ordixnews/ordix-news-archiv/3-2014.html>
- [6] Seminarempfehlung:  
Oracle 12c Neuheiten (Seminar-ID: DB-ORA-49)  
<http://training.ordix.de>

# Wer fragt, der führt!

Die Gesprächsführung in Form eines Interviews kennt man zu Genüge aus der journalistischen Berichterstattung. Inzwischen wird es teilweise schon inflationär in den verschiedenen Medien eingesetzt. Auch bei der Bewerberauswahl ist das Interview die übliche Gesprächsform. Der vorliegende Artikel stellt wirkungsvolle Einsatzmöglichkeiten von Interviews im Projektmanagement-Umfeld vor und gibt einige Tipps aus der Praxis.

## Interview

Die Frageform ist ein mächtiges Instrument in der Gesprächsführung. Durch eine gut formulierte Frage kann ich ein Gespräch entscheidend lenken. Das Interview ist die zugehörige Gesprächsform, die sich dies zu Nutze macht. Im Projektmanagement gibt es viele Einsatzbereiche, in denen ein Interview eine sehr hilfreiche Methode darstellt.

Allein die Tatsache, dass man ein Interview führt, erzeugt bei der Person, die interviewt werden soll, das Gefühl wertgeschätzt zu werden, dass ihre Meinung wichtig genommen wird. Sie werden erleben, dass allein dadurch eine hohe Bereitschaft und Offenheit erzeugt wird. Darüber hinaus wird die Person durch das Interview in das Projekt integriert und fühlt sich beteiligt – gemäß dem wichtigen Leitsatz für Projektleiter „aus Betroffene Beteiligte machen“.

## Im Stakeholder Management

Im Stakeholder Management, wenn es darum geht als Projektleiter die Erwartungen der Stakeholder an das Projekt und deren Informationsbedarf an dem Projekt zu ermitteln, ist ein initiales Interview sehr hilfreich. Ich beziehe durch das Interview die Stakeholder von Anfang an in das Projekt ein und schaffe Vertrauen, dass der Projektleiter spätestens dann braucht, wenn das Projekt einmal in Schieflage geraten sollte.

## Bei der Auswahl der Projektmitarbeiter

Ein gut ausgewähltes Projektteam ist eine wichtige Basis für den Projekterfolg. Insofern sollte ein Projektleiter mit jedem potenziellen Projektmitglied im Rahmen der Projektinitialisierung ein Interview führen, um Erwartungen, Anforderungen, notwendige Skills, Verfügbarkeiten etc. zu klären. Er sollte im Zweifelsfall auch einen Kandidaten ablehnen können, wenn dieser nicht zu den Zielen des Projekts oder zu dem Team passt.

## Beim Anforderungsmanagement

Im Bereich des Anforderungsmanagements hat man häufig das Problem, dass die Personen, die das fachliche Wissen haben, nicht ausreichend in der Lage sind ihr Fachwissen in eindeutige und konsistente Anforderungen zu gießen. Ein Interview ist ein gutes Instrument, um Fachwissen abzufragen und daraus Anforderungen zu generieren.

## Im Rahmen eines Projektaudits

Wenn ein Projekt in eine Schieflage gerät und das Vertrauen in den Projekterfolg verloren gegangen ist, wird häufig eine Revision oder ein (externes) Projektaudit durchgeführt. Um sich ein schnelles Bild über den Status quo des Projekts zu machen, sind Einzelinterviews mit den Projektbeteiligten eine bewährte Vorgehensweise. Aus den einzelnen Gesprächen ergibt sich schnell ein umfassendes Bild. Insbesondere Widersprüche zwischen den verschiedenen Aussagen sind häufig wichtige Indizien für die Ursachen der Projektschieflage.

Die oben beschriebenen Einsatzmöglichkeiten für Interviews illustrieren die Wirksamkeit dieses Instruments im Projektmanagement und sollen als Beispiele an dieser Stelle ausreichen. Im Folgenden geben wir Ihnen einige wertvolle Tipps aus der Praxis, wie man erfolgreich Interviews durchführt:

## Vorbereitung

Bereiten Sie die Interviews gut vor, indem Sie sich eine Liste von Themen, die Sie ansprechen wollen, erstellen. Eine Themenliste soll sicherstellen, dass Sie im Rahmen des Interviews keine wichtigen Aspekte vergessen. Nutzen Sie die Themenliste als roten Faden für das Interview. Formulieren Sie die eigentlichen Fragen aber spontan und fragen Sie ggf. nach, wenn aus Ihrer Sicht weitere Details wichtig sind. Vermeiden Sie es, einen Katalog von ausformulierten Fragen zu verwenden, da dieser Sie unnötig in der Gesprächsführung einengt und wie „abgelesen“ und dadurch nicht authentisch wirkt.

### Dialog zu zweit

Ein Interview sollte immer zu zweit durchgeführt werden – mit dem Interviewer und mit der Person, die interviewt wird. Jeder zusätzliche Teilnehmer würde nur die Offenheit im Gespräch unnötig einschränken und die gemachten Aussagen möglicherweise abschwächen. Insofern gibt es viele Parallelen bezüglich der Erfolgsfaktoren für Interviews und zu denen von Feedback-Gesprächen.

### Offene Fragen

Stellen Sie offene Fragen – d.h. vermeiden Sie Fragen, die man mit „Ja“ oder „Nein“ beantworten kann. Nur dadurch öffnen Sie Ihren Gesprächspartner und schaffen eine positive Gesprächsatmosphäre. Darüber hinaus fördern offene Fragen die Beziehung zwischen den Gesprächspartnern. Ihr eigener Gesprächsanteil sollte möglichst klein gehalten werden (5 - 10%).

Sie „führen“ das Gespräch durch gute Fragen bzw. gezieltes Nachfragen – nicht durch einen hohen Gesprächsanteil. Sollte Ihr Gesprächspartner ausschweifend werden, lenken Sie das Gespräch durch eine neue Frage gezielt auf ein anderes Themengebiet.

Auf keinen Fall sollten Sie Suggestivfragen stellen, denn diese verfälschen schnell die Inhalte. Insofern vermeiden Sie auch im Rahmen des Interviews schon voreilige Interpretationen vorzunehmen, denn diese können leicht auf einen Holzweg führen und Ihre „Antennen“ für die eigentlich gemeinten Aussagen falsch ausrichten.

### Echtes Interesse

Zeigen Sie im Interview echtes Interesse an Ihrem Gesprächspartner und an seinen Antworten. Machen Sie deutlich, dass das Interview wichtig für den Erfolg des Projekts ist. Versuchen Sie sich in die Perspektive des Interviewten hineinzuversetzen und zeigen Sie Verständnis im wahrsten Sinne des Wortes. Machen Sie sich Notizen – nicht nur um die Antworten zu dokumentieren, sondern auch um dem Interviewten die Wichtigkeit seiner Antworten deutlich zu machen.

### Keine manipulative Fragetechnik

Vermeiden Sie manipulative Fragetechniken insofern, dass Sie die Antworten der Interviewten in eine bestimmte Richtung lenken – schließlich wollen Sie ehrliche Antworten bekommen. Einfluss sollten Sie in der Gesprächsführung nur dahingehend nehmen, den Interviewten für sich und das Projekt zu gewinnen. Sie wollen schließlich mit jedem Interview zusätzliche Fürsprecher und Unterstützer für sich und Ihr Projekt generieren.

### Resümee

Holen Sie sich am Ende des Interviews ein kurzes Feedback zu dem Gespräch und formulieren Sie selbst ein kurzes Resümee auf Basis ihrer Notizen. Damit runden Sie das Gespräch ab und schaffen für beide Seiten einen positiven Abschluss.

### Fazit

Die Interviewmethode ist ein sehr wirksames Instrument – auch und vor allem im Bereich Projektmanagement. Oft wird kritisiert, dass solche Gespräche viel zu aufwändig wären und dass im Projekt dafür keine Zeit ist. Auf Basis meiner Erfahrungen sind Interviews insbesondere in der Initialisierungsphase eine gute Investition, die sich durch einen anschließend reibungsloseren Projektverlauf mehrfach auszahlt.



*Benedikt Georgi*  
([info@ordix.de](mailto:info@ordix.de))

Das Ende von „`java.lang.OutOfMemoryError: PermGen space`“

# Auflösung der Permanent Generation

Das Ziel von JEP 122 ist es, die Permanent Generation innerhalb der Hotspot JVM (Java Virtual Machine) zu entfernen. Damit entfällt die Notwendigkeit, dessen Größe zu konfigurieren beziehungsweise zu tunen. Schauen wir zunächst, was hinter den Kulissen passiert ist und was dies für den Betrieb der Hotspot JVM bedeutet.

## Daten innerhalb der Permanent Generation

Die Permanent Generation war bislang ein spezieller Speicherbereich innerhalb der Hotspot JVM, ohne ein Bestandteil des Heap zu sein. In der Permanent Generation wurden beispielsweise Class-Objekte der geladenen Class-Dateien, die so genannten Class-Metadaten, interne Strings oder statische Klassenvariablen abgelegt.

## Warum wurde die Permanent Generation entfernt?

Aufgrund des Bestrebens, die Oracle JRockit JVM und die Hotspot JVM zu vereinheitlichen, musste auch die Permanent Generation angegangen werden. Da es in der JRockit JVM keine Permanent Generation gibt, hat man sich letztendlich mit dem JEP 122 dazu entschlossen, diese auch in der Hotspot JVM zu entfernen. Erste Vorbereitungen dazu wurden bereits im JDK 7 getroffen, indem beispielsweise interne Strings und statische Klassenvariablen bereits in den Heap verschoben wurden. Mit dem JDK 8 wurden schließlich auch die restlichen Daten – u.a. die Class-Metadaten – in den nativen JVM-Speicherbereich ausgelagert. Dieser spezielle Teil des nativen Speicherbereichs wird Metaspaces genannt.

## Auswirkungen für den Betrieb

Da es somit in dem JDK 8 keine Permanent Generation mehr gibt, fallen auch die bekannten Konfigurationsparameter `PermSize` und `MaxPermSize` weg. Beim Start einer Java 8 Hotspot JVM wird daher eine entsprechende Meldung ausgegeben, falls diese Parameter dennoch verwendet werden. Im Gegensatz zur Permanent Generation unterliegt die Größe des neu hinzugekommenen Metaspaces keiner direkten Limitierung durch die JVM. Er ist standardmäßig nur durch den physikalischen Speicher begrenzt. Somit unterscheidet er sich also in Abhängigkeit davon, ob eine 32-Bit oder 64-Bit JVM verwendet wird und wie viel virtuellen Speicher das Betriebssystem zur Verfügung

stellt. Allerdings kann der Metaspaces auch mit speziellen Einstellungen versehen werden, um beispielsweise dessen Größe zu limitieren:

- `-XX:MetaspaceSize=<NNN>`
- `-XX:MaxMetaspaceSize=<NNN>`
- `-XX:MinMetaspaceFreeRatio=<NNN>`
- `-XX:MaxMetaspaceFreeRatio=<NNN>`

## Fazit

Als Betreiber der Hotspot JVM darf man sich freuen, denn `java.lang.OutOfMemoryError: PermGen space` gehört der Vergangenheit an. Dafür sollte man nun darauf achten, die Parameter zur Konfiguration der Permanent Generation nicht mehr zu verwenden. Darüber hinaus ist zu prüfen, ob eine Konfiguration des Metaspaces notwendig ist oder sinnvoll sein könnte.



Thomas Rohde  
([info@ordix.de](mailto:info@ordix.de))

## Glossar

### Hotspot JVM

Ist der Name der ursprünglich von Sun Microsystems entwickelten Java Virtual Machine. Sun wurde im Jahre 2010 Teil der Oracle Corporation.

### JEP

Bei dem JDK Enhancement Proposal handelt es sich um ein Verfahren, das die Entwicklung neuer Funktionen für die Java-Sprache und die Java Virtual Machine erlaubt, ohne dass ein vollständiges Spezifizierungsverfahren (JSR) benötigt wird.

### Heap

Der Java-Heap ist der Bereich des Hauptspeichers, welcher der Java-Laufzeitumgebung zur Verfügung steht.

### Metaspaces

Ein Metaspaces ist ein Teil des nativen Speichers, in dem die Class-Metadaten abgelegt werden.

### Permanent Generation

Die Permanent Generation ist der Speicher, in dem die JVM sich selber verwaltet.

### JRockit JVM

Bezeichnet die ursprünglich von Appeal Virtual Machines entwickelte Java Virtual Machine.

## Links

[1] Webseite Open JDK - Übersicht JDK 8:  
<http://openjdk.java.net/projects/jdk8/>

[2] Webseite Open JDK - Übersicht JEP 122:  
<http://openjdk.java.net/jeps/122>



## ORDIX AG spendet an gemeinnützige Organisationen

Statt der üblichen kleinen Geschenke zu Weihnachten unterstützte die ORDIX AG drei karitative Organisationen mit jeweils einer Spende. In diesem Jahr unterstützt das Paderborner Unternehmen den Verein „Sterntaler - Hilfe für schwerkranke Kinder e.V.“ in Paderborn, die „Bärenherz Stiftung“ in Wiesbaden und das Elisabethstift in Berlin ([www.elisabethstift-berlin.de](http://www.elisabethstift-berlin.de)).

### Spendenübergabe in Paderborn an Sterntaler - Hilfe für schwerkranke Kinder e.V.

Ein Sparschwein als Symbol für die großzügige Spende der ORDIX AG in Höhe von 1.800 € überreichte Benedikt Georgi, Vorstandsmitglied der ORDIX AG an Wilhelm Stute vom Verein „Sterntaler – Hilfe für schwerkranke Kinder e.V.“

„Sterntaler e.V.“ unterstützt schwerkranke und langzeit-erkrankter Kinde sowie deren Eltern und Familien auf regionaler Ebene (Ostwestfalen-Lippe). Die Initiative hilft in individuellen Notlagen, wenn Krankenkassen, Sozialämter und andere Organisationen nicht mehr helfen können – und das ehrenamtlich. Seit Gründung von „Sterntaler“ vor mehr als 15 Jahren sind keinerlei Verwaltungskosten angefallen.

[www.sterntaler-kinder.de](http://www.sterntaler-kinder.de)

### Spendenübergabe in Wiesbaden an die Bärenherz Stiftung für schwerkranke Kinder

Auch am Standort der Geschäftsstelle des Seminarzentrums der ORDIX AG in Wiesbaden wurde einer gemeinnützigen Organisation eine Freude gemacht. Eine Spende in Höhe von 1.800 € überreichten die beiden Leiter der Geschäftsstelle Matthias Jung und Uwe Bechthold an die Geschäftsführerin der Bärenherzstiftung Frau Gabrielle Müller.

Die Bärenherz Stiftung fördert und unterstützt Projekte und Einrichtungen für unheilbar kranke Kinder, die eine

begrenzte Lebenserwartung haben. In Anerkennung ihres herausragenden Engagements hat der Hessische Ministerpräsident Volker Bouffier die Bärenherz Stiftung 2014 mit der Georg-August-Zinn-Medaille ausgezeichnet.

[www.baerenherz.de](http://www.baerenherz.de)

### Spendenübergabe in Berlin beim Elisabethstift



Martin Hoermann, ORDIX AG und Helmut Wegener vom Elisabethstift bei der Spendenübergabe.



In-Memory-Option der Oracle Database 12c

## Einfach und schnell mit dem In-Memory Column Store?

In diesem Artikel stellen wir den In-Memory Column Store der Oracle Database 12c vor. Es wird gezeigt, wie einfach das Setup dieses neuen Speicherbereiches ist und welche Optionen es für den Aufbau gibt. Zudem zeigen wir Ihnen, wie überprüft werden kann, ob der In-Memory Column Store bei einer Abfrage auch tatsächlich genutzt wurde.

### Voraussetzung & Ziel

Oracle Database In-Memory ist eine neue Option der Oracle Enterprise Edition 12c und daher kostenpflichtig.

Das neue Speicherformat im In-Memory Column Store ermöglicht ein schnelleres Ausführen von Scans, Joins und Aggregationen als die traditionellen On-Disk-Formate. Gleichzeitig wird die Performance beim Reporting in OLTP- und Data-Warehouse-Umgebungen gesteigert.

Das neue Format ist besonders geeignet für analytische Applikationen, die im Gegensatz zu OLTP-Operationen auf wenigen Spalten ausgeführt werden und viele Zeilen zurückgeben. Analytische Indizes werden durch das In-Memory-Format überflüssig.

### Dual Format

Der In-Memory Column Store ist ein neuer Pool in der SGA (siehe Abbildung 1). Die für diesen Pool ausgewählten Objekte (Tabellen mit häufigem Zugriff, Partitionen, Subpartitionen) werden in ein Spaltenformat konvertiert und nach einem neuen Kompressionsalgorithmus in drei möglichen Stufen komprimiert. Jedes dieser Objekte befindet sich weiterhin in einer 1:1-Beziehung zu einem Objekt auf Platte im Zeilenformat.

Alle Datenformate auf der Disk für permanente Heap-Tabellen und alle Objekttypen wie Partitionen, Materialized Views, Materialized Join Views, Materialized View Logs und Inline LOBs werden durch den In-Memory Column Store unterstützt. Ausgenommen sind lediglich Clustered Tables und Index Organized Tables (IOT).

Die In-Memory-Datenstruktur wird In-Memory Column Unit (IMCU) genannt. Konventionelle Datenbankblöcke sind 8 KB groß. Die IMCUs sind um ein Vielfaches größer. Die genaue Anzahl der Zeilen pro IMCU wird zur Laufzeit festgelegt und basiert auf der Tabellengröße, der Struktur und den Memory-Beschränkungen.

Beim ersten Zugriff auf eine In-Memory-Tabelle durch eine Abfrage oder nach einem Startup der Datenbank werden die IMCUs allokiert. Bei jedem Neustart der Instanz konvertiert die Tabelle vom Zeilen- in das Spaltenformat, da sich die Kopie der Tabelle im Spaltenformat nur im Memory befindet. Nach der Konvertierung steht die In-Memory-Version der Tabelle nach und nach für Abfragen zur Verfügung. Ist eine Tabelle erst teilweise konvertiert, können Abfragen die konvertierten Teile der In-Memory-Version nutzen und den Rest von der Disk oder aus dem Buffer Cache holen, um nicht auf die komplette Konvertierung warten zu müssen.

Der neue Hintergrundprozess IMCO ist eine Art Scheduler, der für das Befüllen und den Refresh der In-Memory Column Units mit Objekten zuständig ist. Um ein Objekt im In-Memory Column Store zu aktualisieren, werden zusätzlich die Hintergrundprozesse SMCO und Wnnn benötigt.

### Setup

Grundsätzlich muss ein dedizierter Cache innerhalb der SGA definiert sein. Die Größe des In-Memory Column Store wird über den statischen Parameter `INMEMORY_SIZE` festgelegt. Über den Parameter `INMEMORY_FORCE` kann die grundsätzliche Nutzung deaktiviert werden (per Default ist die Nutzung aktiviert). Über den Session Parameter `INMEMORY_QUERY` kann anschließend die eigentliche Nutzung innerhalb einer Session einschränkt werden (siehe Abbildung 2).

### Aktivierungen

Das Enablen des Attributes `INMEMORY` ohne Vergabe einer Priorität bedeutet nicht, dass ein Objekt sofort in das Spaltenformat des In-Memory Column Store konvertiert wird. Es wird erst dann konvertiert, wenn eine Abfrage gestellt wird, die sich auf dieses Objekt bezieht.

Es ist auch möglich, das Attribut `INMEMORY` für eine Untermenge zu setzen oder ihr zu entziehen. Eine Tabelle kann mit dem Attribut `INMEMORY` versehen werden, während einzelne Spalten (im Beispiel `c1`) ausgeschlossen werden können.

```
CREATE TABLE BIG_TABLE (...) INMEMORY NO INMEMORY (c1);
ALTER TABLE BIG_TABLE NO INMEMORY (c1);
```

Dies führt jedoch zu einem Fehler, wenn versucht werden sollte, einzelne Spalten in einer `NO INMEMORY`-Tabelle mit dem Attribut `INMEMORY` zu versehen.

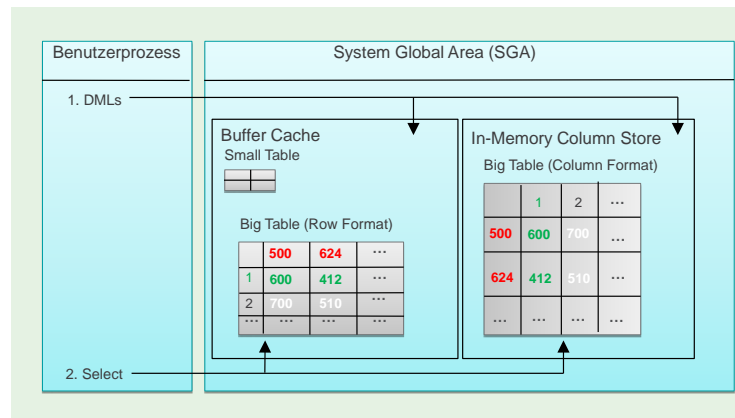


Abb. 1: In-Memory Column Store – Unique Dual Format

Die Größe des In-Memory Column Store wird über folgenden statischen Parameter festgelegt:

```
INMEMORY_SIZE = 100G
```

Die grundsätzliche Aktivierung/Deaktivierung kann über den folgenden Parameter gesetzt werden:

```
INMEMORY_FORCE = OFF | DEFAULT
```

Die Aktivierung in einer Instanz/Session ist wie folgt möglich:

```
ALTER SYSTEM [SESSION] SET INMEMORY_QUERY=ENABLE | DISABLE
```

Abb. 2: Setup des In-Memory Column Store

### Glossar

#### Analytische Applikation

Analytische Applikationen sammeln Daten, transformieren sie in Informationen, fügen intelligente Algorithmen hinzu und unterstützen die Ausführung der nötigen Schritte im System.

#### Analytischer Index

Der analytische Index legt seine Daten in Form eines Stern-Schemas ab. Die Merkmale eines analytischen Index liegen in den Dimensionen, welche verschiedene Attribute tragen können.

#### Bloom-Filter

Ein Bloom-Filter ist eine Datenstruktur, die Mitgliedschaftsabfragen unterstützt - wie z.B. die Frage, ob ein Element Mitglied einer gegebenen Menge ist oder nicht.

#### Clustered Table

Daten verschiedener Tabellen werden anhand eines Schlüssels (Cluster Key) gemeinsam in denselben Datenblöcken gespeichert.

#### Database Buffer Cache

Der Database Buffer Cache ist die Shared-Memory-Struktur der SGA. Sie puffert die Datenblöcke der Daten-Files.

#### Heap-Tabelle

Eine Tabelle ist standardmäßig als Heap-Tabelle gespeichert. Darin sind Daten in einer ungeordneten Reihenfolge abgelegt.

#### Index Organized Table (IOT)

Diese Tabellenstruktur ist intern wie ein B\*-Tree-Index organisiert. Die Daten werden in der Reihenfolge des Primärschlüssels abgelegt.

Die View `V$IM_COLUMN_LEVEL` zeigt, welche Spalten einer In-Memory-Tabelle (nicht) im In-Memory Column Store liegen.

## Glossar (Fortsetzung)

### Inline LOB

Die Large Objects werden zusammen mit den anderen Spalten der Tabelle in demselben Block gespeichert.

### Materialized View

Materialized Views sind Views, in denen Abfrageresultate gespeichert werden. Sie werden häufig genutzt, um umfangreiche Abfragen mit Aggregaten und Berechnungen zu beschleunigen und um Daten zu replizieren.

### OLAP-Würfel

Je nach Anforderungen und technischer Umsetzung gibt es unterschiedliche OLAP-Spielarten. Daten werden dabei nicht in zweidimensionalen Tabellen mit Spalten und Zeilen gespeichert, sondern in mehrdimensionalen Würfeln (Cubes).

### Online Analytical Processing (OLAP)

OLAP-Anwendungen setzen speziell abfrageorientiert organisierte Daten voraus. Im Gegensatz zu OLTP-Systemen hat das einfache, performante Lesen oberste Priorität und nicht die Manipulierbarkeit der Daten. Daten werden nie oder nur in großen, meist zeitbezogenen Tranchen gelöscht. Redundanzen werden gezielt in Kauf genommen. Die Daten werden regelmäßig und nur durch dedizierte Prozesse hinzugefügt. Aggregierte Ergebnisse werden vorausgerechnet und separat gespeichert.

### Online Transactional Processing (OLTP)

OLTP-Anwendungen werden im Gegensatz zum Batch-Betrieb durch kleinere Online-Transaktionen charakterisiert.

### Partitionen

Eine Partition ist die Unterteilung einer großen Tabelle in physisch und logisch kleinere Teile nach einem Partitionierungskriterium (Schlüsselkriterium).

### System Global Area (SGA)

Die System Global Area ist eine Gruppe von Shared-Memory-Strukturen im Hauptspeicher, die Daten sowie Steuerinformationen einer Oracle-Instanz enthalten.

### Star-Schema

Das Star-Schema ist eine besondere Form eines Datenmodells, dessen Ziel nicht die Normalisierung ist, sondern eine Optimierung auf effiziente Leseoperationen. Hauptanwendungsfeld sind Data Warehouse und OLAP-Anwendungen. Die Bezeichnung rührt daher, dass die Tabellen sternförmig angeordnet werden. Im Zentrum steht eine Faktentabelle, um die sich mehrere Dimensionstabellen gruppieren.

```
select display_name, value
  from v$mystat m, v$statname n
  where n.STATISTIC# = m.STATISTIC#
  and display_name in ('session logical reads - IM',
                     'IM scan rows',
                     'IM scan rows valid',
                     'IM scan blocks cache',
                     'IM scan CUs columns accessed');
```

Abb. 3: Monitoring

## Prioritäten

Des Weiteren können Prioritäten für einzelne Objekte vergeben werden, die in das Spaltenformat des In-Memory Column Store konvertiert werden sollen. Durch diese Prioritäten wird der In-Memory Column Store unterschiedlich verzögert befüllt. Hier gibt es vier mögliche Prioritäten: **LOW**, **MEDIUM**, **HIGH** und **CRITICAL**. Das Attribut **PRIORITY** wird beim Statement **CREATE TABLE** oder **ALTER TABLE** mitgegeben, wie das folgende Beispiel zeigt:

```
CREATE [ALTER] TABLE BIG_TABLE (...) INMEMORY
PRIORITY MEDIUM;
```

Die View `V$IM_SEGMENTS` zeigt, ob die Objekte für den In-Memory Column Store vollständig konvertiert wurden.

Falls für ein Objekt keine Priorität vergeben wurde, so wird es erst nach einer Abfrage oder einem Instanz Startup in das Spaltenformat des In-Memory Column Store konvertiert. Gibt es Objekte mit höherer Priorität, die den Platz beanspruchen, so landet ein Objekt wegen niedrigerer Priorität und einem Mangel an Platz eventuell gar nicht oder nur zum Teil im In-Memory Column Store.

## Komprimierung

Durch die Klausel **MEMCOMPRESS** wird ein möglicher Kompressionslevel für ein Objekt im In-Memory Column Store definiert. Folgende Kompressionslevel können gesetzt werden:

- **NO MEMCOMPRESS**  
Es wird nichts komprimiert.
- **MEMCOMPRESS FOR QUERY [LOW|HIGH]**  
Es wird ein Kompressionsalgorithmus genutzt, der zu Gunsten von besserer Performance weniger Platz spart. **FOR QUERY LOW** ist der Default, wenn nur **MEMCOMPRESS** definiert wurde.
- **MEMCOMPRESS FOR CAPACITY LOW**  
Der Zusatz **FOR CAPACITY** nutzt einen Algorithmus, der eine Balance zwischen hoher Kompression und guter Abfrage-Performance sucht. Der Zusatz **LOW** kann auch weggelassen werden.
- **MEMCOMPRESS FOR CAPACITY HIGH**  
Dieser Algorithmus komprimiert noch stärker als die Variante **LOW**.
- **MEMCOMPRESS FOR DML**  
Die Kompression ist minimal und optimiert für die **DML** Performance.

Die View `V$IM_SEGMENTS` zeigt die Kompressionsrate.

## Monitoring

Die Session-Statistiken und der Ausführungsplan einer Abfrage zeigen, ob diese gegen den In-Memory Column Store oder den Buffer Cache ausgeführt wurde (Abbildung 3).



## Was macht die Abfrage so schnell?

Die spaltenbezogene Speicherung hat bei den typischerweise in einem Data Warehouse verwendeten Abfragen auf Wertebereiche und mit wenigen abgefragten Spalten Vorteile gegenüber der satzbezogenen Speicherung. Die satzbezogene Speicherung ist dafür bei OLTP-typischen Einzelsatzabfragen und vielen Spalten vorteilhaft.

Hinzu kommt der höhere Komprimierungsfaktor, sodass geringere Datenmengen aus dem In-Memory-Speicher zu lesen sind. Die Filterprüfung (**WHERE**-Klausel und **JOIN**-Prüfungen) wendet das System direkt auf die komprimierten Daten an, ohne sie zuvor zu dekomprimieren. Oracle nutzt hierfür ein neu entwickeltes Komprimierungsverfahren, bei dem die Daten im komprimierten Zustand interpretiert werden können.

Die **JOIN**-Bedingungen löst das System mittels Bloom-Filtering. Kleine Tabellen, wie die meisten Dimensionstabellen in einem Star-Schema, wandelt das System in einen Bloom-Filter (eine Art Hash-Tabelle) um und sendet sie direkt zur größeren Faktentabelle. Schon bei dem Zugriff auf die Faktentabelle greift der **JOIN**-Filter.

Verfügt das System über mehrere CPU-Kerne, so erhält jeder Kern ein eigenes Bloom-Filter-Objekt und arbeitet direkt die ihm zugeordneten Faktentabellensätze ab.

Aggregationen (**GROUP BY**) löst das System mit so genannten „Report Outlines“. Ein einmal berechnetes Aggregationsergebnis merkt sich das System dynamisch für nachfolgende Abfragen. Praktisch entsteht „On-The-Fly“ ein OLAP-Würfelim Speicher.

## Sinnvolle Anwendungen

Durch die In-Memory-Funktionalität sind Performance-Steigerungen bei bestimmten Abfragen zu erwarten. Zu diesen Abfragen gehören große Scans unter Verwendung von Prädikatfiltern auf eine einzelne Spalten (**=**, **<>**, **<**, **<=**, **>**, **>=**, **IN**, **NULL**, **NOT NULL**, **LIKE**, **SUBSTRING**). Effektiv ist das neue Spaltenformat aber auch bei einer Abfrage weniger Spalten auf eine Tabelle mit vielen Zeilen. Weitere sinnvolle Anwendungen sind ein **JOIN** einer kleinen Tabelle gegen eine große und die im letzten Abschnitt genannten Aggregationsergebnisse.

## Links

[1] Oracle Database 12c Release 12.1.0.2 Documentation:  
<http://www.oracle.com/technetwork/database/in-memory/documentation/index.html>

## Bildnachweis

© istockphoto.com | tiero | file cabinet and folder  
 © psdgraphics.com | computer database psd icon

## Fazit

Die In-Memory-Funktionalität lässt sich schnell und auf den ersten Blick auch sehr einfach umsetzen, da keine Änderungen an bisherigen Anwendungen notwendig sind. Für einzelne Abfragen sind deutliche Performance-Steigerungen möglich. Es stellt sich jedoch die Frage, welche Objekte Kandidaten für diesen Speicherbereich sind.

Auch bei einer Eingrenzung auf die genannten sinnvollen Anwendungen ist der verfügbare Hauptspeicher der limitierende Faktor. Passen nicht alle ausgewählten Kandidaten in den neuen Speicherbereich, so beeinflussen die zu vergebenen Attribute die Kompression und Priorität, ob ein Objekt vollkommen oder nur zum Teil ins Spaltenformat konvertiert wird. Statistiken und Ausführungspläne sollten daher überprüft und der neue Speicherbereich gepflegt werden.



Klaus Reimers  
 (info@ordix.de)

# Was ist neu unter VMware?

In einem vorangegangenen Artikel der ORDIX® news [1] haben wir bereits über die unterschiedlichen Lizenzierungsmodelle für Oracle-Datenbanken berichtet. Im letzten halben Jahr sind weitere wichtige Neuerungen hinzugekommen, die wir Ihnen hier zusammenfassen möchten.

## Virtualisierung

Viele Kunden haben in den letzten Jahren ihre Systeme unter Verwendung von VMware virtualisiert. Oracle definiert VMware grundsätzlich als eine „softpartitioning Virtualization“. Dies bedeutet, dass alle physikalischen Server in einem VMware Cluster in die Berechnung der Kerne (engl. Cores) einbezogen werden mussten und nicht nur die in der dedizierten virtuellen Maschine zugewiesenen Prozessoren. Allerdings ist bei der Entscheidung für eine Oracle-Edition immer der einzelne physikalische Server des VMware Cluster zu betrachten – z.B. SE-One maximal 2 Sockets pro physikalischen Server.

## VMware vSphere

Mit dem Produkt VMware vSphere können virtuelle Server innerhalb eines VMware Cluster verschoben werden. Hat man beispielsweise einen VMware Cluster bestehend aus 6 physikalischen Servern mit je 2 Sockets und jeweils 8 Cores, so müssen auch zwölf SE-One-Processorlizenzen erworben werden.

## VMware vCenter

In einem vCenter werden mehrere VMware Cluster zusammengefasst. Ab der VMware vSphere Version 5.1 können virtuelle Maschinen über vCenter-Grenzen hinweg verschoben werden. Daher muss das komplette vCenter mit allen VMware Clustern bei Oracle lizenziert werden, wenn auch nur eine Oracle-Installation auf einer virtuellen Maschine des vCenter betrieben wird! Leider hat Oracle die detaillierten Lizenzbedingungen bezüglich VMware nirgends genau schriftlich hinterlegt.

## Repository-Datenbanken

Viele Kunden verwenden eine Catalog-Datenbank zur höheren Sicherheit und besseren Verwaltung der Metadaten der RMAN-Sicherungen. Cloud Control benötigt eine Repository-Datenbank, in der die gesammelten Daten zentral abgelegt werden. Diese Datenbanken waren bisher immer lizenzfrei. Daher haben viele Kunden diese Datenbanken unter VMware aufgesetzt. Auch hier gibt es nach neuen Erkenntnissen eine Änderung seitens Oracle in der Auslegung der Lizenzbedingungen.

Sollte eine dieser Datenbanken unter VMware laufen, so müssen ab sofort alle Server im VMware Cluster – außer dem bisher kostenfreien physikalischen Server – lizenziert werden. Ab VMware vSphere Version 5.1 muss sogar das gesamte vCenter mit allen VMware Clustern betrachtet werden.

## Glossar

### VMware vSphere

Virtualisierungslösung von VMware, welche eine virtuelle Maschine für x86-64-basierte Hardware und deren Betriebssysteme bereitstellt.

### VMware Cluster

Hierbei bilden mehrere physikalische Server einen Cluster.

### VMware vCenter

Das VMware vCenter fasst mehrerer VMware Cluster zusammen und ermöglicht einen vollen Grad von Inventarisierung und Automatisierung.

## Links/Quellen

[1] ORDIX® news Artikel 1/2014  
„Oracle-Editionen – Muss es immer Enterprise sein?“:  
<http://www.ordix.de/ordixnews/ordix-news-archiv/12014.html>

[2] Artikel der DOAG zum Thema „Lizensierung von Oracle-Produkten unter VMware vSphere ab Version 5.1 anders gehandhabt“:  
<http://www.doag.org/home/aktuelle-news/article/oracle-aendert-lizensierung-von-oracle-produkten-unter-vmware-vsphere-ab-version-5.1.html>

[3] Blog von Michael Paege  
„Kostenfreie Repositories auf VMware“:  
<https://mpaage.wordpress.com/>

Installationen einer Catalog-Datenbank und der Cloud Control Repository-Datenbank auf einzelnen physikalischen Servern sind nach wie vor lizenzfrei.

### Fazit

Die Unsicherheit und Unzufriedenheit der Oracle-Kunden im Hinblick auf die Lizenzregelungen wird immer größer. Die negativen Überraschungen kommen schließlich bei den Audits, die Oracle immer häufiger bei den Kunden durchführt.

Wir empfehlen daher ab VMware vSphere Version 5.1 ein eigenes vCenter für die Oracle-Installationen. VMware hat in der Version 6.0 die Möglichkeit angekündigt, virtuelle Maschinen sogar über vCenter-Grenzen hinweg zu verschieben. Man kann sich ausmalen was dies für die Oracle-Lizensierung bedeutet.

Gerne unterstützen wir Sie beim Review und beim eventuell daraus resultierenden Redesign Ihrer Oracle-Datenbanklandschaft.

### Hinweis

Die in diesem Artikel vorgestellten Aussagen und Empfehlungen wurden nach bestem Wissen recherchiert und zusammengestellt, erfolgen aber ohne Gewähr.



*Klaus Reimers*  
(info@ordix.de)



*Michael Skowasch*  
(info@ordix.de)

- ORDIX IT-Check -

## Oracle-Lizenzberatung

# Money, Money, Money ...

Ohne spezielles Wissen um die Lizenzmodelle ist es nur schwer möglich, Oracle-Produkte kostengünstig und zugleich lizenzkonform einzusetzen. Zu komplex sind die Fragestellungen bei der Virtualisierung, Auswahl der Oracle-Edition, Verwendung von Optionen und Management-Packs. Zu schnell können Sie bei diesem Thema in eine Lizenzfalle tappen. Oft haben wir bei unseren Kunden eine unnötige Überlizenzierung festgestellt. Hier steckt enormes Einsparpotenzial! Wir kennen die Fallstricke und haben für Ihre Anforderungen die optimale Lösungsmöglichkeit.

Unsere objektive Oracle-Lizenzberatung hilft Ihnen bei der Entscheidungsfindung.

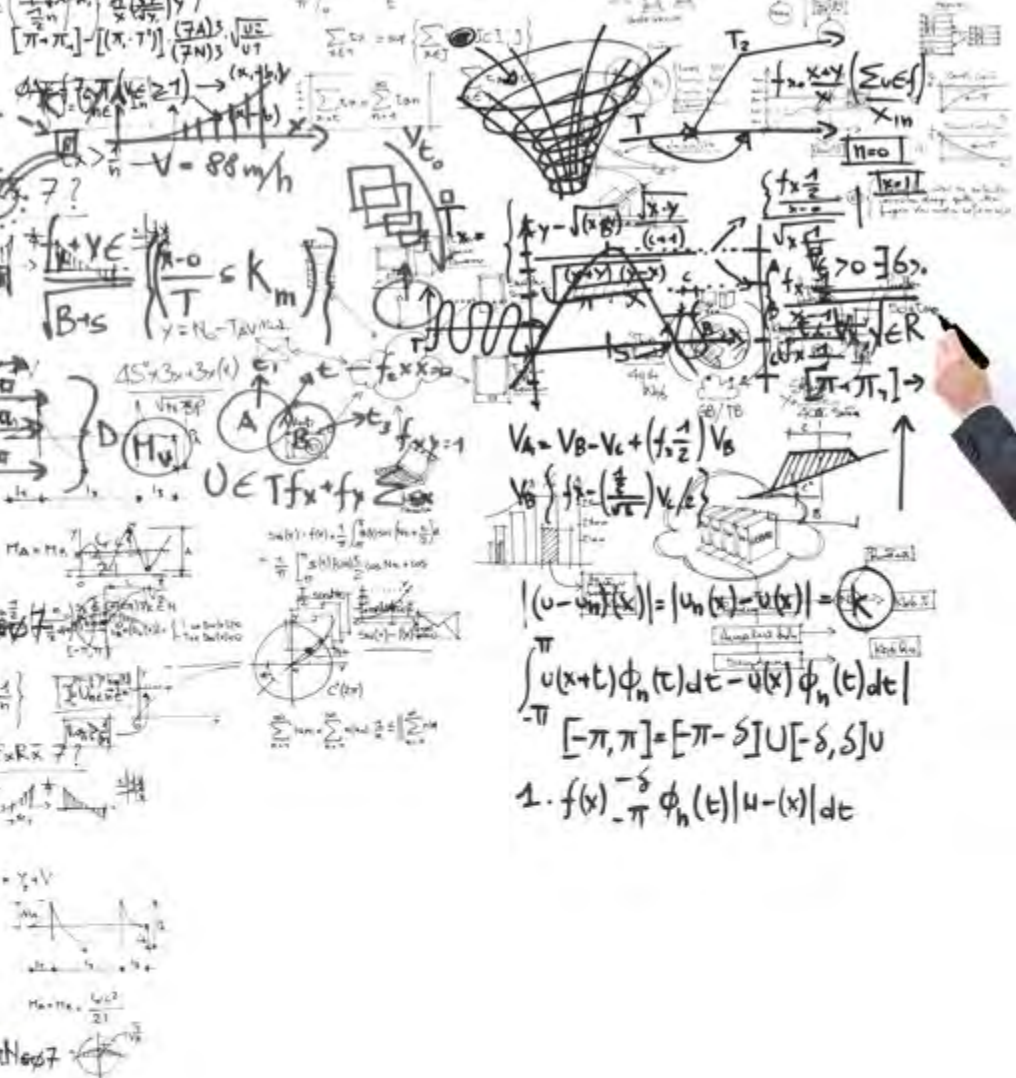
## Unser IT-Check bietet viele Vorteile ...

... aus kaufmännischer Sicht:

- Überschaubarer Aufwand mit großem Nutzen für Ihr Unternehmen
- Große Einsparpotenziale und kostengünstigere Strukturen
- Lizenzkonformes Arbeiten
- Kalkulierbare Kosten auf Basis eines festen Tagessatzes (960 €/Personentag)

... aus technischer Sicht:

- Unabhängige und objektive Analyse Ihrer Oracle-Lizenzumgebung
- Über 20 Jahre Erfahrung mit dem Thema Oracle-Lizenzen
- Erfahrung aus vielen Kunden- und Migrationsprojekten
- Unsere Lizenzberater sind auch technisch sehr kompetente Oracle-Spezialisten



# BIG DATA

INFORMATIONEN NEU GELEBT

Wir bieten Ihnen mit diesem Seminar einen Überblick über Technologien und Konzepte, die unter dem Stichwort Big Data zusammengefasst werden. Sie werden mit den Möglichkeiten und Einsatzgebieten von Big Data Tools und Techniken vertraut gemacht.

## UNSER SEMINAR FÜR SIE:

### Seminarinhalte (Auszug)

- Big Data - wo geht die Reise hin?
- Wie gliedert sich Big Data in die IT-Architektur des Unternehmens ein?
- Welche typischen Anforderungen können adressiert werden? Wo ist Big Data eher ungeeignet?
- Vorstellung eines ausgewählten Produktspektrums von IBM, Oracle und Microsoft

### Termine

26.05.2015  
10.08.2015  
19.10.2015  
07.12.2015

### Seminar ID

DB-BIG-01

### Seminarpreis

590,00 € / Frühbucherpreis: 531,00 €