

Red Stack

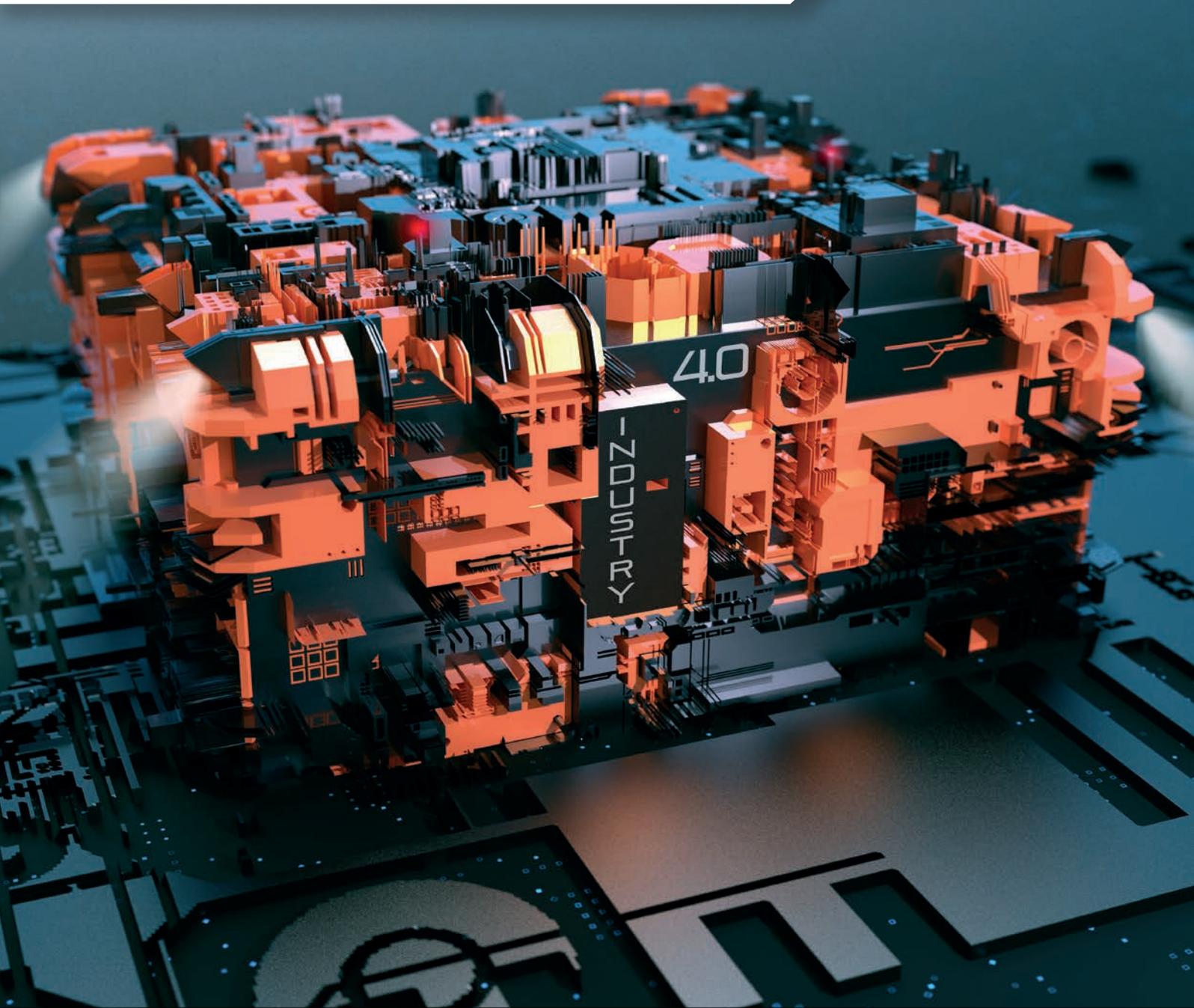
Magazin

DOAG

SOUG
swiss oracle
user group

AOUG
AUSTRIAN ORACLE USER GROUP

Jetzt inklusive BUSINESS NEWS



Aus der Praxis

CI/CD: Theorie meets Praxis

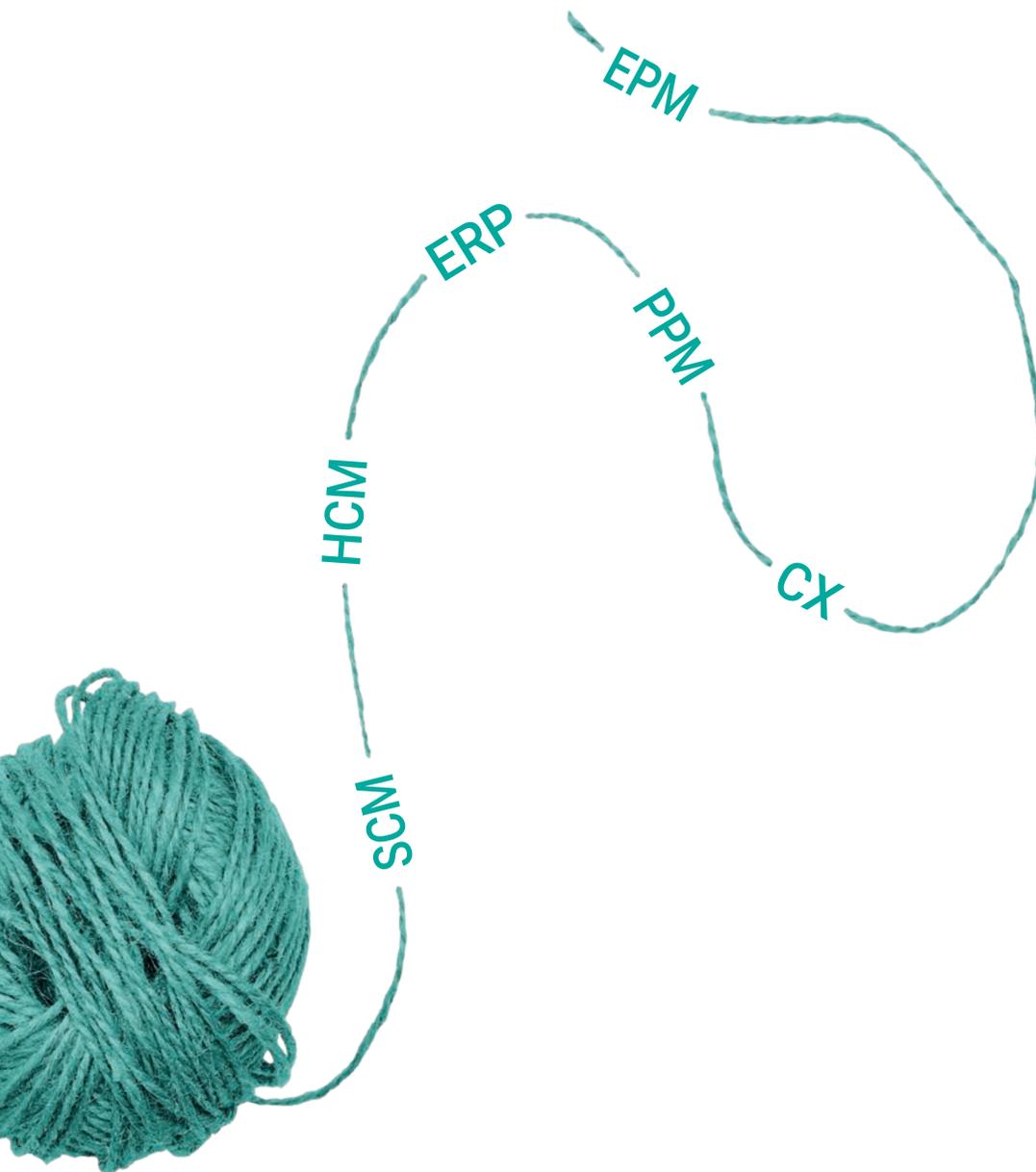


Im Interview

Markus Peppler, Direktor/Abteilungsleiter für Business Analytics bei Universal Investment

BUSINESS NEWS

Smarte Unternehmensapplikationen für Industrie 4.0



Der grüne Faden für Ihre Digitale Evolution

Wir bei PROMATIS folgen einem selbst entwickelten grünen Faden:

Mit professioneller Beratung und innovativen Digitalisierungslösungen schaffen wir exzellente Geschäftsprozesse: agil, bedarfsgerecht, intelligent und zukunftssicher. Nachhaltige Qualität und Wirtschaftlichkeit sichern wir durch kontinuierliche Verbesserung der eingesetzten Verfahren, Produkte und Services.

Mit unserer Digitalisierungskompetenz und unseren Best Practice-Lösungen begleiten wir Sie auf Ihrer Reise in die Oracle Cloud.



Christian Trieb
Vorstand Datenbank,
Leitung Datenbank
Cosmmunity

Liebe Mitglieder, liebe Leserinnen und Leser,

DevOps ist zurzeit ein Thema in der IT, an dem keiner vorbeikommt. Auch für Nutzer von Oracle-Produkten, seien es DBAs oder Entwickler, stellt man sich diesem Thema.

In dieser Ausgabe wird das Thema aus unterschiedlichen Perspektiven betrachtet. So wird das Thema Sicherheit auch in DevOps-Strukturen eine immer wichtigere Bedeutung finden. Auch die Verwendung von Open-Source-Produkten wird in diesem Bereich eine immer stärkere Beachtung finden. Insbesondere Docker findet hier im Datenbankumfeld großes Interesse. Gerade das Zusammenspiel zwischen diesen unterschiedlichen Produkten und der Oracle-Datenbank wird in Zukunft ein sehr spannendes Aufgabengebiet für einen vielseitigen Oracle-DBA werden können. Dabei spielt es meines Erachtens keine Rolle, ob die Software in der Cloud läuft oder im eigenen Rechenzentrum. Die Aufgaben eines DBA unterliegen schon immer einem stetigen Wandel. Insoweit ist und bleibt auch DevOps für DBAs, aber auch für Entwickler, ein sehr spannendes Thema. In diesem Sinne wünsche ich Ihnen einen regen Erkenntnisgewinn beim Lesen der Artikel dieser Ausgabe.

Die im Red Stack Magazin enthaltene Business News beschäftigt sich diesmal mit dem Zukunftsthema „Smarte Unternehmensapplikationen für die Industrie 4.0“. Die interessanten Artikel betrachten das Thema aus unterschiedlichen Perspektiven. Hierbei wird die Verbindung zwischen den Geschäftsvorfällen und der aktuellen Technik gut beschrieben.

Während ich dies hier schreibe, werden täglich neue Meldungen zur Corona-Krise veröffentlicht. Die Prioritäten haben sich inzwischen stark verschoben. In der Hoffnung, dass zum Zeitpunkt ihrer Lektüre das Schlimmste überstanden ist:

Bleiben Sie gesund.

Christian Trieb



Training

Training

MUNIQSOFT
TRAINING

ORACLE
Silver Partner

20 Jahre Oracle-Datenbankschulungen von
Experten Effizient und kundenorientiert!

**Aus aktuellem Anlass können Sie ab sofort ONLINE
an all unseren Schulungen teilnehmen.
Im Livestream verfolgen Sie die gewünschten
Kurse von zu Hause oder Ihrem Büro aus.**

PL /SQL Grundlagen	06.07.-10.07.2020	€ 2.190.- netto
Neuerungen 12c-19c	06.07.-10.07.2020	€ 2.190.- netto
DBA Sommerspecial	20.07.-24.07.2020	€ 1.490.- netto
Postgres DBA	27.07.-29.07.2020	€ 1.690.- netto

☎ 089 679090-40

Website: www.munisoft-training.de

Tipps: www.munisoft-training.de/tipps

Schulungszentrum

Munisoft Training GmbH
Grünwalder Weg 13a
82008 Unterhaching/München

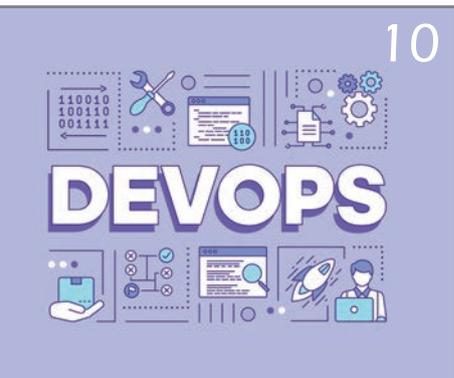
Mehr Oracle Schulungstermine unter
munisoft-training.de

Auf Anfrage bieten wir auch gerne individuelle Inhouse Schulung und Consultingleistungen an!



7

Interview mit Markus Pepler



10

Neben den technischen Aspekten spielen das Mindset der Akteure, deren Arbeitsweise und Ziele jedoch eine weit bedeutendere Rolle.



57

Mithilfe der OCI lassen sich recht einfach und schnell DR-Lösungen in der Cloud realisieren.



73

Die Digitalisierung macht es möglich: Smarte Applikationen steuern zunehmend Anlagen und Prozesse in der Industrie 4.0

Einleitung

- 3 Editorial
- 5 Timeline
- 6 Aus der Ferne betrachtet: Auf was kommt es bei Dev/Ops an?
- 7 „Der primäre Treiber der Umsetzung der definierten Vision sind dann vor allem die Menschen, erst danach kommen die zu definierenden Prozesse und anschließend die Technologie.“
Interview mit Markus Pepler

DevOps

- 10 Culture First – Relevanz der DevOps-Kultur und ein Ansatz für eine nachhaltige Einführung
Carsten Wiesbaum & Kevin Steinhagen
- 15 CI/CD: Theorie meets Praxis
Dennis Hoffmann
- 18 Monitoring mit Prometheus
Markus Bente & Michael Mühlbeyer
- 24 DevOps in Open Source
Wolfgang Nast
- 29 DevSecOps aber sicher! Kontinuierlich sicherer entwickeln
Frank Pientka

Datenbank

- 34 Cyber-Defense mit und in der Oracle Cloud
Michael Fischer
- 40 Oracle Database Indexing Best Practices – Teil 1
Randolf Geist
- 44 Systematisches Oracle Performance Tuning mit kostenlosen 3rd Party Tools
Stefan Koehler
- 50 Analyse von ortsbezogenen Daten in der Oracle DB
Karin Patenge

Infrastruktur

- 57 Hybrid Data Guard mit der OCI
Thomas Rein
- 63 Benchmarking Oracle Cloud Services
Manfred Drozd

Development

- 67 Schnittstellenbündelung mit GraphQL
Philipp Hartenfeller

Aktuelles

- 70 Benachteiligung von Bestandskunden beim Support
DOAG-Redaktion
- 71 Björn Bröhl ist neuer Vorstandsvorsitzender der DOAG
DOAG-Redaktion

Smarte Unternehmensapplikationen für Industrie 4.0

BUSINESS NEWS

- 73 Smarte Technologien in der Social Networked Industry
Dietmar Ebel & Lorenz Kiebler
- 78 Die Bedeutung von Plattformen im IoT-Umfeld
Norbert Gronau & Benedict Bender
- 81 "Vorpreschen und neue Dinge ausprobieren."
Christian Luda sprach mit Dr. Matthias Peissner
- 84 Smarte Anlagenüberwachung für die Industrie 4.0
Daniel Silter
- 88 Vom Sensor zum Prozess
Falk Wolsky

Intern

- 93 Termine + Neue Mitglieder
- 94 Impressum + Inserenten

Timeline

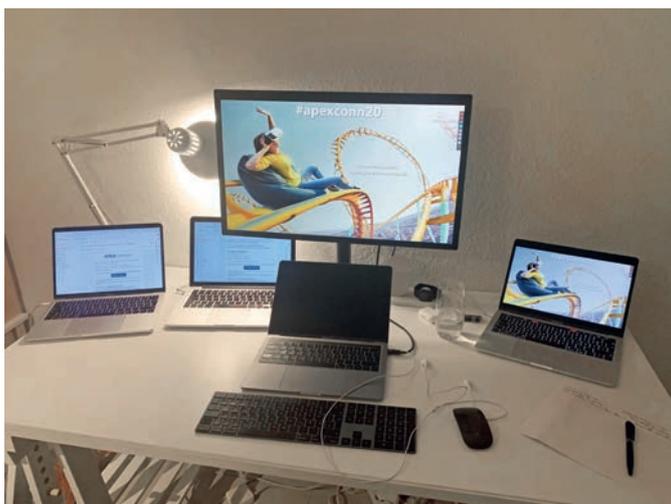
5. Mai 2020

Aufgrund der aktuellen Situation finden die Vorstandssitzungen derzeit nur virtuell statt. Der Vorstand trifft sich an jedem ersten Dienstag im Monat in einem virtuellen Meetingraum. Ziel der Sitzung ist, die neuen Vorstandsmitglieder über die derzeitige Arbeit der Communities zu informieren. Ein weiterer Schwerpunkt ist die Bildung von Programmkomitees für die DOAG 2020 Konferenz + Ausstellung im November sowie für die CloudLand 2021. Aufgrund der großen Unsicherheit, wie die Konferenzen in der Corona-Krise dann stattfinden können, hat der Vorstand eine Strategie eines zweigleisigen Vorgehens beschlossen. Dabei soll die Teilnahme an den Veranstaltungen sowohl vor Ort als auch virtuell stattfinden. Je nach Situation könnten dann zum Beispiel 1.000 Teilnehmer vor Ort in Nürnberg teilnehmen und weitere 1.000 virtuell im Netz. Dabei sollen die beiden Teilnehmergruppen auch miteinander interagieren können. Hierfür werden von der DOAG-Geschäftsstelle technische Konzepte erarbeitet. Im Ergebnis soll die Konferenz für jeden Teilnehmer, ob vor Ort oder virtuell, ein Erlebnis werden.

5. und 6. Mai 2020

Die APEX connect lebt weiter! Am 5. und 6. Mai 2020 findet virtuell über das Internet die APEX connect 2020 [Online] mit reduziertem Programm für 400 Teilnehmer statt – und dies kostenfrei. Aufgrund der Corona-Krise ist dies ein willkommener Ersatz für die nicht stattfindende Veranstaltung im Phantasialand.

Dank zahlreicher hochkarätiger Referentinnen und Referenten, die sich bereit erklärten, ihr Wissen auch online mit den Teilnehmern zu teilen, bietet die Veranstaltung rund 35 Veranstaltungen zu einem breiten Themenspektrum. Die Vortragsunterlagen stehen anschließend wie gewohnt auch als Download zur Verfügung.



8. Mai 2020

Das DOAG Datenbank Webinar mit Christian Ballweg zum Thema "Migration Factory: Oracle to AWS RDS for PostgreSQL" findet online statt. Es beleuchtet exemplarisch die Migration einer Oracle-Datenbank zum Relational Database Service (RDS) von AWS.

19. Mai 2020

An jedem ersten Dienstag im Monat findet eine Telefonkonferenz statt, in der das Programmkomitee Themen für die DOAG 2020 Konferenz + Ausstellung im November sucht und mit der Referentenakquise beginnt.

25. Mai bis 19. Juni 2020

Die DOAG 2020 Datenbank findet aufgrund behördlicher Verordnungen der Stadt Düsseldorf am 25. und 26. Mai leider nicht statt. Ein virtuelles und kostenfreies Angebot der Datenbank Community wird daher als Ersatz ins Leben gerufen.

An jedem Werktag werden zwei Remotevorträge des Programms der DOAG 2020 Datenbank angeboten. Die Teilnehmer nehmen daran regen Anteil.

Das Programm ist unterteilt in die Themenschwerpunkte „Einsteiger“ (25. bis 29. Mai), „Datenbank“ (2. bis 5. Juni sowie 15. bis 19. Juni) und „Engineered Systems“ (8. bis 10. Juni). Die Keynote "The Importance of Failing" hält Frits Hoogland.

2. Juni 2020

Die DOAG-Vorstandssitzung 2/20 findet online statt. Thema ist die IT-Strategie, deren Externalisierung und Neuausrichtung. Ein wesentlicher Aspekt ist, aufgrund der zunehmenden Komplexität, mehr externe Experten einzubinden und auf standardisierte Lösungen zu setzen.

12. Juni 2020

Das DOAG Datenbank-Webinar "SAP on Oracle Development Update" mit Dr. Christian Graf (SAP SE) steht an. Der Vortrag beleuchtet die aus SAP-Sicht bestehenden Möglichkeiten zum Betrieb von SAP-Systemen auf Basis von Oracle-Datenbanken, so wie sie heute bereits bestehen bzw. mittel- bis langfristig aufgesetzt und betrieben werden können. Hierbei wird sowohl auf die hauseigenen Cloud-Angebote als auch auf die vielen verschiedenen Kombinationen von Applikations-, Datenbank- und Betriebssystem-Layer oder die Nutzung von Engineered Systems eingegangen.

24. Juni 2020

Das DOAG Development Webinar „Test-Driven PL/SQL: Ein pragmatischer Einstieg mit utPLSQL“ steht auf der Agenda.

Aus der Ferne betrachtet: Auf was kommt es bei Dev/Ops an?



*Dr. Dietmar Neugebauer
Ehemaliger DOAG-Vorstandsvorsitzender*

Als mir Gedanken zu dieser Frage durch den Kopf gingen, ist mir in Erinnerung gekommen, welche Organisationsformen ich in meiner Berufslaufbahn schon miterlebt habe, obwohl ich eigentlich immer mit derselben Oracle-Datenbank gearbeitet habe. Angefangen habe ich in der zentralen IT und dort Projekte für die Fachabteilungen gemacht, dann wurden bei der nächsten Umorganisation beide Abteilungen zur Anwendungs-IT zusammengelegt. Irgendwann kam die Organisationsform Plan/Build/Run, da wurde sogar die Datenbankgruppe in Build und Run aufgeteilt. Zum Schluss kann ich mich noch an die Aufteilung Infrastruktur, Middleware und Development erinnern.

Und wenn ich es jetzt aus der Ferne betrachte, wurde bei jeder Neuorganisation alles aufgewirbelt, man hat eventuell einen neuen Chef bekommen, nach einer gewissen Zeit hat sich wieder alles eingerüttelt und man hat wieder mit denselben Leuten zusammengearbeitet wie vorher. Denn das Entscheidende war ja doch, wie effektiv, kompetent und vertrauensvoll die einzelnen Personen zusammenarbeiten können, um in den Projekten und in den täglichen Problemstellungen erfolgreich zu sein.

Aber ist die Frage nach Dev/Ops überhaupt zu Zeiten der Corona-Krise eine so wichtige Frage?

Beim Schreiben dieses Artikels habe ich einen Bericht über eine junge Lehrerin in der Süddeutschen Zeitung gelesen, die eine Idee hatte. Sie fragte die Buchhandlung in ihrer Straße, ob sie helfen könnte und eine Monatsmiete übernehmen dürfte. Mit großer Freude willigte die Buchhändlerin ein und die junge Frau bastelte gleich noch eine Webseite, um andere auch dazu zu

motivieren und um „Suche Unterstützung“ und „Biete Unterstützung“ zusammenzubringen.

Wie sich die Corona-Krise auf die DOAG auswirkt, ist wohl noch nicht abzusehen. Aber auch hier wird sicherlich das Engagement jedes Einzelnen gefragt sein. Sowohl im DOAG Office als auch bei jedem einzelnen DOAG-Mitglied. Überall wird es Einschnitte geben, dabei ist die DOAG sicherlich gut beraten, hier für jeden im Einzelfall der jeweiligen Situation angepasste Lösungen zu finden.

Andererseits setzt die DOAG dabei natürlich auch auf die Treue ihrer Mitglieder. Weiter zeigt das zunehmende Angebot von Wissensvermittlung im Internet auch, dass die DOAG hier in Zukunft ihre Online-Präsenz noch wesentlich verstärken muss.

Bei all dem sollten wir die Gemeinschaft der europäischen Usergruppen nicht vergessen. Gerade hier können wir zeigen, dass wir gemeinsam unsere erfolgreiche Zusammenarbeit aus den letzten Jahren auch in schwierigen Zeiten fortsetzen können.

Lasst uns die Fragen und Probleme, die jeden von uns betreffen, gemeinsam lösen. Gemeinsam sollten wir danach streben, die Orte unseres persönlichen Netzwerkes in den Regionalgruppen und auf der DOAG Konferenz so engagiert und schnell wie möglich wieder zu erreichen. Wie war der Spruch von Lukas Podolski aus dem Jahre 2006: „Doppelpass alleine? Vergiss es!“

... Gilt sicherlich auch für Dev/Ops

... Die Webseite heißt übrigens www.helfer-in-der-krise.de



„Der primäre Treiber der Umsetzung der definierten Vision sind dann vor allem die Menschen, erst danach kommen die zu definierenden Prozesse und anschließend die Technologie.“

Oliver Lemm, DOAG-Themenverantwortlicher DevOps, sprach mit Markus Peppler, Direktor/Abteilungsleiter für Business Analytics bei Universal Investment, über den Einsatz von DevOps in Unternehmen.

Was sind Ihre Aufgaben bei Universal-Investment?

Die Entwicklungs-Abteilung, die ich leite, nennt sich Business Analytics & Business Applications Development.

Wir sind für die komplette Business Intelligence & Analytics Infrastruktur in der Universal-Investment Gruppe, deren Weiterentwicklung und Instandhaltung zuständig. Sie umfasst eine zentrale Oracle- Datenbank sowie zahlreiche Satelliten wie SAP, das CRM (Microsoft Dynamics), unzählige Anwendungsdatenbanken, Sharepoints u. v. m. Alles verbunden durch anwendungsbezogene, aber auch eigenentwickelte ETL-Prozesse, Datenbanklinks und APIs.

Darüber hinaus entwickeln wir auf Basis von Oracle APEX für alle Unternehmensbereiche Business- Applikationen. Aktuell

betreuen wir 50 Anwendungen mit knapp 1000 internen und externen Anwendern.

Für zwei Teams agiere ich als Scrum Master, bin Projektleiter eines Cloud-Projektes, Business Relationship Manager für die Bereiche Portfoliomanagement & Immobilien sowie IT Innovation Ambassador.

In welchem Rahmen sind Sie zum ersten Mal mit dem Thema DevOps konfrontiert worden?

Im Jahr 2016 haben wir unsere Produktentwicklung auf das agile Vorgehensmodell Scrum umgestellt. Dieses basiert auf Werten und Prinzipien, stellt aber keinesfalls eine präskripti-

ve Methode dar, die vorschreibt, wie der Prozess aussehen soll.

In der Agile Community wurde das Thema DevOps so ab 2017 mehr und mehr als fester Bestandteil integriert und auf Events und in den entsprechenden Community-Blogs nahezu gehypt. Da wir uns beginnend 2018 auch mit Cloud-Themen beschäftigt haben, sind wir endgültig am Thema DevOps nicht mehr vorbeigekommen und haben begonnen, uns darüber Gedanken zu machen, wie wir davon profitieren könnten.

Scrum konzentriert sich stark auf das, was während des Sprints geschieht. Scrum sagt dir nicht, wie der Prozess innerhalb des Sprints aussehen soll. Außerdem wird lediglich die Entwicklungsphase des Application Lifecycle betrachtet.

Wir sind in der Finanzbranche sehr stark reguliert und haben hohe Dokumentationspflichten an unsere Entwicklungsprozesse. D. h. wir müssen auch für das „Wie“ im Entwicklungsprozess eine Antwort liefern und nachweisen, dass wir uns an diese Vorgaben halten.

Außerdem sind wir in unserer Einheit nicht nur für die reine Entwicklung zuständig, sondern bedienen auch andere Phasen im Application Lifecycle Management (ALM), u. a. das Requirements Engineering oder die Business-Analyse. Zuletzt sind wir auch für die Maintenance und den Betrieb der Applikationen verantwortlich.

Wir haben bei uns im Haus das Glück der kurzen Wege und arbeiten auf dem gleichen Flur wie unsere System- und Datenbank-Administratoren. Die Kollegen haben sich parallel mit un-



Zur Person: Markus Peppler

Markus Peppler ist Abteilungsleiter der Universal IT Services GmbH, dem Full-Service-IT-Provider der Universal Investment GmbH, der größten unabhängigen Investmentgesellschaft im deutschsprachigen Raum. Er ist Diplom-Betriebswirt und Certified EFFAS Financial Analyst. Nach Stationen in der Versicherungsbranche und im Asset Management wechselte er vor 10 Jahren von der Business- auf die IT-Seite und verantwortet dort verschiedene Entwicklungsteams, die Business Analytics-Lösungen, Quantitative Pricing-Modelle sowie Business-Applikationen auf Basis von Oracle Application Express (APEX) entwickeln und betreiben.

Er ist zertifizierter Scrum Master & Product Owner sowie EXIN DevOps Master und begleitet und unterstützt als IT Innovation Ambassador und Initiator der #UIDIGIDAYS aktiv die digitale Transformation der Universal Investment GmbH.

seren Bemühungen hin zu einem agilen Entwicklungsprozess sehr stark mit dem Thema „Infrastructure as Code“ sowie CI / CD beschäftigt. Mit den gemeinsamen Erfahrungen haben wir uns 2019 in ersten Pilotprojekten zusammen dem Thema DevOps gewidmet, sind experimentell vorgegangen, haben uns gleichzeitig durch Teilnahme an Events und Schulungen theoretisch schlaugemacht und Zertifizierungen durchlaufen.

Was sind die Schlüsseltreiber, um DevOps erfolgreich einzusetzen?

Der allererste Schritt sollte sein, zu definieren, was DevOps in der eigenen Organisation eigentlich bedeutet und wie die Vision aussieht, die man zu erreichen versucht.

Noch vor der Einführung von Technologie und der Änderung bestehender Prozesse müssen ein einheitliches Verständnis darüber bestehen und auch klare Grenzen zwischen DevOps und den bestehenden strukturierteren Ansätzen für das IT-Service-Management gezogen werden.

Der primäre Treiber der Umsetzung der definierten Vision sind dann vor allem die Menschen, erst danach kommen die zu definierenden Prozesse und anschließend die Technologie.

Bei der Zusammenstellung eines DevOps-Teams sollten Organisationen Rollen und Verantwortlichkeiten definieren, um eine größere Effizienz zu erreichen.

Der DevOps-Prozess muss auch von Anfang an durch definierte KPIs begleitet werden. Durch diese kann man bewerten, ob die Initiativen erfolgreich sind und ob es Änderungsbedarf gibt. Diese orientieren sich an vereinbarten Zielen, die auf die DevOps-Vision einzahlen und auf dem Weg dorthin als Wegweiser dienen sollen.

Sehr wichtig ist es auch, sich regelmäßig mit Entwicklern und Systemadministratoren, die nicht zum Team gehören, zu treffen und die Ergebnisse objektiv zu bewerten.

Welchen kulturellen Problemen sind Sie zum Thema DevOps begegnet?

Eine der Herausforderungen für Organisationen, die sich auf DevOps hinbewegen wollen, ist der Kampf der Kulturen zwischen Dev und Ops. Meiner Ansicht nach ist diese Kluft nicht anders als andere, die wir in den meisten Organisationen finden, wie z. B. Business vs. IT, agil vs. Wasserfall, On-prem vs. Cloud oder make vs. buy.

In der täglichen Praxis gibt es allerdings keine echten, unüberwindbaren Reibungspunkte. Entwicklern ist es – entgegen der Mär – nicht egal, ob ihre Software später auch wirklich läuft, und dass Admins sich nicht für neue Technologien interessieren und Verständnis für Entwicklungsprozesse haben, ist auch schlicht falsch.

Die Beteiligten sind es lediglich nicht gewohnt, sich über diese Themen auszutauschen. Hier liefert meines Erachtens DevOps auch die Lösung für diese kulturellen Hürden gleich mit. DevOps fördert eine Denkweise der Kommunikation, Zusammenarbeit, Integration und Automatisierung zwischen Softwareentwicklern und IT-Abteilungen, sodass es mehr als nur ein neuer Satz von Werkzeugen ist: Es ist ein kultureller Wandel. Diese Veränderung geschieht nicht sofort oder automatisch, sondern erfordert bewusste Schritte und Maßnahmen sowie kalkulierte Entscheidungen.

Worauf man bei der Einführung aber sehr diszipliniert achten muss, ist, dass sich – gerade die technikverliebten Entwickler – nicht ausschließlich auf die neuen Tools und technischen Möglich-

keiten stürzen. Getreu dem Motto „a fool with tool is still a fool“ gilt es, bei allen Aktivitäten die nötigen organisatorischen und kulturellen Änderungen und vor allem die Vision hinter der DevOps-Organisation stets im Fokus zu behalten.

Wie kann man Vertrauen für das Thema DevOps erzeugen?

Der Wandel erfordert vor allem Engagement vom Management. Von Anfang an muss man aus dem Management heraus ein klares Commitment abgeben und auch immerwährend deutlich den Beteiligten zeigen, dass man selbst mit Eifer bei der Sache ist und das Team auf die Unterstützung durch das Management bauen kann.

Die agilen Prinzipien und Werte wie „right to fail“, Autonomie, Selbstorganisation, Initiative, Subsidiaritätsprinzip, Delegation dürfen keine Lippenbekenntnisse bleiben, sondern müssen von allen Beteiligten als gelebte Praxis empfunden werden.

„Feiern“ ist eine weitere Möglichkeit, eine Verbindung der Teammitglieder untereinander und das Vertrauen in das Vorhaben herzustellen. Das Feiern gerade anfänglicher, auch kleiner Erfolge sowie die Wertschätzung des Beitrages am Erfolg jedes einzelnen.

Dabei ist es wichtig, daran zu denken, als Gesamt-Organisation zu feiern, nicht als einzelne Teams. Die Vorstellung, dass die DevOps-Organisation als Einheit und nicht als einzelne Teams Erfolg hat und scheitert, muss Teil der DNA werden.

Dabei muss es nicht immer ein großes Event sein, schon kleine Aufmerksamkeiten, ein selbstgebackener Kuchen, eine Intranet-Meldung zu errungenen Meilensteinen oder ein Kegelnachdem erfolgreichen Rollout eines neuen Releases.

Was umfasst das Thema DevOps bei Ihnen alles?

Wie erwähnt ist die DevOps-Initiative bei uns aus den Erfahrungen und dem Ansatz der agilen Software-Entwicklung heraus entstanden. Gestartet sind wir zunächst mit dem Product Management, der Planung der Software nach Scrum und der Generierung von Features aus den Requirements, die wir im Zuge von Sprints in Form von User Stories entwickelt und als Releases zur Verfügung gestellt haben.

Nach und nach haben wir diesen Prozess erweitert und neben der kontinuierlichen Planung auch automatische Testverfahren, Continuous Integration und zuletzt auch Continuous Delivery aufgenommen. Die Ausprägungen und der Einsatz dieser Elemente kommt bei uns (noch) nicht flächendeckend in allen Projekten zum Tragen, da dies zum einen mit einigen Alt-systemen mit sehr hohem Migrationsaufwand verbunden wäre und wir andererseits bei diversen Applikationen nur wenige Releases im Jahr benötigen, sodass der Mehraufwand für den vollumfänglichen Einsatz des Prozesses nicht zu rechtfertigen ist.

Unsere Vision ist es dennoch, im Falle der sehr dynamischen Applikationen auch in Zukunft noch einen Schritt weiter zu gehen und den Prozess bis hin zum Continuous Deployment auszuweiten, um den Applikations-Lebenszyklus bis auf wenige Tage verkürzen zu können. Dafür sind aber noch einige Hausaufgaben, gerade im Umfeld der automatischen Tests und Quality Assurance, zu erledigen.

Welche Tools setzen Sie im Bereich DevOps ein?

Zentrales Tool unserer DevOps-Prozesse sind die Azure DevOps Services und das darin integrierte Git Repository. Der Mi-

crosoft-Cloud-Dienst Azure DevOps Services bietet Entwicklern Unterstützung von Teams bei der Arbeitsplanung, der Zusammenarbeit bei der Code-Entwicklung und der Erstellung und Bereitstellung von Anwendungen.

Darüber hinaus setzen wir weitere Tools und Frameworks wie beispielsweise Docker, Kubernetes, PowerShell, Selenium, usw. ein. Diese sind voll in Azure DevOps Services integrierbar und kommen in den CI/CD Pipelines für unterschiedliche Aufgaben im Build, Deployment und Test-Prozess zum Einsatz.

Setzen Sie bereits DevOps in Kombination mit Cloud-Technologien ein?

Wir nutzen zum Verteilen von Open-API-basierten MicroServices die Azure Container Services, welche z. B. Python-basierte Algorithmen in der Cloud zur Verfügung stellen.

Unsere BI-Infrastruktur haben wir kürzlich um den Cloud Services Power BI sowie die Azure Analysis Services ergänzt und unsere aktuelle Data Analytics & Data Science-Plattform ist Databricks.

All diese Cloud Services sind voll integriert in unsere DevOps-Umgebung und durchlaufen hier sowohl Planung, Entwicklung, Source Code-Verwaltung als auch CI- und CD-Prozesse.

Wir setzen insgesamt auf einen hybriden Ansatz aus On-Premises- und Cloud-Modulen und wenden uns verstärkt auch Cloud-gehosteten Infrastrukturen zu.

Welche Veränderungen erwarten Sie in den nächsten Jahren im Bereich DevOps?

Durch die steigende Anzahl an Cloud-Projekten und die Migration ganzer IT-Infrastrukturen in die Cloud wird das Thema DevOps noch stärker in den Fokus geraten. Denn DevOps und Cloud sind eng miteinander verwoben und verfolgen die gleichen Ziele und Strategien wie beispielsweise schnellere Bereitstellung von Anwendungen, um die Bedürfnisse der Geschäftseinheiten schneller zu erfüllen, Benutzeranforderungen, die schnell wieder in die Software einfließen, und niedrigere Kosten für Entwicklung, Tests, Bereitstellung und Betrieb.

Mit der zunehmenden Verbreitung von Cloud-nativen Ansätzen rücken Container Images und Registries sowie die passenden Werkzeuge wie Kubernetes, Docker, Artifactory etc. stark in den Fokus der Entwicklungsprozesse und werden mehr und mehr zentraler Bestandteil der DevOps-Prozesse.

Open-Source erhält immer mehr Aufmerksamkeit, da die Vorteile und die Flexibilität, die es den Entwicklern bringt, immer mehr Beachtung finden. Gerade im DevOps-Umfeld gewinnt das Thema auch in großen Unternehmen stark an Bedeutung, so auch bei uns.

Durch die zunehmende Automatisierung, die schnelleren Produktzyklen und nicht zuletzt das Outsourcen von ganzen Prozessen in die Cloud tritt das Thema Security auch zunehmend in den Mittelpunkt und Sicherheit hat im Entwicklungslebenszyklus mehr denn je Priorität. Sicherheit wird zur Verantwortung aller und nicht nur der Sicherheitsexperten.

Nicht zuletzt bieten die Automatisierung, Cloud Computing sowie der DevOps-Prozess auch die Möglichkeit, neue Technologien in den Entwicklungsprozess zu integrieren und mithilfe von AI-Tests, Fehlerbehebungen und Problemerkennung durch Machine Learning zu unterstützen und damit den Lebenszyklus der Produkte bei höherer Qualität zu verkürzen.

DEVOPS



Culture First – Relevanz der DevOps-Kultur und ein Ansatz für eine nachhaltige Einführung

Carsten Wiesbaum, esentri
Kevin Steinhagen, esentri

Der folgende Artikel gibt eine Einführung in das Thema DevOps mit besonderem Schwerpunkt auf Zusammenarbeit und Kultur. Zunächst wird die Notwendigkeit eines Umdenkens im Bereich der Softwareentwicklung sowie die daraus resultierende agile Arbeitsweise und DevOps-Kultur aufgezeigt. Anschließend werden die Ausprägungen einer solchen Kultur vorgestellt und Parallelen zu agilen Werten gezogen. Schließlich wird die Frage „Wie kann eine DevOps-Transformation im Unternehmen gelingen?“ erörtert.

Das Thema DevOps ist nicht erst seit gestern in aller Munde; wer noch nie etwas davon gehört hat, versteckte sich wohl die letzten Jahre gekonnt hinter dem Mond. Im Allgemeinen versteht man darunter das Aufbrechen der Grenzen zwischen Softwareentwicklung und Softwarebetrieb. Das eigentliche Ziel ist, beide Gruppen von Spezialisten zusammenzuführen, ein gemeinsames Verständnis für die Arbeit des anderen zu schaffen und auf eine gemeinsame Vision hinzuarbeiten. Leider verkümmert dieses anspruchsvolle Ziel allzu oft zu einer reinen Tool- und Technik-Initiative. Die Herausforderungen der Zusammenarbeit werden häufig vernachlässigt. Neben den technischen Aspekten spielen das Mindset der Akteure, deren Arbeitsweise und Ziele jedoch eine weit bedeutendere Rolle. „Wir müssen unbedingt DevOps einführen, das machen jetzt alle!“ (siehe Abbildung 1), „Welche Tools und Lizenzen brauchen wir dafür?“ So oder so ähnlich könnte es in den Chefetagen einiger Softwareschmieden und Unternehmen mit größeren IT-Abteilungen in den letzten Jahren abgelaufen sein.

Die Verschmelzung von Softwareentwicklung – „Dev“ – und Softwarebetrieb – „Ops“ – zu cross-funktionalen Teams mit gemeinsamen Zielen und Werten ist dabei alles andere als ein Selbstläufer. Mit der Umbenennung von Berufsbezeichnungen, dem Einstellen von Lead DevOps Engineers und dem Zusammenwerfen von Teams stellen sich in den seltensten Fällen nachhaltige Verbesserungen und Erfolge ein. Warum sollte man sich dann überhaupt mit dem Thema beschäftigen?

Darum geht es nicht ohne DevOps!

Wer nach wie vor nach dem Wasserfallmodell, dem Rational Unified Process oder Tayloristischen Ansätzen Software entwickelt, wird schnell an die Grenzen der Wirtschaftlichkeit stoßen. Die rasanten Anforderungsänderungen am Markt lassen lange, unflexible Entwicklungsprozesse nicht mehr zu. Was vor wenigen Jahren, wenn nicht Monaten, noch der neueste Schrei war, ist inzwischen veraltet oder – schlimmer – bereits zum Sicherheitsrisiko geworden. Zudem erfordern die stetig steigende Komplexität und Erwartungshaltung von Kunden in Bezug auf Umfang, Sicherheit und Qualität



Abbildung 1: Einführung DevOps (© Anna Klara Cominelli, esentri AG)

neue Wege und Denkweisen. Besonders treffend lässt sich die aktuelle Marktsituation mit dem von Warren Bennis und Burt Nanus geprägten Akronym VUCA, kurz für volatile, uncertain, complex, ambiguous, beschreiben. Als Antwort auf die genannten Herausforderungen führte man iterative Entwicklungsprozesse mit agilen Frameworks wie SCRUM, Nexus oder SAFE ein, die schneller zu potenziell einsatzfähigen Software-Versionen führten. Zudem lassen sich unscharfe und häufig ändernde Anforderungen viel effektiver mit agiler Arbeitsweise abfedern als durch starre, fest vorgeschriebene Arbeitsschritte. Produktionsreife Inkremente werden nun in Wochen fertiggestellt und nicht mehr in Monaten oder Jahren. Beim professionellen Bereitstellen von Infrastruktur für die Stück für Stück wachsende Software muss der Betrieb die entscheidende Rolle einnehmen. Die Kompe-

tenzen beider Bereiche sind nun schon von Beginn eines Software-Projektes gefragt. Es liegt nahe, dass Abteilungen mit hoher Korrelation möglichst eng zusammenarbeiten sollten, um das bestmögliche Ergebnis zu erzielen. Dabei stellt sich die Frage, warum diese Zweckverbindung nicht ohne weiteres Zutun gelingen kann.

Darum brauchen wir eine DevOps-Kultur!

John Willis, Vorreiter und Visionär im Bereich DevOps, antwortete in einem Interview auf die Frage, was wohl die größte Hürde im Zusammenhang mit DevOps sei, mit: Kultur, Kultur, Kultur [1]. Viele Menschen stehen Veränderungen in ihrem Arbeitsalltag oft sehr kritisch gegenüber, gerade wenn diese von der Führungsebe-

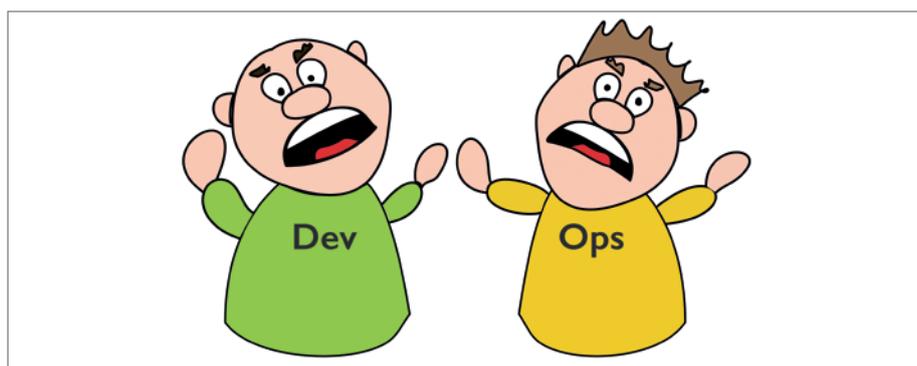


Abbildung 2: Zielkonflikt Development und Operations (© Anna Klara Cominelli, esentri AG)

ne ohne klärende Kommunikation diktiert werden. Es fehlt dann häufig der systemische Zusammenhang oder das Verständnis, Prozesse zu überdenken, alteingesessenes über Bord zu werfen und Neues zu akzeptieren. Im Fall von DevOps spielen weitere Faktoren eine wichtige Rolle. Da die Bereiche Entwicklung und Betrieb über lange Zeiträume hinweg koexistiert haben, wuchsen gedankliche Mauern an den Abteilungsgrenzen. Bei Fehlern ist schnell der Schuldige außerhalb des eigenen Verantwortungsbereichs gefunden, der Fokus auf Ursache und Lösung auftretender Probleme kann dabei schnell verloren gehen. Das sogenannte Silo-Denken lässt sich auf konkurrierende Ziele der Bereiche zurückführen. Während Entwickler stetig neue Funktionalitäten liefern, also Systeme ändern, verfolgt der Betrieb den Erhalt von Stabilität und Sicherheit, ein klassischer Zielkonflikt, den es zu lösen gilt (siehe *Abbildung 2*). Mit Unterstützung der Führungsebene können gemeinsame Ziele eingeführt und Anpassungen der Organisationsstruktur vorgenommen werden. Die Bereitschaft zur Zusammenarbeit und eine positive innere Haltung gegenüber der neuen DevOps-Struktur kann allerdings nur von innen, aus den Köpfen der Mitarbeiter heraus entstehen. Wie kann also eine gemeinsame Arbeitskultur aussehen?

So kann eine DevOps-Kultur aussehen!

Grundlage für eine neue Kultur sind gemeinsame Werte, geteilte Verantwortung und eine gemeinschaftliche Vision der Arbeit, ganz nach dem Motto: „You build it, you run it!“ Es dürfen, anders als bei der Wortkreation DevOps, nicht nur Teile von Development und Operations gestrichen und anschließend die Überreste zusammengefügt werden. Die Akzeptanz für die

jeweils fremde Domäne ist grundlegend und muss in einem beidseitigen Verständnis der Aufgabengebiete münden. Findet ein regelmäßiger Wissenstransfer mit konstruktiver Kritik und kontinuierlichen Feedbackrunden statt, steigt auch die Bereitschaft, die eigenen Fähigkeiten auszubauen und neue Aufgaben zu übernehmen. Während des Kulturwandels verschwimmen die klassischen Entwickler- und Betriebsrollen immer weiter, bis schließlich crossfunktionale Teams etabliert sind. Von ganz allein werden sich diese Vorgänge allerdings nicht einstellen, es bedarf Coaching, um Konfliktpotenziale aufzulösen, alte Denkmuster aufzubrechen und neue Arbeitsweisen zu identifizieren und zu etablieren. Die wichtigsten Werte und Grundprinzipien einer DevOps-Kultur vereinen Kernpunkte bestehender Konzepte wie dem Agilen Manifest, Lean Management und Kaizen. Die Werte Selbstverpflichtung, Mut, Offenheit, Fokus und Respekt werden durch das Prinzip der Vermeidung überflüssiger Tätigkeiten und das Streben nach kontinuierlicher Verbesserung ergänzt. Wie kann eine DevOps-Transformation im Unternehmen gelingen?

So gelingt die Einführung einer DevOps-Transformation im Unternehmen!

Die Umstellung auf eine DevOps-Kultur bedeutet für Firmen, Abteilungen und jeden Mitarbeiter einschneidende Veränderungen im Arbeitsalltag. Generell weiß jeder von sich selbst: Veränderungen fühlen sich erst einmal unangenehm an und werden abgelehnt. Versucht man nun per Diktat eine Veränderung herbeizuführen, stößt man oft auf starke Gegenwehr und die erhofften Ziele werden nicht erreicht. Die Erfahrung zeigt hier: Wenn Mitarbeiter nicht mitgenommen werden, können sie

jede Transformation scheitern lassen. Aus diesem Grund muss die Einführung einer DevOps-Kultur gut geplant und bewusst begleitet werden. Nach unserer Erfahrung kann eine Transformation zu einer DevOps-Kultur nur gelingen, wenn das Management es schafft, eine klare Vision der Ziele zu formulieren und ein ernsthaftes Interesse bei den Mitarbeitern zu wecken. Häufig hilft hierbei eine externe Unterstützung, um die Scheuklappen abzulegen und neue Denkräume zu öffnen. Wie in *Abbildung 3* zu erkennen, kann die Transformation dabei in fünf Schritten aufgebaut werden.

Im ersten Schritt gilt es alle nötigen Vorbereitungen für die Transformation zur DevOps-Kultur zu treffen. Dabei geht es darum, der Transformation einen Rahmen zu geben, geeignete Bereiche und Mitarbeiter zu identifizieren und benötigte Trainings durchzuführen. Für diesen Schritt sind etwa 45 Tage einzuplanen.

Zunächst muss eine Vision mit dem Management formuliert werden. Dabei sollte die Vision unter anderem folgende Fragen beantworten können:

- Was sind die Herausforderungen des Unternehmens?
- Wie hat sich die Situation im Vergleich zur Vergangenheit verändert?
- Welche Grundlage dafür gibt es?
- Welche Auswirkungen könnte es für jeden einzelnen Mitarbeiter und das gesamte Unternehmen haben, wenn nicht reagiert wird?
- Wie kann DevOps bei der Bewältigung der Herausforderungen helfen?
- Was ist das konkrete Ziel für die Einführung einer DevOps-Kultur?

Häufig wird dem Management bei der Formulierung der Vision auch erst klar, welche Tragweite die Transformation hat, aber auch, welche Potenziale erschlossen werden können.

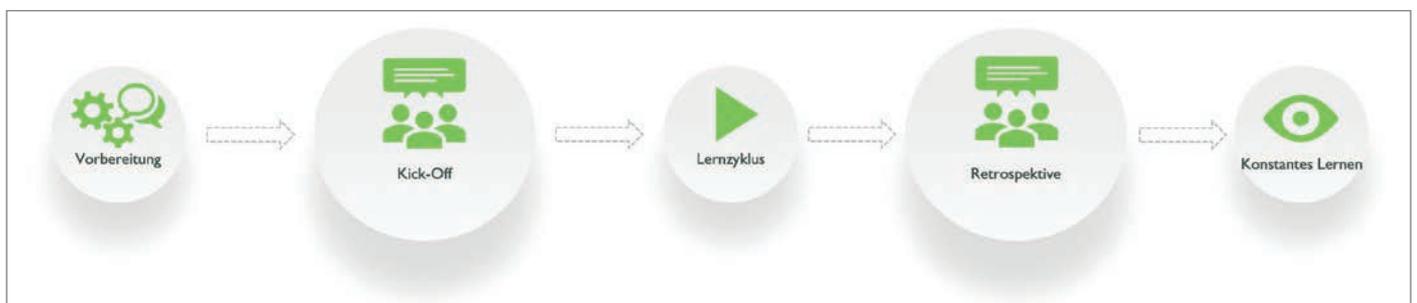


Abbildung 3: Transformationsprozess (© esentri AG)

Ein zweiter wichtiger Aspekt ist die Identifikation geeigneter Projekte oder Produkte, die von einer DevOps-Kultur profitieren werden. Oft wird hierbei der Fehler gemacht, unkritische Bereiche zu nutzen, um das unternehmerische Risiko zu minimieren. Das Problem hierbei ist, dass damit zwar das Risiko, allerdings auch der spürbare Effekt eines Erfolges reduziert wird. Generell lassen sich skeptische Mitarbeiter erst dann überzeugen, wenn sie einen Effekt in ihrem Umfeld bemerken können. Daher ist es wichtig, Produkte und Projekte zu wählen, die genug Aufmerksamkeit innerhalb der Belegschaft haben.

Nun existiert eine Vision und ein geeigneter Bereich für die Umsetzung ist identifiziert. Bisher sollten maximal zwei intensive Wochen in enger Zusammenarbeit mit dem Management vergangen sein. Als Nächstes müssen die Personen, welche die Transformation zu einer DevOps-Kultur gestalten sollen und auch wollen, gefunden werden. Gerade der Aspekt des „Wollens“ ist hier, im Sinne der Motivation und Handlungsbereitschaft, sehr wichtig. Hierzu hat

es sich bewährt, Werbung für das Vorhaben im eigenen Unternehmen zu machen und eine offene Einladung durch das Management auszusprechen. Grundlage hierfür ist die erarbeitete Vision. Das Ziel ist es, in der verbliebenen Vorbereitungszeit, etwa 30 Tagen, genügend Freiwillige für die Transformation zu gewinnen. Hierfür muss das Thema omnipräsent und an neuralgischen Punkten für die Kollegen sichtbar sein. Bewährte Techniken hierfür sind Newsletter, Artikel im Intranet, Aushänge in Korridoren und an Kaffeemaschinen sowie Gespräche unter Kollegen. In der Regel finden sich in dieser Phase genug motivierte Kollegen, die bereit sind, die Reise auf sich zu nehmen und als Pioniere für das Unternehmen voranzugehen.

Nach der Vorbereitung folgt der zweite Schritt, eine Kick-off-Veranstaltung zum Start der gemeinsamen Reise hin zur DevOps-Kultur. Bei diesem Kick-off sollten alle Beteiligten gemeinsam an einem Ort zusammenkommen und die Transformation starten. Für den Kick-off sollten ein bis zwei Tage eingeplant werden. Inner-

halb des Zeitraums sollten mindestens folgende Themen adressiert werden:

- Vorstellung der Vision durch das Management
- Vorstellung des begleiteten Transformationsprozesses
- Vorstellung der zeitlichen Planung & Methodik
- Sammeln von offenen Fragen und Herausforderungen
- Beantwortung offener Fragen durch das Management
- Erarbeitung von Lösungen für Herausforderungen durch alle Beteiligten

Je nach verwendeter Methode für die Produktentwicklung kann sich die konkrete Ausgestaltung der Veranstaltung unterscheiden. Es ist oft hilfreich, erfahrene Moderatoren mit an Bord zu haben, damit die Kick-off-Veranstaltung an die konkreten Rahmenbedingungen angepasst werden kann.

Nach dem Kick-off der Initiative folgt der dritte Schritt, die begleitete Einführung der DevOps-Kultur. In dieser ersten,

ENTDECKEN SIE UNSER KOMBINIERTES ONLINE- UND PRÄSENZ-KURSANGEBOT!

robotron
schulung

Duale Weiterbildung im Robotron Schulungszentrum

Ab Juni begrüßen Sie unsere Dozenten gern wieder persönlich in unseren modernen Schulungsräumen. Wir bieten Ihnen optimale Lernbedingungen, egal ob vor Ort oder online mit eigener Übungsumgebung.

Viele Kurse können ab sofort als Online- und Präsenzveranstaltung gebucht werden. Wir freuen uns auf Ihre Anfrage. Finden Sie jetzt Ihr maßgeschneidertes Weiterbildungsangebot unter www.robotron.de/schulungszentrum



Durchführungs-
garantie



Live-Training mit
erfahrenen Trainern
von Robotron



Persönliche
und individuelle
Betreuung



Lernplattform mit
zusätzlichem
Trainings- und
Übungsmaterial



etwa 100 Tage andauernden Lernphase geht es darum, dass sich Abteilungen und Teams selbst ausprobieren und das in den Trainings gelernte Wissen anwenden. Dabei entwickeln jeder Mitarbeiter und jedes Team eigene Methoden, die auf die jeweilige Zusammenstellung der Charaktere und die verwendeten Technologien passen. Bei diesem Prozess ist es wichtig, konstant über die Methoden und Werkzeuge zu reflektieren. Dies hilft dabei, die DevOps-Kultur und -Prozesse stetig zu verbessern sowie zu festigen. Zusätzlich fördert der offene Austausch zwischen Dev und Ops ein besseres Verständnis für die Bedürfnisse und Fähigkeiten des Gegenübers. Für die Reflektion während dieser Phase hat es sich bewährt, auf Methoden aus der agilen Werkzeugkiste oder den Liberating Structures [2] zurückzugreifen.

Innerhalb dieser drei Phasen wird eine externe Begleitung der Transformation durch erfahrene Coaches empfohlen. Sie helfen allen Beteiligten dabei, alte Denkmuster zu verlassen und den Lernprozess strukturiert zu gestalten.

So wichtig diese Unterstützung für die Vorbereitung und Einführung der DevOps-Kultur auch ist, irgendwann muss die Betreuung enden und die neu gebildeten Strukturen müssen beweisen, dass sie autonom existieren können. Den Übergang hierzu stellt der vierte Schritt dar, eine moderierte Retrospektive über die Lernphase der vergangenen 100 Tage. Wie beim Kick-off zuvor kommen für diese Veranstaltung alle Beteiligten für einen Tag an einem Ort zusammen. Während dieses Tages sollen Erfolge gefeiert, Gelerntes reflektiert und bestehende Herausforderungen gemeinsam gelöst werden. Als weiterer Themenpunkt muss noch einmal klar gemacht werden, dass die intensive Betreuung der Coaches zu diesem Zeitpunkt endet. Diese klare Kommunikation stellt ein Übergangsritual für alle Beteiligten dar. So wird gewährleistet, dass ein Umdenken in den Köpfen stattfindet und die Coaches nicht als Sicherheit immer im Hintergrund stehen.

Nach dem Übergangsritual geht die Transformation über in den fünften und letzten Schritt. Dabei handelt es sich um einen konstanten Lernzyklus von 100 Tagen mit anschließender Retrospektive zur Weiterentwicklung und Festigung der DevOps-Kultur. Sämtliche Coaches sollten im zweiten Lernzyklus zunächst für 30 Tage anderen Tätigkeiten nachgehen. Dadurch

können alle Beteiligten ihre Autonomie akzeptieren und Vertrauen in ihre Methode gewinnen. Nach 30 Tagen können die Coaches zu einzelnen Terminen und Reviews zurückkehren. Dabei werden lediglich Gesundheitschecks der Kultur durchgeführt und Impulse gegeben. Die entstandene Organisation soll allein lebensfähig sein.

Aus Erfahrung ermöglicht das beschriebene Vorgehen eine effektive und nachhaltige Einführung einer DevOps-Kultur in Abteilungen und Unternehmen. Dabei verstreichen von der Initiierung bis hin zur ersten großen Retrospektive etwa 145 Tage. Anschließend kann das etablierte System gefestigt und bei Bedarf im Unternehmen skaliert werden.

Fazit

Das Thema DevOps ist vielschichtig und bietet verschiedenste Diskussionsebenen. Gerade im Bereich der Zusammenarbeit und Arbeitskultur darf dabei nie vergessen werden, dass es um Menschen und deren Miteinander geht. Entscheidungen auf Papier prägen sich in der Realität anders aus als erwartet. Genau deshalb ist es so wichtig, flexibel reagieren zu können, Beteiligte im Prozess abzuholen, Verständnis zu schaffen und auf ein gemeinsames Ziel hinzuarbeiten. Stimmt die Grundeinstellung der Beteiligten, kann ein Transformationsprozess, wie oben beschrieben, gelingen! Die systematische Einführung von DevOps erzeugt eine Aufbruchsstimmung im Unternehmen, die sich nicht nur in hohem Maße auf die Mitarbeiterzufriedenheit, sondern auch positiv auf wirtschaftliche Kennzahlen auswirkt. Verschließen Sie sich deshalb nicht vor Neuem und bleiben Sie neugierig, denn wer stehen bleibt, hat bereits verloren.

Quellen

- [1] Dawn Foster 28.06.2016: What is DevOps? John Willis Explains. Linux.com
- [2] Holistic AG 14.03.2020: www.liberatingstructures.de

Über die Autoren

Carsten Wiesbaum

Immer schon fasziniert von Technologie und Software, habe ich in meinem Berufsleben

verschiedene Rollen wahrgenommen. Was mir auf dem Weg früh klar wurde, ist, dass Technologie und Software meist nicht Ursache oder Lösung für Probleme in Unternehmen sind. Ich glaube, dass Menschen in ihrem Arbeitsalltag Erfüllung und Selbstverwirklichung erreichen können. Hierzu braucht es jedoch ein neues Denken in Unternehmen, das auf agilem Mindset, Empowerment und Transparenz basiert. In dieser Entwicklung unterstütze ich mein eigenes und andere Unternehmen seit vielen Jahren.

Kevin Steinhagen

Als Programmierer und Software-Berater in verschiedenen IT-Projekten erkannte ich mein Talent bei der Optimierung von Kommunikation und Zusammenarbeit im Team. Als Agile Coach liegt mir das Vorbringen der digitalen Kultur und das erfolgreiche Miteinander besonders am Herzen. Durch die Erfahrungen als Programmierer kenne ich die Herausforderungen eines Entwicklers aus erster Hand. Im Projektalltag und bei der Arbeit im eigenen Unternehmen verbinde ich technische Skills der Softwareentwicklung mit den methodischen Anforderungen eines Coaches.



Carsten Wiesbaum
carsten.wiesbaum@esentri.com



Kevin Steinhagen
kevin.steinhagen@esentri.com



CI/CD: Theorie meets Praxis

Dennis Hoffmann, MT AG

In einer immer schnelllebigeren Zeit geht es auch in der IT mehr und mehr darum, flexibler und agiler zu sein. Ein agiles Projektmanagement (zum Beispiel Scrum oder Kanban) schafft hierfür die Grundvoraussetzung, um flexibel auf Kundenwünsche und geänderte Projektumstände zu reagieren. Wie aber lässt sich der entwickelte Code permanent testen und anschließend schnell und automatisch an die Endanwender ausliefern? Hier hört man immer häufiger von CI/CD. Doch was heißt das eigentlich genau und wie funktioniert es in der Praxis?

Dieser Artikel soll eine Brücke zwischen Theorie und Praxis schlagen. Durch meine Erfahrungen aus einem großen Java-Projekt kann ich beurteilen, wie eine optimale Systemumgebung aussehen sollte und welche weiteren Bedingungen erfüllt sein müssen. So werden im Weiteren die folgenden Fragen geklärt: Welche Voraussetzungen werden für CI/CD benötigt? Wie baue ich eine CI/CD-Pipeline in meinem Projekt auf? Welche Hindernisse und Probleme müssen überwältigt werden, um zum Erfolg zu kom-

men? Wie wirkt sich die Einbindung von CI/CD in der Praxis auf den Projekterfolg und vor allem auf die Zeitspanne von der Entwicklung bis zum produktiven Einsatz aus? Was für Vorteile bringt der Einsatz langfristig mit sich?

CI/CD – Was genau ist das eigentlich?

Es handelt sich bei CI/CD um Methoden, bei denen sowohl den Kunden und

Endanwendern als auch dem Entwicklerteam regelmäßig lauffähige Software bereitgestellt wird und alle Phasen der Anwendungsentwicklung automatisiert ablaufen können.

„CI“ bedeutet hierbei Continuous Integration, also der Automatisierungsprozess für das Entwicklerteam. Bei einer erfolgreichen CI werden durch das Team regelmäßig neue Codeänderungen für die Anwendungen entwickelt, geprüft und in einem gemeinsamen zentralen Repository zusammengeführt.

Bei der modernen Anwendungsentwicklung arbeiten mehrere Entwickler an unterschiedlichen Features der gleichen Anwendung. Ohne den Einsatz von Continuous Integration kann die gleichzeitige Zusammenführung aller Quellcode-Branches an einem Tag (auch bekannt als „Merge Day“) einen hohen Arbeits- und

Zeitaufwand bedeuten. Der Grund dafür ist, dass Anwendungsänderungen von getrennt arbeitenden Entwicklern miteinander in Konflikt treten können, wenn sie zeitgleich durchgeführt werden.

Mithilfe der Continuous Integration können Entwickler ihre Codeänderungen in einem gemeinsamen „Branch“ der Anwendung viel häufiger zusammenführen, sogar mehrmals täglich. Sobald die Änderungen eines Entwicklers zusammengeführt werden, werden sie in automatischen Builds und unterschiedlichen Stufen von Automatisierungsprüfungen (normalerweise Unit- und Integrations-tests) validiert. So wird sichergestellt, dass die Funktionsfähigkeit nicht beeinträchtigt wurde. Dabei müssen alle Klassen und Funktionen bis hin zu den verschiedenen Modulen der Anwendung getestet werden. Wenn die automatische Prüfung Konflikte zwischen aktuellem und neuem Code erkennt, lassen sich diese mithilfe

von CI schneller und dadurch einfacher beheben.

„CD“ bedeutet Continuous Delivery beziehungsweise Continuous Deployment. Beide Konzepte sind verwandt und werden im Projektgeschäft teilweise synonym verwendet. Obwohl es bei beiden Konzepten um die Automatisierung weiterer Phasen der Pipeline geht, unterscheiden sie sich tatsächlich aber im Ausmaß der Automatisierung.

Continuous Delivery bedeutet üblicherweise, dass Änderungen eines Entwicklers automatisch auf Bugs getestet und in ein zentrales Repository hochgeladen werden, von wo aus sie vom Team händisch in einer Live-Produktionsumgebung verwendet werden können. Dieser Vorgang ist die Antwort auf Transparenz- und Kommunikationsprobleme zwischen Dev- und Business-Teams. Damit soll sichergestellt werden, dass neuer Code mit minimalem Aufwand implementiert werden kann. Ziel der Continuous Delivery ist eine Codebasis, die jederzeit für die Implementierung in einer Produktionsumgebung bereit ist.

Continuous Deployment wiederum automatisiert zusätzlich auch noch die Freigabe von Entwickleränderungen vom Repository zur Produktionsumgebung, wo sie

direkt vom Endanwender genutzt werden können. Dieser Vorgang soll der Überlastung von Ops-Teams bei manuellen Prozessen entgegenwirken, die die Anwendungsbereitstellung verlangsamen.

Da hier der Produktionsphase in der Pipeline kein manuelles Gate mehr vorgeschaltet ist, müssen beim Continuous Deployment die automatisierten Tests immer sehr gut durchdacht sein.

Wie sieht eine solche Umgebung in der Praxis aus?

Durch die Praxis bei einem großen Kunden habe ich persönlich Erfahrungen mit folgenden Systemen machen können:

Git (Bitbucket), Maven, Jenkins, JFrog Artifactory, Oracle Database, WebLogic Server.

Grundsätzlich lässt sich sagen, dass agile Methoden der Softwareentwicklung prädestiniert für die Verwendung einer CI/CD-Pipeline sind. Sie haben beide die kurzen Intervalle, permanent lauffähige Software und schnelle Reaktionszeiten als Ziel.

Bei jedem Commit im **Git**-Repository wird über einen Bitbucket-Hook der entsprechende **Jenkins**-Build-Job ausgelöst und beginnt nun die neue aktuelle, veränderte Codebasis auszuchecken, den

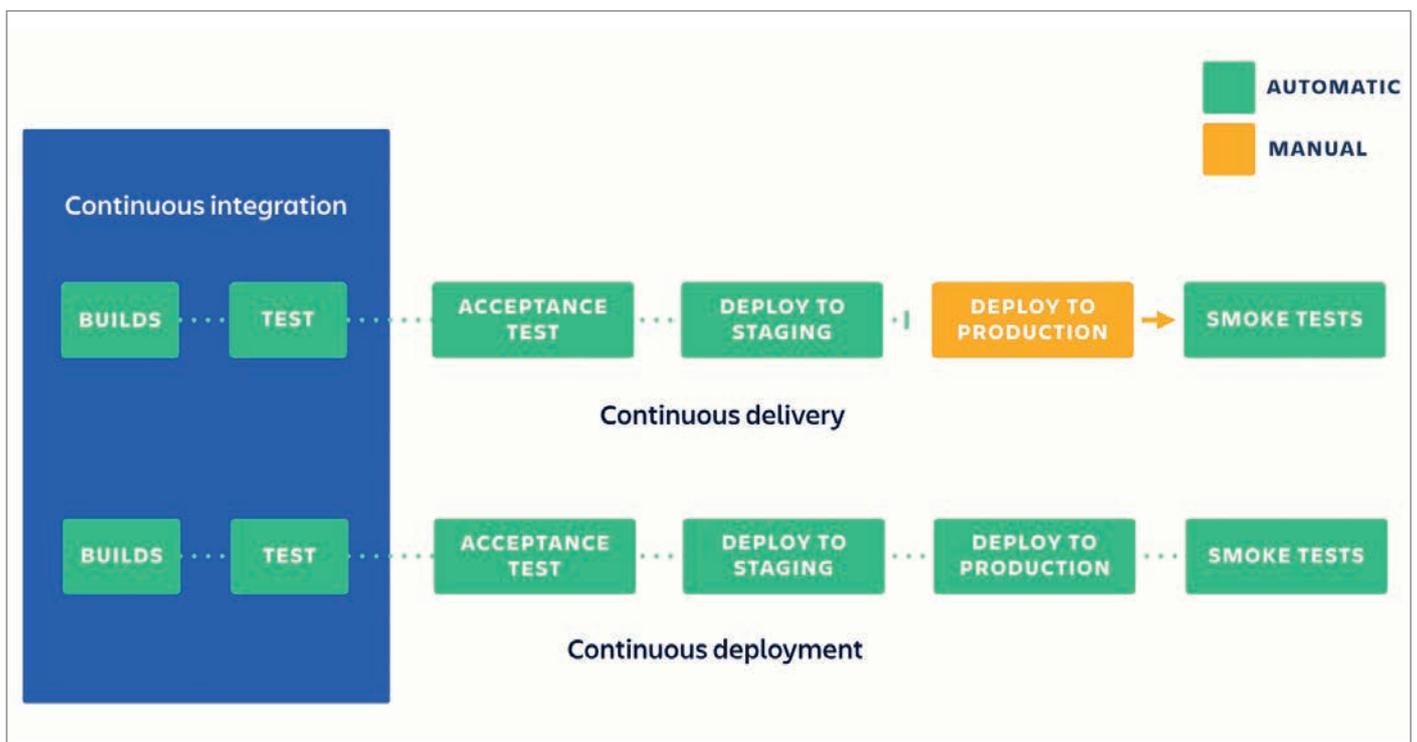


Abbildung 1: Zwei Arten von CI/CD-Pipelines (Quelle © <https://www.atlassian.com/continuous-delivery/principles/continuous-integration-vs-delivery-vs-deployment>)

Code mit **Maven** zu bauen, Tests zu starten und anschließend die gebauten Jars in das zentrale **Artifactory** zu deployen.

Somit ist es jedem Entwickler sofort möglich, die neueste Codebasis zu verwenden und seine eigene Entwicklung kompatibel zu halten. Das Entwicklerteam ist untereinander möglichst stark synchronisiert, dabei arbeiten die einzelnen Entwickler aber weiterhin so autark wie möglich. Dies ist der Kerngedanke der Continuous Integration.

Ebenso hat jeder Entwickler durch die automatisierten Builds des Jenkins die Möglichkeit, seine Implementierung sehr schnell auf einem Testsystem zu verwenden und Probleme so früh wie möglich zu erkennen.

Sollte es im Build-Prozess bereits zu Fehlschlägen bei Tests kommen, erhält der Entwickler automatisch eine E-Mail dazu. Die Jars landen in diesem Fall natürlich auch nicht im Artifactory und behindern somit nicht die Entwicklung des restlichen Teams.

Die „Zusatzstufe“, die komplett gebaute Anwendung direkt auf die Produktion zu deployen, wird aktuell im Projekt meines Kunden nicht verwendet, da ein Großteil der Qualitätssicherung noch aus per Hand durchgeführten Smoketests besteht. Dadurch wird der mögliche Geschwindigkeitsvorteil wiederum ausgebremst und stellt keine optimale Umgebung dar.

Welche Voraussetzungen benötigt ein Projekt dafür?

Um CI/CD effektiv und sicher in seinem Projekt einzusetzen bedarf es einigen Zusatzaufwands. Gerade die Vorabinvestitionen sind beträchtlich, da automatische Tests für die diversen Prüf- und Release-Phasen in der CI/CD-Pipeline geschrieben werden müssen.

Ebenso benötigt man die gesamte Infrastruktur für die Jenkins-Instanzen und Git-Repositories. Gerade bei umfangreichen Projekten braucht man hier schnell eine große Menge an Rechenkapazität. Ohne eine virtualisierte Umgebung stößt man hier sehr schnell an Grenzen der technischen Machbarkeit.

Auch aus meiner Projekterfahrung kann ich sagen, dass der vor Kurzem erfolgte Umstieg in die Cloud das ganze Konzept der CI/CD-Pipeline erst skalierbar gemacht hat.

Ein weiterer unumgänglicher Teil zu einer erfolgreichen Arbeit mit CI/CD besteht in einem Team, das entsprechend geschult und eingestellt ist. Die Entwickler müssen mit agilem Projektmanagement vertraut sein und die eingesetzten Tools sicher beherrschen. Zudem muss ein Bewusstsein für die immense Bedeutung der zu schreibenden Tests geschaffen werden.

Ein umfangreiches, über die Jahre gewachsenes Projekt auf eine CI/CD-Pipeline umzustellen, gestaltet sich hierbei umso schwieriger, da die Anpassungen und Umstellungen im laufenden Projektalltag geschehen müssen.

Wo liegen die Schwierigkeiten von CI/CD?

Wie alle automatisierten Systeme birgt auch CI/CD einige Gefahren. So ist die Pipeline nur so gut und sicher, wie ihre Implementierung dies vorgibt. Wenn zum Beispiel ein Fehler in der Testauswertung besteht und Fehlschläge nicht korrekt behandelt werden, besteht die Gefahr, dass der Code ohne weitere Kenntnis über den Fehler in das Artifactory oder sogar auf die Produktionsumgebung gelangt.

Ebenso hängt die Qualität und Fehlerquote in der fertigen Anwendung in hohem Maße von der Quantität und Qualität der Unit-, Integrations- und auch Smoketests ab. Werden zu einem neuen Feature keine ausreichenden Testcases erkannt oder diese nicht richtig implementiert, fallen diese Fehler womöglich erst auf der Produktion auf.

Gerade im Hinblick auf die Qualitätssicherung muss das Team verinnerlichen, dass ein beträchtlicher Teil der Entwicklungszeit für das Definieren und Implementieren ausreichender Testscenarien benötigt wird. Auch im Zuge der Qualitätssicherung der umgesetzten Features muss ein besonderes Augenmerk auf den Tests liegen.

Welche Grenzen gibt es?

Die Vorteile der automatisierten Softwareentwicklung und -auslieferung liegen auf der Hand. Die Geschwindigkeit, in der entwickelt, getestet und neue Software ausgeliefert werden kann, gerade in großen, dezentralen Projekten, an denen viele Entwickler gleichzeitig arbeiten, ist das schlagende

Argument. Die Entwickler, der Kunde und auch der Endanwender profitieren unmittelbar von der schnellen Reaktionszeit.

Meine Erfahrung im Projekt hat aber auch gezeigt, dass die Verlagerung dieser ganzen Schritte an einen zentralen Ort auch einen gegenteiligen Effekt haben kann. Stehen der Jenkins oder das Artifactory aus trivialen Gründen wie Wartung oder auch Ausfällen nicht zur Verfügung, kann der einzelne Entwickler einen Großteil der Schritte nicht mehr händisch ausführen.

Wie zuvor bereits erwähnt, hängt die Skalierbarkeit der CI/CD-Pipeline in hohem Maße vom Einsatz einer cloudbasierten Lösung ab. Gerade in umfangreichen Projekten mit vielen Tests und großen Datenmengen stößt man ansonsten schnell an die Grenzen der Hardware. Durch die flexible Zuschaltung bei tatsächlich benötigter Rechenkapazität lässt sich der oft zitierte Bottleneck hier vermeiden. Ganz nebenbei kann man durch Abschaltung nicht benötigter Prozessoren auch noch Kosten sparen.

Über den Autor

Dennis Hoffmann arbeitet seit über drei Jahren als IT-Berater der MT AG im Bereich Oracle Java Development.

Durch seine Tätigkeit als Berater begleitet er Firmen in Prozessen der Digitalisierung und Automatisierung. Die immer weiter fortschreitende Digitalisierung stellt gerade große Konzerne und mit ihnen historisch gewachsene IT-Projekte vor immense Herausforderungen, bei denen individuelle, qualitativ hochwertige Lösungen gefordert sind.

2019 war er als Speaker mit dem vorangegangenen Artikel auf der DOAG-Konferenz.



Dennis Hoffmann
Dennis.Hoffmann@mt-ag.com

Monitoring mit Prometheus

Markus Bente, Trivadis
Michael Mühlbeyer, Trivadis Germany

Prometheus bringt, wie sein Namensgeber in der griechischen Mythologie, Licht in das Monitoring von Cloud, Microservices und Container-basierten Umgebungen. Prometheus liefert auf einfache und schnelle Weise Metriken der verfügbaren Systeme. 2012 von SoundCloud entwickelt, wird es seit 2016 als Projekt der Cloud Native Computing Foundation geführt.



Das Monitoring-Tool Prometheus besteht im Kern aus einer Time-Series-Datenbank. Prometheus konzentriert sich auf die Speicherung von Metriken aus den verschiedenen Quellen, dabei wird kein Zustand eines Hosts oder Servers gespeichert, sondern lediglich Name und Wert der entsprechenden Metrik mit einem Label, um den Kontext zur überwachten Komponente herstellen zu können. Die Metriken werden dabei von Prometheus an der jeweiligen Quelle abgeholt (siehe Abbildung 1). Für die Abfrage der Metriken bietet Prometheus eine eigene Abfrage-Sprache: PromQL

Für verschiedene Dienste, wie zum Beispiel Datenbanken, Messaging-Systeme und Webserver, bietet Prometheus die Möglichkeit, mit sogenannten Exportern die Metriken von diesen Systemen abzugreifen (siehe Abbildung 2). Darüber hinaus gibt es eine Vielzahl von Diensten und Software, die die Metriken bereits in einem für Prometheus lesbaren Format bereitstellen. Dazu gehören unter anderem Kubernetes, Kibana, WildFly, Springboot und OpenStack. Eine vollständige Liste aller Exporter gibt es auf der offiziellen Prometheus Website (<https://prometheus.io/docs/instrumenting/exporters/>). Für die Überwachung von HTTP(S), DNS, ICMP oder TCP Endpoints bietet Prometheus den sogenannten Blackbox Exporter. Dieser erlaubt es relativ einfach, die Verfügbarkeit und Performance eines Service zu monitoren. Eine Performance-Metrik ist zum Beispiel die Response Time des überwachten Service. Die Exporter sind Teil des gesamten Prometheus-Stacks, laufen aber als eigenständiger Prozess unabhängig vom Prometheus-Prozess.

Das Alerting im Prometheus-Toolstack erledigt der Alertmanager. Der Alertmanager gehört wie die Exporter zum Prometheus-Stack, läuft aber als separater Prozess neben Prometheus. Standardmäßig sendet Prometheus auftretende Alerts an den Alertmanager (siehe Abbildung 3). Dieser erlaubt eine Gruppierung der Alarme und bietet die Möglichkeit, Alarme zu deduplizieren. Deduplikation basiert auf den Labels des Alarms, der gesendet wurde. Sollten zwei Alarme mit identischen Labels gemeldet werden, so wird nur eine Notifikation ausgelöst. Des Weiteren bietet der Alertmanager die Möglichkeit, Abhängigkeiten zu Alerts mittels Inhibition Rules zu definieren. Dies gibt uns die Möglichkeit, bestimmte Alarm-Notifikationen zu unterdrücken, wenn bestimmte andere Alarme

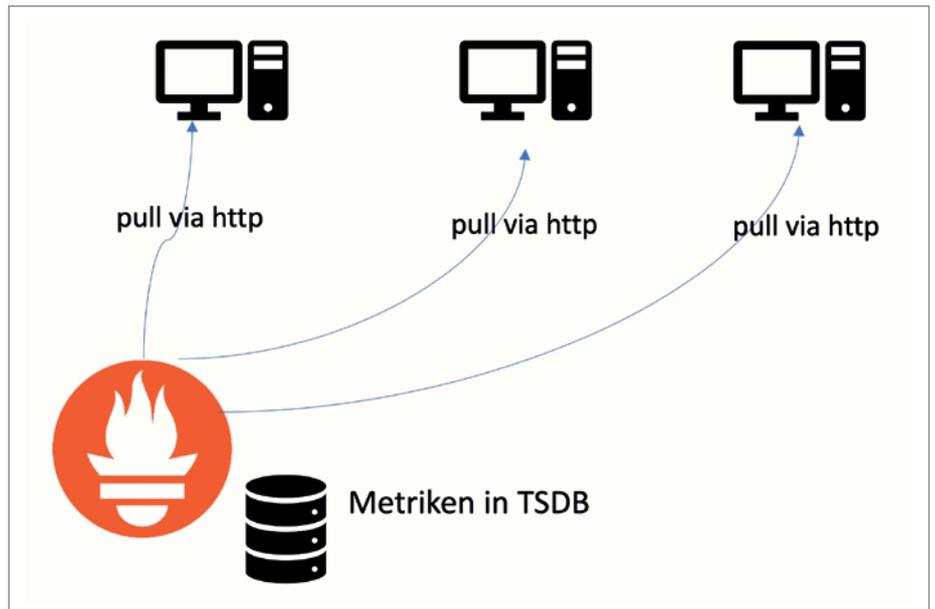


Abbildung 1: Prometheus-Pull-Architektur (© Michael Mühlbeyer)

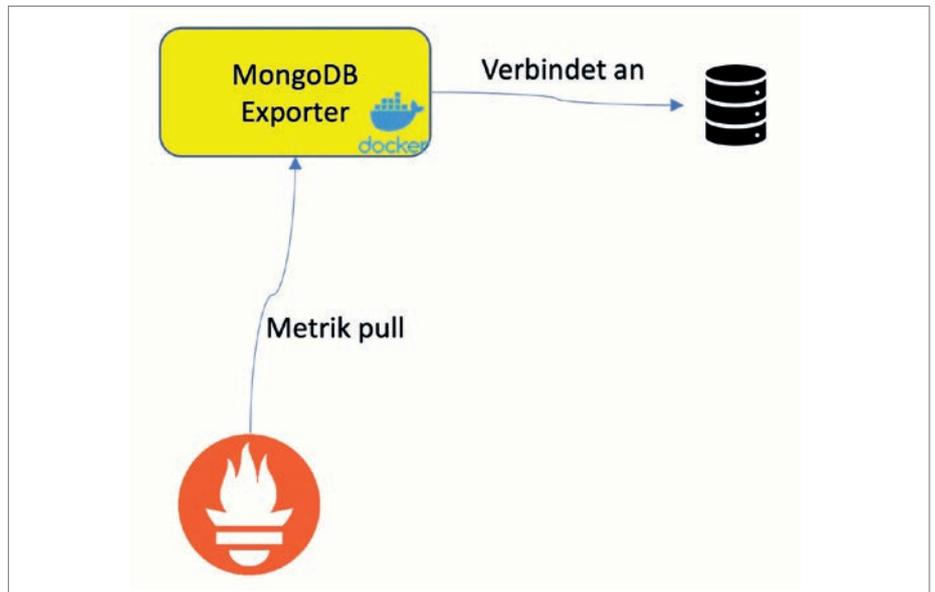


Abbildung 2: Prometheus Exporter (© Michael Mühlbeyer)

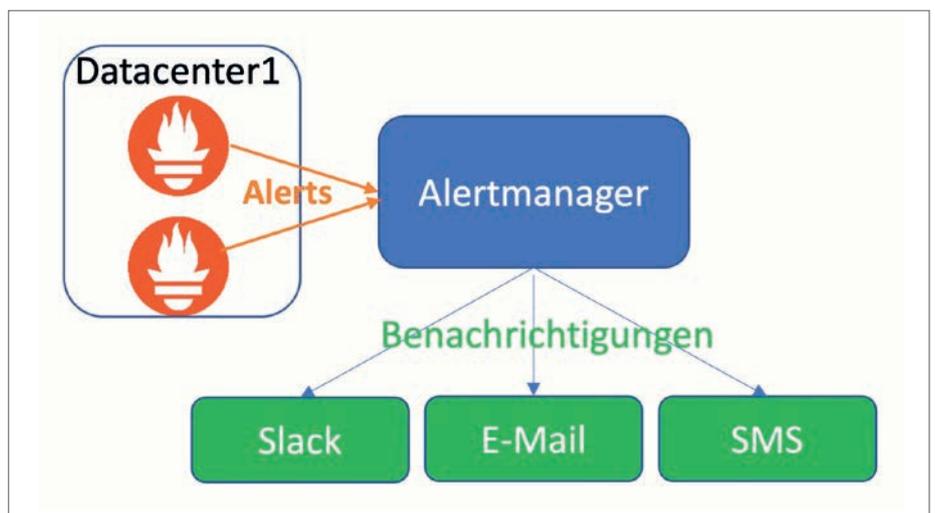


Abbildung 3: Prometheus Alertmanager (© Michael Mühlbeyer)

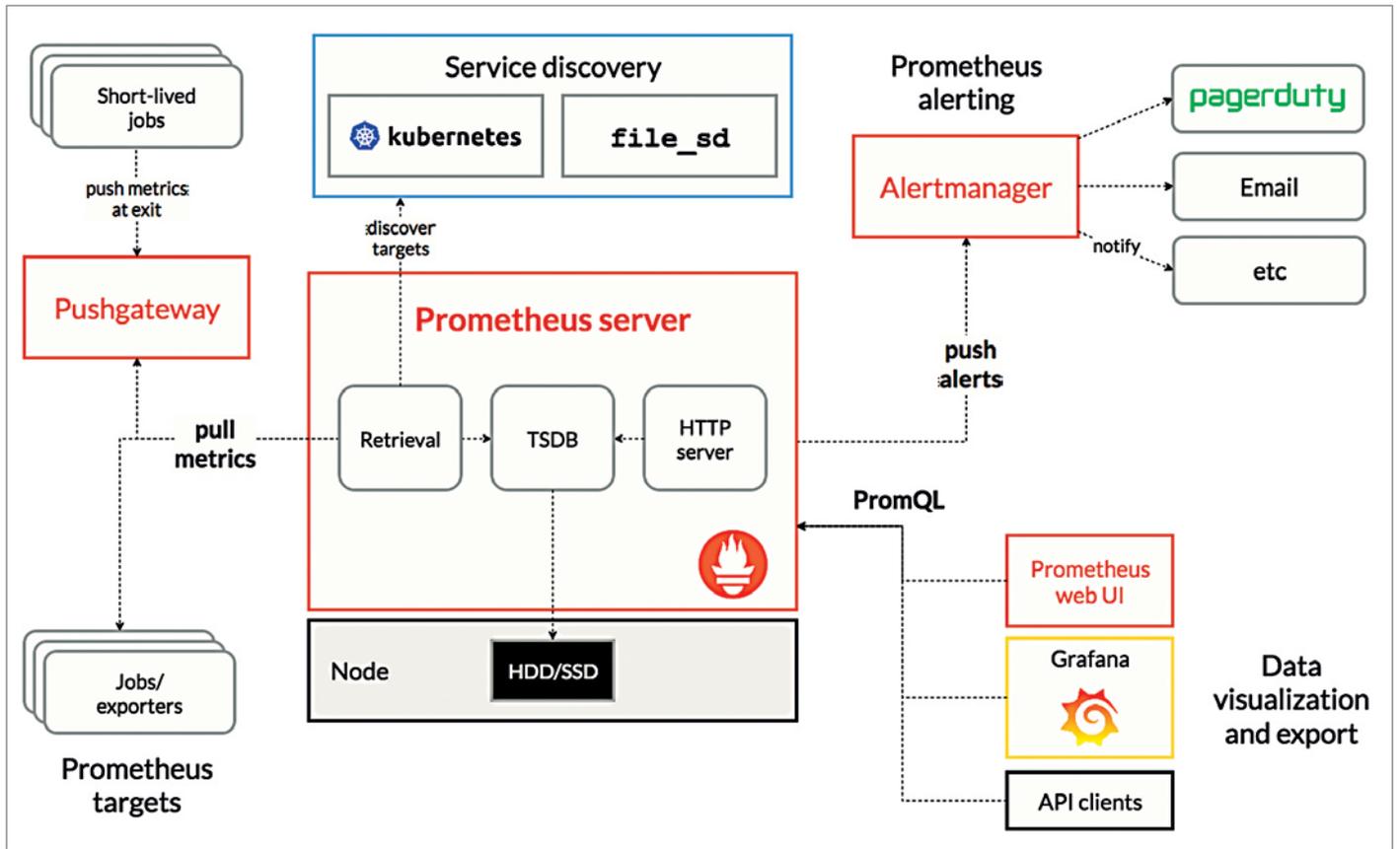


Abbildung 4: Gesamte Prometheus-Architektur (© Prometheus Authors, Linux Foundation, <https://prometheus.io/assets/architecture.png>)

ausgelöst wurden. Darüber hinaus steuert der Alertmanager das Routing der Alarme an verschiedene Ziele. Möglich sind zum Beispiel E-Mail, Slack oder Microsofts Teams (siehe Listing 1).

Die Listings 1 bis 14 finden Sie unter folgendem Link:
http://www.doag.org/go/redstack/202004_listings_bente_muehlbeyer

Zusammenfassend hier noch einmal die Prometheus-Architektur in Gesamtübersicht mit den einzelnen Komponenten (siehe Abbildung 4).

Prometheus-Installation und -Konfiguration

Die Installation von Prometheus ist sehr einfach. Prometheus selbst, die Exporter und der Alertmanager können als TAR-Archiv direkt auf der Prometheus Website geladen werden. Auf Wunsch gibt es das Ganze auch als Quellcode zum Selbstkompilieren oder wahlweise als Docker Image. Die Größe der Software ist recht über-

schaubar, die drei Komponenten Prometheus, Alertmanager und Blackbox Exporter benötigen entpackt nur ca. 200 MB. Das entpackte Archiv enthält alle relevanten Dateien (siehe Listing 2). Prometheus ist mit der ausgelieferten Beispiel-Konfiguration lauffähig und benötigt für einen ersten Start

keine Anpassung an die Konfigurations-Dateien. Die zentrale Konfigurationsdatei ist die Datei prometheus.yml, hier können die Parameter für die globale Konfiguration, das Alerting und die Scrape-Konfiguration gemacht werden. Die Scrape-Konfiguration definiert, bei welchen Targets Metriken ab-



Abbildung 5: Beispiel Dashboard (© Markus Bente)

geholt werden (siehe Listing 3). Weitere Konfigurationsparameter können direkt dem Prometheus Executable übergeben werden (siehe Listing 4).

Für das Setup der Kommunikation zwischen Prometheus und dem Alertmanager muss dieser lediglich über das Konfigurationsfile bekannt gemacht werden. Für den Blackbox Exporter wird ein zweiter Scrape Job definiert, der für unser Beispiel die Verfügbarkeit prüft (siehe Listing 3).

Eine ausführliche Beispiel-Konfiguration findet sich hier: <https://github.com/mmuehlbeyer/prometheus>

Prometheus-Visualisierung mit Grafana

Was wäre ein Monitoring ohne eine schicke Visualisierung? Mit Grafana können die in Prometheus gespeicherten Metriken einfach und schnell dargestellt werden. Dazu muss Prometheus nur als Datenquelle für Grafana definiert werden. Danach können mit den vorliegenden Daten selbst Dashboards gebaut oder fertige Dashboards (siehe Abbildung 5) bei Grafana geladen werden (<https://grafana.com/grafana/dashboards>).

Prometheus und Docker

Eine Prometheus-Umgebung kann auch mit Docker betrieben werden.

Die einzelnen Komponenten (Prometheus, Alertmanager, Blackbox Exporter usw.) werden jeweils als Container gestartet und entsprechend parametrisiert. Dafür bietet sich ein Docker Compose an (siehe Listing 5).

Die Konfiguration der einzelnen Container befindet sich in den entsprechenden Dockerfiles, zum Beispiel für den Service Prometheus (siehe Listing 6). Die einzelnen yaml-Files können mit kleinen Anpassungen aus der Non-Container-Konfiguration übernommen werden.

Prometheus und Kafka

Am Beispiel Kafka möchten wir Ihnen näherbringen, wie Prometheus als Überwachungslösung eingesetzt werden kann. Diese Architektur beinhaltet die Überwachung der zentralen Java-Prozesse sowie deren Custom mBeans von Broker, Zookeeper und Mirror Maker (siehe Abbildung 6).

Als Replication Monitoring kommt ein Custom-build Docker-Container zum Einsatz.

Prozess-Monitoring

Zum Exponieren der mBeans als Rest-Endpoint kommt der Prometheus JMX Exporter zum Einsatz. Dieser kann als Java Agent mit dem Java-Prozess gestartet werden. Alternativ ist auch ein Start als Standalone-HTTP-Server möglich. In unserem Beispiel wird dieser je Prozess als Java Agent mittels Erweiterung des Systemd Service gestartet (siehe Listing 7).

Mithilfe von Konfigurationsdateien kann parametrisiert werden, welche Metriken exponiert werden sollen. Des Weiteren können über diese Konfigurationsdateien Label gesetzt und eine Umbenennung der Metriken vorgenommen werden (siehe Listing 8).

Replication-Monitoring

Der Replication-Monitor ist ein Custom-build Docker-Container, der jede Minute das Consumer Lag einer mitgegebenen Consumer-Gruppe ausliest und in eine Datei schreibt. Im Anschluss wird diese Datei aufbereitet (siehe Listing 9), sodass die Darstellung Prometheus-konform ist. Diese Datei wird mittels Node Exporter als Rest-Endpoint publiziert.

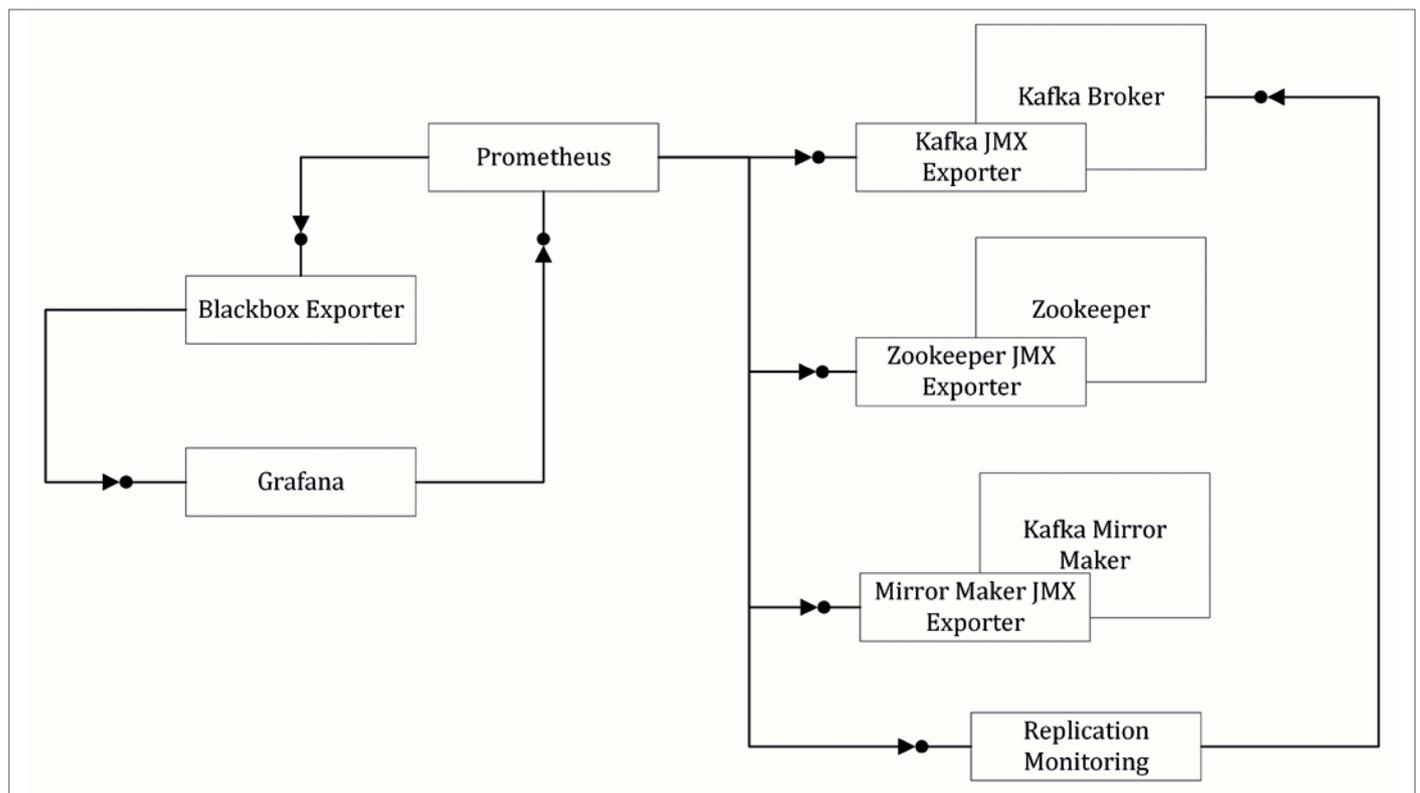


Abbildung 6: Prometheus-Kafka-Monitoring-Architektur (© Markus Bente)

Visualisierung des Kafka-Monitorings

Als Visualisierungs-Frontend kommt Grafana zum Einsatz, das durch den Blackbox Exporter von Prometheus überwacht wird (siehe Abbildung 7).

Alerting

Für das Alerting wird der Prometheus Alertmanager genutzt. Die Konfiguration dafür ist relativ einfach und kompakt, sie besteht aus einzelnen Code-Blöcken. Diese Blöcke bestehen aus mehreren Konfigurationen. Zum einen einer PromQL-Expression und einem Zeitfenster, die definieren, bei welchem Schwellwert der Alert ausgelöst wird (z.B. das Lag einer Replication ist größer als 50 für die Dauer von einer Minute). Zusätzlich gibt es die Möglichkeit, Labels (z.B. Severity) und ein Summary zu definieren (siehe Listing 10).

Prometheus und Oracle-Datenbanken

Für die Überwachung von Oracle-Datenbanken kann Prometheus ebenfalls genutzt werden. Da die Datenbank stan-



Abbildung 7: Grafana Kafka Cluster Dashboard (© Markus Bente)

dardmäßig keine für Prometheus lesbaren Metriken bereitstellt, müssen wir dies der Datenbank mithilfe von PL/SQL und Oracle Rest Data Service (ORDS) beibringen.

Der PL/SQL-Teil wird mit einer Procedure implementiert, die das PL/SQL-Webtool Kit benutzt. Die Procedure wird mit dem User dbsnmp angelegt, da dieser die notwendigen Rechte auf v\$ and dba_ Views besitzt.

Die Metriken werden in einem Cursor gespeichert (siehe Listing 11). Im zweiten Teil der Procedure erfolgt die Ausgabe mittels http.print (siehe Listing 12). In unserem Auszug wird die Oracle-Version als Metrik ausgegeben.

Für das Exportieren der Metriken benötigen wir noch den Oracle Rest Data Service (ORDS). Dieser kann entweder stand-alone oder als WAR-File auf einem Tomcat- oder WebLogic-Server betrieben werden.

Die Installation des ORDS ist dabei nach Standard vorzunehmen (siehe Listing 13). Wichtig ist, dass die Datenbank-Definition angepasst wird. Details dazu gibt es in einem Blog Post [2]

Danach können wir den ORDS-Server starten und unsere Metriken über die URL anzeigen lassen: `http://<servername>:<port>/ords/prometheus_metrics/PROMETHEUS_METRICS`

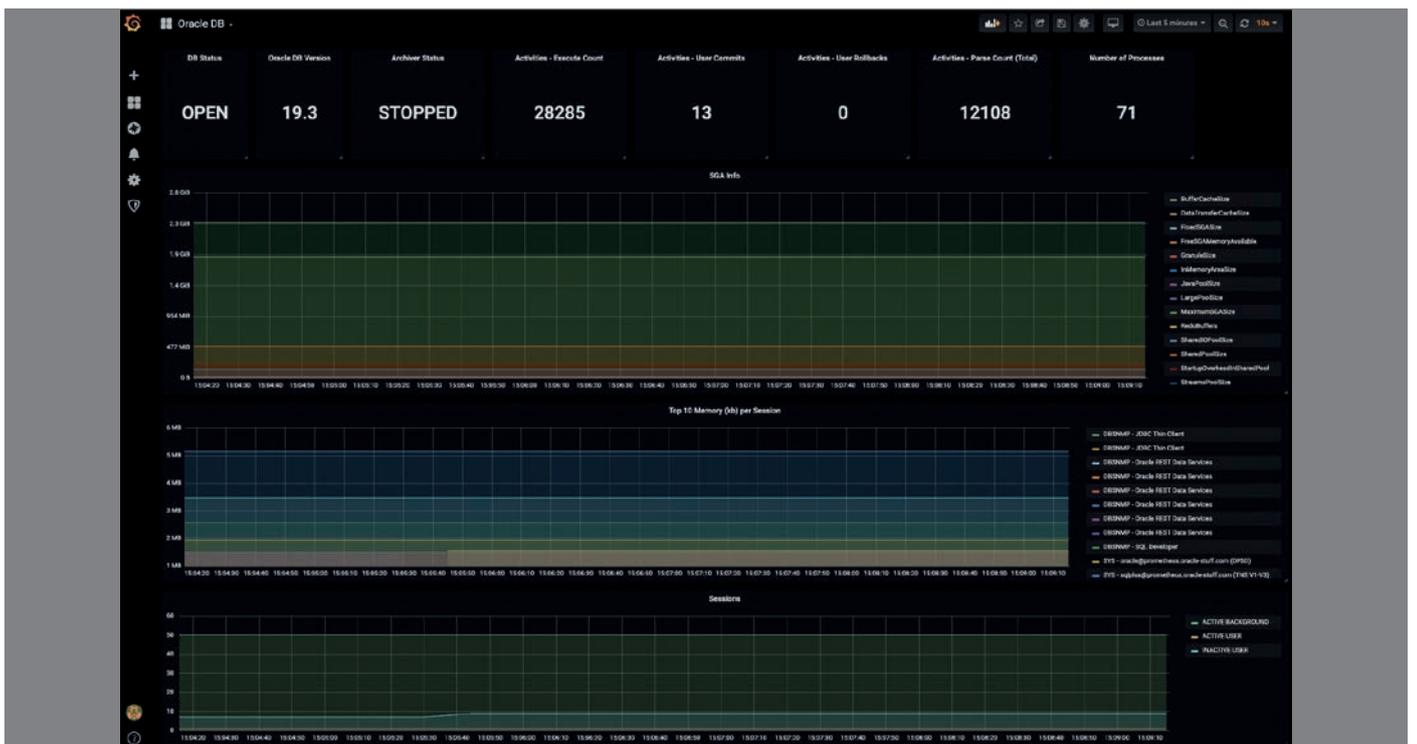


Abbildung 8: Grafana Oracle Dashboard (© Dirk Nachbar, <http://dirknachbar.blogspot.com/2020/03/generating-prometheus-metrics-directly.html>)

Anschließend noch die scrape_config-Konfiguration im prometheus.yml anpassen und den Prometheus-Server neu starten (siehe Listing 14).

Fazit

Prometheus ist eine gute, flexible und performante Lösung für Metrik-basiertes Monitoring. Durch viele verfügbare Exporter, aber auch die Erweiterbarkeit mit eigenen Exportern, kann das System schnell an wechselnde Umgebungen und Systeme angepasst werden. Durch die Visualisierung mit Grafana Dashboards besteht die Möglichkeit, eine 360-Grad-Sicht auf seine Umgebung oder Prozesse zu ermöglichen. Der Prometheus Stack bietet für jedes DevOps-Team eine gute Lösung für das Monitoring der aktuellen und kommenden Komponenten, sei es On-Premises oder in der Cloud.

Quellen

- [1] <https://prometheus.io/>
- [2] <http://dirknachbar.blogspot.com/2020/03/generating-prometheus-metrics-directly.html>

Über die Autoren

Markus Bente ist seit 2017 bei der Trivadis AG am Standort Bern beschäftigt. Sein technologischer Fokus liegt auf FastData- und Data-Analytics-Plattformen sowohl bei der Konzeptionierung der Architektur als auch bei deren Implementierung.

Michael Mühlbeyer ist seit 2012 bei der Trivadis Germany GmbH beschäftigt. Seine Schwerpunkte liegen auf Big Data und Cloud-Architekturen. Ein weiterer Fokus ist die Architektur und der Betrieb von Datenbanksystemen im Oracle- und Open-Source-Umfeld.

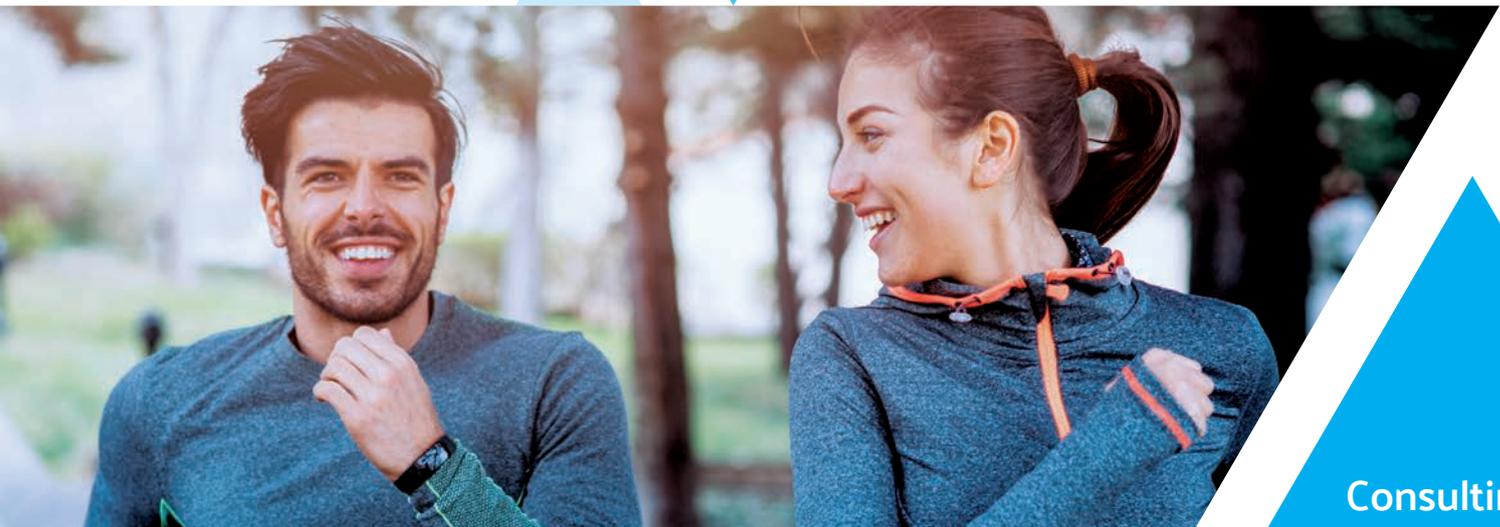


Markus Bente
markus.bente@trivadis.com



Michael Mühlbeyer
michael.muehlbeyer@trivadis.com

MUNIQSOFT
CONSULTING



Consulting

Performance-Tuning mit IQ

Mehr Power für Ihre Oracle Lösungen!

Nutzen Sie unseren proaktiven Datenbank-Healthcheck als Startschuss für die Optimierung Ihrer Oracle Datenbanken.

Ungebremst ans Ziel mit der Muniqsoft Consulting GmbH
www.muniqsoft-consulting.de

ORACLE Gold Partner

Specialized
Oracle Database



Jetzt Beratungstermin vereinbaren:
+49 89 62286789-39

Verfügung stehen. Als kleine Zugabe ist es möglich, die Artefakte zu bereinigen.

Das Hosting

Jetzt zu der wichtigen Frage, wo man sein Open-Source-Projekt ablegen oder veröffentlichen kann oder sollte. Klassisch fallen einem die Namen SourceForge, Gnu, Apache und Eclipse ein. SourceForge und Gnu sind nicht der übliche Ort für Java-Projekte, Apache und Eclipse sind nur für größere Java-Projekte zweckmäßig. Für die Aufgabe eignen sich am besten die freien Repositories von GIT. Hier sind GitHub, GitLab und BitBucket aktuell zu sehen. Alle drei Repositories bieten auch das Erstellen einer Pipeline. Ich habe GitHub gewählt, weil hier viele bekannte Open-Source-Projekte mit ihren Repositories zu finden sind.

Die Tools

Man benötigt ein Java JDK mindestens Version 9. Erste Tests nutzten Java 9, aktuell wird Java 11 verwendet. Bei Maven wird die Version 3.6 eingesetzt. Git ist, wie schon beschrieben, GitHub. IDE ist nicht vorgegeben, hier reicht die POM für IntelliJ, Eclipse und NetBeans. Weitere Tools/Plug-ins werden später vorgestellt.

Die Maven-Vorgaben

Damit man ein Maven Plug-in erstellen kann, wird Maven als Build Tool benötigt und auch als Umgebung, mit der man testet, da das Plug-in in Maven läuft. Es werden folgende Abhängigkeiten benötigt (siehe Listing 1).

In Listing 2 sehen Sie die notwendigen Einstellungen, um ein Maven Plug-in zu erstellen.

Das Packaging ist hier nicht jar wie bei einem normalen Java-Projekt. Die Attribute name, description und url sind auch Teil von Standard-Projekten. Bei der URL ist schon das GitHub-Projekt eingetragen. Neu sind die minimale Version von Maven und das Jahr, ab dem das Plug-in eingeführt wurde.

Des Weiteren sind beim Build auch noch Änderungen vorzunehmen (siehe Listing 3).

Das Maven-Surfire-Plug-in und das Maven-Compiler-Plug-in müssen neuere Versionen als Standard verwenden, damit mit

```
<dependency>
  <groupId>org.apache.maven</groupId>
  <artifactId>maven-plugin-api</artifactId>
  <version>3.6.0</version>
  <scope>provided</scope>
</dependency>
<dependency>
  <groupId>org.apache.maven</groupId>
  <artifactId>maven-core</artifactId>
  <version>3.6.0</version>
  <scope>provided</scope>
</dependency>
<dependency>
  <groupId>org.apache.maven.plugin-tools</groupId>
  <artifactId>maven-plugin-annotations</artifactId>
  <version>3.6.0</version>
  <scope>provided</scope>
</dependency>
```

Listing 1: Die Abhängigkeiten eines Maven Plug-ins

```
<packaging>maven-plugin</packaging>

<name>Plugin to generate jars with module information as Maven Plugin</name>
<description>The maven main core project description</description>
<url>http://github.com/mt-ag/jar-module-maven-plugin</url>

<prerequisites>
  <maven>3.6.0</maven>
</prerequisites>
<inceptionYear>2019</inceptionYear>
```

Listing 2: Die notwendigen Einstellungen zum Erstellen eines Maven Plug-ins

```
<build>
  <plugins>
    <plugin>
      <artifactId>maven-surefire-plugin</artifactId>
      <version>2.22.0</version>
    </plugin>
    <plugin>
      <groupId>org.apache.maven.plugins</groupId>
      <artifactId>maven-compiler-plugin</artifactId>
      <version>3.8.1</version>
      <configuration>
        <source>${java.version}</source>
        <target>${java.version}</target>
      </configuration>
    </plugin>
    <plugin>
      <groupId>org.apache.maven.plugins</groupId>
      <artifactId>maven-plugin-plugin</artifactId>
      <version>3.6.0</version>
    </plugin>
    <plugin>
      <groupId>org.apache.maven.plugins</groupId>
      <artifactId>maven-site-plugin</artifactId>
      <version>3.7.1</version>
    </plugin>
    <plugin>
      <groupId>org.apache.maven.plugins</groupId>
      <artifactId>maven-project-info-reports-plugin</artifactId>
      <version>3.0.0</version>
    </plugin>
  </plugins>
</build>
```

Listing 3: Die Änderungen beim Build

```
<groupId>com.mt-ag.tools.maven</groupId>
```

Listing 4: Die GroupId

```
<scm>
  <connection>scm:git:git://github.com/mt-ag/jar-module-maven-plugin.git</connection>
  <developerConnection>scm:git:ssh://github.com:mt-ag/jar-module-maven-plugin.git</developerConnection>
  <url>http://github.com/mt-ag/jar-module-maven-plugin/tree/master</url>
</scm>
```

Listing 5: Die SCM-URL zum GitHub-Projekt

```
<plugin>
  <groupId>org.apache.maven.plugins</groupId>
  <artifactId>maven-javadoc-plugin</artifactId>
  <version>3.1.0</version>
  <configuration>
    <quiet>true</quiet>
    <nohelp>true</nohelp>
    <additionalOptions>-html5</additionalOptions>
  </configuration>
  <executions>
    <execution>
      <id>make-assembly</id>
      <phase>package</phase>
      <goals>
        <goal>jar</goal>
      </goals>
    </execution>
  </executions>
</plugin>
```

Listing 6: JavaDoc generieren

```
<plugin>
  <groupId>org.apache.maven.plugins</groupId>
  <artifactId>maven-source-plugin</artifactId>
  <version>3.0.1</version>
  <executions>
    <execution>
      <id>attach-sources</id>
      <phase>verify</phase>
      <goals>
        <goal>jar</goal>
      </goals>
    </execution>
  </executions>
</plugin>
```

Listing 7: Java Source als Zip-Artefakt

```
<plugin>
  <groupId>net.nicoulaj.maven.plugins</groupId>
  <artifactId>checksum-maven-plugin</artifactId>
  <version>1.8</version>
  <executions>
    <execution>
      <id>checksum-artifacts</id>
      <phase>verify</phase>
      <goals>
        <goal>files</goal>
      </goals>
      <configuration>
        <attachChecksums>true</attachChecksums>
        <csvSummary>>false</csvSummary>
        <fileSets>
          <fileSet>
            <directory>target</directory>
            <include>*.jar</include>
            <include>*.pom</include>
          </fileSet>
        </fileSets>
      </configuration>
    </execution>
  </executions>
</plugin>
```

Listing 8: Checksummen für Artefakte

Java 9, 10 oder 11 gebaut werden kann. Die beiden anderen Maven Plug-ins werden benötigt, um das neue Plug-in zu bauen und um die Dokumentation zu generieren.

Vorgaben vom Maven Central Repository

Aktuell gelten die Vorgaben von Sonartype [5], hier die wichtigsten:

1. Metadaten nach Vorgabe
2. JavaDoc
3. Java Sources
4. Checksummen
5. Signaturen mit GPG/PGP
6. Repositories von Sonartype

Metadaten nach Vorgabe heißt, der Package-Name in Java beginnt mit der umgekehrten URL der Firma oder Organisation, bei mir ist es com.mt_ag, abgeleitet

von `www.mt-ag.com`. Der `,` kommt als Ersatz für das nicht erlaubte `,` für Java-Packages. Der Name der GroupId beginnt auch mit der umgekehrten URL ähnlich wie beim Java-Package (siehe Listing 4).

Hier ist das `,` erlaubt, deshalb nicht `,` in der GroupId.

Der SCM-Eintrag muss auch angegeben werden (siehe Listing 5).

Des Weiteren sind noch die Lizenz, die Organisation und die Entwickler anzugeben. Die Lizenz muss eine Open-Source-Lizenz sein.

Das JavaDoc kann noch einfach mit einem Plug-in generiert werden (siehe Listing 6).

Für Java 9 und 10 ist die zusätzliche Option `-html5` zu setzen, bei Java 11 nicht mehr.

Java Source wird mit einem Plug-in als Zip hinzugefügt (siehe Listing 7).

Vorgegeben von Sonatype ist die Verwendung von Checksummen, dies geschieht mit diesem Plug-in (siehe Listing 8).

Wichtig hierbei ist, dass nur die Jars und die POM Checksummen bekommen. Es wird keine Zusammenfassung benötigt.

Die Artefakte und POM müssen noch signiert werden, dafür wird das PGP-Plug-in verwendet (siehe Listing 9).

Hier gibt es noch eine gute Beschreibung für GitHub Actions, siehe [8]. Der Private Key zum Signieren müsste noch in der `settings.xml` hinterlegt werden. Das ist jedoch zu unsicher, da man die `settings.xml` noch zum Projekt hinzufügen muss, um sie zu verwenden. Auch wenn man das Passwort mit dem Mitteln von Maven verschlüsseln kann, ist dies nicht zu empfehlen, da dann auch noch die Datei `settings-security.xml` Teil des Projektes würde.

Code-Qualität

Hierfür gibt es keine besonderen Vorgaben. Es empfiehlt sich, Checkstyle und ja-coco zu verwenden und passende Tests zu hinterlegen.

Sicherheit

Es muss darauf geachtet werden, dass keine Zugangsdaten miteingecheckt werden. Besonders nicht in der Konfiguration, wie zum Beispiel `settings.xml`, oder in der Dokumentation.

```
<plugin>
  <groupId>org.apache.maven.plugins</groupId>
  <artifactId>maven-gpg-plugin</artifactId>
  <version>1.4</version>
  <executions>
    <execution>
      <id>sign-artifacts</id>
      <phase>verify</phase>
      <goals>
        <goal>sign</goal>
      </goals>
      <configuration>
        <skip>>false</skip>
      </configuration>
    </execution>
  </executions>
</plugin>
```

Listing 9: GPG Signing

```
name: Java Ubuntu Build

on:
  push:
    branches:
      - master

jobs:
  build:

    runs-on: ubuntu-latest

    steps:
      - uses: actions/checkout@v2
      - name: Set up JDK 11
        uses: actions/setup-java@v1
        with:
          java-version: 11
      - name: Cache Maven packages
        uses: actions/cache@v1
        with:
          path: ~/.m2
          key: ${{ runner.os }}-m2-${{ hashFiles('**/pom.xml') }}
          restore-keys: ${{ runner.os }}-m2
      - run: chmod a+x mvnw
      - name: Maven clean
        run: ./mvnw clean --file pom.xml
      - name: Maven compile
        run: ./mvnw compile --file pom.xml
      - name: Maven test
        run: ./mvnw test --file pom.xml
      - name: Maven install
        run: ./mvnw install -Dmaven.test.skip=true --file pom.xml
      - run: mkdir staging && cp target/*.jar staging
      - uses: actions/upload-artifact@v1
        with:
          name: PluginJars
          path: staging
```

Listing 10: GitHub-Konfiguration

Die CI/CD-Konfiguration

Unter GitHub wird die Konfiguration durch die `maven.yml`-Datei vorgenommen (siehe Listing 10).

Hier werden für den Cache die Emp-

fehlungen von GitHub berücksichtigt und der Maven Build wird in mehrere Stages unterteilt.

Unter GitLab [9] geschieht die Konfiguration mit der `.gitlab-ci.yml`-Datei (siehe Listing 11).

Beide Konfigurationen stellen am Ende der Pipeline ein Artefakt mit den Jars zur Verfügung. Ein direktes Deployen ins Maven Central Repository wäre zwar möglich, es wäre mir allerdings noch zu unsicher, die Zugangsdaten zu hinterlegen.

Zusammenfassung

Bei Open-Source-Projekten ist viel im Vorfeld zu beachten und es gibt eine Auswahl von guten Webseiten, um ein Projekt zu veröffentlichen. Dabei wird auch eine CI/CD-Pipeline direkt unterstützt. Es ist nicht mehr nötig, die CI/CD-Pipeline über eine andere Webseite anzubinden. Schön bei GitHub ist, dass die generierte Dokumentation gleich mitveröffentlicht werden kann (*siehe im Beispielprojekt [1]*).

Quellen

- [1] <https://github.com/mt-ag/jar-module-maven-plugin> Das Maven Plug-in des Artikels
- [2] <https://maven.apache.org/> DasApache-Maven-Projekt
- [3] <https://maven.apache.org/plugin-developers/index.html> Die Plug-ins für Apache Maven
- [4] <https://maven.apache.org/repository/index.html> Das Maven Central Repository
- [5] <https://central.sonatype.org/pages/requirements.html> Die Vorgaben von Sonatype für Maven Repositories
- [6] <https://github.com/> Das GitHub-Projekt
- [7] https://github.com/mt-ag/jar-module-maven-plugin/new/master?filename=.github%2Fworkflows%2Fmaven.yml&workflow_template=maven Hier die Action, um Maven aufzurufen
- [8] <https://help.github.com/en/actions/language-and-framework-guides/publishing-java-packages-with-maven> Hier die Action, um auf Maven Central Repository zu veröffentlichen
- [9] <https://gitlab.com/> Das GitLab-Projekt

Über den Autor

Wolfgang Nast arbeitet seit der Version 1.0 mit Java und hat die unterschiedlichsten Unternehmen bei der Einführung und Umsetzung mit Java begleitet. Aktuell unterstützt er bei der Einführung von Micro-Services und CI/CD-Pipelines.



Wolfgang Nast
Wolfgang.nast@mt-ag.com

```
image: openjdk:11.0.6-jdk

variables:
  MVN_OPTS: "--batch-mode -Dmaven.repo.local=.m2/repository"

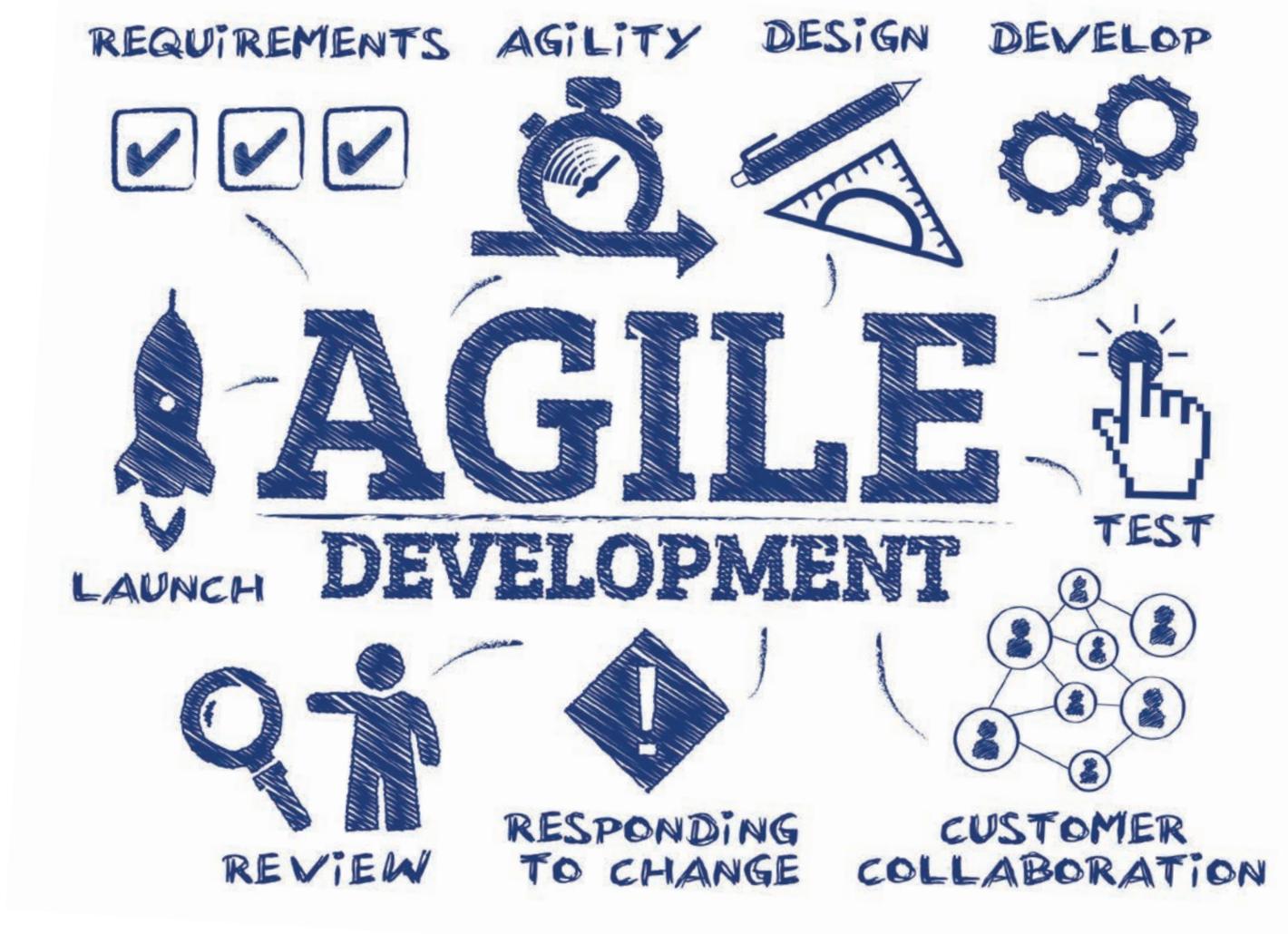
cache:
  paths:
    - .m2/repository/
    - target/

build:
  stage: build
  script:
    - chmod a+x ./mvnw
    - ./mvnw $MVN_OPTS compile
  artifacts:
    paths:
      - target

test:
  stage: test
  script:
    - chmod a+x ./mvnw
    - ./mvnw $MVN_OPTS test

deploy:
  stage: deploy
  script:
    - chmod a+x ./mvnw
    - ./mvnw $MVN_OPTS -Dmaven.test.skip=true install
  artifacts:
    paths:
      - target/*.jar
  only:
    - master
```

Listing 11: GitLab-Konfiguration



DevSecOps aber sicher! Kontinuierlich sicherer entwickeln

Frank Pientka, Materna

Die Softwareentwicklung hat sich die letzten Jahre dynamisch weiterentwickelt. Agile Methoden haben vor allem die Zusammenarbeit zwischen Business und Development verbessert, sodass Anforderungen schneller und besser getestet umgesetzt werden konnten. DevOps verbessern den gesamten Bau-, Test- und Auslieferungsprozess.

Vom agilen zum DevSecOps-Wandel

Letztendlich geht es darum, mehr geschäftlichen Wert bei reduziertem Risiko und gleichbleibender Qualität schneller zum Einsatz und zum Kunden zu bringen.

Oft konnte die Software zwar kontinuierlich gebaut und integriert werden, doch wurde sie auf dem Weg in die Produktion meist ausgebremst. Das lag oft

an den unflexiblen und wenig automatisierten Betriebsprozessen.

Die Wand zwischen Dev(elopment) und Op(eration)s, die einer optimalen Unterstützung der Geschäftsprozesse im Wege steht, existiert trotz agiler Entwicklungsprozesse leider oft immer noch und wurde um eine weitere Wand, das vorgelagerte DevOps-Team, sogar noch verstärkt.

Die Idee, dass interdisziplinäre Teams Verantwortung für die kompletten Ent-

wicklungs- und Betriebsprozesse ihrer Services übernehmen, setzt sich trotzdem immer mehr unter dem DevOps-Gedanken in vielen Unternehmen durch. Leider ist gerade beim Thema Sicherheit eine gewisse „Müdigkeit“ [1] eingetreten. Das liegt an den gerade durch die Cloud und verteilte Microservices gestiegenen Anforderungen, aber auch an den durch die Agilität beschleunigten Prozessen, so dass oft mehrere Versionen eines Dienst-

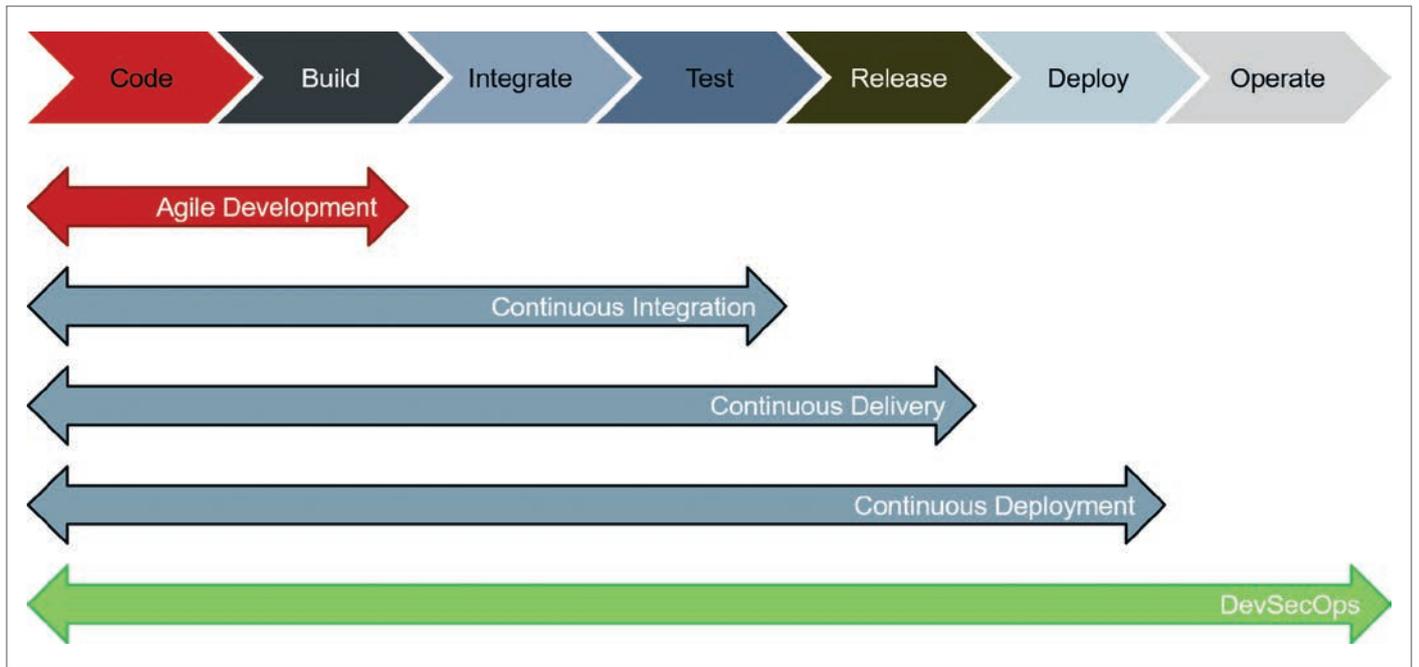


Abbildung 1: DevSecOps: Einordnung in die agile Softwareentwicklung (Quelle: Frank Pientka)

tes parallel betrieben werden können. Verschärfend kommt zum Tragen, dass Sicherheit weder beim Betrieb noch bei der Entwicklung bisher eine große Priorität hatte, sodass sowohl das Wissen, wie man damit umzugehen hat, als auch die entsprechenden Werkzeuge nicht weit verbreitet waren.

Beim Thema Sicherheit handelt es sich um das schon klassische Hase- und-Igel-Phänomen, dass der Verteidiger seinen

kriminellen Herausforderern oft hinterherhinkt.

Deswegen ergibt ein einmaliger und extern durchgeführter Pentest am Ende der Releasefreigabe wie bei Wasserfallprojekten für Monolithen heute keinen Sinn mehr, da Sicherheit permanent und über alle Phasen gewährleistet sein muss. Nur so kann kontinuierliches Deployment in Produktion mit einer hohen Frequenz stattfinden.



Abbildung 2: Welche Disziplinen spielen bei DevSecOps eine Rolle? (Quelle: Frank Pientka)

DevOps ist das wahre Agile - Entwickelst du noch oder lieferst du schon?

Aktuelle Auswertungen [2, 3] im jährlich stattfindenden „State of DevOps Report“ zeigen eindrücklich, wie Firmen, die schon länger konsequent auf DevOps setzen, erhebliche Effizienz- und Qualitätsvorteile gegenüber anderen Marktbegleitern haben. Nicht nur, dass ihre Teams konsequenter schneller und öfter liefern. Sie können auch schneller auf Fehlersituationen reagieren und das echte Feedback unter Live-Bedingungen in die Weiterentwicklung einfließen lassen. Im Gegensatz zu agilen Methoden lassen sich die Erfolge der Einführung von DevOps-Prinzipien durch das Messen von vier Hauptmetriken (Lead Time, Change Fail, Deployment Frequency, Time to Restore) aus dem sechsjährigen Forschungsprogramm des DevOps Research and Assessment (DORA) sogar wissenschaftlich belegen.

Mit dem Building Security In Maturity Model (BSIMM) [7, 8] gibt es sogar ein verlässliches Modell, das man auf seine DevSecOps-Pipeline anwenden kann und mit dem man sich in zwölf Gebieten bewerten und mit den Daten anderer Unternehmen vergleichen kann.

DevSecOps als nächste Stufe von DevOps

Letztendlich ergänzt DevSecOps die Entwicklungs- und Betriebsprozesse um

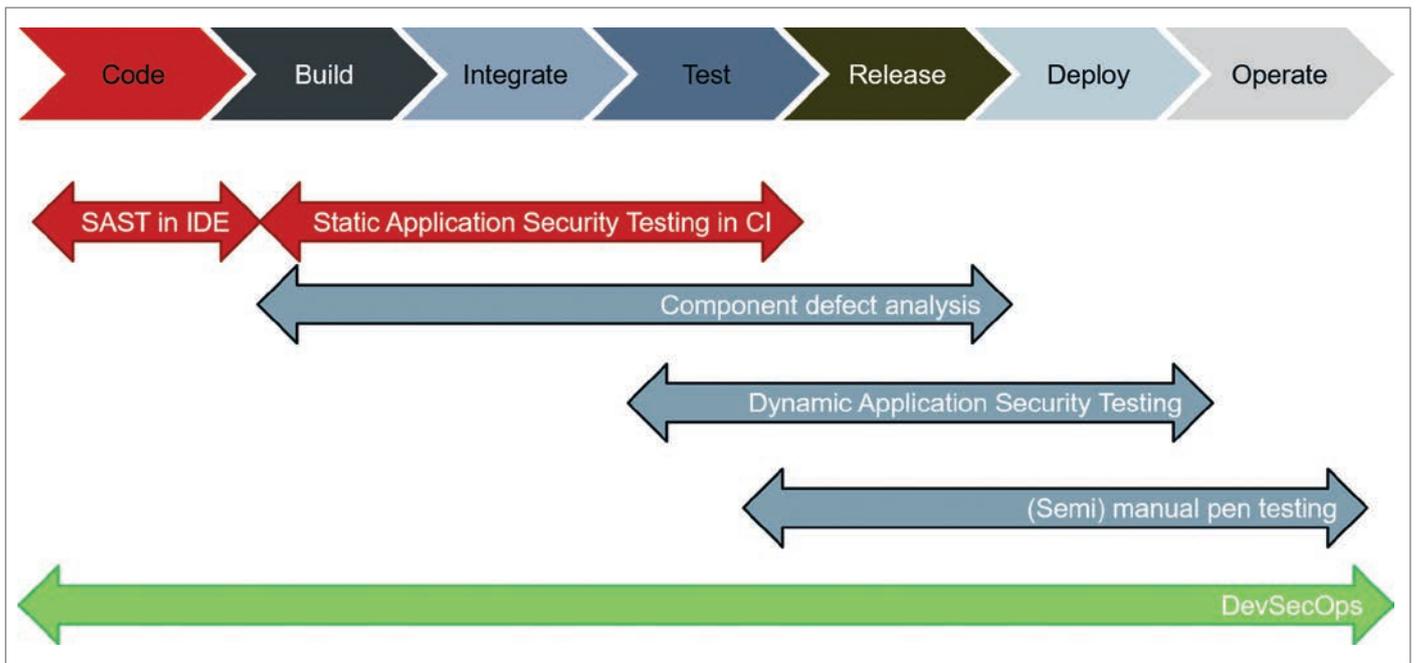


Abbildung 3: Welche Methoden können bei DevSecOps in welcher Phase eingesetzt werden? (Quelle: Frank Pientka)

zusätzliche möglichst weitgehend automatisierte Qualitätsprozesse (siehe Abbildung 2). Dadurch wird ein höheres Maß an Sicherheit gewährleistet und es hilft, Kopf-Monopole sowie halbherzig und sporadisch durchgeführte Sicherheitsmaßnahmen zu vermeiden.

Durch das Aufkommen von Containern und der Tatsache, dass die Bereitstellung von Infrastruktur vom DevOps-Team selbst übernommen wird, kommen dafür immer mehr bereits in der Entwicklung bewährte Prinzipien und Werkzeuge auch im Betrieb (GitOps, Infrastructure-as-Code) zum Einsatz. Zusätzlich gibt es jedoch mehr Abhängigkeiten und Werkzeuge, die integriert, konfiguriert und gepflegt werden müssen (siehe Abbildung 3).

Der Shift-Left-Ansatz setzt auf die bei jeder Code-Änderung automatisch durchgeführten statischen Anwendungssicherheitstests (Static Application Security Testing, SAST) in den frühen Entwicklungsphasen. Oft werden diese bereits in der Umgebung des Entwicklers und zusätzlich auf dem nachgelagerten Buildserver ausgeführt. Sie werden in der nächsten Phase um dynamische Anwendungssicherheitstests (Dynamic Application Security Testing, DAST) ergänzt. Manchmal kommen noch interaktive Anwendungssicherheitstests (Interactive Application Security Testing, IAST) und Runtime-Anwendungsselbstschutz (Run-

time Application Self-Protection, RASP) hinzu.

Die am häufigsten eingesetzten Sicherheitstests sind SAST und DAST, da hier schon länger ausgereifte und einfach zu verwendende Werkzeuge existieren. Bei SAST werden der Quellcode, Bibliotheken oder Binärdateien auf bereits bekannte Schwachstellen überprüft.

Gerade Docker-Container altern oft schneller als gedacht und die dort verwendeten Bibliotheken enthalten innerhalb kurzer Zeit oft Hunderte Schwachstellen nach der Erstellung eines Container-Image. Deswegen müssen diese regelmäßig auf Sicherheitslücken überprüft und aktualisiert werden.

Da gerade Open-Source-Komponenten sehr zügig bekannte Sicherheitslücken beheben, bietet es sich an, die auszuliefernden Dienste automatisch mit aktuellen Versionen der extern verwendeten Komponenten parallel zu bisherigen Versionen in der Regression zu testen und dann wieder in den Hauptzweig zu integrieren, wenn sie sich als stabil erwiesen haben. So bleibt die ausgelieferte Software immer frisch. Man kann sich auf das Liefern neuer Funktionen konzentrieren und das Anhäufen von technischen Schulden vermeiden. Eine frühzeitige Codehygiene-Vorsorge vermeidet spätere teure Sicherheitskorrekturen und zahlt sich auch langfristig aus, da sie bereits Teil des normalen Auslieferprozesses ist.

Shift-Left und Shift-Right gleichzeitig (siehe Abbildung 4)

Doch auch die DevOps-Pipeline möchte gepflegt werden. Dies ist beim Aufbau ein wichtiges Thema, wo die Zugangsdaten auf die verschiedenen Umgebungen mit Rechten vergeben und verwaltet werden (Shift-Right). Ebenso sollen Erkenntnisse aus erkannten Schwachstellen in der Produktion auch wieder zurück in die Entwicklungspipeline fließen. Beim Zusammenstecken der DevOps-Pipeline kann man entweder die besten Werkzeuge für die jeweilige Phase und Aufgabe von verschiedenen spezialisierten Herstellern oder eine integrierte Pipeline von einem Hersteller, oft basierend auf dem Git-Versionierungsprodukt, einsetzen. Der Best-of-Breed-Ansatz ist sicher in der Breite und Tiefe das bevorzugte Modell. Er kommt jedoch mit zusätzlichen Integrations- und Pflegekosten, da man bei jedem Update einer Teilkomponente prüfen muss, ob diese mit den Nachbarkomponenten noch korrekt funktioniert. Deswegen bietet sich der integrierte Ansatz für die DevOps-Pipeline von einem Hersteller (alles aus einer Hand) an, um die Abhängigkeiten und Pflegeaufwände zu reduzieren. Der integrierte Ansatz, wie ihn die sehr populären Plattformen GitLab und GitHub anbieten, hat den weiteren Vorteil, dass diese als Container vorliegen und so einfacher in der Private

und Public Cloud ohne Zusatzkosten eingesetzt werden können. Deswegen sind diese für eine Cloud-native-Entwicklung in einer Multi-Cloud-Welt besser geeignet als die in der On-Premises-Welt bewährten DevOps-Produkte (siehe Abbildung 5).

Regeln für erfolgreiche DevSecOps

Als Regeln für den erfolgreichen Einsatz von DevSecOps-Prinzipien und -Werkzeugen lässt sich zusammenfassen:

- Beginne mit Sicherheit von Anfang an
- Versuche möglichst viele Sicherheitstest über alle Phasen zu automatisieren
- Halte deine DevSecOps-Pipeline einfach, sauber, aktuell und sicher
- Überwache kontinuierlich deine Software-Komponenten auf bekannte Schwachstellen
- Kenne deine Abhängigkeiten und überprüfe diese regelmäßig
- Gieße deine Sicherheitsrichtlinien in automatisiert ausführbare Prüfregrn
- Vertraue nur geprüften Quellen, fische nicht in trüben Gewässern

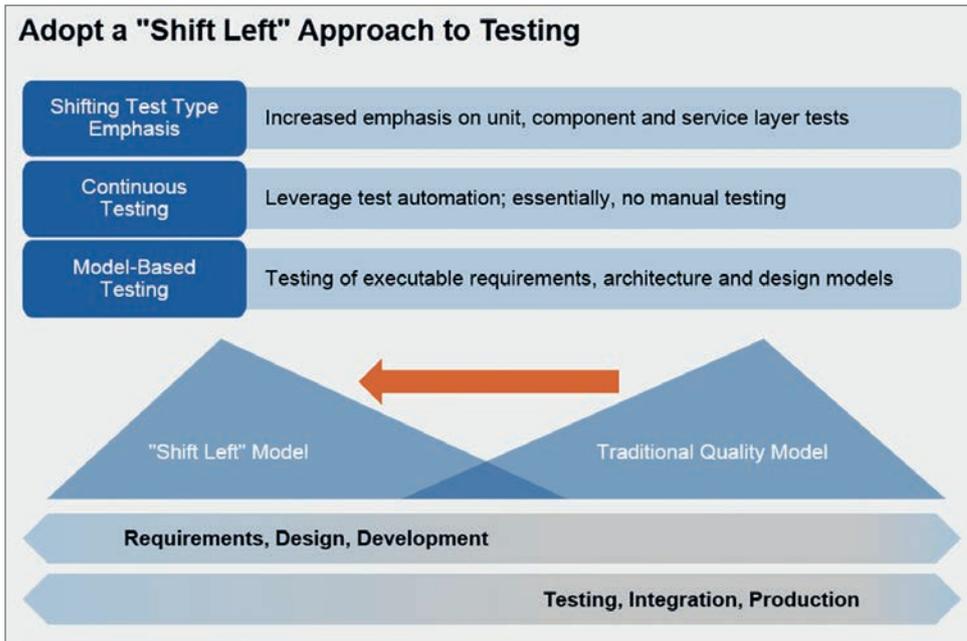


Abbildung 4: Der Shift-Left-Ansatz nach Gartner (Quelle: Frank Pientka)

SAST		cloudbees Plugins	
DAST		cloudbees Plugins	
Secrets Detection		cloudbees Plugins	
Dependency Scanning		cloudbees Plugins	
Container Scanning		cloudbees Plugins	
License Compliance		cloudbees Plugins	
Vulnerability Database		cloudbees Plugins	
IAST		cloudbees	
Fuzzing		cloudbees	

Abbildung 5: Vergleich der Sicherheitsfunktionen von Jenkins und GitLab [6] (Quelle: GitLab)

DevSecOps bedeutet weniger Risiko und mehr Konzentration auf das Wesentliche

Mit DevSecOps steht der nächste größere Schritt an, um schneller auf Änderungen oder Sicherheitslücken reagieren zu können.

Der durch DevSecOps eingeleitete Kulturwandel (Shift-Left) [4, 5] ist ein wichtiger Baustein auf dem Weg in die Cloud. Nur wer DevSecOps wirklich lebt, kann Cloud-native-Services kontinuierlich liefern und verbessern.

Da Sicherheit der Weg und nicht ein Ziel ist, kann man sich diesem nur schrittweise annähern. Hier schaffen DevSecOps-Pipelines mit ihren vorkonfigurierten Werkzeugen und Prozessen ein Netz von Tests und Best Practices, von denen jeder profitieren kann. Dadurch, dass Sicherheitstests in allen Phasen automatisiert durchgeführt werden, und die Bestrebung, durch den Shift-Left- und Shift-Right-Ansatz diese Tests immer früher und auch für die DevOps-Pipeline auszuführen, kann man Fehler früher erkennen und damit günstiger und schneller beheben. So lässt sich die Lücke zwischen bekannten Schwachstellen und ihrer Behebung erheblich besser schließen. Außerdem vermeidet man das Anhäufen von technischen oder Design-Schulden, indem man kontinuierlich und systematisch (security-by-design, by-default) das Sicherheitsrisiko minimiert.

Der durch DevSecOps eingeleitete Kulturwandel ist ein wichtiger Baustein auf dem Weg in die Cloud.

Quellen

- [1] Brian Stanton et. Al. Security Fatigue, IEEE IT Professional, Issue: 5 2016
- [2] Forsgren / Humble / Kim, Das Mindset von DevOps. Accelerate, Vahlen, 2019
- [3] Accelerate: State of DevOps 2019, <https://services.google.com/fh/files/misc/state-of-devops-2019.pdf>
- [4] Neil MacDonald, Ian Head, DevSecOps: How to Seamlessly Integrate Security Into DevOps, Gartner, 2016
- [5] Joachim Herschmann, Jim Scheibmeir Take a 'Shift Left' Approach to Testing to Accelerate and Improve Application Development, Gartner, 2019
- [6] Vergleich Sicherheitsfunktionen Jenkins vs. GitLab <https://about.gitlab.com/devops-tools/jenkins-vs-gitlab.html>
- [7] Building Security In Maturity Model (BSIMM) Report <https://www.bsimm.com/content/dam/bsimm/reports/bsimm10.pdf>

- [8] BSIMMsc Applying the BSIMM to the software supply chain <https://www.bsimm.com/about/bsimm-for-vendors.html>

Über den Autor

Frank Pientka (@fpientka) arbeitet als Dipl.-Informatiker (TH Karlsruhe) bei der MATERNA in Dortmund. Als Principal Software Architect sorgt er für mehr Qualität in der Software und kümmert sich, als Gründungsmitglied des iSAQB, um eine verbesserte Ausbildung und Zertifizierung von Architekten.

Seit mehr als drei Jahrzehnten unterstützt er Firmen bei der Umsetzung tragfähiger Software-Architekturen.

Dabei führt er externe Reviews und Architekturbewertungen als Gutachter durch. Dazu hat er auch schon mehrere Fachartikel veröffentlicht und Vorträge gehalten.



Dipl. Inform. Frank Pientka
Frank.Pientka@materna.de



Oracle Database In-Memory ab sofort bis zu 16 GB kostenfrei

DOAG-Redaktion

Die Option steht den Nutzern ab der Version 20c als „Base Level“-Feature zur Verfügung.

Oracle Database In-Memory wurde 2014 mit der Oracle Database 12c als kostenpflichtige Option eingeführt. Mit ihrer spaltenorientierten Datenhaltung im Hauptspeicher können für bestimmte Workloads Performance-Vorteile erzielt werden. In einem Blogpost hat Oracle Product Manager Andy Rivenes nun bekanntgegeben, dass es mit der Veröffentlichung

von Oracle Database 20c erstmals ein kostenfreies Feature der Option geben wird.

Das "Base Level"-Feature ermöglicht die Nutzung von Database In-Memory mit bis zu 16 GB Column Store, ohne dass die Option hierfür lizenziert werden muss. Die 16-GB-Grenze gilt auf der Ebene der Container-Datenbank (CDB). Somit nutzen alle Pluggable Databases (PDBs)

diese Grenze gemeinsam. Bei RAC-Datenbanken ermöglicht das Feature jedoch die Zuweisung eines 16 GB Column Store auf jede RAC-Instanz.

Auf YouTube finden Sie eine Zusammenfassung der Ask TOM Office Hours Session, in der das „Base Level“-Feature genauer beschrieben wird (<https://youtu.be/yoHIV7eH708>).



Cyber-Defense mit und in der Oracle Cloud

Michael Fischer, Oracle Deutschland

Die Sicherheit und Verfügbarkeit von Rechenzentren und Cloud-Diensten hat für Oracle als weltweitem Anbieter von Cloud Services höchste Priorität. Um dies zu gewährleisten, kommen in der Oracle Cloud verschiedene Security-Ansätze zum Tragen. Neben den Basismechanismen wie DDoS-Abwehr oder Verschlüsselung sind dies das Prinzip der Isolierung, Security-by-Default in der Cloud, dedizierte Security Services, die Automatisierung für die Wahrung der Security und Abwehr von Angriffen. Alle diese Mechanismen sind Bestandteile der Cyber-Defense der Oracle Cloud, die im folgenden Artikel kurz vorgestellt werden.

Zu aktuellen Bedrohungen, inklusive der erfolgreichen unrechtmäßigen Datenabflüsse (Threats und Breaches), gibt es unzählige Aufstellungen verschiedener renommierter Sicherheitsdienstleister und Hersteller. Der Grundtenor aller Berichte ist die stetig stark steigende Anzahl von Angriffen, die zunehmende Intelligenz dabei und letztendlich auch deren Erfolg. Die Einschätzung von Befragten bezüglich einer doch erfolgreich implementierten IT-Sicherheit beziehungsweise Cyber-Defense wird in vielen Fällen durch die Wirklichkeit

widerlegt. Sehr frappierend wiegt dabei die Analyse, dass viele Angriffe nur deswegen erfolgreich waren, weil sie ungepatchte Systeme – teils mit seit fünf Jahren nicht eingespielten Patches – kompromittieren konnten. Dies ist in den einschlägigen Data Breach Reports dargestellt.

Die Angriffsvektoren werden typischerweise nicht mehr singulär ausgeführt, sondern im Rahmen eines Angriffes orchestriert. Dem eigentlichen Angriff geht dabei eine teils lang andauernde Auskundschafts- und Analysephase voraus. Am

Beispiel von Emotet lässt sich das aktuell nachverfolgen. Nach wie vor ist die Entdeckung von Breaches nicht wirklich effektiv, die einschlägigen Studien sehen bei der Erkennung einen Durchschnitt von 100-150 Tagen. Diese Zahl hat sich nur geringfügig über die letzten Jahre verkürzt.

IT-Sicherheit besteht aus verschiedenen Mechanismen, die für erfolgreiche Cyber-Defenses reibungslos zusammenspielen müssen. Statische Abwehrmechanismen müssen mit kontextbasierten Mechanismen unter Einbezug von

künstlicher Intelligenz (Machine Learning) fortwährend sich selbst aktualisieren und weiterentwickeln. Über die typischen Angriffsvektoren, wie in *Abbildung 1* dargestellt, lassen sich die Cyber-Defense Controls in und mit der Oracle Cloud nachvollziehen:

Die Betrachtung der Threats und zugeordneten Cyber-Defense Controls wird im Uhrzeigersinn beschrieben:

- Internal Threats:** Die ältesten und bekanntesten Threats sind die, die durch Innentäter verursacht werden. Beispielsweise sind das unzufriedene Mitarbeiter, die einfach die Daten abgreifen, die durch sie zugreifbar sind. Möglich durch mangelnden Schutz auf Verzeichnisebene oder auch dadurch, dass beispielsweise Datenbanken nicht verschlüsselt und somit auf Fileebene oder im Backup lesbar sind. Die Verteilung von Innentätern zu Abgriffen über „External Threats“ ist über die Jahre gleich geblieben. Dies zeigt, dass hier durchaus Lücken sind. Interne Abgriffe werden durch das sogenannte „Defense-in-Depth“-Konzept abgewehrt. Dies ist das älteste Konzept und Oracle-Kunden seit über 20 Jahren bekannt. Es erfolgen eine Reihe von Maßnahmen, die den Zugriff kontrollieren beziehungsweise ein-

schränken. Dazu gehört eine grundlegende Authentifizierung (Login), Autorisierung beim Zugriff (Least Privilege, SoD), Audit (Kontrolle) und Verschlüsselung der Daten/Datenbank. So können Daten nicht mehr einfach abgegriffen werden. Es kommt zudem eine kontextbasierte Ausprägung zum Tragen, um beispielsweise eine Steuerung von Zugriffen während Bürozeiten oder im Firmennetz versus außerhalb zu unterscheiden.

- Top-Down Threats:** In diesem Bereich finden sich Threats, die sich durch gezieltes Ausnutzen von Sicherheitslücken über das Netzwerk verbreiten. Einmal eingefangen durch zum Beispiel präparierte Mails, die dann Schadcode einpflanzen oder Accounts ausspionieren. Die Abwehr erfolgt hier durch Isolation sowohl auf Server- und Hypervisor- als auch auf Netzwerkebene. Angriffe werden so in ihrer Verbreitung (Lateral Movement) verhindert. Mehr dazu im nächsten Kapitel. Neben den dargestellten Maßnahmen kommen hier auch IDS- und IPS-Systeme zum Ansatz (siehe Abschnitt SOC).

- Bottom-Up Threats:** Von „unten“ kommen die Angriffe, die sich über infizierte BIOS (Server, Controller, HW-Erweiterungen usw.) oder kompromit-

tierte Betriebssysteme (Server, Router, ...) verbreiten.

Eine solche Kontamination wird schlicht verhindert, indem beim Rücksetzen oder Neu-Aufsetzen von Systemen kryptografisch gesicherte BIOS und Betriebssystemimages aufgespielt werden. Diese werden in einem separaten Netz auf speziellen Devices vorgehalten.

- External Threats:** Diese Angriffe erfolgen beispielsweise über Hacker oder Bot-Netzwerke, die systematisch Schwächen prüfen und ausnutzen. Erstaunlicherweise steigt die Anzahl der erfolgreichen Angriffe, obwohl die Abwehrmaßnahmen seit vielen Jahren bekannt sind.

Neben dem selbstverständlichen Sicherheits-Patching verhindern hier verschiedene Mechanismen das Eindringen von außen. Dazu gehören eine Abwehr „schlechter“ Bots, Erkennung von verdächtigem Verhalten (UEBA – User Entity Behaviour Analytics mithilfe von Machine Learning), Erkennung und Abwehr der OWASP Top 10, eigene Regelwerke sowie ein permanentes Monitoring des Internets und dessen Auswertung beziehungsweise Einbezug der Erkenntnisse darüber. Beispielsweise, um einer ungewollten Routing-Umleitung zu entgehen.

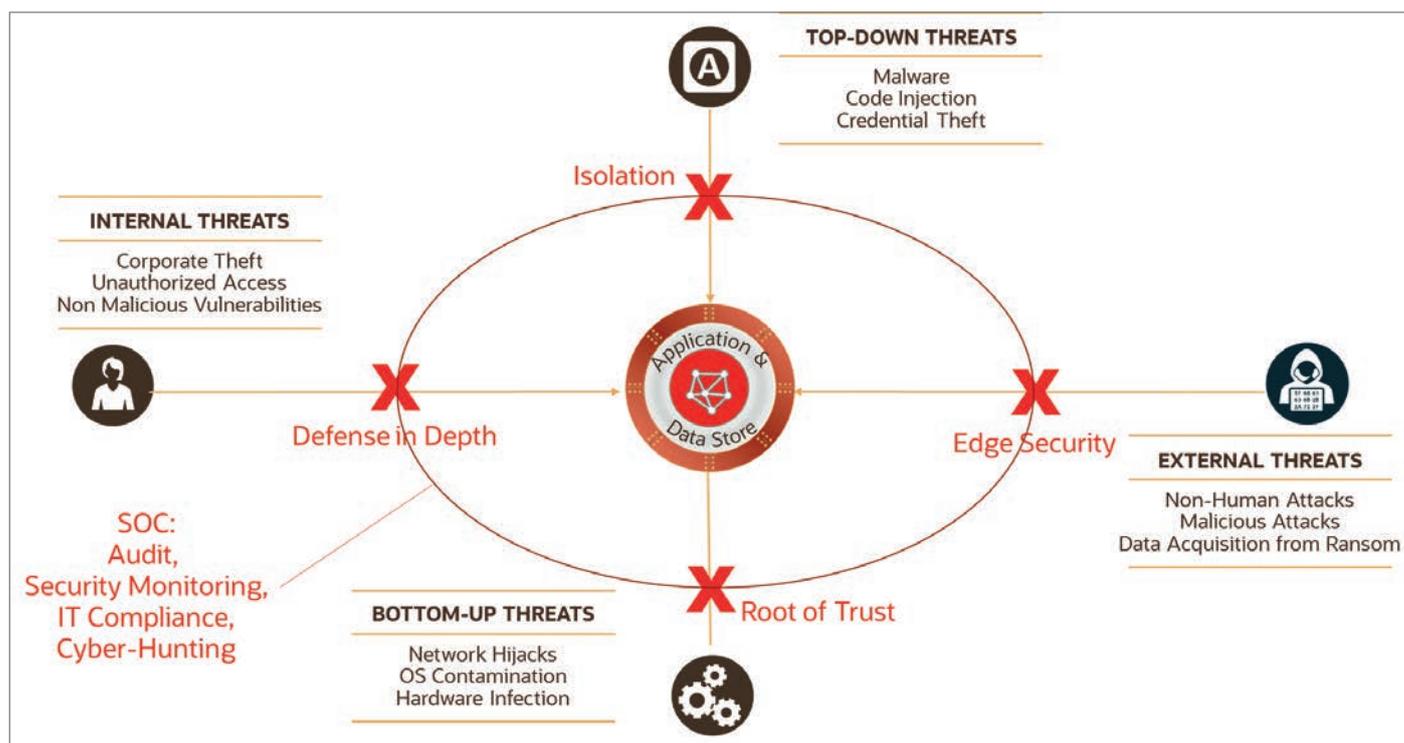


Abbildung 1: Typische Angriffsvektoren und IT-Schutzmaßnahmen (Quelle: Michael Fischer)

SOC (Security Operations Center): Audit, Security Monitoring, IT Compliance:

Viele Angriffe können nur durch die Auswertung einer Vielzahl von beteiligten Systemen erkannt werden. Dies erfolgt in diesem Bereich klassischerweise durch SIEM- und IT-Compliance-Systeme. Externe Quellen werden hier mit angebunden.

Die aufgeführten Mechanismen zeigen die Cyber-Defense Controls in der Oracle Cloud. Nicht dargestellt sind die in der Cloud vorhandenen Mechanismen zur Verfügbarkeit der Services (HA, DR, Backup/Restore) sowie eine transparente redundante Speicherung und redundante Netzwerkanbindung. Die Security-Prinzipien der Cloud lassen sich zusammenfassen als „Security First & Security by Default“. Dies ist das Design der Oracle Cloud Generation 2 wie zum Beispiel in Frankfurt; es implementiert damit einen durchgängigen Sicherheitsansatz. Benefit dieser Architektur ist ein schnelles „non-blocking“-Netzwerk, das bedeutet, es gibt Zusagen zu Latenzzeiten und Performance.

Die beschriebenen Cyber-Defense Controls werden in Abhängigkeit von ihrem Zugriff teilweise hintereinander durchlaufen, wie in *Abbildung 2* dargestellt.

Automatisierung beziehungsweise Machine Learning ist bei Sicherheit entscheidend und wird sowohl innerhalb der Services eingesetzt als auch über Services hinweg. Einsatzgebiete dafür sind neben Betriebsprozessen auch Security-Monitoring und Bekämpfung von Threats. Das Security-Patching lässt normalerweise die Funktionen/Version unangetastet, sodass

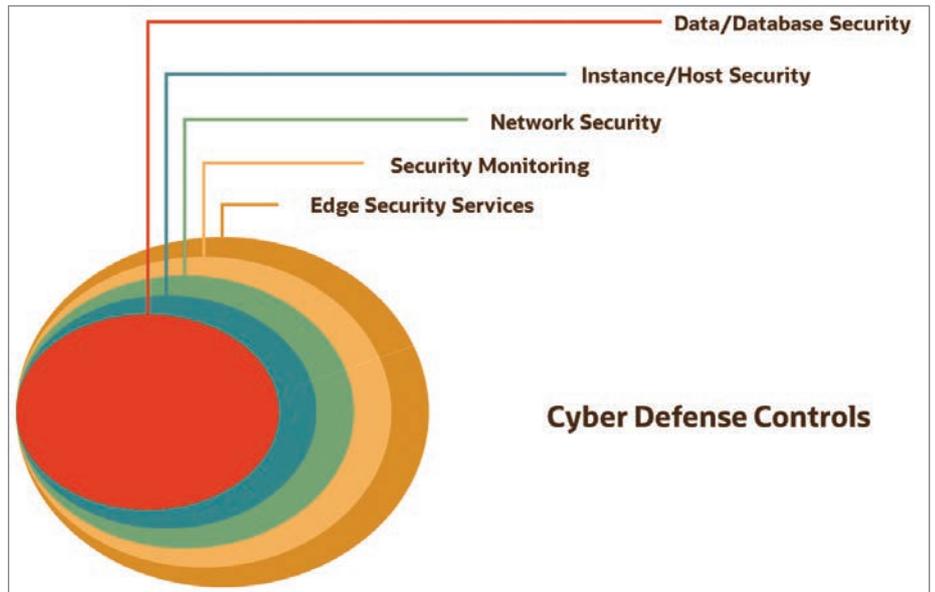


Abbildung 2: Cloud Cyber-Defense Controls (Quelle: Michael Fischer)

eine unterstützte Version bestehen bleibt. Beim Security-Monitoring werden Auffälligkeiten, schon bedingt durch die Menge an Daten und die Änderungsfrequenz von Systemen, mithilfe von Machine Learning entdeckt. So werden Sachverhalte erkannt, die vorher nicht bekannt waren. Das Machine Learning kommt neben der Anomalie-Erkennung und dem Security-Patching auch beim Threat Hunting und der automatischen Abwehr zum Tragen.

Im Folgenden werden einige der Cyber Security Controls weiter ausgeführt.

Cyber-Defense Control: Isolierung

Die Oracle Cloud baut auf einem „Least Trust Design“ auf. Isolierungen sind ge-

genüber den Mandaten, im Hypervisor und im Netzwerk implementiert. Der Grad der Isolierung kann durch Konfiguration erhöht werden, indem in der Cloud dedizierte Maschinen genutzt werden.

Die *Abbildung 3* zeigt die Isolierung (siehe *Top-Down Threats*) in der Cloud. Die Trennung der Netzwerkfunktion aus dem Hypervisor beziehungsweise dem Bare-Metal Server in eine eigene, von Oracle entwickelte Hardware, das sogenannte SmartNIC, verhindert ein Verbreiten von Threats außerhalb des Hypervisors auf der Maschine beziehungsweise beim Einsatz von Bare Metal außerhalb dieser Maschine. Ein Cloud-Netzwerk des Mandanten wird nicht wie bei anderen Lösungen nur definiert, sondern auf der Netzwerk-Hardware konfiguriert/durchgeschrieben. Eine logische Darstellung

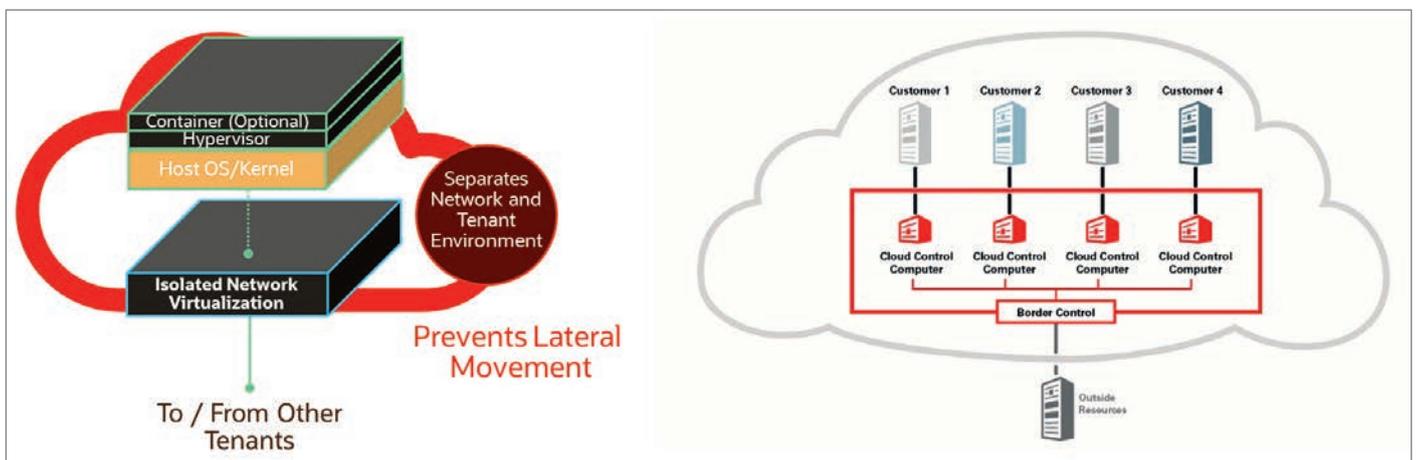


Abbildung 3: Isolierung / Zero Trust in der Oracle Cloud (Quelle: Michael Fischer)

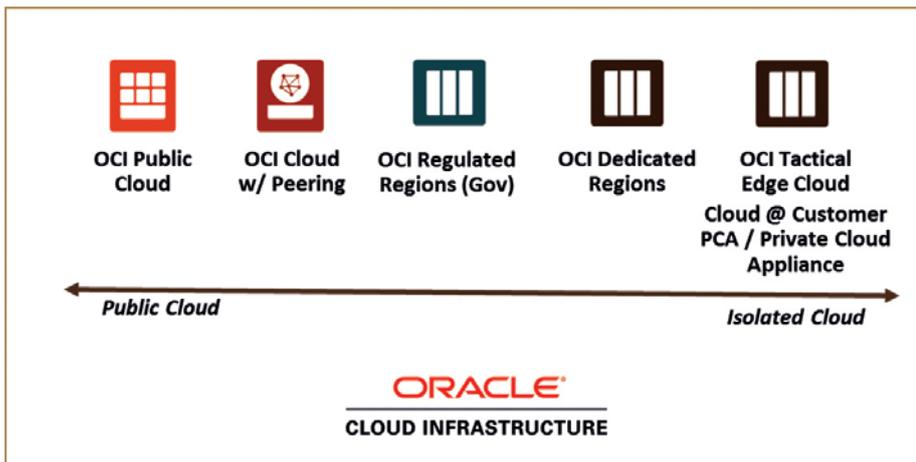


Abbildung 4: Cloud-Umgebungen (Quelle: Michael Fischer)

mit mehreren Systemen findet sich in der rechten Bildhälfte von *Abbildung 3*.

Reicht diese Art der Isolierung nicht aus, gibt es weitere Modelle der Cloud, die die *Abbildung 4* zeigt.

Gegenstand der bisherigen Beschreibung war die OCI Public Cloud. Diese kann so abgeschottet werden, dass sie ausschließlich über ihr Rechenzentrum erreichbar ist (sogenanntes „Peering“). Eine weitere Isolierung (sogenannte „Government Cloud“) kann über das Betriebspersonal – nur sicherheitsüberprüfte Staatsbürger des jeweiligen Landes – erreicht werden. Zusätzliche Einschränkungen bezüglich Behördennetzwerk und darin verfügbarer Services bestehen ebenfalls. Das Modell ist in Deutschland aktuell nicht verfügbar. Oracle bietet über „Dedicated Regions“ die Möglichkeit, eine Oracle Cloud in Ihrem eigenen oder einem „Wunsch-“

Rechenzentrum aufzubauen. Darüber hinaus gibt es verschiedene Oracle Cloud Appliances, die einen Teil der Oracle Cloud zur Verfügung stellen. On- und Offline-Modi sind dabei ebenso möglich, wie verschiedene der dargestellten Cloud-Modelle zu mischen.

Cyber-Defense Control: Verschlüsselung und Schlüsselmanagement (KMS)

Eine Verschlüsselung erfolgt sowohl am Speicherort als auch auf dem Transportweg.

Alle gespeicherten Daten im externen Blockstorage einschließlich Bootvolumes, Objektstorage und File Storage, werden standardmäßig durch Verschlüsselungen geschützt. Die Verschlüsselungs-Keys

können auf ein von Oracle bereitgestelltes Hardware Security Modul (HSM, FIPS compliant) in der Oracle Cloud ausgelagert werden, sodass der Nutzer die Schlüssel selbst verwalten kann.

Neben den oben angeführten Verschlüsselungen des Speichers unterliegen die Schlüssel der Oracle-Cloud-Datenbanken der Verwaltung des Kunden, sofern nicht anders spezifiziert. Die Datenbanken sind ohne Mehrkosten immer „zwangsverschlüsselt“. Ein verwaltbarer Key der Datenbank kann auch auf ein On-Premises Key-Management-System ausgelagert werden.

Cyber-Defense Control: Edge Security (DNS, WAF, WWW-Monitoring)

Die sogenannten Edge Services (*siehe Abbildung 5*) haben folgende Cyber-Defenses-Merkmale:

- DNS mit Ausfallsicherheit und Schutz vor DDoS-Angriffen durch globales Anycast-Netzwerk mit mehreren über Kontinente verteilten Rechenzentren und redundanten Tier-1-Anbindungen.
- Berücksichtigung des aktuellen Routings im Internet (WWW-Monitoring) durch Traffic-Analyse mithilfe von 600 Monitoring-Punkten, das Leistungsabfälle, Änderungen beim Internet-Routing, Netzwerksicherheitswarnungen und DDoS-Schutzaktivierungen von DDoS-Anbietern auswertet.

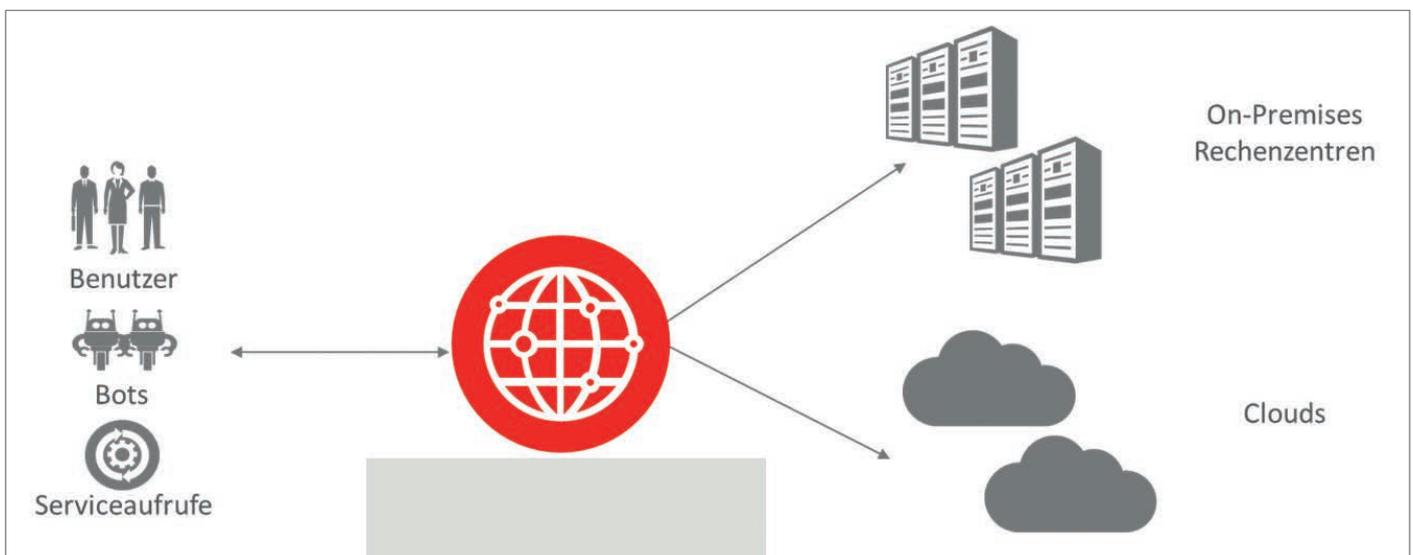


Abbildung 5: Überblick und Einordnung Edge Services (Quelle: Michael Fischer)

- Web Application Firewall (WAF) zum DDos-Schutz auf Layer 7 sowie Regelwerke gegen Botnets, Cross Site Scripting, SQL Injection und OWASP-Angriffen (CRS 3.0). Es können zusätzlich eigene Policies definiert werden, um den Zugriff basierend auf der Geografie, der URL oder der Signatur zu beschränken.
- Optimierung für die Erreichbarkeit und Latenz für Workloads in Clouds und On-Premises, Einbindung von Health Checks von verschiedenen Punkten im WWW für eigene URLs, um die Verfügbarkeit von Anwendungen und Services zu monitoren.

Es handelt sich hierbei um durch das globale Security Operations Center (SOC) von Oracle gemanagte Services, die neben dem Betrieb auch die Regelwerke aktualisieren.

Werden alle Edge Services genutzt, findet zuerst eine DNS-Steuerung bei der Auflösung der URL in eine IP statt. Der nächste Kontaktpunkt ist die WAF, die entsprechende Prüfungen vornimmt. Erfolgreiche Prüfungen der WAF lassen dann den Request auf den adressierten Server oder (API) Gateway oder Loadbalancer, in der Cloud oder On-Premises, durch. Für das Monitoring stehen die Log-Informationen aus den aufgerufenen Services, der WAF und den „Health Checks“ zur Verfügung.

Im Zusammenhang mit den OWASP Top10 gibt es neben der Entdeckung beziehungsweise Abwehr von ausgenutzten Schwachstellen auch Empfehlungen von Architektur-Blueprints, gerade auch

aus der Analyse von Security-Problemen in Microservices-Architekturen. Ein solcher Blueprint sieht vor, dass ein Request zuerst an einer WAF geprüft wird, dann auf ein API Gateway trifft, das einerseits den Request „kennt und steuert“ und andererseits die Authentifizierung und Autorisierung sicherstellt beziehungsweise einholt. Hierfür wird ein Identity- und Access-Management-System einbezogen (IdaaS, CIAM). Services werden anschließend Node- oder Container-basiert abgearbeitet. Alle Services in dieser Kette werden von der Oracle Cloud ebenfalls bereitgestellt.

- **Oracle Identity Cloud Service** stellt eine cloudbasierte Identity- und Access-Management-Plattform zur Verfügung (IdaaS und CIAM). Damit können Mitarbeiter, Externe, Partner, Kunden und Dinge (zum Beispiel Geräte oder IoT) verwaltet und angebunden werden. Durch vorgefertigte Integrationen bezüglich der Anbindung von sowohl On-Premises IDM- Systemen als auch SSO-Systemen ist eine einfache Nutzung möglich. Risikofaktoren bei der aktiven Session können bei der Authentifizierung und Autorisierung berücksichtigt und Services können spezifisch geschützt werden. Merkmale sind beispielsweise registrierte Devices, gerootete Devices, freigegebene Uhrzeiten, Geolocations, auffälliges Verhalten (UEBA) usw. Die Standards wie OpenIDConnect, OAuth, SAML werden bereitgestellt.
- **Oracle API Gateway**, um Serviceendpunkte wie Node-, Container- und

Functionaufrufe zu publizieren und dabei die Authentifizierung und Autorisierung sicherzustellen. Mithilfe dieses Service kann gewährleistet werden, dass allein der richtige User die richtigen Ressourcen nutzt. Er hilft ebenso zur Vorbereitung bei der Abwehr von potenziellen Attacken auf die Infrastruktur

- **Blockchain Cloud Service** bietet die Nutzung von permission-based Blockchains auf Basis von Hyperledger. Blockchain löst das Problem des Vertrauens zwischen Organisationen, indem es eine unabhängige Validierung durch ein manipulations sicheres, peer-distributed Ledger ermöglicht und somit die Notwendigkeit eines Offline-Abgleichs eliminiert. Vorkonfigurierte Blockchain-Codes sind für viele Standardgeschäftsprozesse einschließlich ERP-Transaktionen zur Nutzung vorkonfiguriert. Das Benutzer-/ Vertrauensmanagement erfolgt über den bereitgestellten Identity Cloud Service.

Cyber-Defense Control: Cyber Attack Hunting

Die Aktivitäten in der Cloud werden in Audit- beziehungsweise Logfiles mandantenspezifisch protokolliert (siehe Abbildung 6). Dabei wird unterschieden zwischen Services, die von Oracle bereitgestellt und „betrieben“ werden, und Services, die ein Mandant aufsetzt und selbst betreibt. Von Oracle bereitgestellt werden die Auditlogs vom Cloudmanage-

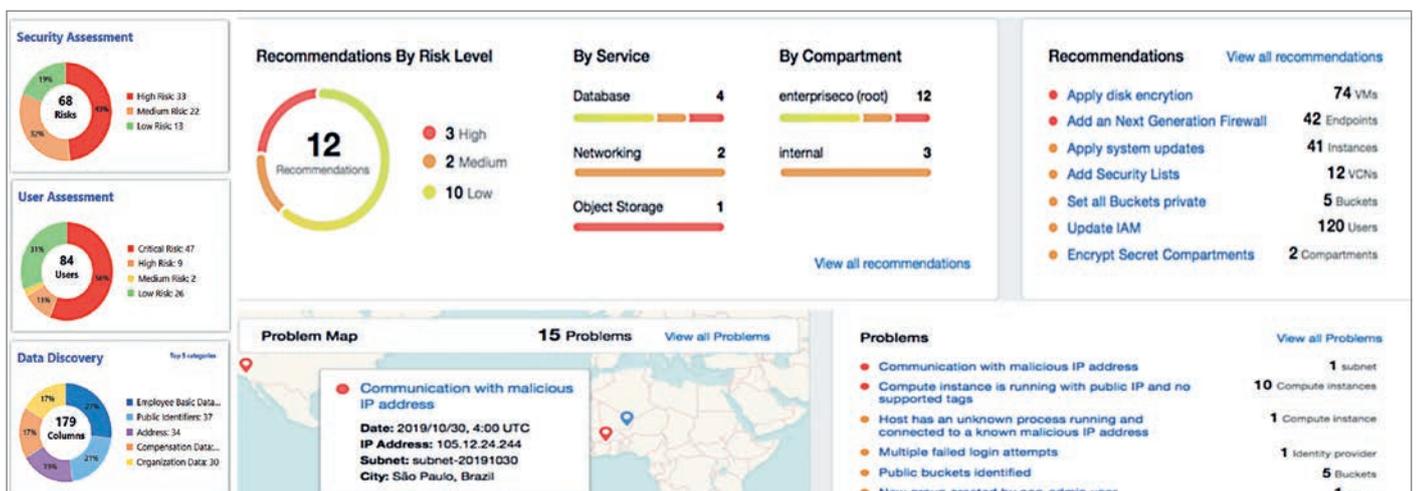


Abbildung 6: Auswertungen Cyber-Defense (Quelle: Michael Fischer)

ment (z.B. Löschen einer Instanz) sowie Logs vom Netzwerk (VCN Flows inklusive der Aktionen wie Blockieren). Spezielle Services liefern weitere Loginformationen wie das Datenbankauditlog oder syslog. Diese Informationen liegen in Rohform vor und können mit Oracle-Werkzeugen durch den Kunden oder 3rd-Party-Werkzeugen ausgewertet werden. Dies erfolgt unabhängig der Verantwortung von Oracle, die Verfügbarkeit der Cloud sicherzustellen.

Von Oracle beigestellte Werkzeuge für Kunden sind:

- **Oracle Data Guard** zum Security Monitoring von Oracle-Datenbanken bezüglich Security-Konfiguration (Assessment gegen Best Practices, STIG, GDPR), risikoreichen Benutzern und risikoreichen Aktivitäten sowie Dokumentieren sensibler Daten, Daten-Discovery und Daten-Masking.
- **Oracle Cloud Guard** zum Monitoring der Oracle Cloud. Auditlogs, Data Safe, OS Management und Network Flow Logs werden aggregiert und ausgewertet, um Threats und Fehlkonfigurationen zu erkennen und gegebenenfalls Threats automatisiert zu behandeln, zum Beispiel Public IPs zu entfernen, auffällige Instanzen herunterzufahren, Benutzerberechtigungen zu entziehen usw. Der Service ist aktuell (2/2020) noch nicht verfügbar.
- **Oracle Cloud Access Security Broker** (CASB) zum cloudübergreifenden Monitoring von Aktivitäten und Konfigurationen (UEBA, Policies, IT-Compliance). Hier werden neben der Oracle Cloud auch Azure, AWS, Salesforce und weitere Clouds unterstützt.
- **Oracle Management Cloud** sind Services zur Überwachung, Verwaltung und Analyse der IT-Systeme und Anwendungen, On-Premises und in der Cloud. Unter anderem sind Leistungsschwankungen diagnostizierbar, Verhaltens- und Fehleranalysen (top-down) möglich. Machine-Learning-basierte Auswertungen ermöglichen zum Beispiel Clustering, Anomalie-Erkennung und Vorhersagen. Eine Vielzahl von Adaptoren ermöglicht die einfache Integration von Oracle und 3rd-Party-Systemen.

Ein Spezialfall für Services mit Security Automation und eingebautem Cyber-Hunting

sind die sogenannten **Autonomous** Systeme, die aktuell für die Oracle-Datenbank und Oracle Linux verfügbar sind. Die Systeme stellen das automatische Security-Patching sicher, bei dem der Kunde explizit Abschwächungen konfigurieren muss. Security-Patching wehrt 85% der Angriffe ab, die gegen ungepatchte Systeme erfolgreich waren (siehe *Data Breach Reports*). Patching erfolgt generell ohne Downtime. Daher kann Oracle im Falle der Datenbank bis zu 99,995% Verfügbarkeit inklusive geplanter Maintenance zusagen. In der jüngsten Vergangenheit mit den Patches zu Spectre/Meltdown erfolgte der Praxisnachweis mit 150 Millionen Patchvorgängen innerhalb von 4 Stunden. Mit User Entity Behaviour Analytics (UEBA) wird abweichendes Verhalten von Innentätern, Externen und Systemen erkannt. Alle weiteren Standard Cyber-Defense Controls wie Verschlüsselung, Zugriffskontrolle durch Operations, Auditlogs usw. bleiben erhalten.

Fazit

Oracle bietet mit den dargestellten Security Controls umfassenden Schutz. Alle vorgestellten Mechanismen sind in der Oracle Cloud im Standard vorhanden und aktiv. Entsprechende Zertifizierungen wie ISO27001, ISO270017/18, BSI C5 oder Fedramp IL5 weisen eine effektive Umsetzung nach.

Die besten Security Controls nützen nichts, wenn die Grundregeln der IT-Sicherheit nicht eingehalten werden, so wie das heute bei vielen IT-Umgebungen der Fall ist. Um dieses Problem zu minimieren, setzt Oracle hier auf die hochgradig KI-basierte Automatisierung der Security. Um einen Start aus einer misslichen Ausgangslage zu verhindern, setzt Oracle auf Security-by-Default, einen Deny-All-Ansatz und ein entsprechendes Monitoring, damit die vorhandenen Einstellungen nicht von der erwarteten Security abweichen. Natürlich können die Security Controls auch um eigene oder 3rd Party Security Appliances erweitert werden.

Das große Problem mit dem Security-Patching kann heutzutage automatisiert mit den sogenannten Autonomous Systemen von Oracle für Oracle-Datenbanken und Oracle Linux elegant gelöst werden. Dabei ist das Patching nicht der einzige

Mehrwert von Autonomous. Damit einher gehen das Patching im laufenden Betrieb und die Zusicherung von SLAs, die die Wartung bereits enthalten. Autonomous ist nicht äquivalent zu Automatisierung. Autonomous arbeitet zusätzlich KI-basiert, um hier eine möglichst hohe Autonomie und Reaktionsgeschwindigkeit zu erreichen. Die „Erfahrungen“ der KI kommen dabei aus den Betriebserfahrungen und Supportfällen von Oracle sowie vom Oracle Engineering.

Durch die Möglichkeit, auch Oracle Clouds im eigenen oder „Wunsch-“ Rechenzentrum oder als Appliances zu betreiben, ergibt sich hier die Möglichkeit, die Vorteile der implementierten und fortwährend aktualisierten Security-Mechanismen zu nutzen.



Michael Fischer
michael.fischer@oracle.com



Oracle Database Indexing Best Practices – Teil 1

Randolf Geist, Unabhängiger Berater

Auch wenn die Version 19c der Oracle-Datenbank die automatische Indizierung als neues Feature mit sich bringt, wird die Indizierung auch in Zukunft ein Dauerbrenner-Thema bleiben – auch schon aufgrund der Lizenzpolitik seitens Oracle, die das neue Feature in naher Zukunft nur einem Bruchteil der Kunden zukommen lassen wird. Daher soll dieser Artikel mit einigen Mythen im Bereich Indizierung aufräumen und hilfreiche Tipps und Tricks beschreiben, die helfen, besser zu verstehen, was im Bereich Indizierung sinnvoll und nützlich ist. In diesem ersten Teil zeige ich, warum es in aller Regel ein Mythos ist, Indizes regelmäßig neu bauen zu müssen.

Index-Rebuild-Mythen

Es ranken sich sicherlich auch heutzutage noch einige Mythen um das Thema „Indizes neu erstellen“ – auch im Internet finden sich weiterhin viele Treffer, die auf Halbwissen und teilweise schlichten Fehlannahmen beruhen.

Eine der grundlegenden Fehlannahmen beschreibt als Grund für eine Neuerstellung eines Index eine mögliche „Schräglage“ der Index-Struktur, bei der der Index angeblich unterschiedliche Höhen beziehungsweise Tiefen in seiner Struktur enthält und daher neu gebaut werden sollte, um dies zu berei-

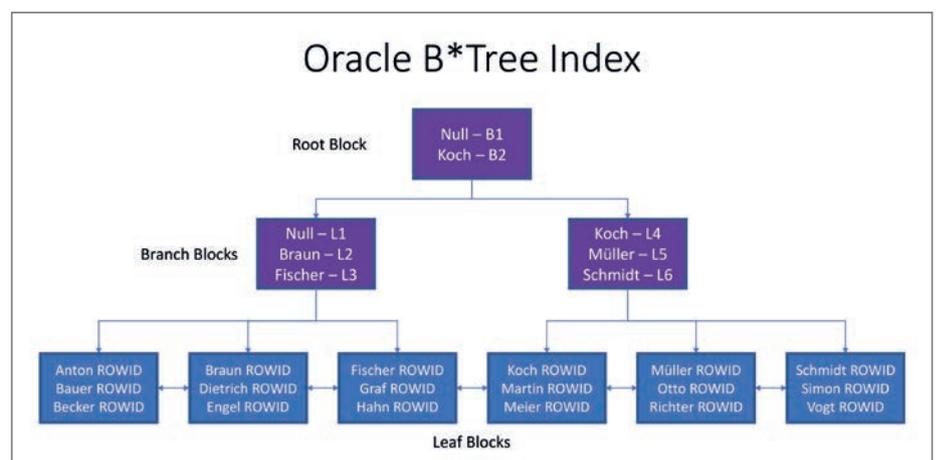


Abbildung 1: Oracle B*Tree Index (© Randolf Geist)

nigen. Grundsätzlich kann man sich so eine B*Tree-Index-Struktur als Baum vorstellen, der so aussehen könnte – bei einer ausreichend großen Anzahl an Einträgen besteht die Struktur dann aus einem „Root-Block“, einer oder mehreren Ebenen an „Branch-Blöcken“, die alle jeweils die minimal notwendige Information beinhalten, um in der Struktur navigieren zu können, und schließlich der untersten Ebene der „Leaf-Blöcke“, in denen die eigentlich zu indizierenden Daten stehen, zusammen mit der ROWID, die auf den entsprechenden Eintrag in der Tabelle verweist (siehe Abbildung 1).

Die beschriebene Schiefelage könnte dann ungefähr so aussehen – eine Mischung von Leaf- und Branch-Blocks auf der gleichen Ebene sowie Teilbäume der Struktur mit unterschiedlicher Tiefe (siehe Abbildung 2).

Wenn man sich aber vor Augen führt, welche Möglichkeiten der Veränderungen einer B*Tree-Index-Struktur es gibt, wird schnell klar, dass dies so gar nicht passieren kann – die entscheidende Frage in diesem Zusammenhang ist ja, wann sich die Höhe der Struktur verändert.

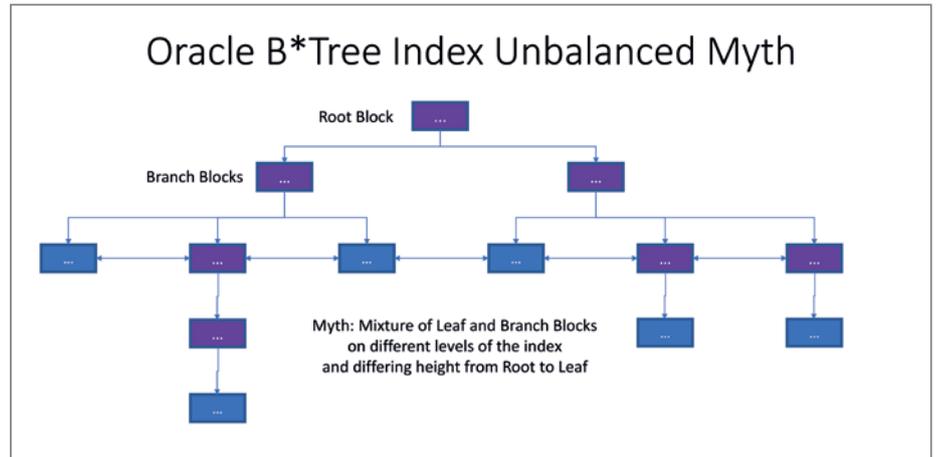


Abbildung 2: Oracle B*Tree Index Unbalanced Myth (© Randolph Geist)

Schauen wir uns dazu die verschiedenen Fälle an, wie eine Index-Struktur gepflegt wird. Da ein Index ja eben eine vorsortierte Struktur darstellt, können neue Einträge nicht einfach an beliebiger Stelle platziert werden (das ist genau die Eigenschaft der Standard Oracle „Heap“-Tabelle – wo Platz gefunden wird, kann ein Eintrag hinzugefügt werden), sondern ein

Eintrag muss an einer genau definierten Stelle in der Index-Struktur eingefügt werden – zumindest aus logischer Sicht.

Leaf-Block-Split

Daher kann es zu dem Fall kommen, dass in dem Block, der den neuen Eintrag aufzu-



Das E-3 Magazin

Information und Bildungsarbeit von und für die SAP-Community

Wir leben alle unter dem gleichen Himmel, aber wir haben nicht alle den gleichen Horizont.

Konrad Adenauer



E-3 – die Wissensplattform für die SAP-Community

nehmen hat, schlicht kein ausreichend freier Platz mehr vorhanden ist. In diesem Fall muss Oracle sich etwas einfallen lassen, was sich dann „Index-Block-Split“ nennt – einer neuer Block wird in die Index-Struktur eingefügt und die Daten werden zwischen dem ursprünglichen und dem neuen Block aufgeteilt (siehe Abbildung 3).

In dem abgebildeten Szenario handelt es sich um einen sogenannten 50-50-Block Split, bei dem ein neuer Block in die Struktur eingefügt wird und die Daten volumemäßig 50 zu 50 zwischen den beiden Blöcken aufgeteilt werden. Die Index-Struktur wird also auf einer existierenden Ebene „breiter“ durch diese Operation – die Höhe ändert sich nicht. Es sei noch angemerkt, dass es sich hier um eine recht aufwendige Aktion handelt, da verschiedenste Dinge angepasst werden müssen – so müssen die Zeiger, die von Leaf-Block zu Leaf-Block zeigen, in den logisch folgenden und vorherigen Blöcken angepasst und natürlich auch in den Branch-Blöcken darüber entsprechende Anpassungen vorgenommen werden, um den neuen Block in der Struktur bekannt zu machen.

Branch-Block-Split

Denkt man die Operation weiter, kann es natürlich bei den Anpassungen in den darüber liegenden Branch-Blöcken ebenso passieren, dass kein ausreichender Platz mehr vorhanden ist – hier anhand eines anderen „Leaf-Block-Split“-Szenarios durchgespielt (siehe Abbildung 4).

Hier wird jetzt ein sogenannter 90-10-Block-Split auf Leaf-Block-Ebene durchgeführt. Oracle verwendet diese Operation, wenn der einzufügende Eintrag sich oberhalb oder unterhalb des derzeit höchsten beziehungsweise niedrigsten Eintrags befindet, logisch gesehen also am Ende oder Anfang des Index eingefügt wird. In diesem Fall wird keine Aufteilung der Daten vorgenommen, sondern einfach ein neuer Block der Struktur hinzugefügt und der neue Eintrag in den neuen Block übernommen. Zu beachten ist, dass diese unterschiedlichen Formen der Block-Splits wesentlichen Einfluss auf den Füllgrad der Blöcke haben. Beim 50-50-Block-Split verbleiben eben ca. 50% Freiplatz in den Blöcken, die je nachdem, welche Daten in die Struktur weiterhin eingefügt werden, entweder frei bleiben oder weiter aufge-

füllt werden. Beim 90-10-Block-Split, der technisch gesehen ein 99-1 oder ähnlicher Split ist, werden die Blöcke jeweils zu 100% gefüllt, bis auf den letzten am Anfang beziehungsweise Ende der Struktur. Daher schwankt der typische Füllgrad einer B*Tree-Index-Struktur zwischen 70 und 75 Prozent, je nach Art und Weise, wie die Einträge in den Index eingefügt werden (siehe Abbildung 5).

Ist jetzt nicht ausreichend Platz in dem darüber liegenden Branch-Block, findet auf dieser Ebene auch ein Block-Split statt und ein neuer Block wird in die Struktur aufgenommen mit den dazugehörigen Vertei-

lungs- und Pflegemaßnahmen. Auch in diesem Fall ändert sich die Höhe der Struktur nicht, sondern nur die „Breite“ auf der vorhandenen Ebene (siehe Abbildung 6).

Root-Block-Split

Schließlich kann es dazu kommen, dass im Root-Block auch nicht mehr ausreichend Platz besteht, um neue Einträge hinzuzufügen zu können, und auch dann findet ein sogenannter „Root-Block-Split“ statt. Nur in diesem Fall ändert sich tatsächlich die Höhe der Index-Struktur, indem zwei neue

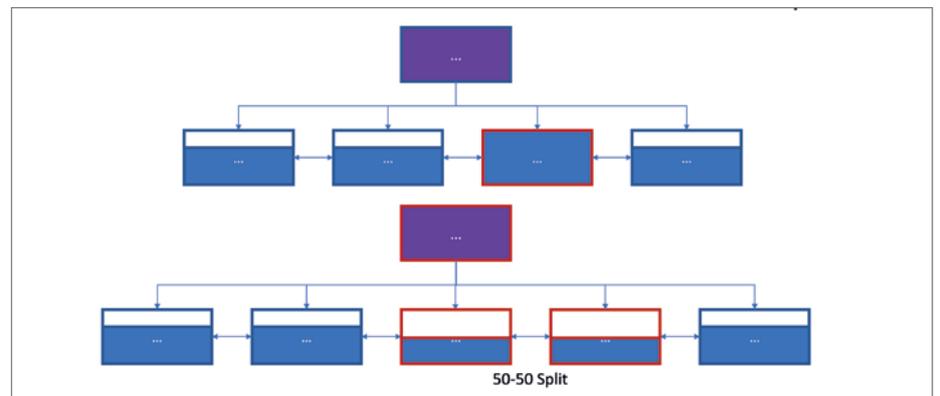


Abbildung 3: Oracle B*Tree Index True Behaviour On Split (© Randolph Geist)

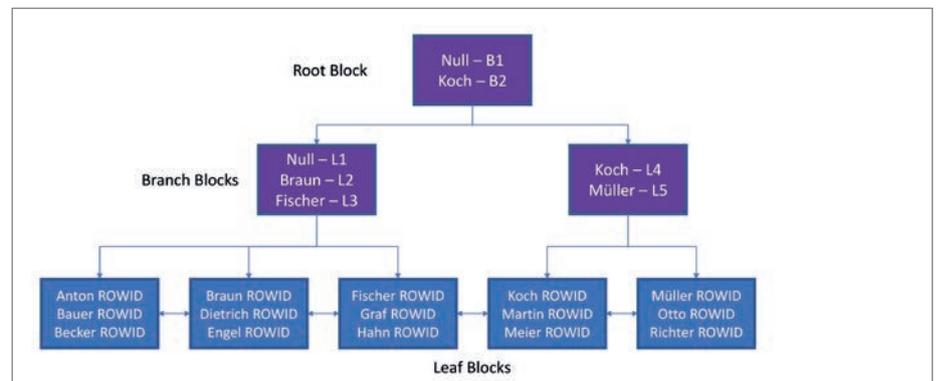


Abbildung 4: Oracle B*Tree Index Branch Block Split (1) (© Randolph Geist)

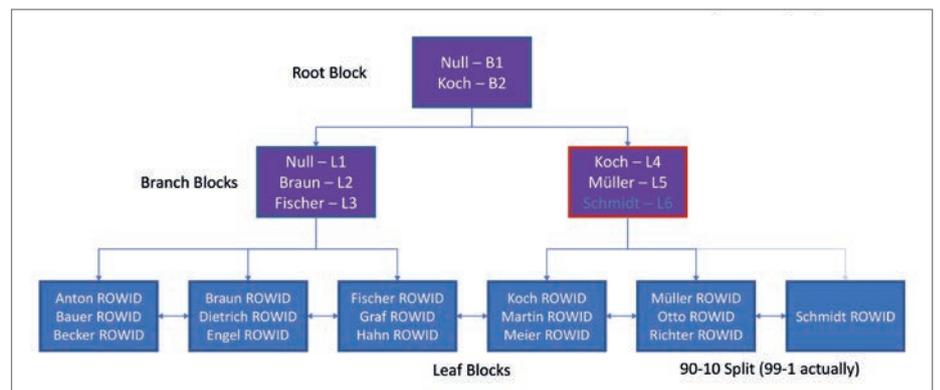


Abbildung 5: Oracle B*Tree Index Branch Block Split (2) (© Randolph Geist)

Branch-Blöcke unterhalb des Root-Blocks eingefügt werden und somit eine weitere Ebene dem Index hinzugefügt wird. Durch dieses Verfahren bleibt der Index immer „balanciert“ und die Entfernung von Root- zu Leaf-Blöcken im gesamten Index immer gleich (siehe Abbildung 7 und 8).

Man kann sich von dieser Tatsache auch auf andere Art und Weise überzeu-

gen, indem man einen sogenannten Index-Tree-Dump erzeugt – das kann mit dem folgenden Befehl angefordert werden (siehe Listing 1).

„nnnnn“ ist die OBJECT_ID des Objektes – im Fall von (sub)partitionierten Indizes wird die OBJECT_ID der jeweiligen Index-(Sub-)Partitionen eingesetzt.

Die Datenbank erzeugt aufgrund des „ALTER SESSION“-Befehls eine entsprechende Trace-Datei, die dann zum Beispiel so aussehen kann:

```
branch: 0x3c00781 62916481 (0:
nrow: 3, level: 2)
  branch: 0x3c00a35 62917173
(-1: nrow: 690, level: 1)
  leaf: 0x3c00782 62916482
(-1: row:46.46 avs:7356)
  leaf: 0x3c00783 62916483
(0: row:46.46 avs:7356)
  leaf: 0x3c00784 62916484
(1: row:43.43 avs:7362)
  leaf: 0x3c00785 62916485
(2: row:43.43 avs:7355)
  leaf: 0x3c00786 62916486
(3: row:43.43 avs:7356)
```

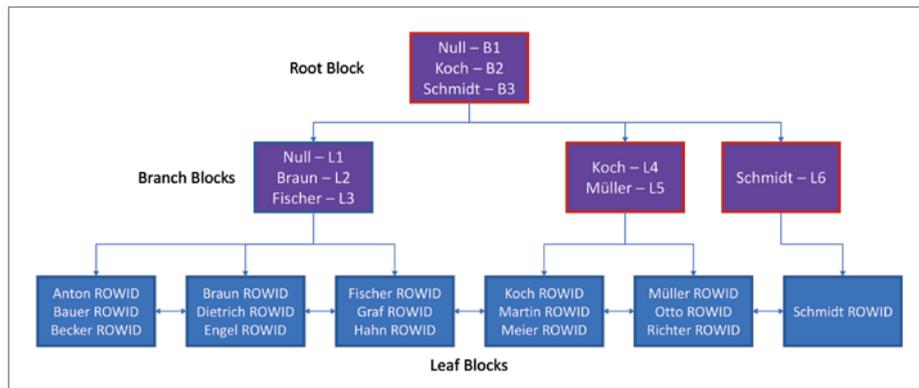


Abbildung 6: Oracle B*Tree Index Branch Block Split (3) (© Randolf Geist)

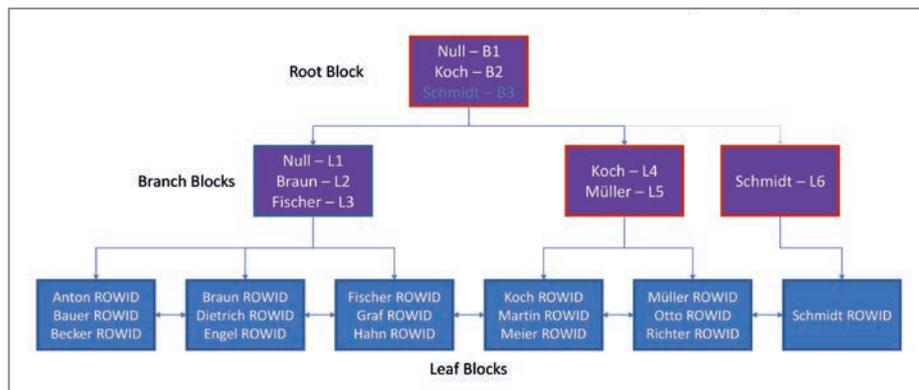


Abbildung 7: Oracle B*Tree Index Root Block Split (1) (© Randolf Geist)

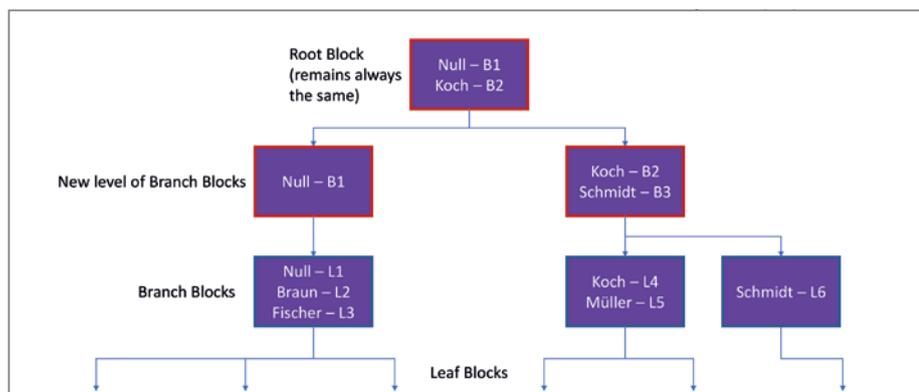


Abbildung 8: Oracle B*Tree Index Root Block Split (2) (© Randolf Geist)

```
SELECT OBJECT_ID FROM DBA/ALL/USER_OBJECTS
WHERE OBJECT_NAME = '<INDEX_NAME>' AND OBJECT_TYPE = 'INDEX';
ALTER SESSION SET EVENTS 'immediate trace name treedump level nnnnn';
```

Listing 1: Befehl zur Erzeugung eines Index-Tree-Dumps

Der Tree-Dump zeigt ganz klar, dass die Entfernung von Root- zu Leaf-Blöcken an allen Stellen des Index gleich ist.

In den kommenden Teilen 2 und 3 gehe ich darauf ein, was beim Design von Mehrspalten-Indizes zu beachten ist.

Über den Autor

Randolf Geist ist seit 20 Jahren als freiberuflicher Oracle Datenbank-Experte tätig und auf Performance-Themen spezialisiert. Im Bereich der Oracle Optimizer-Technologie und SQL-Performance-Analyse gehört er zu den Top-Experten weltweit. Er veröffentlicht regelmäßig entsprechende Beiträge in Fachzeitschriften sowie auf seinem Blog und seinem YouTube-Kanal, und gibt sein Wissen in Seminaren / Webinaren, den Oracle OTN-Foren und auf Konferenzen weiter. Des Weiteren werden seine Experten-Seminare bei der Oracle University und der DOAG angeboten. Randolf ist Mitglied des OakTable Networks und war bis 2016 Mitglied des Oracle ACE Director-Programms.



Randolf Geist
randolf.geist@oracle-performance.de



Systematisches Oracle Performance Tuning mit kostenlosen 3rd Party Tools

Stefan Koehler, Freiberuflicher Oracle Performance Berater & Researcher

In meiner täglichen Arbeit als freiberuflicher Oracle-Berater stehe ich immer häufiger vor der Herausforderung, dass ich u.a. aus Sicherheitsgründen keinen direkten Zugriff auf die entsprechenden Oracle-Datenbanken erhalte oder akute Performance-Probleme ad hoc ohne Mail- bzw. Information-Ping-Pong analysieren muss. Letzteres ist auch mein eigener Anspruch bei der Arbeit mit Kunden - eine schnelle, unkomplizierte und effiziente Kommunikation sowie Analyse des Problems.

Um alle relevanten Performance- und Diagnosedaten einfach, schnell und vor allem vollständig zu sammeln, stehen am Markt sowohl kostenpflichtige als auch kostenfreie Tools zur Verfügung. Zwei dieser Tools, die ich selbst täglich in der Zusammenarbeit mit meinen Kunden nutze, möchte ich Ihnen in diesem Artikel vorstellen.

Systematisches Performance Tuning mithilfe des Method-R-Ansatzes

Selbst die besten Performance-Tools benötigen einen systematischen Ansatz zur Performance-Analyse, um diese effektiv einsetzen zu können und hier empfiehlt sich aus meiner Erfahrung der letzten 18

Jahre im Oracle Performance-Umfeld die „Method-R“-Methodik [1] von Cary Millsap - das „R“ steht dabei für „Response Time“ - da sich diese Methodik auf die Endbenutzer/Businessprozess-Antwortzeit fokussiert und nicht auf die Oracle Database Time, welche von Oracle wie folgt definiert wird: „DB Time = Total time in database calls by foreground sessions

Rule #1: If you can trace, trace it & analyze with appropriate profiler

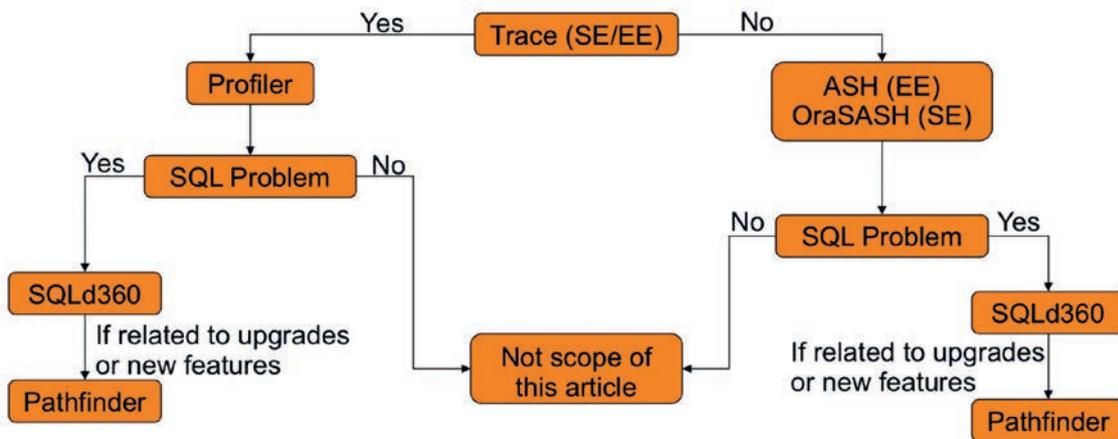


Abbildung 1: Vorgehensweise mit Tools bei Method-R (© Stefan Koehler)

and includes CPU time, IO time and non-idle wait time“.

Weiterführende Informationen zu diesem methodischen Ansatz und warum dieser zum Beispiel globalen Performance-Analysen mit AWR oder STATSPACK Reports überlegen ist, finden Sie in dem als Referenz angegebenen Buch [1].

Method-R besteht im Wesentlichen aus vier einzelnen Schritten:

1. Select the user actions for which the business needs improved performance.
2. **Collect properly scoped diagnostic data that will allow you to identify the causes of response time consumption for each selected user action while it is performing sub-optimally.**
3. Execute the candidate optimization activity that will have the greatest net payoff to the business. If even the best net-payoff activity produces insufficient net payoff, then suspend your performance improvement activities until something changes.
4. Go to step 1.

Ein entscheidender Erfolgsfaktor zur schnellen, effektiven und kostengünstigen Performance-Analyse ist Punkt 2 und hierbei unterstützen unter anderem die zwei Tools, die nun im weiteren Verlauf vorgestellt werden.

Method-R-Ansatz mithilfe von Tools

Die wichtigste Regel beim Sammeln von exakten und Endbenutzer-/Businesspro-

zess-spezifischen Analysedaten ist: „Wenn die Endbenutzeraktion bzw. der Businessprozess getracet werden kann - dann unbedingt tracen und mit einem passenden Profiler analysieren.“

Ein großer Vorteil beim Tracen ist, dass auch nicht ganz offensichtliche Ursachen (zum Beispiel CPU Scheduling oder Implementierung der Applikation) mit aufgezeichnet und analysiert werden können, die sonst mit dem klassischen 1-/10-Sekunden-Sampling (z.B. mittels Oracles ASH) unter anderem nicht erfasst werden. Des Weiteren ist das Tracing in jeder Oracle-Edition verfügbar und an Hand der sehr genauen Analysedaten können entsprechend präzise Performance Forecasts für geplante/mögliche Anpassungen getroffen werden (siehe Abbildung 1).

Um einen Endbenutzer/Businessprozess spezifischen Extended SQLTrace zu aktivieren, bietet die Oracle-Datenbank verschiedenste Möglichkeiten, wie beispielsweise das PL/SQL-Package DBMS_MONITOR oder die mit Oracle Release 11 komplett überarbeitete „Kernel Debug, Diagnostics & Tracing Infrastructure“. Welches der beste und einfachste Weg für das Erzeugen der Tracedaten ist, hängt von verschiedenen Faktoren, wie zum Beispiel „Ist die Applikation instrumentiert?“ oder „Wird Connection-Pooling verwendet?“ ab.

Nachdem die Endbenutzeraktion bzw. der Businessprozess getracet worden ist, muss die Trace-Datei mit einem passenden Profiler wie beispielsweise dem kostenpflichtigen Method-R Profiler [2] oder dem kostenfreien TVD\$XTAT [3] analysiert werden.

An Hand der aufbereiteten Trace-Datei ist nun zu entscheiden, ob ein SQL-Problem (zum Beispiel aufgrund eines unpassenden Ausführungsplans) oder ein anderweitiges Performance-Problem vorliegt. Eine erste annähernde Einschätzung kann zum Beispiel auf Basis der Logical Reads pro zurückgelieferter Zeile auf Zugriffspfads-Ebene erfolgen [4], das bedeutet, dass man folgende Daumenregel pro (nicht aggregiertem) Zugriffspfad verwendet:

- Zugriffspfade mit weniger als 5 Logical Reads pro zurückgelieferter Zeile sind wahrscheinlich gut
- Zugriffspfade zwischen 5 und 15 Logical Reads pro zurückgelieferter Zeile sind wahrscheinlich akzeptabel
- Zugriffspfade mit über 20 Logical Reads pro zurückgelieferter Zeile sind wahrscheinlich ineffizient und können auf ein Ausführungsplan bzw. Daten(modell)-Problem hindeuten

In der Praxis gibt es (leider) auch Fälle, die ein Tracen der Endbenutzeraktion beziehungsweise des Businessprozesses nicht zulassen. Darunter fallen unter anderem nicht mehr reproduzierbare Performance-Probleme (zum Beispiel nach Data Guard Switch-/Failover oder nach Applikationsanpassungen), Performance-Probleme, die nur sehr sporadisch auftreten oder Performance-Probleme, die zwar noch auftreten, aber aufgrund der Anwendungsarchitektur nicht dediziert für einen Endbenutzer/Businessprozess getracet werden können. In diesen Fällen muss man auf historische und gesampelte ASH-Analysedaten (die

per Default auf Oracles DB Time Definition beruhen) zurückgreifen, sofern die zusätzlich benötigte Oracle EE-Option dafür vorhanden ist. Für die Oracle Standard sowie Oracle Enterprise Edition ohne die entsprechende Option gibt es kostenfreie ASH-Replika wie etwa OraSASH (Oracle Simulation ASH) von Marcin Przepiorowski [5].

Die Oracle-Datenbank liefert von Haus aus einige Tools wie beispielsweise das SQL-Skript „@?/rdbms/admin/ashrpti.sql“ zur ASH-Datenanalyse mit - ich persönlich bevorzuge und nutze für ASH-Analysen aber 3rd-Party-Skripte wie etwa „ashtop.sql“ oder „dashtop.sql“ von Tanel Poder [6], da diese wesentlich flexibler bei Aggregation sowie Output sind und über Parameter dynamisch angepasst werden können.

Die „Method-R“-Methodik an sich bleibt auch mit ASH-Daten die gleiche, wobei man Abstriche (vor allem im Hinblick auf die Genauigkeit sowie Qualität) bei den zur Verfügung stehenden Analysedaten machen muss.

Dieser Artikel behandelt im weiteren Verlauf ausschließlich den Fall eines SQL-Problems und stellt zwei kostenlose Tools vor, die beim Sammeln und Analysieren von SQL-spezifischen Daten unterstützen.

Tool „SQLd360“ [7]

Wenn die vorhergehenden Analysen der Trace-Dateien (oder mittels ASH) ergeben haben, dass wahrscheinlich ein SQL-Problem vorliegt, müssen im weiteren Verlauf der Analyse alle relevanten Informationen (wie zum Beispiel Ausführungspläne, Bind-Variablen, involvierte Objekte, Statistiken, PL/SQL-Prozeduren, historische sowie aktuelle Ausführungsstatistiken etc.) zu diesem spezifischen SQL gesammelt werden, um anschließend mögliche Lösungsansätze aufzeigen zu können.

Aspect	SQLTXPLAIN	SQLd360
Installation	Requires installation. Creates two schemas that serve as metadata and code repository, they need to be present every time the tool is executed	No installation, at the end of the execution no evidence is left behind
Source code	Mainly PL/SQL, unwrapped (clear text). Code is Oracle property	SQL scripts, clear text, true open source released under GPL. Code available on GitHub
Main Goal	Automate every aspect of diagnostic collection for SQL tuning	Same as SQLT, but less "obsessive" 😊
Main Focus	Execution plan generation, why CBO generated plan X,Y,Z (includes custom testcase, very flexible)	Execution plan generation (but a little less than SQLT) and how plans execute at runtime
Main advantage	Includes every bit of info needed to drill into CBO decisions, unlikely the need to go back and collect anything else	Makes diagnostic consumption faster and easier because of data visualization, lots of focus on runtime
Main disadvantage	Requires installation. Not much focus on execution time ("same plan different performance" type of cases harder to deal with)	Might miss some information under rare circumstances

Abbildung 2: Vergleich von SQLd360 mit Oracle's SQLT [8] (© Stefan Koehler)

Das Tool „SQLd360“ unterstützt den DBA bzw. externen Berater genau in diesem Punkt, da es eine 360-Grad-Übersicht in Form von HTML-Dateien, Grafiken, Trace-Dateien und ggf. notwendigen Dump-Dateien für ein spezifisches SQL Statement erzeugt (siehe Abbildung 3).

Das Tool wurde ursprünglich von Mauro Pagano entwickelt, der unteren anderem auch an der (Weiter-)Entwicklung von Oracles SQLT-Tool beteiligt war. Im Prinzip bieten beide Tools einen ähnlichen Funktionsumfang, allerdings hat SQLd360 den entscheidenden Vorteil, dass keine Installation in der Datenbank erfolgen muss, d.h. unter anderem auch, dass das Tool von jedem beliebigen Client sowie auf dem Datenbankserver direkt via SQL*Plus ausgeführt werden kann (siehe Abbildung 2).

SQLd360 selbst ist vor einiger Zeit mit dem ebenfalls kostenlosen Tool eDB360

(welches eine 360-Grad Übersicht für die komplette Datenbank liefert) vereinigt und unter dem neuen Tool-Namen „sqldb360“ veröffentlicht worden. Das aktuelle Release kann von GitHub [7] als ZIP-Datei heruntergeladen werden. Anschließend wird die ZIP-Datei auf dem Datenbankserver oder einem Client mit einer entsprechenden SQL*Net-Verbindung extrahiert. SQLd360 liefert bereits eine Default-Konfiguration (Datei „.../sqldb360-19.3/sql/sqld360_00_config.sql“) mit, die in den allermeisten Fällen ausreichend ist – gegebenenfalls ist es notwendig, den Analysezeitraum (Parameter „sqld360_conf_days“) anzupassen oder mithilfe des SQL Test Case Builder (Parameter „sqld360_conf_incl_tcb“ & „sqld360_conf_tcb_sample“) Testfälle zu integrieren, aber dies sind eher die Ausnahmen als die Regel.

Der Aufruf des Tools (SQL Skript sqld360.sql) ist mit lediglich zwei Übergabeparametern an sich sehr simpel:

1. SQL-ID des zu analysierenden SQL Statements
2. [T = Tuning | D = Diagnostic | N = NONE] zur Definition, welche Lizenz für die Datenbank, in der das zu analysierende SQL ausgeführt wurde, vorhanden ist

Der Output von SQLd360 ist auch ohne Oracle EE-Optionen oder nur mit der Oracle Standard Edition hilfreich, da weiterhin alle

```
T193DB [oracle@OEL76 sqldb360-19.3]$ sqlplus / as sysdba

SQL*Plus: Release 19.0.0.0.0 - Production on Thu Nov 7 13:23:04 2019
Version 19.3.0.0.0

Copyright (c) 1982, 2019, Oracle. All rights reserved.

Connected to:
Oracle Database 19c Enterprise Edition Release 19.0.0.0.0 - Production
Version 19.3.0.0.0

SQL> @sqld360 1rn4029rjzfa3 T
```

Abbildung 3: Aufruf von SQLd360 (© Stefan Koehler)

relevanten Objekt-/Statistikinformationen zu dem SQL enthalten sind und ein manuelles Sammeln entfällt. Des Weiteren kann im Falle von NONE der Ausführungsplan mit zusätzlichen Informationen angereichert werden (Stichwort „GATHER_PLAN_STATISTICS Hint“), welche dann ebenfalls im Output von SQLd360 enthalten sind.

Die Laufzeit von SQLd360 kann abhängig von der Anzahl involvierter Objekte sowie der Menge historischer Informationen stark variieren. Im Alltag konnte ich beobachten, dass die Gesamtlaufzeit zum Großteil durch die Sammlung der „SQL Plan Directives“ beziehungsweise der „SQL Plan Directives Objects“-Information beeinflusst wird - dies ist aber nicht dem Tool an sich, sondern eher der Oracle DDIC- Struktur und den damit verbundenen notwendigen Abfragen zur Datenextraktion anzulasten.

Das Tool SQLd360 kann jederzeit ausgeführt werden, da es nicht Laufzeit, sondern Last-optimiert ist, das bedeutet, dass die Gesamtlaufzeit an sich keine große Rolle spielt.

Das Ergebnis des Tools ist eine ZIP-Datei, die alle relevanten Informationen enthält (siehe Abbildung 4). Zur Analyse des Outputs öffnet man die HTML-Datei „00001_

sqlid360_<db_name>_<sql_id>_index.html“, von der man anschließend in die verschiedensten Bereiche navigieren kann. Ich persönlich starte meistens mit der Sektion 4b Punkt 122/122, um die historischen Laufzeiten (inklusive Ressourcen-/Datenbankzeitverteilung) mit der berichteten/gemessenen Antwortzeit des Endbenutzers/Businessprozesses abzugleichen und abhängig von dem Analyseoutput in die weiteren Sektionen zu springen (zum Beispiel Analyse von unterschiedlichen Ausführungsplänen, Analyse der bearbeiteten Datenmengen etc.).

Der Nachbau und die Ausführung des SQLs in einer isolierten Testumgebung ist abhängig von der Fragestellung unter anderem auch relativ zügig mithilfe der Sektionen 2a Punkt 23 und 3f (ohne den SQL Test Case Builder Output) realisierbar.

Ein kleiner, aber wichtiger Hinweis: Die Sektion 5a Punkt 145 (Cost Based Optimizer / 10053 Trace) ist unter anderem mit Vorsicht zu genießen, da der CBO Trace basierend auf einem bereits existierenden Parent/Child Cursor mittels DBMS_SQLDIAG.DUMP_TRACE erzeugt wird und diese PL/SQL-Prozedur bei komplexeren SQLs/Umgebungen leider nicht zuverlässig funktioniert.

Tool „Pathfinder“ [7][9]

Das Tool Pathfinder wurde ebenfalls von Mauro Pagano entwickelt und dient nur einem einzigen Zweck: „(bessere) Ausführungspläne für ein spezifisches SQL unter verschiedensten Cost Based Optimizer-Einstellungen zu finden“. Infolgedessen kommt das Tool unter anderem in folgenden Szenarien zum Einsatz:

- nach Datenbankupgrades
- es gibt einen besseren Ausführungsplan mit einem alten OFE- (OPTIMIZER_FEATURES_ENABLE) Wert
- identifizieren der Ursache für die Performance-Verschlechterung aufgrund eines (neuen) Cost Based Optimizer Features

Beispiel: Wenn die vorherigen SQLd360-Analysen ergeben haben, dass sich die Laufzeit eines SQLs aufgrund eines neuen Ausführungsplans (zum Beispiel nach einem Datenbank-Upgrade) verschlechtert hat, möchte man das (neue) Feature, das dafür verantwortlich ist, ohne großes Cost Based Optimizer Know-how identifizieren.

SQLd360 v192: SQL 360-degree view				
sqlid:1rn6029rjafa3 dbname:979647 version:19.0.0.0.0 host:904976 license:7 days:1 today:2019-11-07/13:23:18				
1/5	2/5	3/5	4/5	5/5
<p>1a. Database Configuration</p> <ol style="list-style-type: none"> System Under Observation html (19) Identification html (1) Version html (1) Database html (1) Instance html (1) Modified Parameters html (9) Non-default Parameters html (19) All Parameters html (451) Parameter File html (19) System Parameters Change Log html (9) System Stats html (13) IO Calibration html (9) Non-default Fix Controls html (9) All Fix Controls html (149) Optimizer Processing Rate html (28) <p>1e. NLS Settings</p> <ol style="list-style-type: none"> Session Parameters html (17) Instance Parameters html (17) Database Parameters html (29) <p>1f. Observations</p> <ol style="list-style-type: none"> System-wide observations html (1) Cursors and Plans specific observations html (2) Table-level observations html (9) 	<p>2a. Identification</p> <ol style="list-style-type: none"> Formatted SQL sql Standalone SQL sql Expanded SQL sql SQL Text html (1) SQL Text from AWR html (1) <p>2b. Performance</p> <ol style="list-style-type: none"> SQL Performance Summary html (4) SQL Statistics from Memory html (4) SQL Statistics from Memory (SQLSTATS) html (1) SQL Plan Statistics from Memory html (26) SQL Plan Statistics from Memory (SQLSTATS) html (2) SQL Statistics from History html (4) SQL Plan Statistics from History html (16) <p>2d. Execution Plans</p> <ol style="list-style-type: none"> Plans Details page (2) Plans from Memory text Plans from History text Plan from Plan Table text Plans from Baseline text Plans from SQL Tuning Sets text <p>2e. Plan Control</p> <ol style="list-style-type: none"> SQL Profiles html (4) SQL Profile Data html (9) SQL Plan Baselines html (4) SQL Patches html (4) SQL Plan Directives html (4) SQL Plan Directives Objects html (4) SQL ReOptimization Hints html (4) Mapped SQL html (4) <p>2f. SQL Monitor</p> <ol style="list-style-type: none"> SQL Monitor html (16) SQL Plan Monitor html (156) <p>2g. Binds</p> <ol style="list-style-type: none"> Binds Summary from Memory html (9) Binds List from Memory html (9) Binds Summary from History html (4) Binds List from History html (4) Captured Binds Details page (4) Binds with unstable datatype html (4) <p>2h. Cursor Sharing</p> <ol style="list-style-type: none"> Non-sharing reasons summary html (1) Non-sharing reasons details html (4) ACS Histogram html (4) ACS Statistics html (4) ACS Selectivity html (4) 	<p>3a. Objects</p> <ol style="list-style-type: none"> Tables html (2) Indexes html (4) Index Columns html (4) Index Usage html (4) Columns html (4) Column Usage html (4) Column Stats Extensions html (4) Expression Statistics html (4) Histograms page (4) Histograms on long strings html (4) Constraints html (4) Views html (4) LOBs html (4) Materialize Views html (4) Clusters html (4) Partition Key Columns html (4) Unindexed Partition Key Columns html (4) Table Partition Info html (4) Table Partitions html (4) Index Partition Info html (4) Index Partitions html (4) LOB Partition Info html (4) LOB Partitions html (4) Partitions Columns page (4) Table Subpartitions html (4) Index Subpartitions html (4) LOB Subpartitions html (4) Table Modifications html (4) Table Stats Preferences html (4) Triggers html (4) Policies html (4) Avd Policies html (4) Segments html (2) Objects html (2) Tablespaces html (1) <p>3c. Statistics History</p> <ol style="list-style-type: none"> System Statistics History html (4) Tables Statistics History html (4) Indexes Statistics History html (4) Columns Statistics History html (4) Partitions Statistics History html (4) <p>3d. In-Memory</p> <ol style="list-style-type: none"> In-Memory Area html (4) In-Memory FastStart Area html (1) Automatic In-Memory Tasks html (4) Automatic In-Memory Task Details html (4) In-Memory Segments html (4) In-Memory Header html (4) In-Memory Column Level html (4) In-Memory Join Groups html (4) <p>3e. Big Table Caching</p> <ol style="list-style-type: none"> BT Scan Cache html (1) BT Scan Objects Temps html (4) <p>3f. Metadata</p>	<p>4a. System Impact</p> <ol style="list-style-type: none"> SQL Execute Time for Instance 1 html line (1) Peak Demand for recent executions for Instance 1 html line (4) Peak Demand for historical executions for Instance 1 html line (4) Distinct number of sessions executing for Instance 1 html line (4) Total Buffer Gets for Instance 1 html line (4) Total Rows Processed per for Instance 1 html line (4) Total number of Executions for Instance 1 html line (4) Version Count and Shareable Memory for Instance 1 html line (4) <p>4b. Execution metrics</p> <ol style="list-style-type: none"> SQL Execute Time by Wait Class for Instance 1 html line (4) Elapsed Time per recent executions for Instance 1 html line (4) Elapsed Time per historical executions for Instance 1 html line (4) Avg Buffer Gets/Execution for Instance 1 html line (4) Avg Rows Processed/Execution for Instance 1 html line (4) Avg Elapsed Time/Execution (total) for Instance 1 html line (4) Avg Elapsed Time/Execution (delta) for Instance 1 html line (4) Avg Elapsed Time/Execution (moving 1d) for Instance 1 html line (4) Avg and Median Elapsed Time/Execution for recent executions for Instance 1 html line (4) Avg and Median Elapsed Time/Execution for historical executions for Instance 1 html line (4) Streak of non-executing SQL for historical executions for Instance 1 html line (4) Streak of non-executing SQL for historical executions for Instance 1 html line (4) <p>4c. Execution metrics per PHV</p> <ol style="list-style-type: none"> Total number of Executions per PHV for Instance 1 html line (4) Number of execution per PHV from AWR for Instance 1 html line (4) Total Elapsed Time for recent executions per PHV for Instance 1 html line (4) Total Elapsed Time per PHV from AWR for Instance 1 html line (4) Avg Elapsed Time/Execution per PHV for recent executions for Instance 1 html line (4) Avg Elapsed Time/Execution per PHV for historical executions for Instance 1 html line (4) Avg Elapsed Time/Execution per PHV (total) from AWR for Instance 1 html line (4) Avg Elapsed Time/Execution per PHV (delta) from AWR for Instance 1 html line (4) Avg Elapsed Time/Execution per PHV (moving 1d) from AWR for Instance 1 html line (4) Avg Buffer Gets/Row/Execution per PHV from AWR for Instance 1 html line (4) Execution Plan Cost per PHV from AWR for Instance 1 html line (4) 	<p>5a. 10053 Trace</p> <ol style="list-style-type: none"> 10053 Trace text <p>5b. SQL Monitor Reports</p> <ol style="list-style-type: none"> SQL Monitor Reports zip <p>5f. ASH Raw Data</p> <ol style="list-style-type: none"> Raw Data html (791) <p>5g. eAdam</p> <ol style="list-style-type: none"> eAdam html (14) <p>5h. SQL Tuning Advisor</p> <ol style="list-style-type: none"> STA report text

Abbildung 4: Output von SQLd360 (© Stefan Koehler)

Das aktuelle Release von Pathfinder kann von GitHub [7][9] als ZIP-Datei heruntergeladen werden, ist ebenfalls in dem Tool-Merge „sqldb360“ enthalten und kann sowohl mit der Oracle Standard als auch Enterprise Edition eingesetzt werden. Nach dem Download von GitHub wird die entsprechende ZIP-Datei auf dem Datenbankserver oder einem Client mit einer entsprechenden SQL*Net-Verbindung extrahiert. Das Tool besteht aus nur 2 SQL-Skripten und hat ebenso wie SQLd360 den Vorteil, dass keine Installation in der Datenbank erfolgen muss:

- pathfinder.sql - SQL Skript zum Start des Tools mit entsprechendem Connect-String zum Applikations-/Parsing-Schema (siehe Abbildung 6).
- script.sql - Datei, in der das zu analysierende SQL mithilfe des Kommentars /* ^^pathfinder_testid */ gespeichert wird. Im Prinzip sind hier keine Grenzen gesetzt, das heißt, dass das SQL exakt so wie von der eigentlichen Applikation ausgeführt werden kann (zum Beispiel mit Bind Variablen, sessionabhängigen Parametern, OCI oder NLS-Einstellungen etc.).

In Abbildung 5 sieht man eine exemplarisch zu analysierende Abfrage (in diesem Fall das zugegebenermaßen etwas unspektakuläre SQL „select * from dual“) und den notwendigen Kommentar („/*

^^pathfinder_testid */“), der von Pathfinder zur Laufzeit durch eine eindeutige ID ersetzt wird - mit diesem Verfahren kann jede erdenkliche Cost Based Optimizer-Einstellung separat und dediziert getestet werden.

Das Ergebnis des Tools ist eine ZIP-Datei, die alle relevanten Informationen (u.a. einzeln getestete Cost Based Optimizer-Einstellungen und deren Ausführungspläne, Laufzeit des SQLs unter den jeweiligen Cost Based Optimizer-Einstellungen etc.) enthält (siehe Abbildung 7). Zur Analyse des Outputs öffnet man die HTML-Datei „00001_pathfinder_<dbname>_<date>_index.html“, von der man anschließend zu den einzelnen Ausführungen mit der jeweiligen Cost Based Optimizer-Einstellung navigieren kann.

Die Gesamtlaufzeit des Tools ist abhängig von der Laufzeit der einzelnen SQL-Ausführungen unter der jeweiligen Cost Based Optimizer-Einstellung, d.h. wenn es sich um einen Langläufer (zum Beispiel Ø Ausführungszeit von 2 Minuten) handelt, wird auch das Tool entsprechend lange (zum Beispiel Ø 2 Minuten x 2.091 Ausführungen unter Oracle 19c = 4.182 Minuten) benötigen - im Prinzip handelt es sich bei Pathfinder um eine Brute-Force-Methode zum Auffinden und Testen der möglichen Ausführungspläne.

Wenn die Laufzeitstatistiken für die einzelnen Testläufe beziehungsweise zu den entsprechenden Ausführungsplänen primär nicht relevant sind und nur die Ausführungspläne (u.a. Hash Value, Predicate Information, etc.) analysiert bzw.

```

/*
 * PLNFND script example
 * Name of the script must be script.sql
 */

/*
 * Add here the ALTER SESSIONS
 */

/*
 * Add here the SQL text, remember the mandatory
 * comment ^^pathfinder_testid
 */
select /* ^^pathfinder_testid */ * from dual;

```

Abbildung 5: Von Pathfinder zu analysierendes SQL (Quelle: Stefan Koehler)

```

T193DB [oracle@OEL76 pathfinder-master]$ sqlplus / as sysdba

SQL*Plus: Release 19.0.0.0.0 - Production on Thu Nov 7 14:35:32 2019
Version 19.3.0.0.0

Copyright (c) 1982, 2019, Oracle. All rights reserved.

Connected to:
Oracle Database 19c Enterprise Edition Release 19.0.0.0.0 - Production
Version 19.3.0.0.0

SQL> @pathfinder.sql

Parameter 1:
Full connect string of the database to run the SQL into
If the database is remote or a PDB then you must include
the TNS alias i.e. scott/tiger@orcl

Enter value for 1: test/test@T193DB
test/test@T193DB
Building Pathfinder driver scripts
Connected.
1) "pathfinder_{ 20191107_1435 (00001)" 14:35:52 BASELINE
2) "pathfinder_{ 20191107_1435 (00002)" 14:35:53 "_adaptive_window_consolidator_enabled" = FALSE
3) "pathfinder_{ 20191107_1435 (00003)" 14:35:54 "_add_stale_mv_to_dependency_list" = FALSE
4) "pathfinder_{ 20191107_1435 (00004)" 14:35:54 "_allow_level_without_connect_by" = TRUE
5) "pathfinder_{ 20191107_1435 (00005)" 14:35:55 "_always_anti_join" = 'CHOOSE'
6) "pathfinder_{ 20191107_1435 (00006)" 14:35:55 "_always_anti_join" = 'CUBE'
7) "pathfinder_{ 20191107_1435 (00007)" 14:35:56 "_always_anti_join" = 'HASH'

```

Abbildung 6: Aufruf von Pathfinder (Quelle: Stefan Koehler)

Pathfinder v1601 (2016-01-03): Plan Finder

dbname:upgr version:19.0.0.0.0 connect string:// as sysdba starttime:2019/05/16 12:15:01

Test#	CFB Reparse#	Setting	Plan Hash Value	Elapsed Time	CPU Time	Buffer Gets	Rows Processed	Execution Plan	V\$SQL Details
502		ALTER SESSION SET approx_for_aggregation = TRUE;	589055361	.044462	.04221	113	0	plan	details
503		ALTER SESSION SET approx_for_count_distinct = TRUE;	589055361	.015754	.014451	35	0	plan	details
291		ALTER SESSION SET "_optimizer_partial_join_eval" = FALSE;	954326358	.061731	.058793	1188	0	plan	details
555		ALTER SESSION SET optimizer_features_enable = '11.2.0.1';	954326358	.029231	.028356	35	0	plan	details
556		ALTER SESSION SET optimizer_features_enable = '11.2.0.2';	954326358	.021456	.020419	39	0	plan	details
557		ALTER SESSION SET optimizer_features_enable = '11.2.0.3';	954326358	.014509	.013715	41	0	plan	details
558		ALTER SESSION SET optimizer_features_enable = '11.2.0.4';	954326358	.015812	.015	41	0	plan	details
156		ALTER SESSION SET "_optimizer_allow_all_access_paths" = FALSE;	1835045712	.03494	.032207	100	0	plan	details
595		ALTER SESSION SET optimizer_index_cost_adj = 10000;	2194469899	.070813	.064245	1759	0	plan	details
232		ALTER SESSION SET "_optimizer_filter_pushdown" = FALSE;	2602649900	.019279	.017964	44	0	plan	details
246		ALTER SESSION SET "_optimizer_generate_transitive_pred" = FALSE;	2602649900	.01331	.012829	33	0	plan	details
1		BASELINE	3300316041	.026976	.024519	34	0	plan	details
2		ALTER SESSION SET "_adaptive_window_consolidator_enabled" = FALSE;	3300316041	.009946	.008316	51	0	plan	details
3		ALTER SESSION SET "_add_stale_mv_to_dependency_list" = FALSE;	3300316041	.00632	.005683	32	0	plan	details
4		ALTER SESSION SET "_allow_level_without_connect_by" = TRUE;	3300316041	.010221	.008471	32	0	plan	details
5		ALTER SESSION SET "_always_anti_join" = 'CHOOSE';	3300316041	.011335	.010089	32	0	plan	details
6		ALTER SESSION SET "_always_anti_join" = 'CUBE';	3300316041	.014757	.012705	51	0	plan	details
7		ALTER SESSION SET "_always_anti_join" = 'HASH';	3300316041	.010657	.009042	51	0	plan	details
8		ALTER SESSION SET "_always_anti_join" = 'MERGE';	3300316041	.01215	.010064	51	0	plan	details
9		ALTER SESSION SET "_always_anti_join" = 'NESTED_LOOPS';	3300316041	.022302	.019826	51	0	plan	details
10		ALTER SESSION SET "_always_anti_join" = 'OFF';	3300316041	.018607	.016365	51	0	plan	details
11		ALTER SESSION SET "_always_semi_join" = 'CHOOSE';	3300316041	.013667	.01291	39	0	plan	details
12		ALTER SESSION SET "_always_semi_join" = 'CUBE';	3300316041	.021739	.019734	51	0	plan	details
13		ALTER SESSION SET "_always_semi_join" = 'HASH';	3300316041	.054037	.0489	87	0	plan	details
14		ALTER SESSION SET "_always_semi_join" = 'MERGE';	3300316041	.022608	.020372	53	0	plan	details
15		ALTER SESSION SET "_always_semi_join" = 'NESTED_LOOPS';	3300316041	.030332	.027345	53	0	plan	details

Abbildung 7: Output von Pathfinder (Quelle: Stefan Koehler)

verglichen werden sollen, kann Pathfinder auch nur mit der „EXPLAIN PLAN“ Option ausgeführt werden [10] - hierbei sind die Limitierungen (zum Beispiel kein Bind Peeking etc.) von „EXPLAIN PLAN“ aber unbedingt zu beachten, damit das zu analysierende SQL im Skript „script.sql“ entsprechend angepasst werden kann.

Fazit

Mithilfe des systematischen „Method-R“-Tuning-Ansatzes lassen sich Performance-Probleme sehr zielgerichtet sowie effizient analysieren und entsprechend verlässliche Performance Forecasts der geplanten Tuningmaßnahmen erstellen. Das Tool SQLd360 unterstützt den DBA oder involvierten externen Berater bei weiterführenden Analysen, falls die Ursache des Performance-Problems an einzelnen SQLs oder deren Ausführungsplänen liegt.

Das Tool Pathfinder unterstützt den DBA bei der weiterführenden Analyse von Ausführungsplänen und den entsprechend dafür verantwortlichen Cost Based Optimizer Features in ganz bestimmten Szenarien (Upgrades/Updates) ohne dass der DBA dabei tiefes Cost Based Optimizer-Wissen mitbringen muss.

Quellen

- [1] Buch: Optimizing Oracle Performance - ISBN: 978-0596005276
- [2] Software Method-R: <https://method-r.com/software/workbench/>
- [3] Software TVD\$XTAT: <https://antognini.ch/2017/03/tvdxtat-4-0-beta-11/>
- [4] Buch: Troubleshooting Oracle Performance - ISBN: 978-1-4302-5759-2
- [5] Software OraSASH: <https://pioro.github.io/orasash/>
- [6] ASH Analyse Skripte: <https://github.com/tanelpoder/tpt-oracle/blob/master/ash/ashtop.sql> & <https://github.com/tanelpoder/tpt-oracle/blob/master/ash/dashtop.sql>
- [7] Software SQLd360: <https://github.com/sqldb360/sqldb360/releases>
- [8] Vergleich SQLd360/SQLTXPLAIN: <https://mauro-pagano.com/2017/04/29/sqltxplain-vs-sql360-differences-and-similarities/>
- [9] Software Pathfinder: <https://github.com/mauropagano/pathfinder>
- [10] Blog Post „EXPLAIN PLAN“ Pathfinder: <https://medium.com/@FranckPachot/fixing-execution-plan-regression-with-optimizer-features-enable-what-to-do-next-ddb99bb7ba15>

Über den Autor

Stefan Koehler folgt seiner Passion für die Oracle-Datenbank seit nun bereits über 18 Jahren und hat sich auf das Thema Oracle

Performance Tuning, insbesondere dem Oracle Cost Based Optimizer (CBO) und den Oracle RDBMS Core Internals spezialisiert. Heutzutage arbeitet er weltweit als freiberuflicher Oracle Performance Berater & Researcher in größtenteils "Mission Critical"-Umgebungen für marktführende Firmen in unterschiedlichen Branchen. Er unterstützt seine Kunden bei der Analyse, dem Verständnis und beim Lösen von komplexen Oracle- Performance-Problemen sowie beim Troubleshooting von nicht trivialen Oracle- Datenbankproblemen auf Basis von sehr kurzfristigen Beratungseinsätzen. Er ist ebenso Mitglied des OakTable-Netzwerks und wurde durch Oracles ACE-Programm als Oracle ACE im Bereich "Database Management & Performance" ausgezeichnet.



Stefan Koehler
contact@soocs.de



Analyse von ortsbezogenen Daten in der Oracle DB

Dipl.-Inf. Karin Patenge, Oracle Deutschland

Mit Daten zu arbeiten, diese in Mehrwerte umzusetzen, das ist wohl tägliches Geschäft für sehr viele von uns, nicht nur in der IT. Die typischen Arbeitsschritte für die Analyse von Daten mit Ortsbezug unterscheiden sich dabei nicht von denen anderer Daten. Ortsbezogene Daten werden geladen, validiert, angereichert um Geokoordinaten durch Umrechnung von Adressen oder anderen Ortsinformationen, räumlich analysiert und zumeist in Form interaktiver Karten dargestellt. Die aktuell wohl am meisten Beachtung findenden Beispiele für solche Karten beziehen sich auf Daten zu und Ergebnisse von Analysen zur weltweiten Verbreitung von COVID-19. Dieser Artikel soll in gestraffter Form einen Workflow für die Verarbeitung von Geodaten zeigen. Als Werkzeuge werden dafür die Oracle-Datenbank sowie Oracle Spatial Studio verwendet (siehe Abbildung 1). Letzteres ist eine Webanwendung, mit der Sie deklarativ und somit auf einfache Weise die einzelnen Arbeitsschritte wie das Laden von ortsbezogenen Daten, deren Analyse sowie das Erstellen und Publizieren in Form von interaktiven Karten abarbeiten können.

Das Wichtigste zur Geodatenunterstützung in der Oracle-Datenbank

Geodaten, häufig auch räumliche oder ortsbezogene Daten genannt, sind für effiziente Prozesse auf nahezu allen Ebenen der Wirtschaft, Verwaltung und Gesellschaft relevant. Ihre Verarbeitung und Nutzung im Zusammenhang mit anderen Daten ermöglicht es, komplexe Sachverhalte einfacher, schneller und transparenter zu gestalten sowie die Analyse-Ergebnisse intuitiv verständlich darzustellen. Eine große Anzahl aller Planungs- und Entscheidungsprozesse hat einen Ortsbezug. Daher sind Geoinformationen eine universell anwendbare Schlüsselressource der Informationsgesellschaft. Anwendungsbeispiele lassen sich in allen Branchen finden. Sei es das Validieren von Adressdatenbeständen bei Behörden, die Vorhersage von Wetter auf der Basis von Satellitenbildern und Messdaten, die Planung und Optimierung von Vertriebsgebieten, das Finden der besten Standorte für Sonnenkollektoren, die Sicherung von Großveranstaltungen oder die Suche nach der ökonomischsten Route, um von A nach B zu kommen unter Berücksichtigung der aktuellen Verkehrssituation sowie der Verfügbarkeit von Ladestationen. Jede Edition der Oracle-Datenbank stellt für die Arbeit native Datentypen (Vektor- oder Rasterdaten) und Datenmodelle (Netzwerke, Punktwolken, und mehr) für alle gängigen Arten von Geodaten bereit. Diese werden komplettiert durch Domänen-spezifische Indizes sowie Operatoren und Funktionen für die Prozessierung und Analyse dieser Daten.

Die ersten Schritte mit Geodaten in der Oracle DB

Ausschlaggebend für den Einstieg in die Welt der Geodaten sind sehr häufig das Vorliegen von Adressdaten auf der einen Seite oder fachliche Fragestellungen, die mit dem Wort „Wo“ anfangen, auf der anderen Seite. Sollen Adressdaten auf einer Karte angezeigt werden, müssen diese Adressen immer in Geokoordinaten umgerechnet werden. Das gilt gleichermaßen auch für die Beantwortung der „Wo“-Fragen. Die Umrechnung von Adressen in Geokoordinaten kann entweder direkt

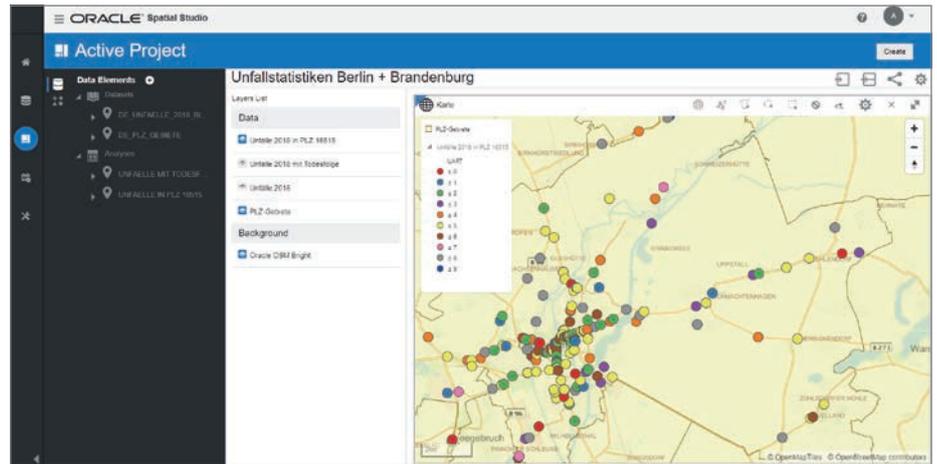


Abbildung 1: Projekt mit Geodaten in Oracle Spatial Studio (Datenquelle Unfalldaten: <https://opendata-esri-de.opendata.arcgis.com>)

in der Datenbank mittels SQL erfolgen [1] oder als Service über XML Requests / Responses [2]. Beide Varianten setzen jeweils einen Referenzdatenbestand voraus. Das Geocoding Plug-in als Teil des Sample Geolocation Showcase (Autor: Carsten Czarski) für Oracle Application Express beispielsweise nutzt die Service-Variante [3] ebenso wie Oracle Spatial Studio, ein weiteres kostenfreies Werkzeug aus dem Oracle-Software-Portfolio. Spatial Studio ist eine Browser-basierte Anwendung, die Adressen im Batch-Verfahren in die Oracle-Datenbank lädt, geocodiert und

gleich die räumlichen Indizes anlegt. Fehler oder nicht eindeutige Ergebnisse bei der Geokodierung können auf Datensatzebene nachvollzogen und, wenn nötig, leicht korrigiert werden.

Hinweis:

In diesem Artikel gebe ich keine detaillierte Beschreibung zur Installation von Oracle Spatial Studio. Eine solche finden Sie unter anderem im Oracle Technology Network [4]. Die im Artikel gezeigten Abbildungen beziehen sich auf eine Oracle-Spatial-Studio-Installation auf einer Micro-

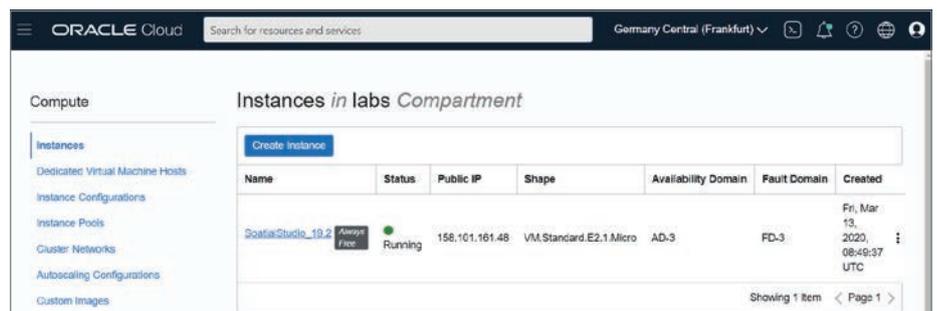


Abbildung 2: Spatial Studio auf Compute-Instanz in Oracle Cloud (Free Tier Account) (Quelle: Karin Patenge)

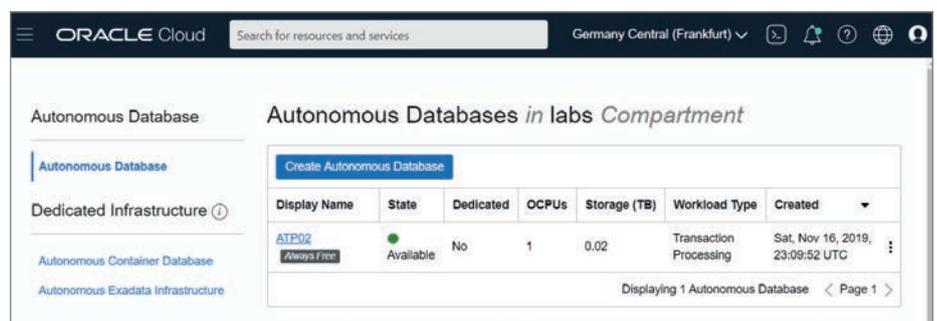


Abbildung 3: Autonomous Database in Oracle Cloud (Free Tier Account) (Quelle: Karin Patenge)

```

Verkaufsstellennr;Name;Strasse;PLZ;Ort
501072;Reisezentrum Berlin Alexanderplatz;Dircksenstraße 7;10178;Berlin
501148;Reisezentrum Berlin Ostbahnhof;Am Ostbahnhof 9;10243;Berlin
509273;Reisezentrum Berlin Hbf;Europaplatz 1;10557;Berlin
501270;Reisezentrum Berlin Zoologischer Garten;Hardenbergplatz 11;10623;Berlin
509380;Reisezentrum Berlin Südkreuz;General-Pape-Straße 1;12101;Berlin
    
```

Listing 1: Beispieldaten (Quelle: Open Data-Portal der Deutschen Bahn AG)

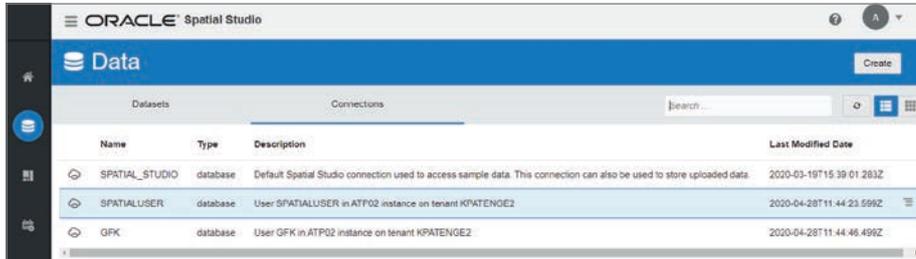


Abbildung 4: Oracle Spatial Studio – bereits definierte „Connections“ (Quelle: Karin Patenge)

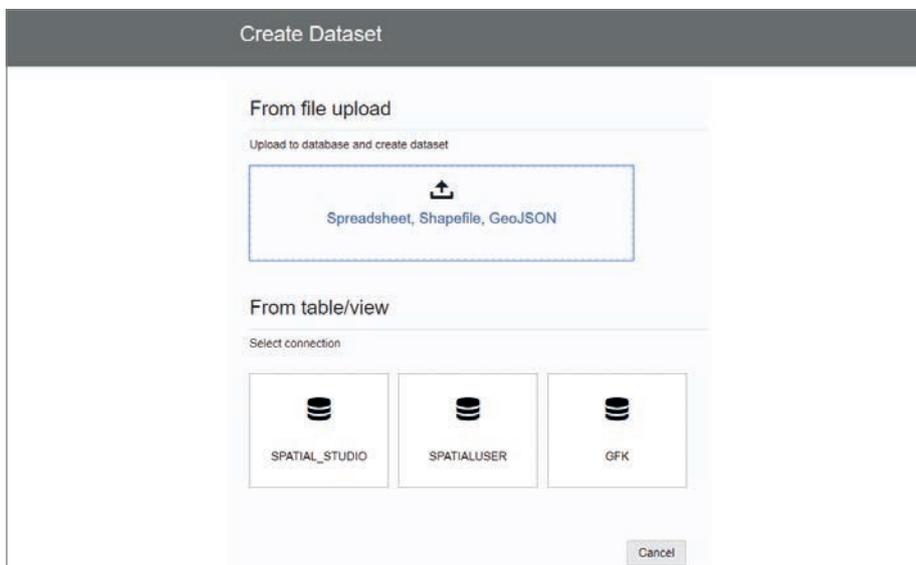


Abbildung 5: Oracle Spatial Studio – Datenset erzeugen (Quelle: Karin Patenge)

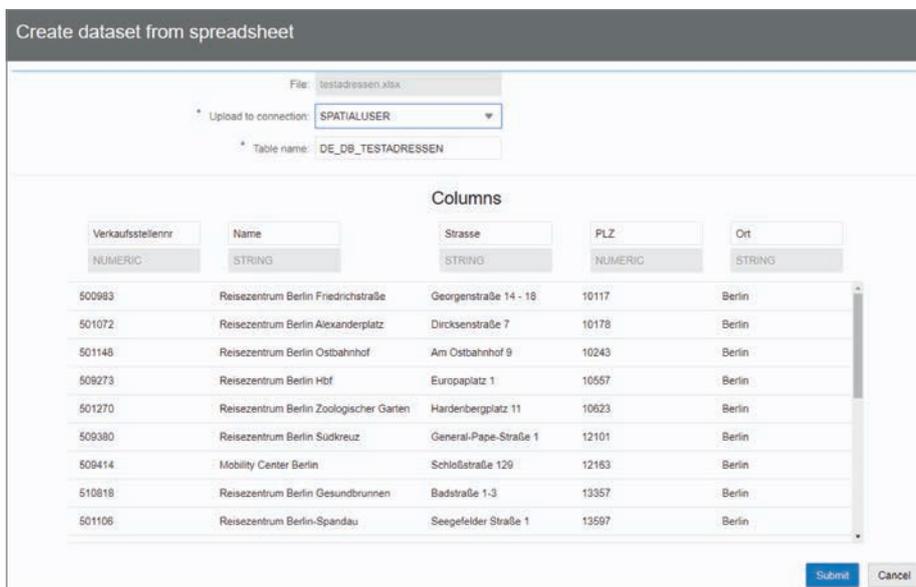


Abbildung 6: Oracle Spatial Studio – Auswahl des DB-Anwendungsschemas und Angabe des Tabellennamens (Quelle: Karin Patenge)

Compute-Instanz (siehe Abbildung 2) in der Oracle Cloud (Free Tier) [5]. Die Metadaten für Oracle Spatial Studio liegen ebenso wie die Anwendungsdaten in einer Oracle-Autonomous-Database-Instanz im gleichen Free Tier Account (siehe Abbildung 3). Die verwendete Umgebung ist somit komplett (lizenz-)kostenfrei.

Geokodieren von Adressdaten einfach gemacht mit Oracle Spatial Studio

Adressdaten sind häufig bereits in Datenbank-Tabellen erfasst und gepflegt. Wahrscheinlich genauso häufig werden sie in Form von Textdateien (.csv) oder Excel-Tabellen bereitgestellt. Oracle Spatial Studio kann Excel-Tabellen direkt verarbeiten, also ohne vorherige Konvertierung. Für eine erste Geokodierung nehme ich Adressdaten von DB-Reisezentren, die im Open-Data-Portal der Deutschen Bahn [6] bereitgestellt werden. Einen Auszug davon finden Sie in Listing 1.

Diese Daten können im „Data“-Bereich von Oracle Spatial Studio geladen werden. Das Laden setzt allerdings voraus, dass zuvor eine „Connection“ auf das Oracle-DB-Schema (meine Autonomous-DB-Instanz) eingerichtet wurde, welches die Daten aufnehmen soll. Eine solche Verbindung ist in Abbildung 4 zu sehen.

Per Drag & Drop kann die Excel-Tabelle mit den Beispieldaten in den „From file upload“-Bereich gezogen werden (siehe Abbildung 5).

Danach sind noch das DB-Anwendungsschema auszuwählen, in dem die Beispieldaten abgelegt werden sollen, sowie der Name für die zu erzeugende Tabelle anzugeben (siehe Abbildung 6).

Bevor es mit der Geokodierung weitergeht, muss sichergestellt werden, dass für die angelegte Tabelle ein eindeutiger Schlüsselwert ausgewählt ist beziehungsweise wird. Hierfür gibt es im Kontextmenü den Punkt „Properties“ und die Ansicht

„Columns“. Im Beispiel für das Laden einer Excel-Tabelle fügt Spatial Studio bereits eine eindeutige ID mit dem Namen STUDIO_ID hinzu (siehe Abbildung 7).

Über „Prepare“ und „Geocode Address“ im Kontextmenü für das neu erstellte Datenset wird eine Maske geöffnet, um die Attribute auszuwählen, die für die Geokodierung benötigt werden. In Abbildung 8 ist zu sehen, in welcher Weise die Spalten des Datensets für die Geokodierung zugeordnet werden:

- Straße -> „Address“
- Ort -> „City“
- PLZ -> „Postal/Zip Code“

„Country“ wird zentral für alle Adressdatensätze auf den Wert „Germany“ gesetzt.

Die Geokodierung selbst erfolgt im Batch-Modus. Das Ergebnis der Verarbeitung selbst wird auch in einer Tabelle gespeichert. Diese Tabelle lässt sich einfach am Präfix SGTECH\$TABLE erkennen. In Listing 2 werden die Ergebnisse mittels SQL ausgelesen. Alternativ kann das Ergebnis auch im Status-Bereich des Geokodierungsaufrufs (siehe Abbildung 9) eingesehen und nachbearbeitet werden.

Schauen wir das Ergebnis mal etwas genauer an: Für die Beispieldaten gab es insgesamt zweiunddreißig partielle sowie einen nicht eindeutigen (engl.: ambiguous) Treffer. Bei geringer Qualität der Ausgangsdaten und auch in Abhängigkeit von den im Hintergrund verwendeten Referenzdaten kann die Anzahl der nicht eindeutigen Treffer auch höher ausfallen. Spatial Studio bietet allerdings Hilfestellung in Form einer Auswahl- und Editierungsmaske an, um die passende Adresse auszuwählen oder eine Adresse manuell zu korrigieren und damit die Geokodierung möglichst komplett abzuschließen. Abbildung 10 zeigt die Auswahlmaske an, über welche die richtige aus den möglichen Adressen des Referenzdatenbestandes übernommen werden kann.

Die Geokodierung erzeugt in der Tabelle mit den Adressdaten eine neue Spalte mit dem Namen GC_GEOMETRY. Diese ist vom Typ SDO_GEOMETRY, der native, standardkonforme Datentyp in der Oracle DB für alle Geodaten, die geometrisch gesehen einen Punkt, eine Linie oder Fläche abbilden. Noch ein wenig genauer hingeschaut (dieses Mal mithilfe des SQL Developer), ist zu sehen, dass auch gleich ein räumlicher

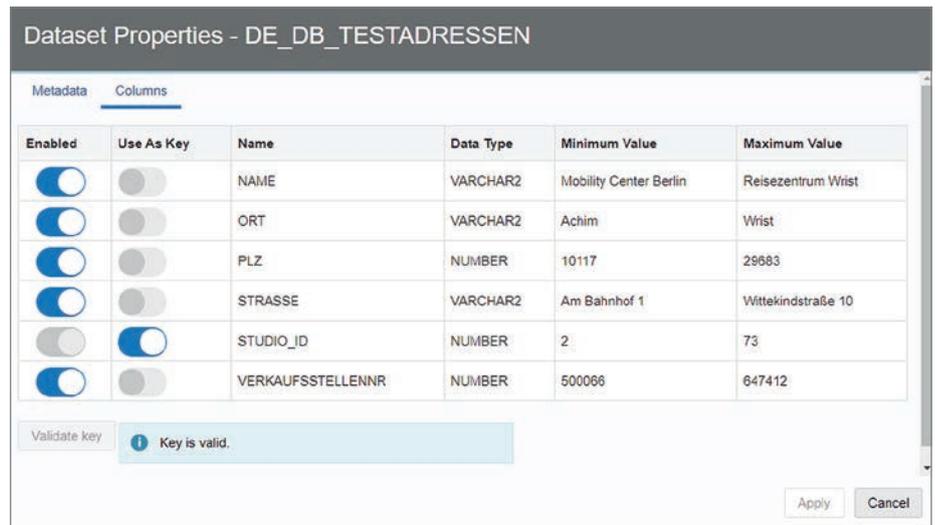


Abbildung 7: Oracle Spatial Studio – Eigenschaften des Datensets (Quelle: Karin Patenge)

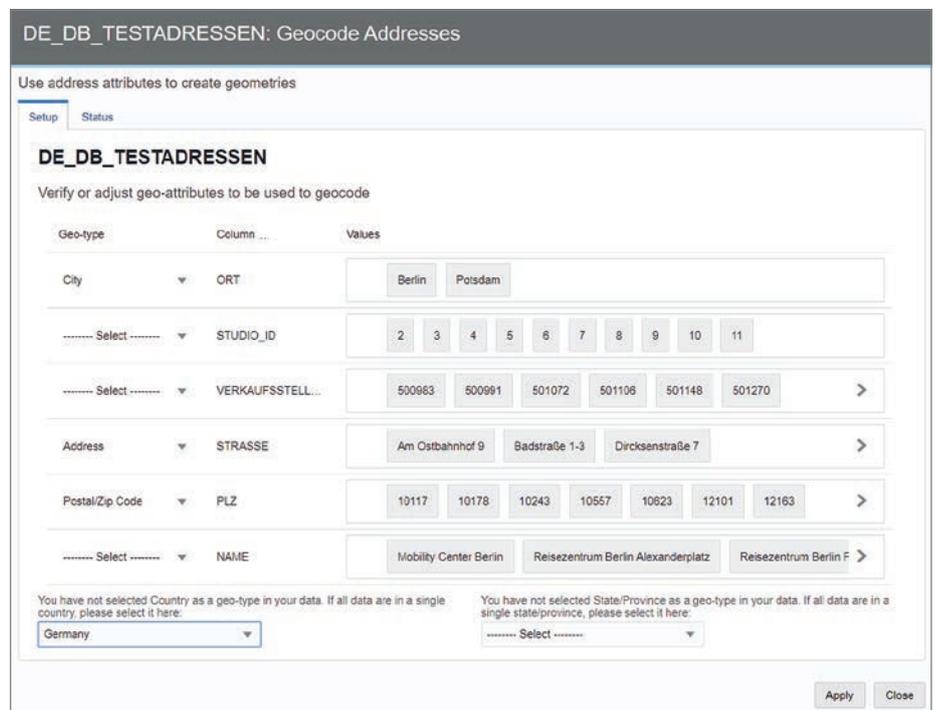


Abbildung 8: Oracle Spatial Studio – Attribut-Mapping für Geokodierung (Quelle: Karin Patenge)

```
SELECT *
FROM DE_DB_TESTADRESSEN dst
LEFT JOIN SGTECH$TABLE8DE6508AF8454945396AD96FC356716E temp
ON dst.ROWID = temp.ROWID_COLUMN
WHERE temp.GC_STATUS != 'MATCH'
```

Listing 2: Oracle Spatial Studio – Abfrage für Ergebnis einer Geokodierung

Index auf der Spalte GC_GEOMETR angelegt wurde. Diesen Index erkennt man am Typ „DOMAIN“. Und noch etwas passiert im Hintergrund: Für jede Spalte vom Typ SDO_GEOMETRY werden Metadaten registriert werden, die unter anderem das verwendete Koordinatenbezugssystem und

die räumliche Ausdehnung angeben. Das Koordinatenbezugssystem ist für die geladenen Adressdaten einheitlich „4326“, auch bekannt als WGS84. Wenn es genauer interessiert, findet Informationen zu Koordinatenbezugssystemen unter anderem auf <http://epsg.io>.

Die erste Karte

Die Adressdaten sind jetzt also geokodiert, mit einem räumlichen Index versehen. Mit wenigen Klicks lässt sich nun daraus eine interaktive Karte bauen. Dazu wechselt man in den „Projects“-Bereich in Spatial Studio und legt mit „Create“ ein neues Projekt an. Die Adressdaten werden über das Plus-Zeichen bei „Data Element“ hinzugefügt und dann als zusätzlicher Layer einfach auf die automatisch einblendete Hintergrundkarte gezogen. Ohne „Verschönerungsarbeiten“ sieht das Projekt mit der Karte aus wie in *Abbildung 11*.

Eine Karte setzt sich also immer zusammen aus einem Hintergrund und einem oder mehreren Layers. Der Layer mit Adressdaten wird als rote Punkte auf der Karte dargestellt. Symbolik, Beschriftung, Darstellung und mehr kann für jeden Layer angepasst werden. Ebenso können Filter definiert und der gesamten Kartenansicht Kontrollelemente hinzugefügt werden. Eine genauere Beschreibung dieser Möglichkeiten soll aber nicht Gegenstand dieses Artikels sein.

Vielmehr soll es im vorletzten Abschnitt darum gehen, das erstellte Projekt um räumliche Analysen zu erweitern.

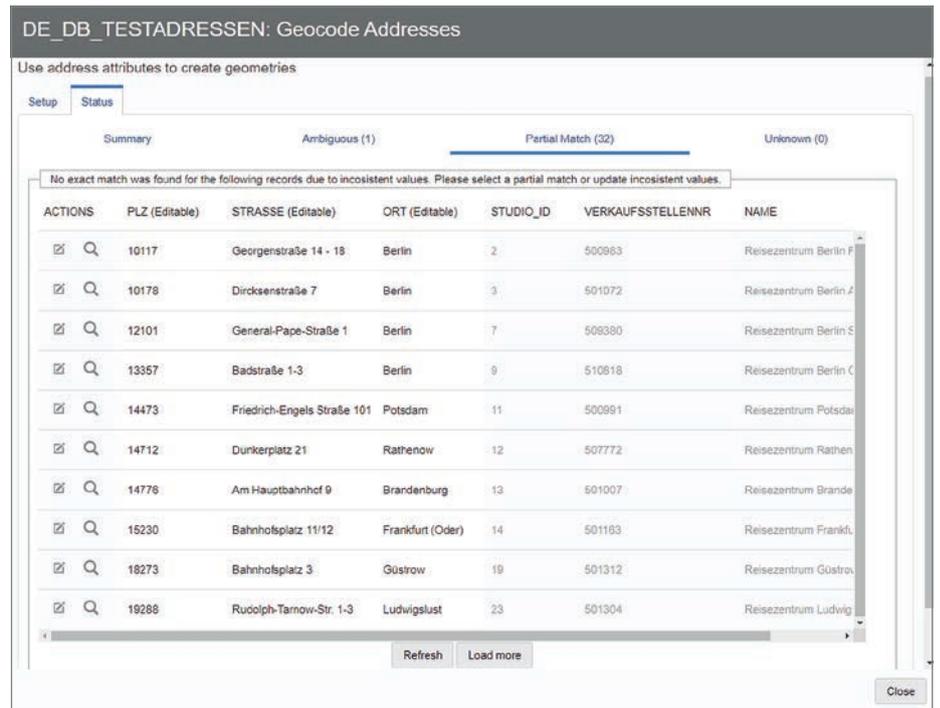


Abbildung 9: Oracle Spatial Studio – Anzeige Ergebnis der Geokodierung (Quelle: Karin Patenge)



Abbildung 10: Oracle Spatial Studio – Bereinigung von nicht eindeutigen Treffern bei Geokodierung (Quelle: Karin Patenge)

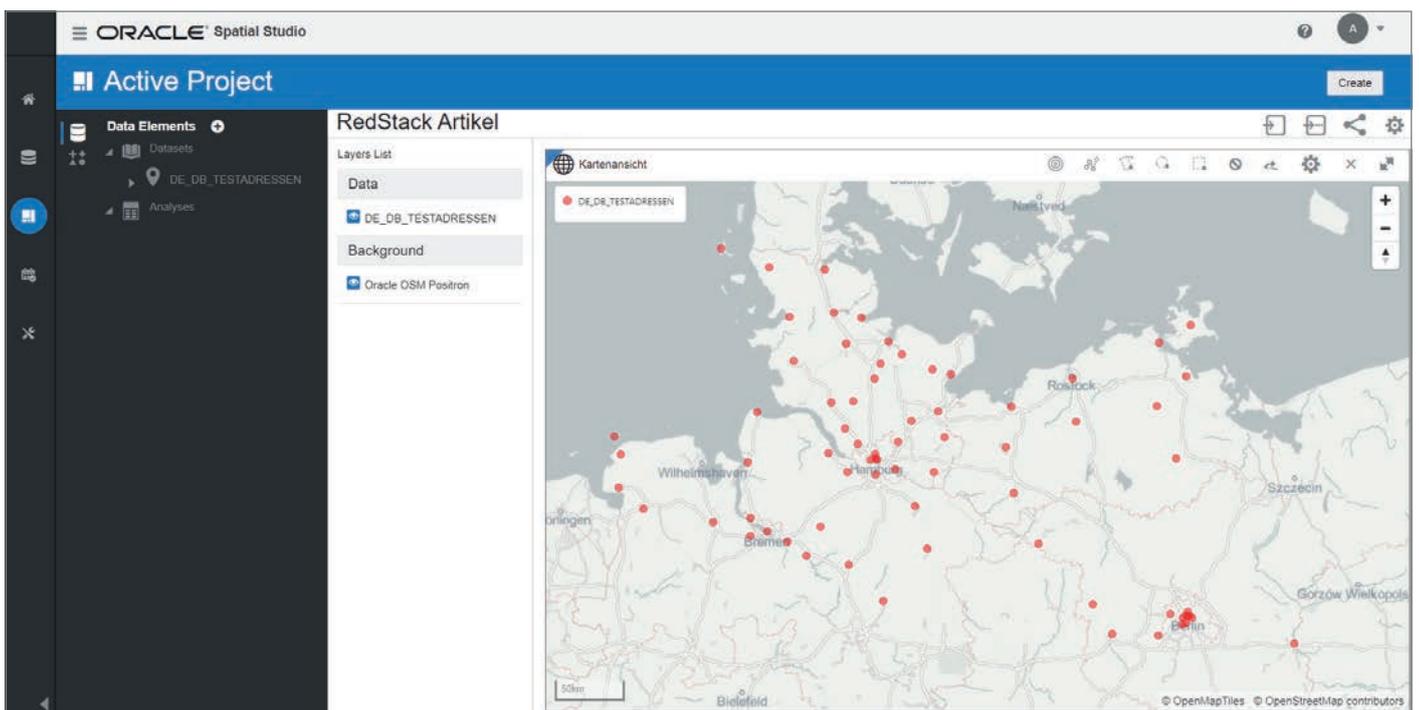


Abbildung 11: Projekt mit interaktiver Karte (Quelle: Karin Patenge)

```

SELECT A.NAME
  FROM DE_DB_TESTADRESSEN A, DE_DB_TESTADRESSEN B
 WHERE B.NAME = 'Reisezentrum Berlin Hbf'
    AND SDO_WITHIN_DISTANCE(a.GC_GEOMETRY, B.GC_GEOMETRY, 'DISTANCE=10 UNIT=KM') = 'TRUE'

```

Listing 3: SQL-Abfrage „Within Distance“

Eine erste räumliche Analyse

Bei räumlichen Analysen geht es zumeist darum, wie ein Ort räumlich in Bezug zu anderen Orten steht. Also zum Beispiel für die vorliegenden Adressdaten:

Welche anderen DB-Reisezentren liegen in maximal zehn Kilometer Entfernung zu einem bestimmten DB-Reisezentrum?

Wer sich mit Geodaten in der Oracle DB schon mal beschäftigt hat und zudem die SQL-Syntax kennt, führt eine SQL-Abfrage

wie in Listing 3 aus unter Verwendung des Operators SDO_WITHIN_DISTANCE.

Wer sich (noch) nicht auskennt, zudem keine Lust auf Handbuch [7] oder Lernen der Syntax hat, nutzt die in Spatial Studio vorgefertigten Analyse-Elemente (siehe Abbildung 12).

Zunächst wird in der Karte der Bezugspunkt, also ein DB-Reisezentrum, selektiert.

Über „Data Elements“ kann jetzt eine neue Analyse erzeugt werden, indem aus der Liste der „Spatial Analysis Operations“ das passende Analyse-Element ausgewählt wird. Abbildung 13 zeigt die Parametrisierung für eine WITHIN DISTANCE -Analyse und Abbildung 14 das Ergebnis der Analyse, welcher als weiterer Layer auf die Karte gezogen wurde.

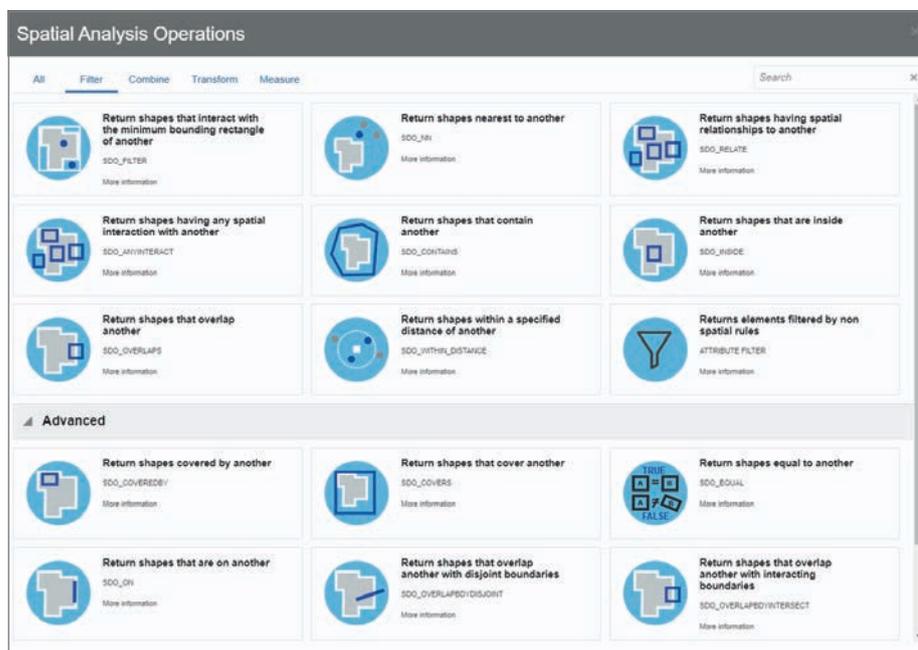


Abbildung 12: Oracle Spatial Studio – Ausschnitt der verfügbaren Spatial Analysis Operations (Quelle: Karin Patenge)

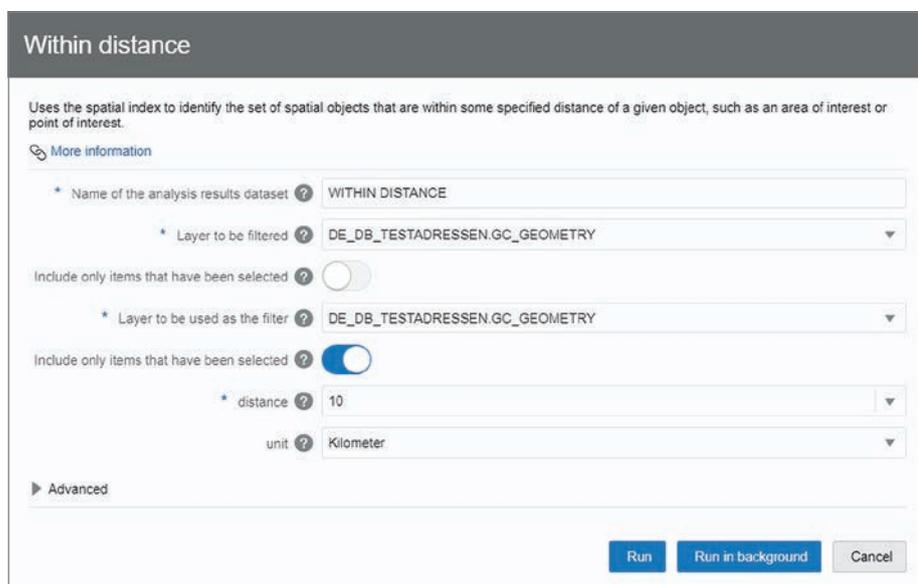


Abbildung 13: Oracle Spatial Studio – räumliche Analyse definieren (Quelle: Karin Patenge)

Zusammenfassung

Die Oracle-Datenbank hat mit der lizenzzeitigen Freigabe der bisherigen beiden Datenbank-Optionen Spatial and Graph sowie Advanced Analytics für alle DB- Editionen im Dezember 2019 wesentlich an Attraktivität für die Developer Community gewonnen. Unter diese Freigabe fallen auch Werkzeuge wie Spatial Studio oder die Map Visualizati-on Components, die sich auf die Arbeit mit Geodaten in der Oracle DB fokussieren. Klares Ziel dabei ist, die Entwicklung von Datenbank-zentrischen Anwendungen zu unterstützen, die es ermöglichen, Daten unterschiedlichster Arten und Formate nahtlos zu verarbeiten und zu analysieren.

Quellen

- [1] Oracle Spatial Developer´s Guide 19c, SDO_GCDR Package (Geocoding) Data: https://docs.oracle.com/en/database/oracle/oracle-database/19/spati/SDO_GCDR-reference.html
- [2] Oracle Spatial Developer´s Guide 19c, Geocoding Address Data, Kapitel 12.7: <https://docs.oracle.com/en/database/oracle/oracle-database/19/spati/geocoding-address-data-concepts.html>

- [3] Oracle Maps Cloud (ehemals Oracle eLocation Service):
<http://maps.oracle.com>
- [4] Oracle Technology Network, Get Started with Oracle Spatial Studio:
<https://www.oracle.com/database/technologies/spatial-studio/get-started.html>
- [5] Oracle Cloud Free Tier:
<https://www.oracle.com/cloud/free>
- [6] Deutsche Bahn AG, Open-Data-Portal, Datensätze:
<https://data.deutschebahn.com>
- [7] Oracle Database 19c Spatial Developer's Guide, Indexing and Querying Spatial Data:
<https://docs.oracle.com/en/database/oracle/oracle-database/19/spatl/indexing-querying-spatial-data.html>
- [8] SlideShare, Upload-Bereich von Karin Patenge:
<https://www.slideshare.net/kpatenge>
- [9] Blog „Geodaten für Alle: Oracle Spatial“:
<https://oracle-spatial.blogspot.com>

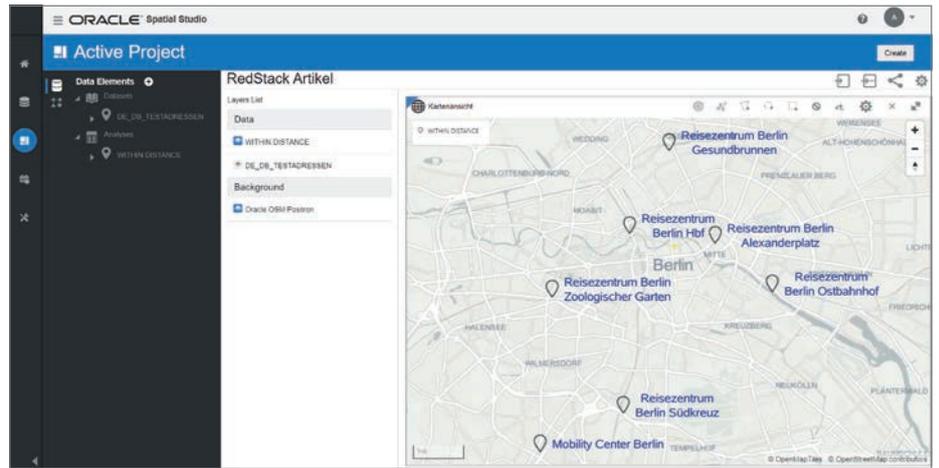


Abbildung 14: Oracle Spatial Studio – Ergebnis einer räumlichen Analyse (Quelle: Karin Patenge)

Über die Autorin

Karin Patenge arbeitet seit 2007 als Solution Engineer für Core & Cloud Technologies bei Oracle in Deutschland. Sie berät Kunden und Partner insbesondere im Umfeld Datenmanagement-Plattformen und Datenbank-zentrische

Anwendungsentwicklung. Karin ist regelmäßige Sprecherin auf verschiedenen Veranstaltungen der DOAG und auch anderer Konferenzen. Einen Ausschnitt an Foliensätzen von solchen Veranstaltungen finden Sie auf SlideShare [8]. Karin schreibt darüber hinaus Artikel für Fachzeitschriften und ist verantwortliche Autorin des Blogs „Geodaten für Alle: Oracle Spatial“ [9].



Karin Patenge
karin.patenge@oracle.com



Oracle VM erhält noch drei Jahre Extended Support

Zunächst hatte Oracle das Produkt- und Support-Ende für März 2021 angekündigt.

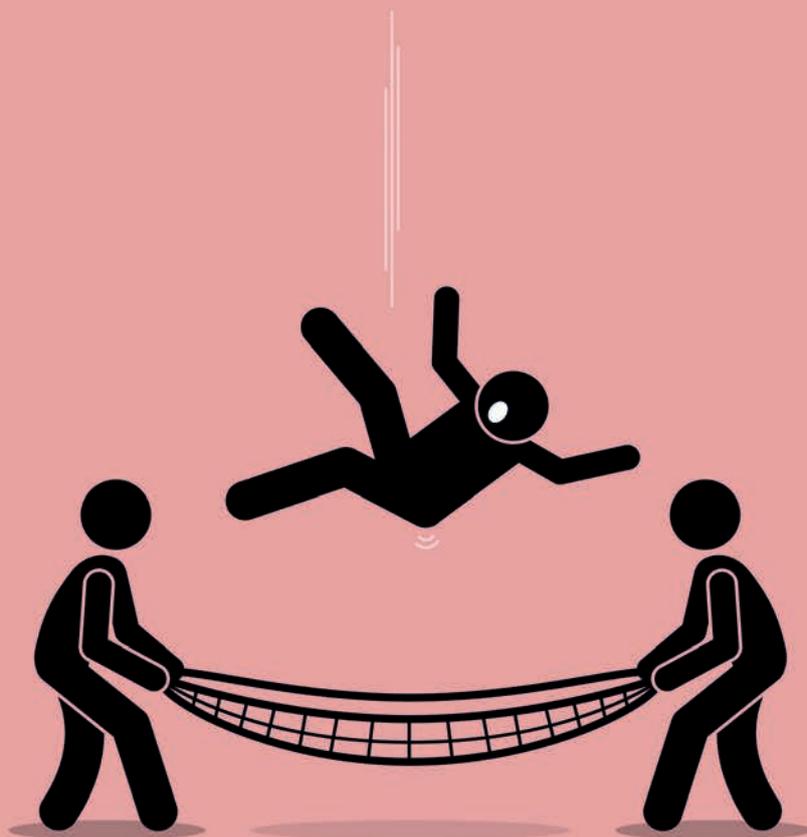
Gute Nachrichten für Anwender von Oracle VM: Michael Paege, DOAG-Vorstand und Leiter des Competence Center Lizenzierung, berichtet auf seinem Blog, dass Oracle entgegen bisheriger Verlautbarungen den Extended Support für Oracle VM nun doch für drei weitere Jahre gegen Mehrkosten (Zuschlag von 10 %

im ersten Jahr und jeweils 20 % im zweiten und dritten Jahr) anbietet. Zudem soll der ursprünglich im Juni 2020 endende Extended Support für Oracle Linux 5 bis November 2020 verlängert werden.

Wie wir von Oracle erfahren haben, wird es hierzu Ende der kommenden Woche eine offizielle Ankündigung mit

entsprechenden Support Policy Documents geben. Sobald diese veröffentlicht sind, werden wir sie an dieser Stelle verlinken.

Update (6.5.2020): Oracle hat nun mit etwas Verspätung das entsprechende Dokument veröffentlicht.



Hybrid Data Guard mit der OCI

Thomas Rein, dbi services

Immer wieder trifft man auf produktive Kundendatenbanken, die praktisch nicht gegen einen Totalausfall abgesichert sind. Häufige Argumente gegen DR-Lösungen sind höhere Kosten (Lizenzen und Betrieb), zu große Komplexität oder einfach „brauchen wir nicht, wir haben ja einen Backup“. Klar, nicht jede produktive Datenbank muss innerhalb von Minuten wieder online sein und manche Anwendungen können auch Datenverluste verkraften. Dennoch sollten sich alle Stakeholder über Kosten und andere Folgen eines Totalausfalls im Klaren sein. Hier können die modernen Cloud Provider – insbesondere Oracle mit der OCI – weiterhelfen, ist es doch mit deren Werkzeugen relativ einfach, ein „virtuelles Datacenter“ für Standby-Datenbanken bereitzustellen.

Mit der Oracle Cloud Infrastructure (OCI) ist es auf einfache Art möglich, eine DR-Lösung in der Cloud aufzubauen. Es werden nicht nur die Releases EE 11g bis 19c unterstützt, es ist ebenso möglich, über Drittprodukte wie DBvisit

auch für eine Standard Edition eine DR-Lösung bereitzustellen. Eine Herausforderung stellt allerdings die Encryption in der Cloud dar, da viele Enterprise-Edition-Anwender keine Advanced-Security-Option lizenziert haben und sie

für die Standard Edition On-Premises nicht verfügbar ist. Hier gibt es die Möglichkeit, im Cloud Marketplace auf vorgefertigte Oracle-VM-Images zurückzugreifen und die Standby-Systeme ohne TDE aufzusetzen.

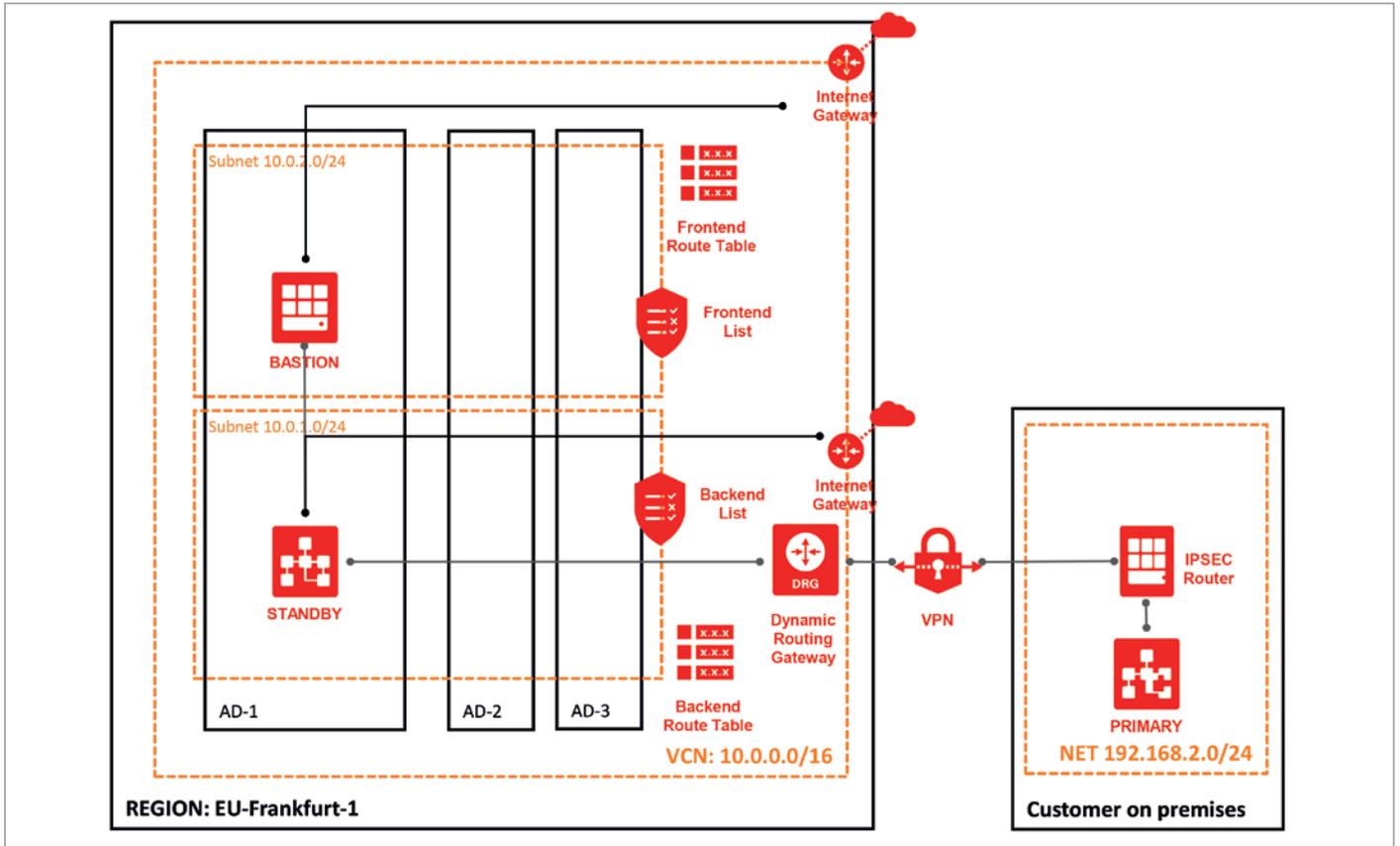


Abbildung 1: Architektur (Quelle: Thomas Rein)

Architektur

In einem ersten Schritt wird die Architektur der neuen Infrastruktur festgelegt. Minimal wird ein VCN (Virtual Cloud Network, z.B. 10.0.0.0/16) mit einem Subnetz (10.0.1.0/24 regional oder AD-bezogen) benötigt, in welches das Standby-System eingebunden wird. Für die Anbindung zum On-Premises-Netzwerk (z.B. 192.168.2.0/24) müssen ein Dynamic Routing Gateway, ein CPE-Objekt (Customer Premises Equipment) und ein entsprechender IPSec-Tunnel definiert werden. Das CPE-Objekt ist nichts anderes als eine logische Repräsentation des Rou-

ters, der im lokalen Netzwerk zum Einsatz kommt (siehe Abbildung 1). Eine Liste aller zertifizierten Router kann unter <https://docs.cloud.oracle.com/de-de/iaas/Content/Network/Reference/CPEList.htm> abgerufen werden. Sollte ihr Gerät nicht auf der Liste erscheinen, gibt es immer noch die Möglichkeit, On-Premises einen IPSec-Server als Bare Metal oder als VM zu erstellen und den Netzwerkverkehr zum VCN über dieses System zu routen.

Damit Systeme aus dem Backend-Netzwerk bei Bedarf (z.B. Updates) auch das Internet erreichen können, benötigt es ein NAT-Gateway und eine entsprechende Routingtabelle (siehe Tabelle 1).

Abschließend wird noch die Security-Liste definiert. Hierbei wird sämtlicher Datenverkehr nach außen zugelassen. Bei den eingehenden Regeln werden aus dem On-Premises-Netzwerk alle Ports geöffnet, im Minimum jedoch die Ports 22 und 1521 für SSH und SQL*Net sowie ICMP Typ 3 Code 4, um MTU-Fragmentierungen zuzulassen.

Soll das Datenbanksystem auch über das Internet erreichbar sein, ist ein weiteres Subnetz (z.B. 10.0.2.0/24) einzurichten, das über ein Internet Gateway (IGW) erreichbar ist. Die Routing-Tabelle für dieses Frontend-Netzwerk wird analog zum Backend-Netzwerk definiert. Nur

Destination	Target Type	Target
0.0.0.0/0	NAT-Gateway	NatGW
192.168.2.0/24	Dynamic Routing Gateway	DRG

Tabelle 1: Backend Routing Table

Destination	Target Type	Target
0.0.0.0/0	Internet Gateway	IGW
192.168.2.0/24	Dynamic Routing Gateway	DRG

Tabelle 2: Frontend-Routing-Tabelle

Source	Protocol	Source Port	Target Port
192.168.2.0/24	All	All	All
0.0.0.0/0	TCP	All	22

Tabelle 3: Frontend-Security-Liste (Ingress Rules)

Source	Protocol	Source Port	Target Port
192.168.2.0/24	All	All	All
10.0.2.0/24	TCP	All	22

Tabelle 4: Backend-Security-Liste

das NAT-Gateway wird durch das Internet Gateway ersetzt (siehe Tabelle 2).

In der Security-Liste wird sämtlicher ausgehender Datenverkehr zugelassen (siehe Tabelle 3). Eingehende Verbindungen aus dem Internet sind nur auf Port 22 (SSH) erlaubt. Aus dem On-Premises-Netzwerk können alle Ports freigeschaltet werden.

Die SSH-Verbindung vom Bastion Host zum Backend-Subnetz muss dann über eine entsprechende Ingress-Regel in der Backend-Security-Liste zugelassen werden (siehe Tabelle 4).

Damit ist die Netzwerkinfrastruktur vorbereitet. Wichtig bei der Wahl der Netzwerkadressen ist, dass private Adressbereiche (10.0.0.0/8, 172.16/12 und 192.168/24) für das VCN verwendet werden und keine Überschneidung mit dem lokalen Netzwerk vorliegt.

Konfiguration des VPN-Tunnels

Erfahrungsgemäß gibt es die häufigsten Probleme bei der Konfiguration der IPSec-Verbindung zwischen beiden Netzwerken. Für die gängigsten Router bietet Oracle zertifizierte Konfigurationen an (<https://docs.cloud.oracle.com/de-de/iaas/Content/Network/Reference/libreswanCPE.htm>).

Besitzen Sie keine unterstützte Hardware, kann der Tunnel über das Open-Source-Produkt Libreswan (<https://libreswan.org>) implementiert werden. Alternativ kann der Tunnel auch komplett manuell aufgesetzt werden (siehe Doc ID 2495541.1). In diesem Fall benötigen Sie einen IPSec-

Server sowohl im lokalen Netz als auch in der Oracle Cloud. Die Tunnelkonfiguration kontrollieren Sie in diesem Fall selbst. Meist werden die Probleme durch eine falsch gewählte MTU (Maximum Transfer Unit) verursacht (Fragmentierung von Paketen) oder der Router/IPSec-Server kann nicht mit den redundanten Tunneln auf OCI-Seite umgehen (asymmetrisches Routing). Auf das MTU-Problem sei hier kurz eingegangen.

Die OCI, wie mittlerweile auch viele On-Premises-Netzwerke, verwendet eine maximale Transfer Unit von 9000 Bytes. Im Internet werden je nach Provider 1500 oder auch etwas weniger verwendet (siehe Abbildung 2). Durch die Encapsulation im Tunnel stehen dort tatsächlich weniger Bytes zur Verfügung. Durch diese Einschränkung kann es zu fragmentierten Paketen kommen, wobei nur das erste Fragment den Ziel- und Sender-Port enthält und damit nachfolgende Fragmente von den Security-Listen verworfen werden. Die Verbindung hängt, obwohl andere Services noch durch den Tunnel geroutet werden. In solchen Fällen sollten die Security-Listen (ICMP Type 3 Code 4) beziehungsweise die MTU-Größen untersucht und angepasst werden. Eventuell werden auch spezielle Routingeinträge mit speziellen MTU-Größen benötigt.

Eine weitere Problemquelle ist das asymmetrische Routing. Wenn ein VPN in der OCI angelegt wird, erhält man zwei IPSec-Tunnel, die in der Cloud an verschiedenen Routern anliegen. Ein Datenpaket kann auf dem einen Weg in das VCN geschickt werden und die Antwort

auf dem anderen Weg zurückkommen. Nicht jeder Router / jede Firewall kann damit umgehen. Hier kann man die Routen mit entsprechenden Metriken versehen, sodass ein Tunnel bevorzugt wird. Der zweite Tunnel wird dann als Standby verwendet. In jedem Fall sollte man sich für die Konfiguration der VPN-Anbindung Zeit nehmen und deren Stabilität ausführlich testen.

Vorbereitung des Standby-Systems

Die Auswahl und Konfiguration des Standby-Systems werden bestimmt durch die Datenbank-Version, die Datenbank-Edition, Datenbank-Optionen und natürlich durch die vorhandenen Lizenzen. Der nachfolgende Überblick soll die Auswahl des Systemtyps erleichtern. Bei der Auswahl des Standby-Systems gibt es mehrere Möglichkeiten:

- Verwendung eines DBaaS-Systems (VM/BM): Schnelles, einfaches Setup, TDE erforderlich
- Verwendung eines von Oracle bereitgestellten VM Image (Marketplace): OS bereits konfiguriert, GI und RDBMS-Software können bei der Erstellung ausgewählt werden. Nach der Bereitstellung wird das System manuell konfiguriert
- Manuelles Setup einer VM: Maximale Freiheit bezüglich OS, GI und RDBMS-Versionen. Gleicher Aufwand wie in einer On-Premises-Installation

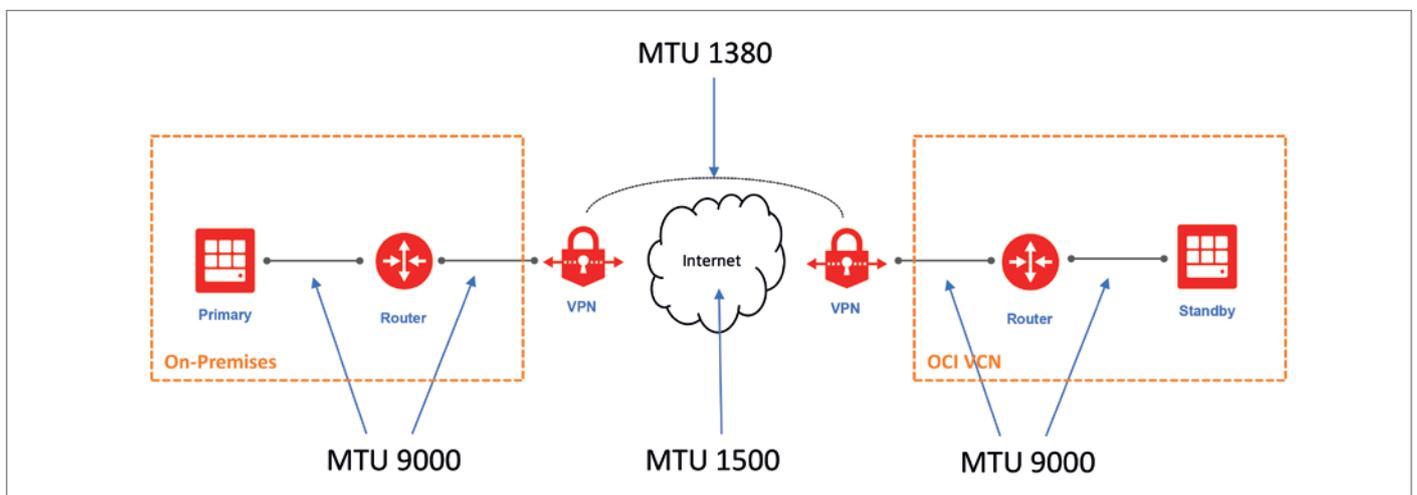


Abbildung 2: MTU-Größen (Quelle: Thomas Rein)

Database Releases	New Unencrypted Tablespace Creation Allowed?	Open Unencrypted Tablespace?	Migrated Unencrypted Tablespaces can be Opened?	Hybrid DR with Unencrypted Tablespaces can be Opened?
11.2	Yes	Yes	Yes	Yes
12.1	Yes	Yes	Yes	Yes
12.2 (Pre-April'17 PSU)	Yes	No	No	No
12.2 (Post-April '17 PSU) and higher	No	Yes	Yes	Yes

Abbildung 3: Unencrypted Tablespace Support in OCI (Oracle Doc ID 2359020.1) (Quelle: Thomas Rein)

DBaaS-Systeme

Die Vorteile eines DBaaS-Systems liegen auf der Hand: Mit wenig Aufwand kann ein fertiges Datenbanksystem konfiguriert werden. Die Versionen reichen von 11.2.0.4.190416 bis zu 19.6. Die verfügbaren Editionen sind SE, EE (ohne Options), EE High Performance (notwendig für Multitenant) und EE Extreme Performance (alle Optionen, insbesondere RAC und Active Data Guard). Für 1-Node-Systeme kann zwischen ASM oder Blockvolume Storage gewählt werden, RAC-Systeme gibt es nur unter ASM. Allen Versionen und Editions ist gemeinsam, dass sie die TDE (Transparent Data Encryption) verwenden, allerdings in unterschiedlichen Ausprägungen:

- Für 11.2- und 12.1-Datenbanken wird die Encryption über den Parameter `ENCRYPT_NEW_TABLESPACES = CLOUD_ONLY|DDL|ALWAYS` gesteuert. Bei obigen Versionen darf dieser auf DDL gesetzt werden, damit können Non-TDE Tablespaces angelegt werden.
- Ab 12.2.0.1. (genauer, ab PSU April 17) können nur noch Encrypted Tablespaces in der Cloud angelegt werden

In *Abbildung 3* ist das Verhalten der Versionen bezüglich Encrypted und Unencrypted Tablespaces ersichtlich.

Hieraus ergibt sich, dass minimal eine Enterprise Edition mit Advanced Security Option (ASO) im On-Premises-Umfeld zu lizenzieren ist. Ansonsten ist nicht sichergestellt, dass nach einem Switchover oder Failover die Datenbank noch ordnungsgemäß geöffnet werden kann. Für „normale“ EE- oder SE-Anwender ist die DBaaS-Variante also keine Option.

Oracle VM Images

Im Cloud Marketplace stellt Oracle ein Datenbank VM Image bereit, das zur Erstellung von Compute-Instanzen verwendet werden kann. Ein Oracle Linux 7 ist bereits vorinstalliert, die RDBMS-Version kann bei der Erstellung ausgewählt werden (12.2, 18.8 und 19.5). Die GI 19c wird ebenfalls mit bereitgestellt, sodass sie bei Bedarf installiert werden kann. Bei der RDBMS-Software sind sowohl EE als auch SE möglich. Eine neue Datenbank-VM kann damit schnell bereitgestellt werden, Release Updates und One-Off Patches können über das bekannte Prozedere nachinstalliert werden. Die Bereitstellung von Storage für die Datenbank erfolgt über das Attachment von Block Volumes. Diese sind entweder mit einem von Oracle verwalteten Schlüssel oder durch einen eigenen Schlüssel encrypted. Die Volumes selbst können entweder in eine ASM Diskgroup oder nativ über das Filesystem als Datenbank-Storage bereitgestellt werden. Soll eine Compute-Instanz skaliert werden, so geschieht dies mit Downtime über die Zuweisung eines neuen VM Shape. Eine Größenänderung von Volumes ist im laufenden Betrieb nicht möglich, da die Volumes für diesen Vorgang nicht an eine VM angebunden sein dürfen. Deshalb ist es hier empfehlenswert, entweder mit dem Linux LVM oder gleich mit dem ASM zu arbeiten. Eine Änderung der Volume Performance ist allerdings im laufenden Betrieb durchaus möglich. Wer sich genauer über den Elastic Volume Service der OCI informieren möchte, findet unter <https://docs.cloud.oracle.com/en-us/iaas/Content/Block/Concepts/blockvolumeperformance.htm> weiterführende Informationen und Performanceauswertungen.

Die Oracle-VM-Image-Variante ist besonders interessant für diejenigen, die

eine Enterprise Edition oder eine Standard Edition in 12.2, 18c oder 19c verwenden. Die Encryption erfolgt auf Volume-Ebene, eine TDE wird also nicht benötigt. Sofern bereits eigene Encryption Keys vorhanden sind, können diese verwendet werden.

Manuelles Setup einer VM

Wer andere RDBMS-Versionen benötigt, seine Datenbank auf Windows betreibt oder mit anderen Linux-Versionen arbeiten muss, der kann immer noch sein Standby-System als normale Compute-Instanz erstellen. Gewisse Einschränkungen mag es bei älteren Linux-Versionen und/oder bei Windows hinsichtlich des Attachments von Block Storage geben (nur iSCSI Attachment, keine Paravirtualisierung), ansonsten gilt jedoch das Gleiche wie bei den Oracle VM Marketplace Images. Die „Selbstbau“-Variante ist vor allem für Windows-Anwender interessant oder wenn spezielle Anforderungen an das Filesystemlayout (OFA, Datafiles usw.) vorliegen. Auch für Standard-Edition-Anwender, die sich eine Standby-Datenbank mithilfe von Drittprodukten wie DBvisit aufbauen wollen, kommt diese Variante infrage.

Bereitstellen der Standby-Instanz

Werden die Compute-Instance-basierten Varianten gewählt, kann man sofort mit der Erstellung der Standby-Datenbank loslegen. Der Prozess unterscheidet sich nicht vom Anlegen einer Standby Database im On-Premises-Umfeld:

- Spfile erstellen und Passwortfile kopieren, Auxiliary-Instanz starten

- Duplicate oder Restore from Service, je nach Version
- Standby Redologs anlegen, Flashback und Force Logging aktivieren
- Statische Listener definieren, Broker konfigurieren
- Gegebenenfalls Observer einrichten, Fast Start Failover aktivieren

In der DBaaS-Variante müssen vor der Erstellung der Standby-Datenbank noch ein paar Vorarbeiten erledigt werden. Beim Anlegen des Systems wird automatisch auch eine Datenbank miterstellt (DB_NAME müssen in OCI und lokal gleich sein, der Unique-Name wird von der OCI vergeben). Es werden nur die Datenfiles gelöscht (siehe Listing 1).

Die GI-Konfiguration für die Datenbank sollte ebenfalls gesichert werden. Danach können die DB gestoppt und die Datenfiles gelöscht werden. Abschließend wird das TDE Wallet von der lokalen DB auf der VM im richtigen Verzeichnis bereitgestellt (siehe Listing 2).

Danach kann die Standby-Datenbank wie gewohnt erstellt werden. Zu beachten ist hier, dass der DB_UNIQUE_NAME, der in der OCI vergeben wurde, beibehalten wird. Da ein VPN-Tunnel verwendet wird, muss der Redo-Transport nicht unbedingt verschlüsselt werden. Ansonsten sind die entsprechenden Einträge in den sqlnet.ora-Files noch anzupassen.

Testing und Tuning

Unabhängig von der gewählten Variante sollten die üblichen Tests wie Switchover oder Convert to snapshot standby durchgeführt werden. Insbesondere die Performance des Redo-Transports sollte geprüft werden (V\$MANAGED_STANDBY). Eventuell ist es hier nötig, die SDU- und RECV_BUF_SIZE/ SEND_BUF_SIZE-Größen anzupassen. Um Latenzproblemen aus dem Weg zu gehen, sollte das System im MAX PERFORMANCE (ASYNC) betrieben werden. Zusätzliche Redundanz in Form eines

weiteren Routers im lokalen Netzwerk und die Definition eines weiteren IPSec-Tunnels können ebenfalls helfen, die Ausfallsicherheit zu erhöhen.

Lizenzierung

Abschließend noch ein paar Anmerkungen zur Lizenzierung.

Der Licensing Guide von Oracle sagt klar: „**Data Recovery Environments using Copying, Synchronizing or Mirroring** Standby and Remote Mirroring are commonly used terms to describe these methods of deploying Data Recovery environments. In these Data Recovery deployments, the data, and optionally the Oracle binaries, are copied to another storage device. In these Data Recovery deployments all Oracle programs that are installed and/or running must be licensed per standard policies documented in the Oracle Licensing and Services Agreement (OLSA). This includes installing Oracle pro-

Unsere e-Services machen es einfacher.



Entdecken Sie unsere Remote-Services für

- Platform Audits
- License Reviews
- Workshops



grams on the DR server(s) to test the DR scenario. Licensing metrics and program options on Production and Data Recovery/ Secondary servers must match.“

Was heißt das nun für die oben beschriebenen Varianten? Die Enterprise Edition und die Standard Edition unterscheiden sich hier grundsätzlich. Bei der Core-basierten Lizenzierung der EE muss man für die OCI die gleiche Anzahl OCPUs wie bei On-Premises lizenzieren. Hierbei gilt: 2 vCPUs in der VM = 1 OCPU. Hat man also beispielsweise 2 EE-CPU-Lizenzen, so sind das (mit Intel Cores) 4 Cores, auf denen die EE betrieben werden kann. Dies entspricht wiederum 4 OCPUs in der OCI (= 8vCPUs in der VM). Bei einer Standard Edition One sind 4 OCPUs erlaubt. Eine Standard Edition Two wird, unabhängig von der Anzahl OCPUs, sowieso nur 16 Threads nutzen.

In der DBaaS-Variante können die Lizenzen in die monatlichen Kosten miteinbezogen werden (unabhängig ob Standard- oder Enterprise Edition), ein BYOL-Lizenzierungsmodell ist ebenfalls möglich.

Im Gegensatz dazu müssen bei Compute-Instanzen grundsätzlich die eigenen Lizenzen mitgebracht werden. Auch hier gilt: Eine SE1 darf dann mit maximal 4 OCPUs betrieben werden, für eine SE2 greift die 16-Threads-Grenze.

Kosten

Generell lassen sich die Kosten im Cloud-Umfeld nie ganz einfach berechnen. Jeder Anbieter hat hier seine Eigenheiten. Am konkreten Beispiel eines Kunden in der OCI lässt sich jedoch sagen, dass ein DBaaS-System mit 4 Cores Enterprise Edition High Performance pro Jahr etwa 32.000 Euro kostet und damit deutlich unter den Kosten liegt, den der Betrieb eines zusätzlichen Datacenters zuzüglich der EE-Lizenzen und -Options verursachen würde. Bei den Compute-Instanzen ist die Berechnung insofern komplizierter, da die Lizenzen (BYOL) in das Modell mit eingerechnet werden müssen – aber eine VM 4 Core VM kostet ohne Datenbank-Lizenzen etwa 3.000 Euro pro Jahr.

Fazit

Mithilfe der OCI lassen sich recht einfach und schnell DR-Lösungen in der Cloud

```
{
  sqlplus -S / as sysdba << EOF
    set verify off head off lines 999
    set pages 0 feedback off trimspool on

    spool /tmp/delete.sh
    select 'asmcmd rm '||name from v\datafile
    union all
    select 'asmcmd rm '||name from v\stempfile
    union all
    select 'asmcmd rm '||member from v\logfile
    union all
    select 'asmcmd rm '||name from v\controlfile;

    spool off

    create pfile='/tmp/D01.ora' from spfile;
EOF
  chmod 777 /tmp/delete.sh
}
```

Listing 1: ASM-Kommandos erstellen

```
ENCRYPTION_WALLET_LOCATION=
(
  SOURCE=(METHOD=FILE)
  (
    METHOD_DATA=(DIRECTORY=/opt/oracle/dcs/commonstore/wallets/
tde/$ORACLE_UNQNAME)
  )
)
```

Listing 2: Wallet Location

realisieren. Die Vorteile liegen nicht nur in der Absicherung eines Totalausfalls, sondern auch in der Flexibilisierung von Migrationen und Tests (Standby First Patching, Snapshot Standby). Was die Sicherheit anbetrifft, so können – unabhängig von der gewählten Variante – die eigenen Keys verwendet werden. Für den Datentransfer stehen VPN-Tunnel und SQL*Net Encryption zur Verfügung. Die Kosten sind aus Sicht des Autors geringer als beim Betrieb eines zweiten Datacenters im On-Premises-Umfeld. Eine ausreichende Lizenzierung ist natürlich Voraussetzung für solche Lösungen. Selbstverständlich ist eine DR-Lösung für die Datenbank nur die halbe Miete – auch Anwendungen sollten abgesichert werden. Es ist aber sicher ein erster Schritt in die richtige Richtung.

- [2] <https://docs.cloud.oracle.com/en-us/iaas/Content/Block/Concepts/blockvolumeperformance.htm>
- [3] <http://www.oracle.com/us/corporate/pricing/cloud-licensing-070579.pdf>
- [4] <https://www.oracle.com/assets/data-recovery-licensing-070587.pdf>
- [5] <https://www.oracle.com/technetwork/database/availability/hybrid-dg-to-oci-5444327.pdf>



Thomas Rein
thomas.rein@dbi-services.com

Quellen

- [1] Oracle Database Tablespace Encryption Behavior in Oracle Cloud (Doc ID 2359020.1)



Benchmarking Oracle Cloud Services

Manfred Drozd, Peakmarks

Mittlerweile stehen unzählige Cloud-Angebote zur Verfügung. Die technischen Kapazitäten und Preise sind gut dokumentiert. Aber über die Leistungsfähigkeit dieser Plattformen ist nichts bekannt. Es fehlen verständliche und allgemeingültige Performance-Kennzahlen, wie wir sie in anderen Industrien, beispielsweise der Automobilindustrie, schon lange kennen. Die Peakmarks AG hat eine Benchmark-Software entwickelt, um schnell und zuverlässig verständliche und repräsentative Performance-Kennzahlen von Oracle-Datenbank-Plattformen zu ermitteln. Diese Vorgehensweise wurde auf der letztjährigen TPC Technology Conference einer breiteren Benchmark Community vorgestellt [1]. Peakmarks verfolgt dabei einen ganzheitlichen Ansatz: Mit einem Benchmark Tool werden innerhalb weniger Tage sowohl die wichtigsten Komponenten einer Plattform (Server, Storage) als auch kritische Datenbank-Hintergrundprozesse (Log Writer, Database Writer) und typische Datenbank-Operationen (Data Load, Data Analytics, Transaction Processing) vermessen. Alle Workloads werden mit Oracle SQL generiert, die Performance-Kennzahlen stammen weitgehend aus dem Oracle Workload Repository. Dieses Benchmark-Verfahren ist nicht für den Vergleich von Datenbank-Systemen geeignet, sondern lediglich für den Vergleich verschiedener Infrastrukturen, auf denen ein Oracle-Datenbanksystem läuft.

Verständliche Performance-Kennzahlen

Performance-Kennzahlen von Database Services sind für verschiedene Aufgaben notwendig.

Evaluation. Performance-Kennzahlen können für Preis-Leistungs-Überlegungen verschiedener Database Services, Technologien oder Konfigurationen verwendet werden.

Kapazitätsplanung. Bei der Migration

von Datenbank Services auf neue Plattformen oder Cloud Services helfen Performance-Kennzahlen bei einer seriösen Kapazitätsplanung.

Qualitätssicherung. Cloud Services können

nen ihre Infrastruktur jederzeit und ohne Vorankündigung ändern. Durch regelmäßige Performance-Überprüfungen wird die Leistungsqualität sichergestellt. Leistungsaussagen der Anbieter können schnell überprüft werden.

Lizenzkostenoptimierung. Durch Wahl geeigneter Infrastruktur können Oracle Lizenz- und Wartungskosten erheblich reduziert werden. Oracle Lizenz- und Wartungskosten übersteigen häufig bei Weitem die Infrastrukturkosten.

Nachfolgend möchten wir die wichtigsten Workloads und Performance-Kennzahlen vorstellen, die für Evaluation, Kapazitätsplanung, Qualitätssicherung und Lizenzkostenoptimierung von Oracle Cloud Services notwendig sind.

Server Performance

Server Workloads ermitteln die Leistung eines Servers mit seinen Prozessoren, seinem Hauptspeicher und seinen internen Speicherkanälen im Datenbankbetrieb. Diese Workloads sind besonders dann wichtig, wenn es um die Optimierung von Lizenzkosten geht, denn die Lizenzkosten hängen oft von der Anzahl der eingesetzten Prozessoren ab. In diesem Fall wird ein Server mit der höchsten Leistung pro Prozessor gesucht.

Server Workloads zeigen auch die Effizienz von Multithreading- und Virtualisierungs-Technologien und geben Hinweise auf Skalierbarkeit bei hoher Anzahl von Cores und Sockets (NUMA-Effekte). Wird eine Datenbankverschlüsselung verwendet, kann auch deren Auswirkung auf

die gesamte Datenbankanleistung ermittelt werden.

Alle Server Workloads greifen über SQL mit unterschiedlichen Zugriffsmustern auf Tabellen zu. Die betroffenen Tabellen werden im Buffer-Cache fixiert. Es gibt keine I/O-Operationen, sodass diese Workloads vollständig durch die CPU-Leistung limitiert sind.

Die primären Performance-Kennzahlen sind Abfragen pro Sekunde (qps), die Antwortzeit und die Scanrate der Abfragen (Speicherbandbreite). Eine sekundäre Leistungskennzahl ist die Anzahl der logischen Blockzugriffe pro Sekunde.

Table 1 zeigt einen Workload, bei dem über einen Primärschlüssel genau auf einen Datensatz zugegriffen wird. Typische Beispiele für diesen Workload sind Abfragen wie: Selektiere Kunde, Auftrag, Produkt etc.

Die Spalte „Jobs“ zeigt die Anzahl der Prozesse an, die die Last erzeugen. Der Test wurde auf einem Server mit 2x Intel Xeon 8260 durchgeführt. Auffallend ist, dass die Leistung pro CPU (gemäß Oracle-Konfigurationsparameter CPU_COUNT) mit zunehmender Last kontinuierlich sinkt und die Antwortzeit ansteigt. Ein nicht untypisches Verhalten für Intel-Prozessoren mit höherer Anzahl von Cores pro Socket.

Storage Performance

Herkömmliche I/O-Benchmark Tools wie fio, vdbench, iometer, Orion zeigen oft Leistungswerte an, die im realen Datenbankbetrieb nicht erreicht werden. Der Grund dafür liegt in der Komplexität von Datenbank-I/O-Operationen.

Wird ein Datenblock gelesen, muss die Buffer-Cache-Verwaltung der Datenbank viele Aufgaben erfüllen:

- a) einen freien Platz für den Block finden
- b) wenn kein freier Platz vorhanden ist, ältere Blöcke ersetzen
- c) alle Datenbankprozesse synchronisieren, die gleichzeitig versuchen, freie Plätze im Buffer-Cache einzunehmen
- d) bei Verwendung einer Shared-Disk Cluster-Architektur (Oracle Real Application Cluster) muss die Synchronisation clusterweit erfolgen, was im besten Fall den Austausch von Nachrichten zwischen Cluster-Knoten erfordert
- e) und schließlich werden Blöcke während der I/O-Übertragung auf ihre Integrität und Konsistenz überprüft.

Storage Workloads, die die I/O-Last mithilfe von SQL Statements erzeugen, liefern deutlich repräsentativer Performance-Kennzahlen für ein Speichersystem im Oracle-Betrieb.

Storage Workloads zeigen die Effizienz des gesamten I/O-Stacks (I/O-Scheduler, Queues, Multipathing, Virtualisierung), der verwendeten Technologien (HDD, SSD, Flash, SAS oder PCI, NVMe etc.) und von speicherspezifischen Funktionalitäten (Deduplizierung, Kompression, Verschlüsselung, Snapshots, Redundanz, Spiegelung, SQL-Offloading).

Die primären Performance-Kennzahlen sind Durchsatz und Servicezeit bei sequential and random I/O, gemessen in Oracle.

Nachfolgender Test zeigt die „random I/O“ Performance (8 KByte, 100% read) auf einer Exadata X8M (siehe Tabelle 2).

Test Workload	Nodes	Jobs	CPU busy [%]	CPU user [%]	CPU sys [%]	CPU idle [%]	Queries total [qps]	Queries per cpu [qps]	Response time [ms]	Log reads total [dbps]	Log reads per cpu [dbps]	BuCache read [%]
1 SRV-LIGHT	1	1	2	1	0	98	111,772	111,772	0.009	335,474	335,474	100.0
2 SRV-LIGHT	1	8	9	9	0	91	772,148	96,519	0.010	2,316,537	289,567	100.0
3 SRV-LIGHT	1	16	17	17	0	83	1,453,571	90,848	0.011	4,359,858	272,491	100.0
4 SRV-LIGHT	1	24	25	24	0	75	1,809,592	75,400	0.013	5,428,961	226,207	100.0
5 SRV-LIGHT	1	32	33	33	0	67	2,100,826	65,651	0.015	6,303,010	196,969	100.0
6 SRV-LIGHT	1	40	41	40	0	59	2,375,747	59,394	0.016	7,127,950	178,199	100.0
7 SRV-LIGHT	1	48	49	49	0	51	2,591,160	53,983	0.018	7,774,885	161,977	100.0
8 SRV-LIGHT	1	56	58	57	0	42	2,743,805	48,997	0.020	8,232,539	147,010	100.0
9 SRV-LIGHT	1	64	66	65	0	34	2,842,314	44,411	0.022	8,526,861	133,232	100.0
10 SRV-LIGHT	1	72	73	72	0	27	2,909,352	40,408	0.024	8,730,033	121,250	100.0
11 SRV-LIGHT	1	80	80	80	1	20	2,977,330	37,217	0.026	8,933,416	111,668	100.0
12 SRV-LIGHT	1	88	90	89	1	10	3,115,151	35,399	0.027	9,346,891	106,215	100.0
13 SRV-LIGHT	1	96	96	95	1	4	3,176,528	33,089	0.029	9,531,313	99,285	100.0

Table 1: Workload, bei dem über einen Primärschlüssel genau auf einen Datensatz zugegriffen wird

Test	Workload	Nodes	Jobs	CPU busy [%]	CPU user [%]	CPU sys [%]	CPU idle [%]	CPU iow [%]	Phys reads total [dbps]	Phys reads total [IOPS]	Service time [us]	Phys reads total [MBps]	BuCache read [%]	FlCache read [%]
35	STO-RANDOM	4	4	3	2	1	97	0	117,231	114,811	189	975	48.2	56.9
36	STO-RANDOM	4	16	8	6	2	92	0	440,083	436,855	121	3,644	17.1	40.5
37	STO-RANDOM	4	32	14	10	3	86	0	800,953	800,353	77	6,626	9.2	27.5
38	STO-RANDOM	4	48	19	14	4	81	0	1,127,177	1,125,477	65	9,313	3.7	22.4
39	STO-RANDOM	4	64	24	18	4	76	0	1,754,061	1,752,672	60	14,473	3.5	20.6
40	STO-RANDOM	4	72	26	21	4	74	0	2,166,357	2,166,019	61	17,867	3.5	20.3
41	STO-RANDOM	4	88	30	25	3	70	0	3,038,247	3,038,309	67	25,035	3.5	20.3
42	STO-RANDOM	4	96	30	26	3	70	0	3,193,351	3,193,081	72	26,312	3.5	20.0
43	STO-RANDOM	4	112	32	27	3	68	0	3,555,520	3,552,780	74	29,282	3.1	20.0
44	STO-RANDOM	4	128	33	28	4	67	0	3,668,738	3,665,469	79	30,210	3.0	19.9
45	STO-RANDOM	4	144	34	29	4	66	0	3,678,228	3,673,467	86	30,300	2.9	19.8
46	STO-RANDOM	4	160	38	32	4	62	0	3,897,980	3,891,664	94	32,095	2.7	19.9
47	STO-RANDOM	4	176	40	34	4	60	0	3,922,048	3,916,677	106	32,321	3.1	19.9
48	STO-RANDOM	4	192	44	38	5	56	0	4,206,920	4,201,380	120	34,657	3.0	20.0

Tabelle 2: „random I/O“-Performance (8 KByte, 100% read) auf einer Exadata X8M

Die Servicezeiten sind außergewöhnlich gut! Da nicht alle Daten im „Persistent Memory Cache“ der Storage Server gehalten werden können, gibt es allerdings schwankende Servicezeiten. Typisch für multi-tiered Storage-Systeme.

Log-Writer-Service-Prozesse

Die Log-Writer-Prozesse sind in erster Linie für die Transaktionsprotokollierung und die Wiederherstellung der Datenbank nach Systemausfällen verantwortlich. Diese Prozesse sind entscheidend für die Gesamtleistung von Oracle bei der Verarbeitung von Transaktionen. Die Latenzzeit der Transaktionsprotokollierung kann einen erheblichen Einfluss auf die Antwortzeit von Benutzertransaktionen haben.

Optional wird die Log-Protokollierung auch für die Datenbankreplikation verwendet, um Standby-Datenbanken zu synchronisieren. Diese Technologie ist sehr beliebt bei Disaster-Recovery-Lösungen. Die Replikation kann im synchronen oder

asynchronen Modus erfolgen. Der Datentransfer zwischen Primär- und Standby-Datenbank kann optional verschlüsselt und/oder komprimiert werden. Bei der synchronen Replikation müssen lokale Transaktionen warten, bis auch die Standby-Datenbanken das Transaktionsprotokoll erhalten haben. Dies kann die lokale Transaktionsverarbeitung erheblich verzögern.

Performance-Kennzahlen für Log Writer sind die Latenzzeit beim COMMIT von Transaktionen und die Datenmenge, die innerhalb eines bestimmten Zeitraums in REDO Logfiles geschrieben werden kann.

Das in *Tabelle 3* dargestellte Beispiel zeigt einen extremen Stresstest für die Log-Writer-Prozesse. Pro Transaktion wird ein Datensatz eingefügt und mit COMMIT abgeschlossen.

Database-Writer-Service-Prozesse

Die Database Writer sind für die Verwaltung des Buffer-Cache verantwortlich.

Diese Prozesse sind entscheidend für die Gesamtleistung von Oracle, wenn viele Blöcke geändert werden, wie beispielsweise beim konventionellen Laden von Daten über den Buffer-Cache oder bei intensiver Transaktionsverarbeitung.

Die wichtigste Leistungskennzahl für Database Writers ist die Anzahl der Datenbankblöcke, die innerhalb eines bestimmten Zeitraums in die Datenbank-Dateien zurückgeschrieben werden können.

Data Load Performance

Systemarchitekten und Kapazitätsplaner benötigen Performance-Kennzahlen von Database Services bezüglich ihrer Fähigkeit, Daten zu laden. Dies ist besonders wichtig für Data-Warehouse- und Data-Analytics-Systeme, bei denen die Datenmengen immer größer und die Zeitfenster für das Laden immer kleiner werden.

Die wichtigste Performance-Kennzahl für Data Load ist die Datenmenge, die pro Zeiteinheit geladen werden kann.

Test	Workload	TX size [rpt]	Nodes	Jobs	CPU busy [%]	CPU user [%]	CPU sys [%]	CPU idle [%]	CPU iow [%]	Transactions total [tps]	Response time [ms]	REDO blocks [rbps]	REDO data [MByte]	REDO sync [ms]	FlCache write [%]
113	LGWR-LAT	1	2	2	5	3	2	95	0	6,874	0.289	36,329	15	0.191	95.8
114	LGWR-LAT	1	2	8	9	6	3	91	0	21,296	0.372	110,391	47	0.267	97.1
115	LGWR-LAT	1	2	16	14	10	3	86	0	37,484	0.424	194,349	83	0.308	96.6
116	LGWR-LAT	1	2	24	17	14	3	83	0	47,946	0.491	250,037	107	0.362	94.8
117	LGWR-LAT	1	2	32	21	17	4	79	0	57,674	0.550	301,073	128	0.413	95.0
118	LGWR-LAT	1	2	40	25	21	4	75	0	67,338	0.589	350,228	150	0.443	96.5
119	LGWR-LAT	1	2	48	30	25	4	70	0	73,575	0.640	380,678	164	0.480	97.3
120	LGWR-LAT	1	2	56	36	30	5	64	0	78,279	0.698	401,671	175	0.526	97.9

Tabelle 3: Extremer Stresstest für die Log-Writer-Prozesse

Test	Workload	Loc	Nodes	Jobs	DOP	CPU busy [%]	CPU user [%]	CPU sys [%]	CPU idle [%]	Scanned user data [rps]	Scanned user data [MBps]	BuCache read [%]	FlCache read [%]
173	DA-LOW	EXA	2	2	4	5	3	2	95	37,886,120	12,932	0.0	100.0
174	DA-LOW	EXA	2	8	4	11	6	4	89	124,690,705	42,561	0.0	100.0
175	DA-LOW	EXA	2	16	4	18	10	6	82	187,670,124	64,058	0.0	100.0
176	DA-LOW	EXA	2	24	4	22	13	8	78	219,624,237	74,965	0.0	100.0

Test	Workload	Loc	Nodes	Jobs	DOP	CPU busy [%]	CPU user [%]	CPU sys [%]	CPU idle [%]	Scanned user data [rps]	Scanned user data [MBps]	BuCache read [%]	FlCache read [%]
185	DA-LOW	COL	2	2	1	6	5	1	94	89,469,204	28,189	0.0	0.0
186	DA-LOW	COL	2	8	1	19	17	2	81	346,009,573	109,015	0.0	0.0
187	DA-LOW	COL	2	16	1	36	32	3	64	554,274,948	174,632	0.0	0.0
188	DA-LOW	COL	2	24	1	50	44	4	50	718,084,140	226,243	0.0	0.0
189	DA-LOW	COL	2	32	1	59	53	5	41	809,941,927	255,184	0.0	0.0
190	DA-LOW	COL	2	40	1	64	56	5	36	828,407,105	261,002	0.0	0.0
191	DA-LOW	COL	2	48	1	66	59	6	34	828,391,828	260,997	0.0	0.0

Tabelle 4: Zwei Technologien für Data Analytics im Vergleich

Oracle stellt verschiedene Technologien zum Laden von Daten zur Verfügung, die erheblichen Einfluss auf die Performance haben können: konventionelles Laden über den Buffer-Cache, direktes Laden unter Umgehung des Buffer-Cache und (ab Oracle 19c) einen neuen Stream-Modus für das Laden von IOT-Daten. Die Peakmarks Benchmark Software stellt für jedes Verfahren einen entsprechenden Workload zur Verfügung.

Data Analytics Performance

Systemarchitekten und Kapazitätsplaner benötigen Performance-Kennzahlen von Database Services bezüglich ihrer Fähigkeit, große Datenmengen zu durchsuchen. Diese Anwendungen basieren typischerweise auf „full table scan“-Operationen.

Die wichtigste Performance-Kennzahl für Data Analytics ist daher die Datenmenge, die pro Zeiteinheit durchsucht werden kann.

Die Performance von „full table scans“ hängt von verschiedenen Faktoren ab: der Position der Daten in der Speicherhierarchie (Storage, Hauptspeicher) und der Technologie zur Steigerung der Scanleistung (SQL Offload, In-Memory Column Store). Manche Technologien erfordern spezialisierte Hardware und/oder zusätzliche Softwarelizenzen, was sich wiederum auf das Preis-Leistungs-Verhältnis einer Plattform auswirkt. Auch hier stehen verschiedene Workloads zur Verfügung, um die unterschiedlichen Situationen vermessen zu können.

Die Tabelle 4 zeigt zwei Technologien für Data Analytics im Vergleich. In der oberen Tabelle wird die „SQL Offload“-Technologie der Exadata verwendet, im unteren Beispiel der Column Store der In-Memory-Option.

Transaction Processing Performance

Systemarchitekten und Kapazitätsplaner benötigen Performance-Kennzahlen von Database Services bezüglich ihrer Fähigkeit, typische Transaktionen ausführen zu können. Die Peakmarks Benchmark Software stellt verschiedene Workloads mit unterschiedlicher Transaktions-Komplexität zur Verfügung.

Die wichtigsten Kennzahlen für Transaction Processing Workloads sind der Durchsatz in Transaktionen pro Sekunde und die Antwortzeit.

Die Leistung dieser Workloads hängt von verschiedenen Faktoren ab, darunter dem Verhältnis von Datenbankgröße zu Buffer-Cache-Größe. Je höher die Trefferrate im Buffer-Cache, desto höher die Transaktionsrate und desto geringer die Antwortzeit. Dies gilt insbesondere für einen niedrigen Anteil von Mutationen, bei denen der Anteil der Schreibvorgänge im Verhältnis zur Anzahl der Lesevorgänge gering ist.

Fazit

Verständliche und repräsentative Performance-Kennzahlen sind für Evaluation,

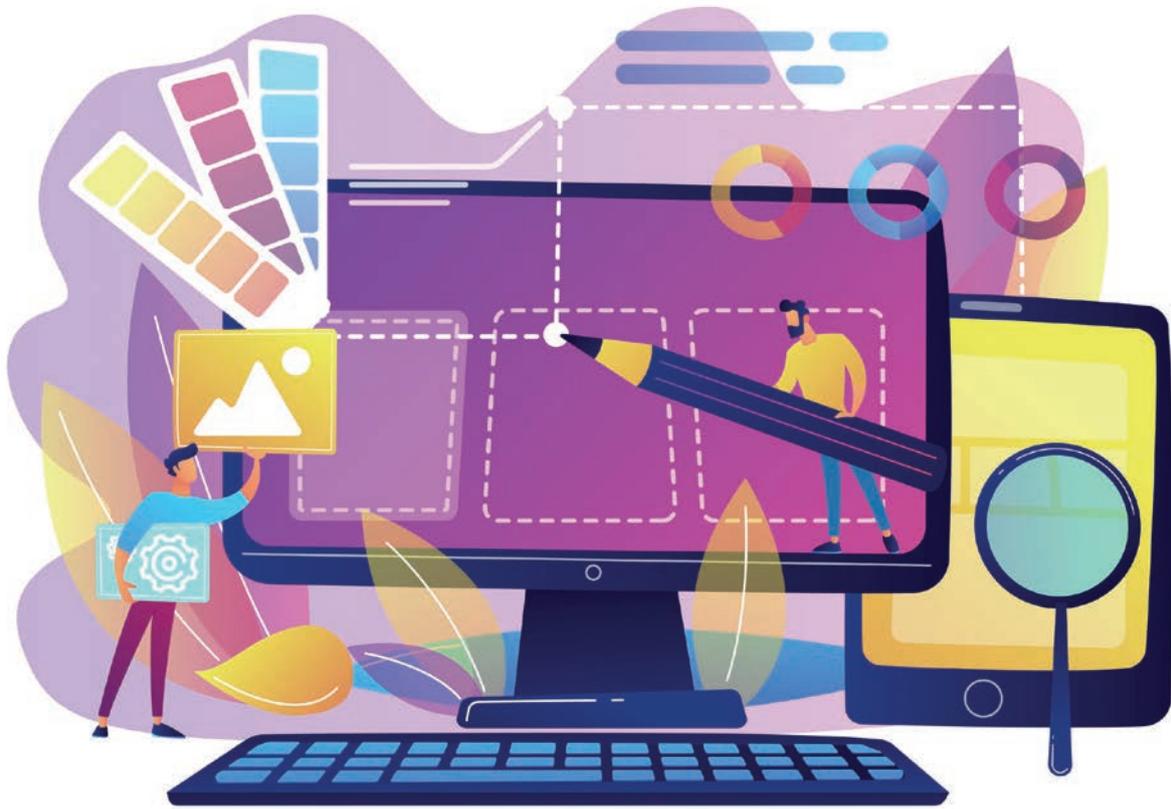
Kapazitätsplanung, Qualitätssicherung und Lizenzkostenoptimierung von Oracle Database Services – egal ob On-Premises oder in der Cloud – unerlässlich. Sie ermöglichen einen fairen und nachvollziehbaren Vergleich von Technologien, Konfigurationen, Komponenten und kompletten Systemen unterschiedlicher Hersteller und Service-Anbieter.

Quellen

[1] <http://www.tpc.org/tpctc/tpctc2019/default.asp>



Manfred Drozd
manfred.drozd@peakmarks.com



Schnittstellenbündelung mit GraphQL

Philipp Hartenfeller, MT AG

Die Schnittstellentechnologie GraphQL wird von immer mehr Firmen eingesetzt. Die Technologie bringt besondere Eigenschaften mit, die sie von den bewährten Technologien wie REST und SOAP abgrenzt. In diesem Artikel wird näher darauf eingegangen, wie diese Technologie dazu verwendet werden kann, um die bestehenden Schnittstellen gebündelt bereitzustellen.

GraphQL

GraphQL ist eine von Facebook intern für die eigene Plattform entwickelte Schnittstellentechnologie, sowie die zugehörige Abfragesprache. Seit der Open-Source-Veröffentlichung im Jahre 2015 bekommt die Technologie immer mehr Zuspruch und wird bereits von großen Unternehmen wie Twitter, PayPal und Airbnb produktiv eingesetzt. Ende 2018 hat Facebook die GraphQL-Foundation gegründet, bei der neben den bereits genannten Firmen auch Technologieriesen wie Amazon und IBM die Technologie fördern.

GraphQL unterscheidet sich zu den üblichen Schnittstellentechnologien wie SOAP und REST darin, dass keine Endpunkte mit vordefinierten Datenstrukturen verwendet, sondern die Daten ähnlich wie bei SQL ex-

plizit abgefragt werden. Daten können außerdem graph-orientiert, also über deren Beziehungen, abgefragt werden. Aus diesen beiden Eigenschaften setzt sich auch der Name „GraphQL“ zusammen (QL steht für Query Language).

Die Hauptvorteile für diesen Ansatz sind, dass Over- und Underfetching vermieden werden können. Bei Restendpunkten wird vordefiniert, welche Daten beziehungsweise Spalten zurückgegeben werden. Dies führt dazu, dass mehr als eigentlich benötigte Daten transportiert werden (Overfetching). Besonders für Anwendungen, die auch mobil genutzt werden, können daraus spürbare Ladeverzögerungen resultieren. Ein Lösungsansatz für Overfetching ist die Menge an Daten, auf mehrere kleinere Endpunkte aufzuteilen. Was daraus wiederum resultieren

kann, ist, dass mehrere Anfragen gesendet werden müssen, um alle benötigten Daten zu erhalten (Underfetching). Mehrere Anfragen zu versenden, ist gegenüber einer einzigen ebenfalls ineffizient. In Folge dessen muss entweder für jeden spezifischen Anwendungsfall ein Endpunkt eingerichtet oder diese Ineffizienz in Kauf genommen werden.

GraphQL löst diese beiden Probleme mit der Abfragesprache, sodass die benötigten Spalten explizit angefordert werden müssen. So erhält der Konsument der Schnittstelle nur die Daten, die er wirklich benötigt. Außerdem lassen sich über die Beziehungen der Entitäten flexibel zusammenhängende Daten abfragen. Falls so immer noch nicht alle Daten abgefragt werden können, kann eine GraphQL-Anfrage mehrere Abfragen enthalten.

Um diese Flexibilität zu ermöglichen, muss in der Schnittstelle ähnlich wie bei einer relationalen Datenbank, ein Schema definiert werden. Schemata bestehen aus den folgenden fünf Bausteinen:

1. Types
2. Queries
3. Mutations
4. Subscriptions
5. Resolvers

Types definieren Datenstrukturen und deren Datentypen. Außerdem wird angegeben, ob es sich bei einer Spalte um ein Pflichtfeld handelt und ob ein Datensatz oder ein Array aus mehreren zurückgegeben werden muss. Im Gegensatz zu REST überprüft GraphQL auf Datentypen und Pflichtfelder und gibt bei einem Verstoß eine Fehlermeldung zurück. Dies hat den Vorteil, dass so Folgefehler bei der Verarbeitung vermieden werden. Beziehungen können abgebildet werden, indem als Datentyp eine andere GraphQL-Datenstruktur referenziert wird. In *Abbildung 1* sind die zwei Typen, **Movie** und **Actor**, in der GraphQL-Schema-Definitionssprache dargestellt.

Queries, Mutations und **Subscriptions** bilden die verfügbaren Operationen ab. Queries sind Abfragen, die dem Konsumenten bereitgestellt werden und somit

```
type Movie {
  id: Int!
  title: String!
  description: String
  releaseYear: Int
  genres: [String]
  cast: [Actor]
}

type Actor {
  id: Int!
  firstName: String
  lastName: String
  age: Int
  photos: [Image]
  movies: [Movie]
}
```

Abbildung 1: GraphQL-Types (© Philipp Hartenfeller)

auch Einstiegspunkte in das Schema. Eine Query muss immer einen Datentypen zurückgeben, über deren Beziehungen dann weitere Datentypen abgefragt werden können. In *Abbildung 2* wird über die Query „searchMovie“ mit dem Parameter „Matrix“ für den Titel ein Film zurückgegeben. Über die definierte Beziehung zu Schauspielern können so Informationen zu beispielsweise Keanu Reeves abgefragt werden und über die Beziehung von Schauspielern zu Filmen wiederum weitere Filme, in denen der Schauspieler mitspielt. Dieses Spiel könnte bis ins Unendliche getrieben werden.

Mutations sind datenverändernde Operationen. Ähnlich zu Queries kann eine Mutation auch Daten zurückgeben. Über Subscriptions können Konsumenten Ereignisse abonnieren und bekommen Datenän-

derungen anschließend über Websockets gepusht. Dies ist nützlich für Anwendungsfälle wie Chats, wo die Konsumenten so schnell wie möglich neue Nachrichten erhalten müssen und periodische Abfragen und damit verbunden unnötige Last vermieden werden kann.

Resolver enthalten die Logik, die von GraphQL ausgeführten Operationen zu erfüllen. Diese Resolver müssen eigenständig programmiert werden und somit ist es für den GraphQL-Server völlig irrelevant, woher die Daten kommen oder wohin Daten geschrieben werden. In diesem Beispiel könnte für die Abfrage „searchMovie“ eine simple Datenbankabfrage hinterlegt werden, welche Filminformationen und Schauspieler-IDs zurückgibt. Über einen Resolver, welcher für Schauspieler-IDs deren Informationen zurückgibt, können dann die Schauspielerdaten gesammelt werden. Die Resolver selbst haben keine Vorgaben, müssen lediglich die erwarteten Daten zurückgeben. So können auch einfach verschiedene Datenquellen für das GraphQL-Schema vermischt werden.

Die offizielle Referenz für GraphQL wurde von Facebook für JavaScript entwickelt. Durch die Community wurden allerdings viele Implementierungen in Sprachen wie Go, C#, Python und Java umgesetzt. Unter folgendem Link ist eine Übersicht der verfügbaren Softwarepakete aufgelistet: <https://graphql.org/code>

Serviceorientierte IT-Landschaften sind überall

In den meisten IT-Landschaften von Unternehmen befinden sich eine Vielzahl verschiedener Schnittstellen. Ob Datenbanken, ERP-Systeme oder DevOps-Server, mittlerweile wird fast jedes System mit einem SOAP oder REST-Webservice ausgestattet.



Abbildung 2: GraphQL-Abfrage und Antwort (© Philipp Hartenfeller)

Über eine daraus entstehende serviceorientierte Architektur können somit flexibel über ein standardisiertes Protokoll auf fremden Systemen Prozesse gestartet und Informationen angefragt werden. Daraus resultiert, dass die gesamte Systemlandschaft eine hohe Interoperabilität aufweist.

Allerdings bringt so ein „Service-Dschungel“ auch einige Nachteile mit sich. Wenn nun viele Anwendungen und Systeme über diese Schnittstellen mit anderen Systemen kommunizieren, entstehen viele Punkt-zu-Punkt-Verbindungen. Dies hat zur Folge, dass ein großes Netz aus Abhängigkeiten über die Webservices entstehen. Meistens sind die Systeme sehr stark gekoppelt, was zur Folge hat, dass viele Anwendungen nicht mehr korrekt funktionieren, wenn eine wichtige Schnittstelle ausfällt. Wartungen oder Updates können somit ein großes Risiko für die Konsumenten darstellen.

Ein weiterer Faktor ist, dass oftmals in Unternehmen relativ intransparent ist, welche Schnittstellen überhaupt existieren, wie diese genau funktionieren und wer diese konsumiert. Letzterer Punkt kann zum Beispiel dazu führen, dass die Abschaltung alter Systeme unvorhergesehene Folgen mit sich bringt. Meistens werden Webservices nicht zentral aufgelistet und dokumentiert, sodass Entwickler nicht von der Existenz einiger Services wissen. Dies ist problematisch, wenn deswegen bereits vorhandene Logik doppelt programmiert wird.

GraphQL als Middleware zur Bündelung von Schnittstellen

Um den im vorherigen Kapitel genannten Problemen entgegenzuwirken, könnte GraphQL dazu eingesetzt werden, Webservices im Unternehmen in einer zentralen Middleware zu bündeln und darüber bereitzustellen. Wie eine solche Systemlandschaft aussehen könnte, ist in *Abbildung 3* dargestellt.

Der Vorteil einer zentralen Middleware, welche Schnittstellen bündelt, ist, dass alle Konsumenten im Optimalfall nur noch die eine Abhängigkeit zur Middleware haben. Somit muss nur noch sichergestellt werden, dass die Konsumenten im Rahmen ihrer Berechtigungen auf diese Middleware zugreifen können und nicht auf die hinterlegenden Systeme. Die Konsumenten müssen außerdem nur noch eine Schnittstellentechnologie implementieren. So eine Implementierung bringt meistens ei-

nen kleinen Overhead mit sich, wenn etwa XML-Anfragen für SOAP-Requests zusammengebaut werden müssen.

Für die Entwickler ist zu jeder Zeit ersichtlich, welche Daten und Operationen verfügbar sind. Die hinterlegenden Systeme und die Logik zu deren Anbindung an die Middleware ist wie eine Blackbox vollkommen nach außen abgekapselt, was die Komplexität runterbricht. Die Entwickler können sich folglich rein auf den Konsum der Middleware-Schnittstelle konzentrieren, und nicht darauf was technisch dahinter liegt.

An dieser zentralen Stelle können dann auch Logging- und Monitoring-Funktionalitäten implementiert werden, um einen Überblick zu erhalten, welche Daten wann, wie häufig von wem verwendet werden. Auch Caching und Authentifizierung sowie Autorisierung kann an dieser zentralen Stelle umgesetzt werden.

Kleine Datenverarbeitungen können gezielt in der Middleware platziert werden, damit die Konsumenten die Daten so wenig wie möglich selbst weiterverarbeiten müssen. Diese Logiken sind dann wiederverwendbar und verhindern somit, dass Abläufe für verschiedene Konsumenten doppelt programmiert werden.

Ein weiterer Vorteil ist, dass in der Middleware Fehlerroutrinen implementiert werden können. Wenn zum Beispiel ein System ausfällt, aber ein zweites System mit ähnlichen Daten vorhanden ist, könn-

ten die Abfragen auf dieses System umgeleitet werden und so Störungen global für alle Konsumenten vermieden werden.

Diese Vorteile sind beispielsweise auch mit Enterprise-Service-Bussen erzielbar. GraphQL bietet für einen solchen Ansatz zusätzliche Vorteile. Dadurch, dass bei GraphQL gezielte Abfragen getätigt werden, können Entwickler deutlich flexibler die benötigten Informationen abrufen. Es muss keine hohe Anzahl an Endpunkten definiert werden, welche auf die Konsumenten zugeschnitten sind. Daten aus verschiedenen Quellen können gebündelt unter einem GraphQL-Type angeboten werden. Da durch den graph-orientierten Ansatz der Technologie zusammenhängende Daten über deren Beziehung abgefragt werden, kann so flexibel durch das Schema navigiert werden. Da in der Middleware das Schema sehr groß werden sollte, verschaffen die Beziehungen einen Überblick über die Datenzusammenhänge.

GraphQL ist zusätzlich quasi selbstdokumentierend. Da ein Schema strikt definiert werden muss, wird diese programmierte Struktur verwendet, um Informationen über die verfügbaren Abfragen, Operationen und Datentypen nach außen freizugeben. In der Regel stellt ein GraphQL-Server auch eine Weboberfläche bereit, in der diese generierte Dokumentation einsehbar ist und wo Abfragen und Operationen auch getestet werden können.

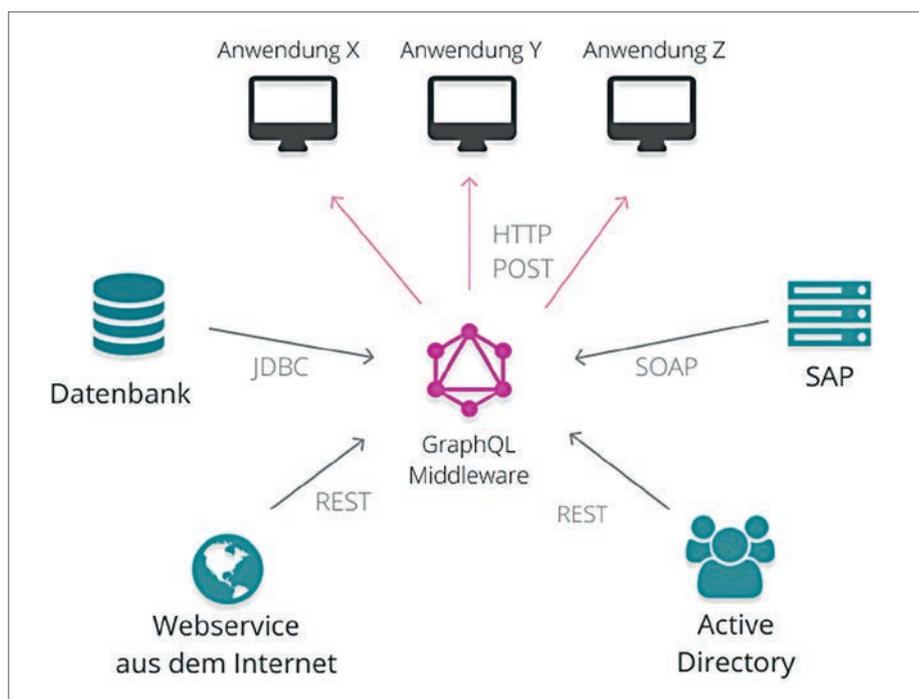


Abbildung 3: Beispielhafte Systemlandschaft mit GraphQL als Middleware (© Philipp Hartenfeller)

Allerdings sollte auch beachtet werden, dass eine GraphQL-Schnittstelle deutlich aufwendiger zu programmieren und wartungsbedürftiger ist, als beispielsweise eine REST-Schnittstelle. Die erzielte Flexibilität kann also dazu führen, dass die Schnittstelle unperformant wird und auf Grund dessen aufwendige Optimierungen vorgenommen werden müssen.

Über den Autor

Philipp Hartenfeller ist als IT-Berater bei der MT AG mit dem Schwerpunkt Oracle Application Express tätig. Neben Datenbanken interessiert er sich für Technologien rund um das Web wie Frontend-Frameworks, Node.js, Docker und User Experience.



Philipp Hartenfeller
philipp.hartenfeller@mt-ag.com

Benachteiligung von Bestandskunden beim Support von „Spatial Graph“ und „Advanced Analytics“

DOAG-Redaktion

Am 5. Dezember 2019 hat Oracle die Funktionalitäten der beiden bis dato separat zu lizenzierenden Optionen in die Datenbank integriert, ohne dass hierfür zusätzliche Kosten entstehen. Bestandskunden, die daher den Support für „Spatial Graph“ und „Advanced Analytics“ nicht verlängern wollen, sehen sich nun mit Problemen konfrontiert.

Berlin, 15. Mai 2020. Die beiden Optionen waren auf der am 6. Dezember 2019 neuerschienenen Preisliste nicht mehr aufgeführt, während die Lizenz- und Supportpreise der Oracle Database Standard Edition 2 (DB SE2) und Oracle Database Enterprise Edition (DB EE) jeweils gleichgeblieben sind. Im Zuge dessen wurden die zum Vertragswerk gehörenden versionsabhängigen Licensing Informations überarbeitet. Demnach sind die Funktionalitäten von „Spatial Graph“ und „Advanced Analytics“ (ehemals „Data Mining“) ab der Version 11.2 in der Oracle DB EE und ab der Version 12.1 auch in der Oracle DB SE 2 enthalten. Diese Zuordnung gilt auch rückwirkend. Somit haben Kunden, die die Funktionalitäten bereits zuvor ohne Lizenz in einer der entsprechenden Datenbank-Versionen genutzt haben, in einem heutigen Audit keine Nachzahlung zu befürchten. Dies wurde der DOAG von Oracle License Management Services (LMS) entsprechend bestätigt.

Während Neukunden vom kostenlosen Zuschlag der Optionen profitieren, ergeben sich Probleme für Bestandskun-

den, die eine oder beide Optionen bereits lizenziert und unter Support haben. Eine Kündigung von Supportverträgen sieht Oracle nicht vor, stattdessen hat der Kunde die Möglichkeit, seinen Supportvertrag nicht zu verlängern. Für einzelne Komponenten ist dies allerdings nicht möglich. Enthält der Supportvertrag neben „Spatial Graph“ und „Advanced Analytics“ weitere Produkte, so werden die Supportgebühren für die Ex-Optionen entweder der Datenbank zugeschlagen oder – wenn ansonsten etwa nur Middleware-Produkte im Supportvertrag enthalten sind – ohne Einzelpreise als Gesamtpreis ausgewiesen, der den Supportbetrag der Ex-Optionen enthält. DOAG-Mitglieder, die sich hierüber beschwert haben und die Positionen der Ex-Optionen monetär aus ihrem Vertrag entfernt haben wollen, berichten, dass Oracle "aus Gutdünken" lediglich Preisnachlässe in Höhe von wenigen Prozent des Supportpreises der Ex-Option gewähre.

Christian Trieb, Leiter des DOAG Competence Center Support und Vorstand der DOAG Datenbank Community, und

Michael Paege, Leiter des DOAG Competence Center Lizenzierung und stellv. Vorsitzender der DOAG, haben versucht, hierüber mit den für Support zuständigen Ansprechpartnern von Oracle zu sprechen. Leider verweigert Oracle ein gemeinsames Telefongespräch und verweist auf die Oracle Technical Support Policies: <https://www.oracle.com/us/assets/057419.pdf>

„Bestandskunden werden gegenüber Neukunden benachteiligt, da sie bereits vor Jahren die Lizenzen beschafft und seitdem auch jährliche Supportzahlungen geleistet haben“, meint Christian Trieb. „Da sich die Grundpreise der Datenbank nicht geändert haben, muss man wirklich von einer deutlichen Schlechterstellung der Bestands- gegenüber den Neukunden sprechen“, unterstützt Michael Paege.

Sollte Oracle seinen Bestandskunden die Nicht-Verlängerung des Supports der Ex-Optionen weiterhin verweigern, hat der DOAG-Gesamtvorstand beschlossen, die Situation juristisch prüfen zu lassen und gegebenenfalls juristische Schritte einzuleiten.

Björn Bröhl ist neuer Vorstandsvorsitzender der DOAG

Christian Luda, DOAG-Redaktion

Am 18. April wurde auf einer informellen virtuellen Delegiertenversammlung die Neubesetzung des Vorstands vorgestellt.

Die für den 18. April 2020 geplante ordentliche Delegiertenversammlung konnte aufgrund der SARS-CoV-2-Eindämmungsverordnung (SARS-CoV-2-EindV) nicht stattfinden und wurde auf unbestimmte Zeit verschoben. Daher wurde für denselben Tag eine informelle Delegiertenversammlung einberufen und online abgehalten.

Zunächst berichtete der Vorstand über die Auswirkungen der Corona-Krise. Aufgrund der Absage aller Veranstaltungen bis mindestens zur Sommerpause sieht sich die DOAG mit erheblichen finanziellen Einbußen konfrontiert. Um die daraus resultierende Finanzlücke zu schließen, wurden bereits im März Maßnahmen zur Kostenreduzierung ergriffen. Hierzu zählen der Shutdown der Geschäftsstelle sowie die Anordnung von Kurzarbeit. Gleichzeitig wird die Neuausrichtung der DOAG vorangetrieben.

Die Maßnahmen haben das erklärte Ziel, sowohl den Verein als auch die GmbH optimal für ein erfolgreiches Jahr 2021 vorzubereiten. Während in diesem Jahr der Fokus auf der Durchführung der DOAG 2020 Konferenz + Ausstellung im November liegt, wird sich die DOAG im kommenden Jahr auf seine drei Hauptevents fokussieren: Die DOAG Konferenz + Ausstellung sowie die JavaLand und die CloudLand. Weitere Fachkonferenzen finden maximal einmal pro Jahr statt und Tagesveranstaltungen sowie andere Formate für Themen, deren Zielgruppen auf den Hauptevents nicht ausreichend repräsentiert sind, werden nach Bedarf durchgeführt.

Ein weiterer Tagesordnungspunkt war die Neubesetzung der aufgrund von Rücktritten freigewordenen Vorstandsposten. Eine formale Beschlussfassung war rechtlich in der virtuellen Versammlung nicht möglich, daher hatte sich der Vorstand für eine Kooptation entschieden und die Delegiertenversammlung ledig-

lich um ein Meinungsbild gebeten. Die Delegiertenversammlung stimmte dem Vorschlag des Vorstands in einem Meinungsbild mit 92 Prozent der abgegebenen Stimmen zu. Im Anschluss an die virtuelle Versammlung kooptierte der Vorstand die freigewordenen DOAG-Vorstandsposten durch Vorstandsbeschluss bis zur nächsten ordentlichen Delegiertenversammlung einstimmig.

In drei DOAG-Communitys wurde ein neuer Vorstand und Leiter kooptiert:

- Vorstand und Leiter der Data Analytics Community: Armin Wildenberg (IKB Deutsche Industriebank AG)
- Vorstand und Leiter der Business Solutions Community: Thomas Karle (Promatis Software GmbH)
- Vorstand und Leiter der Infrastruktur und Middleware Community: Andre Lünsmann (Barmenia Krankenversicherung AG)

Nach dem Rücktritt von Stefan Kinnen wurde Björn Bröhl als neuer Vorstandsvorsitzender kooptiert. Er war zuvor Vorstand

und Leiter der Data Analytics Community und ist – mit kurzer Unterbrechung – seit 2010 DOAG-Vorstandsmitglied. „Um bei der immer schneller werdenden Entwicklung und Transformation von Business und Technologie am Ball zu bleiben, braucht es ein starkes Netzwerk – zum Austausch und Wissenstransfer. Die DOAG bildet dieses Netzwerk und vertritt die Interessen der Anwender gegenüber den Herstellern“, so der neue DOAG-Vorstandsvorsitzende.

Neben dem Engagement bei der DOAG ist Björn auch im Vorstand des EOUC e.V. – dem Verbund der Oracle User Groups in Europa, dem Mittleren Osten und Afrika. Als Head of Product Design & Trends bei Trivadis verantwortet er die Weiterentwicklung des Angebot-Portfolios und identifiziert neue Trends.

Die ordentliche Delegiertenversammlung wird nach aktueller Lage frühestens im Herbst 2020, eventuell im Rahmen der DOAG 2020 Konferenz + Ausstellung, stattfinden.

Die DOAG bedankt sich bei den bisherigen Vorständen und Leitern für die hervorragende Vorstandsarbeit.



BUSINESS NEWS

NEWS

04/2020



Smarte Technologien in der Social Networked Industry

Dietmar Ebel und Lorenz Kiebler, Fraunhofer IML

Smarte Technologien und Systemanwendungen übernehmen auf Basis digitaler Plattformen in Netzwerken künftig die Rolle klassischer Unternehmensapplikationen, wie zum Beispiel von ERP-Systemen. Insbesondere in der Logistik werden smarte Technologien im Rahmen einer Social Networked Industry entwickelt [1]. Durch die zunehmende Globalisierung werden Wertschöpfungsnetzwerke komplexer und Kundenwünsche individueller. Ebenfalls werden die Produktlebenszyklen kürzer und die generelle Dynamik des wirtschaftlichen Handelns steigt an [2]. Gerade auf hohe Stückzahlen, starre Automatisierungen und tiefe Wertschöpfungsstrukturen ausgelegte zentralisierte Produktionssysteme haben mit diesen Trends zu kämpfen. Zusammenfassend steigt die Forderung nach Flexibilität und Dynamik, wodurch die allgemeine Komplexität von Prozessen, Informationen und Planungsaufgaben stetig wächst. Die Beherrschung dieser Komplexität – intern sowie extern – lässt sich nur durch eine durchgängige Digitalisierung der Prozesse und eine Daten- und Informationstransparenz meistern. Unerlässlich hierfür ist die Speicherung und Analyse von Daten mit hierfür ausgelegten IT-Systemen.

In der zukünftigen Arbeitswelt der Social Networked Industry werden Beschäftigte über smarte Devices mit intelligenten und vernetzten Maschinen interagieren. Mit der Weiterentwicklung und dem Einsatz verschiedener Technologien und Ansätze nimmt die Zukunftsvision der Social Networked Industry Gestalt an. Es werden, in Anlehnung an die bekannten sozialen Netzwerke, Kommunikationsplattformen entstehen, die Mensch und Technik auf neuartige Weise in der Arbeitswelt miteinander verbinden – im Ergebnis steht eine sozial vernetzte Industrie. Dieses Konzept geht über das Internet der Dinge (IoT) – bei dem die reine Vernetzung von Dingen im Fokus steht – hinaus, da hier im Rahmen eines serviceorientierten Ansatzes auch angebotene Dienstleistungen oder Services einzelner Entitäten integriert werden.

Langfristiges Ziel einer solchen Weiterentwicklung der Wertschöpfungssysteme stellt das Ideal eines dynamisch und unternehmensübergreifend vernetzten Wertschöpfungsnetzwerks dar. Dieses steuert und optimiert sich selbst durch eine dezentrale Selbstorganisation von intelligenten Systemen und Objekten, die in Verbindung zu Menschen stehen [3]. Diese Dezentralisierung von Entscheidungen in einzelne autonome Einheiten, die vernetzt und intelligent handeln, wird als ein wichtiger Ansatz der Komplexitätsbeherrschung angesehen [4].

Diese Entwicklungen werden in der Logistik unter dem Begriff der „Silicon Eco-

nomy“ zusammengefasst [5]. Die Silicon Economy ist die Daten- und Plattformökonomie, in der Menschen, Unternehmen, autonome Fahrzeuge und IoT-Devices miteinander interagieren. Der Begriff der Silicon Economy spielt mit der Nähe zum Silicon Valley, doch tatsächlich steht er für einen Paradigmenwechsel: Die proprietären, also herstellergebundenen Systeme aus dem Silicon Valley werden in der Silicon Economy aus dem privaten in das betriebliche Umfeld übertragen und durch offene, föderale Strukturen abgelöst [6].

Smarte Anwendungen in der Social Networked Industry

Die Transformation von Unternehmen wird nicht sprunghaft stattfinden, sondern sich Schritt für Schritt, Funktion für Funktion entwickeln. Smarte Unternehmensapplikationen sind ein unerlässlicher Baustein zur digitalen Transformation: Neue Systeme müssen eingebunden, Sensordaten gespeichert und interpretiert sowie entscheidungsunterstützende Informationen generiert und für den Menschen aufbereitet werden. Sowohl stark spezialisierte Anwendungen und Systeme für einzelne Aufgaben als auch unternehmensweite Softwarearchitekturen und Softwarelösungen müssen dahingehend betrachtet und entwickelt werden.

In der Social Networked Industry wird die Arbeitswelt der Zukunft mit einer engen Kollaboration zwischen Mitarbeitern und intel-

ligenten Maschinen oder Assistenzsystemen betrachtet und konzeptionell sowie technisch weiterentwickelt. Unter Anlehnung an die verbesserten Möglichkeiten von Kommunikation und Zusammenarbeit durch soziale Netzwerke im privaten Raum wird das Konzept digitaler Netzwerke so auf den gewerblichen und industriellen Sektor übertragen. Somit wird der Vernetzungsgedanke der Industrie 4.0 über den reinen Ansatz einer technischen Automation hin zu einer intuitiven Kommunikation zwischen Mensch und Maschine weiterentwickelt.

Für die Etablierung und Weiterentwicklung der Social Networked Industry sind neben den namensgebenden Social Networks, welche die dynamische Organisation von Mensch und Maschine ermöglichen, weitere Bausteine notwendig. So müssen Assistenzsysteme und Smart Devices für die Unterstützung des Menschen als Entscheider entwickelt, die Mensch-Technik-Interaktion verbessert, die digitale Gestaltung durch digitale Abbilder der realen Welt ermöglicht und die betriebsübergreifende Koordination durch die horizontale Vernetzung vertieft werden [7]. Besonders in den Bausteinen der intelligenten Assistenzsysteme und intelligenten Endgeräte (Smart Devices), der Mensch-Technik-Interaktion und dem Management sowie der Erstellung von digitalen Abbildern der realen Systeme und Objekte ermöglichen smarte Unternehmensapplikationen die fortschreitende Um-

setzung dieser Vision, wie beispielhafte Use Cases aus dem Fraunhofer IML darstellen:

Mit dem Ziel der Digitalisierung des stark analog organisierten Verpackungsprozesses wurde im Rahmen eines Transferprojekts der intelligente Packarbeitsplatz iPackAssist entwickelt und umgesetzt. Basierend auf zwei unabhängigen Technologien wurde die Unterstützung der Mitarbeiter konzipiert, um den steigenden Anforderungen der Versandlogistik durch höhere Anforderungen an die Verpackung und stetig zunehmende Paketvolumina durch den Onlinehandel zu begegnen. Durch die intelligente Software PUZZLE, welche die optimale Paketgröße sowie die ideale Packanordnung und Reihenfolge einer Packliste ermittelt, lassen sich Packvorschriften dynamisch im Verpackungsprozess generieren. Das Assistenzsystem pAsst bildet diese Packvorschriften intuitiv und leicht verständlich durch optische und interaktive Darstellungen in Abhängigkeit des Packprozesses auf dem Arbeitsplatz ab. Die Mensch-Maschine-Schnittstelle visualisiert Ausrichtung und Anordnung der Produkte durch zweidimensionale illuminierte Achsen in dem Packtisch (siehe Abbildung 1) [7].

Im Rahmen der Produktion und Montage sind neue Assistenzsysteme für Mitarbeiter notwendig, um den Anforderungen der Individualisierung von Kundenwünschen und Produkten begegnen zu können. Durch die hohe Varianz an Varianten und Produkten sind auch Arbeitsanweisungen je nach Produkt unterschiedlich und nicht mehr standardisiert anzuwenden. Durch das vom IML entwickelte Kamera-Assistenzsystem für Montagearbeitsplätze wird der Mitarbeiter oder die Mitarbeiterin monitorgestützt durch den individuellen Montageprozess geführt.



Abbildung 1: Intelligenter Packarbeitsplatz iPackAssist (© Michael Neuhaus – Fraunhofer IML)

Durch Schritt-für-Schritt-Bauanleitungen sowie eine Kontrolle der verwendeten Teile und der Ergebnisse durch einen Bildabgleich per Kamera werden Fehler vermieden und die Produktkomplexität beherrscht. Eine beispielhafte Integration des Assistenzsystems in einen Montagearbeitsplatz ist in Abbildung 2 dargestellt. Durch das Systemwissen über den Erfahrungsstand der eingesetzten Mitarbeiter können Unterstützungen und Anleitungen in Abhängigkeit vom individuellen Mitarbeiterwissen eingesetzt oder nicht eingesetzt werden, um nur dort zu unterstützen, wo eine Unterstützung notwendig ist. [8]

Mit dem Ziel der Nachverfolgbarkeit von Lieferungs- und Warenbewegungen wurde der Low Cost Tracker in Zusammenarbeit im

Telekom Open IoT Lab entwickelt. Ausgestattet mit Sensoren zur Bewegungs- und Temperaturmessung ist der Tracker in das NarrowBand IoT eingebunden, welches das Maschinen- und Sensornetz der deutschen Telekom darstellt. Vorteil des Trackers ist die lange Batterielaufzeit von fünf bis zehn Jahren durch den geringen Verbrauch und regelmäßige, aber nicht ständige Update-meldungen (ca. 2 pro Tag) und die einfache Befestigung der Tracker in dafür vorbereiteten Euro-Paletten (siehe Abbildung 3). Durch dieses Nachrüsten von intelligenten Sensor- und Vernetzungsfähigkeiten an bestehende Gegenstände wird das Internet der Dinge ausgebaut und auch auf vorher nicht digitalisierte Bereiche ausgeweitet [9].

Zusätzlich zu den dargestellten Anwendungsfällen existieren weitere Technologien und Entwicklungen in Bereichen wie der Produktion, der Logistik, der Instandhaltung oder dem Handel. Diese einzelnen Applikationen und deren integrative Betrachtung in komplexeren Use Cases oder Anwendungsszenarien stellen beispielhafte Ansätze dar. Hierbei ist besonders die Integration von Praxisanwendungen wichtig, um eine sinnvolle und den bestehenden Herausforderungen und Restriktionen entsprechende Digitalisierung sicherzustellen.

Digitale Transformation der IT-Architekturen

Neben der Entwicklung einzelner Applikationen oder Systeme ergibt sich die Notwendigkeit des Aufbaus einer modernen IT-Systeminfrastruktur. Häufig basieren Ge-



Abbildung 2: Kamerabasiertes Assistenzsystem für Montagearbeitsplätze (© Michael Neuhaus – Fraunhofer IML)

schäftsprozesse auf nicht geeigneten oder veralteten Systemen und Strukturen, die den neuen Anwendungsfällen nicht gewachsen sind. Eine durchgehende, transparente und digitale Erfassung, Analyse und Entscheidungsgenerierung mit IT-Unterstützung ist mit einer solchen Infrastruktur nicht möglich [10]. Somit muss zuerst eine moderne IT-Infrastruktur konzeptioniert und implementiert sein, um mit smarten Unternehmensapplikationen auf den bestehenden Informationen aufzubauen oder neue Datenquellen anzuschließen. Werden diese Informationen und Datenquellen unternehmensübergreifend strukturiert und geregelt gebündelt, entstehen Möglichkeiten für neue Geschäftsmodelle. Vergleichbar mit den Entwicklungen im privaten Bereich werden auch im B2B-Bereich Plattformen entstehen, auf denen verteilte Systeme und Teilnehmer Informationen, Dienste und Waren anbieten, über diese mit Interessenten verhandeln und Verträge abschließen. Unter dem Begriff der Plattformökonomie wird diese kommunikationsbezogene und datenintensive Steuerung von Geschäftsprozessen und Geschäftsvorfällen zusammengefasst. In einer solchen Plattformökonomie werden klassische Geschäftsmodelle wie der einfache Verkauf von Produkten durch digitale Zusatzservices erweitert und komplett neue servicebasierte Geschäftsmodelle entstehen. Für Gebrauchsgüter fallen hierunter Services wie zum Beispiel die Abrechnung



Abbildung 3: Low Cost Tracker (© Michael Neuhaus – Fraunhofer IML)

der tatsächlichen Maschinen- und Anlagenutzung in Mietmodellen oder das Angebot von vorbeugenden Instandhaltungs- und Wartungsleistungen durch die Fernanalyse und -auswertung von Sensordaten auf Basis von KI- und Predictive-Maintenance-Ansätzen. Für Verbrauchsgüter werden neue Services entstehen, in denen die Nutzungsgewohnheiten der Endverbraucher integriert werden, insbesondere durch die Verknüpfung von produktbezogenen und produktneutralen Features.

Ermöglicht wird diese Entwicklung durch die neu generierten, kommunizierten, gespeicherten und durch Big-Data-Analysen und KI-Algorithmen ausgewerteten Daten und Infor-

mationen. Die Beachtung von Aspekten der Datensouveränität und der notwendigen Geheimhaltungsregelungen muss bei der Konzeption von Plattformen und Geschäftsmodellen zwingend mit beachtet werden.

Durch die immer weiter voranschreitende Auflösung der Automatisierungspyramide und die sich damit verändernden Aufgaben, Zuständigkeiten und Funktionsweisen stehen klassische Unternehmensapplikationen vor einer großen Anpassungsherausforderung [11]. Der Wandel in der inner- und überbetrieblichen IT wird sich weg von einer starren Integration und hin zu einer dynamischen Kommunikation der verschiedenen Funktionen und Applikationen eines Systems entwi-

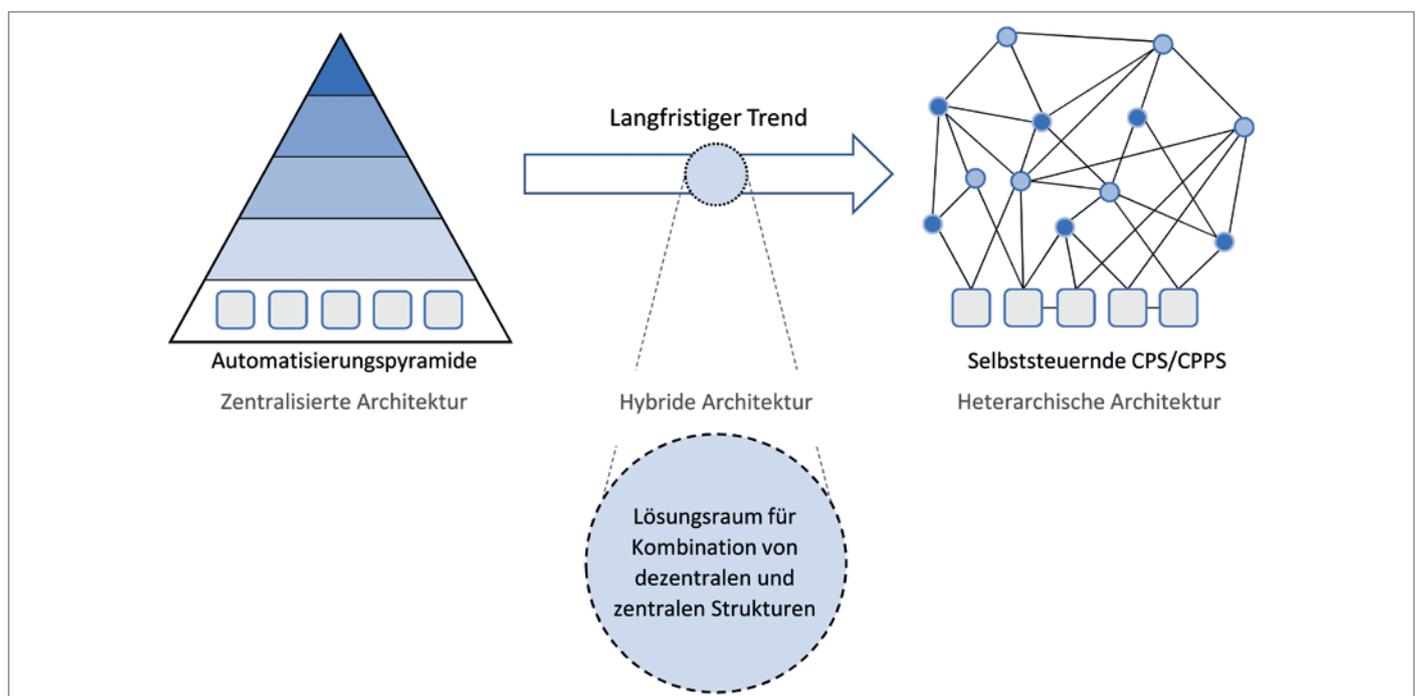


Abbildung 4: Auflösung der Automatisierungspyramide und hybride Zwischenformen (in Anlehnung an [3, 11])

ckeln. Aus den streng hierarchischen Steuerungsebenen von der Unternehmensspitze hin zum Shopfloor entstehen Netzwerke aus dezentral organisierten Entscheidungsentitäten, die auf Basis ihrer eigenen Zielsetzungen autonom Handlungsalternativen auswählen und so die Wertschöpfung steuern. Diese über das Internet kommunizierenden und mit eingebetteten Softwaresystemen ausgestatteten Objekte, Geräte oder Anlagen werden als cyberphysische Systeme (CPS) beschrieben. CPS können ihre Umwelt durch Sensorik wahrnehmen, diese Informationen gemeinsam mit weiteren weltweit verfügbaren Daten zielgerichtet auswerten, Handlungen ableiten und durch eine entsprechende Aktorik die Umwelt wieder beeinflussen. [12] Wie in *Abbildung 4* dargestellt, entstehen somit langfristig dynamische und modular erweiterbare Netzwerke, in denen Entscheidungen dezentral getroffen werden. Netzwerke aus einzelnen cyberphysischen Systemen werden je nach Anwendungsdomäne als cyberphysische Produktions- oder Logistiksysteme (CPPS/CPLS) bezeichnet. Auszuführende Aufgaben, Bearbeitungen oder sonstige Dienstleistungen werden als sogenannte Services angeboten und nachgefragt, sodass eine serviceorientierte Architektur (SOA) entsteht [4]. Auf dem Weg zu dieser echtzeitnahen Steuerung und einer, auf durchgehenden und offenen Standards basierenden, dynamischen Vision werden verschiedene Zwischenziele und hybride Architekturkonzepte entstehen, da eine schnelle Umstellung in einem Schritt sowohl aus technischer als auch aus organisatorischer und finanzieller Sicht nicht oder nur schwer möglich ist.

Besonders in dieser Übergangsphase haben die bestehenden ERP-Systeme eine entscheidende Rolle inne. Da Unternehmens-, Produktions- und Bestandsdaten häufig in ERP-Systemen gespeichert, Prozesse durch ERP-Workflows gesteuert und das Reporting aus ERP-Daten aufgebaut werden, lässt sich langfristig eine Anpassung der Softwarestrukturen nur unter starker Einbindung der aktuell zentralen Unternehmensplanungssysteme umsetzen.

Zukünftig werden statt eines vollintegrierten Lösungsbausteins dynamische, Appbasierte und echtzeitfähige Applikationsnetzwerke entstehen. Durch die zentrale und integrierende Rolle in den aktuellen IT-Architekturen müssen besonders ERP-Systeme durch eine erhöhte Wandlungsfähigkeit schnell an zukünftige Änderungen angepasst werden können. Außerdem müssen die an-

gebundenen Drittsysteme aufwandsarm integriert sowie Massendaten effizient gespeichert und auch analysiert werden können. Beispielhaft für allgemeine Anforderungen an Unternehmensapplikationen sind in *Abbildung 5* die funktionalen sowie die technischen Anforderungen an ERP-Systeme im Kontext der Industrie 4.0 dargestellt. Die Systeme müssen echtzeitfähig, mobil verfügbar und modular aufgebaut sein, um den veränderten Anforderungen genügen zu können. Die Dematerialisierung umschreibt die Trennung von Software und Maschinenhardware, also die Entkopplung der Steuerung von den eigentlichen Aktoren auf dem Shopfloor. Diese Verlagerung der Steuerung ermöglicht eine umfassendere und einfachere Maschinendatenerfassung, hat allerdings auch einen großen Einfluss auf die grundlegende IT-Architektur. Entsprechende Entwicklungen laufen bereits bei vielen ERP-Anbietern, um die Grundlage für eine zunehmende digitale Transformation der Wertschöpfungssysteme bieten zu können [10].

Die großen Veränderungen von innerbetrieblichen Applikationen und IT-Strukturen enden jedoch nicht an den juristischen Unternehmensgrenzen. Gerade entlang komplexer Supply Chains werden große Veränderungspotenziale entstehen, um gestiegene

Komplexitätsgrade und Flexibilitätsanforderungen auch im überbetrieblichen Anwendungsfall bewältigen zu können. Neue Innovationsprozesse, aber auch die verstärkte überbetriebliche Kommunikation und Vernetzung müssen entwickelt, abgestimmt und implementiert werden [13].

Mit dem Ziel der proaktiven Gestaltung dieser überbetrieblichen Wertschöpfungsnetzwerke unter Beachtung der digitalen Transformation und den gestiegenen Anforderungen wurde das Zielbild einer Silicon Economy entworfen. Das Zielbild stellt die unternehmensübergreifende Verallgemeinerung und Erweiterung der Social Networked Industry dar. In Systemen wächst die Dezentralität und die Selbstorganisation mit der Komplexität [14], weswegen vor allem im komplexen überbetrieblichen Fall die verteilte Lösungskompetenz betont wird. Die zentralen Prozesse werden immer weiter aufgebrochen und durch agile und dezentrale Entscheidungen ersetzt. Basierend auf dieser Erkenntnis müssen dezentrale Managementansätze für die überbetriebliche Ebene entworfen, aber vor allem in den einzelnen teilnehmenden Unternehmen umgesetzt werden.

Besonders durch die Kombination der Technologien des IoT und der Blockchain-Technologie lassen sich dezentrale Wert-



Abbildung 5: Technische und funktionale Anforderungen an ERP-Systeme (in Anlehnung an [10])

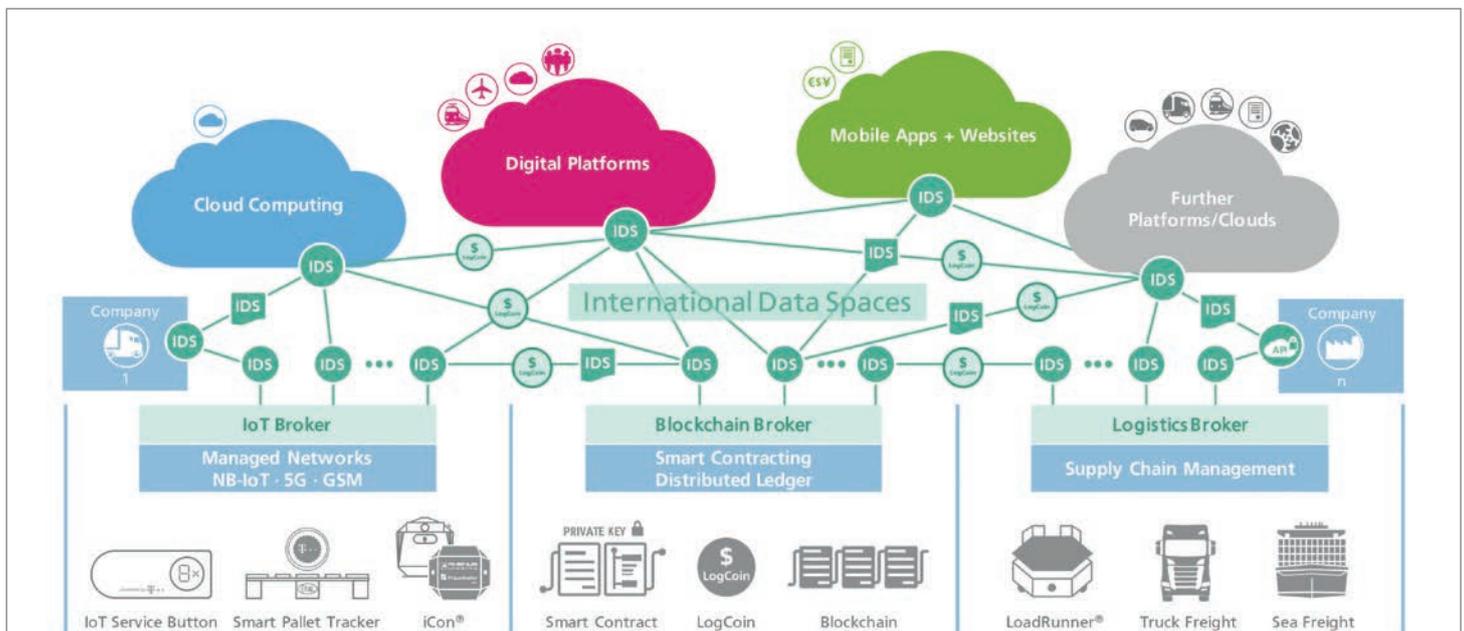


Abbildung 6: Big Picture der Silicon Economy (© Fraunhofer IML)

schöpfungsnetzwerke ohne regelnde oder steuernde Zwischeninstanzen etablieren. Das in *Abbildung 6* abgebildete Big Picture der Silicon Economy stellt die Vernetzung der verschiedenen Unternehmen unter Nutzung von Cloud-Computing, digitalen Plattformen, mobilen Anwendungen und Big Data dar. In verschiedenen Brokern werden die drei relevanten Anwendungsaspekte in das Modell integriert. CPS der physischen Ebene werden in das Netzwerk eingebunden (IoT-Broker), die rechtlichen und vertraglichen Aspekte der plattformgebundenen Interaktionen werden durch die Blockchain abgebildet (Blockchain-Broker) und die dezentralen Planungs- und Steuerungsansätze und die Einbindung von Organisation und Geschäftsmodellen werden durch den Logistics-Broker sichergestellt. Die wichtigen Aspekte der Datensouveränität im Zuge der steigenden Transparenz- und Informationsbedarfe werden durch die Modelle und Ansätze der Industrial Data Spaces Association betrachtet und gelöst, da je Information eindeutige Regelungen für die Weitergabe und Nutzung festgelegt werden können [15].

Durch die sich aktuell entwickelnden Technologien wie der künstlichen Intelligenz oder den additiven Fertigungsverfahren – wie dem 3D-Druck – stehen die nächsten Technologien bereit, um in IT-Systeme und Geschäftsprozesse integriert zu werden. Hierdurch werden sich die IT-Architekturen weiter anpassen, um näher an das Ideal der intelligenten Wertschöpfungsnetzwerke der Industrie 4.0 zu rücken.

Literaturangaben

- [1] Social Networked Industry ganzheitlich gestalten. Future Challenges in Logistics and Supply Chain Management, Whitepaper, Tüllmann, C., ten Hompel, M., Netzträger, A. u. Prasse, C., 2017
- [2] Obermaier, R.: Industrie 4.0 als unternehmerische Gestaltungsaufgabe. Strategische und operative Handlungsfelder für Industriebetriebe. In: Obermaier, R. (Hrsg.): Industrie 4.0 als unternehmerische Gestaltungsaufgabe. Betriebswirtschaftliche, technische und rechtliche Herausforderungen. Wiesbaden: Springer Fachmedien 2016, S. 3–34
- [3] BITKOM e.V., VDMA e.V. u. ZVEI e.V. (Hrsg.): Umsetzungsstrategie Industrie 4.0. Ergebnisbericht der Plattform Industrie 4.0. Berlin: Bitkom Research GmbH 2015
- [4] Bauernhansl, T.: Die Vierte industrielle Revolution – Der Weg in ein wertschaffendes Produktionsparadigma. In: Vogel-Heuser, B., Bauernhansl, T. u. ten Hompel, M. (Hrsg.): Handbuch Industrie 4.0 Bd.4. Allgemeine Grundlagen. Springer Reference Technik. Berlin: Springer Vieweg 2017, S. 1–31
- [5] Silicon Economy: Wie künstliche Intelligenz die Welt in eine Plattformökonomie führt. LOGISTIK entdecken 20 (2020), S. 6–14
- [6] Fraunhofer-Institut für Materialfluss und Logistik IML: Logistik ruft Silicon Economy aus. Durchbruch mit digitalen Plattformen und Verfahren Künstlicher Intelligenz, 2020. <https://www.ima.fraunhofer.de/de/silicon-economy.html>
- [7] Lehmann, S.: Verpackung und Versand: Passt so! Logistik Heute 41 (2019) 9, S. 60–61
- [8] Schier, A.: Innovationslabor Hybride Dienstleistungen in der Logistik. LOGISTIK entdecken 19, S. 10–17
- [9] Kerner, S.: Klein, aber Oho. LOGISTIK entdecken 19, S. 22–25
- [10] ERP-Marktstudie 2016: Auswirkungen von Industrie 4.0 auf die Anforderungen an ERP-Systeme, Klink, P., Mertens, C. u. Kompalka, K., Dortmund 2016
- [11] Ebel, D. u. Kiebler, L.: Digitale Transformation. Auswirkungen dezentraler und autonomer Planungsarchitekturen auf ERP-Systeme. ERP Management 14 (2018) 4, S. 33–35
- [12] Bauernhansl, T.: Die Vierte Industrielle Revolution – Der Weg in ein wertschaffendes Produktionsparadigma. In: Vogel-Heuser, B., Bauernhansl, T. u. ten Hompel, M. (Hrsg.): Handbuch Industrie 4.0 Bd.4. Allgemeine Grundlagen. Springer Reference Technik. Berlin: Springer Vieweg 2017, S. 1–31
- [13] Henke, M., Besenfelder, C. u. Kaczmarek, S.: Dortmund Management-Modell. In: ten Hompel, M., Vogel-Heuser, B. u. Bauernhansl, T. (Hrsg.): Handbuch Industrie 4.0. Produktion, Automatisierung und Logistik. Springer Nachschlagewissen. Wiesbaden: Springer Fachmedien Wiesbaden 2019, S. 1–17

[14] ten Hompel, M. u. Henke, M.: Logistik 4.0 – Ein Ausblick auf die Planung und das Management der zukünftigen Logistik vor dem Hintergrund der vierten industriellen Revolution. In: Vogel-Heuser, B., Bauernhansl, T. u. ten Hompel, M. (Hrsg.): Handbuch Industrie 4.0 Bd.4. Allgemeine Grundlagen. Springer Reference Technik. Berlin: Springer Vieweg 2017, S. 247–268

[15] Industrial Data Space - Digitale Souveränität über Daten. White Paper, Otto, B., Auer, S., Cirullies, J., Jürjens, J., Mend, N., Schon, J. u. Wenzel, S., München 2016



Dietmar Ebel

dietmar.ebel@iml.fraunhofer.de

Dietmar Ebel ist Teamleiter des Teams ERP und Digital Design am Fraunhofer Institut für Materialfluss und Logistik.



Lorenz Kiebler

lorenz.kiebler@iml.fraunhofer.de

Lorenz Kiebler ist wissenschaftlicher Mitarbeiter im Team ERP und Digital Design am Fraunhofer Institut für Materialfluss und Logistik.



Die Bedeutung von Plattformen im IoT-Umfeld

Norbert Gronau und Benedict Bender, Universität Potsdam

Produkte werden zunehmend internetfähig. Gerade im Kontext der Industrie 4.0 statten beispielsweise Maschinen- und Anlagenbauer sowie Komponentenhersteller ihre Produkte mit Option zur Netzwerk- und Internetanbindung aus. Neben der technischen Realisierung beschäftigt Unternehmen die Frage, wie diese neue Geschäftsmodelle realisieren können. Gerade bei Produkten, die bei Kunden im Einsatz sind, kommt der Nutzung von IoT-Plattformen eine wichtige Bedeutung zu.

Das Internet der Dinge (engl. Internet of Things, kurz IoT) bietet Unternehmen zahlreiche Möglichkeiten, neuartige Geschäftsmodelle zu realisieren. Die zunehmende Vernetzung von Produkten bietet Herstellern ebenso wie Kunden weitreichende Möglichkeiten [1]. Im Maschinen- und Anlagenbau können Kunden beispielsweise von einer Fernüberwachung ihrer Anlagen oder Manufacturing Analytics profitieren [4]. Hersteller können gezielte Wartungsservices, zum Beispiel in der Form einer Ferndiagnose oder durch den Vertrieb von Ersatzteilen bei Ausfall, anbieten. Weiterhin können Hersteller Informationen über den Einsatz ihrer Maschinen (Nutzungsdaten) gewinnen, die als Grundlage für die gezielte Produktverbesserung genutzt werden können [2].

Die Etablierung zusätzlicher Geschäftsmodelle stellt viele Unternehmen jedoch vor Herausforderungen. Einerseits fragen sich Unternehmen, welche Geschäftsmodelle realisiert werden können. Andererseits besitzen viele Unternehmen nicht die notwendigen Kompetenzen um die integrierte Konzeption von technischer Lösung sowie neuartigem Geschäftsmodell zu realisieren. Bei der Realisierung neuartiger Geschäftsmodelle kommt IoT-Plattformen eine gesteigerte Bedeutung zu.

IoT-Plattformen ermöglichen die Anbindung einer Vielzahl von Endgeräten sowie die Erhebung derer Statusparameter. Zudem bieten Plattformen die Möglichkeit, diese Informationen (zum Beispiel Sensorwerte) zu speichern, auszuwerten sowie Regelwerke zu definieren. Durch die Anbindung unterschiedlicher Maschinen können auch übergreifende Steuerungsaufgaben realisiert werden [5]. Dies bietet zahlreiche Mehrwerte für Kunden und Hersteller.

Eigene Plattform oder Anbindung

Beabsichtigen Unternehmen, eine digitale Plattform für ihre Produkte anzubieten, bestehen zwei grundlegende Lösungsszenarien. Unternehmen können eine eigene Plattform bereitstellen oder ihre Produkte für die Anbindung an eine am Markt verfügbare Plattform optimieren (z.B. AWS IoT Core). Im ersten Lösungsszenario ist der Hersteller auch gleichzeitig Anbieter der Plattform und kann somit ein integriertes Portfolio anbieten. Im zweiten Lösungsszenario stellt der Hersteller sein Produkt her, das lediglich für die Nutzung mit externen Plattformen vorbereitet ist. Im Folgenden werden drei wesentliche Unterschiede der zwei Lösungsszenarien systematisiert.

Realisierungsaufwand und Know-how

Die zwei Lösungsmöglichkeiten unterscheiden sich weitgehend in dem Aufwand der damit verbundenen Realisierung. Bei der Anbindung an eine externe Plattform ist lediglich die Umsetzung der Spezifikationen der gewählten Plattform oder die Implementierung von offenen Standards zu bewerkstelligen. Deutlich höher ist der Aufwand für die Realisierung einer eigenen Plattform. Auch hierfür existieren unterschiedliche Möglichkeiten. Neben der vollständigen Eigenentwicklung können etablierte Technologiebausteine für die Realisierung genutzt werden, ebenso wie auf sogenannte Whitelabel-Plattformen (zum Beispiel Cumulocity) zurückgegriffen werden kann. Während eine Eigenentwicklung vollständige Freiheit hinsichtlich der Realisierung bietet und individuelle Anforderungen realisiert werden können, so ist der damit verbundene Aufwand hoch und kommt insbesondere für KMU nicht infrage [1]. Insgesamt ist der Aufwand für den Betrieb einer eigenen Plattform höher als

bei der Anbindung an eine externe Plattform, die von Dritten betrieben, gepflegt und weiterentwickelt wird. Weiterhin ist das für die Realisierung der Lösung benötigte Know-how stark unterschiedlich. Während die Anbindung an eine externe Plattform lediglich die Umsetzung gegebener Spezifikationen erfordert, so sind für die Konzeption, Umsetzung und Pflege einer eigenen Plattform weitreichende Kenntnisse und Kapazitäten erforderlich. Auch das Management der Plattform während des Betriebs erfordert domänenspezifisches Wissen und Marktkenntnis.

Gestaltungsfreiheit

Der Grad der Lösungsfreiheit und Autonomie unterscheidet sich ebenfalls gemäß der gewählten Lösung. Im Falle einer eigenen Plattform hat der Anbieter weitreichende Gestaltungsfreiheit im Hinblick auf die Lösung. Bei einer externen Plattform hingegen hat der Anbieter nahezu keinen Einfluss auf die Weiterentwicklung der Plattform. Zudem verfolgen insbesondere große Plattformanbieter zumeist eigene Interessen und sind aufgrund der Marktmacht auch nicht auf die Berücksichtigung der Interessen einzelner Produkthersteller angewiesen. Weiterhin entwickeln sich Plattform sowie das zugehörige Angebot über die Zeit weiter. Auch die funktionalen Gestaltungsmöglichkeiten auf der Plattform sind stark eingeschränkt. Während einzelne Plattformen die Entwicklung von Add-ons ermöglichen, so sind doch die grundlegenden Strukturen, Datenformate und Lösungsmöglichkeiten vorgegeben. Damit einhergehend sind auch die Möglichkeiten zur Erweiterung des Geschäftsmodells auf externen Plattformen stark eingeschränkt. Insbesondere wenn diese mit den Angeboten des Plattformanbieters konkurrieren, entstehen Interessenskonflikte.

Langfristig besteht die Gefahr der Übernahme beziehungsweise Ausgrenzung des Herstellers durch den Plattformanbieter in Bezug auf die angebotenen Zusatzleistungen [3]. Dementsprechend sind im Hinblick auf die Angebote zusätzlicher, eigener Services mit Geschäftsbezug eigene Plattformen vorzuziehen.

Marktperspektive

Im Rahmen der Entscheidung ist auch die Marktperspektive angemessen zu berücksichtigen. Etablierte Plattformen zeichnen sich zumeist durch ein umfangreiches Ökosystem von Nutzern und Anbietern aus [6]. Insbesondere Kunden profitieren durch die Netzwerkeffekte zum Beispiel in Form von Kompatibilität mit anderen eingesetzten Maschinen im Rahmen der betrieblichen Wertschöpfung. Entscheidet sich ein Unternehmen für den Aufbau einer eigenen Plattform, sind derartige Ökosystem aufzubauen und zu entwickeln. Meist fokussieren sich Hersteller jedoch auf die Anbindung eigener Produkte, wohingegen spezialisierte Plattformanbieter die Anbindung unterschiedlicher Hersteller fokussieren. Auch aus Kundenperspektive ist die Frage zu stellen, ob die angebotenen Produkte zum Beispiel im Kontext von Industrie 4.0 mit einer Vielzahl weiterer Produkte zu vernetzen sind oder es weitestgehend isolierte Produkte sind, die von Spezialfunktionen profitieren (zum Beispiel Windkraftanlagen).

Für Unternehmen stellt sich die zentrale Frage, ob eine eigene IoT-Plattform bereitgestellt oder Produkte an bereits im Markt verfügbare Lösungen angehängt werden sollen. Welches die geeignete Strategie für ein Unternehmen ist, hängt im Wesentlichen von der Zielstellung des Unternehmens ab. Ausgehend von den Zielen kann die Bedeutung der Plattform ermittelt und die entsprechende Strategie entwickelt werden. Die Anbindung an externe Plattformen zeichnet sich durch geringen Aufwand, jedoch auch geringe Gestaltungsmöglichkeiten aus. Eine eigene Plattform bietet umfangreiche Möglichkeiten zur Gestaltung und Erweiterung der Geschäftstätigkeit, erfordert jedoch mehr Aufwände sowie entsprechende Kompetenzen. Soll die IoT-Plattform ein zentrales Element zur Differenzierung im Wettbewerb darstellen, so bietet beispielsweise eine eigene Plattform bessere Möglichkeiten zur individuellen Ausgestaltung. Auch für die Realisierung neuartiger Geschäftsmodelle bietet eine eigene Plattform mehr Möglichkeiten

und ist durch weniger direkte Konkurrenz auf der Plattform geprägt. Beabsichtigt ein Unternehmen jedoch nur die Erfüllung einzelner Kundenwünsche, die für das Unternehmen von geringer Bedeutung sind, so stellt die Anbindung an etablierte Plattformen eine aufwandsarme Möglichkeit dar. Dennoch sollten Unternehmen die Bedeutung von Plattformen für ihr Geschäft fortwährend evaluieren und gegebenenfalls notwendige Anpassungen vornehmen.

Der Weg zum Angebotsportfolio

Im Hinblick auf die Frage, welche Geschäftsmodelle Unternehmen realisieren sollen, gibt es keine allgemeingültige Antwort. Dennoch sind Unternehmen angehalten, mögliche Geschäftsfelder systematisch zu erarbeiten. Allgemein sollten mögliche Alternativen evaluiert, bewertet und entsprechend priorisiert werden. Bei der Identifikation möglicher Geschäftsfelder kann die Expertise einschlägiger Branchenexperten oder die gezielte Kooperation mit Kunden helfen. Gegenwärtige Kunden bieten gute Impulse für die Erweiterung des bestehenden Geschäftsmodells auf Basis aktueller und zukünftiger Kundenbedürfnisse. Hingegen bietet die Interaktion mit potenziellen Kunden in strategischen Zielsegmenten die Option, Auskunft über die Attraktivität und Nützlichkeit eines geplanten Angebots zu erhalten.

Zusammenfassung

Die Potenziale digitaler Plattformen sind im Hinblick auf die Realisierung technischer Neuheiten ebenso wie die Realisierung von neuartigen Geschäftsmodellen vielfältig. Unternehmen sollten sich der Potenziale für ihren konkreten Anwendungsfall bewusst werden sowie darauf basierend eine geeignete Strategie entwickeln. IoT-Plattformen sind im produktbezogenen Umfeld bestens geeignet, um produktbegleitende Dienstleistungen und Zusatzservices anzubieten. Gegenwärtige wie auch potenzielle Zielkunden sind für die Konzeption des Angebotsportfolios von großer Wichtigkeit. Nach Definition der Ziele muss geklärt werden, ob die Plattform selbst angeboten oder eine Anbindung an bestehende Plattformen realisiert werden soll. Hierfür sind insbesondere die Ziele und die strategische Bedeutung im Portfolio des Unternehmens ausschlaggebend. Und nun: Viel Erfolg bei der Erarbeitung und Umsetzung Ihrer neuen Geschäftsmodelle!

Quellen

- [1] Bender, B., Grum, M., Gronau, N., Alfa, A., & Maharaj, B. T. (2019) Design of a Worldwide Simulation System for Distributed Cyber-Physical Production Networks. In 2019 IEEE International Conference on Engineering, echnology and Innovation (ICE/ITMC) IEEE, Nizza, Frankreich.
- [2] Bender, B., Habib, N., Gronau, N. (2020) Digitale Plattformen: Strategien für KMU. Wirtschaftsinformatik & Management, Springer, (in Druck).
- [3] Bender, B. und Gronau, N. (2017) Coring on Digital Platforms – Fundamentals and Examples from the Mobile Device Sector in: International Conference on Information Systems (ICIS). Seoul, Südkorea.
- [4] Verhaelen, B., Thomas, K., Häfner, B., Lanza, G., & Schuh, G. (2019). Potenziale datenbasierter Produktallokationen. ZWF Zeitschrift für wirtschaftlichen Fabrikbetrieb, 114(3), 96-100.
- [5] Gronau, N., Grum, M., und Bender, B. (2016). Determining the optimal level of autonomy in cyber-physical production systems. In 2016 IEEE 14th International Conference on Industrial Informatics (INDIN) (pp. 1293-1299). IEEE.
- [6] Müller-Stewens, G., und Stonig, J. (2019). Digitale Transformation: Werttreiber beim Aufbau plattformbasierter Ökosysteme in etablierten Unternehmen. Controlling, 31(6), 4-10.



Prof. Dr.-Ing. Norbert Gronau
n Gronau@lswi.de

Prof. Dr.-Ing. Norbert Gronau ist Inhaber des Lehrstuhls für Wirtschaftsinformatik, Prozesse und Systeme an der Universität Potsdam. Durch seine Tätigkeiten als Direktor des Center for Enterprise Research sowie Gründer der Potsdam Consulting Advisory GmbH finden seine Forschungsergebnisse direkten Einzug in die betriebliche Praxis. Als Trusted Advisor begleiten er und sein Team Unternehmen bei der Konzeption und Realisierung zukunftsfähiger Geschäftsmodelle.



Benedict Bender
Benedict.Bender@wi.uni-potsdam.de

Benedict Bender ist wissenschaftlicher Mitarbeiter am Lehrstuhl für Wirtschaftsinformatik, Prozesse und Systeme an der Universität Potsdam. Der Praxistransfer seiner Forschungsergebnisse wird u.a. durch seine Tätigkeiten als Managementberater und Coach erreicht. Herr Bender unterstützt Unternehmen aller Größen bei der Erarbeitung zukunftsfähiger Geschäftsmodelle sowie bei der Realisierung von digitalen Plattformen.



"Vorpreschen und neue Dinge ausprobieren."

Die Entwicklungen rund um künstliche Intelligenz (KI) und maschinelles Lernen werden vielfach mit einschneidenden Veränderungen für Wirtschaft und Gesellschaft verbunden. Viele Menschen machen sich Sorgen um die Zukunft Deutschlands als führenden Technologie-Standort, um das zukünftige Rollenverhältnis von Mensch und KI und um ihren Job. Dr. Matthias Peissner vom Fraunhofer IAO interessiert sich vor allem für die Chancen positiver Veränderungen durch eine menschenzentrierte Entwicklung von KI-Technologien. Christian Luda von der DOAG-Redaktion sprach mit ihm.

Sie leiten am Fraunhofer IAO den Bereich Mensch-Technik-Interaktion. Woran forschen Sie aktuell?

Wir sind insgesamt über 40 Mitarbeiter aus ganz unterschiedlichen Fachrichtungen: Ingenieure, Informatiker, Designer, Wirtschaftswissenschaftler, Psychologen und ein paar Exoten, die an verschiedenen Fragestellungen der Mensch-Technik-Interaktion arbeiten. Wir möchten die Technik so gestalten, dass sie intuitiv bedienbar wird und die Menschen eine Beziehung zur Technik aufbauen, die über eine reine Werkzeug-Beziehung hinausgeht. Wir haben vier Teams. Das erste beschäftigt sich schwerpunktmäßig mit der Interaktion im Fahrzeug. Beim zweiten Team geht es um neue Technologien und die Gestaltung von Nutzerinteraktion – auch um die Frage, welche neuen Technologien aus dem Bereich der Erkennungstechnologie wie Sprach- und Gesenerkennung so eingesetzt werden können, dass die Interaktion attraktiver und effizienter, vielleicht spielerisch interessanter wird. Beim dritten Team geht es um die menschlichen Faktoren, also die User Experience. Wie können wir die Interaktion mit Technik zu einem positiven Erlebnis machen? Wir haben einige Methoden entwickelt, mit denen sich gezielt positive Nutzungserlebnisse in Produkte und Services im-

plementieren und dann auch nutzerzentrierte Innovationsprozesse durchführen lassen. In diesem Team betreiben wir seit einigen Jahren ein Neuro-Lab. Dort setzen wir Verfahren aus den Neurowissenschaften ein, um genauer an mentale und emotionale Prozesse während der Interaktion heranzukommen. Wir arbeiten dort zudem an Lösungen, Neurosignale im ganz normalen Leben, während der Arbeit oder im Fahrzeug live zu erfassen, sie zu klassifizieren und dann an ein System zurückzuspielen. Dieses soll auf emotionale Regungen in Echtzeit reagieren und sich entsprechend an Nutzerbedürfnisse oder verschiedene Situationen anpassen. Im vierten Team geht es um Identitätsmanagement. Wir betrachten unter anderem Rollen, differenzierte Zugänge zu IT-Systemen oder Personalisierungsfragen. Hier geht es uns nicht nur um technische, sondern auch um betriebswirtschaftliche Fragestellungen der IT-Sicherheit sowie um Akzeptanz und Usability-Aspekte. Denn das größte IT-Sicherheitsrisiko sind immer noch die Menschen selbst, die kreativ Sicherheitsmechanismen umgehen, die ihnen zu aufwendig sind oder die sie nicht richtig verstehen. Aktuell beschäftigen wir uns auch sehr viel mit der Frage, wie wir die Interaktion mit immer intelligenteren

Systemen gestalten können. Es gibt ja momentan ein riesengroßes Interesse für künstliche Intelligenz und Machine Learning. Ich bin jetzt seit 20 Jahren am Institut und seitdem beschäftige ich mich eigentlich immer auch mit Fragen der Interaktionen mit KI-Systemen. Da geht es um die Interaktion mit Sprachdialogsystemen oder Systemen, die Entscheidungsvorschläge machen, sowie um die Frage, wie man KI-Systeme so erklärbar machen kann, dass der Mensch versteht, auf welcher Datengrundlage ein Systemverhalten entsteht.

Wie weit ist die Technik im Bereich KI, auch in Bezug auf Anwendbarkeit und Akzeptanz?

Das lässt sich nicht einheitlich beantworten. In Smartphones steckt schon relativ viel KI. Manchmal fahre ich dann doch mit dem Auto zur Arbeit und wenn ich einsteige, sagt mir mein Smartphone, wie lange ich noch brauche. Ich bin vor Kurzem umgezogen und interessanterweise hat das System in dem Moment gewusst, wo ich hinfahren will. Es gibt ja auch immer wieder Geschichten von Leuten, die erstaunt sind, dass sie sich mit jemandem über ein Thema unterhalten und danach genau zu diesem Thema Werbung bekommen. Ich glaube, in diesem Bereich ist man schon

sehr stark an KI-Mechanismen oder -Funktionen gewöhnt. Man nimmt sie in Kauf, weil sie Komfort bieten – beispielsweise, wenn es um KI-Systeme zur Verbesserung von Smartphone-Kameras geht. Da sind wir schon sehr weit. Auf der anderen Seite gibt es in der Arbeitswelt noch sehr große Vorbehalte, vielleicht auch etwas unrealistische Erwartungen und Befürchtungen, wie sich die Arbeitswelt durch den Einsatz von KI entwickeln wird. Da gibt es noch sehr viel an Forschung und Entwicklung zu tun – insbesondere, wenn die Systeme komplexer werden und mehrere Module oder Stakeholder in einem großen Netzwerk von vernetzter Wertschöpfung zusammenspielen.

Sind deutsche Unternehmen im internationalen Vergleich zurückhaltender?

Das würde ich nicht unbedingt sagen. Ich glaube eher, dass wir in Deutschland sehr gute Jahre hatten. Viele Unternehmen sind in den letzten Jahrzehnten mit ihren traditionellen Geschäftsmodellen sehr gut gefahren. Daher haben sie teilweise noch nicht die Notwendigkeit gesehen, sich zu wandeln und neue Technologien auszuprobieren. Ich glaube aber schon, dass es sehr viel Interesse gibt. Es ist eher eine Frage der Priorisierung oder der wirtschaftlichen Potenziale, die man sich verspricht. Klar ist: Wenn eine Volkswirtschaft eher angreift und sich neue Felder sucht – wie wir das momentan in China sehen –, dann gehen die da mit einer riesigen Power rein und sagen: Das ist unser Markt von morgen. Wir müssen in Deutschland sehr gut aufpassen, dass wir diese Entwicklung nicht verschlafen. Wir sollten technische Möglichkeiten und neue Entwicklungen sehr sensibel aufnehmen. Gerade im industriellen Bereich, wo wir in Europa schon sehr viele Daten verfügbar haben, sollten wir vorpreschen und auch neue Dinge ausprobieren.

Welche Rolle spielt der Datenschutz? Inwiefern bremst er Innovation, etwa im Vergleich zu China?

Ich glaube, dass unser verantwortungsvoller und vorsichtiger Umgang mit personenbezogenen Daten ein Innovationstreiber sein kann und dass wir diese Karte im internationalen Wettbewerb tatsächlich auch spielen müssen. Wir sollten unsere westlichen Werte nicht einfach aufgeben und irgendwelchen radikalen Kapitalisten hinterherrennen. Trotzdem müssen wir umdenken. Wir sollten damit aufhören, die Sammlung von personenbezogenen Daten komplett zu vermeiden – in vielen Fällen besteht ja momentan die Strategie darin,

Daten aus verschiedenen Datenschutzgründen nicht zu erfassen. Wir müssen Strategien entwickeln, verantwortungsvoll mit personenbezogenen Daten umzugehen. Wir arbeiten heute schon in einem großen Netzwerk, im industriellen Internet der Dinge: Maschinen, Anlagen, Sensoren und Computer sind miteinander vernetzt, kommunizieren miteinander und optimieren Prozesse – und der Mensch ist eigentlich mittendrin. Wir können es uns meiner Meinung nach auf Dauer nicht leisten, diese Daten nicht zu erfassen und im Internet der Dinge zu integrieren. Heute ist gerade in Deutschland die Grundhaltung, dass diese personenbezogenen Daten möglichst nicht erfasst werden dürfen, da Arbeitgeber ansonsten Rückschlüsse auf die Leistungsbereitschaft ihrer Mitarbeiter ziehen könnten. Ich glaube, dass in der Zukunft die Generationen ein bisschen anders ticken werden. Die Arbeitnehmenden werden sagen: Schau mal, lieber Arbeitgeber, du hast die Möglichkeit zu erkennen, wenn ich eine Pause brauche, irgendwo Qualifizierungsbedarfe habe, mich bei der Arbeit langweile oder überfordert bin. Nutze diese Daten, um mir einen möglichst gesundheitsförderlichen Arbeitsplatz sicherzustellen, um mich weiterzuentwickeln und mein Wohlbefinden bei der Arbeit zu steigern. Ich erwarte also, dass sich diese Vermeidungshaltung in Richtung Anforderung entwickelt: Verwende die Daten, aber verwende sie zu meinem Vorteil. Da gibt es noch sehr viel zu tun. In der High-Level Expert Group der EU wurden jetzt entsprechende Guidelines zum verantwortungsvollen Umgang mit KI entwickelt. Ich glaube, dass es da ganz viel Neues – neue Strategien, neue Technologien und neue Ansätze – braucht, um personenbezogene Daten verantwortungsvoll zu nutzen.

Wie groß werden die Veränderungen sein, die uns in den nächsten Jahren erwarten?

Die Veränderung wird so groß sein, wie ich zumindest – ich bin jetzt 46 Jahre alt – noch keine Veränderung erlebt habe. Die Veränderungen haben wir in den letzten Jahren schon gemerkt und der Trend wird sich weiter fortsetzen. Es wird noch intensiver und schneller werden und jeder einzelne wird davon sehr stark betroffen sein. Das wird vielleicht im Umfeld der privaten Techniknutzung durch immer besser werdende Benutzerfreundlichkeit von entsprechend intelligenten Systemen noch einigermaßen gut abgefedert, aber aufseiten des Arbeitsmarkts sehe ich da eine ganz große Veränderung auf uns zukommen. Das Modell, dass die Menschen einen

Beruf lernen und dann ein Leben lang ausführen, werden wir radikal aufgeben müssen, hin zu einer Veränderungsbereitschaft und Lernbereitschaft jedes einzelnen. Das birgt die Gefahr, dass wir viele Menschen verlieren werden, die vielleicht nicht ausreichend motiviert sind oder nicht die ausreichenden kognitiven Fähigkeiten haben. Es ist ja auch anstrengend, sich weiterzuentwickeln. Wir müssen in den Unternehmen Mechanismen finden, um diese Entwicklung möglichst so zu gestalten, dass die Menschen mit Freude lernen und Veränderung nicht als Bedrohung wahrnehmen, sondern als Gewinn für jeden einzelnen. Die Anstrengung, Neues zu lernen, muss anders interpretiert werden. Dabei ist auch ein Wertewandel notwendig. Wir brauchen zudem entsprechende Technologien, die dabei helfen, benötigte Kompetenzen in den Unternehmen rechtzeitig aufzuspüren. Die Unternehmen müssen viel genauer wissen, was ihre Mitarbeiter eigentlich können – auch Dinge, die sie heute nicht unbedingt in ihrem täglichen Job machen. So könnte man den Menschen schnell und sozialverträglich zu einer Zukunft mit guten Perspektiven auf dem Arbeitsmarkt verhelfen.

Müsste man nicht heute schon beginnen, die Menschen zu schulen?

Das Problem ist, dass wir nicht wissen, welche Kompetenzen wir in fünf Jahren brauchen. Wir müssen heute vor allen Dingen beginnen, die Voraussetzungen für Veränderungsfähigkeit und -bereitschaft zu schaffen. In den Unternehmen muss man damit jetzt starten, um die Voraussetzungen für Kompetenzmanagement und Weiterqualifizierungsinfrastrukturen zu entwickeln. Das bedeutet auch einen Wertewandel, ein Mitnehmen der Unternehmen. Es gibt viele Kollegen in älteren Semestern, die nicht unbedingt sehen, warum sie sich jetzt noch einmal etwas komplett anderes aneignen sollten. Das ist eine riesige Gefahr.

Werden wir denn in Zukunft überhaupt genügend Jobs für alle haben?

Davon bin ich ziemlich fest überzeugt. Zum einen haben wir heute schon einen riesengroßen Fachkräftemangel in etlichen Bereichen. Zum anderen stehen wir vor einer ziemlich heftigen demografischen Entwicklung, die etliches an Härtefällen mit sich bringen wird. Außerdem ist es so: Wir haben in den letzten Jahrzehnten eigentlich mit jeder Automatisierungs- und Digitalisierungswelle gesehen, dass die absolute Anzahl der Arbeitsplätze nicht gesunken, sondern sogar eher gestiegen ist. Wir

erwarten mit den großen Digitalisierungs- und KI-Entwicklungen, dass dort natürlich etliche Jobs automatisiert werden können, aber der Bedarf an anderer Stelle wird dafür umso mehr wachsen. Ich denke, dass das aktuelle Thema Digitale Transformation weniger ein zahlenmäßiges Thema ist, sondern eine Frage, wie wir arbeiten, also eher eine qualitative Frage.

Entstehen dank Bedienhilfen, an denen Sie ja forschen, auch neue einfache Jobs?

Wir versprechen uns, dass wir durch die Technik, an der wir arbeiten, Möglichkeiten der Teilhabe für möglichst alle schaffen, dass wir beispielweise ältere Menschen länger in der Arbeit halten und Menschen mit anderem sprachlichen Hintergrund leichter ins Arbeitsleben integrieren können. Da gibt es etliches an Bemühungen und ich denke, da werden wir auch vorankommen. Aber tatsächlich gehen viele Prognosen dahin, dass die Arbeit weniger ausführende, sondern mehr überwachende und planerische Tätigkeiten haben wird. Das sind natürlich in der Regel auch verantwortungsvollere Tätigkeiten, vielleicht auch Tätigkeiten, die ein höheres Qualifizierungsniveau brauchen. Im Bereich der weniger qualifizierten, einfacheren Berufe sehe ich schon ein Problem, dort weiterhin genügend Jobs zu haben. Wenn man sich beispielsweise den Gesundheitsbereich oder die Pflege anschaut, gibt es einen großen Bedarf an Jobs, die man vielleicht als einfachere Tätigkeiten bezeichnen kann. Aber man darf sich auch nichts vormachen: Diese Tätigkeiten erfordern ein ganz großes Maß an Empathie und Organisation. Die Unterteilung in einfachere und anspruchsvollere Tätigkeiten ist sehr schwierig. Ist einfach immer das, was man auch automatisieren kann? Ich glaube nicht. Ich glaube, dass auch viele Expertentätigkeiten durch KI ersetzt werden können. Heute schon entstehen im Finanzbereich sehr viele Expertensysteme, die möglicherweise auch sehr gut ausgebildete Berufe gefährden. Insofern ist die Diskussion immer ein bisschen schwierig – was ist einfach, was ist anspruchsvoll? Ich denke aber, dass wir einen ganz großen Bedarf haben werden an Kompetenzen, die eben in absehbarer Zeit durch KI nicht übernommen werden können: Kreativität, Empathie, unternehmerisches Denken, intrinsische Motivation und soziale Kommunikationsfähigkeit. Insofern würde ich eher nach Kompetenzprofilen schauen: Für Tätigkeiten, die ein großes Potenzial der Substituierung durch Technik haben, wird sicher ein bisschen weniger Bedarf bestehen. Dahingegen sind andere Tätigkeiten – etwa im Handwerk

mit seinen unvorhersehbaren Rahmenbedingungen – auf absehbare Zeit nicht so schnell automatisierbar.

Was sind aktuell die spannendsten Entwicklungen im Bereich Bedienhilfen?

Eine spannende Entwicklung ist „Training on the Job“: Hier werden Qualifizierungsangebote direkt in die Arbeitsaufgabe integriert, so dass das Lernen keinen Schulbankaufwand bedeutet, sondern wirklich Spaß macht. Die Leute können schon beim Lernen produktiv sein, sich Lernerfolge direkt merken und nebenbei neue Qualifikationen aneignen. Ein anderer Bereich, in dem ich ebenfalls einen sehr großen Bedarf sehe, ist die Frage, wo die Inhalte für die Bedienhilfen eigentlich herkommen. Wir sehen heute sehr viele Experimentierfälle mit Augmented Reality, teilweise schon im Realbetrieb, etwa AR-Brillen. Der Einsatz scheitert oft daran, dass die Inhalte bereitgestellt werden müssen. Das ist zum einen eine Frage der Content-Management-Systeme, also wie solche Inhalte ohne große technische Vorerfahrungen bereitgestellt werden können. Zum anderen ist es aber auch ein inhaltliches Thema. Hier sehe ich einen großen Bedarf im Zusammenhang mit dem demografischen Wandel: Wir werden in den nächsten Jahren sehr viel Fachexpertise in den Ruhestand verlieren. Eine riesige Herausforderung besteht darin, Mechanismen zu entwickeln, um ein Kompetenz- oder Wissensmanagement in den Unternehmen zu implementieren. Das Wissen, das die in den Ruhestand scheidenden Mitarbeiter haben, muss in die Arbeitsumgebung transferiert und dort auch an andere, möglicherweise neue Mitarbeiter weitergegeben werden. Da haben wir schon etliches an Forschungsarbeit geleistet. Dazu zählt die Frage, wie anspruchsvolle Tätigkeiten dokumentiert werden können, ohne dass ein Mitarbeiter hierfür viel Zeit aufwenden muss. Ein weiteres großes Thema ist die Frage, wie Mitarbeiter motiviert werden können, dass sie überhaupt bereit sind, ihr Wissen zu teilen. Das hat sehr viel mit der Unternehmenskultur und dem Maß der Identifikation mit den Unternehmenszielen zu tun. Außerdem setzen wir uns mit der Frage auseinander, wie sich von Mitarbeitern erstellte Inhalte validieren, strukturieren und dann auch aufbereiten lassen, sodass daraus wirklich gute Bedienhilfen oder Anleitungen werden.

Wie werden sich unsere Arbeitswelten ändern?

Wir werden ein viel breiteres Spektrum an verschiedenen Arten der Arbeit haben. Man wird nicht generell sagen können: Die Flexibilisie-

rung der Arbeit führt generell zu einem Work-Life-Blending, dass jeder, egal wo er ist, mal ein bisschen arbeitet, mal ein bisschen lebt, sich ein bisschen um die Familie kümmert, und das alles ineinanderfließt. Ich denke, das ist vielleicht für manche Berufe und für manche individuellen Arbeitnehmenden ganz ok, andere werden ihren Arbeitsalltag aber anders, vielleicht eher etwas konventioneller organisieren. Mir zum Beispiel ist es ganz wichtig, da auch eine möglichst klare Trennung, auch eine örtliche Trennung zu haben, weil man dann vielleicht auch nicht so gut abschalten kann. Aus meiner Sicht wird diese Flexibilisierungsdiskussion in Bezug auf Arbeitsorte und Arbeitszeiten dahingehen, dass wir in der Zukunft eine viel größere Heterogenität haben werden. Wir werden von unseren Mitarbeitenden nicht mehr erwarten, dass sie jeden Tag immer an einem bestimmten Ort sind, sondern es wird eine ganz große Bandbreite geben. Wir sehen ja auch schon die Themen Home-Office oder mobiles Arbeiten nicht mehr nur in den klassischen Wissensarbeitsplätzen, sondern auch in Bereichen wie Engineering und Produktion. Vielleicht können wir in der Zukunft mobil Produktionsanlagen steuern oder überwachen, dafür gibt es ja heute schon Ansätze.

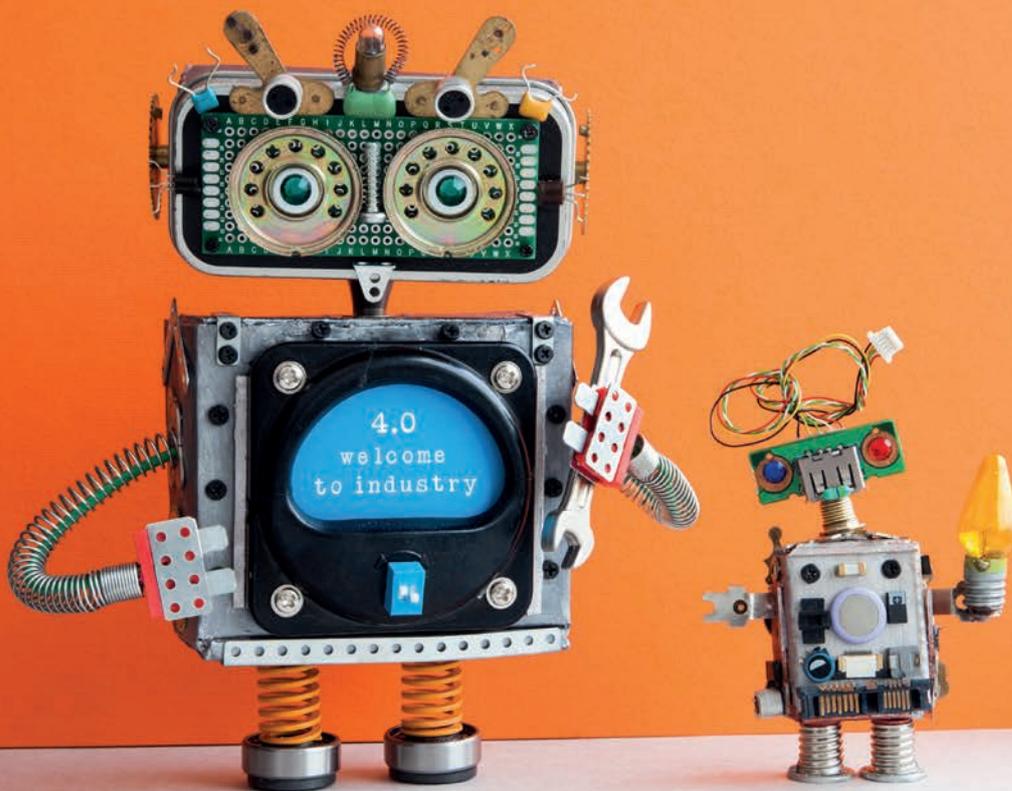
Herr Dr. Peissner, vielen Dank für das Gespräch.



Dr. Matthias Peissner

matthias.peissner@iao.fraunhofer.de

Dr. Matthias Peissner ist Institutsdirektor und leitet den Forschungsbereich Mensch-Technik-Interaktion am Fraunhofer IAO. Seine interdisziplinären Teams arbeiten an Lösungen, die ein effizientes Zusammenspiel von Mensch und intelligenter Technik ermöglichen. Schwerpunkte seiner Arbeit sind anpassungsfähige Systeme, zukünftige Arbeitsumgebungen sowie die Gestaltung positiver Nutzungserlebnisse. In ISO TC159 SC4 WG5 ist er in der internationalen Normung von Benutzungsschnittstellen tätig. Zur menschengerechten Gestaltung von KI-Systemen engagiert er sich unter anderem in der Plattform Lernende Systeme und der Fokusgruppe »KI in der Arbeitswelt« des Bundesministeriums für Arbeit und Soziales (BMAS). Er ist einer der Gründer und langjähriges Vorstandsmitglied der »German Usability Professionals Association e.V.« (G-UPA). An der Hochschule für Gestaltung Schwäbisch-Gmünd lehrt Dr. Peissner zur menschenzentrierten Gestaltung interaktiver Systeme.



Smarte Anlagenüberwachung für die Industrie 4.0

Daniel Silter, PROMATIS Gruppe

Längst ist das „Internet of Things“ (IoT) vom Trend zum Mainstream geworden. Was in Privathaushalten in Form von „Smart Devices“ erfolgreich wurde, ist nun auch in den Unternehmen nicht mehr nur eine abstrakte technische Möglichkeit. Konkrete Anwendungsfälle versprechen einen echten Mehrwert für die Unternehmen. Auch Oracle positioniert sich mit eigenen IoT-Applikationen im Markt, wie beispielsweise der Asset Monitoring Cloud. Doch wie kann es gelingen, dieses Versprechen einzulösen?

Digitalisierung ist für viele Unternehmen ein strategisches Ziel geworden. Dabei gilt es, diesen Gedanken nicht aus reinem Selbstzweck heraus zu erfüllen. Die Umsetzung eines Ziels und dessen Erfolg muss zum einen messbar sein und zum anderen einen unternehmerischen Mehrwert generieren. Digitalisierung um der Digitalisierung willen kann sogar kontraproduktiv sein. Es werden Ressourcen verbraucht, die anderswo besser genutzt werden könnten,

und schlimmstenfalls wird der Blick auf die wirklichen Potenziale der Digitalisierung verbaut. Dies gilt insbesondere für einen wesentlichen Baustein der Digitalisierungsagenda: das „Internet of Things“ (IoT).

Von der technischen Möglichkeit zum konkreten Use Case

Die Umsetzung eines IoT Use Case erfordert nicht nur die Investition in die klassische Implementierung, sondern auch in die

Infrastruktur der entsprechenden „intelligenten“ Geräte. Wo also liegen die Potenziale von IoT-Anwendungen, um solch eine Investition effektiv in eine Digitalisierungsstrategie einbringen zu können?

Nach einer aktuellen IoT-Studie von Telefonica mit über 300 Unternehmen, die IoT-Anwendungsfälle planen oder durchführen, ist die Qualitätskontrolle der derzeit meistgenannte Anwendungsfall für die Unternehmen – verliert aber in Zukunft relativ

an Bedeutung. Smart Grid (also intelligente Verbrauchssteuerung) und Predictive Maintenance nehmen zukünftig am stärksten an Bedeutung zu (siehe [1]). Diese Ergebnisse deuten einen Wandel an, vom reinen Erfassen und Auswerten der Daten hin zu realen Anwendungsfällen.

Zwar ist das Ziel, die Qualitätskontrolle zu verbessern, durchaus mit einem unternehmerischen Mehrwert verbunden. Aufgrund seiner Allgemeingültigkeit gilt es allenfalls als Platzhalter für einen IoT-Anwendungsfall (sofern das Hauptgeschäftsfeld nicht auf die Qualitätssicherung zielt). Letztendlich geht es hier über das Sammeln und Auswerten von Daten nicht hinaus.

Erfreulicherweise gewinnen mit den Themen Smart Grid und Predictive Maintenance zwei Anwendungsfälle an Bedeutung, in denen IoT-Technologien einen direkten Mehrwert erzeugen können. Nicht zufällig liegen diese IoT-Anwendungen ganz weit vorn im Trend. So sind die mit Abstand meistgenannten Erfolgskriterien von IoT-Projekten: Produktivitätssteigerung, Kostensenkung, Steigerung von Umsätzen und geringere Ausfallzeiten/höhere Auslastung (vgl. [2]). Gefragt nach dem konkret erzielten Nutzen, werden bei den Unternehmen, die überhaupt eine Auswirkung bemessen konnten, die genannten Erfolgskriterien jedoch nur zu je 19% bis 29% erfüllt (siehe Abbildung 1). In [3] wurde die Kos-

tenreduktion etwa lediglich von 27,3% der Unternehmen genannt – wohlgemerkt von den Unternehmen, die überhaupt einen Erfolg messen konnten.

Der Grund hierfür ist in der wenig zielgerichteten Anwendung von IoT-Technologien zu vermuten. Im Gegensatz dazu wird ein konkreter IoT-Einsatz zur Optimierung des Energieverbrauchs die Kosten ebenso nachweislich senken, wie durch einen IoT-gestützten präventiven Wartungsplan die Ausfallzeit von störanfälligen Anlagen gesenkt wird.

IoT in der Unternehmensarchitektur

Eine gewinnbringende IoT-Strategie erfordert demnach eine Übereinstimmung der technischen Möglichkeit mit einem konkreten Anwendungsfall. Der wesentliche Faktor für das Gelingen besteht darin, dass diese Bewertung mit der Identifizierung und Betrachtung der Anwendungsfälle startet. Das heißt, abgeleitet von dem Geschäftsprozess werden die Anwendungsfälle identifiziert, in denen durch IoT-Technologie ein Mehrwert erzielt werden kann. Es geht explizit nicht darum zu prüfen, wie eine bestimmte Technologie oder Hardware in den bestehenden Geschäftsprozess integriert werden kann, um daraus dann einen theoretischen Mehrwert zu erzielen.

Zur Veranschaulichung dieser Methode eignet sich das RAMI-4.0-Modell (Reference

Architecture Model Industrie 4.0, visualisiert in Abbildung 2). Zwar zeigt das Modell nicht unmittelbar die Methodik für eine IoT-Implementierung auf, es lässt sich jedoch eine geeignete Vorgehensweise bei der Einbindung von IoT-Technologien in das Unternehmen daraus ableiten.

Das RAMI 4.0 beschreibt strukturiert die wesentlichen Elemente eines IoT-Assets mithilfe eines aus drei Achsen bestehenden Schichtenmodells. Die Kombination der drei Achsen beschreibt die jeweilige Ausprägung der Aspekte „Architektur“, „Lebenszyklus“ und „Hierarchie“ für ein Asset (siehe [4]). In der vertikalen Architektur-Achse ist das Asset als Bestandteil der Unternehmensarchitektur dargestellt und in Bezug zu den anderen Architektur-Komponenten auf unterster Ebene eingeordnet. Auf oberster Ebene, dem Business Layer, ist der Anwendungsfall verortet. Im Detail, laut RAMI 4.0, sind in dieser Ebene neben den „monetären Bedingungen“ unter anderem auch die „Abbildung von Geschäftsmodellen und den sich daraus ergebenden Gesamtprozessen“ in [5] genannt.

Es ist naheliegend, dass diese Architektur für ein existierendes Geschäftsmodell nicht von der unteren Ebene heraus aufgebaut werden kann. Ein zielgerichteter unternehmerischer Nutzen oder ein Anwendungsfall wären kaum zu realisieren, allein dadurch, dass Daten und Funktionen zur Verfügung

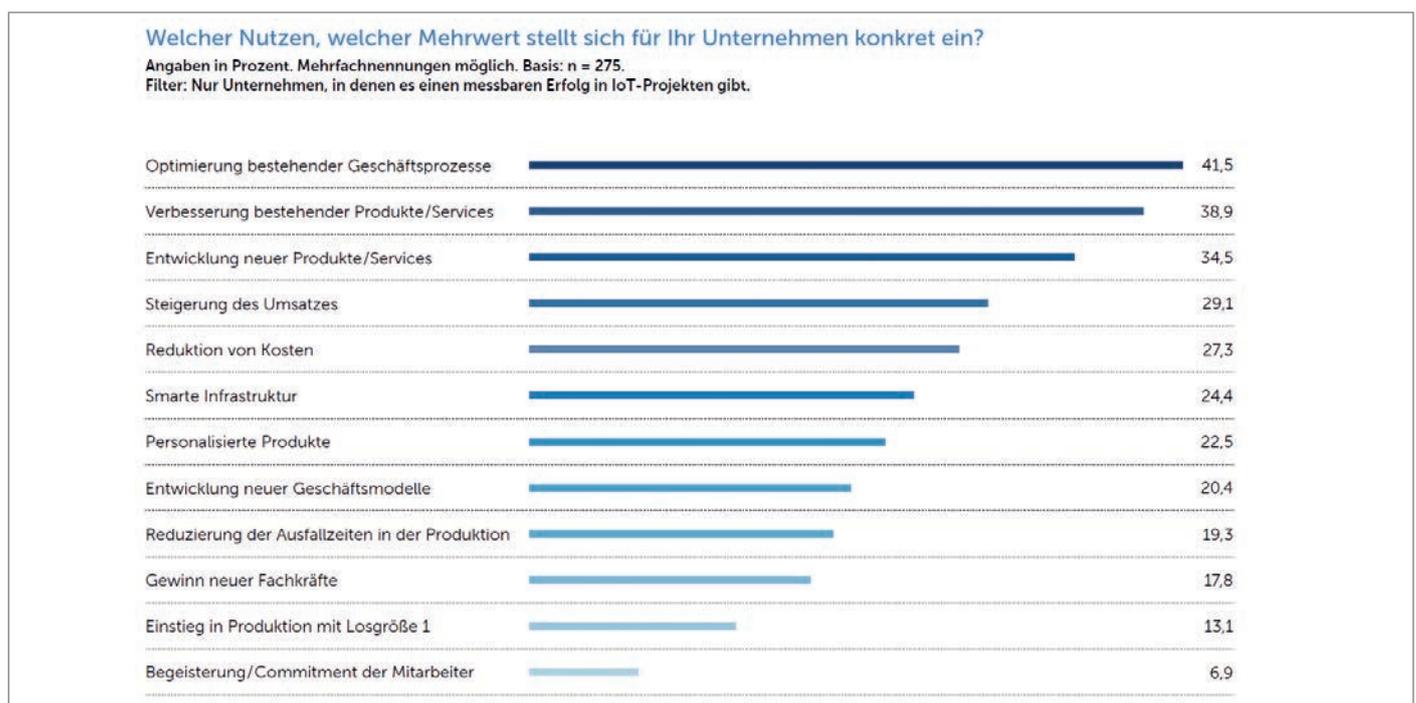


Abbildung 1: Studie zum Mehrwert von IoT-Projekten (Quelle: Telefonica-IoT-Studie 2020, Seite 16)

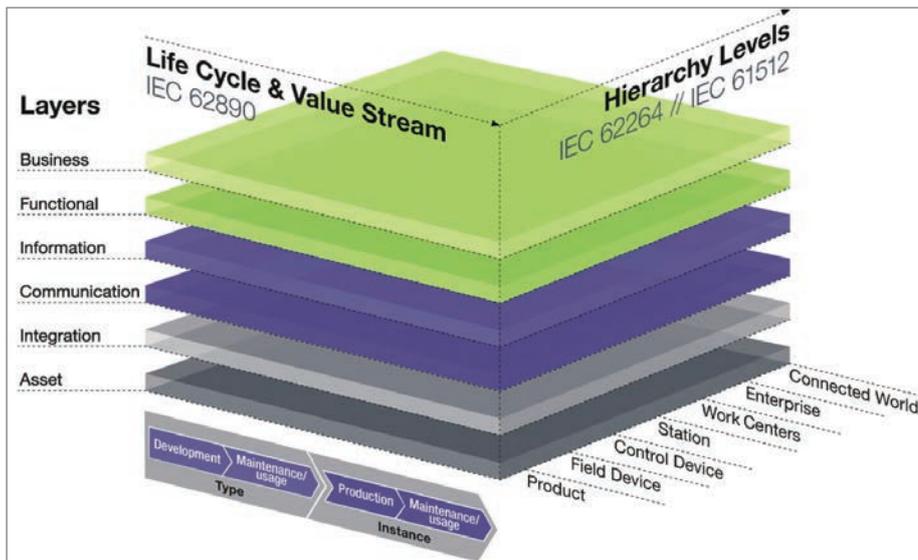


Abbildung 2: RAMI 4.0 Modell (Quelle: DIN Spezifikation 91345, 2016, Beuth Publishing DIN, Seite 19)

stehen. Ebenso wenig ist es sinnvoll, einen existierenden Geschäftsprozess mit einer existierenden Anlage zu verbinden, nur weil es technisch möglich ist. Diese beiden genannten Herangehensweisen sind eher im Kontext der Produktentwicklung denkbar, das heißt für die Entwicklung von IoT-Produkten. Für den Einsatz von IoT-Technologien sollten sie allerdings kein Maßstab sein. Wie aber ist die Identifizierung eines

tatsächlichen IoT-Anwendungsfalles möglich und welche Applikationen sollten genutzt werden?

Anwendungsfälle für die Asset Monitoring Cloud

Der Schlüssel für eine erfolgreiche IoT-Strategie liegt also zunächst einmal in dem Erkennen der Anwendungsfälle. Das erfordert ein detailliertes Wissen über die eigenen Ge-

schäftsprozesse. Im RAMI-4.0-Modell werden für den Ausgangspunkt des Business Layer im Einzelnen die Elemente „Orchestrierung von Diensten des Functional Layer“, „Verbindungselement zwischen verschiedenen Geschäftsprozessen“ und „Empfang von Ereignissen für die Weiterschaltung eines Geschäftsprozesses“ referenziert (vgl. etwa [5]). Der Geschäftsprozess muss also im Detail mit seinen Ereignissen, den Teilprozessen und deren Verbindungen dokumentiert sein. Ein solch strukturiertes Geschäftsprozess-Modell enthält zudem idealerweise die Weiterschaltungslogik für die einzelnen Ereignisse sowie eine übergeordnete Orchestrierung aller Teilprozesse. Durch Priorisierung einzelner Ereignisse und Festlegen von Kennzahlen lassen sich anschließend Prozess-Schritte herausfiltern, die einen Bottleneck bilden. Über die Abhängigkeiten der einzelnen Ereignisse und Prozesse wird eine ganzheitliche Bewertung abgeleitet. Diese Beurteilung sollte nach klaren, unternehmerischen Zielen stattfinden, wie etwa der eingangs angeführten Reduzierung von Kosten oder Ausfallzeiten. Durch eine Verknüpfung des Geschäftsprozessmodells mit entsprechenden Kennzahlen und Wahrscheinlichkeiten wird so eine Bewertung möglich

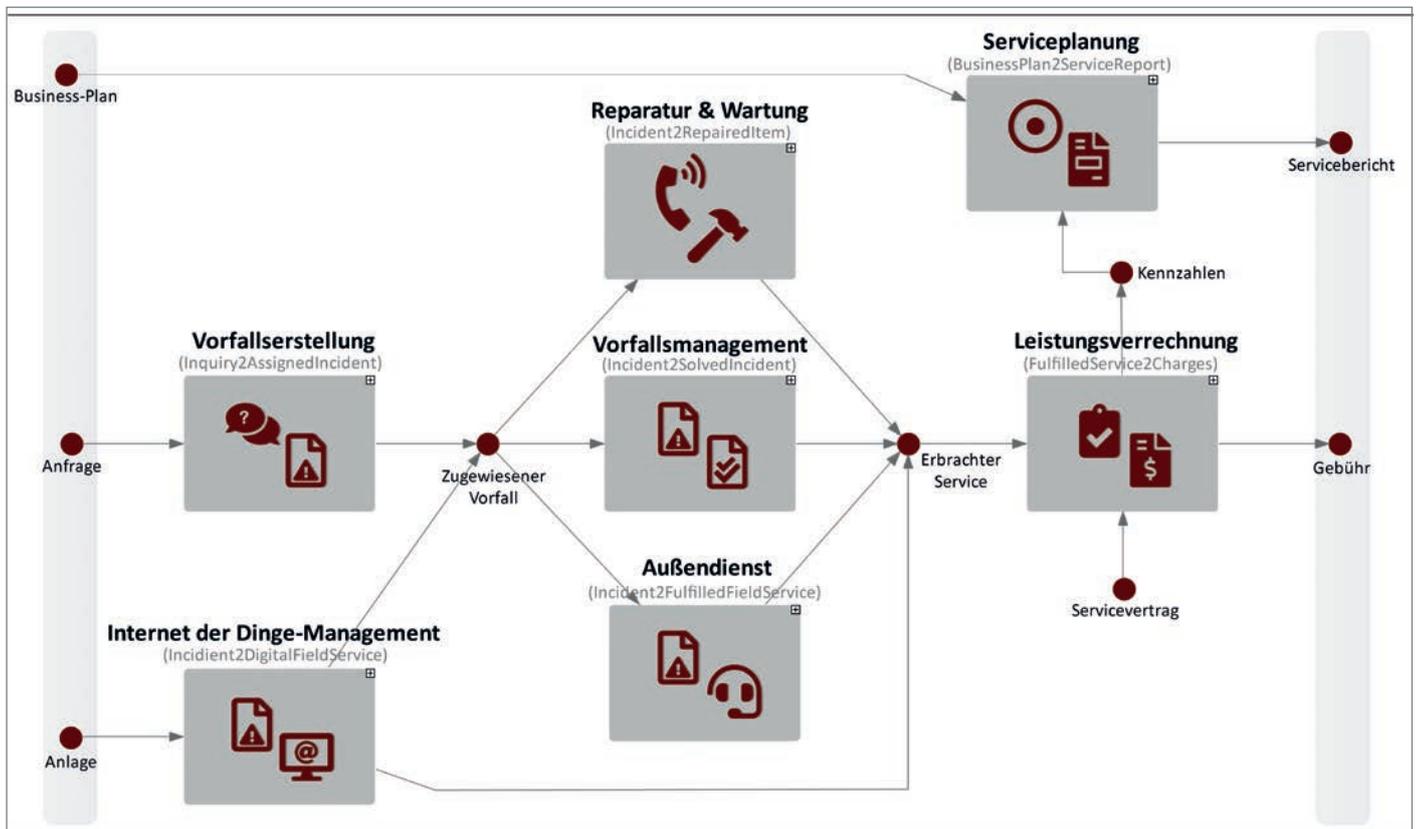


Abbildung 3: Best-Practice-Service-Geschäftsprozess (Quelle: Service Best Practice Modell, 2020, Horus software GmbH)

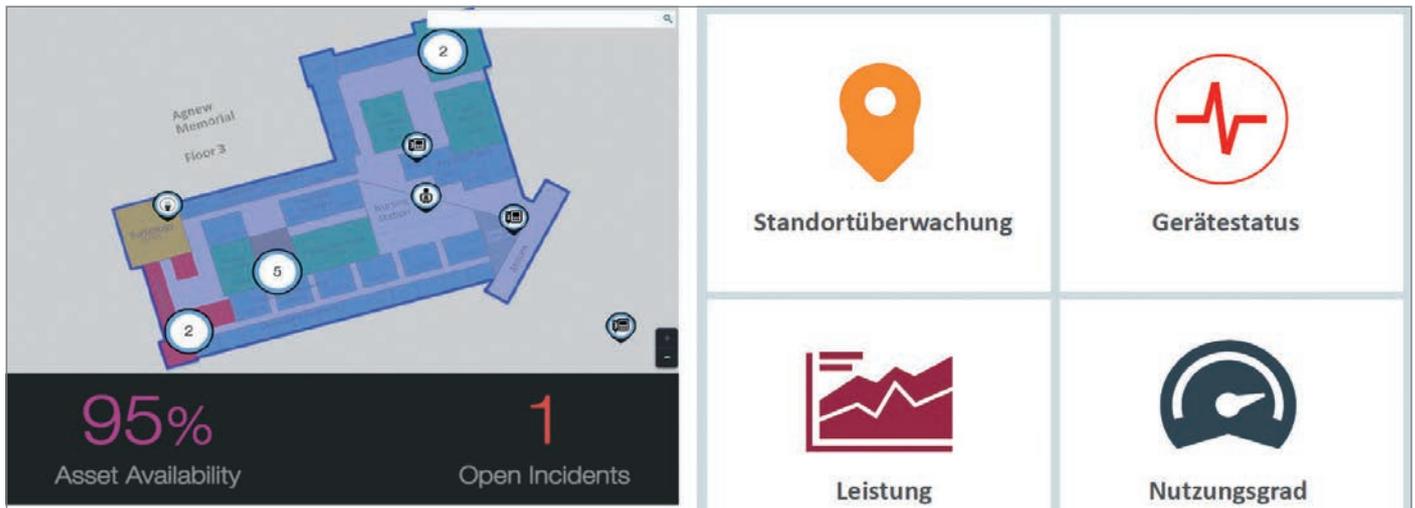


Abbildung 4: Oracle Asset Monitoring Cloud (Quelle: Präsentation „Das Internet der Dinge als Cloud-Anwendung in der modernen Supply Chain“, 2020, Oracle)

gemacht und erlaubt gegebenenfalls sogar – je nach Modellierungs-Werkzeug – eine Simulation verschiedener Szenarien. Auf die Art werden Rückschlüsse gezogen, für welche Prozesse und Ereignisse IoT-Technologien einen Mehrwert generieren oder überhaupt relevant sind. Bewusst ist anfangs bei der Nennung IoT-relevanter Anwendungsfälle von „störanfälligen“ Anlagen die Rede. Eine Maschine, die in den letzten zehn Jahren nie ausgefallen ist und keiner nennenswerten Wartung bedarf, wird durch den Einsatz von IoT-Technologie nicht leistungsfähiger. Ein Asset, dessen präziser Standort keine Rolle spielt und das aller Wahrscheinlichkeit nach nie gestohlen wird, muss nicht auf eine Gebietsüberschreitung hin überprüft werden. Auch wenn dies mit IoT-Technik leicht möglich wäre. Ein bestimmter Sensor oder eine bestimmte Applikation, und seien sie technisch noch so ausgereift, bieten keinen Mehrwert, wenn der Anwendungsfall nicht gegeben ist. Anders gesagt: Ein relevanter Anwendungsfall mit Potenzial ist die Basis für ein erfolgreiches und auch weitgehend akzeptiertes IoT-Projekt!

Hierbei treten oftmals die internen Service-Prozesse auf den Aktionsplan, da sich hier sowohl der konkrete Handlungsbedarf sehr gut aus dem internen Know-how ableiten lässt als auch ein realer Kostenvorteil möglich ist. Die anfangs erwähnten IoT-Trends Smart Grid und Predictive Maintenance finden beispielsweise in der Gebäudeautomation und Instandhaltung ihre natürlichen Anwendungsfälle. Selbstverständlich gilt das bei entsprechender Ver-

netzung mit dem Kunden auch für externe Anbieter solcher Dienstleistungen.

Die identifizierten IoT-Anwendungsfälle werden abschließend mit ihren jeweiligen Anknüpfungspunkten in das Geschäftsprozessmodell eingebettet. In *Abbildung 3* ist das für einen allgemeinen Service-Anbieter exemplarisch dargestellt.

Nach der Betrachtung des Business Layer erfolgt die Analyse des Functional Layer, also der Applikation, die den Anwendungsfall abbildet. Gegebenenfalls getroffene Annahmen und definierte Anforderungen sind in der Applikation abzubilden und in der Bewertung der Geschäftsprozesse zu berücksichtigen.

Grundsätzlich besteht bei der Auswahl der Applikation die Wahl zwischen Standard- und Individualsoftware. Während für Spezial-Anwendungsfälle eine Individuallösung sinnvoll sein kann, ist für Prozesse wie etwa die Gebäudeautomation und oder Instandhaltung eine Cloud-basierte Standardapplikation die günstigere und auch skalierbare Wahl. Es wird zum einen die RAMI-4.0-Architektur bis zur Kommunikationsebene durch standardisierte Methodiken abgedeckt. Zum anderen basiert die Applikation auf modernen, marktüblichen Technologien. So gehören der Telefonica-IoT-Studie zufolge – neben der IoT-Hardware und den Netzwerktechnologien – die IoT-Plattformen, Cloud-Computing und Analytics-Funktionen zu den fünf meistgenannten unverzichtbaren IoT-Technologien für Kunden (*siehe [6]*).

Die Asset Monitoring Cloud bringt die drei zuletzt genannten Technologien im

Lösungspaket mit. Außerdem kann auf ein umfangreiches Lösungs-Know-how sowie Best-Practice-Ansätze für die verschiedenen Anwendungsfälle zurückgegriffen werden. Die Szenarien erstrecken sich von Anlagen in der Fertigung und dem Bergbau bis hin zu Krankenhäusern und der bereits erwähnten Gebäudeautomation. Ein weiterer Vorteil ist die Integration mit anderen Systemen. Was in der Geschäftsprozess-Sicht bereits prozessübergreifend gemacht wurde, wird in der Applikation durch vordefinierte Integrationspunkte, die sich an den Standardgeschäftsprozessen ausrichten, ebenso ermöglicht. Damit kann die Asset Monitoring Cloud leicht in eine bestehende Systemlandschaft eingebunden werden oder zu einem späteren Zeitpunkt um weitere Cloud-Module ergänzt werden. Auf die Art können zum Beispiel Vorfälle an einer Anlage oder errechnete Wartungszyklen automatisch als Wartungsauftrag beziehungsweise Wartungsplan an das Instandhaltungsmodul übergeben und dort abgewickelt werden. Die Integration zum ERP (Enterprise Resource Planning) triggert anschließend die Leistungserfassung und -abrechnung. Das ist IoT mit echtem Mehrwert!

Fazit

Unternehmen stehen vor der Herausforderung, Digitalisierungsstrategien zu entwickeln und umzusetzen. Der Einsatz von IoT-Technologie spielt dabei eine wesentliche Rolle. Damit eine solche Strategie nicht nur zum Imagegewinn beiträgt, aber schlimmstenfalls zusätzliche Kosten verur-

sacht, sollten folgende Faktoren beachtet werden:

Die Identifizierung von Anwendungsfällen – für die IoT-Technologie einen echten unternehmerischen Mehrwert erzeugt – ist der Ausgangspunkt für eine erfolgreiche Strategie. Um diese Anwendungsfälle sichtbar zu machen und bewerten zu können, ist eine detaillierte Kenntnis und strukturelle Aufbereitung der Geschäftsprozesse elementar. Insbesondere Anwendungsfälle in der Gebäudeautomation und Instandhaltung erweisen sich als ausgesprochen kompatibel, um diesen Mehrwert mittels IoT-Technologie zu erzielen. Die Asset Monitoring Cloud ist hierfür ein leistungsfähiges

Standardprodukt, basierend auf modernen Technologien und vielfältigen Integrationsmöglichkeiten.

Quellen

- [1] Telefonica-IoT-Studie 2020, <https://iot.telefonica.de/download-iot-studie-2020>, Seite 9
- [2] Telefonica-IoT-Studie 2020, <https://iot.telefonica.de/download-iot-studie-2020>, Seite 15
- [3] Telefonica-IoT-Studie 2020, <https://iot.telefonica.de/download-iot-studie-2020>, Seite 16
- [4] DIN SPEC 91345 2016, Beuth Publishing DIN, Berlin, Seite 19
- [5] DIN SPEC 91345 2016, Beuth Publishing DIN, Berlin, Seite 20
- [6] Telefonica-IoT-Studie 2020, <https://iot.telefonica.de/download-iot-studie-2020>, Seite 23



Daniel Silter
daniel.silter@promatis.de

Daniel Silter ist Senior Principal Consultant bei der PROMATIS Gruppe. Seine fachlichen Schwerpunkte sind Field Service, Service-Logistik und IoT.



Abbildung 1: Value Chain eines Metall-produzierenden Unternehmens (© Falk Wolsky)

Vom Sensor zum Prozess

Falk Wolsky, Chief Innovation Officer, Innogate Technologies

Während viele von digitaler Transformation reden und schillernde Pilotprojekte auf den Weg bringen, gibt es für die meisten Unternehmen noch viel zu tun. Am Beispiel eines der größten Energie-produzierenden Unternehmen der Ukraine möchte ich zeigen, welche Ideen, Maßnahmen und Technologien rasch zu messbaren Ergebnissen und zur erwünschten Steigerung der Produktivität führen.

Nicht ohne Grund halte ich eine Vortragsreihe zum Thema „Hype und Realität in Blockchain, Big Data und Machine Learning“. Es ist zu viel *Buzzwording* und oft zu wenig

konkretes Verständnis für Kernthemen der Digitalisierung vorhanden. Zugleich drängt uns die Zeit, denn die Bedeutung der Digitalisierung zeigt sich gerade auch in schwierigen

Zeiten wie der aktuellen Corona-Krise. In diesem Artikel möchte ich auf technischer Ebene einen Weg mit Ihnen gehen. Wir schauen uns gemeinsam an, welches

die Fundamente für digitale Transformation sind und wie Mehrwerte aus Sensorsignalen (IoT) anhand von IT-Maßnahmen und passender Infrastruktur realisiert werden können.

Ganz konkret lässt sich ein gutes Fallbeispiel anhand eines Projektes für eines der größten Strom-produzierenden Unternehmen der Ukraine zeigen. Sicher nicht für alle Unternehmen eins zu eins übertragbar, zeigt das Vorhaben dennoch eine gut verwendbare Schablone zum Themenbereich der digitalen Transformation. Denn der Weg vom Papier-Prozess am Werk zu durchgängig digitalen und automatisierten Prozessen ist lohnenswert. So kann zum Beispiel der Bezug von Ersatzteilen für Maschinen dank Predictive Maintenance komplett automatisiert werden.

Netzwerkabdeckung als erste wichtige Maßnahme der Digitalisierung

Man möchte schmunzeln, doch der erste Schritt, noch vor den Sensoren, ist auch für Unternehmen eine gute Netzwerk-Abdeckung: WiFi, LPWAN, ZigBee. Die Netzwerke für Sensoren, Steuerung und Mitarbeiter sind getrennt, müssen aber an allen Arbeitsplätzen im Betriebsgelände in guter Abdeckung vorhanden sein. Gute Netzwerk-Abdeckung klingt zunächst selbstverständlich. Zugleich wird einem beim Anblick eines Bergbauplatzes beziehungsweise in verstreuten Betriebsstädten in der Realität rasch klar, dass dies nicht selbstverständlich ist und der Ausbau der Netzwerk-Abdeckung zur ersten wichtigen Infrastrukturmaßnahme wird.

Sensorsignale und Sensing Strategy

Während in der Investitionsgüterindustrie durch den hohen Automatisierungsgrad bereits eine gute Sensordichte herrscht, ist in anderen Industrien noch entsprechender Nachholbedarf zu bemerken. In *Abbildung 1* ist zum Beispiel die Value Chain eines Metall-produzierenden Unternehmens dargestellt. Es wird rasch klar, dass es eine Fülle von Stellen gibt, an denen Sensoren wertvolle Daten liefern können.

Eine Sensing Strategy umfasst zum Beispiel folgende Arten von Sensorik:

- Kameras und Lidar-Sensoren
- Optische Sensoren, Lichtsensoren
- Positions-, Geo-Positions- und Abstandssensoren
- Geschwindigkeits-, Beschleunigungs- und Bewegungssensoren
- Temperatursensoren
- Feuchtigkeitssensoren
- Geräusch- und Vibrationssensoren
- Gas- und chemische Sensoren
- Druck-, Zug- und Drehmomentsensoren

Die Sensoren haben unterschiedliche Charakteristika, die beim Sammeln und Verarbeiten der Daten relevant sind. Sensoren haben in aller Regel die Charakteristik, Zeitreihen-Daten zu erzeugen. Von kleinen stündlichen Datenpaketen (*siehe Listing 1*) bis hin zum konstanten Datenstrom von Kameras.

Der Strom von Daten wird anhand verschiedener Strategien bereits an der Quelle (Edge) beziehungsweise durch gestufte lokale Vorverarbeitung reduziert, sodass nur relevante Signale in die Geschäftsanwendungen gelangen.

Die Zeitreihen-Daten sind in ihrer Natur und Menge gänzlich unterschiedlich zu Geschäfts-Transaktionsdaten. Dies ist bei der Planung der IT-Infrastruktur von großer Bedeutung. Schauen wir uns dies nochmal genauer am Beispiel des Temperatursensors an. Dieser sendet zum Beispiel alle 30 Sekunden ein Datenpaket (*siehe Listing 2*).

Wir werden auf dieses Beispiel im Verlauf dieses Artikels nochmals zurückkommen. An dieser Stelle ist es wichtig zu verinnerlichen, dass die Vielzahl der Sensoren ausschlaggebend für die Planung der IT-Infrastruktur ist. Denn wenn ein Sensor diese Datenpakete sendet, erscheint es uns einfach und wir erwarten, sie in normalen Datenbanken ohne Mühe speichern zu können. In der Realität treten für Unternehmen jedoch Tausende, Zehntausende und auch Hunderttausende Sensoren gleichzeitig auf. Es entsteht damit eine Herausforderung in zwei Dimensionen: zum einen die Bandbreite und zum anderen

das Herausfiltern der wichtigen Signale, die für Geschäftsprozesse relevant sind.

In einem Projekt zur Einführung von Smart-Metern (die neue Generation von digitalen Strom-Messgeräten) konnte ich diesen nahezu typischen Verlauf der Projekte gut beobachten. Ein großes Consulting-Unternehmen sorgte für ein Team, welches das Microservice-Backend und die Datenbank für Smart-Meter bereitstellen sollte. Es war „sofort klar“, dass eine moderne „IoT-Datenbank“ zum Einsatz kommen sollte. Einen Monat später wurde dieser Ansatz zu den Akten gelegt. Man fand heraus, dass die IoT-Datenbank der Bandbreite des eingehenden Datenstroms nicht gewachsen war. Danach wurde *Cosmo DB* auf *Azure* getestet. Auch dieser Ansatz musste, wieder nach einem Monat, als vergebens betrachtet werden. Man fand am Ende heraus, dass Managed MSSQL-Server auf *Azure* zumindest in der Lage waren, die Bandbreite des Test-Szenarios mit nur einhunderttausend Smart-Metern (MvP) abzudecken. Der Elektrizitätsanbieter hat zugleich aber viele Millionen Kunden. Die Überlegung, alles auf einem klassischen RDBMS abbilden zu wollen, schlägt allzu häufig fehl. Schlussfolgernd braucht es eine entsprechende skalierbare Big-Data-Infrastruktur.

Zentrale Device-Verwaltung mittels IoT-Gateways

Die Wegstrecke von den Sensoren zu den Geschäftsprozessen entscheidet über den Erfolg und die Kosten der digitalen Transformation. Schauen wir also auf ein Szenario, das sich bewährt:

Sensoren werden über ein IoT-Gateway gebündelt. IoT-Gateways stellen das Device Management und die Konnektivität bereit. Denn eines ist sicher: Die Datenverbindung von einer Zentrale oder einem Microservice zum Device ist nicht per se langfristig stabil. IoT-Gateways helfen die Konnektivität zu sichern, ohne dass dies zur Herausforderung beziehungsweise zum Show Stopper wird. Von großer Bedeutung ist ebenso die Absicherung der Kommunikation im Netzwerk. Der *Azure IoT Hub* hat sich hier in vielen Projekten bewährt [1]. Aber

```
<LISTING 1>
{ "temperature": 23.20, "humidity": 43.70 }
```

Listing 1: Beispiel Datenpaket eines Temperatursensors

```
12:20:00 { "temperature": 23.20, "humidity": 43.70 }
12:20:30 { "temperature": 23.30, "humidity": 43.80 }
12:21:00 { "temperature": 23.30, "humidity": 43.80 }
12:21:30 { "temperature": 23.20, "humidity": 44.00 }
12:22:00 { "temperature": 24.10, "humidity": 45.60 }
```

Listing 2: Beispiel Zeitreihen-Daten eines Temperatursensors

auch Open-Source-Alternativen wie etwa *Eclipse Kura* oder das *Thingsboard IoT-Gateway* können eine gute Alternative sein [2]. In vielen Fällen ergibt es Sinn, eine große Anzahl von Sensoren schon „vor dem IoT-Gateway“ mittels Edge Computing zu bündeln. Hier gibt es eine große Anzahl von Anbietern und Devices. Letztlich folgt man dem Gedanken, kleinere Embedded Computer „im Feld“ zu verwenden und Sensoren direkt vor Ort mit den Edge-Computern zu verbinden. Diese können dann lokal bereits Datenanalyse, Aufbereitung und Komprimierung vornehmen. Es ist über den Anwendungsfall abzuwägen, ob diese weitere Infrastruktur-Schicht zum Tragen kommen soll, denn auch die Edge-Computer stellen einen nicht zu unterschätzenden Pflegeaufwand dar.

Über das IoT-Gateway werden die Datenströme von den Sensoren gebündelt und deren Konnektivität abstrahiert. Der Datenstrom kann nun der Verarbeitungsschicht zugeführt werden.

Von der zentralen Sammlung der Daten von Sensoren zur Auswertung von Historiendaten

Während man in der Vergangenheit sinngemäß „one size fits all“ sagte und viele Aufgaben der klassischen relationalen Datenbank zugeschrieben hat, sind die Herausforderungen heute mit anderen Mitteln besser zu lösen.

Als nächste Infrastruktur-Schicht nach dem IoT-Gateway folgt das „Auffangen“ des Datenstroms. Hier empfiehlt sich der Einsatz von *Kafka* beziehungsweise *HiveMQ*. Es

handelt sich dabei um Systeme, deren Aufgabe es ist, Datenströme (Streams) sicher aufzunehmen, gegebenenfalls kurzfristig zwischenspeichern und an weitere Systeme als Stream zu verteilen.

Eines der nachfolgenden Systeme kann ein zum Beispiel auf *Hadoop* basierendes Big-Data-Ledger-System sein, das die Archivierung der eingehenden Daten übernimmt. Anhand des Archives lassen sich parallelisiert auf einem Hadoop Cluster (zum Beispiel in einer Kubernetes-Umgebung) verschiedenste Analysen auf den Historiendaten durchführen. Hier kann Apache Spark eingesetzt werden, das in letzter Zeit viel vom Apache-Flink-Projekt „übernommen“ hat.

Die Organisation des Datenschemas kann mit *Avro* erfolgen. Während Daten generalisiert und Schema-frei auf *Hadoop* gespeichert werden können, empfiehlt es sich, mittels *Avro* eine Datenschema-Organisation zu unterstützen (siehe Listing 3). So kann zum Beispiel ein Datenarchiv nicht nur einfach die Payload abspeichern, sondern zugleich auch die Schema-Information passend zu den Daten. Dies erleichtert eine automatisierte Verarbeitung in weiteren Schritten und Prozessen erheblich.

Sensordatenanalyse in Echtzeit

Nachdem Daten von Sensoren sicher aufgegriffen und im Archiv gespeichert wurden, kann man auf längeren Historiendaten Analysen laufen lassen. Dazu eignet sich das Hadoop-/Spark-System, wie auch andere

```

{
  "type": "record",
  "namespace": "Tutorialspoint",
  "name": "Employee",
  "fields": [{
    "name": "temperature",
    "type": "decimal"
  },
  {
    "name": "humidity",
    "type": "decimal"
  }
]
}
    
```

Listing 3: Beispiel Avro-Schema, passend zum Temperatursensor-Daten-Output

vergleichbare Distributed-Workload-Systeme, hervorragend.

Zugleich ist eine der wichtigen Ideen der Digitalisierung, rascher auf Ereignisse reagieren zu können, im Idealfall in nahezu Echtzeit. Um dies zu erreichen, müssen die eingehenden Datenströme nicht nur gespeichert, sondern im Fluss, im Stream direkt analysiert werden. Hierzu kommen direkt an *Kafka* angegliedert Streaming-Analytics-Systeme zum Einsatz. Eine spannende Lösung ist *Airflow*. Sie ermöglicht die visuell unterstützte Erstellung von Datenverarbeitungs-Pipelines, die im Stream genutzt werden können. Das bedeutet, dass Daten nicht mehr zu festgelegten Zeiten als Stapel, sondern direkt bei ihrem „Eintreffen ins System“ im Fluss verarbeitet werden können. Mithilfe einer zeitlich fest-

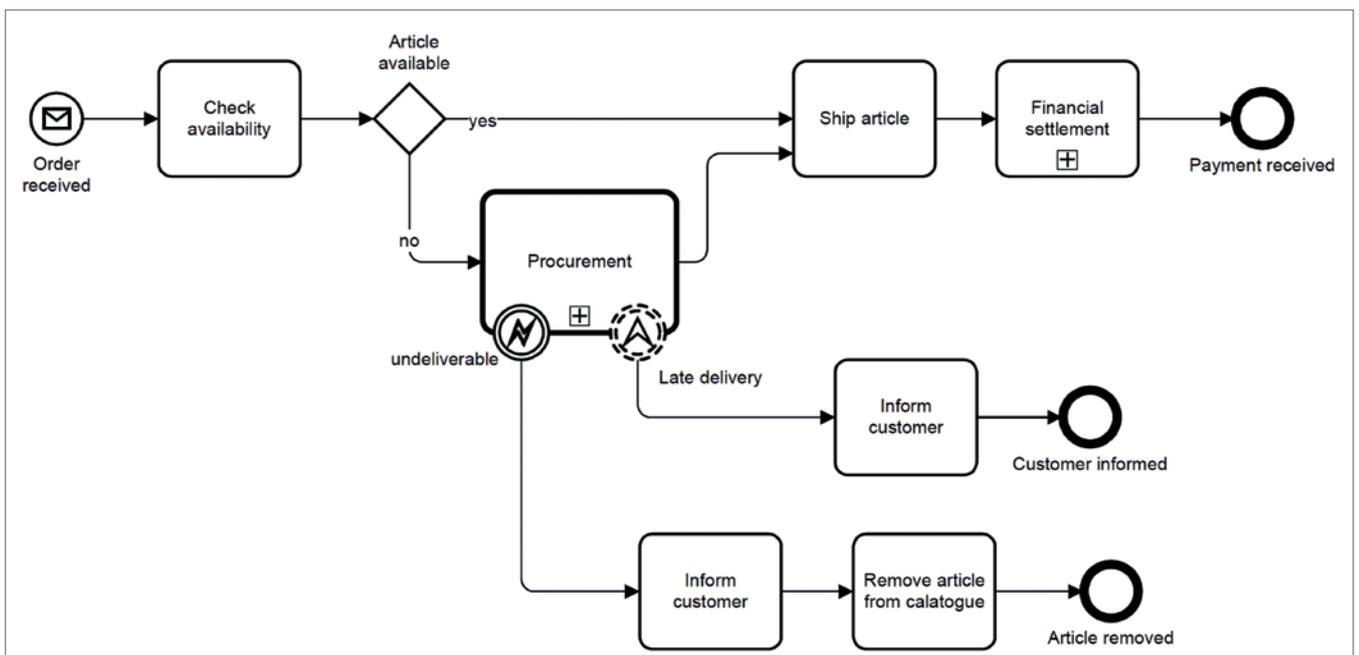


Abbildung 2: Order-Fulfillment-Prozess (© Object Management Group® <http://www.omg.org/spec/BPMN/2.0/examples/ZIP>)

gelegten Window-Funktion lassen sich dann Auswertungen innerhalb eines sich mit der Zeit mitbewegenden Zeitabschnitts machen.

Ganz konkret können wir auf unseren Temperatursensor schauen und einen Trigger formulieren, der ausgelöst wird, wenn die Abweichung der Temperatur einen Schwellenwert überschreitet. Auch können wir das Archivieren der Daten auf Veränderungen reduzieren. Wenn wir uns nochmals *Listing 2* anschauen, so ändert sich die Temperatur in der Zeit von 12:20:30 bis 12:21:30 nicht. Insofern würden wir Daten speichern, die keinen direkten Mehrwert bieten.

Aus dem Streaming-Analytics-System können wir bei einem ausgelösten Trigger wieder zurück nach *Kafka* kommunizieren, in ein eigenes *Kafka*-Topic. Auf dieses den Sensor-Trigger zugeordnete *Kafka*-Topic können wir eine Subscription anwenden, um immer dann informiert zu werden, wenn solche Trigger ausgelöst werden. *Kafka* wird dann hier also nicht nur von der eingehenden Seite der Daten genutzt, sondern zugleich auch als Lieferant von Daten und Ereignissen.

Einbindung der Sensorsignale und Trigger in Geschäftsprozesse

Nachdem wir sowohl die Sensor-Historien-daten als auch relevante Ereignisse im Daten-Stream nutzen können, ist es möglich, diese Daten in Geschäftsprozesse einzubinden. Schon seit Langem sind die Ereignis-gesteuerten Prozessketten bekannt und haben ihren Ruf. Später über BPM (Business Process Management) und nun mit der BPMN-2.0-Spezifikation (Business Process Model and Notation) ist eine sehr gute Evolution dieser Gedanken und Systeme zu beobachten. Es gibt inzwischen sehr schnelle BPMN 2.0 Engines (Interpreter) und Modellierungswerkzeuge auf Web-Basis.

In *Abbildung 2* ist eine Geschäfts-Prozess-Modellierung zu sehen. Selbst ohne IT-Fachwissen lässt sich die Funktion anhand der visuellen Repräsentation ablesen. Die Möglichkeit, dass Business- und Fachabteilung die inneren Funktionen einer Geschäftsanwendung nachvollziehen können, ist von entscheidendem Vorteil. Auf diese Weise können auch Mitarbeiter der Business- und Fachabteilung einen Teil der Anwendungen selbst erstellen, die benötigt werden.

Mit einer BPMN Engine lassen sich Geschäftsprozesse und -funktionen sehr gut abbilden und automatisieren. Zugleich ist BPMN nicht geeignet, um komplexere Kalkulationen durchzuführen, geschweige denn die Logik

dafür zu „modellieren“. BPMN kann dort sinnvoll genutzt werden, wo es um die Orchestrierung von Funktionen geht. Das bedeutet, dass BPMN die „inneren Zusammenhänge“, Datenflüsse und Abhängigkeiten von Funktionen eines Geschäftsprozesses abdecken kann.

Auf diese Weise kann eine BPMN Engine ein hervorragendes Werkzeug sein, um auf die Signale von Sensoren beziehungsweise genauer auf die von der Streaming Analytics erzeugten Trigger-Signale zu reagieren.

Die Brücke hin zu bestehenden Anwendungen und Microservices kann von BPMN durch die Webservice-Activities erstellt werden. In einem BPMN-Prozess lassen sich an beliebiger Stelle nicht nur Nachrichten versenden und empfangen, sondern auch Webservices und Rest-APIs aufrufen.

Machine Learning und die Smart Factory als Ziel der Digitalisierung

Mittels BPMN lässt sich auf einfache Weise auf Sensorsignale und Trigger in Geschäftsanwendungen reagieren. Damit ist eines der Ziele der Digitalisierung und Automatisierung leichter erreichbar. Ein zweites Ziel ist, zugleich diese Infrastruktur und Daten

für die Abschätzung zukünftiger Ereignisse einzubeziehen. Machine Learning bietet hier eine Reihe von Möglichkeiten.

Durch den aktuellen Hype rund um künstliche Intelligenz ist eine gewisse Ablenkung von der Verschiedenheit möglicher Machine-Learning-Grundlagen zu beklagen. Im Kern gibt es eine Reihe von Algorithmen-Familien, die sich bereits seit Langem bewähren: *Linear Regression, Logistic Regression, Support Vector Machines, Decision Trees, Random Forest, K-Means Clustering, K-Nearest Neighbour, Naive Bayes Classifier, Ensemble Learning*. Die besondere Aufmerksamkeit für künstliche Intelligenz ist jedoch den Entwicklungen im Bereich der Artificial Neural Networks geschuldet.

Die IT-Infrastruktur für Sensordaten-Verarbeitung ist also nach dem Sammeln und Archivieren der Daten und dem Bilden von Trigger-Signalen ebenso für die Erstellung von Vorhersagen (Predictions) zu nutzen. Das geschickte Anwenden von „Nahe-Echtzeit-Daten und Vorhersagen“, die automatisiert beziehungsweise halbautomatisiert auf Geschäftsprozesse und Fertigung Einfluss nehmen können, ist der Schlüssel zu einer erfolgreichen digitalen Transformation.

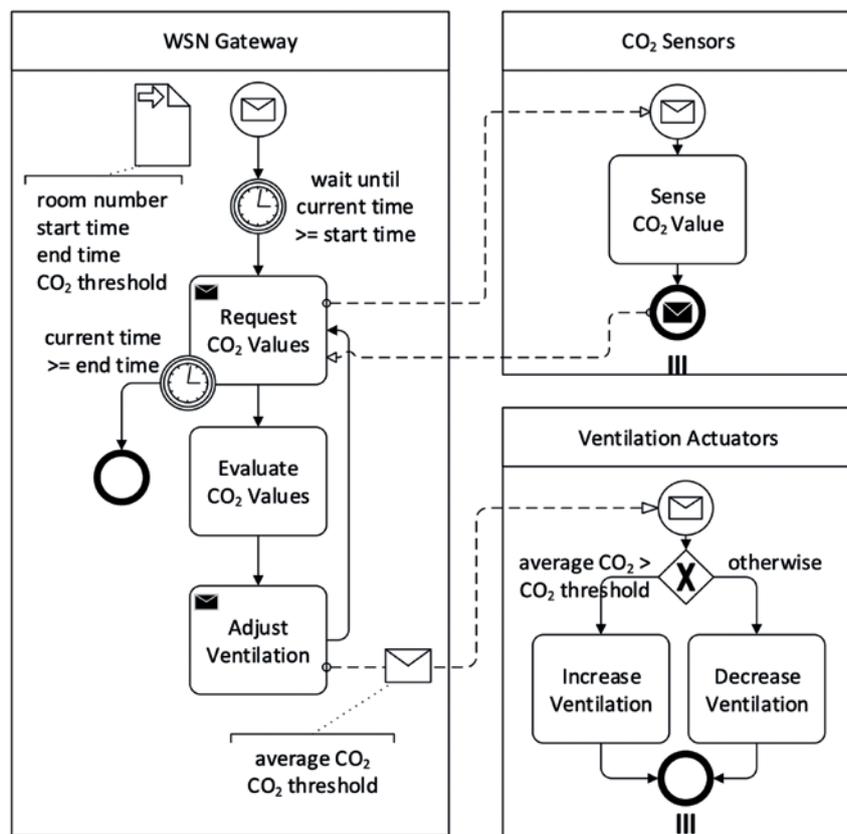


Abbildung 3: Orchestrating Actuators by sensor signals with BPMN Engine (© SAP Research, <https://www.iaas.uni-stuttgart.de/publications/INPROC-2013-37-Extending-BPMN-for-Wireless-Sensor-Networks-.pdf>)

Neue Version des DOAG-Lizenzguide veröffentlicht

DOAG-Redaktion

Die überarbeitete, 114 Seiten umfassende Version 4.0 enthält wichtige Aktualisierungen sowie neue Fachartikel.

Die Lizenzierungsregeln von Oracle sind nicht nur für Endanwender, die sich nicht regelmäßig mit dem Thema befassen, schwer zugänglich, sondern selbst für Fachleute in konkreten Fällen nicht immer leicht zu durchschauen. Aus diesem Grund stellt unser Arbeitskreis Lizenzierung unter der Leitung von Michael Paege, stellvertretender DOAG-Vorsitzender und Leiter Competence Center Lizenzierung, und unter Mitwirkung

der AOUG und SOUG seit 2013 einen kompakten Lizenzguide zum Download bereit. Dieser steht ausschließlich Mitgliedern der AOUG, DOAG und SOUG zur Verfügung.

Das Thema Lizenzierung ist niemals gänzlich abgeschlossen. Daher wird der DOAG-Lizenzguide auch zukünftig aktualisiert und erweitert werden. Unser Arbeitskreis freut sich über Feedback, Themenvorschläge und weitere Fachartikel.

DOAG-Lizenzguide steht zum Download bereit

DOAG-Mitglieder können sich den DOAG-Lizenzguide 4.0 unter <https://www.doag.org/de/themen/competence-center/lizenzierung/#c18974> herunterladen.

Voraussetzung für den kostenlosen Download der Vollfassung ist die Mitgliedschaft in der DOAG.

Termine

Juli

07

10.07.2020

DOAG Datenbank Webinar

Online

29.07.2020

DOAG Development Webinar: Datenschnittstellen mit GraphQL

Online

August

08

14.08.2020

DOAG Datenbank Webinar: Oracle technical support policies und Oracle lifetime support policy

Online

26.08.2020

DOAG Development Webinar: Kerberos, OAuth2 oder doch lieber SAMLV2: beliebte authentifizierungsmethoden unter der Lupe

Online

Wir begrüßen unsere neuen Mitglieder

Persönliche Mitglieder

- Armin Wildenberg
- Saeed Hassanpour
- Uwe Helten
- Thomas Kistler
- Robert Marz
- Marcus Mönning



Impressum

Red Stack Magazin inkl. Business News wird gemeinsam herausgegeben von den Oracle-Anwendergruppen DOAG Deutsche ORACLE-Anwendergruppe e.V. (Deutschland, Tempelhofer Weg 64, 12347 Berlin, www.doag.org), AOUG Austrian Oracle User Group (Österreich, Lassallestraße 7a, 1020 Wien, www.aoug.at) und SOUG Swiss Oracle User Group (Schweiz, Dornacherstraße 192, 4053 Basel, www.soug.ch).

Red Stack Magazin inkl. Business News ist das User-Magazin rund um die Produkte der Oracle Corp., USA, im Raum Deutschland, Österreich und Schweiz. Es ist unabhängig von Oracle und vertritt weder direkt noch indirekt deren wirtschaftliche Interessen. Vielmehr vertritt es die Interessen der Anwender an den Themen rund um die Oracle-Produkte, fördert den Wissensaustausch zwischen den Lesern und informiert über neue Produkte und Technologien.

Red Stack Magazin inkl. Business News wird verlegt von der DOAG Dienstleistungen GmbH, Tempelhofer Weg 64, 12347 Berlin, Deutschland, gesetzlich vertreten durch den Geschäftsführer Fried Saacke, deren Unternehmensgegenstand Vereinsmanagement, Veranstaltungsorganisation und Publishing ist.

Die DOAG Deutsche ORACLE-Anwendergruppe e.V. hält 100 Prozent der Stammeinlage der DOAG Dienstleistungen GmbH. Die DOAG Deutsche ORACLE-Anwendergruppe e.V. wird gesetzlich durch den Vorstand vertreten; Vorsitzender: Stefan Kinnen. Die DOAG Deutsche ORACLE-Anwendergruppe e.V. informiert kompetent über alle Oracle-Themen, setzt sich für die Interessen der Mitglieder ein und führen einen konstruktiv-kritischen Dialog mit Oracle.

Redaktion:

Sitz: DOAG Dienstleistungen GmbH
(Anschrift s.o.)
ViSdP: Mylène Diacquenod
Redaktionsleitung Red Stack Magazin:
Martin Meyer
Redaktionsleitung Business News:
Marcos López
Kontakt: redaktion@doag.org
Weitere Redakteure (in alphabetischer Reihenfolge): Benedict Bender, Markus Bente, Manfred Drozd, Dietmar Ebel, Michael Fischer, Randolph Geist, Norbert Gronau, Philipp Hartenfeller, Dennis Hoffmann, Lorenz Kiebler, Stefan Koehler, Oliver Lemm, Christian Luda, Michael Mühlbeyer, Wolfgang Nast, Dietmar Neugebauer, Karin Patenge, Matthias Peissner, Markus Peppler, Frank Pientka, Thomas Rein, Daniel Silter, Kevin Steinhagen, Carsten Wiesbaum, Falk Wolsky

Titel, Gestaltung und Satz:

Alexander Kermas
DOAG Dienstleistungen GmbH
(Anschrift s.o.)

Fotonachweis:

Titel: © pedrosek | <https://de.123rf.com>
S. 7: © Aurielaki | <https://stock.adobe.com>
S. 10: © Руслан Нестеренко | <https://de.123rf.com>
S. 15: © Ognyan Chobanov | <https://de.123rf.com>
S. 18: © apinan | <https://de.123rf.com>
S. 24: © Marek Uliasz | <https://de.123rf.com>
S. 29: © Jörg Schiemann | <https://de.123rf.com>
S. 33: © jossdiim | <https://de.123rf.com>
S. 34: © Leo Wolfert | <https://de.123rf.com>
S. 40: © tashatuvango | <https://de.123rf.com>
S. 44: © Tanawat Pontchour | <https://de.123rf.com>
S. 50: © Thanakorn Phanthura | <https://de.123rf.com>

S. 57: © Khoon Lay Gan | <https://de.123rf.com>
S. 63: © artinspiring | <https://de.123rf.com>
S. 67: © rastudio | <https://de.123rf.com>
S. 72: © O_du_van | <https://de.123rf.com>
S. 78: © Denis Putilov | <https://de.123rf.com>
S. 81: © Michal Bednarek | <https://de.123rf.com>
S. 84: © besjunior | <https://de.123rf.com>
S. 93: © gmast3r | <https://de.123rf.com>

Anzeigen:

Simone Fischer,
DOAG Dienstleistungen GmbH
(verantwortlich, Anschrift s.o.)
Kontakt: anzeigen@doag.org
Mediadaten und Preise unter:
www.doag.org/go/mediadaten

Druck:

WIRmachenDRUCK GmbH,
www.wir-machen-druck.de

Alle Rechte vorbehalten. Jegliche Vervielfältigung oder Weiterverbreitung in jedem Medium als Ganzes oder in Teilen bedarf der schriftlichen Zustimmung des Verlags.

Die Informationen und Angaben in dieser Publikation wurden nach bestem Wissen und Gewissen recherchiert. Die Nutzung dieser Informationen und Angaben geschieht allein auf eigene Verantwortung. Eine Haftung für die Richtigkeit der Informationen und Angaben, insbesondere für die Anwendbarkeit im Einzelfall, wird nicht übernommen. Meinungen stellen die Ansichten der jeweiligen Autoren dar und geben nicht notwendigerweise die Ansicht der Herausgeber wieder.

Inserentenverzeichnis

B4Bmedia.net AG https://e-3.de	S. 41	MuniQsoft Consulting GmbH www.muniqsoft-consulting.de	S. 23	Robotron Datenbank-Software GmbH www.robotron.de	S. 13
DOAG e.V. www.doag.org	U 3	MuniQsoft Training GmbH www.muniqsoft-training.de	S. 3	Trivadis AG www.trivadis.com	U 4
dbi services sa www.dbi-services.com	S. 61	PROMATIS software GmbH www.promatis.de	U 2		



2020
DOAG
Konferenz + Ausstellung

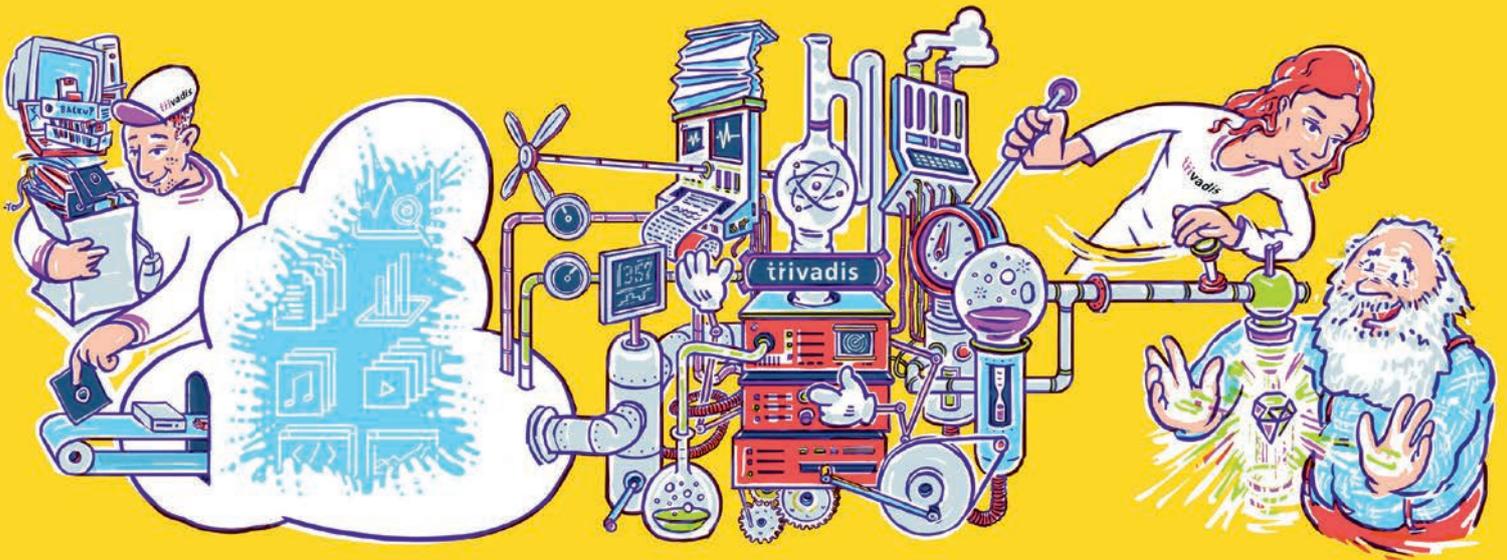
SAVE THE
DATE

17. - 20. Nov 2020



Einfach clever.

Deine
Cloud-Lösungen
von Trivadis.



Von Rohdaten zu Diamanten – mit Trivadis und Microsoft Azure

Mit unseren Cloud-Lösungen holst du das meiste aus deinen Daten. Doch wie schaffst du den Weg in die Cloud? Wie stellst du sicher, dass du die Übersicht behältst und die Kosten im Griff hast? Und wie bindest du auch Technologien wie IoT oder KI sinnvoll ein?

Mit unserem „Way to Azure“ gelingt dir der Weg in die Microsoft Cloud ganz selbstverständlich. Und mit unserer „Azure Foundation“ kannst du die Microsoft Cloud als sichere und produktive Plattform unkompliziert in deiner Unternehmung einführen. **Einfach clever!**

