

Red Stack

Magazin

DOAG

SOUG
swiss oracle
user group

AOUG
AUSTRIAN ORACLE USER GROUP

Jetzt inklusive BUSINESS NEWS



DIGITALISIERUNG

Aus der Praxis

Mit Lift & Shift
in die
Oracle Cloud



Im Interview

Dr.-Ing. Maximilien Kintz,
Leiter Team Angewandte KI,
Fraunhofer IAO

BUSINESS NEWS

Data Science:
Messbare Mehrwerte
und neue Produkte

DOAG 2021 Datenbank

17. und 18. Mai 2021 online

datenbank.doag.org



DOAG

Save the Date:





Niels de Bruijn
Leiter der Development
Community

Liebe Mitglieder, liebe Leserinnen und Leser,

die Digitale Transformation bezeichnet einen fortlaufenden, in digitalen Technologien begründeten Veränderungsprozess, der uns alle betrifft. Vieles wird digitalisiert und ermöglicht neue Dienste: von der Babyphone App bis zum selbstfahrenden Auto. Viele Unternehmen sind gezwungen, diesen Wandel mitzumachen, ansonsten werden sie früher oder später durch andere innovative Unternehmen überholt. Die bekannten Gewinner sind hier Amazon, Lego und Uber, die Verlierer sind beispielsweise Kodak, Nokia und Quelle. Der technologische Fortschritt stellt das aktuelle Geschäftsmodell in Frage. Innovative Unternehmen schaffen den Wandel, und wer sich auf bisherigen Erfolgen ausruht, hat irgendwann das Nachsehen.

In dieser Ausgabe befassen wir uns mehrfach mit dem spannenden Thema „Digitale Transformation“ begleitet durch Artikel über Oracle APEX, Postgres, MySQL, Requirements Engineering und vieles mehr.

In der Business News ab Seite 68 blicken wir auf den aktuellen Stand von Data Science und klären einmal mehr darüber auf, was Data Science eigentlich genau ist und kann – und was es nicht kann. Darüber hinaus bildet der spannende Leitartikel ein Data-Science-Projekt ab, dem mithilfe einer speziellen Kundensegmentierung und feingliedrigen Warenkorbanalyse die Identifikation von Kundenbedürfnissen gelingt – und zugleich die Steigerung des Umsatzes. Angereichert wird das Ganze mit einem Ausflug in die „digitale Evolution in Zeiten von VUCA“.

Viel Spaß beim Lesen, wünscht Euch

Euer Niels de Bruijn



Training

Training

MUNIQSOFT

TRAINING

20 Jahre Oracle-Datenbankschulungen von Experten, effizient und kundenorientiert!

Sie können an all unseren Schulungen auch ONLINE teilnehmen.

Im Livestream verfolgen Sie die gewünschten Kurse von zu Hause oder Ihrem Büro aus.

SQL I Grundlagen	15.03.-17.03.2021	€1.190.-netto
PL/SQL II Fortschritt	22.03.-26.03.2021	€2.990.-netto
DBA I Grundlagen	19.04.-23.04.2021	€2.190.-netto
Postgres	26.04.-28.04.2021	€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!



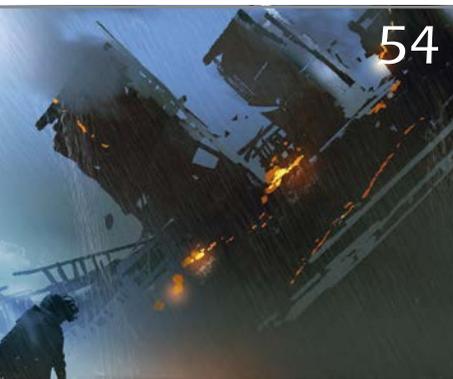
7

Interview mit Dr.-Ing. Maximilien Kintz



14

Enterprise Cloud Readiness – was ist zu tun?



54

Agile Entwicklung ohne automatisiertes Testen = Schiffbruch



68

Ein junges Team rollt den Lebensmittelgroßhandel neu auf – mit Data Science

Einleitung

- 3 Editorial
- 5 Timeline
- 6 Aus der Ferne betrachtet: Wir können Digitalisierung!
- 7 „Eine organisationsweite Strategie im Bereich Digitalisierung und KI muss von oben getrieben werden, aber auch alle Mitarbeitenden – und deren unterschiedliche Bedürfnisse, Vorkenntnisse, auch Ängste – mitnehmen können.“
Interview mit Dr.-Ing. Maximilien Kintz

Digitale Transformation

- 10 Digitale Transformation – Agilität überall um jeden Preis?
Carsten Wiesbaum

Datenbank

- 14 Enterprise Cloud Readiness – was ist zu tun?
Nisanth Muthukurushnasamy
- 18 Mit Lift & Shift in die Oracle Cloud
Stefan Kühnlein
- 25 OCI Architecture Patterns
Michael Fischer
- 29 Tuning einer Multitenant-Datenbank
Clemens Bleile
- 35 Warum Postgres immer mehr an Bedeutung in IT-Landschaften gewinnt
Borys Neselovskyi
- 39 Von Oracle zu Postgres im Handumdrehen
Andreas Kother & Markus Klenke
- 44 MySQL 8 Best Practices – Tuning für Dummies
Carsten Thalheimer & Mario Beck

Data Analytics

- 49 Datenschutz: Anonymisierung und Pseudonymisierung
Andreas Buckenhofer

Strategie & Innovation

- 54 Agile Entwicklung ohne automatisiertes Testen = Schiffbruch
Jan Ott
- 59 Requirements Engineering – wieso, weshalb, warum?
Tobias Schweiker

MariaDB/MySQL

- 63 MariaDB-/MySQL-Stolperfallen – und wie komme ich da wieder raus? – Teil 1
Olli Sennhauser

Data Science: Mit den richtigen Daten messbare Mehrwerte und neue Produkte

BUSINESS NEWS

- 68 Die Bedürfnisse des Kunden identifizieren – Ein Data-Science-Projektbericht: Von der Datenanreicherung bis zur gezielten Kundenansprache
Jürgen Günter, Andre Stapper, Karolina Weber, Hermann Wilde
- 77 Optimierung und Automatisierung maschinellen Lernens
Malena Reiners
- 80 Digitale Evolution in Zeiten von VUCA: Agilität, das Gebot der Stunde
Dr. Frank Schönthaler & Marion Rosemann
- 86 Data Science, KI, maschinelles Lernen: Was kann das alles (nicht)?
Maximilien Kintz
- 89 Process Mining: Daten für ganzheitliches Geschäftsprozessmanagement
Janna Meyer & Claus Engel
- 98 Compliance und Corona: Neue Normen und Technologien?
Alexey Zakharov

Intern

- 101 Neue Mitglieder + Termine
- 102 Impressum + Inserenten

Timeline

8. Januar 2021

Mit seiner Präsentation "Durch die Dornen zu den Sternen: Migration von Oracle nach Postgres" zeigt Borys Neselovskyi beim DOAG Datenbank Online-Event die für eine einfache und problemlose Migration notwendigen Vorkehrungen und Schritte.

20. Januar 2021

Die Generalversammlung der Austrian Oracle User Group, auf der der Vorstand für das Jahr 2021 neu gewählt wird, findet statt. Klaus-Michael Hatzinger (DBConcepts) wird für ein weiteres Jahr zum Präsidenten gewählt und Bernhard Halbetel (SoftwareOne) wird erneut das Amt des Vizepräsidenten des Vereins übernehmen. Neben weiteren bereits bekannten Vorstandskollegen wird der Vorstand dieses Jahr um drei neue Gesichter erweitert: Martin Veren (primeone business solutions) für den Veranstaltungsbe- reich SaaS Cloud, Martin Seiser (Immofinanz AG) für PaaS Cloud und Paul Hudson (Agrarmarkt Austria) für Application Security.

21. Januar 2021

Das Online-Event DOAG CC Support zum Thema "Oracle's Autonomous Health Framework: How TFA Collector and ExaChk will improve your relationship with Oracle Support" wird von den Oracle-Referenten Bill Burton und Bryan Vongray präsentiert.

26. Januar 2021

Beim ersten Termin der vierteiligen Veranstaltungsreihe NetSuite User Days 2021 steht geballtes Praxiswissen auf dem Programm. Nach zwei erfolgreichen Präsenzveranstaltungen in Berlin und Hamburg wird das Event-Format an die aktuelle Lage angepasst und 2021 in neuem Gewand präsentiert: online! Der Grundgedanke ist jedoch geblieben: Top-Vorträge für einen optimalen Wissensaustausch.

1. Februar 2021

Die AOUG startet das Call for Papers für die diesjährige Anwenderkonferenz. Da sich die Corona-Situation voraussichtlich frühestens im Herbst lockern wird, wird die Anwenderkonferenz unter dem Motto „AOUG Unplugged 2021“ im September geplant. Durch zahlreiche Reisebeschränkungen ist davon auszugehen, dass diese Veranstaltung sowohl live als auch per Video-Zuschaltung stattfinden wird.

2. Februar 2021

Am zweiten Tag der NetSuite User Days 2021 lauten die Themen "Keeping Your NetSuite Tidy" mit Beth Jose (OLX) sowie "Professionelle Unternehmensplanung und Balanced Scorecard für NetSuite-Anwender" mit Philipp Werner (Promatis). Anschließend wird es ein Get-together geben.

9. Februar 2021

Der dritte Tag der Tag der NetSuite User Days 2021 bietet erneut zwei spannende Präsentationen: "Effizienz im Rechnungseingangs-

prozess mit Fast Four Scan & Capture" (Akito Suzuki, Spryker) und "Oracle NetSuite Lokalisierungs News (Deutschland) und SuiteTax Neuerungen" (Mathias Reinecke, Oracle NetSuite).

16. Februar 2021

Am vierten NetSuite User Day stehen die "Handhabung von großen Transaktionsmengen mit NetSuite Premium Tier" (Aris Brauchart, Waterdrop) und ein Roundtable zu Themen aus den vorherigen Sessions sowie unter anderem Service-Tier und OCI2 auf der Agenda.

27. Januar 2021

Zu den spannendsten neuen Features, die Oracle 18c für Entwickler bietet, zählen polymorphe Tabellenfunktionen und qualifizierte Ausdrücke. Der Vortrag von Robert Marz im Rahmen des DOAG Development Online-Events stellt die Konzepte der Polymorphic Table Functions und Qualified Expressions vor und zeigt Anwendungsfälle und Szenarien aus der Praxis.

12. Februar 2021

Das DOAG Datenbank Online-Event "Oracle 18c/19c Microsoft Active Directory Integration" mit dem Referenten Christian Pfundtner findet statt.

24. Februar 2021

Das DOAG Development Online-Event mit dem Thema „EBR oder Exception Handling“ mit dem Referenten Sven Weller steht auf dem Programm.

12. März 2021

Der Referent Christian Floreck referiert beim DOAG Datenbank Online-Event zum Thema „Autoupgrade von Datenbanken mittels Autoupgrade“.

16. und 17. März 2021

Aufgrund der andauernden Pandemie-Situation findet die JavaLand 2021 als zweitägige Online-Konferenz statt. Mehr als 120 Vorträge in acht Streams rund um viele beliebte Themen aus dem Java-Bereich stehen auf der Agenda! Um den Teilnehmern dennoch ein persönliches Networking bestmöglich zu ermöglichen, haben wir das allen Teilnehmern aus den vergangenen Jahren bekannte JavaLand grafisch nachgebildet. Mit Avataren kann es durch die Besucher erkundet werden und bei Begegnungen mit anderen Teilnehmern können per Video-Chat Gespräche geführt werden. Auch viele Community-Aktivitäten finden so zur Begeisterung der Teilnehmer im virtuellen Park statt.



Aus der Ferne betrachtet: Wir können Digitalisierung!



Günther Stürner

Zum Wort des Jahres hat es ‚Digitalisierung‘ bisher nicht geschafft. In den Reden und Verlautbarungen von Politikern und Industriekapitänen ist dieser Begriff aber omnipräsent.

Digitalisierung ist der Inbegriff für Modernes, für Zukunft, für Silicon Valley, für Cooles halt.

Wie sagte eine bekannte Kanzlerin einst: „Internet ist für uns alle Neuland“; damit wollte sie sich dafür entschuldigen, dass die Sache mit der Überwachungssoftware PRISM aus dem Ruder lief. Aber das ist ja auch schon ‚ewig‘ her, das war damals, zwanzig-dreizehn.

Heute ist alles anders.

Heute gibt es eine Staatsministerin für Digitales, und Digitalisierung ist Chefinnen-Sache. Das ist gut so! Auch wenn Digitalisierung neben Netzausbau, IT-Sicherheitsgesetzen und ‚iPads für alle‘ noch viele weitere unterschiedliche Facetten beinhaltet, die eher selten in Parteiprogrammen und Reden auftauchen.

Es ist keine Frage, wir als Hochtechnologieland werden unseren Lebensstandard und unsere Stellung in der Welt nur dann halten können, wenn wir technologisch an der Spitze mitmischen. Software und Daten spielen die maßgeblichen Rollen in diesem Spiel. Software, Software und nochmals Software wird das Maß der Dinge in der Zukunft sein. Warum sollten sich VW, Mercedes oder BMW sonst ein eigenes Betriebssystem bauen. Kein Betriebssystem für einen klassischen Rechner oder für ein Telefon, nein, ein Betriebssystem für ihre Automobile. Software als Differenzierungsmethode für den nächsten Neuwagen! Fahrzeuge, die alles Mögliche können, sogar uns komfortabel von A nach B bringen.

Oder die klassischen Maschinenbauer, die bauen heute High-Tech-Maschinen, vollgestopft mit Software, Computer-Chips, Sensoren, angeschlossen an Datenbanksysteme und deren Machine-Learning-Komponenten, um die Überwachung und die Qualitätssicherung in Real-Time auszuführen. Closed-Loop zum Anfassen.

Ganz zu schweigen von den vollautomatisierten Fertigungsstraßen. Software, soweit das Auge reicht. Digitalisierung vom Feinsten. Und wir, die privilegierte Klasse der Informatiker und IT-ler sind mittendrin. Voll im prallen Leben, meist gut bezahlt und mit einem Job, der einen oft nicht loslässt.

Digitalisierung ist schon da, wenn auch nicht gleichmäßig verteilt. Es gibt noch genügend Bereiche, die noch in den 80er

oder 90er Jahren verharren. Ja, es gibt immer noch IT-Systeme, die aus diesen Jahren stammen und brav ihre Arbeit als Cobol-Programme (Cobol==Programmiersprache für kommerzielle Aufgabestellungen) verrichten. Ganz zu schweigen von handschriftlichen Listen, die in manchen Amtsstuben, aber nicht nur dort, als Basis für die Buchhaltung dienen. Oder Schulen, die mit Museums-PCs ausgestattet sind oder an denen sich manche Lehrer als Hardcore-WLAN-Verweigerer gerieren.

Apropos Schulen. Hier wird es höchste Zeit, dass die Informatik Einzug hält. Ich habe Respekt davor, dass sich immer wieder Lehrer und Lehrerinnen (meist Mathe oder Physik) als Interims-Informatik-Lehrkräfte zur Verfügung stellen. Aber das ist weniger als eine Krücke. Für die Lehrer nicht befriedigend und für die Schüler nicht zielführend. Und wer denkt, dass iPads für alle Schüler und intelligente Wandtafeln den bildungspolitischen und digitalen Durchbruch bringen, hat den Knall überhört.

Was wir brauchen, ist ein exzellenter Lehrplan für das Fach Informatik, der von Fachleuten erstellt und periodisch überprüft wird (Informatik ist verdammt schnell und dynamisch), und wir brauchen die zugehörigen Geräte, Services, Kontakte zu Praktikern und Firmen. Wir brauchen gut ausgebildete Informatik-Lehrer, die in der Lage sind, dieses breite und wichtige Fachgebiet den Schülern nahezubringen. Denn seien wir mal ehrlich. Die sogenannten Digital Natives haben bis auf einige Ausnahmen keine Ahnung, was „hinter dem iPhone-Bildschirm“ eigentlich passiert. Wir sind und wir werden immer mehr eine Gesellschaft, in der die große Mehrheit halt nicht weiß, was sich IT-technisch tut. Nicht jeder kann ein Spezialist sein, das ist nicht die Forderung. Aber jeder sollte in Grundzügen verstehen, was sein Leben maßgeblich beeinflusst und steuert. Nur so schaffen wir Akzeptanz für das viele Neue, das auf uns zukommt.

Digitalisierung der Wirtschaft und der Gesellschaft wird und muss voranschreiten. Es ist wichtig, dass Organisationen wie die DOAG und Firmen wie Oracle, neben vielen anderen Akteuren, aktiv an diesem Prozess mitarbeiten. Es ist wichtig, dass die Protagonisten sich nicht verstecken hinter Datenbanken, SQL, JSON und Co. Es ist wichtig, dass wir uns, neben vielen anderen, einbringen mit unserer Erfahrung, unserem Wissen.

Dann wird Digitalisierung zu einem Erfolg.

E-Mail: guenther.stuerner@dbms-publishing.de



„Eine organisationsweite Strategie im Bereich Digitalisierung und KI muss von oben getrieben werden, aber auch alle Mitarbeitenden – und deren unterschiedliche Bedürfnisse, Vorkenntnisse, auch Ängste – mitnehmen können.“

Martin Meyer, Redaktionsleiter des Red Stack Magazin, sprach mit Dr.-Ing. Maximilien Kintz, Leiter Team Angewandte Künstliche Intelligenz am Fraunhofer-Institut für Arbeitswirtschaft und Organisation IAO über Digitalisierung und den Einsatz von Künstlicher Intelligenz.

Herr Dr. Kintz, was ist Ihre berufliche Verbindung zum Thema Digitalisierung und KI? Mit welchen Fragen und Aufgaben beschäftigen Sie sich?

Im Team „Angewandte Künstliche Intelligenz“ am Fraunhofer IAO helfen wir Unternehmen und Organisationen, ihre Arbeitsabläufe mit Software und spezifisch mit KI-Ansätzen effizienter zu machen. Konkret heißt das: Herausfinden, wo und wie man Menschen mit KI-basierter Software unterstützen kann, was man automatisieren kann, wie diese neuen Technologien in die

aktuellen Prozesse passen. Wir bieten dazu sowohl Recherche oder Beratung als auch eigene Technologieentwicklung oder auch Schulungen an.

Was sind die Gründe für eine digitale Transformation? Wer benötigt Künstliche Intelligenz und Digitalisierung oder wo liegt das Potenzial?

Wer benötigt das nicht! Einfach gesagt sind wir alle von der Digitalisierung und auch von Fortschritten im Bereich KI betroffen,

sowohl im beruflichen als auch im privaten Bereich (da kann man einfach an Smartphones denken oder an Betrugserkennung bei Bezahlungen, das ist sowohl für Banken als auch für Endkunden wichtig). Warum Digitalisierung auch an Schulen und Universitäten wichtig ist, hat das letzte Jahr ganz deutlich gemacht.

Als Gründe kann man einerseits technologische Fortschritte nennen, die Dinge ermöglichen, die früher nicht möglich waren (weil Hardware, Algorithmen und Vernetzung alle effizienter, schneller und leichter nutzbar werden). Andererseits gibt es weitere Gründe, die vermutlich mehr oder weniger abhängig voneinander sind: regulatorische Aspekte, Druck von Wettbewerbern, neue Erwartungen der Kunden oder Mitarbeitenden (z.B. wenn Home Office beliebter wird, müssen Firmen das auch ermöglichen, um attraktiv zu bleiben, und das erfordert einen gewissen Digitalisierungsgrad).

Wer ist für Digitalisierung und KI verantwortlich? Wer oder was treibt sie voran? Ist die Corona-Pandemie ein Beschleuniger?

Eine organisationsweite Strategie im Bereich Digitalisierung und KI muss von oben getrieben werden, aber auch alle Mitarbeitenden – und deren unterschiedliche Bedürfnisse, Vorkenntnisse, auch Ängste – mitnehmen können. Das ist jedoch einfacher gesagt als getan! Dafür spielen Schulungen, Weiterbildungsmaßnahmen, Innovationsstrategien oder Kommunikation eine wichtige Rolle.

Entwicklungen in der Forschung und neue Technologien (Algorithmen, Hardware) sind sicherlich ein Treiber – und bessere Antworten als Reaktionen auf Aktivitäten von Wettbewerbern oder neue Regelungen!



Zur Person: Maximilien Kintz

Dr.-Ing. Maximilien Kintz hat in Nantes (Frankreich) und Stuttgart Ingenieurwissenschaften und Informatik studiert. Er forscht seit 2009 am IAT der Universität Stuttgart und am Fraunhofer IAO, wo er das Team „Angewandte Künstliche Intelligenz“ leitet. Im Team werden Lösungen untersucht und entwickelt, die Unternehmen und Organisationen bei der Optimierung und Atomisierung von Arbeitsabläufen unterstützen. Schwerpunkte der Arbeit von Herrn Kintz sind die Analyse von Textdokumenten, Dashboards und die Überwachung von Prozessen.

Die Corona-Pandemie hat – wie wir alle feststellen konnten – bestimmte Aspekte der Digitalisierung beschleunigt: Videokonferenzen sind heute verbreiteter als noch vor einem Jahr, dasselbe gilt für Fernunterricht an Schulen. Gleichzeitig hat die Pandemie aber auch andere Projekte gebremst, weil die nötigen Ressourcen für langfristige Aktivitäten nicht mehr verfügbar waren, da man sich um eine Krisensituation kümmern musste. Kurzfristig wurden also Änderungen beschleunigt, mittelfristig glaube ich eher, dass wichtige Projekte gebremst wurden.

Wo führt digitale Transformation und KI hin?

Unter anderem führen Digitalisierung und Automatisierung der Aktivitäten mit KI zu einer Änderung der Aufgaben und Tätigkeiten, die für Menschen wichtig sind. Computer können immer besser und schneller analysieren, auswählen, bei Entscheidungen unterstützen, Vorhersagen machen. Menschen haben immer noch einen Vorteil bei kreativeren Aufgaben, und vor allem dann, wenn Menschen sich Nähe und soziale Kontakte wünschen. Die Bereiche, wo eine KI nicht gut genug ist, um Menschen zu ersetzen, werden immer weniger. Aber es bleiben viele Bereiche, in denen eine KI sich finanziell nicht lohnt, und noch mehr Bereiche, in denen keine erwünscht ist, weil wir lieber mit anderen Menschen zusammenarbeiten. Es ist wichtig, diese Bereiche und Aufgaben zu identifizieren.

Welche Konsequenzen und Probleme können sich aus dem Einsatz von KI und digitalen Technologien ergeben? Welche Fragen sind zu lösen?

Menschen tendieren dazu, Fehler von Menschen deutlich leichter zu akzeptieren als Fehler von Computern. Das führt dazu, dass die Anforderungen an KI-Lösungen generell sehr hoch sind und dass die Akzeptanz der KI-Entscheidungen manchmal niedrig ist. Viele KI-Verfahren sind auch Black-Box-Verfahren, die für Menschen ohne Weiteres nicht nachvollziehbar oder transparent sind (da diese einfach eine riesige Menge an Parametern berücksichtigen). Erklärbare KI (Explainable AI, kurz XAI) oder der Versuch, diese Entscheidungen für Menschen verständlich zu machen, ist also derzeit ein wichtiges Forschungsthema. Die Nutzung von Trainingsdaten als Basis für neue KI-Anwendungen, die die Vielfalt unserer Welt richtig abbilden, ist auch ein Thema. Bestimmte Personengruppen sind manchmal in den genutzten Daten unterrepräsentiert, was zu Diskriminierungen bei der Anwendung von KI-Lösungen führen kann. Und Themen wie Datenschutz und Datensouveränität (Was kann man in der EU machen? Welche personenbezogenen Daten werden unkontrolliert oder schlecht kontrolliert in die USA oder nach China geschickt?) sollen auch betrachtet werden.

Welche Aufgaben kommen auf die deutschen Unternehmen, Verwaltung und Politik zu? Wo sehen Sie Hemmnisse?

Für Unternehmen: Verstehen, dass KI-Projekte datenintensiv sind, dass Daten wichtig und wertvoll sind und dass man dafür Zeit und Geld investieren muss! Ohne gute Daten kann man immer noch keine KI bauen. Und bereit sein, nicht sofort große Erfolge zu sehen, sondern erst nach Monaten oder Jahren Arbeit... Für die Politik: Sicherstellen, dass sensible Daten geschützt werden, ohne dass es zu große Probleme für Unternehmen gibt (klare, einfache Regeln sind wichtig). Datenschutz soll Daten schützen, nicht Innovation behindern. Gleiches gilt für

Ethik und faire oder transparente KI: Anspruchsvolle, aber auch realistische Anforderungen und Vorgaben sind wichtig.

Wo steht Deutschland beim Einsatz von KI und Digitalisierung? Gibt es regionale Unterschiede?

Was Kompetenzen und Fachkräfte angeht, ist Deutschland sicherlich sehr gut dabei. Die USA oder China haben jedoch einfach dank deren Größe einen Vorteil; für Deutschland scheint also die EU eine sinnvolle Ebene zu sein, um aktiv zu werden. In der Praxis wird aber sehr viel auf Landes- oder Bundesebene gemacht, es sollten öfter EU-weite Kooperationen genutzt werden. Regionale Unterschiede kann ich nicht beurteilen. In größeren Städten oder in der Nähe von Universitäten sind sicherlich mehr Start-ups aktiv als in anderen Gebieten.

Welchen Ausblick können Sie hinsichtlich der Digitalisierung geben? Können Sie eine Prognose wagen?

Eine einfache Prognose: Digitalisierung und KI werden noch wichtiger und verbreiteter. Spezifische KI-Bereiche wie die Transparenz von KI-Verfahren oder automatisiertes Lernen, neue Anwendungen in der Biologie und Medizin oder die Analyse natürlicher Sprache werden wichtiger; wir werden dort in den nächsten Jahren verstärkt Erfolge sehen. Die Grundlagen sind bereits da und die Übertragung in die Praxis hat angefangen. Die Performance von Hardware wird weiterhin steigen und immer weniger eine Limitierung sein – auch billige oder kleine Geräte sind heutzutage leistungsfähig genug, um (einfache) Modelle zu trainieren. Dieser Trend wird sich fortsetzen.



JavaLand 2021: Anpassung der Ticketpreise

Erfreuliche Neuigkeiten für Java-Fans: Wir haben die regulären Ticketpreise für die JavaLand 2021 angepasst! Diese Reduzierung ist auch dank der großzügigen Unterstützung unserer Sponsoren möglich.

Auf einer jeden JavaLand steht ihr, die Community, im Mittelpunkt. Daher bietet euch die Konferenz in diesem Jahr, trotz der besonderen Umstände und räumlicher Trennung, neben dem breitgefächerten Vortragsprogramm weiterhin auch interaktive Community-Aktivitäten und Workshops und jede Menge Networking-Möglichkeiten, um den Wissenstransfer und die Verknüpfung innerhalb der Community noch weiter zu stärken.

Die JavaLand ist seit ihren Anfängen Community-getrieben. Was bedeutet das? Viele Personen wirken ehrenamtlich mit und investieren viel Zeit und Liebe zum Detail, damit dieses Angebot zustande kommt.

In letzter Zeit wurden wir häufig gefragt, wie die Preise der Online-Konferenz zu-

stande kommen. Deswegen möchten wir unseren Ansatz hier erläutern: Wir schöpfen unseren finanziellen Ressourcen nicht aus einem Marketing-Topf. Wir finanzieren uns aus den Einnahmen, die durch Teilnehmertickets und Sponsoring-Aktivitäten generiert werden. Überschüsse investieren wir wieder in die Veranstaltungen und Aktivitäten der Community im iJUG. Dazu zählt auch unsere gemeinsame Publikation Java aktuell. Für die Koordination dieser Aktivitäten greifen wir auf ein Backoffice zurück, das das ganze Jahr über dafür sorgt, dass unsere Aktivitäten vorankommen. Die Geschäftsstelle organisiert und koordiniert die Aktivitäten der Mitwirkenden aus der Community und kümmert sich um alles Nötige im Hintergrund, vom Teilnehmermanagement bis hin zur Akquise von Sponsoren.

Aktuell beobachten wir trotz der pandemiebedingten Einschränkungen einen großen Zuspruch bei unseren Sponsoren. Wir sind überwältigt von ihrem Commitment gegenüber der JavaLand und ihrer Teilnehmer. Wir tun alles, um auch ihnen mit unserem Online-Konzept eine angemessene Sichtbarkeit zu geben. Deswegen freuen wir uns, wenn die Teilnehmer der JavaLand 2021 auch ihnen einen virtuellen Besuch abstatten. Denn auch sie wollen mit der Community in Kontakt treten und ihr Netzwerk erweitern.

Wir freuen uns, trotz Nicht-Durchführbarkeit einer physischen Veranstaltung, über die mehr als positiven Aussichten der Online-Konferenz, die uns ermöglichen, euch einen neuen, reduzierten Ticketpreis anzubieten.



Digitale Transformation – Agilität überall um jeden Preis?

Carsten Wiesbaum, esentri

Der Begriff der „Digitalen Transformation“ läuft uns heutzutage im Geschäftsumfeld und in der Politik häufig über den Weg. Im Kern geht es darum, durch die Nutzung digitaler Infrastruktur und damit bereitgestellter Anwendungen das Verwertungspotenzial digitaler Geschäftsmodelle zu nutzen. Ein wichtiger Aspekt ist dabei die Volatilität des Marktes und wie diese von Unternehmen beherrscht, aber auch bedient werden kann. In diesem Kontext wird „Agilität“ oft als Silberkugel für alle Herausforderungen in Unternehmen ins Feld geführt. Dabei wird leider oft der Fehler gemacht, Prozesse und Methoden aus der agilen Praxis blind auf alle möglichen Bereiche anzuwenden, ohne die Sinnhaftigkeit für den Bereich zu hinterfragen. Ist es sinnvoll, Entwicklung und Betrieb auf Scrum umzustellen, wenn sich am Beauftragungsprozess und beim Controlling nichts ändert? Muss jedes Sachbearbeiterteam auf agile Führung umgekrempelt werden und ist dies zielführend?

Mit der digitalen Transformation beschäftigt sich unsere Gesellschaft nicht erst seit Kurzem. Die Digitalisierung unserer täglichen Arbeit und der Prozesse in Unternehmen ist schon vor Jahrzehnten gestartet. Kommunikation in Unternehmen geschieht durch E-Mails oder Chat-Werkzeuge wie Slack oder MS Teams. Dokumente werden fast überall nicht mehr per Hauspost versendet, sondern durchlaufen digitalisierte Prozesse. Ganze Unternehmensbereiche, wie die Logistik oder Buchhaltung, wären in den aktuellen Dimensionen ohne die Unterstützung von Software gar nicht möglich. Was hat sich also geändert, dass die digitale Transformation in den letzten Jahren so präsent zu sein scheint?

Hierfür gibt es eine ganze Reihe von Gründen, wobei der Autor zwei als besonders relevant für die beschriebene Präsenz sieht. Zum Ersten hat sich die Entwicklung von Technologien in den letzten Jahren stetig beschleunigt. Insbesondere im Bereich der Softwareentwicklung gibt es immer neue Innovationen und Trends, die zur Erfüllung des Kundenbedarfs genutzt werden können. Junge Unternehmen und Start-ups nutzen diese Technologien, um ihren Kunden ein gutes Erlebnis bei der Nutzung ihres Produktes zu bieten. Ihr Vorteil dabei ist, dass sie

nur wenige oder keine existierenden Produktivsysteme und Prozesse haben, die es zu integrieren gilt. Moderne Funktionen, eine hohe Qualität und eine rasante Weiterentwicklungsgeschwindigkeit sind für viele in unserer Gesellschaft bereits der Standard.

Der zweite relevante Grund sind die oftmals komplett digitalen Geschäftsmodelle, die auf der Verwendung von digitaler Infrastruktur aufbauen. Entsprechende Geschäftsmodelle ersparen sich die Bindung an die physische Welt und die damit verbundene Trägheit. Die Produktion realer Güter ist langwierig und kostet Geld. Die Beschäftigung einer Belegschaft von qualifizierten Facharbeitern, das Bereitstellen eines geeigneten Arbeitsplatzes und entsprechendes Arbeitsmaterial ebenso. Kleine Unternehmen starten oft mit nichts und zielen auf den Plattform- oder Service-Markt ab. Initial werden lediglich eine Idee, ein paar Softwareentwickler und etwas Kapital benötigt. Im besten Fall entstehen zunächst nur Kosten für den Betrieb einer Webseite, auf der potenzielle Kunden ein Abo abschließen können. Erst wenn der erste Kunde diesen Schritt gemacht hat, werden reale Instanzen des Produktes gestartet und weiteres Geld in die Entwicklung investiert. Durch von vornherein vorgesehene

Mechanismen können das Nutzerverhalten analysiert und kurzfristig auf den Bedarf des Kunden reagiert werden. Auch hier lernen wir als Kunden etwas Neues, nämlich dass unsere Wünsche relevant sind und deren Umsetzung durchaus sehr kurzfristig möglich ist.

Beide Punkte stellen etablierte Unternehmen vor große Herausforderungen. Die bereits angebotenen digitalen Prozesse basieren häufig auf bereits lange existenten analogen Geschwistern. Gleichmaßen sind die unterstützenden Entwicklungsprozesse nach klassischem Projektmanagement und betriebswirtschaftlichen Prinzipien gestaltet worden. Mit diesen Rahmenbedingungen ist es schwierig, die neue Selbstverständlichkeit der Kunden hinsichtlich Qualität und Umsetzungsgeschwindigkeiten zu bedienen. Dabei ist an den existierenden Produkten und den dazugehörigen Prozessen nicht unbedingt alles falsch und auch niemand schuld. Sie kommen lediglich aus einer anderen Zeit und hatten damals einen völlig anderen Fokus bei ihren Zielen und ihrer Entwicklung.

Dennoch, Unternehmen spüren den großen Druck des Marktes, schauen oftmals mit Verwunderung auf die Kleinen und versuchen durch Dienstreisen ins Silicon Valley herauszufinden, was transfor-

miert werden muss. Ein wiederkehrendes Muster dabei ist, dass die meisten der Kleinen mit agilen Vorgehensmodellen arbeiten. Entsprechend ist aktuell die Umstellung von Teams, Abteilungen und teilweise (fast) ganzen Firmen auf „Agile“ ein Trend. Ein Segen für die Branche der Unternehmensberater, für viele transformierende Unternehmen am Ende aber eine schmerzhaft und ernüchternde Reise.

Wieso erscheint Agile wie eine Silberkugel für Unternehmen?

Betrachten wir zur Beantwortung dieser Frage einmal das „Manifesto for Agile Software Development“ und seinen Ursprung [1]. Niedergeschrieben im Jahr 2001 von 17 erfahrenen Softwareentwicklern, verarbeiteten sie damit ihre Erlebnisse während der Dotcom-Blase. Damals wurde viel Geld in die Softwareindustrie gesteckt und es gab viele neue Ideen für Produkte. Oftmals waren dabei, unter anderem aufgrund der jungen Technologie und fehlender Erfahrung, die Kundenwünsche völlig ungewiss. Der Fokus lag primär auf der Produktentwicklung, die jedoch nach bis dato bekannten Best Practices durchgeführt wurde. Wer die Historie kennt, weiß, es ging deutlich schief und es wurden Unmengen an Geld verbrannt [2]. Einer der Gründe, wenn auch sicher nicht der einzige, war die Unfähigkeit der bekannten Projektmanagement- und Softwareentwicklungs-Methoden, in einem volatilen Markt mit großer Ungewissheit das zu liefern, was der Kunde benötigte. Das agile Manifest und alle darauf basierenden Vorgehensmodelle sind daher darauf ausgelegt, eine Produktentwicklung innerhalb eines volatilen Marktes im Rahmen eines iterativen Prozesses zu ermöglichen. Das Ziel ist es nicht, den Entwicklungsprozess generell schneller zu machen, sondern als Organisation die Fähigkeit zu entwickeln, Kundenwünsche Schritt für Schritt zu erforschen und zu bedienen. Die Stärke agiler Methoden liegt darin, Verschwendung in Form von Investition in unnötige Funktionen innerhalb der Produktentwicklung zu vermeiden und in kurzen Zyklen stetigen monetären Wert zu generieren.

Die heutige Marktsituation ähnelt der damaligen. Ständig gibt es neue Tech-

nologien und Funktionen, die Unternehmen zur Umsetzung ihrer digitalen Produkte nutzen können. Die Meinungen und Wünsche der Kunden ändern sich stetig. Was heute noch relevant ist, kann morgen schon veraltet sein. Unternehmen und ihre Mitarbeiter befinden sich zwangsweise in einem ständigen Lernprozess und können sich keine langwierigen Entwicklungszyklen leisten. Das Angebot und der Wettkampf sind zu groß. Eben genau für diese Situationen bieten sich agile Vorgehensmodelle an, da sie genau für diese Situationen entwickelt worden sind. Es wird eine qualitativ hochwertige und schnelle Produktentwicklung mit motivierten Teams und zufriedenen Kunden versprochen [3, 4, 5]. Wer würde da schon ablehnen? Dies gilt mittlerweile auch nicht mehr nur für den Bereich der Softwareentwicklung. In immer mehr Bereichen von Unternehmen halten agile Vorgehensmodelle Einzug. Hier werden Human Resources (HR) auf agile Prinzipien umgestellt und an anderer Stelle ein Team von Sachbearbeitern einer Versicherung umgekrempelt. Aber inwiefern passen agile Denkweisen und agile Vorgehensmodelle denn auf all diese Bereiche?

Jeder nimmt an der Messe teil, aber nur wenige wissen, warum...

An dieser Stelle soll erwähnt sein, dass der Autor dieses Artikels voll und ganz hinter einer agilen Denkweise und den verbundenen Vorgehensmodellen steht. Nimmt man sich der Entwicklung oder Weiterentwicklung eines neuen Produktes an, ist die Formulierung von Hypothesen bezüglich der Kundenwünsche und eine iterative Validierung eben jener Hypothesen unabdingbar für den Erfolg. Allerdings gibt es immer häufiger Situationen, in denen sich der Autor doch fragt, ob eine blinde Adaption von agilen Methoden an dieser Stelle sinnvoll ist.

Das klassische Beispiel für eine blinde Adaption ist der Start einer neuen softwarebasierten Produktentwicklung und könnte folgendermaßen ablaufen. Ein Unternehmen sieht sich den oben beschriebenen Herausforderungen gegenüber und möchte die Kundenwünsche bedienen. Dazu wird einem neu gegründeten Produktteam Freiheit im Sinne ei-

nes „grüne Wiese“-Projektes eingeräumt. Da gerade jemand, der etwas zu sagen hat, aus dem Silicon Valley oder von einer spannenden Konferenz zurückgekommen ist, wird beschlossen, das Projekt agil zu machen. Agil bedeutet Projektmanagement nach Scrum, was ein schlankes Framework ist, mit dem die Entwicklungsgeschwindigkeit gesteigert werden kann. Das Tolle daran ist, dass es viel einfacher ist als die schwergewichtigen Prozesse, die bisher eingesetzt werden. Die gesamte Beschreibung des Frameworks ist gerade einmal 14 Seiten lang und steht frei zum Download bereit. Zur Sicherheit werden noch ein bis zwei der Lead Developer auf eine Scrum-Master-Schulung geschickt. So können sie das Wissen an die übrigen Produktteammitglieder weitergeben und auch direkt die Rolle nebenbei wahrnehmen. Einer der erfahrenen Produktmanager wird die Rolle des Product Owner übernehmen. Da dieser bisher mit viel komplexeren Prozessen gearbeitet hat, sollte die Arbeit innerhalb eines agilen Projektes kein Problem sein. Außerdem kann er durch die Nutzung seiner bereits bewährten Methoden sogar eher noch zu besseren Ergebnissen beitragen und eine Integration in die bestehenden Unternehmensprozesse ermöglichen.

Dies liest sich zunächst wie eine sehr überspitzte Beschreibung der Situation, wird so oder ähnlich allerdings häufiger im Alltag vorgefunden. Vielleicht hat der ein oder andere Leser sogar ein paar Parallelen zur eigenen Realität erkannt? Fakt ist jedoch, dass viele agile Vorhaben und Transformationen daran scheitern, dass die aktuelle Situation innerhalb von Unternehmen nicht betrachtet wird und Teams oder Abteilungen einfach andere Prozesse ohne eine professionelle und erfahrene Begleitung übergestülpt werden. Sicherlich, Strategie und Strukturen sind notwendig für die Bildung von Kultur, aber man darf die Kultur eben auch nicht außer Acht lassen. Unternehmen können nicht darauf hoffen, dass sich Menschen einfach nebenbei ändern und plötzlich anders denken. Schon der britische Ökonom John Maynard Keynes schrieb 1936: „Die Schwierigkeit liegt nicht so sehr in den neuen Gedanken, als in der Befreiung von den alten [...]“ [6].

Auch die Einführung agiler Methoden und Werkzeuge in anderen Bereichen als der Softwareentwicklung birgt Tücken.

Die meisten Informationen zu diesen Themen, die man bei einer schnellen Internetrecherche findet, sind noch immer Theorien, Best Practices und Erfahrungen aus dem Bereich der Softwareentwicklung. Daher ist es ein grober Fehler, entsprechendes Wissen und Methoden einfach eins zu eins in andere Bereiche zu übernehmen. Sie müssen immer im Kontext betrachtet und an die Rahmenbedingungen, zum Beispiel im Bereich HR, angepasst werden. Die Herausforderung dabei ist, dass man hierfür die Konzepte im ursprünglichen Bereich verstanden und gemeistert haben und zusätzlich genügend Erfahrung im neuen Bereich mitbringen muss, um sie gemäß den neuen Rahmenbedingungen anpassen zu können. Es reicht eben nicht aus, nur Rituale zu kopieren und diese ohne entsprechendes Hintergrundwissen anzuwenden. Nur wer die Beweggründe dahinter richtig verstanden hat und interpretieren kann, ist in der Lage, wirklich etwas zu bewegen und die erhofften Ergebnisse zu erreichen.

Agilität überall um jeden Preis?

Die bisherige Diskussion hat gezeigt, dass agile Denkweisen und Vorgehensmodelle durchaus in der aktuellen Marktsituation den entscheidenden Vorteil bringen können. Sie ermöglichen innerhalb eines volatilen Marktes eine kurzfristige und stetige Erzeugung von Wert. Was Agilität jedoch nicht ist, ist ein Allheilmittel, das alle Probleme löst. Unternehmen müssen sich immer vor Augen halten, dass agile Vorgehensmodelle in erster Linie konstante Lernzyklen darstellen, mit denen man versucht, Komplexität und Ungewissheit zu beherrschen. Die Einführung agiler Vorgehensmodelle wird von allein weder die Entwicklungsqualität noch die Entwicklungsgeschwindigkeit erhöhen. Auch wird sie die Innovationskraft der Mitarbeiter nicht sprunghaft steigern, sondern lediglich einen Rahmen schaffen, um eine Denkweise zu etablieren, die eine konstante Reflektion und Weiterentwicklung fördert. Die den Status quo infrage stellt und Experimente zulässt, die vorher nie möglich gewesen wären.

Innerhalb der Produktentwicklung sind genau diese Fähigkeiten sehr rele-

vant. In anderen Bereichen aber wiederum auch nicht. Letzten Endes muss immer die aktuelle Aufgabe betrachtet und erst dann entschieden werden, welches Werkzeug das richtige ist. Es muss auch nicht zwangsweise immer die Einführung eines ganzen Vorgehensmodells sein. Manchmal reicht es auch schon, die aktuelle Situation zu analysieren, einzelne Methoden aus der agilen Werkzeugkiste zu nehmen und so für Verbesserung zu sorgen. So kann es teilweise schon sehr befreiend für ein Team sein, wenn regelmäßige Retrospektiven durchgeführt und die daraus entstehenden Maßnahmen umgesetzt werden. An anderer Stelle kann die durch ein Kanban Board erzeugte Transparenz innerhalb eines Teams oder einer Abteilung wahre Wunder hinsichtlich der Leistung dieser Einheit sorgen.

Wichtig bei jeglicher Veränderung von Strategie, Struktur oder Kultur ist immer, dass es niemals nur ein monetäres Investment ist. Gleichmaßen ist es immer auch ein Investment auf soziologischer Ebene. Neue Methoden und Werkzeuge einzuführen, war immer schon die einfache und schnelle Variante. Den beteiligten Menschen in den Mittelpunkt zu stellen und ihm die Möglichkeit und nötige professionelle Hilfe zu geben, seine über Jahre und Jahrzehnte gefestigten Denkweisen und Verhaltensmuster zu verändern, ist jedoch die eigentliche und viel mühseligere Aufgabe. Aber nur jene Unternehmen und Führungspositionen, die diesen Weg gemeinsam mit den Kollegen gehen, werden auf lange Sicht auf jedem Markt bestehen und immer neue Höhen erreichen können.

Quellen

- [1] Kent Beck et al. (2001): Manifesto for Agile Software Development, <http://agilemanifesto.org/>
- [2] Wikipedia (abgerufen 7. Januar 2021): Dotcom-Blase, <https://de.wikipedia.org/wiki/Dotcom-Blase>
- [3] Scrum.org (abgerufen am 7. Januar 2021): What is Scrum?, <https://www.scrum.org/resources/what-is-scrum>
- [4] Scaled Agile, Inc. (Dezember 2019): Achieving Business Agility with SAFe® 5.0, <https://www.scaledagile.com/?download=47510>
- [5] Matthew Heusser (2012, abgerufen am 7. Januar 2021): How the Kanban Meth-

od Changes Software Engineering, <https://www.cio.com/article/2393677/how-the-kanban-method-changes-software-engineering.html>

- [6] John Maynard Keynes (2017): Allgemeine Theorie der Beschäftigung, des Zinses und des Geldes - Aus dem Englischen neu übersetzt von Nicola Liebert, Duncker & Humblot, Berlin

Über den Autor

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.



Carsten Wiesbaum
carsten.wiesbaum@esentri.com



Enterprise Cloud Readiness – was ist zu tun?

Nisanth Muthukirushnasamy, Trivadis

Technisch betrachtet ist eine Cloud-Plattform im Handumdrehen eingeführt. Damit ein Unternehmen allerdings langfristig von ihr profitiert und nicht schneller, als es ihm lieb ist, mit ressourcenraubenden und kostspieligen Problemen konfrontiert wird, müssen neben den technischen auch die strategischen sowie organisatorischen Herausforderungen identifiziert und gemeistert werden.

Die Vorteile und die einfache Einführung einer Cloud-Plattform haben sich längst über die IT-Abteilungen hinaus herumgesprochen. Viele Fachabteilungen sind mittlerweile sogar in der Lage, ganz leicht und ohne Rücksprache mit der IT eigene Lösungen für sich aufzubauen. Doch diese stehen meist auf wackeligen Beinen und sind der Anfang einer Reihe von Problemen. Aber auch ganz allgemein kann das falsche Vorgehen beim Aufbau einer Cloud-Plattform Sicherheitslücken, Ausfälle von produktiven Applikationen und Systemen oder unkontrollierten Kostenzuwachs zur Folge haben.

Damit es gar nicht erst so weit kommt, braucht es bei der Einführung der Cloud ein strukturiertes Vorgehen, das abgesehen von technischen auch organisatorische und strategische Aspekte berücksichtigt. Dazu hat Trivadis ein eigenes Verfahren entwickelt, das im „Cloud Adoption Plan“ erarbeitet wird. Dieser teilt sich in die zwei Subphasen „Cloud Readiness“ und „Cloud Foundation“ auf.

Phase 1: Cloud Readiness prüfen

Bevor die Einführung einer Cloud-Plattform stattfindet, müssen demnach in der Phase „Cloud Readiness“ zuerst der IST-Zustand und die Anforderungen von verschiedenen Bereichen im Unternehmen geklärt und priorisiert werden. Zusätzlich zu den Anforderungen sollte auch eine Vision für und mit Cloud-Computing erarbeitet werden. Aus dieser sollte ebenfalls eine

Strategie für den Umgang mit den Cloud-Services abgeleitet werden können.

Die Aufnahme des IST-Zustands sowie der Anforderungen kann in vier Hauptbereiche (*siehe Abbildung 1*) und einen optionalen Bereich aufgeteilt werden:

1. Strategisch: Wie sehen die strategische Ausrichtung für die Nutzung von Cloud-Computing und die Zukunft der IT-Landschaft aus? Konkrete Fragen, die es dabei zu beantworten gilt, sind beispielsweise:
 - Haben Sie eine Firmenvision, die die Themen Digitalisierung, IT oder Cloud berücksichtigt?
 - Wollen Sie einen Cloud-Service-Provider oder wollen Sie mehrere einsetzen (Multicloud-Strategie)?
2. Business: Welche Schwierigkeiten, Einschränkungen und Anforderungen stellen sich Business-Abteilungen, um moderne Technologien aus der Cloud zu nutzen und ihre Prozesse zu optimieren? Konkrete Fragen, die es dabei zu beantworten gilt, sind beispielsweise:
 - Gibt es vertragliche Vereinbarungen, die eine Nutzung der Cloud erschweren oder verhindern können?
 - Gab es schon Fälle, bei denen Businessvorhaben aufgrund von IT-Mittel-Limitationen nicht umgesetzt werden konnten?
3. Organisatorisch: Wie ist die aktuelle IT-Organisation aufgebaut? Ist die Organisation noch stark auf einzelne Infrastruktur-Bereiche ausgelegt oder schon

Richtung BI-Modale-IT beziehungsweise auf Service/Produkten oder Lösungen basierend organisiert? Konkrete Fragen, die es dabei zu beantworten gilt, sind beispielsweise:

- Gibt es einen Ausbildungsplan für das Team, um die notwendigen Skills zu erhalten?
 - Sind die Rollen des Informationssicherheits-Beauftragten und des Datenschutz-Beauftragten besetzt?
4. Technisch: Welche Voraussetzungen sind bereits für die technische Integration gegeben und welche müssen noch geschaffen werden? Konkrete Fragen, die es dabei zu beantworten gilt, sind beispielsweise:
 - Sind Anforderungen für eine zentrale Monitoring-Lösung vorhanden?
 - Werden IT-Security-Komponenten genutzt, um das Netzwerk abzusichern, zum Beispiel Firewall, Web Application Firewalls, Proxy etc.?
 5. Cloud Journey (Optional): Hat die Organisation bewusst oder vielleicht unbewusst (Schatten-IT) schon mit der Nutzung von Cloud Services gestartet? Falls bewusst gestartet, wo steht man jetzt? Konkrete Fragen, die es dabei zu beantworten gilt, sind beispielsweise:
 - Haben Sie bereits erste PoCs bei einem oder mehreren Cloud-Service-Providern durchgeführt?
 - Werden Cloud-Services genutzt, sind aber nicht in die bestehende Unternehmensorganisation integriert?

Was sind die Stolpersteine?

- Keine ganzheitliche Aufnahme des IST-Zustands, fehlende Informationen über den aktuellen Zustand der IT-Landschaft
- Fehlende Angaben zu Governance, Compliance oder Security
- Fehlendes Know-how über das Cloud-Computing oder über die einzelnen Cloud-Provider sowie die Möglichkeiten bei den jeweiligen Providern
- Der Versuch, alle relevanten Themen für Cloud Adoption von Anfang an richtig und vollständig aufgegleist zu haben
- Cloud-Computing an die On-Premises-Welt anpassen, womit die Vorteile von Cloud-Computing von Anfang an eliminiert werden

Phase 2: Mit der Cloud Foundation eine Grundlage zur Lösung erarbeiten

Die Cloud Foundation legt unabhängig vom Provider Baustein für Baustein die Grundlage für eine stabile, ausbaufähige und sichere Infrastruktur in der Cloud.

Dabei werden alle Governance-, Security- und Compliance-Anforderungen des Unternehmens von Beginn an berücksichtigt.

Die Cloud Foundation gliedert sich in die drei Bausteine „Governance“, „Core-Infrastruktur“ und „Operations“ (siehe Abbildung 2).

Der erste Schritt besteht darin, im Baustein „Governance“ alle nötigen Richtlinien und Leitplanken festzulegen und folgende Fragen zu beantworten:

- Wie werden die Ressourcen innerhalb der Cloud-Umgebung organisiert?
- Welche Richtlinien gelten in Sachen Sicherheit?
- Wie sieht die Auditierbarkeit der Cloud-Umgebung und einzelner Ressourcen aus?
- Wie wird die Kostenkontrolle gewährleistet?

Je nach Cloud-Provider stehen unterschiedliche Hilfsmittel für die Umsetzung der festgelegten Governance-Richtlinien und Leitplanken zu Verfügung.

Im zweiten Schritt geht es im Baustein «Core-Infrastruktur» darum, das Lösungsdesign aus den folgenden Komponenten zu erarbeiten und umzusetzen:

- „Identity & Access Management“ für die Verwaltung der digitalen Identitäten und deren Zugriff auf die Ressourcen
- „Connectivity und Network“ für die Anbindung der lokalen Rechenzentren und die Umgebung in der Cloud sowie auch das Netzwerkdesign in der Cloud-Umgebung
- „Security Management“ für die Überwachung, Verwaltung und Steuerung der Sicherheitsaspekte

Gleichermaßen oder für eine nachhaltige Cloud-Lösung sogar noch wichtiger als die Evaluation, Konzeption und Umsetzung der beiden Bausteine „Governance“ und „Core-Infrastruktur“ sind die betriebliche Zuordnung, die Optimierung und die Weiterentwicklung der Lösungen. Um diese sicherzustellen, müssen die IT-Service-Management-Prozesse an die neuen Lösungen angepasst oder neu eingeführt werden. Der dritte Baustein „Operations“ beinhaltet daher die folgenden Bereiche:

- „System Management“, um die betrieblichen Prozesse festzulegen und durchzuführen.

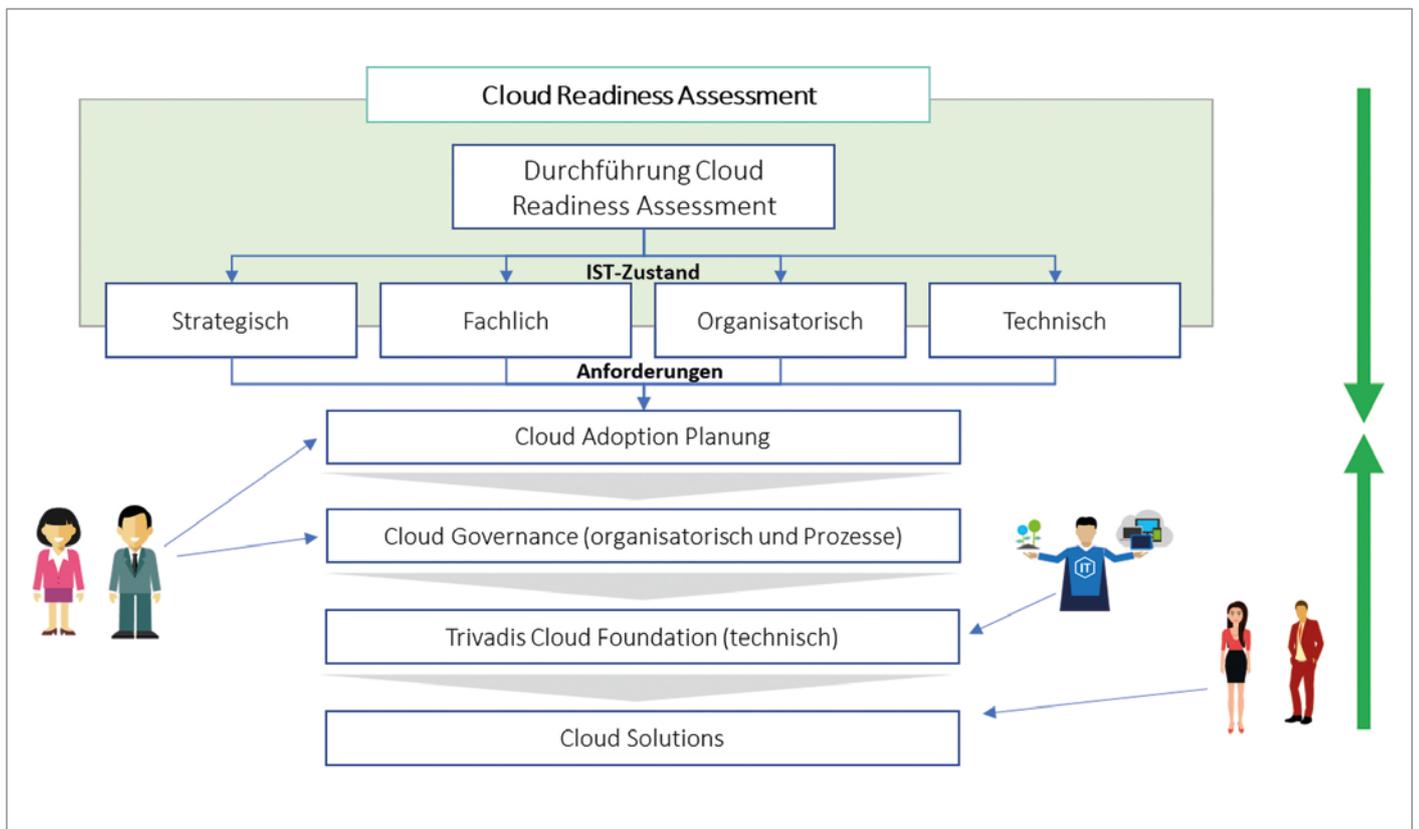


Abbildung 1: Trivadis-Cloud-Adoption-Vorgehensmodell (Quelle: © Nisanth Muthukirushnasamy, Trivadis AG)

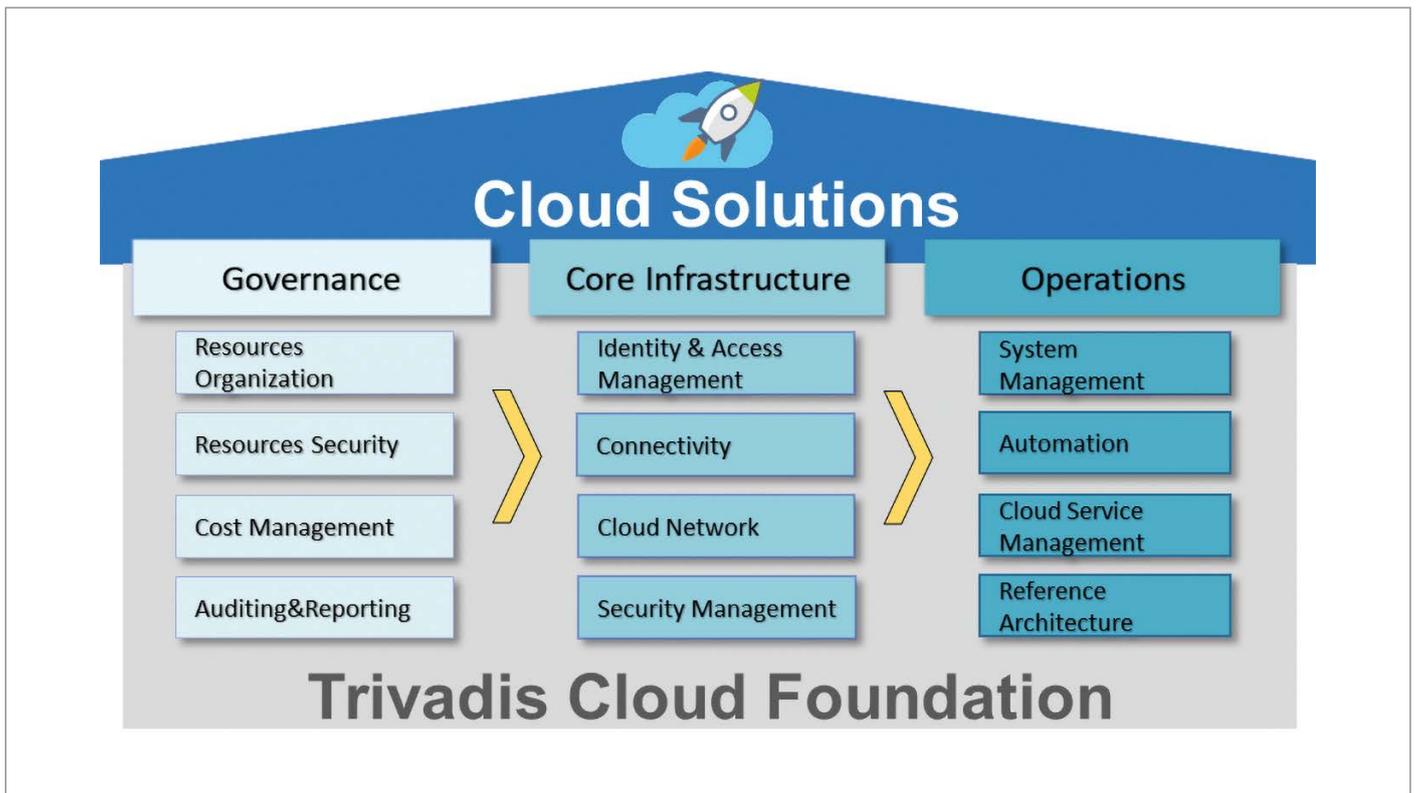


Abbildung 2: Trivadis Cloud Foundation (Quelle: © Nisanth Muthukirushnasamy, Trivadis AG)

- „Automatisierung“, um die betrieblichen Prozesse, soweit möglich, zu automatisieren und den Betriebsaufwand zu reduzieren.
- „Service Management“, um den Lifecycle der Services zu verwalten und neue Anforderungen zu bearbeiten.

**Zusammenfassung:
Wie wird es in der Praxis gemacht?**

- Kleines Team für Cloud Adoption bereitstellen.
- Allgemeines Wissen zur Cloud und spezifisches zum Provider aufbauen.
- Den IST-Zustand und die Anforderungen strukturiert aufnehmen.
- Ein erstes Business-Projekt als treibende Kraft für die Cloud Adoption identifizieren, wenig kritisch und geringe Abhängigkeit.
- Die ersten relevanten Themen aus der Cloud Foundation umsetzen, das kann je nach Organisation sehr unterschiedlich sein, zum Beispiel Governance MVP.
- Mithilfe laufender Projekte und Migrationen die Themen aus der Cloud Foundation vervollständigen beziehungsweise

- ausweiten, dazu auch die notwendigen organisatorischen Anpassungen vornehmen.
- Das Ganze in kurze, aber effektive Iterationen planen und umsetzen.

Für Interessierte bietet Trivadis ein Cloud Readiness Assessment sowie auch alle beschriebenen Dienstleistungen der Cloud Foundation an.

Über den Autor

Nisanth Muthukirushnasamy arbeitet als Principal Consultant im Bereich Cloud-Architektur bei der Trivadis AG. Dabei bereitet er Kunden technisch sowie organisatorisch auf die Einführung einer Cloud-Plattform vor und konzipiert und implementiert mit ihnen zusammen spezifische Lösungen, insbesondere im Bereich Microsoft Azure. Außerdem unterstützt er IT-Abteilungen dabei, ihre Rolle als Managed-Service-Provider und -Broker wahrzunehmen und für Business-Abteilungen innovative und passende Technologien bereitzustellen.



Nisanth Muthukirushnasamy
Nisanth.Muthukirushnasamy@trivadis.com



Mit Lift & Shift in die Oracle Cloud

Stefan Kühnlein, Opitz Consulting Deutschland

Im November 2019 erhielten wir einen Anruf: Ein österreichischer Kunde bat uns, ihn bei der Migration eines On-Premises-Rechenzentrums in die Oracle Cloud Infrastructure (OCI) zu unterstützen und den anschließenden Betrieb zu übernehmen. Auf Basis der zur Verfügung gestellten Unterlagen erarbeiteten wir ein erstes Grobkonzept. Im Juni 2020 war es so weit, wir bekamen den Zuschlag für das Projekt und die Migration konnte starten. In diesem Artikel stellen wir wichtige Architektur Aspekte und praktische Herausforderungen vor, die uns bei diesem Migrationsprojekt begegnet sind.

Unser Kunde, das Spinout eines europäischen Logistikdienstleisters und seit acht Jahren in Folge im Magic Quadrant von Gartner gelistet, stand kurz vor einer Umfirmierung. Dafür galt es, die virtuellen Server sowie die zentrale Exadata aus dem derzeitigen Rechenzentrum in eine Public Cloud zu überführen. Für die Migration von bestehenden Anwendungen, die in einer On-Premises-Umgebung laufen, können je nach identifizierten Handlungsbedarfen eine oder auch eine Kombination von unterschiedlichen Migrationsstrategien angewendet werden.

Folgende Migrationsstrategien haben wir dazu im Vorfeld betrachtet:

- **Lift & Shift**
Migration der Anwendung auf eine neue Laufzeitplattform mit nur minimalen Änderungen am Code. Codestruktur, Features und Funktionen werden nicht geändert.
- **Rearchitect**
Erhebliche Änderungen des Codes und der Architektur, um die Funktionen der neuen Anwendungsplattform voll auszunutzen.
- **Rebuild**
Neuentwicklung der Anwendung von Grund auf, wobei bisherige Anforderungen

und die Spezifikation beibehalten werden.

Aufgrund des kurzen Zeitfensters – die Migration sollte bis Ende 2020 abgeschlossen sein – entschied sich der Kunde für eine Lift-&-Shift-Migration.

Die Ausgangslage

Die zentrale Anwendung, um die es in diesem Migrationsprojekt ging, besteht aus einer Reihe von unterschiedlichen Modulen, die jeweils in einer separaten JBoss-Instanz bereitgestellt werden. Das Web Front-

end für die Module wird mit einem Apache-Server abgebildet. Dieser nimmt die Requests der Webanwendung entgegen und leitet sie mithilfe des Apache-JServ-Protokolls (AJP) an die jeweiligen Module weiter. Neben eines Webzugangs besteht zusätzlich die Möglichkeit des Datenaustauschs mittels FTP, SFTP, OFTP2, AS2 oder JMS. Unterstützt wird die zentrale Anwendung durch eine Reihe von zusätzlichen Anwendungen. Insgesamt werden für die unterschiedlichsten Anwendungsfälle wie Entwicklung, Test, Schulung, Demo, Produktion usw. bis zu 15 unterschiedliche Umgebungen betrieben.

Die On-Premises-Architektur stellte sich zu Beginn des Projekts wie folgt dar: Apache-Server und virtuelle Server für den elektronischen Datenaustausch befinden sich innerhalb der DMZ und stellen das Web Frontend der jeweiligen Module bereit. Als Firewall und Loadbalancer für den eingehenden Traffic wird eine BigIP F5 Appliance eingesetzt. Die Anbindung der Endanwender und Partner erfolgt entweder direkt über das Internet oder mithilfe von dedizierten Standleitungen.

Die JBoss-Instanzen der Anwendungsmodul sowie die unterstützenden Anwendungen werden in einem separaten Netzwerksegment betrieben. Die zentrale Speicherung aller Daten erfolgt in zwei Exadata Appliances, die über zwei physi-

kalische Rechenzentren und als Active-Standby-Cluster betrieben werden.

Abbildung 1 zeigt einen Auszug der On-Premises-Architektur.

Herausforderung 1: Die neue Netzwerkarchitektur

Die Erstellung einer wohldefinierten Netzwerkarchitektur ist der erste und zugleich wichtigste Schritt beim Entwurf einer Cloud-Architektur. Beim Design des Virtual Cloud Network haben wir die folgenden Anforderungen berücksichtigt:

- Anbindung des On-Premises-Rechenzentrums über FastConnect
- Anbindung der Partner über dedizierte Standleitungen
- Leitung des eingehenden Traffics komplett über die Firewall der BigIP F5
- Trennung der Produktionsumgebung von der Test- und Entwicklungsumgebung
- Berücksichtigung des Berechtigungskonzepts

Diese Anforderungen führten uns zu einer sogenannten Hub-Spoken-Netzwerkarchitektur mit Transitrouting [1]. Mit diesem Netzwerksetup war es während der Migration möglich, die Compute-Instan-

zen oder Services der OCI aus der On-Premises-Umgebung direkt zu adressieren.

Wie wir das Setup für das Transitrouting zwischen dem On-Premises-Netzwerk und den VCNs in der Cloud konfiguriert haben, zeigt *Abbildung 2*.

Bei der Definition der Subnetze mussten wir drei Punkte berücksichtigen:

- Abbildung der bestehenden Zonen
- Alle Compute-Instanzen, die einer beziehungsweise mehreren Umgebungen zugeordnet werden können, liegen im gleichen Subnetz
- Alle Compute-Instanzen, deren Software Services für alle Umgebungen anbieten, liegen in einem separaten Subnetz

Daher entschieden wir uns, alle Subnetze als regionale Subnetze aufzusetzen. Regionale Subnetze erstrecken sich über alle Availability-Domänen (AD) einer Region und bieten daher eine höhere Flexibilität als AD-spezifische Subnetze.

Herausforderung 2: BigIP F5

Im On-Premises-Rechenzentrum wurde eine Big-IP Appliance von F5 als Web Application Firewall und Loadbalancer eingesetzt (vgl. Gesamtarchitektur, *Abbildung 1*). Um si-

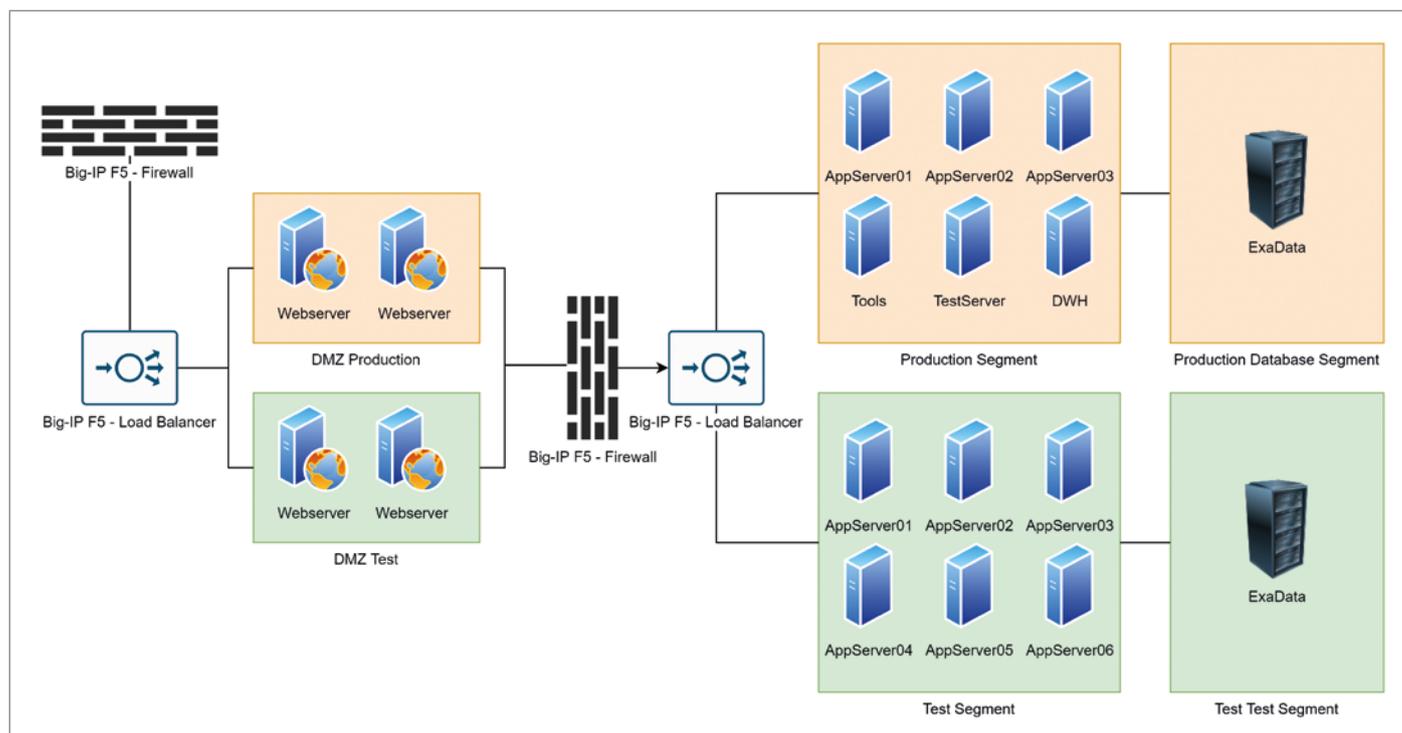


Abbildung 1: Auszug der On-Premises-Architektur des Rechenzentrums vor der Migration (Quelle: Stefan Kühnlein)

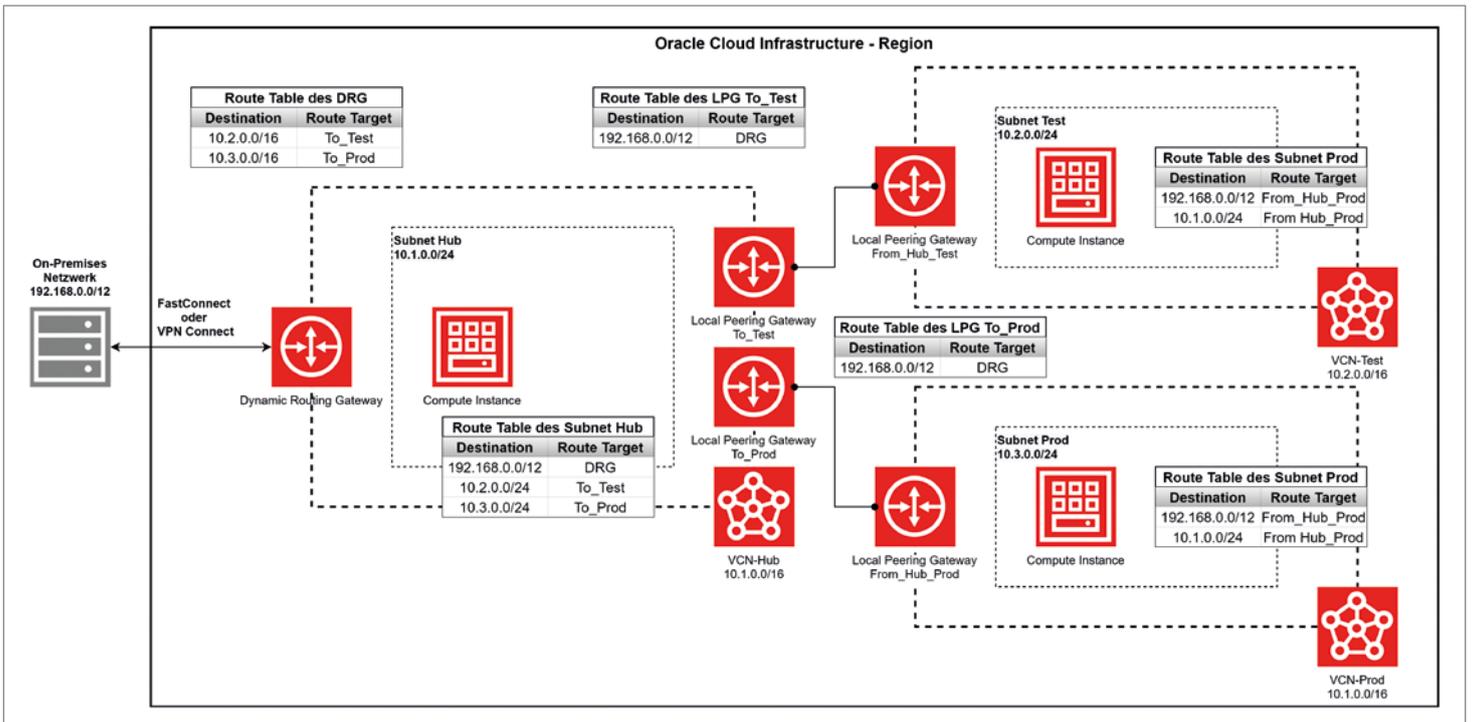


Abbildung 2: Setup einer Hub-Spoke-Netzwerkarchitektur (Quelle: Stefan Kühnlein)

herzustellen, dass die bestehenden Regeln nahezu 1:1 übernommen werden können, war es wichtig, in der Cloud ebenfalls eine Big-IP einzusetzen. Für den Betrieb einer Big-IP in der Oracle Cloud stellt F5 das fertige Image einer Big-IP Virtual Edition (Big-IP VE) [2] mit zwei Varianten bereit; die erste Variante für den Betrieb einer Big-IP VE mit einer Netzwerkkarte, die zweite Variante mit drei zusätzlichen Netzwerkkarten.

In der Big-IP VE sind folgende Komponenten verfügbar:

- BIG-IP Local Traffic Manager (LTM)
- Access Policy Manager (APM)
- Anwendungsfirewall (ASM)
- Netzwerkfirewall (AFM)
- DNS

Für die Lift-&-Shift-Migration wählten wir die Big-IP VE mit drei zusätzlichen Netzwerkkarten. Durch die Verwendung dieses Image

werden die Netzwerkkarten in unterschiedliche Subnetze installiert und ermöglichen somit eine Trennung der internen und extern zugänglichen Netzsegmente in der Cloud. *Tabella 1* gibt eine Übersicht über die benötigten Subnetze, die Zuordnung der jeweiligen Netzwerkkarten und die Zuordnung des Interface der virtuellen Netzwerkkarte (VNIC) in der Big-IP-VE-Konfiguration.

Jeder öffentliche oder private Endpunkt, der durch einen virtuellen Server in der Big-IP VE abgebildet werden soll, benötigt im Netzwerk Big-IP Public eine eigene zusätzliche private IP-Adresse im Netzwerk Big-IP Public. Soll diese IP-Adresse öffentlich zugänglich sein, so muss diese in der OCI noch mit einer öffentlichen IP-Adresse verknüpft werden.

Um die Ausfallsicherheit der Big-IP VE sicherzustellen, muss eine zweite Compute-Instanz mit der gleichen Konfiguration in einer anderen Availability Domain be-

reitgestellt werden. Lediglich die zusätzlichen IP-Adressen dürfen nur an eine der beiden Instanzen gebunden werden. Fällt eine Big-IP VE-Instanz aus, erfolgt ein Switch Over auf die zweite Big-IP VE. Hierbei werden alle zusätzlichen IP-Adressen und deren zugehörigen virtuellen Server auf die zweite Instanz übertragen.

Herausforderung 3: Webserver-Migration

Alle eingehenden HTTP(s)-Requests einer Umgebung werden durch einen dedizierten Apache Server verarbeitet. Die Apache Server der einzelnen Umgebungen sind – bis auf die Produktionsumgebung – in einer virtuellen Maschine installiert. Da alle Apache Server auf Ports 80 oder Port 443 hören, benötigt jeder Apache Server seine eigene IP-Adresse.

Netzwerk	Subnet	VNIC/Interface	Netzwerkkarte	Zusätzliche IP-Adressen
Big-IP Public	public	1.1	Erste zusätzliche Netzwerkkarte	Bis zu 31 weitere IP-Adressen
Big-IP Management	public oder private	1.0	Primäre Netzwerkkarte der ComputeInstance	Keine
Big-IP Internal	private	1.2	Zweite zusätzliche Netzwerkkarte	Keine
Big-IP High Availability	private	1.3	Dritte zusätzliche Netzwerkkarte	Keine

Tabella 1: Übersicht der Netzwerkkarten der Big-IP VE

```
#cloud-config
write_files:
  - content: |
      # Configure private IP's for Persistence
      DEVICE="ens3:0"
      BOOTPROTO=static
      IPADDR=10.130.10.11
      NETMASK=255.255.255.0
      BROADCAST=10.130.10.255
      ONBOOT=yes
    path: /etc/sysconfig/network-scripts/ifcfg-ens3:0
    permissions: '0644'
    owner: root:root
```

Listing 1: Persistente Konfiguration mit cloud-init

Um die Compute-Instanzen bei der Lift-&-Shift-Migration analog in die Cloud zu überführen, wurden entsprechende Compute-Instanzen erzeugt und der primären VNIC noch weitere zusätzliche sekundäre private IP-Adressen zugewiesen. Nach dem Hinzufügen einer weiteren Netzwerkadresse an einer Compute-Instanz muss diese für die Verwendung im Betriebssystem konfiguriert werden. Dies erfolgt mit dem Befehl:

```
ip addr add 10.130.10.11/24 dev ens3 label ens3:0
```

Damit die Konfiguration der zusätzlichen Netzwerkadressen über den Neustart hinweg persistent zur Verfügung steht, werden die Informationen in einer Konfigurationsdatei im Verzeichnis `/etc/`

`sysconfig/network-scripts/ifcfg-<phys_dev>:<adrs_seq:num>` gespeichert. Bei Deployments mit Infrastructure-as-Code kann diese Datei direkt beim Setup der Compute-Instanz über cloud-init [3] erzeugt werden. Das Listing 1 zeigt, wie eine weitere Netzwerkadresse mit cloud-init persistent konfiguriert wird.

Herausforderung 4: Migration der Anwendungsmodule

Die Kernanwendung unseres Kunden basiert, wie bereits erwähnt, auf einer Reihe von JEE-Modulen, die in einem JBoss

bereitgestellt werden. Jedes Modul wird über einen separaten Port angesprochen. Damit die Anwendungen bei einem Systemausfall verfügbar bleiben, wurden sie im On-Premises-Rechenzentrum auf zwei virtuellen Servern installiert, die als Active-Active-Cluster konfiguriert sind. Die Lastverteilung zwischen den JBoss-Instanzen erfolgte durch eine weitere Big-IP.

Die Session-Replikation zwischen den JBoss-Instanzen in die Cloud zu überführen, war nicht ganz trivial. Der Austausch der Informationen über die jeweiligen Sessions erfolgte über Multicast und eine spezielle Multicast-Adresse. In IPv4 ist hierfür der Adressbereich 224.0.0.0 bis 239.255.255.255 reserviert. Da dieser Bereich in der Cloud nicht zur Verfügung steht, musste die bestehende Multicast-Session-Replikation durch eine Unicast-Session-Replikation ersetzt werden.

Eine weitere nicht einfache Aufgabe bestand darin, bei der Migration der Anwendungsmodule die bestehende Big-IP zwischen den JBoss-Modulen durch einen OCI-Loadbalancer auszutauschen. In der On-Premises-Umgebung wurde in der Big-IP ein virtueller Server konfiguriert, der Anfragen an die entsprechenden Module weiterleitet. Je nach Komplexität der Umgebung wurden hier bis zu 28 verschiedene Module und Ports angesprochen. Für

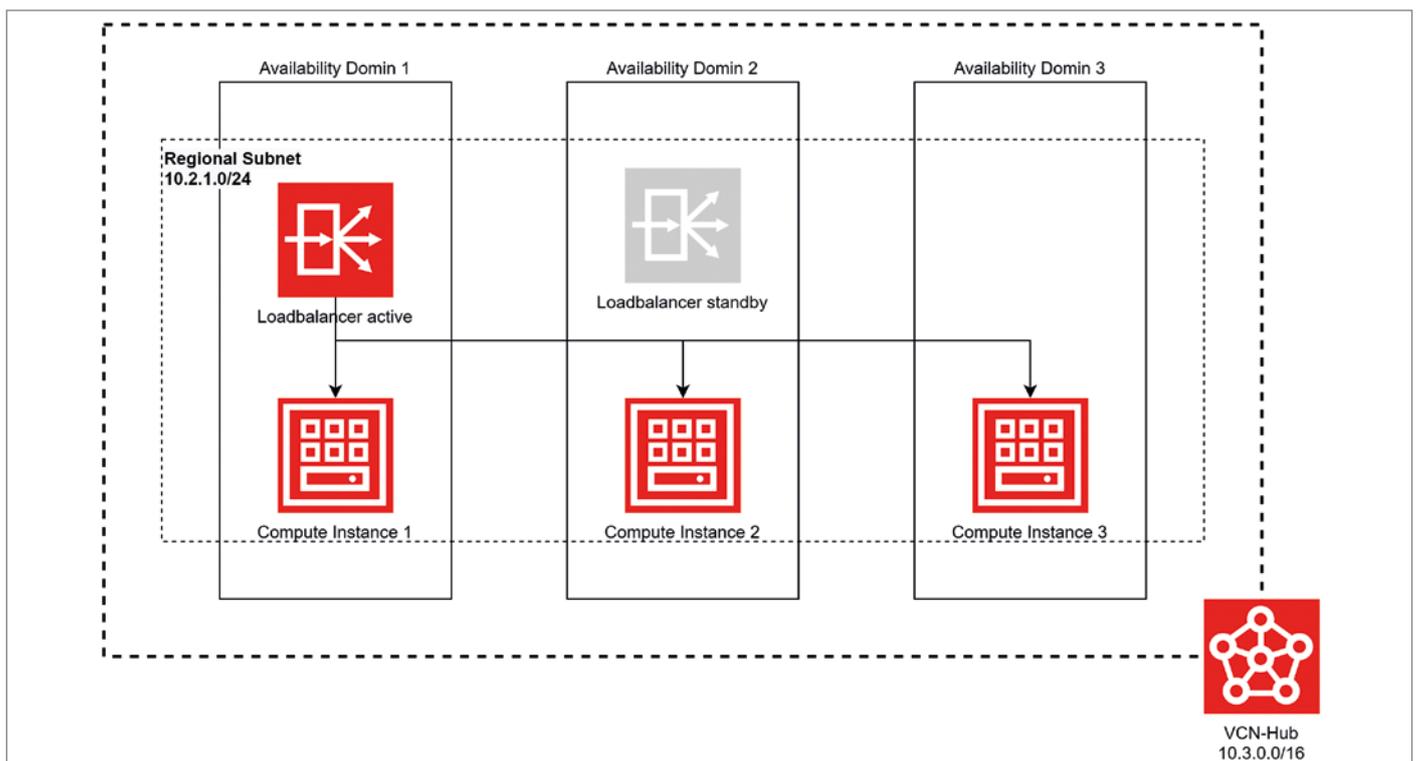


Abbildung 3: Subnetz und Compute-Instanzen für die Module (Quelle: Stefan Kühnlein)

Typ	Quelle	IP-Protokoll	Port	Zustandslos	Beschreibung
Ingress	10.2.1.0/24	TCP	80	Nein	Eingehende HTTP-Nachrichten
Ingress	10.2.1.0/24	TCP	443	Nein	Eingehende HTTPS-Nachrichten
Egress	10.2.1.0/24	TCP	80	Nein	Ausgehende HTTP-Nachrichten
Egress	10.2.1.0/24	TCP	443	Nein	Ausgehende HTTPS-Nachrichten

Tabelle 2: Ingress- und Egress-Regeln des Subnetzes

einen Loadbalancer in der OCI können jedoch maximal je 16 Listener und Backends konfiguriert werden. Der bestehende Big-IP Loadbalancer musste somit durch insgesamt 26 OCI Loadbalancer ersetzt werden. Ein OCI Loadbalancer wird bei der Erstellung immer als hochverfügbarer Loadbalancer ausgelegt. Daher muss dieser entweder in einem regionalen Subnetz oder in zwei Availability-Domain-spezifischen Subnetzen installiert werden.

An dieser Stelle sei auch noch erwähnt, dass jeder OCI Loadbalancer drei private IP-Adressen benötigt, die nicht von außen definiert werden können. Das ist bei der Auslegung der Subnetze also zu berücksichtigen. Im Rahmen der Cloud-Architektur wurden Compute-Instanzen und Loadbalancer im gleichen regionalen Subnetz bereitgestellt (siehe Abbildung 3). Auch wenn sich die Compute-Instanzen und der Loadbalancer im gleichen Subnetz befinden, kann der Loadbalancer die eingehenden Requests nicht an die jeweilige Compute-Instanz weiterleiten. Damit dies möglich wird, sind die Ingress- und Egress-Regeln des Subnetzes entsprechend zu setzen. Tabelle 2 zeigt, wie dies im Einzelnen aussehen kann.

Herausforderung 5: Erstellung konsistenter Backups

Die Erstellung von Backups zur Sicherung der Daten ist ein essenzieller Bestandteil des Rechenzentrumsbetriebs und darf auch in der Cloud auf keinen Fall vernachlässigt werden. Für die Sicherung der Daten von Block Volumes oder Boot Volumes stellt die OCI zwei unterschiedliche Verfahren bereit:

- **Manuelle Backups**

Manuelle Backups sind einmalige On-Demand-Backups, die entweder über die Konsole der OCI oder über das API der OCI gestartet werden.

- **Policy-basierte Backups**

Bei dieser Variante werden die Backups automatisch nach einem regelmäßigen Zeitplan durchgeführt. Der gewünschte Zeitplan für die Durchführung der Backups sowie für die Aufbewahrungsdauer der Backups wird in einer sogenannten Backup-Policy definiert.

Bei beiden Backup-Verfahren kann zwischen zwei Backup-Typen gewählt werden:

- **Inkrementell**

Bei der Erstellung eines inkrementellen Backups werden alle Änderungen seit dem letzten Backup gesichert.

- **Vollständig**

Bei der Erstellung eines vollständigen Backups werden alle Änderungen seit der Erstellung des Volume gesichert.

Für ein manuelles oder ein Policy-basiertes Backup in der OCI wird ein Point-In-Time-Snapshot der Daten auf dem zu sichernden Block-Volume erstellt. Im Anschluss werden ausschließlich diese Daten auf dem Block-Volume gesichert. Daten, die gegebenenfalls von den installierten Anwendungen verändert und noch nicht persistiert wurden, werden hingegen nicht mitgesichert. Die Backups sind daher also immer Crash-konsistent, nicht aber anwendungskonsistent. Anwendungskonsistente Backups müssen je nach Anwendungsfall und Verwendungszweck individuell umgesetzt werden.

Die Definition der Zeitpläne für eine Backup-Policy kann sehr flexibel gewählt werden. Die Oracle Cloud Infrastructure stellt hierzu die folgenden Möglichkeiten bereit:

- **Täglich:**

Backups werden täglich erstellt. Für die Auslösung des Backups kann eine bestimmte Stunde gewählt werden.

- **Wöchentlich:**

Backups werden wöchentlich erstellt. Für die Auslösung des Backups können ein Wochentag sowie die Stunde des Tages definiert werden.

- **Monatlich:**

Backups werden monatlich erstellt. Für die Auslösung des Backups können der Tag sowie die Stunde des Tages definiert werden.

- **Jährlich:**

Backups werden jährlich erstellt. Für die Auslösung des Backups können der Monat, der Tag sowie die Stunde für die Auslösung des Backups definiert werden.

Achtung: Die Definition einer Backup-Policy beschreibt lediglich, wann OCI-intern ein Auftrag zur Erstellung eines Backups generiert wird. Mit einer Backup-Policy kann kein garantierter Zeitpunkt für die Erstellung eines Backups definiert werden. Bei der Migration des Rechenzentrums unseres Kunden konnten wir beobachten, dass sich die Erstellung eines Backups über drei Stunden hinzog, während andere in viel kürzerer Zeit erledigt wurden, obwohl für alle zu sichernden Block-Volumes die gleiche Backup-Policy verwendet wurde. Die Backup-Policies scheinen also lediglich das Event zu generieren. Ein oder mehrere zentrale Backup-Prozesse verarbeiten daraufhin die Backups für alle Tenants in einer Region.

Herausforderung 6: Namensauflösung der Compute-Instanzen

Die Namensauflösung innerhalb der OCI erfolgt auf Ebene der VCNs durch einen internen DNS-Resolver. Der DNS-Resolver ist über die IP-Adresse 169.254.169.254

erreichbar. Jeder Compute-Instanz wird ein Fully Qualified Domain Name (FQDN) zugewiesen nach dem Schema host.subnet.vcn.oraclevcn.com. Der OCI-interne Resolver kann diesen Hostnamen anhand der FQDN auflösen und die IP-Adresse der Compute-Instanz ermitteln. Ebenso können interne Resolver auch öffentliche IP-Adressen wie www.opitz-consulting.com auflösen. Die Namensauflösung erfolgt jedoch nur innerhalb des VCN und nicht über die VCN-Grenzen hinweg.

Eine der wesentlichen Anforderungen, die bei der Migration berücksichtigt werden mussten, war die Ansprache der neuen Compute-Instanzen, die weiterhin über den bereits verwendeten internen Domainnamen funktionieren sollte.

Mit dem DNS-Service der OCI kann eine interne Zone aufgesetzt werden, in der die privaten IP-Adressen und der FQDN für die Compute-Instanzen eingetragen werden. Die autorisierenden Nameserver der privaten Zone werden dann über die

DHCP-Option an den jeweiligen Subnetzen registriert.

Diese Lösung hatte allerdings einen Nachteil: Es konnten keine öffentlichen IP-Adressen aufgelöst werden, weil der DNS-Service der OCI im Gegensatz zu den Cloud Services von Amazon und Microsoft keine hybriden DNS anbietet.

Dieses Problem konnten wir lösen, indem wir im VCN Database ein zusätzliches Subnetz einführt und zwei zusätzliche Compute-Instanzen in unterschiedlichen Availability Domains einrichteten. Auf diesen beiden Compute-Instanzen installierten wir dnsmasq, der als DNS-Forwarder zu den Nameservern der internen DNS-Zone konfiguriert ist. Die IP-Adressen des DNS-Forwarder trugen wir in den jeweiligen DHCP-Options ein (siehe Abbildung 4).

Mit dieser Architektur können sowohl die Compute-Instanzen aus den unterschiedlichsten VCNs über ihren FQDN angesprochen als auch öffentliche IP-Adressen ermittelt werden.

Weitere Details zu Private, Public und Hybrid DNS in der OCI können im Blog von Maximilian Jahn [4] nachgelesen werden.

Herausforderung 7: Migration der Oracle Exadata

Die Applikationsdaten waren im On-Premises-Rechenzentrum des Kunden in zwei Exadata Appliances gespeichert. Die Exadata Appliances standen räumlich getrennt und wurden als Active-Standby-Cluster betrieben.

In der OCI haben wir analog hierzu zwei Exadata Cloud Services mit dem OCI Shape Exadata.Quarter3.100 installiert und entsprechend konfiguriert. Die Datenübertragung zwischen der On-Premises-Datenbank und der Exadata in der Cloud erfolgte mit dem Zero Downtime Migration Tool (ZDM) [5] von Oracle. Mit diesem Tool können Datenbanken direkt in die Cloud migriert werden, sodass die

Unsere e-Services machen es einfacher.



Entdecken Sie unsere Remote-Services für

- Platform Audits
- License Reviews
- Workshops



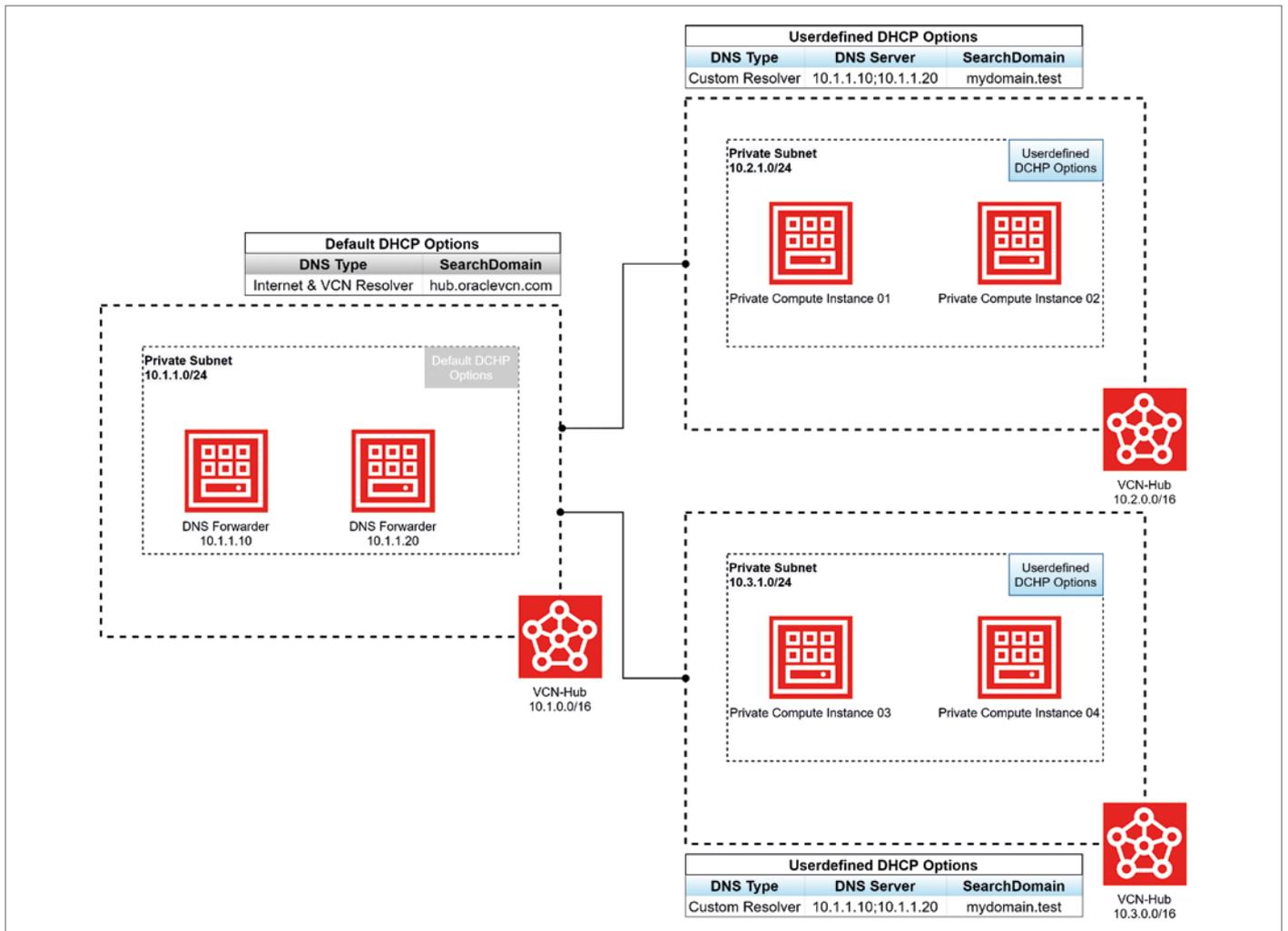


Abbildung 4: DNS-Forwarder (Quelle: Stefan Kühnlein)

Migration der Daten nur minimale bis gar keine Auswirkungen auf die Produktionsdatenbank hat.

Das ZDM-Tool hatte Oracle erst wenige Wochen vor unserem Projektstart aufgesetzt. Wir waren daher einer der ersten, der dieses Tool verwendete. Bei der Installation und der Einrichtung stießen wir auf eine Reihe von Anfangsproblemen, die wir zum Teil selbst beziehungsweise in enger Zusammenarbeit mit dem Entwicklungsteam in San Francisco beheben konnten.

Fazit

Die Migration eines On-Premises-Rechenzentrums via Lift & Shift in die Oracle Cloud ist eines von mehreren möglichen Szenarien, wenn es darum geht, ein Rechenzentrum in die Cloud zu bringen. Mit dem Lift-&-Shift-Ansatz ist dies schnell und einfach möglich. Doch auch bei dieser Migrationsvariante sind vorab eine Reihe von

Überlegungen zu treffen, um zu entscheiden, wie die zukünftige Architektur in der Cloud aussehen soll.

Quellen

- [1] <https://docs.cloud.oracle.com/en-us/iaas/Content/Network/Tasks/transitrouting.htm>
- [2] https://clouddocs.f5.com/cloud/public/v1/oracle_index.html
- [3] <https://cloud-init.io/>
- [4] <https://blog.maxjahn.at/2020/09/working-with-public-private-hybrid-dns-in-oracle-oci/>
- [5] <https://www.oracle.com/database/technologies/rac/zdm.html>

Über den Autor

Stefan Kühnlein ist bei der OPITZ CONSULTING Deutschland GmbH als Senior Solution Architekt im Bereich Software Development tätig. Der Schwerpunkt seiner derzeitigen Arbeit liegt im Entwurf von Microservice- und nativen Cloudarchitekturen sowie in der Auswahl moderner

Softwarearchitekturen und entsprechender technischer Frameworks. Zusätzlich beschäftigt er sich noch intensiv mit den unterschiedlichsten Cloud-Umgebungen und berät Kunden bei der Migration in die Cloud. Das in diesem Artikel beschriebene Projekt hat er als technischer Verantwortlicher maßgeblich mitgestaltet.



Stefan Kühnlein
stefan.kuehnlein@opitz-consulting.com



OCI Architecture Patterns

Michael Fischer, ORACLE Deutschland B.V. & Co.

Oracle bietet unter dem Begriff „Oracle Cloud Infrastructure Architecture Center“ eine Online-Sammlung von technischen Informationen und Beschreibungen zur Nutzung der Oracle Cloud. Dies erstreckt sich von How-Tos bis hin zu Schulungsunterlagen. Ein nahtloser Übergang zu Zertifizierungen mit weiterem Material von Oracle University ist möglich, wofür erst dann ein Login benötigt wird. Bis auf die Zertifizierungen ist die Nutzung zum Zeitpunkt der Artikelerstellung kostenfrei.

Die sogenannte Oracle Cloud Infrastructure wurde 2016 mit den klassischen Infrastructure-as-a-Service-Diensten gestartet. Seitdem wurden viele weitere IaaS-, PaaS- und SaaS-Dienste in über 30 Oracle-Cloud-Regionen weltweit bereitgestellt. Dank der umfangreichen Funktionalität können hochverfügbare, sichere und modernste Cloud-Anwendungen erstellt, konfiguriert beziehungsweise genutzt werden. Die umfangreiche Dokumentation im Web beschreibt je Service dessen Funktionalitäten. Um einen schnelleren und leichteren Einstieg zu ermöglichen, hat Oracle ein sogenanntes Architecture Center mit Referenzarchitekturen, Quick-Starters, Soluti-

on Playbooks, Getting-Started-Anleitungen, Self-Learnings und Blogs bereitgestellt. Der zentrale Link, über den die Themen erreichbar sind, lautet: <https://www.oracle.com/cloud/architecture-center.html>

Im Folgenden erläutere ich kurz den Inhalt der einzelnen Sammlungen:

Referenzarchitekturen

Die Bibliothek mit Referenzarchitekturen enthält einige der gängigsten Anwendungsmustern. Diese sind kurz beschrieben und mit einem Architekturdiagramm visualisiert. Zum Beispiel eine klassische,

datenbankbasierte Web-Applikation, die über Loadbalancer hochverfügbar ist und Auto-Scaling anhand der CPU-Auslastung durchführt. Die Datenbank ist ebenfalls hochverfügbar ausgelegt.

Alle Ressourcen werden in einer AD, einem Oracle Cloud Datacenter, in einer Oracle-Cloud-Region bereitgestellt. Die Region ist zum Beispiel Frankfurt. Die Loadbalancer, die Web-Tier und die Datenbank sind in separaten Subnetzen innerhalb eines einzelnen VCN isoliert. Security-Lists (=Firewalls) regeln den Netzwerkverkehr zu und von den Ressourcen in jedem Subnetz. An die Subnetze ist das Routing angehängt.

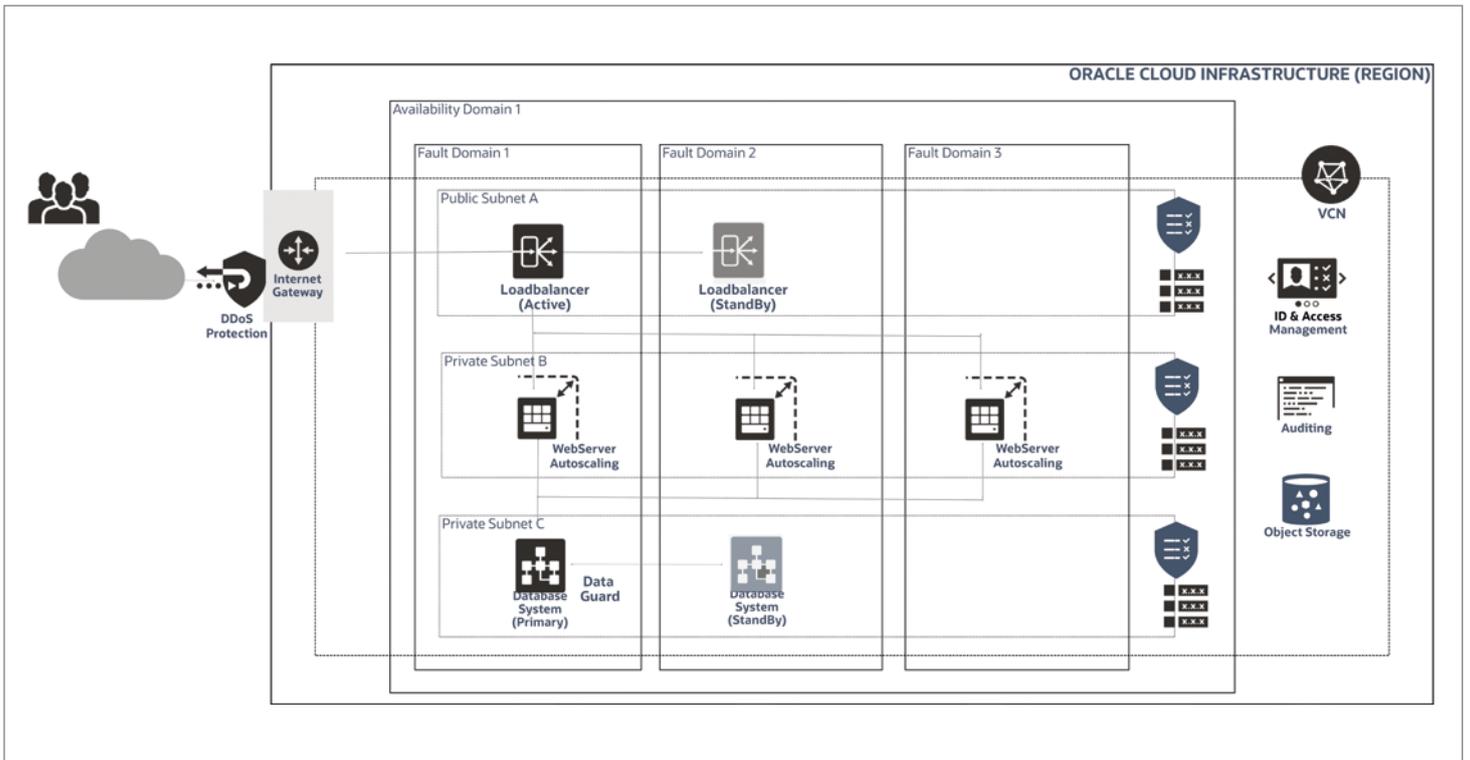


Abbildung 1: Oracle-Cloud-Referenzarchitektur (Quelle: Michael Fischer)

Das Diagramm in *Abbildung 1* zeigt diese Referenzarchitektur.

Die vollständige Beschreibung zu diesem Pattern ist hier zu finden: <https://docs.oracle.com/en/solutions/autoscale-webapp/index.html>

Weitere Patterns zu Datenbanken (Oracle, Microsoft, MySQL, MongoDB...), Kubernetes, Functions, WebSphere, Tomcat, Spark, HPC, Machine Learning, ... SaaS (ATG, HCM, EBS, Flexcube...) sowie technische Integrationen wie zum Beispiel Oracle-Azure Interconnect, VPN, Splunk-Anbindung oder OnPrem Backup sind ebenfalls zu finden. (siehe Beispiele unter <https://docs.oracle.com/en/solutions/index.html?type=reference-architectures>)

Quick-Starters: Terraform

Zu einigen Referenzarchitekturen gibt es Terraform Code, der heruntergeladen und als Ausgangsbasis verwendet werden kann. Terraform-Beispiele sind neben den Referenzarchitekturen auch für weitere Services/Szenarien verfügbar. Terraform kann nativ aus einer eigenen Umgebung oder im vorhandenen Cloud-Tenant über die enthaltene Cloud-Shell oder über den Oracle Cloud Infrastructure Resource Manager, einem auf Terraform basierenden UI, ausgeführt werden.

Der Code zum vorhergehenden Beispiel ist hier zu finden: <https://github.com/oracle-quickstart/oci-arch-web-ha>

Weitere Vorlagen enthalten Lösungen für Essbase, H2O, Kinetica, SAS, Splunk, Hortonworks, ... und erleichtern den Einstieg in die Verwendung der Oracle Cloud Infrastructure. Sie können diese Lösungen natürlich anpassen, da sie Terraform-bedingt im Source vorhanden sind. (siehe Beispiele unter: <https://github.com/oracle-quickstart>)

Mehr Informationen bezüglich Terraform: OCI Registry von Terraform: <https://registry.terraform.io/providers/hashicorp/oci/latest/docs>

OCI Registry von Terraform: <https://registry.terraform.io/namespaces/oracle-terraform-modules>

Zu Terraform und Oracle gibt es auf Git weitere Quellen, die über die Suche zu finden sind.

Solution Playbooks

Solution Playbooks sind zusammen mit den Referenzarchitekturen in einem Repository abgelegt. Die Playbooks haben den Charakter einer Erklärung von Themen, die zur Lösung führen, und sind damit eher Tutorials für das Eigenstudium.

Als Beispiel hier die Authentifizierung einer Python-Anwendung mit Oracle Identity Cloud Service. Dazu wird das Python SDK vom Oracle Identity Cloud Service verwendet. Dieses SDK erleichtert die Integration, da es alle erforderlichen REST Calls kapselt, um den Benutzer mit Oracle Identity Cloud Service mithilfe von OpenID zu authentifizieren. Der Oracle Identity Cloud Service bietet Identity Management, Authentifizierung beziehungsweise Single Sign-On (SSO) und Autorisierungen für eigene und Oracle-Cloud-Anwendungen. Der Service ist in der Oracle Cloud standardmäßig enthalten.

Anstatt eine Python-Webanwendung von Grund auf neu zu erstellen, stellt Oracle eine Beispielanwendung bereit, die natürlich nicht für den Produktiveinsatz gedacht ist. Diese Anwendung wurde mit dem Django-Framework von Python entwickelt.

Der Flow (authorization code grant type) in der Anwendung ist in *Abbildung 2* dargestellt.

Die Schritte dabei sind:

1. Der Benutzer fordert eine geschützte URL an.
2. Der Oracle Identity Cloud Service zeigt die Anmeldeseite an.

3. Der Benutzer gibt seine Anmeldeinformationen ein.
4. Oracle Identity Cloud Service gibt einen AuthZ-Token an die Python-Anwendung aus.
5. Die Python-Anwendung verwendet das SDK, um den AuthZ-Token gegen ein Access-Token auszutauschen.
6. Die Python-Anwendung zeigt den geschützten Inhalt für den Benutzer an.

Mehr zu diesem Beispiel, auch Erklärungen zum verwendeten OAuth Flow, unter folgendem Link: <https://docs.oracle.com/en/solutions/authenticate-sample-python-app-with-identity-cloud/index.html>

Weitere Beispiele umfassen Datenbankmigration, Chatbot, Private Access, WebLogic-Migration, Lösungen zu VMware, Cloudera, Essbase, Kubernetes, SaaS Extensions, PLM, Siebel usw. (siehe: <https://docs.oracle.com/en/solutions/index.html?type=solution-guides>)

Best Practices Frameworks

Oracle Cloud Infrastructure bietet mit den angebotenen Services eine Vielzahl von Funktionen und Konfigurationsmöglichkeiten. Oracle empfiehlt eine Reihe von

Best Practices zum Entwerfen und Betreiben von Cloud-Topologien. Diese sind in folgende vier Themen gruppiert:

- Security und Compliance mit User Authentication, Isolierung, Zugriffssteuerung, Sicherheit bei Compute, Sicherheit für Datenbanken, Schutz von Daten und Netzwerksicherheit.
- Reliability und Resilienz mit fehlertoleranten Netzwerkarchitekturen, Service Limits und Quotas, Backup und Skalierung.
- Optimierung bei Performance und Kosten mit Sizing bei Compute, Verwendung unterschiedlicher Speichertypen, Netzwerk-Monitoring und -Tuning, Kostentracking und Management.
- Operationale Effizienz mit Deployment-Strategien, Monitoring der Workloads, Management der Betriebssysteme und Support.

Link zu den Frameworks finden sich unter: <https://docs.oracle.com/en/solutions/oci-best-practices>

Self-Learning

Mit dem Material in diesem Bereich kann das Know-how zu Oracle Cloud aufge-

baut und auf Wunsch auch zertifiziert werden. Die Schulungsunterlagen, Beschreibungen, Folien, Videos und Labs sind teilweise kostenfrei, lediglich eine Zertifizierung/Prüfung ist mit Kosten verbunden. Es sind Lernpfade bezüglich der Zertifizierungen vorgeschlagen und auf der Webseite dokumentiert. Die Foundation-Zertifizierung überprüft lediglich das Verständnis bezüglich der Oracle Cloud Infrastructure, alle weiteren Zertifizierungen benötigen ein gewisses Know-how im praktischen Umgang mit der Oracle Cloud Infrastructure.

Angeborene Schulungen und Zertifizierungen sind im Folgenden aufgeführt.

Die Oracle Cloud Infrastructure **Foundation** vermittelt grundlegende Kenntnisse der wichtigsten Cloud-Computing-Konzepte und ein Verständnis dieser Dienste in der Oracle Cloud Infrastructure. Die Zertifizierung ist geeignet für Personen mit technischem und nichttechnischem Hintergrund, die ihr Verständnis der Oracle-Cloud-Infrastruktur überprüfen möchten.

Die **Associate**-Zertifizierung Oracle Cloud Infrastructure **Architect** bietet Cloud-Architekten ein Verständnis der Oracle Cloud Infrastructure Services.

Die Oracle Cloud Infrastructure **Operations-Associate**-Zertifizierung bestä-

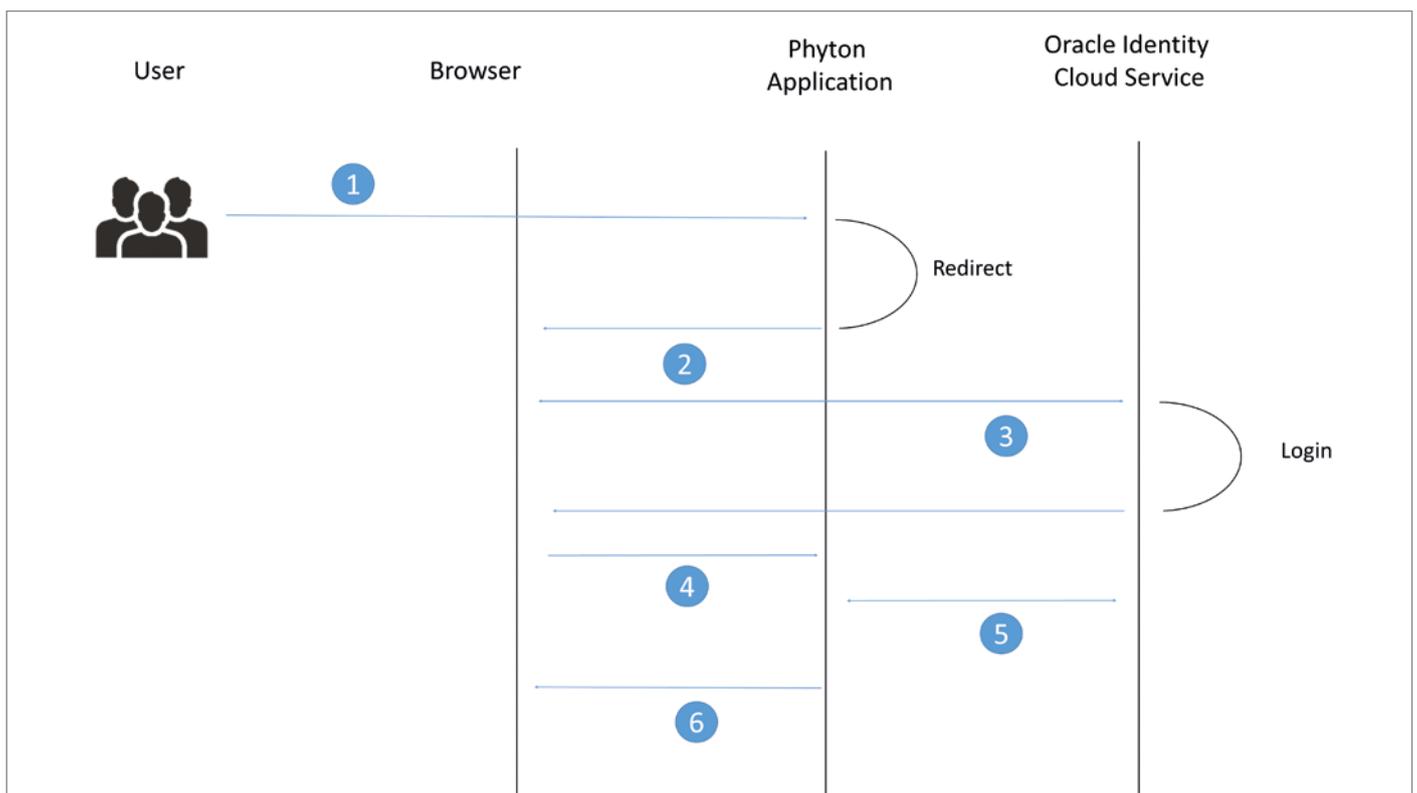


Abbildung 2: OAuth Flow Python-IDCS (Quelle: Michael Fischer)

tigt ein tiefes Verständnis der Verwaltung beziehungsweise des Betriebes von Oracle Cloud Services.

Die Oracle Cloud Infrastructure **Developer-Associate**-Zertifizierung zielt darauf ab, Cloud-native Anwendungen auf Oracle Cloud Infrastructure zu entwerfen beziehungsweise zu erstellen.

Die **Professional**-Zertifizierung Oracle Cloud Infrastructure **Architect** bestätigt ein tiefes Verständnis der Oracle Cloud Infrastructure Services und ist die praktische Vertiefung der Associate-Architect-Zertifizierung.

Link zu den Schulungen/Zertifizierungen finden sich unter <https://www.oracle.com/cloud/iaas/training/>

Blogs

Speziell zu Oracle Cloud gibt es einen von Oracle verwalteten Blog. Das sind in der Regel praktisch angewendete neue Cloud Services oder Funktionalitäten, wie etwa CloudGuard, HPC oder Erfahrungsberichte wie zum Beispiel SDN Disaster Recovery mit VMware. Dabei wird nicht nur der reine Cloud Service betrachtet, sondern auch das Aufsetzen beziehungsweise der Betrieb von Oracle-On-Premises-Produkten oder 3rd-Party-Produkten in der Oracle Cloud, wie zum Beispiel Fortinet oder Trend Micro. Beschreibungen rund um die Cloud wie beispielsweise CIS Benchmarks ergänzen die Aufstellungen. Wie bei Blogs üblich,

sind sie chronologisch sortiert und erlauben freie Suchen beziehungsweise das geordnete Filtern über die vordefinierten Kategorien Performance Solutions, Product News, Strategy, Cloud Workloads, Customer Stories, Partner Solutions, Developer Tools and Solutions, Security Strategies, Events, Technical Solutions, Public Sector and Government und Government Cloud Infrastructure.

Link zu den Blogs: <https://blogs.oracle.com/cloud-infrastructure/>

Geführter Einstieg zu Infrastructure as Code

Neben den aufgeführten Hilfestellungen und Informationsquellen gibt es eine Step-by-Step-Beschreibung, um mit der Oracle Cloud in das Thema Infrastructure as Code einzusteigen und dieses dann für eine echte Umsetzung als Basis weiterverwenden zu können. Dabei werden die von Oracle bereitgestellten OCI-Basismodule verwendet, die es erlauben, auch komplexe Enterprise-Architekturen aufzubauen und zu verwalten. Das Modulkonzept von Terraform erlaubt die „parametrisierte“ Verwendung von zentral bereitgestellten Funktionalitäten. Es kommen beim Tooling die bekannten Werkzeuge wie Terraform, Packer, Git usw. zum Einsatz. Ausprobieren lässt sich das auch mit dem sogenannten Free Tier, einer kostenfreien Testumgebung (siehe unter <https://www.oracle.com/de/cloud/free/>).

www.oracle.com/de/cloud/free/). Als Basis kann die im Free Tier enthaltene Cloud Shell genutzt werden, sodass außerhalb der Cloud keine Installation von Werkzeugen oder Umgebungen erforderlich ist. Der Link zur Beschreibung des Einstiegs und zum Repository ist unter folgendem Link erreichbar: <https://blogs.oracle.com/cloudtec-de/oraclecloudtooling>

Fazit

Mit dem bereitgestellten Know-how ist ein schneller Einstieg in die Oracle Cloud möglich. Der Free Trial der Oracle Cloud Infrastructure ermöglicht in dessen Rahmen Tests ohne kommerziellen Aufwand. Informationen zu den Trials inklusive Zugangsregistrierung zum Test finden Sie unter <https://www.oracle.com/de/cloud/free/>



Michael Fischer
michael.fischer@oracle.com

Die neuen Features von APEX 20.2

DOAG Online

Im Oktober wurde das neueste APEX-Release veröffentlicht, das eine Reihe von Neuerungen mit sich bringt.

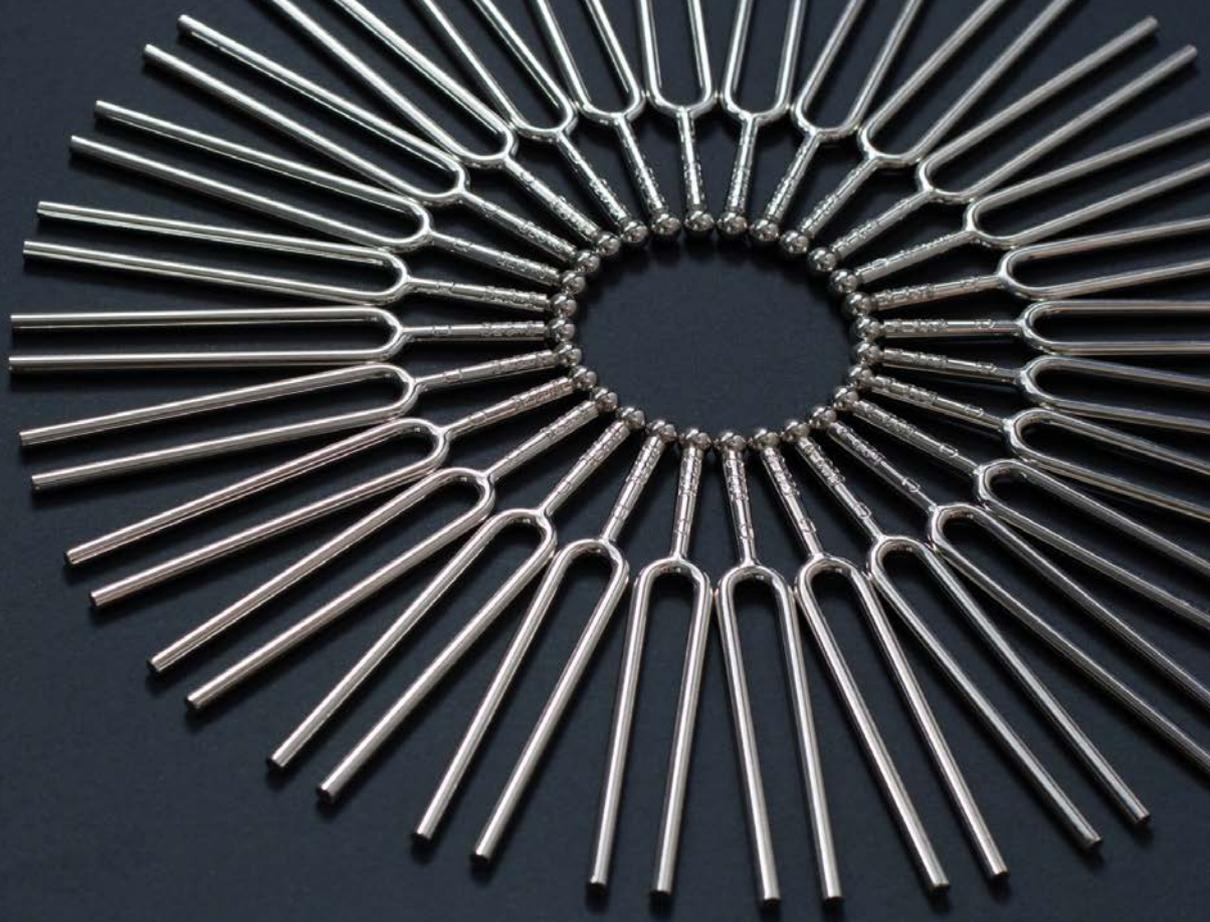
Carsten Czarski, seit 2016 Mitglied des APEX-Entwicklungsteams, gibt in der aktuellen ORAWORLD-Ausgabe #23 einen detaillierten Überblick über die neuen Möglichkeiten von APEX 20.2. Neben großen Neuerungen wie der Cards Region, den Automations oder den neuen Druckmöglichkeiten beleuchtet der Experte zudem

einige kleinere Verbesserungen und Features, die ab sofort APEX-Anwendungen bereichern und den Entwicklern die Arbeit erleichtern.

Der Artikel ist der mittlerweile sechste Teil von Carsten Czarskis APEX-Serie in der ORAWORLD. In der kommenden Ausgabe 24, die Anfang März erscheint, er-

wartet Sie dann der siebte Teil der Serie zum Thema "APEX PL/SQL APIs".

Die englischsprachige ORAWORLD ist ein kostenloses eMagazine, das vom EOUC herausgegeben wird und aktuell viermal im Jahr erscheint. Auf der ORAWORLD-Website (www.oraworld.org) finden Sie sämtliche Ausgaben zum Download.



Tuning einer Multitenant-Datenbank

Clemens Bleile, dbi services

Seit Oracle 19c darf man in Multitenant-Datenbanken bis zu 3 Pluggable DBs ohne Multitenant-Option-Lizenz nutzen. Dies und die Tatsache, dass in Oracle 20 die Non-Container-Architektur nicht mehr unterstützt wird, führt dazu, dass Container-DBs stärker verbreitet sind. In dem Zusammenhang stellt sich die Frage, was man beim Tuning von Multitenant-DBs beachten muss. Auch im Zeitalter autonomer DBs wird das Tuning von DBs und SQL-Abfragen nicht verschwinden. Dieser Artikel will sich dem Thema Tuning in einer Multitenant-Umgebung widmen.

Mit Oracle 12.1 wurde 2013 die Multitenant-Option eingeführt. Mit ihr kam der Begriff Container-DB (CDB) in die Oracle-Sprachwelt. Die Container-DB besteht aus mehreren Daten-Containern (Pluggable DBs, PDBs) und einem Root-Container, der die statischen Metadaten enthält. Den einzelnen Containern werden Identifikationsnummern zugewiesen, wobei die Container-ID 0 für die gesamte Container-DB steht. Die Container-ID 1 ist der Root-Con-

tainer, Container-ID 2 ist der Seed-DB vorbehalten, die als Template für neue Container fungiert. Ab Container-ID 3 spricht man dann von den Pluggable DBs, da diese wie USB-Sticks mittels „Unplug“ und „Plug-In“ in eine andere Container-DB „gesteckt“ werden können. Seit Oracle 19c ist die Verwendung von 3 Pluggable DBs lizenztechnisch in der Standard-Edition-2- und der Enterprise-Edition-Lizenz enthalten. Wenn man die Architektur (siehe Abbildung 1) be-

trachtet, so sieht eine Oracle-Instanz einer Container-DB genauso aus wie die einer Non-Container-DB. Der Unterschied zur Non-Container-Architektur wird eher durch die Dateien der DB deutlich. Jeder Container hat seine eigenen Tablespaces.

Aus Performance-Sicht stellen sich mit der geänderten Architektur einige Fragen: Wo werden Automatic-Workload-Repository-Daten gespeichert? In welchem Container kann ich Performance-Metriken

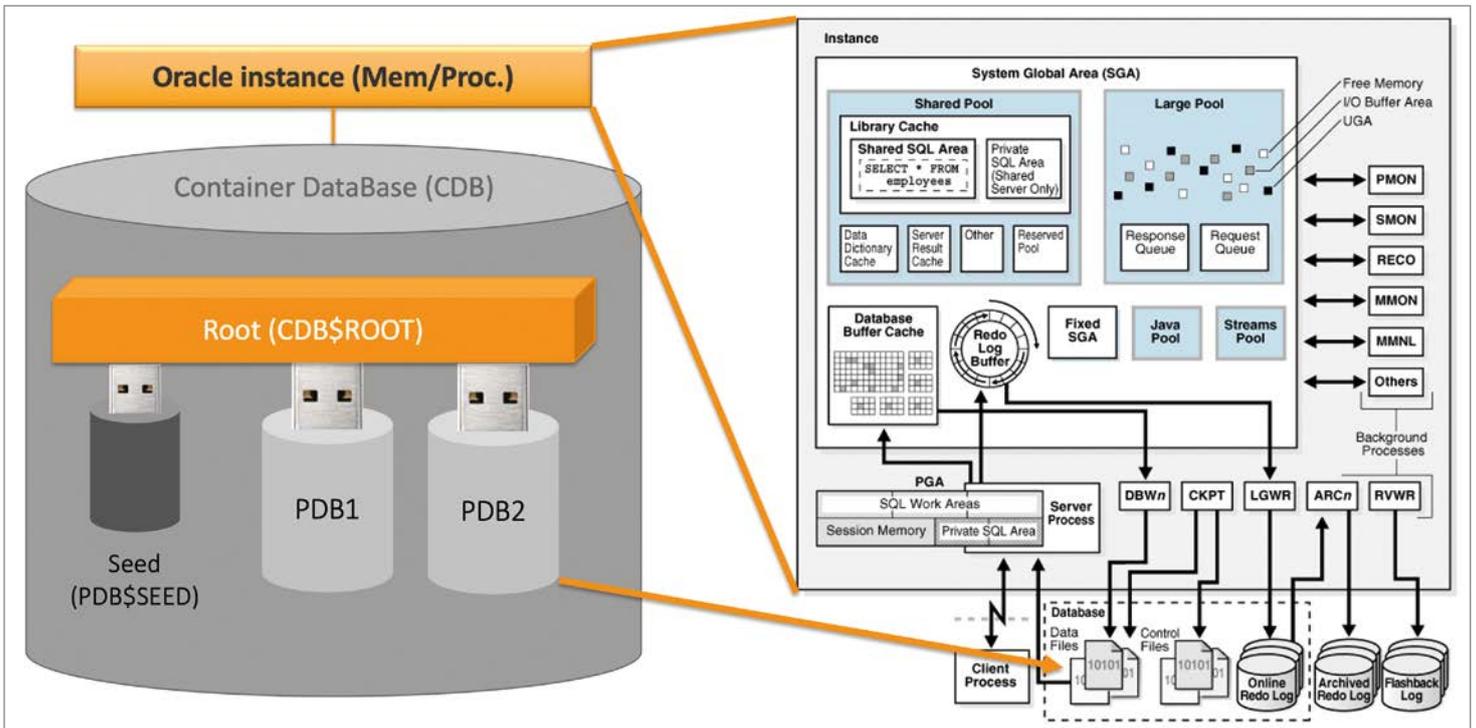


Abbildung 1: Multitenant-Architektur mit Instanz (© [Instance, Oracle-Dokumentation])

abfragen? Gibt es Konflikte im „Shared Pool“, wenn mehrere Pluggable DBs die gleichen SQL-Statements verwenden?

SQL_TRACE

Parameter, die im Root-Container systemweit gesetzt wurden, werden in den PDBs übernommen, falls dort nicht explizit ein anderer Wert gesetzt wurde. Das heißt beispielsweise, dass ein Kommando

```
alter system set sql_trace=true;,,
```

im Root-Container ausgeführt, zum SQL-Tracing in allen Containern führt. Die entstehenden Trace-Dateien stehen alle im gleichen Verzeichnis, das mit folgendem SQL ausgegeben werden kann:

```
SELECT value FROM v$diag_info
WHERE name = 'Diag Trace';.
```

Versucht man hingegen mit einem Kommando

```
alter system set events 'sql_trace[sql: 8n9f7w0d6mhu5]
level=12';
```

ein Tracing für das Statement mit SQL_ID 8n9f7w0d6mhu5 automatisch zu aktivieren,

wenn dieses Statement ausgeführt wird, so gilt dieses Setting nur für den Container, in dem man sich gerade befindet. Das bedeutet, dass im Root-Container das Setting nur für Statements gilt, die im Root-Container ausgeführt werden. Entsprechend verhält es sich, wenn zum Beispiel in einer PDB ein „alter system set events“ abgesetzt wurde. Es gilt nur im Kontext dieser PDB.

In den Trace-Dateien steht die Container-ID. Leider reicht es nicht, Trace-Dateien einer PDB zu identifizieren, da eine Trace-Datei Daten aus mehreren Containern (PDBs) enthalten kann. Zur Identifikation von Trace-Dateien wird geraten, mit dem TRACEFILE_IDENTIFIER zu arbeiten (z.B. alter session set tracefile_identifizier=slow_HR;). Alternativ können Trace-Daten einer PDB mit dem Werkzeug trcsess gefiltert werden, indem ein Service angegeben wird. In dem Zusammenhang sei angemerkt, dass man immer einen Applikations-Service erstellen und nicht den von Oracle standardmäßig erstellten PDB-Service nutzen sollte.

Setzt man tkprof ein, so hat der Parameter PDBTRACE nichts mit PDBs zu tun. Vielmehr erlaubt der Parameter PDBTRACE, Trace-Daten via SQL aus der DB statt vom Filesystem zu lesen.

Im Grunde genommen verhält sich SQL_TRACE so wie in Non-Container-

DBs. Man schaltet das Tracing auf PDB-Ebene ein und wandelt die gefilterten (trcsess) rohen Trace-Dateien mithilfe eines Werkzeugs wie tkprof in besser lesbare Form.

Automatic Workload Repository (AWR)

In der Standardeinstellung werden AWR-Snapshots im Root-Container ausgeführt und beinhalten die Daten aller Container. Mit der DB-Parameter-Einstellung AWR_PDB_AUTOFLUSH_ENABLED = TRUE auf der PDB kann man das explizite Erstellen von PDB-AWR-Snapshots aktivieren. Dies hat den Vorteil, dass ein PDB-Admin „eigene“ Snapshots zur Verfügung hat und diese beim Umzug in eine andere Container-DB nicht verloren gehen.

Die Erstellung eines AWR-Reports kann in 3 Varianten durchgeführt werden:

1. AWR-Report im Root-Container: Dieser enthält Daten aus allen PDBs.
2. AWR-Report in der PDB basierend auf den Snapshots im AWR_ROOT: Dieser enthält Daten der lokalen PDB und basiert auf den Snapshots des Root-Containers.
3. AWR-Report in der PDB basierend auf den Snapshots der AWR_PDB: Dieser

enthält Daten der lokalen PDB und basiert auf den Snapshots der lokalen PDB.

Variante 1 ist dem Container-DB-Admin vorbehalten und hat den Vorteil, dass der Report mehr Daten enthält:

- Buffer-Pool-Statistiken
- Advisory-Statistiken
- Latch-Statistiken
- Dictionary-Cache-Statistiken
- Library-Cache-Statistiken
- Operating-System-Statistiken
- Background Wait Events
- usw.

Enthalten sind also alle Arten von Statistiken, die global in der Container-DB gesammelt werden und nicht einer PDB zugewiesen werden können.

Die AWR-Reports aus Variante 2 und 3 sind nahezu identisch, können aber dadurch variieren, dass die Snapshots zu unterschiedlichen Zeitpunkten gezogen wurden. Wie schon erwähnt, werden lokale PDB-Snapshots (Variante 3) bei einem Umzug der PDB mitgenommen.

Active Session History (ASH)

Eine Abfrage auf v\$active_session_history im Root-Container zeigt die Daten aller Container. Entsprechende Einträge einer PDB werden mit der Spalte CONTAINER_ID identifiziert (siehe Abbildung 2).

Ein ASH-Report (ashrpt.sql), gestartet im Root-Container, zeigt die Top-Container, sortiert nach Aktivität an. Dies ist sehr nützlich, wenn bei hoher Last die verursachenden PDBs zu identifizieren sind.

```
SQL> select con_id, min(sample_time), count(*)
       2 from v$active_session_history group by con_id order by 1;
```

CON_ID	MIN(SAMPLE_TIME)	COUNT(*)
1	16-MAR-20 03.09.50.710 PM	1600
2	16-MAR-20 03.09.55.825 PM	1
3	16-MAR-20 03.09.56.849 PM	114
4	16-MAR-20 03.09.56.849 PM	62
5	16-MAR-20 03.09.56.849 PM	54
6	16-MAR-20 03.09.56.849 PM	51

6 rows selected.

Abbildung 2: Active-Session-History-Daten (Quelle: Clemens Bleile)

Führt man den ASH-Report in einer PDB aus, so werden einem wieder die Alternativen AWR_ROOT oder AWR_PDB zur Auswahl angezeigt. Unterschiede gibt es nur dadurch, dass man für die eine oder andere Variante eventuell eine längere Historie eingestellt hat beziehungsweise die Snapshot-Häufigkeit unterschiedlich sein kann. Der Unterschied ist aber nur dann relevant, wenn man auf historisierte ASH-Daten zurückgreifen muss (aus DBA_HIST_ACTIVE_SESS_HISTORY statt V\$ACTIVE_SESSION_HISTORY).

Automatic Database Diagnostic Monitor (ADDM)

Standardmäßig wird ADDM nach jedem „globalen“ Snapshot im Root-Container ausgeführt und enthält somit Hinweise für alle PDBs. Will man auch ADDM für die lokalen PDB-Snapshots ausführen, so müssen lokale Snapshots eingestellt sein (AWR_PDB_AUTOFLUSH_ENABLED = TRUE). Die ADDM-Reports der lokalen Snapshots enthalten aber wieder nur eine Sicht auf die PDB. Hinweise zu Instanz-weiten Proble-

men, wie in folgender Liste, werden nicht angezeigt:

- IO-Probleme
- Hard-Parsing-Probleme
- SGA-Sizing-Hinweise
- Log File Switch Waits
- Free Buffer Waits
- etc.

SQLTXPLAIN

Zum Werkzeug SQLTXPLAIN ist nicht viel zu sagen, außer dass es für jede PDB separat installiert werden muss. Auch die Auswertungen im Haupt-Report zeigen nur Daten der jeweiligen PDB an.

SQL Management Base (SMB)

Die SQL Management Base besteht aus SQL-Profilen, SQL Patches, dem SQL Statement Log (SQL IDs) und der SQL Plan History (SQL Plan Baselines). Daten der SMB werden lokal in der PDB gespeichert. Es

```
sys@orclcdb@CDB$ROOT> select con_id, child_number, plan_hash_value, sql_plan_baseline, sql_profile
       2 from v$sql where sql_id='7ws837zyp1zv' order by 1;
```

CON_ID	CHILD_NUMBER	PLAN_HASH_VALUE	SQL_PLAN_BASELINE	SQL_PROFILE
3	3	3722429161		
4	1	3722429161		
4	2	3722429161	SQL_PLAN_a5ymwanr0pxvrab13ea89	
5	0	2597291669		coe_7ws837zyp1zv_2597291669

4 rows selected.

Abbildung 3: Gleiche SQL_ID in mehreren PDBs mit unterschiedlichem SQL-Management (Quelle: Clemens Bleile)

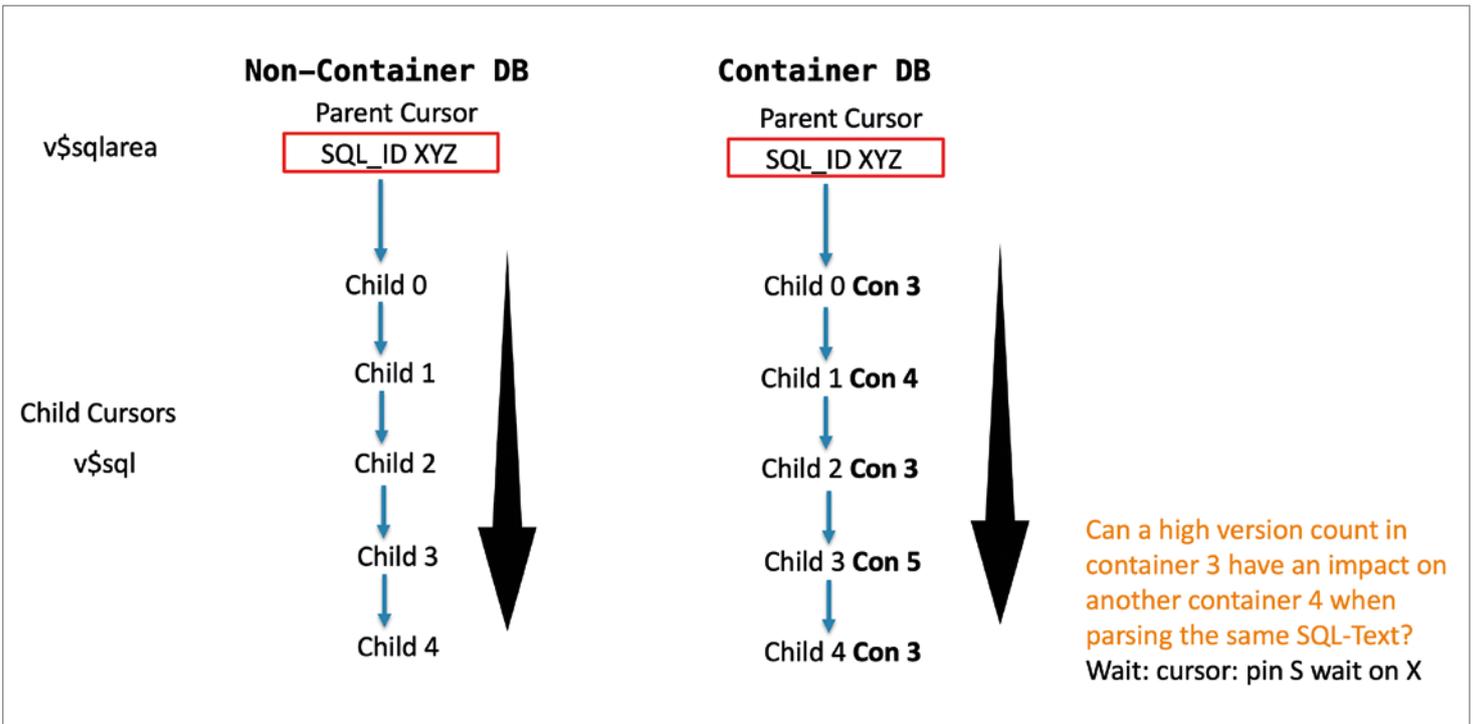


Abbildung 4: Cursor-Liste in Non-Container- und in Container-DB (Quelle: Clemens Bleile)

ist somit beispielsweise möglich, für eine SQL_ID in PDB1 einen SQL-Patch, in PDB2 ein SQL-Profil und in PDB3 ein SQL Plan Baseline aktiviert zu haben, ohne dass diese sich gegenseitig beeinflussen (siehe Abbildung 3).

Potenzielles Performance-Problem in einer CDB

Da es pro Parent-Cursor in einer Container-DB nur eine Child-Cursor-Liste gibt, ist es vorstellbar, dass es zu Performance-Problemen führen kann, wenn mehrere PDBs die gleichen SQL-Statements parsen (siehe Abbildung 4).

Frits Hoogland hat sich mit der Problematik auseinandergesetzt [1]. Gemäß seinen Tests sollten keine nennenswerten Probleme auftreten:

„The main reason for waiting for a cursor/library cache entry is a child pinned in exclusive mode during creation (‘cursor:pin S wait on X’). When an identical statement is executed in another container, it will skip a child that is created or in the process of being created in another container, even when it’s pinned in exclusive mode.

This means that the multi-tenant option in my opinion does not significantly increase the risk of waiting because of library cache concurrency, specifically because it can skip

child cursors pinned exclusively in another container.“

Um dies zu verifizieren, wurde ein Test-szenario aufgebaut:

Die Swingbench-Applikation wurde gegen 3 PDBs gleichzeitig ausgeführt. Zuerst ein Test mit Standardeinstellungen. Dies führte im globalen AWR-Report zu folgenden Top-Events (siehe Abbildung 5).

In einem zweiten Test wurden in PDB1 künstlich Child-Cursor erzeugt. Dies wurde erreicht, indem jede Session auf PDB1 mittels Logon-Trigger eigene OPTIMIZER-Einstellungen bekam. Im AWR-Report führte dies zu folgenden Top-Events (siehe Abbildung 6).

Optimal (low version count): 603 TPS

Event	Waits	Total Wait Time (sec)	Avg Wait	% DB Wait time Class
db file sequential read	23,770	317.1	13.34ms	42.3 User I/O
DB CPU		250.4		33.4
log file sync	74,500	61	818.68us	8.1 Commit
read by other session	17	7.1	419.92ms	1.0 User I/O
library cache: mutex X	1,075	4.8	4.42ms	.6 Concurrency
PGA memory operation	126,518	2.3	18.21us	.3 Other

Abbildung 5: Top-10-Events mit Swingbench gegen 3 PDBs (Quelle: Clemens Bleile)

High version count on PDB1: 601 TPS

Event	Waits	Total Wait Time (sec)	Avg Wait	% DB Wait time	Class
db file sequential read	10,943	648	59.21ms	60.8	User I/O
DB CPU		265.7		24.9	
log file sync	75,516	72.8	.96ms	6.8	Commit
cursor: pin S wait on X	1,991	36.9	18.51ms	3.5	Concurre
read by other session	12	13.2	1100.90ms	1.2	User I/O
library cache: mutex X	1,103	5.3	4.83ms	.5	Concurre

Abbildung 6: Top-10-Events mit Swingbench gegen 3 PDBs und künstlich hoher Child-Cursor-Anzahl in PDB1 (Quelle: Clemens Bleile)

Man sieht eine erhöhte Anzahl von „cursor: pin S wait on X“-Waits. Sollten diese sich auf PDB1 beschränken, so gäbe es keinen Einfluss eines hohen Version-Count in einer PDB auf eine andere PDB. Dies bestätigte sich. In den AWR-Reports von PDB2 und PDB3 waren keine „cursor: pin S wait on X“-Waits zu sehen. Alle Waits bezogen sich auf PDB1 (siehe Abbildung 7).

Tatsächlich lässt sich das Traversieren durch die Child-Cursor logisch so wie in

Abbildung 8 darstellen.

Somit konnte gezeigt werden, dass keine Shared-Pool-Probleme aufgrund gleicher SQL-Statements in unterschiedlichen Containern (PDBs) zu erwarten sind.

Sonstiges – Statistiken

- Globalgesetzte Statistik-Präferenzen (z.B. `dbms_stats.set_global_prefs(<METHOD_`

`OPT,»FOR ALL INDEXED COLUMNS SIZE AUTO»);`) gelten in der lokalen PDB. Global bedeutet hier also global für die jeweilige PDB.

- Jede PDB hat ihr eigenes Wartungsfenster für das Sammeln von Statistiken. Mit dem Parameter `AUTOTASK_MAX_ACTIVE_PDBS` (Default 2) kann definiert werden, wie viele PDBs gleichzeitig Statistiken sammeln, um die Auslastung besser zu kontrollieren.

PDB with low version count: Wait events

Event	Waits	Total Wait Time (sec)	Avg Wait	% DB Wait time	Class
db file sequential read	3,395	201.5	59.35ms	63.9	User I/O
DB CPU		74.5		23.6	
log file sync	25,262	25.2	1.00ms	8.0	Commit
read by other session	9	11.4	1263.16ms	3.6	User I/O
enq: TX - index contention	19	1.9	102.03ms	.6	Concurre
library cache: mutex X	326	1.2	3.79ms	.4	Concurre

PDB with high version count

Event	Waits	Total Wait Time (sec)	Avg Wait	% DB Wait time	Class
db file sequential read	3,810	221.9	58.23ms	53.8	User I/O
DB CPU		107.7		26.1	
cursor: pin S wait on X	1,990	36.9	18.52ms	8.9	Concurre
log file sync	24,902	22.7	913.49us	5.5	Commit
kksfbc child completion	87	4.5	51.64ms	1.1	Other
library cache: mutex X	457	2.9	6.43ms	.7	Concurre

Abbildung 7: Top-10-Events im AWR-Report der PDB2 (oben) und im AWR-Report der PDB1 (unten) (Quelle: Clemens Bleile)

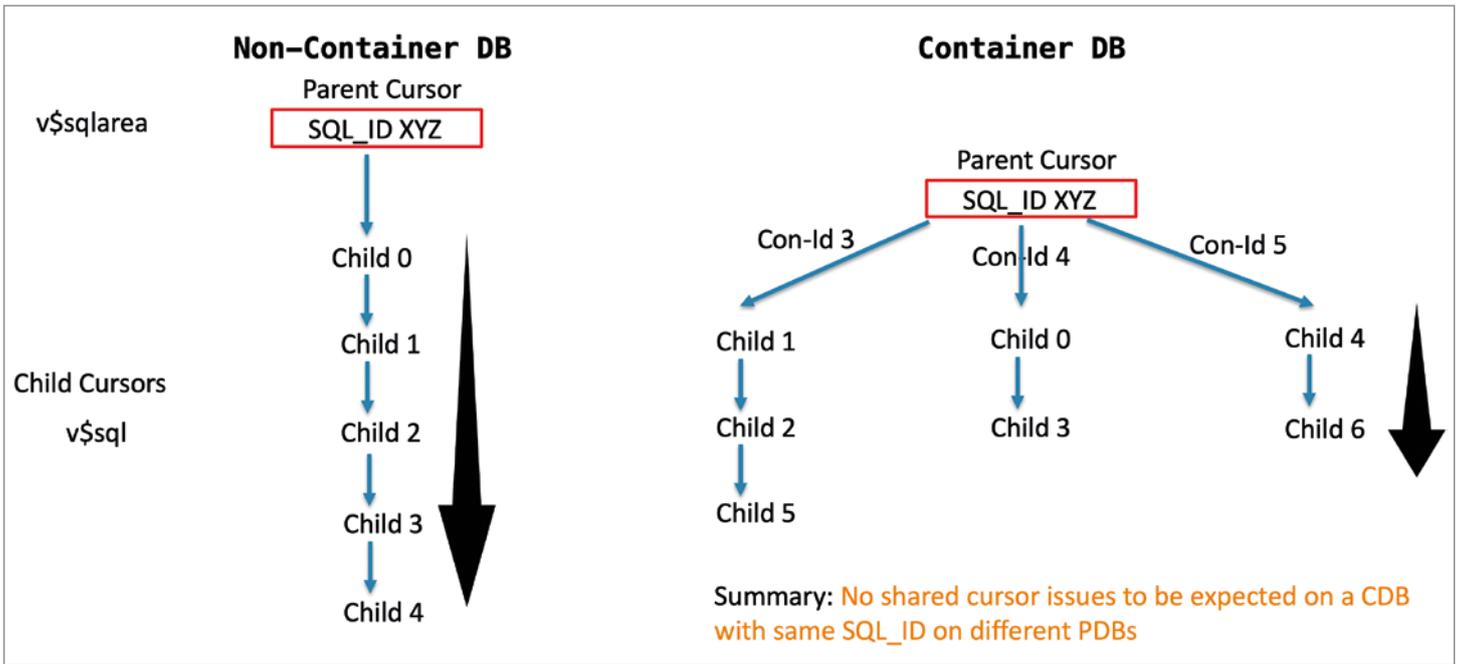


Abbildung 8: Logische Darstellung des Traversierens durch die Child-Cursor in einer Non-CDB und einer CDB (Quelle: Clemens Bleile)

Sonstiges – Ressourcen Manager

- Ab 12.2. kann mit den Parametern MAX_IOPS und MAX_MBPS die maximale IO-Last pro PDB kontrolliert werden.
- Ab 19c kann bei eingeschaltetem Ressourcen-Manager ein PDB-Caging mit den Parametern CPU_COUNT und CPU_MIN_COUNT aufgesetzt werden. Zusätzlich kann im Root-Container ein CPU_COUNT für das Instance-Caging sorgen.

Zusammenfassung

Tuning in einer Multitenant-DB ist dem Tuning in einer Non-Container-DB ähnlich. AWR/ADDM und ASH-Reports im Root-Container verschaffen einen globalen Überblick mit Statistiken, die auf PDB-Ebene nicht sichtbar sind (z.B. Instanz-Daten, Redo-Daten etc.). Insofern sollte der Container-DB-DBA seine Reports im Root-Container erstellen.

Auf PDB-Ebene kann sich der PDB-DBA eigene AWR-Snapshots erstellen und diese dann beim Umzug in einen anderen Container auch mitnehmen. Einige Tuning-Werkzeuge werden nur auf PDB-Ebene eingesetzt. Dazu zählen

- SQL Trace
- SQLTXPLAIN
- SQL Management Base

- dbms_stats und dessen globale Präferenzen

Erfreulich ist, dass keine Shared-Pool-Probleme zu erwarten sind, wenn mehrere PDBs die gleichen SQL-Statements parsen. Seit Oracle 12.2. wurden im Bereich Ressourcen-Management einige Features hinzugefügt, die es erlauben, PDBs besser zu kontrollieren.

Quellen

[1] Frits Hoogland, 2020: Oracle multi-tenant and library cache isolation. Blog, <https://fritshoogland.wordpress.com/2020/04/03/oracle-multi-tenant-and-library-cache-isolation/>

[2] Clemens Bleile, 2016: Enable 10046 Tracing for a specific SQL. Blog, <https://blog.dbi-services.com/enable-10046-tracing-for-a-specific-sql/>

[3] Tim Hall: Multitenant: Resource Manager PDB Performance Profiles in Oracle Database 12c Release 2 (12.2). Blog, <https://oracle-base.com/articles/12c/multitenant-resource-manager-pdb-performance-profiles-12cr2>

[4] Oracle-My-Oracle-Support-Dokument: High Version Counts For SQL Statements (>1024) Post Upgrade To 12.2 and Above Causing Database Slow Performance (Doc ID 2431353.1)

und 14 Jahre im Consulting. Er hat sich auf Oracle-Datenbank-Performance-Tuning (SQL Tuning, DB Tuning) und die Entwicklung von Oracle-DB-IT-Architekturen (hochverfügbar, wartungsarm, kosteneffiziente Speicherung von Daten) spezialisiert und ist Experte im Bereich der Problem-Analyse und -Lösung.



Über den Autor

Clemens Bleile hat mehr als 29 Jahre IT-Erfahrung, 13 davon im Oracle-Support

Clemens Bleile
clemens.bleile@dbi-services.com



Warum Postgres immer mehr an Bedeutung in IT-Landschaften gewinnt

Borys Neselovskyi, EnterpriseDB

In Zeiten der Digitalisierung stehen Unternehmen oft vor der Herausforderung, sich dem Markt schnell anzupassen, um wettbewerbsfähig zu bleiben. Die Anforderungen an Sicherheit und Datenschutz, Verfügbarkeit und Performance steigen. Kosten dafür sollen transparent gehalten und nach Möglichkeit reduziert werden. Dieses Dilemma, mehr Leistungen und Services anzubieten, ohne dafür die Kosten signifikant zu erhöhen, treibt Unternehmen, intelligente und innovative Lösungen zu suchen.

Die Begriffe „Open Source“, „DevOps“ und „Cloud“ sind nicht mehr wegzudenken. Firmen optimieren betriebliche Prozesse und Abläufe und sorgen für die Automatisierung von Standardaufgaben, zu denen Installation, Aktualisierung und Wartung von Infrastruktur-Komponenten gehören. Zahlreiche DevOps-Tools haben Open-Source-Natur. Sie werden von freien Communities gebaut und vorangetrieben. Ein Vorteil von freien Produkten ist die Tatsache, dass Kunden keine Lizenzkosten für deren Einsatz tragen. Die Verwendung von

Open-Source-Produkten bringt große Vorteile in Umgebungen, in denen die dynamische Skalierung von Ressourcen eine tragende Rolle spielt. Nehmen wir als Beispiel einen Webshop. Zu Weihnachtszeiten werden mehr Käufe erwartet und dafür müssen kurzfristig mehr IT-Ressourcen bereitgestellt werden. Die zusätzlichen Kapazitäten müssen extra lizenziert werden, wenn hier kommerzielle Software zum Einsatz kommt.

Anders und viel einfacher stellt sich der Betrieb mit der Einführung von Open-

Source-Produkten dar: Diese müssen nicht lizenziert werden. Die Endnutzer können unbesorgt zusätzliche Ressourcen einschalten, ohne Sorge zu haben, Lizenzregeln zu verletzen.

Warum wollen Unternehmen Postgres einführen und nutzen?

Wie sieht es in der Datenbank-Welt aus? Neben den bekannten kommerziellen Da-

tenbanken von Oracle, IBM und Microsoft mischen zunehmend einige Open-Source-Datenbanken im Markt mit. Die populärste und meistverbreitete Open-Source-Datenbank trägt den Namen PostgreSQL. Sie hat eine lange Geschichte und gehört neben Oracle und DB2 zu den ältesten relationalen Datenbanken, die alle ihren Ursprung in der ersten relationalen Datenbank namens SystemR haben. Die größte Datenbank-Community der Welt ist für die Entwicklung und strategische Ausrichtung der Postgres-Datenbank verantwortlich.

Die Postgres-Datenbank ist die am schnellsten wachsende Datenbank. Die nachfolgende *Abbildung 1* spiegelt die rasante Verbreitung der Datenbank-Engine wider.

Viele Unternehmen wollen Postgres einführen. Zahlreiche Kunden streben die Migration von kommerziellen Datenbanken nach Postgres an. Warum findet diese Entwicklung statt?

Am häufigsten sind die Gründe für einen solchen Wandel finanzieller Art. Die Lizenz- und Supportkosten bei herkömmlichen Datenbanken sind oft übersteuert. Die mit Kunden abgeschlossenen Verträge sind kompliziert formuliert, mehrere Seiten lang und beinhalten viele Passagen, die nicht eindeutig und für den Kunden oft nur schwer zu verstehen sind. Firmen wissen oft nicht, ob die gekauften Lizenzen für die aktuellen Umgebungen ausreichen. Die

von Softwareherstellern eingeordneten Lizenzaudits führen häufig zu Nachzahlungen, die hoch ausfallen können.

Unternehmen wollen klare Regeln und eindeutige Verträge. Beim Einsatz von Open-Source-Produkten fallen in der Regel keine Lizenzkosten an. PostgreSQL wird unter der PostgreSQL-Lizenz vertrieben. Diese liberale Open-Source-Lizenz kann frei erworben werden und ist mit BSD- beziehungsweise MIT-Lizenzen vergleichbar [1].

Viele Kunden verfolgen eine Cloud-Strategie, wollen die beste Lösung vom jeweiligen Anbieter nutzen und dabei frei entscheiden können.

Softwarehersteller stellen oft Bedingungen, die es schwer bis nahezu unmöglich machen, eigene Produkte bei anderen Cloud-Anbietern zu betreiben. Als Beispiel kann die Oracle-Datenbank genannt werden: Der Betrieb einer Oracle-Datenbank in der Oracle Cloud wird preiswerter, als wenn man die gleiche Datenbank in Cloud-Lösungen von Amazon oder Microsoft betreiben will. Oracle zählt die Prozessor-Lizenzen in eigener Cloud anders als in AWS oder in Azure [2]. Dies führt zu erhöhten Ausgaben für Oracle-Datenbank-Lizenzen, wenn diese in AWS oder Azure betrieben werden.

Der Einsatz von Open Source gibt Kunden die notwendige Freiheit und Flexibilität bei der Implementierung der zukünftigen

Plattform. Unternehmen genießen die Unabhängigkeit vom Softwarehersteller und können beste Produkte flexibel einsetzen.

Funktionalität der PostgreSQL

Die Postgres-Community hat eine Datenbank entwickelt, die eine breite Palette an Optionen und Eigenschaften bietet. Entwickler können bei der Implementierung von Datenbankanwendungen zahlreiche Datentypen, Funktionen und Routinen verwenden, die in der Postgres-prozeduralen Umgebung namens PL/pgSQL [3] ablaufen werden. So können zum Beispiel in Postgres Mengenoperationen auf einer oder mehreren Tabellen ausgeführt werden. Die Abfragen können sehr intelligent gestaltet werden und eine oder mehrere Subqueries oder Joins enthalten.

Zahlreiche in Postgres implementierte Treiber (zum Beispiel: jdbc, odbc, node.js und andere) ermöglichen die Kommunikation zwischen der Datenbank und Middleware-Schichten. PostgreSQL bietet Schnittstellen zu mehreren Programmiersprachen, darunter C, C++, Python und Java. So können in Postgres beispielsweise ganze JSON-Dokumente als Tabellen geladen, manipuliert und gespeichert werden.

Für die Erhöhung der Verfügbarkeit können logische oder physikalische Repli-

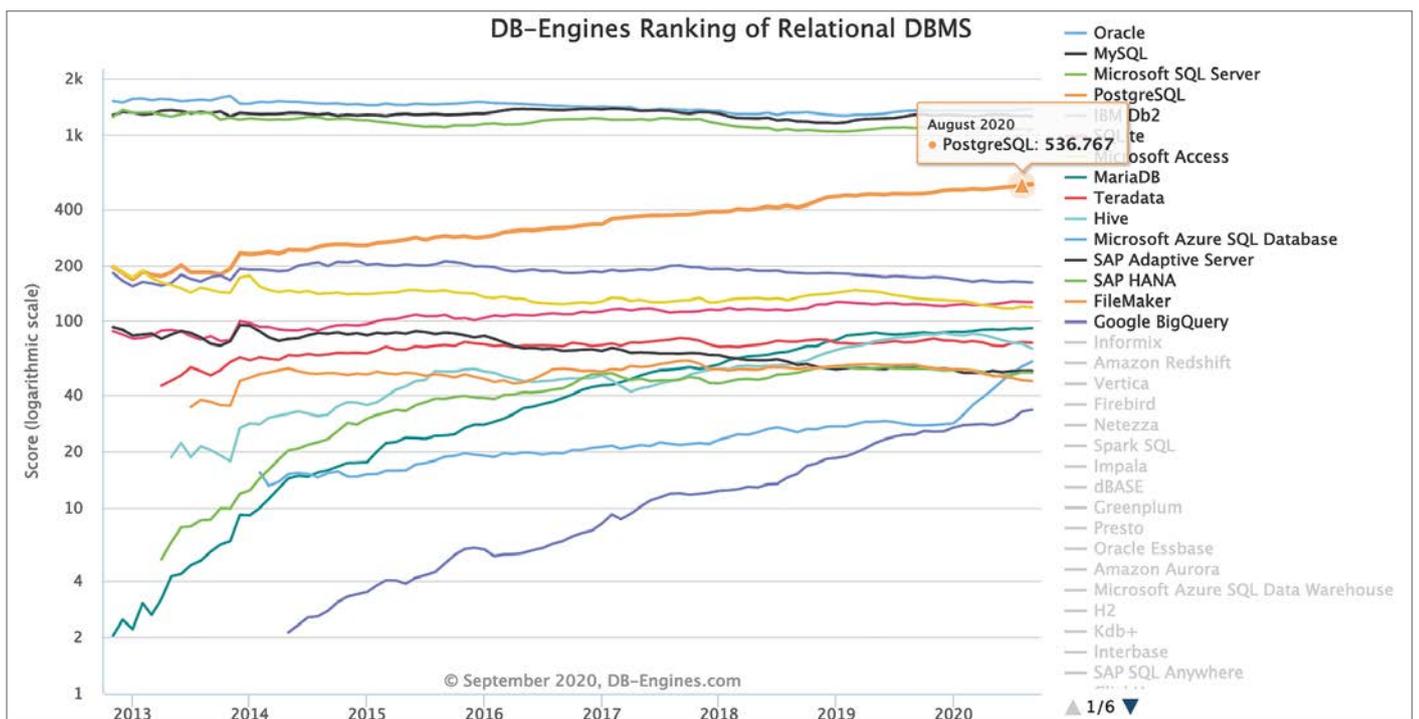


Abbildung 1: DB-Engines Ranking of Relational DBMS – Stand September 2020 (Quelle: Borys Neselovskyi)

kation konfiguriert werden. Die primäre Datenbank wird mit einer oder mehreren Standby-Datenbanken synchron gehalten. Wenn die primäre Seite ausfällt, kann die Standby-Datenbank die Aufgabe der primären Datenbank übernehmen.

Die Datenbank kann mithilfe vieler Kommandozeilen-Befehle administriert und gewartet werden. Mit dem grafischen Tool pgAdmin können die Aufgaben wie Erstellung, Administration und Wartung einer Datenbank sehr vereinfacht werden.

Die Datenbank bietet Werkzeuge zur Einhaltung des hohen Sicherheitsstandards. Die goldene Regel für die Sicherheit bei der Benutzerverwaltung besteht darin, den Benutzern das erforderliche Minimum an Zugangsrechten zu gewähren. Mit einem gut durchdachten User-Role-Konzept wird in Postgres sichergestellt, dass nur autorisierte Benutzer auf Datenbanken den Zugriff bekommen. Für die sichere Authentifizierung kann das Kerberos-Protokoll eingesetzt werden.

Diese und viele andere Eigenschaften sind bereits in der Basisversion der Postgres-Datenbank enthalten.

Zusätzlich zu Core-Eigenschaften implementieren viele Entwickler und Unternehmen zahlreiche Erweiterungen, die an der einen oder anderen Stelle die Funktionalität der Datenbank um einige Optionen ergänzen und erweitern. So zum Beispiel ermöglicht die Erweiterung PostGis [4], geografische Objekt- und Geolokationsdaten in einer Postgres-Datenbank zu speichern und zu bearbeiten.

Die Erweiterung Partony bietet mehr Funktionen, um die Verfügbarkeit einer Datenbank zu erhöhen. Ich habe nur zwei Erweiterungen genannt. Es existieren mehrere andere Anwendungen, die eine Postgres-Datenbank funktionaler, mächtiger und optionsreicher machen.

Welchen Mehrwert bietet EDB Advanced Server für Kunden?

Open-Source-Software bietet Endnutzern Freiheit und Flexibilität. Viele Unternehmen brauchen aber für die eingesetzte Software einen professionellen Support.

In unternehmenskritischen Landschaften dürfen in der Regel keine oder sehr kurze Ausfälle geduldet werden. Fehler und Bugs sollen sehr schnell gelöst werden. Bei einem

Systemausfall sind oft sehr kurze Wiederherstellungszeiten erlaubt. Die Community kann in Fehlerfällen nur bedingt helfen, weil diese keine Support-Strukturen besitzt und keine SLA-Zeiten einhalten kann. Man kann die aufgetretenen Probleme an die Community melden und hoffen, dass Bugs schnell repariert werden. Das reicht allerdings in der Regel nicht aus. Firmen brauchen für Postgres einen Ansprechpartner, der den Support leisten und SLAs einhalten kann.

Die Firma EnterpriseDB (Abkürzung: EDB) hat die notwendigen Strukturen aufgebaut und bietet seit mehreren Jahren einen professionellen Support der Postgres-Datenbank für Kunden an. Ein ganzes Team beschäftigt sich rund um die Uhr mit Problemen, die im Datenbankumfeld auftreten. EDB garantiert Lösungs- beziehungsweise Wiederherstellungszeiten und sorgt für einen sicheren Betrieb der PostgreSQL-Datenbank in unternehmenskritischen Umgebungen.

Warum kann EDB einen Herstellersupport anbieten?

Es ist wichtig zu wissen, dass mehrere EnterpriseDB-Mitarbeiter sich sehr intensiv innerhalb der Postgres-Community engagieren. Meine Kollegen Bruce Momjian und Dave Page sind in dem 5-köpfigen Community-Core-Team vertreten. Amit Langote, Devrim Gündüz und Robert Haas arbeiten in der Community als Major Contributors. Viele weitere Kollegen arbeiten in der Community mit und entwickeln PostgreSQL. Meine Kollegen haben in der Postgres *Materialized Views* und *Just In Time Compilation* parallele Ausführung und Partitionierung eingeführt. Das Tool pgAdmin wurde von Dave Page entwickelt.

Dank der ständigen Weiterentwicklung der Postgres-Datenbank durch meine Kollegen ist EDB in der Lage, einen professionellen Softwaresupport für die Datenbank anzubieten.

Zusätzlich bietet EDB für Kunden eine Betriebsunterstützung an: Erfahrene DBAs verwalten Kundensysteme und sorgen für den reibungslosen Datenbankbetrieb.

Oracle-Kompatibilität

Auch in den frühen 2000er Jahren wollten viele Unternehmen Datenbanken von Oracle nach Postgres migrieren. Die

se fragten die Community, ob es möglich wäre, einige Oracle-Funktionalitäten in Postgres zu implementieren. Die Community hat diese Anfragen abgelehnt mit der Begründung, dass Postgres eine eigenständige Datenbank ist und man einen eigenen Entwicklungsweg einschlagen möchte.

EnterpriseDB hat großes Potenzial in diesem Markt gesehen und baut seit 2006 eine Oracle-kompatible Version von Postgres. Im EDB Postgres Advanced Server sind viele nützliche Oracle-Entwicklungen implementiert.

Viele Oracle-Datentypen, Funktionen und andere Ressourcen sind in dem Advanced Server verfügbar: Sie können Optimizer Hints in ihren Abfragen verwenden. Sehr beliebte Oracle-Funktionen wie *DUAL*, *SYSDATE* und *CURRENT_TIMESTAMP* sind im Advanced Server erstellt. Oracle Konstrukt *ROWNUM* kann auch im Postgres Advanced Server in Abfragen verwendet werden.

Advanced Server bietet erweiterte Sicherheits-Features. EDB hat das *Password Policy Management Framework* implementiert, das die Komplexität, Länge und Lebenszeit eines Passwortes bestimmt. Darüber hinaus haben wir im Postgres Mechanismen implementiert, die Daten nach dem PCI (Payment Card Industry)-Standard sicher in Postgres aufbewahren können. EDB hat auch erweitertes Auditing, Data-Redaktion und Schutz von SQL-Injektion im Advanced Server implementiert.

Advanced Server beinhaltet mehr als 200 Packages, Funktionen und andere Ressourcen, die eine hohe Kompatibilität zu Oracle-Ressourcen aufweisen. Viele DBMS- und UTL-Pakete sind erstellt und können in Postgres verwendet werden. Zum Beispiel ermöglicht das Paket *DBMS_AQ* die Nutzung von Advanced Queuing in Postgres. Das Package *DBMS_LOB* bietet eine Schnittstelle zum Erstellen, Löschen und Verwalten von Large-Objekten (LOBs). Viele Data Dictionary Views, die wir aus Oracle-Datenbanken kennen, sind auch im Advanced Server implementiert.

Die Migration von Oracle zu Postgres stellt für Unternehmen ein großes Hindernis dar. Wenn Sie von der Oracle-Datenbank zur Community-Version von Postgres migrieren, müssen Sie die oben genannten Artefakte in Postgres anpassen beziehungsweise neu erstellen. Advanced-Server-Oracle-Kompatibilität ermöglicht eine sehr einfache Migration



Abbildung 2: Postgres Enterprise Manager: Performance Analyse Dashboard (Quelle: Borys Neselovskyi)

von diesen Objekten nach Postgres, ohne diese neu zu erstellen. Ihre Anwendungen müssen, wenn überhaupt, punktuell geändert werden.

Da wir über die Oracle-Kompatibilität reden, reden wir auch über die Tools, die ebenfalls Oracle-kompatibel sind. EDBLoader und EDBPlus arbeiten ähnlich wie Oracle SQLPlus und Oracle SQL Loader. Wenn Sie sich mit Oracle Tools bestens auskennen, können Sie mit EDB-Alternativen auch im Advanced Server sehr schnell zurecht kommen. Da im realen Leben sehr oft Oracle- und Postgres-Datenbanken parallel betrieben werden, können DBAs mit ähnlichen Tools beide Welten bequem und auf die gleiche Weise betreiben.

Die Oracle-Kompatibilität der Postgres-Datenbank vereinfacht den Weg von Oracle zu PostgreSQL, verringert den Migrationsaufwand und hilft Oracle-Administratoren und -Entwicklern, sich schnell in Postgres einzuarbeiten.

Neben der Oracle-Kompatibilität stellt EDB einige Tools zur Verfügung, die die Funktionalität der Postgres-Datenbank erweitern und helfen, die Postgres in unternehmenskritischen Umgebungen effizient und sicher zu betreiben.

Sehr beliebt bei Nutzern ist Postgres Enterprise Manager (Abkürzung: PEM). Die Überwachung und Verwaltung von kleinen Datenbank-Umgebungen kann auch über die Kommandozeile erfolgen. Sobald jedoch die Anzahl der Datenbanken wächst, braucht man ein Tool, das die Verwaltung, die Optimierung und das Monitoring mehrerer Datenbanken ermöglicht. Mit der Einführung von Postgres Enterprise Manager hat die Firma EnterpriseDB ein grafisch-basiertes Tool entwickelt, das die Verwaltung und Überwachung von großen Datenbank-Landschaften ermöglicht. Oracle-DBAs und Entwickler, die gerne mit dem Oracle Enterprise Manager arbeiten, finden in PEM ähnliche Funktionalität. Postgres Enterprise Manager sammelt Leistungs- und Statusdaten, aggregiert diese und spiegelt sie in Diagrammen und Dashboards wider (siehe Abbildung 2). PEM bietet mehr als 200 integrierte Metriken mit Schwellwerten. Damit können Serverstatus, Speicherüberfüllung sowie Datenbankprobleme überwacht und per E-Mail oder SNMP Traps an DBAs gemeldet werden. Entwickler können SQL ausführen und Abfragen optimieren. Administrative Tätigkeiten, wie beispielsweise Backup-

Konfiguration oder Datenbank-Erstellung, können bequem über die GUI-Oberfläche erledigt werden.

Das nächste Tool, das ich vorstellen möchte, ist EDB Failover Manager (Abkürzung: EFM). Die Lösung dient zur Bereitstellung hochverfügbarer Datenbanken unter Verwendung einer Master-Standby-Architektur mit automatischem Failover. Der Failover Manager ist geeignet für die Replikationen von Datenbanken sowohl lokal in demselben Rechenzentrum als auch zwischen Standorten und kann für die geographische Absicherung von wichtigen Datenbankumgebungen verwendet werden.

Fazit

PostgreSQL ist eine Open-Source-Datenbank, die dank zahlreicher Tools und Erweiterungen den höchsten Anforderungen an Sicherheit, Verfügbarkeit, Performance und Wartbarkeit entspricht. Einige Unternehmen bieten für Postgres einen Herstellersupport an. Die Kombination aus breiter Funktionalität, guten Werkzeugen und professionellem Support ermöglicht den Einsatz der PostgreSQL in unternehmenskritischen Umgebungen, wo strenge Servicevereinbarungen bezüglich Verfügbarkeit und Performance eingehalten werden müssen.

Quellen

- [1] <https://www.postgresql.org/about/licence/>
- [2] <https://www.oracle.com/assets/cloud-licensing-070579.pdf>
- [3] <https://de.wikipedia.org/wiki/PL/pgSQL>
- [4] <https://de.wikipedia.org/wiki/PostGIS>



Borys Neselovskyi
borys.neselovskyi@enterprisedb.com



Von Oracle zu Postgres im Handumdrehen

Andreas Kother & Markus Klenke, TEAM

Mit den weiteren Fokussierungen des Business-Plans seitens Oracle auf die Cloud wird für viele Kunden der Wunsch immer größer, das Herzstück ihrer Business-Anwendung, die Datenbank, nicht dem Zufall zu überlassen. Mit der PostgreSQL-Datenbank existiert eine Plattform, die nicht nur einen angemessenen Ersatz für die existenten Datenbankstrukturen bietet, sondern ebenfalls eine prozedurale Entwicklungsumgebung bereitstellt.

Überblick und Vergleich

Beginnen wollen wir unseren Beitrag mit einer Darstellung der Gemeinsamkeiten und Unterschiede der beiden Datenbankmanagementsysteme (DBMS) Oracle und PostgreSQL (oft kurz Postgres genannt). Bei beiden handelt es sich um objektrelationale DBMS, die sich weitestgehend konform zu dem jeweils aktuellen SQL-Standard verhalten. Manche Funktionen, wie beispielsweise Partitioning, sind bei Postgres Basisbestandteile und bei Oracle eine zusätzliche Option auf die Enterprise Edition (EE).

Lizenzierung und Support sind ebenfalls ein wenig anders organisiert als bei Oracle. Für den Betrieb von Oracle sind Lizenzen käuflich zu erwerben und für den Oracle Support ist ein jährlicher Beitrag zu entrichten. Postgres wird unter der PostgreSQL License bereitgestellt. Das ist ein offenes Lizenzmodell ähnlich der BSD-Lizenz. Der Support läuft über Communities, Mailing-Listen und Ähnliches.

Einen groben Überblick über die technischen Gemeinsamkeiten liefert die *Abbildung 1*.

Einige Details

Bei allen technischen Gemeinsamkeiten gibt es in der Umsetzung doch Unterschiede. Einige wollen wir hier kurz skizzieren.

Datentypen

Bei Oracle kennen wir bei der Verarbeitung von Zahlen im Wesentlichen den Datentyp NUMBER. Damit lassen sich so-

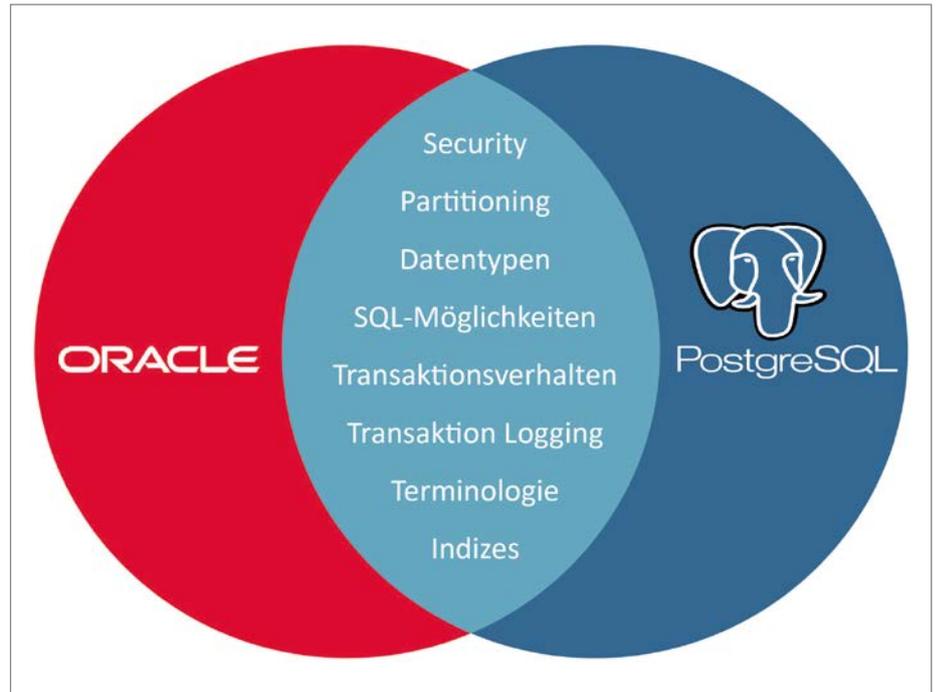


Abbildung 1: Technische Gemeinsamkeiten (Quelle: TEAM GmbH)

wohl ganzzahlige als auch Fest- und Fließkommazahlen verarbeiten. Daneben gibt es speziell für Fließkommazahlen noch die Datentypen BINARY_FLOAT und BINARY_DOUBLE.

Bei Postgres gibt es zur Verarbeitung von Zahlen hingegen eine größere Anzahl verschiedener Datentypen, wie die folgende Aufzählung zeigt:

smallint	2 bytes
integer	4 bytes
bigint	8 bytes
decimal	variable
numeric	variable
real	4 bytes
double precision	8 bytes

Tablespaces

Tablespaces bestehen bei Oracle entweder aus einer (Bigfile Tablespace)- oder aus mindestens einer (Smallfile Tablespace)-Datei. Tabellen werden physikalisch innerhalb des Tablespace, also auch innerhalb der Datei(en) abgelegt.

Bei Postgres bedeutet das Anlegen eines Tablespace nicht das Erzeugen einer oder mehrerer Dateien, sondern das Hinzufügen eines Verzeichnisses auf Filesystem-Ebene. Darunter wird dann für jede neue Tabelle und jeden neuen Index eine neue Datei angelegt. Wird das betreffende Objekt größer als 1 GB, so wird automatisch eine weitere Datei dazu angelegt.

Gemeinsamkeiten

Neben den gerade dargestellten Unterschieden gibt es, wie oben bereits kurz erwähnt, auch eine Menge an Gemeinsamkeiten. Dazu zählen die Art des Verbindungsaufbaus zwischen Client und Server (siehe *Abbildung 2*) sowie für viele vor allem die Möglichkeit, in der Datenbank mit einer prozeduralen Erweiterung namens PL/pgSQL, ähnlich dem PL/SQL bei Oracle, Stored Procedures abzulegen (auch wenn es in PL/pgSQL beispielsweise keine Packages gibt).

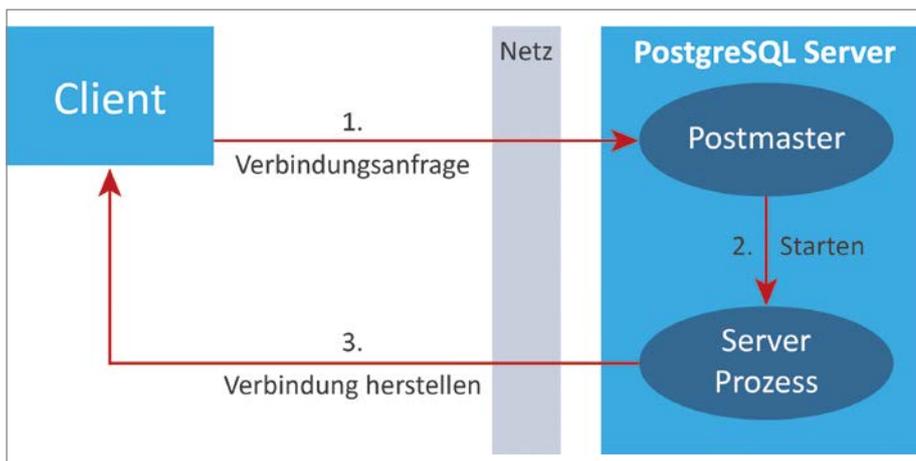


Abbildung 2: Verbindungsaufbau (Quelle: TEAM GmbH)

```

...
# Ersetzt alle Oracle-Datentypen Number und Char mit Precision 1 sowie die spezielle Spalte COL3 in der Ta-
belle TB1 durch den Postgres-Datentyp boolean
REPLACE_AS_BOOLEAN          NUMBER:1 CHAR:1 TB1_:COL3 TB_NAME1:COL_NAME2

# Setzt die migrierten Foreign Keys auf Deferrable (zur Vereinfachung von Datenimporten)
FKKEY_DEFERRABLE 1
...

```

Listing 1: Beispieleinträge in der ora2pg.conf

Von der Theorie zur Praxis: Ein Migrationsansatz

Wie im Vorherigen beschrieben, gibt es sehr viele Ähnlichkeiten zwischen den Datenbankanbietern im schematischen und auch datenspezifischen Aufbau. Diese Ähnlichkeiten erlauben es, dass viele Migrationsszenarien durch zwei Konzepte durchgeführt werden können:

- 1:1-Migration der Oracle-Schema-Objekte in ihre Postgres-Pendants
- Datenmigration nach Umstellung der Schemaobjekte

Dieser Beitrag soll insbesondere den ersten Punkt dieser Auflistung beschreiben. Zur Unterstützung gibt es auf dem Markt diverse Tools, die das Mapping zwischen den Strukturen einer Oracle- und einer Postgres-Datenbank vereinfachen beziehungsweise automatisieren. Eine für uns sinnvolle Option ist die Nutzung des Open-Source-Tools **ora2pg**,

eines Perl-basierten Skripttools, das sequenziell die Überführung der einzelnen Schemaobjekte durchführt. Es befindet sich seit 2001 in der Entwicklung und hat etliche Iterationen von Updates hinter sich, erhält hier und da allerdings immer noch Updates, ist also entsprechend gehärtet.

Für die Verwendung des Tools ist grundsätzlich nur eine aktuelle Perl-Installation mit diesen Modulen notwendig:

- DBI (Allgemeine Datenbank-Interfaces)
- DBD-Oracle (Ora-Client)
- DBD-Pg (Postgres Client)
- ora2pg

Sind die Module installiert, kann über das Kommando `ora2pg --project_base <base_dir> --init_project <project_name>` eine Projektstruktur für die Migration erstellt werden. Diese Ordner enthalten zu den exportierten SQL-Dateien sowohl ausführbare Dateien für die vereinfachte

Ausführung einer Migration als auch eine Konfigurationsdatei für die Migration, die `ora2pg.conf`.

Die eigentliche Arbeit besteht jetzt darin, die richtige und passende Konfiguration für die bestehende Migration zu finden. Mit über 150 verschiedenen Justierungsparametern kann die Durchführung der Migration an die eigenen Projektanforderungen angepasst werden. Ob es einfache Änderungen wie Tabellenumbenennungen oder komplexere Themen wie die Verarbeitung und Überführung von Spatial Data sind, alles kann durch das Setzen von Variablen in der `ora2pg.conf` eingestellt werden (siehe Listing 1).

Um einen minimalen Test einer Migration durchzuführen, müssen aber wenigstens die Verbindungsparameter für die Oracle- sowie die Postgres-Datenbank über `DataSourceNames` eingetragen werden. Sind diese Parameter gesetzt, kann die `ora2pg`-Installation durch

Erste Prüfungen

```
ora2pg -h    ## Hilfe
```

```
ora2pg -t SHOW_TABLE
```

```
ora2pg -t SHOW_COLUMN -a DEPARTMENTS
```

```

[root@OL7oracle107 ora2pg]# ora2pg -t SHOW_COLUMN -a DEPARTMENTS
[1] TABLE DEPARTMENTS (owner: HR, 27 rows)
    DEPARTMENT_ID : NUMBER(4) => smallint
    DEPARTMENT_NAME : VARCHAR2(30) => varchar(30)
    MANAGER_ID : NUMBER(6) => integer
    LOCATION_ID : NUMBER(4) => smallint
-----
Total number of rows: 27

Top 10 of tables sorted by number of rows:
  [1] TABLE DEPARTMENTS has 27 rows
Top 10 of largest tables:
  [1] TABLE DEPARTMENTS: 0 MB (27 rows)

```

```

[root@OL7oracle107 ora2pg]# ora2pg -t SHOW_TABLE
[1] TABLE COUNTRIES (owner: HR, 25 rows)
[2] TABLE DEPARTMENTS (owner: HR, 27 rows)
[3] TABLE EMPLOYEES (owner: HR, 107 rows)
[4] TABLE JOBS (owner: HR, 19 rows)
[5] TABLE JOB_HISTORY (owner: HR, 10 rows)
[6] TABLE LOCATIONS (owner: HR, 23 rows)
[7] TABLE REGIONS (owner: HR, 4 rows)
-----
Total number of rows: 215

Top 10 of tables sorted by number of rows:
  [1] TABLE EMPLOYEES has 107 rows
  [2] TABLE DEPARTMENTS has 27 rows
  [3] TABLE COUNTRIES has 25 rows
  [4] TABLE LOCATIONS has 23 rows
  [5] TABLE JOBS has 19 rows
  [6] TABLE JOB_HISTORY has 10 rows
  [7] TABLE REGIONS has 4 rows
Top 10 of largest tables:
  [1] TABLE LOCATIONS: 0 MB (23 rows)
  [2] TABLE DEPARTMENTS: 0 MB (27 rows)
  [3] TABLE JOB_HISTORY: 0 MB (10 rows)
  [4] TABLE EMPLOYEES: 0 MB (107 rows)
  [5] TABLE REGIONS: 0 MB (4 rows)
  [6] TABLE JOBS: 0 MB (19 rows)
[root@OL7oracle107 ora2pg]#

```

Abbildung 3: Ergebnis eines Testaufrufs von ora2pg (Quelle: TEAM GmbH)

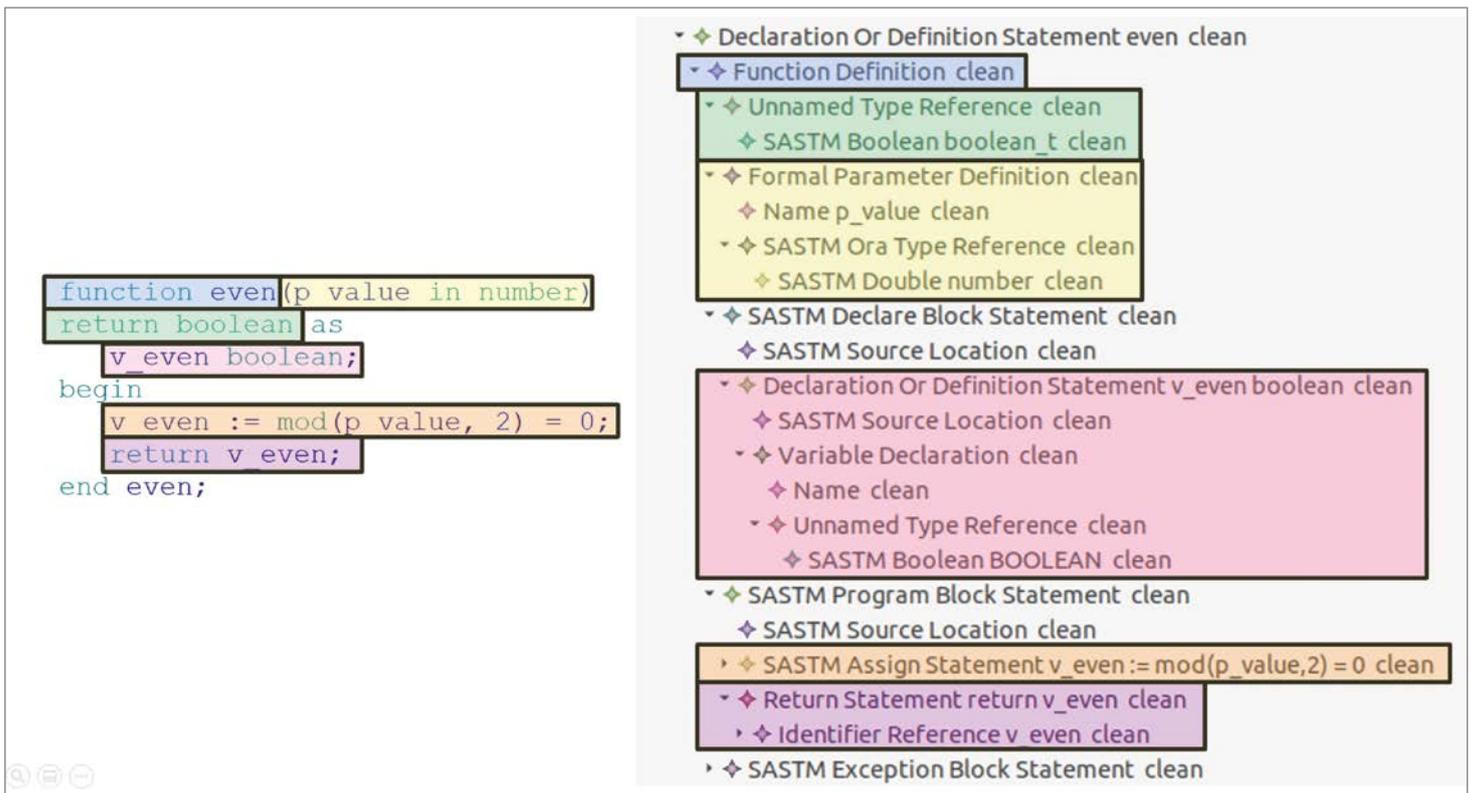


Abbildung 4: Beispiel einer syntaktischen Zerlegung von PL/SQL-Code (Quelle: TEAM GmbH)

einfache Befehle wie `ora2pg -t SHOW_COLUMN -a DEPARTMENTS` getestet werden. Bei korrekter Konfiguration sollten alle Spalten der Tabelle DEPARTMENTS des in der Konfiguration hinterlegten Schemas dargestellt werden (siehe Abbildung 3).

Die Migration wird dabei grundsätzlich iterativ nach folgendem Muster ausgeführt:

1. Konfiguration anpassen
2. Datenexport via `ora2pg -t COPY -o data.sql -b ./data -c ./config/ora2pg.conf`
3. Migrationsimport via `./<Projekt-Pfad>/import_all.sh -d <migrations_db> -U <login_user> -o <db_owner>`

Danach wird das migrierte Datenbankschema überprüft. Stimmt die Migration noch nicht mit dem gewünschten Ziel überein (und bei einer echten Migration wird das auch nach mehreren Iterationen noch nicht der Fall sein), so wird die Konfiguration angepasst und das iterative Verfahren wird erneut ausgeführt, bis eine gewünschte Quote der Migration erreicht ist. Einzelne Aspekte der Migration können dann an dieser Stelle gegebenenfalls manuell durchgeführt werden.

Wo sind die Stolpersteine?

Natürlich gibt es, wie zu Beginn des Artikels beschrieben, auch nicht kanonische Überführungsszenarien. Was zum Beispiel passiert mit üblichen PL/SQL Design Patterns wie Package-Variablen oder plattformspezifischen Aufrufen?

Hier ergeben sich im Prinzip zwei große Fragestellungen:

1. Soll weiterhin der Benutzerzustand in der Datenbank gehalten werden?
2. Soll fachliche Logik weiterhin in der Datenbank bleiben oder als andere Sprache / anderer Service implementiert werden?

Ist man rein an einem Wechsel des Datenbankproviders interessiert, so würde man wahrscheinlich beide Fragen mit „Ja“ beantworten. In dem Fall könnten Package-Variablen (da es ja de facto keine Packages mehr gibt) durch temporäre Tabelleneinträge innerhalb von Schemata realisiert werden. Für die Überführung von plattformspezifischen Packages (dbms_..., utl_...) gibt es beispielsweise ein Open-Source-Projekt [orafce\[1\]](#), das es sich zur Aufgabe gemacht hat, die Oracle-spezifischen Funktionen in Postgres als Wrapper zu implementieren.

Will man beispielsweise seine monolithisch konzipierte Datenbank in Richtung einer Cloud-basierten Infrastruktur bringen (also beispielsweise für Cloud-Native-Entwicklung vorbereiten), so würde man gegebenenfalls die Fragen mit „Nein“ beantworten.

In dem Fall ergibt es Sinn, dass beispielsweise die Sprachkonstrukte wie Functions oder Procedures nicht einfach 1:1 umgesetzt werden, sondern in ihrer syntaktischen Struktur aufgebrochen und für andere Plattformen abstrahiert überführt werden (siehe Abbildung 4). So haben wir beispielsweise in Migrationsprojekten durch Syntaxbaumzerlegung und semantische Anreicherungen ganze PL/SQL-Anwendungen in Java-Webservices umgestellt.

Fazit

Aufgrund der hohen strukturellen Ähnlichkeiten sollte bei einer Planung zur Migration von Oracle-Datenbanken zu anderen DBMS in den meisten Fällen Postgres in Betracht gezogen werden. Viele Funktionalitäten sind entweder 1:1 umsetzbar oder können durch kleine Anpassungen auf die jeweiligen Pendanten überführt werden.

Mit zusätzlichen Tools wie beispielsweise ora2pg lassen sich diese schematischen Überführungen sicher und nachvollziehbar durchführen, auch wenn gerade bei größeren Datenbanksystemen mit dem iterativen Ansatz nicht zu schnell aufgeben werden sollte. Gerade wenn die Hauptarbeit nicht auf der Migration von Tabellen, sondern auf der von komplexeren Strukturen liegt, kann der erste wirkliche Erfolgsfall eine gewisse Zeit und einige Durchläufe dauern. Aber dieser Fall wird eintreten und der Lohn der Mühen ist ab der ersten Sekunde (insbesondere im Geldbeutel) spürbar.

Quellen

[1] <https://github.com/orafce/orafce> - Orafce - Oracle's compatibility functions and packages

Über die Autoren

Markus Klenke unterstützt als Consultant unterschiedliche Kunden bei der Entwicklung von Business Software und der Erstellung und Analyse von Softwarearchitektur. Dabei liegt der Fokus auf modellgetriebenen Migrationen und Entwicklung sowie auf der Oracle Fusion Middleware, wobei insbesondere die deklarativen Frameworks Oracle ADF und Oracle VBCS hier die zentralen Interessengebiete darstellen.

Andreas Kother betreut Kunden im Umfeld Remote Services für Oracle-Datenbanksysteme und Application-Server. Den Schwerpunkt bildet dabei die Datenbank mit ihren Hochverfügbarkeitsmöglichkeiten, aber auch Betriebssystem und Storage sind keine Fremdwörter.



Markus Klenke
info@team-pb.de



Andreas Kother
info@team-pb.de

MUNIQSOFT
— CONSULTING —

Consulting

Hochverfügbarkeit mit IQ

Sicherheit vor teuren Ausfallzeiten:

Mit dem richtigen Konzept sind Ihre Daten und Server vor Systemausfällen optimal geschützt.

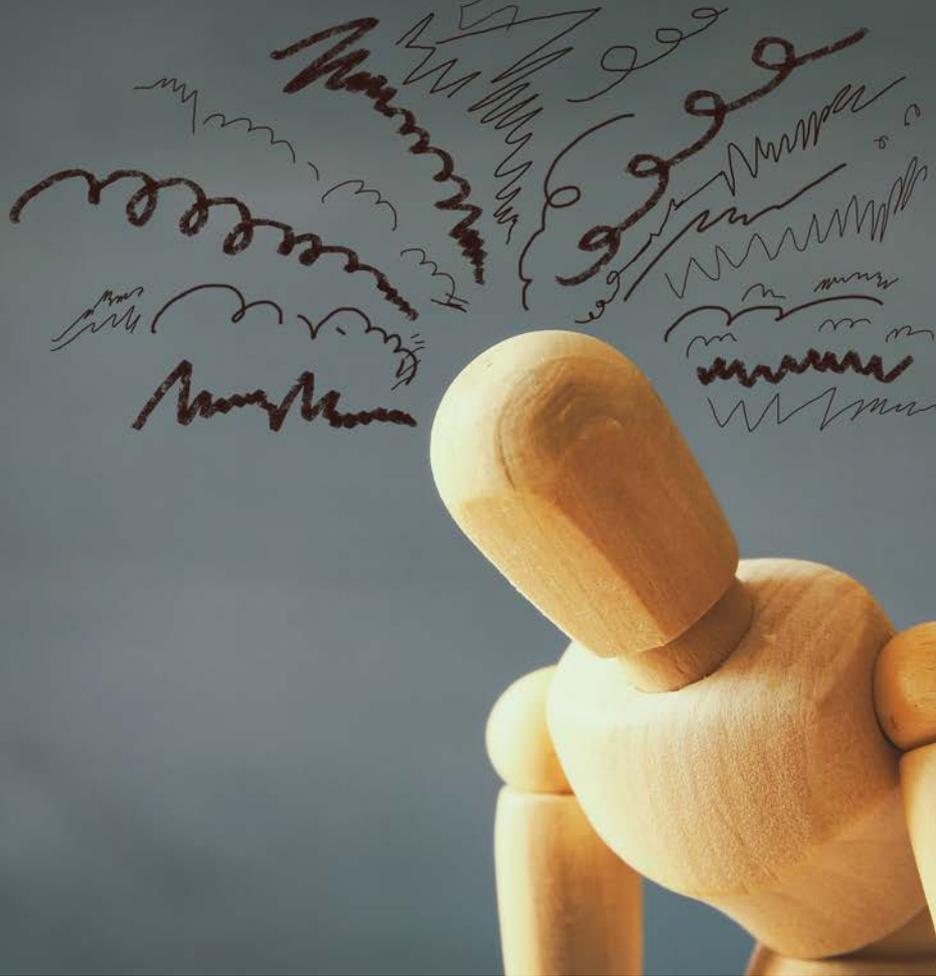
Nutzen Sie die Erfahrung der Muniqsoft Consulting GmbH
www.muniqsoft-consulting.de

ORACLE

Partner



Jetzt Beratungstermin vereinbaren:
+49 89 62286789-39



MySQL 8 Best Practices – Tuning für Dummies

Carsten Thalheimer & Mario Beck, Oracle, Oracle MySQL GBU

Oracle MySQL ist nach wie vor eine der populärsten Datenbanken überhaupt. 2018 erschien Oracle MySQL 8 nach fast 4-jähriger Entwicklung. MySQL 8 steht, genauso wie sein Vorgänger (MySQL 5.7), als Open Source (GPLv2) und als kommerzielle Enterprise-Version zur Verfügung. Oracle bietet für beide Versionen den Oracle Premier Support an. Auch in der Oracle Cloud kann ein vollständig gemanagter MySQL-Datenbankservice gemietet werden [1]. Wenn wir aber MySQL ‚klassisch‘ betreiben wollen, müssen wir selbst Hand anlegen: Die Installation ist denkbar einfach und geht schnell von der Hand. Aber was dann?

MySQL kann vielfältig installiert und betrieben werden. In den 90er und den 2000er Jahren wurde MySQL aus Performancegründen vor allem auf eigenen physischen Servern parallel mit den Applikationen (LAMP [2]) ausgeführt. Seit einigen Jahren hat sich allerdings der Betrieb von MySQL stark auf die Verwendung von virtuellen und/oder Containerinstanzen

verschoben. Dementsprechend vielfältig sind heute die Installationsmöglichkeiten. Wir werden in diesem Artikel auf die Installationsmöglichkeiten nicht näher eingehen, sondern gehen von einer erfolgreichen MySQL-8-Installation auf Basis von Oracle Linux 7 (Minimum Install) aus. Ziel ist es, bei allen Änderungen möglichst nahe an den Systemstandards zu bleiben.

Anpassungen am Betriebssystem

Wie eingangs schon erwähnt, ist MySQL nicht sonderlich wählerisch, wenn es um die Ausführungsumgebung geht. Unterstützt werden neben verschiedenen Linux-Varianten auch Windows, Solaris, macOS und FreeBSD. Linux, im Spe-

ziellen Red Hat und deren Nachbauten, sind für MySQL äußerst populär. Es empfiehlt sich, für MySQL und Anwendungen eine eigene (virtuelle) Betriebssystemumgebung bereitzustellen. Somit kann im Nachgang sehr einfach eine unabhängige Skalierung der Anwendungen und/oder der Datenbanken erfolgen oder als InnoDB Cluster erweitert werden. Grundsätzlich kann MySQL auf allen Plattformen auch ohne OS-Änderungen betrieben werden. Linux bietet die Möglichkeit, Systemressourcen für einzelne Benutzer/Prozesse zu begrenzen, um parallelen Benutzern/Prozessen auf der Instanz „auch“ Ressourcen zur Verfügung stellen zu können. Da auf unserem Server nur MySQL betrieben werden soll, möchten wir sicherstellen, dass der MySQL-Prozess möglichst viele Ressourcen bekommt, um Engpässe im späteren Betrieb zu vermeiden. Es empfehlen sich kleinere Änderungen auf Basis des Betriebssystems.

Linux bietet die Möglichkeit, durch „Optimistic Memory Allocation“ die Menge des benötigten Arbeitsspeichers über „*malloc()*“ dynamisch zuzuweisen. Das System geht allerdings davon aus, dass laufende Anwendungen niemals den gesamten zugewiesenen Speicher auf einmal benötigen werden. Wenn dieser doch benötigt wird, versucht Linux den zusätzlichen Speicher zunächst in den Swap auszulagern.

„*Swappiness*“ ist der Prozess, der auftritt, wenn das System einen Teil des Arbeitsspeichers in den Swap (spezieller Speicherplatz auf der Festplatte) verschiebt. Das passiert, wenn Ihrem System der Arbeitsspeicher ausgeht und das System die Informationen auf die Platte schreiben muss. Bekanntermaßen ist die Festplatte viel langsamer als Ihr Arbeitsspeicher. MySQL reagiert auf diesen Zustand sehr problematisch.

Swappiness kann sehr einfach als *root/sudo* reduziert werden (siehe Listing 1)

An dem Punkt, an dem der gesamte Speicher ausgehen könnte, sucht der *OOM-Killer* einen Prozess (auf Basis einer heuristischen Entscheidung), den er beenden könnte. Vereinfacht gesagt: Prozesse, die sehr viel Speicher verwenden (wie zum Beispiel MySQL), sind wahrscheinlich von einer Zwangsbeendigung durch das Betriebssystem betroffen. Diesen Prozess wollen wir zugunsten von MySQL über *OOMScoreAdjust* beeinflussen (siehe Listing 2).

Die Gesamtzahl aller konkurrierenden Client-Verbindungen, Tabellenzugriffe, Datenbankdateien und Logdateien darf den *Maximum File Descriptor* des Betriebssystems (*ulimit -n*) nicht überschreiten (siehe Listing 3). In älteren Linux- und MySQL-Versionen gab es hier häufig Probleme (Standard war 1024). Durch das Data Dictionary in MySQL 8 und den dadurch verbundenen Wegfall der FRM-Dateien ist auf Basis *systemd* und MySQL 8 das Problem entschärft.

Für viele Anwendungsfälle ist dies ausreichend. Beispielhaft ändern wir die Variable für den Fall, dass wir sehr viele parallele Verbindungen auf sehr vielen Tabellen erwarten. Als grobe Kalkulation können wir hier die Anzahl der maximalen MySQL-Verbindungen (z.B. 300) multipliziert mit der Anzahl der gleichzeitig zu öffnenden Tabellen dieser Verbindungen (40) veranschlagen (300x40=12000) (siehe Listing 4).

MySQL-Anpassungen

MySQL 8 ist das derzeit aktuelle Release. Es wurde in der Version 8.0.11 vor etwa drei Jahren der Allgemeinheit zur Verfügung gestellt. Allerdings stellt Oracle etwa alle drei Monate für MySQL 5.7 und

MySQL 8 ein Update zur Verfügung (Versionsnummer nach dem zweiten Punkt, z.B. 8.0.22). Typischerweise werden in diesem Update nicht nur Fehler behoben, sondern auch kleinere Erweiterungen eingebaut. Hierbei kommt es auch immer wieder zu Optimierungen, die teilweise deutliche Verbesserungen in der Datenbankperformance zeigen. Die vollständigen Details zu den Updates können wir einfach auf dem Entwicklerblog oder in den Release Notes verfolgen [3]. Es empfiehlt sich, keine MySQL-Versionen zu verwenden, die älter als etwa 4-6 Monate sind.

Es gibt zwei wesentliche Konfigurationsdateien für MySQL:

In *my.cnf* unter */etc* werden die Startoptionen und Basiskonfigurationen für MySQL bereitgestellt. Seit MySQL 8 gibt es eine weitere Konfigurationsdatei (*mysqld-auto.cnf* im *datadir*). Diese enthält persistente Systemvariablen-Einstellungen im JSON-Format, die vom Server bei der Aus-

```
# sysctl vm.swappiness
vm.swappiness = 30
# sysctl -w vm.swappiness=1
vm.swappiness = 1
```

Listing 1: Anpassen des Swap-Verhaltens in Linux

```
# cat /proc/$(pidof mysqld)/oom_score
104
# cat /proc/$(pidof mysqld)/oom_score_adj
0
# systemctl edit --full mysqld.service
###add lines###
Adjust OOM
OOMScoreAdjust=-1000
# systemctl restart mysqld.service
```

Listing 2: Anpassen des OOM-Scores

```
# cat /proc/$(pidof mysqld)/limits |grep files
Max open files      10000 10000 files
```

Listing 3: Anzeigen der Maximalwerte

```
# systemctl edit --full mysqld.service
###search in file and change accordingly###
# Sets open_files_limit
LimitNOFILE = 12000
# systemctl restart mysqld.service
```

Listing 4: Änderung der Maximalwerte

führung von *SET-PERSIST*-Anweisungen erzeugt werden und nicht manuell bearbeitet werden sollten.

Die Standardkonfiguration unter `/etc/my.cnf` ist überschaubar (siehe Listing 5).

MySQL belegt über 600 Variablen mit Standards. Diese können wir uns in einer MySQL-Session mittels `show variables` anzeigen lassen beziehungsweise mit `show variables like „%max_connection%“` filtern.

Die Idee ist, mit einer einfachen Konfiguration zu starten und mit einem „Schuss ins Blaue“ zu beginnen. Die Konfiguration erfolgt hierbei natürlich nicht willkürlich, sondern folgt Ideen, die wir sehr häufig antreffen.

Die wichtigsten MySQL-Änderungen betreffen hierbei die richtige Zuweisung des zur Verfügung stehenden Arbeitsspeichers. Die Grundidee ist hierbei, dass wiederkehrende Daten für Lesezugriffe im Speicher gehalten werden und nicht von langsameren Festplatten gelesen werden müssen. Auf der anderen Seite können Daten natürlich nur auf Festplatten persistiert werden. In unserem Beispiel gehen wir von einem InnoDB-basierten MySQL aus. Im InnoDB Buffer (Arbeitsspeicher) werden viele Datenstrukturen, Buffer, Caches, Indizes und Daten gehalten. Die `innodb_buffer_pool_size` (default 128 MB) ist der MySQL-Konfigurationsparameter, der die Menge an Speicher angibt, die dem InnoDB-Pufferpool von MySQL zugewiesen wird. Dies ist die wahrscheinlich wichtigste Einstellung für schnelle SQL-Abfragen und muss auf der Grundlage des verfügbaren Arbeitsspeichers konfiguriert werden. Allerdings ist InnoDB nicht der einzige Service auf dem Linuxsystem, der auf schnellen Speicher angewiesen ist. Daher können wir den Arbeitsspeicher nur begrenzt exklusiv an InnoDB vergeben. In der Praxis haben sich hier Werte zwischen 50-90% des zur Verfügung stehenden Arbeitsspeicher auf dedizierten MySQL-Instanzen etabliert. Für InnoDB gibt es weitere wichtige Parameter, die standardmäßig konservativ angesetzt sind:

Redo Logs / `innodb_log_file_size` (default 2x50 MB): Die kombinierten Größen der Protokolldateien sollten so groß sein, dass Lastspitzen beim Schreiben ausgeglichen werden können. Je größer die Protokolldateien sind, desto weniger Checkpoints beziehungsweise DiskIO-Operationen werden benötigt. Auf der anderen Seite können MySQL Restarts deutlich länger

dauern, weil zunächst Redo Logs abgearbeitet werden müssen.

`innodb_flush_method` (default `fsync`): Linux verwendet standardmäßig den Filesystem Cache. Dieser ist für viele Linux Services sehr hilfreich, allerdings ist für InnoDB ein *double buffering* kontraproduktiv und unnötig (gilt nur für *battery-backed-writcache*-Festplatten).

Die Werte für `innodb_buffer_pool_size`, `innodb_log_file_size` und `innodb_flush_method` können seit MySQL 8 sehr einfach automatisch während des Startens mittels `innodb_dedicated_server=on` optimal gesetzt werden. Beim Betrieb von MySQL in virtuellen Umgebungen/Containern wird so automatisch auf den vom Betriebssystem erkannten Arbeitsspeicher reagiert. Wie der Name schon impliziert, kann diese Einstellung nur sinnvoll für einen exklusiven MySQL-Betrieb auf einem Server verwendet werden. Es stehen praktisch keine weiteren Ressourcen für andere Anwendungen zur Verfügung.

Die durch `innodb_dedicated_server=on` vorgenommenen Änderungen können sehr einfach in einer MySQL-Session mit `show variables` angezeigt werden und vorab mithilfe der Dokumentation abgeschätzt werden [4] (siehe `## [b] ##` in Listing 8).

Eine weitere empfohlene InnoDB-Änderung betrifft den MySQL-Betrieb auf NUMA/vNUMA CPU Cores. Für Server mit mehreren NUMA-Kernen wird empfohlen, den NUMA-Modus auf *Interleaved* zu

setzen, um die Speicherzuweisung an allen Knoten auszugleichen. Ob NUMA konfiguriert wurde oder nicht, können wir einfach über `numactl -H` herausfinden. Sollte die Ausgabe `available nodes >1` sein, empfiehlt sich die Verwendung von MySQL mit `innodb_numa_interleave=on`.

Am Ende dieses Artikels haben wir unser Beispiel `my.cnf` (siehe Listing 8) angehängt. Die Änderung für `innodb_numa_interleave=on` ist mit dem Kommentar `## [a] ##` gekennzeichnet.

Machen wir an dieser Stelle eine kurze Zwischenbilanz über den von uns verwendeten Arbeitsspeicher (am Beispiel 32 GB):

1. Betriebssystem OS (~1,5 GB)
2. `innodb_buffer_pool_size` (75% von 32 GB +10% Overhead = 26,40 GB)
3. `innodb_log_buffer_size` (16 MB – wird vernachlässigt)
4. `slave_pending_jobs_size_max` (128 MB)
5. `key_buffer_size` (8 MB – wird vernachlässigt)
6. `performance_schema.memory` (vernachlässigen wir)
7. MEMORY tables (wir verwenden keine MEMORY tables)

Von den ursprünglichen 32 GB stehen uns nun noch 32 GB – 28 GB = 4 GB für die Client-Verbindungen zur Verfügung.

Eine weitere wichtige Variable für MySQL definiert die Anzahl der gleichzeitigen Verbindungen mit der Datenbank. Er

```
# cat/etc/my.cnf
[mysqld]
datadir=/var/lib/mysql
socket=/var/lib/mysql/mysql.sock
log-error=/var/log/mysql.log
pid-file=/var/run/mysql/mysql.pid
```

Listing 5: Defaults in MySQL 8.0

```
<Peak gleichzeitige Verbindungen>

mysql> show global status like "%Max_used_connection%";
+-----+-----+
| Variable_name | Value |
+-----+-----+
| Max_used_connections | 68 |
| Max_used_connections_time | 2020-11-21 17:14:40 |
+-----+-----+
2 rows in set (0.01 sec)
```

Listing 6: Eine Anpassung finden wir in Listing 8 unter `## [g] ##` vorgenommen

wäre natürlich vorteilhaft, wenn es hier kein Limit gäbe. Wenn jedoch zu viele Verbindungen zu MySQL aufgemacht werden, führt dies zu Performance-Problemen der Datenbank, da jede Verbindung von unserem endlichen Arbeitsspeicher Gebrauch macht. Daher wird standardmäßig eine obere Grenze von maximal 151 Verbindungen definiert (*show variables like „%max_connections%“*).

Wenn alle 151 Clients verbunden sind, müssen diese sich die verbliebenen 4000 MB teilen. Für jede Verbindung stehen somit 26,5 MB zur Verfügung.

Eine Clientverbindung setzt sich maximal zusammen aus (Defaults):

1. read_buffer_size (128 KB)
2. read_rnd_buffer_size (262 KB)
3. join_buffer_size (262 KB)
4. sort_buffer_size (262 KB)
5. binlog_cache_size (32 KB)
6. net_buffer_length (16 KB) – (Zwischensumme 1.- 6. ~ 1 MB)
7. tmp_table_size (16 MB)

Wenn wir davon ausgehen, dass eine Verbindung jeweils nur eine temporäre Tabelle verwendet, sollte unsere Rechnung aufgehen. Wenn aber mehrere temporäre Tabellen von vielen Verbindungen verwendet werden (z.B. sort tables, derived tables etc.), können wir davon ausgehen, dass unser Speicher knapp wird.

Wir könnten dann eventuell den *innodb_buffer_pool* anpassen (von der 75%-Regel abweichend) oder die Anzahl der Verbindungen reduzieren.

Bevor wir die Anzahl der maximalen Verbindungen reduzieren, sollten wir anhand eines laufenden Systems den Peak der gleichzeitigen Verbindungen überprüfen (*Max_used_connections*) und *max_connections* an einen realistischen Wert anpassen (siehe Listing 6).

MySQL ist ausgelegt für sehr viele gleichzeitige schreibende und lesende Verbindungen. Wenn wir aber nur sehr wenige parallele, schreibende Verbindungen erwarten (<10), können wir den Overhead von dedizierten Log-Writer Threads (*innodb-log-writer-threads*) sehr einfach umgehen (siehe ## [e] ## in Listing 8). Wie die meisten neuen Variablen kann auch dieser Wert dynamisch jederzeit zurückgeändert werden.

Bereits mit MySQL 5.7 wurden viele sicherheitsrelevante Variablen standardmäßig gesetzt. Wenn sich ein Client zu

```
mysql> show global status like "%Innodb_buffer_pool_pages%";
+-----+-----+
| Variable_name          | Value      |
+-----+-----+
| Innodb_buffer_pool_pages_data      | 126224     |
| Innodb_buffer_pool_pages_dirty     | 0          |
| Innodb_buffer_pool_pages_flushed  | 158        |
| Innodb_buffer_pool_pages_free     | 201451     |
| Innodb_buffer_pool_pages_misc     | 5          |
| Innodb_buffer_pool_pages_total    | 327680     |
+-----+-----+
```

Listing 7: Status des InnoDB Buffers

MySQL verbindet, führt MySQL ein Reverse-DNS-Lookup durch und vergleicht den Hostnamen und die IP-Adresse. Sollte es allerdings ein Problem mit dem DNS-System geben, wird auch die Datenbankverbindung beeinträchtigt. Zusätzlich erlauben viele Anwender den Zugriff auf Basis von Usernames in Verbindung mit IP-Adressen und nicht Hosts. Für viele MySQL-Instanzen kann daher dieses zusätzliche Lookup einfach durch skip-name-resolve deaktiviert werden (siehe ## [h] ## in Listing 8).

Ob die *innodb_buffer_pool*-Größe passend ist oder nicht, ist etwas schwieriger zu erkennen und benötigt idealerweise ein Monitoring, wie den MySQL Enterprise Monitor. Eine erste grobe Tendenz können wir auf Basis einer laufenden MySQL-Instanz wie folgt erkennen (siehe Listing 7).

Wenn der *innodb_buffer_pool_pages_free*-Wert dauerhaft nicht unter 20% des *innodb_buffer_pool_pages_total*-Werts liegt, können wir über eine Verkleinerung des Buffer Pool nachdenken. Natürlich kann MySQL auch mit sehr viel größeren Daten arbeiten, nur müssen diese bei Bedarf von der Festplatte langsamer gelesen werden. Anpassungen des Buffers haben wir exemplarisch in Listing 8 unter ## [c] ## eingefügt.

MySQL ist eine Standardkonfiguration einer ACID-konformen Datenbank, das heißt, eine beendete Transaktion wird ganz oder gar nicht ausgeführt. Tatsächlich ist diese Eigenschaft vor allem bei geschäftskritischen Prozessen zwingend erforderlich. Im Webumfeld gibt es viele Anwendungsfälle, bei denen eine 100%ige Konsistenz nicht benötigt wird und ein Verlust von Daten über einige Sekunden im Fehlerfall akzeptiert wird. Im Umkehrschluss läuft das System dafür schneller im alltäglichen Betrieb.

Im Internet wird hierfür sehr häufig auf zwei Einstellungen hingewiesen:

Standardmäßig werden die Redo-logs bei jeder beendeten Transaktion geschrieben und auf die Festplatte übertragen (*innodb_flush_log_at_trx_commit=1*). Wenn nun schreibende Transaktionen (für etwa eine Sekunde) in Batches zusammengefasst werden, erfolgt die Übertragung deutlich effizienter. Es kommt zu deutlich höheren Schreibraten. Sollte es allerdings während der Zusammenfassung der Transaktion zu einem Absturz kommen, gehen eventuell die bereits beendeten Transaktionen der letzten Sekunde verloren, da sie zu diesem Zeitpunkt nur im Arbeitsspeicher gehalten wurden. Das geänderte Transaction Commit finden wir in ## [d] ## in Listing 8.

InnoDB bietet alle vier Transaktionsisolationsebenen, die durch den SQL-1992-Standard beschrieben werden. Standardmäßig benutzt MySQL *repeatable-read*. Die zweimalige Ausführung derselben Anfrage innerhalb einer Transaktion gibt immer die gleichen Werte zurück. *dirty reads* (nicht wiederholbare Lesevorgänge) sind dann nicht möglich. Diese strenge Isolierung bedeutet in der Regel umfassendere Sperren von Transaktionen. Dies wiederum beeinträchtigt die Leistung.

Einige andere Datenbanksysteme verwenden *read-committed* als Standard. Da bei Transaktionen die Wahrscheinlichkeit von Deadlocks durch *read-committed* stark reduziert wird, wird dies mittlerweile auch für MySQL zunehmend verwendet.

Auch diese Änderung können wir leicht in der *my.cnf* als Standard eintragen (siehe ## [k] ## in Listing 8).

Übrigens: Sollten einzelne Transaktionen eine erhöhte Isolation erfordern, kann der Transaktions-Isolationslevel dynamisch für jede Transaktion gesetzt werden. Die oben beschriebene Änderung ist nur der Default-Fall.

MySQL-Server schreiben in der Standard-Konfiguration auch das sogenannte Binary Log, kurz Binlog. Hier werden alle Änderungen an Tablespace Row-basiert protokolliert. Das Binlog ist notwendig beispielsweise für die Replikation oder eine Point-in-Time Recovery (PITR). Wenn diese Anforderungen jedoch nicht bestehen, kann auf diese zusätzliche IO-Last verzichtet werden, indem das Binlog mit `disable-log-bin` abgestellt wird (siehe ## [i] ## in Listing 8).

MySQL bietet verschiedene Storage-Engines. Inzwischen hat sich die InnoDB-Storage Engine jedoch so weit durchgesetzt, dass die Benutzung anderer Storage-Engines eher unbeabsichtigt erfolgt und nur selten wirklich erwünscht ist. Eine Mischung von Storage-Engines kann zu spürbaren Performance-Verlusten führen (z.B. bei Mischung von MyISAM- und InnoDB-Tabellen in einem Statement) oder bestimmte Architekturen unmöglich machen, wie zum Beispiel den InnoDB Cluster für Hochverfügbarkeit. Um solche unbeabsichtigte Nutzung anderer Storage-Engines zu vermeiden, können andere Storage Engines aktiv abgeschaltet werden (siehe ## [f] ## in Listing 8).

Jede Tabelle in einer relationalen Datenbank sollte einem Primary Key zugewiesen werden. Der Primärschlüssel ist eine Spalte oder eine Menge von Spalten, deren Werte jede Zeile in der Tabelle eindeutig identifizieren. MySQL-Replikation ohne einen Primary Key führt regelmäßig zu Problemen und ein InnoDB-Cluster-Betrieb ist ohne diese nicht möglich. Ein Primary Key kann für neue Tabellen (und bei Änderungen an bestehenden Tabellen) einfach mit `sql_require_primary_key` forciert werden. Die Konfiguration hierzu finden wir unter ## [j] ## (siehe Listing 8)

SQL-Queries optimieren

Neben den bisher beschriebenen Optimierungen des Betriebssystems und des MySQL-Servers ist ein großes Potenzial für Performance-Verbesserungen die Optimierung von SQL-Queries. Ein fehlender Index oder eine schlecht formulierte Query ist auch mit der besten Hardware, dem „optimiertesten“ Betriebssystem und der ausgefeiltesten Server-Konfiguration nicht vollständig zu retten.

```
<DOAG my.cnf>

[mysqld]
### Default my.cnf, Kommentare gelöscht ###
datadir=/var/lib/mysql
socket=/var/lib/mysql/mysql.sock
log-error=/var/log/mysqld.log
pid-file=/var/run/mysqld/mysqld.pid

### InnoDB Anpassungen ###
innodb_numa_interleave=on          ## [a] ##
innodb_dedicated_server=on        ## [b] ##
# Manuelle Config des Buffers
# (von 75% Regel abweichend)
innodb_buffer_pool_size=4G        ## [c] ##
# innodb_flush_log_at_trx_commit=1 (St) ## [d] ##
# 0 und 2 sind nicht ACID konform!
innodb_flush_log_at_trx_commit=2  ## [d] ##
innodb-log-writer-threads=off     ## [e] ##
# Andere Storage-Engines vermeiden ## [f] ##
disabled_storage_engines= MyISAM,MRG_MYISAM,FEDERATED,CSV,ARCHIVE,-
BLACKHOLE,MEMORY,NDB

### Andere Anpassungen ###
max_connections = 100              ## [g] ##
skip-name-resolve                  ## [h] ##
disable_log_bin                    ## [i] ##
sql_require_primary_key=on         ## [j] ##
# Transaction isolation level
transaction-isolation=read-committed ## [k] ##
```

Listing 8: Konfiguration der beschriebenen Variablen

Ein beliebtes Werkzeug, um problematische Queries überhaupt erst zu identifizieren, ist der MySQL Enterprise Monitor. Dessen Funktion des Query Analyzer bietet umfangreiche Statistiken zu den ausgeführten Queries. Darüber ist es leicht möglich, Queries mit hoher Laufzeit oder auch Queries mit hohem Ressourcenverbrauch zu identifizieren. Diese Queries versprechen den höchsten Gewinn bei einer Optimierung.



Mario Beck
Mario.Beck@Oracle.com

Quellen

- [1] <https://www.mysql.com/de/cloud/>
- [2] [https://de.wikipedia.org/wiki/LAMP_\(Softwarepaket\)](https://de.wikipedia.org/wiki/LAMP_(Softwarepaket))
- [3] <https://mysqlserverteam.com/>
- [4] <https://dev.mysql.com/doc/refman/8.0/en/innodb-dedicated-server.html>



Carsten Thalheimer
Carsten.Thalheimer@Oracle.com



Datenschutz: Anonymisierung und Pseudonymisierung

Andreas Buckenhofer

Die Zukunft ist digital – daran besteht kein Zweifel. Der Umgang mit den Daten ist entscheidend. Eine zentrale Rolle spielt nicht nur die Einhaltung der allgemeinen Datenschutzgrundverordnung (DSGVO). Jeder ist für den vertrauensvollen Umgang mit personenbezogenen Daten verantwortlich: „We need to defend the interests of those whom we’ve never met and never will“ (Jeffrey D. Sachs). Der Artikel fasst meinen Vortrag über Anonymisierungstechniken auf der DOAG-Konferenz zusammen.

Personenbezogene Daten

Die Verarbeitung personenbezogener Daten muss gesetzeskonform erfolgen. Daten wie beispielsweise Name, Geburtsdatum, E-Mail-Adresse oder Wohnort sind offensichtlich personenbezogen. Daten wie dynamische IP-Adressen oder die Fahrzeugidentifikationsnummer (VIN) gehören aber auch zu personenbezogenen Daten. Diese Daten müssen anonymisiert werden, sofern keine Zustimmung zum Verarbeitungszweck gegeben wurde.

Pseudonymisierung und Anonymisierung sind sehr unterschiedlich, werden

aber oft verwechselt. Das Diagramm fasst Anonymisierung und Pseudonymisierung zusammen. Pseudonymisierung basiert auf Techniken wie Hashing oder Tokenisierung, während Anonymisierungstechniken die persönliche Referenz unwiederbringlich entfernen. Die DSGVO gilt nicht mehr für anonyme Daten. Bei Pseudonymisierung ist es weiterhin möglich, Personen mit einem „Schlüssel“ zu identifizieren, und die DSGVO ist weiterhin anwendbar.

Die folgenden Risiken sind bei der Validierung eines anonymisierten Datensatzes von entscheidender Bedeutung. Keines der Kriterien sollte gültig sein:

- Herausgreifen: Möglichkeit, einen oder mehrere Datensätze zu isolieren, die eine Person identifizieren.
- Verknüpfbarkeit: Verknüpfen von zwei oder mehr Datensätze mit (derselben Gruppe von) Personen.
- Inferenz: Ableiten eines Werts für ein Attribut aus vorhandenen Attributen.

Es besteht weiterhin das Restrisiko, dass Daten nach der Anonymisierung wieder identifiziert werden können. Der Aufwand für die Re-Identifizierung in Bezug auf Zeit, Geld oder Komplexität muss hoch genug sein, um das Risiko auf ein Minimum zu reduzieren.

Das folgende Beispiel zeigt die drei Varianten:

- Personenbezogene Daten: „Max Mustermann fährt das Fahrzeug mit dem Kennzeichen UL-WB 134“.
- Pseudonymisierte Daten: „Herr fe435rat5 fährt das Fahrzeug mit der Zulassungsnummer a9f2ebfa70b02d97f7“ (falls ein Entschlüsselungsschlüssel vorhanden ist).
- Anonymisierte Daten: „Ein Autobesitzer aus Ulm fährt einen A180“ (vorausgesetzt, die Referenzgruppe ist groß genug).

Konsequenzen fehlerhafter Anonymisierung

Eine fehlerhafte Verarbeitung personenbezogener Daten kann zu erheblichen Geldstrafen führen, ganz zu schweigen von einer Schädigung des Ansehens des Unternehmens.

Ein Taxiunternehmen in Dänemark hat die Daten seiner Kunden nicht ausreichend anonymisiert und wurde von den Behörden mit einer Geldstrafe belegt. Die Behörden lehnten die Argumentation des Taxiunternehmens ab, dass die Telefonnummer Teil des Primärschlüssels sei und nicht gelöscht

werden könne. Die ungenügende „Anonymisierung“ durch Beibehaltung von Telefonnummern ist offensichtlich. Für das Design physischer Datenbanken kann daraus abgeleitet werden: Überlegen Sie zweimal, bevor Sie personenbezogene Daten als Primärschlüssel verwenden.

Professor Lataney Sweeney demonstrierte 1997, dass 87% der Amerikaner re-identifiziert werden können, wenn Geburtsdatum, Geschlecht und Postleitzahl in einer Datenquelle angegeben werden. Sie benutzte ein Wählerregister und verknüpfte die Datenquellen miteinander, um Einzelpersonen identifizieren zu können. Einige Jahre später veröffentlichte die Facebook-App myPersonality Daten zusammen mit Geburtsdatum, Geschlecht und Postleitzahl. Daher sind diese veröffentlichten Daten der myPersonality-App nur pseudonymisiert, nicht anonymisiert.

Anwendungsfälle

Ein wesentliches Prinzip der DSGVO ist die Datenminimierung. Anonymisierung kann dazu beitragen, dieses Prinzip umzusetzen. Die DSGVO gilt nicht mehr für anonymisierte Daten. Typische Anwendungsfälle sind unten aufgeführt. Anwendungsfälle

relevante Anonymisierungstechniken werden ebenfalls erwähnt. Diese Anonymisierungstechniken werden weiter unten im Artikel beschrieben.

- Datennutzung in Test- und Entwicklungsumgebungen.
Daten in solchen Umgebungen müssen anonym sein bis auf wenige begründete Ausnahmen. Die Techniken „synthetische Daten mit Lookups oder Randomisierung“ können betrachtet werden, um eine Datenkopie zu erstellen.
- Datenanalysen wie Visualisierung, explorative Analysen oder maschinelles Lernen.
Die Analyse kann unter Verwendung von Techniken wie „Differential Privacy“ oder mit einem „Grouping / Clustering“ -Ansatz durchgeführt werden. Alternativ ist auch eine Analyse synthetischer Daten möglich, die von Modellen für maschinelles Lernen erstellt wurden, um statistische Eigenschaften zu erhalten.

Da eine einzelne Methode oft nicht ausreicht, können viele Techniken kombiniert werden. Gartner hat Techniken zur Wahrung der Privatsphäre wie Differential Privacy oder synthetische Daten in seinen Hype Cycle for Privacy (2020) aufgenommen. Es gibt auch Open-Source-

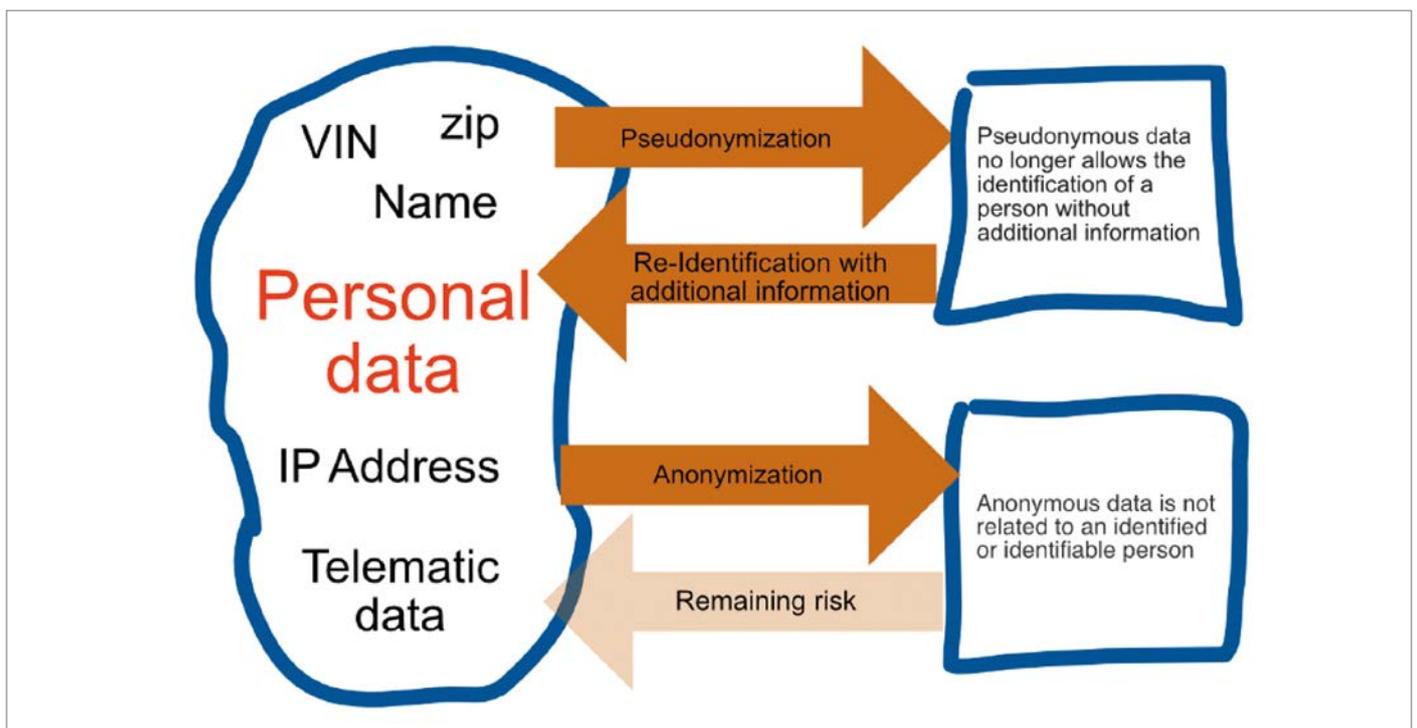


Abbildung 1: Übersicht über Anonymisierung und Pseudonymisierung von Daten (Quelle: Andreas Buckenhofer)

Projekte wie Google Rappor für Differential Privacy oder ARX.

Techniken

Die DSGVO legt nicht fest, welche Anonymisierungstechniken verwendet werden müssen. Die Artikel-29-Arbeitsgruppe hat eine Stellungnahme zu verschiedenen Maßnahmen veröffentlicht, die je nach Anwendungsfall in Betracht gezogen werden können.

Es gibt zwei allgemeine Ansätze: In-Place- und Out-of-Place-Anonymisierung. Out-of-Place-Methoden erstellen einen neuen Datensatz (Datenkopie), während In-Place-Methoden direkt mit den Produktionsdaten arbeiten.

In-Place-Anonymisierung

Durch „Data Redaction“ werden Daten wie Kreditkartennummern vor Benutzern

ausgeblendet, während die Originaldaten in der Datenbank verbleiben. Die Technik wird normalerweise in Produktionssystemen basierend auf den Anmeldeinformationen des Benutzers angewendet. Dieses Verfahren ist ein besonderer Fall, um nur bestimmte sensible Daten von Verbrauchern zu verbergen, zum Beispiel vor einem Call-Center-Agenten, der nicht alle Daten sehen darf. Oracle verfügt über integrierte Funktionen zur Datenredaktion, die der individuellen Programmierung

```
SQL> CREATE TABLE customer (
  customer_id NUMBER NOT NULL
, credit_card VARCHAR2(19) NOT NULL
, CONSTRAINT customer_pk PRIMARY KEY (customer_id)
);
2 3 4 5
Table created.

SQL> INSERT INTO customer (customer_id, credit_card) VALUES (1, '1111-1111-1111-1111');
1 row created.

SQL> INSERT INTO customer (customer_id, credit_card) VALUES (2, '1234-5678-1234-5678');
1 row created.

SQL> COMMIT;

Commit complete.

SQL> -- Original data
SELECT customer_id, credit_card FROM customer ORDER BY customer_id;
SQL>
CUSTOMER_ID CREDIT_CARD
1          1111-1111-1111-1111
2          1234-5678-1234-5678

SQL> -- Configure redaction for credit card column
BEGIN
  DBMS_REDACT.add_policy (
    object_name => 'customer'
  , policy_name => 'credit_card_pol'
  , column_name => 'credit_card'
  , function_type => DBMS_REDACT.PARTIAL
  , function_parameters => 'VVVVFVVVVFVVVVFVVVV, VVVV-VVVV-VVVV-VVVV, *, 1, 12'
  , expression => '1=1');
END;
/

SQL> 2 3 4 5 6 7 8 9 10
PL/SQL procedure successfully completed.

SQL> -- Redacted data
SELECT customer_id, credit_card FROM customer ORDER BY customer_id;
SQL>
CUSTOMER_ID CREDIT_CARD
1          ****.****.****.1111
2          ****.****.****-5678

SQL>
```

Listing 1: Oracle-Lösung mit zwei Abfragen in der Spalte credit_card

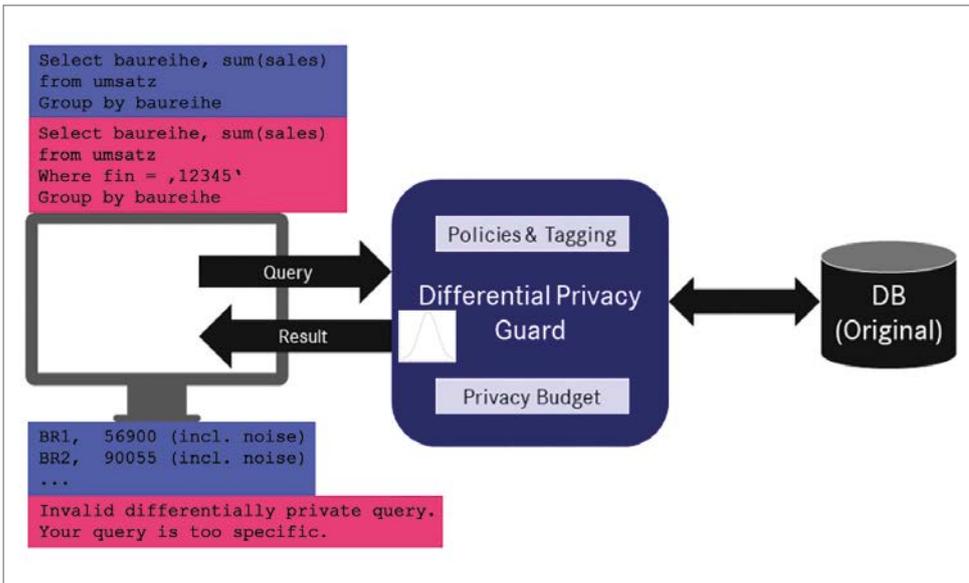


Abbildung 2: Nicht blockierte (in blau) und blockierte Abfrage (in rot) (Quelle: Andreas Buckenhofer)

Out-of-Place-Anonymisierung

Eine generierte synthetische Datenkopie mit Lookups oder Randomisierung kann die sensiblen Teile der Originaldaten verbergen. Lookup-Daten werden bereitgestellt für zum Beispiel Namen, Bankkonten und andere personenbezogene Daten. Eine Datenkopie wird mit synthetischen Daten erstellt, indem die Originaldaten durch Lookup-Werte ersetzt werden. Ein ähnlicher Ansatz ist die Verwendung von Randomisierungstechniken wie Mischen, Maskieren oder Löschen von Daten. Dazu werden Regeln erstellt, wie beispielsweise das Mischen aller Datenzellen in einer Spalte. Lookups und Randomisierungsmethoden werden häufig kombiniert. Oracle Data Masking stellt hierfür Funktionen zur Verfügung. Darüber hinaus können Modelle für maschinelles Lernen die Erzeugung synthetischer Daten mit ähnlichen stochastischen Eigenschaften wie die Originaldaten unterstützen. Teile der Originaldaten werden zum Trainieren von Modellen für maschinelles Lernen verwendet. Das folgende Diagramm (siehe Abbildung 3) zeigt den Gesamtansatz zum Erstellen einer synthetischen Datenkopie mit maschinellem Lernen. Typische Produkte sind Static oder mostly.ai.

- K-Anonymität, L-Diversität und ähnliche Methoden gruppieren Daten (Clus-

von Funktionen in der Anwendung weit überlegen sind. Die Oracle-Lösung wird im folgenden Beispiel (siehe Listing 1) mit zwei Abfragen in der Spalte credit_card gezeigt. Eine Policy steuert die zweite Abfrage und gibt nur die letzten vier Ziffern der Kreditkartennummer zurück.

- Differential Privacy infiltriert „Rauschen“ in die Ergebnisse. Der Benutzer (z. B. ein Data Scientist) schreibt SQL-Abfragen gegen die Originaldaten, während die Software zu detaillierte Abfragen über-

wacht und verhindert. Abfragen, die viele Zeilen gruppieren, werden mit einem zusätzlichen Rauschen zurückgegeben, während Abfragen für einzelne Zeilen blockiert werden. Das folgende Diagramm (siehe Abbildung 2) zeigt eine nicht blockierte (in blau) und eine blockierte Abfrage (in rot). Apple, Google und viele andere verwenden diese Technik auch in Browsern oder Mobilgeräten, bevor Daten an einen zentralen Server gesendet werden. Typische Produkte sind Immuta oder Aircloak.

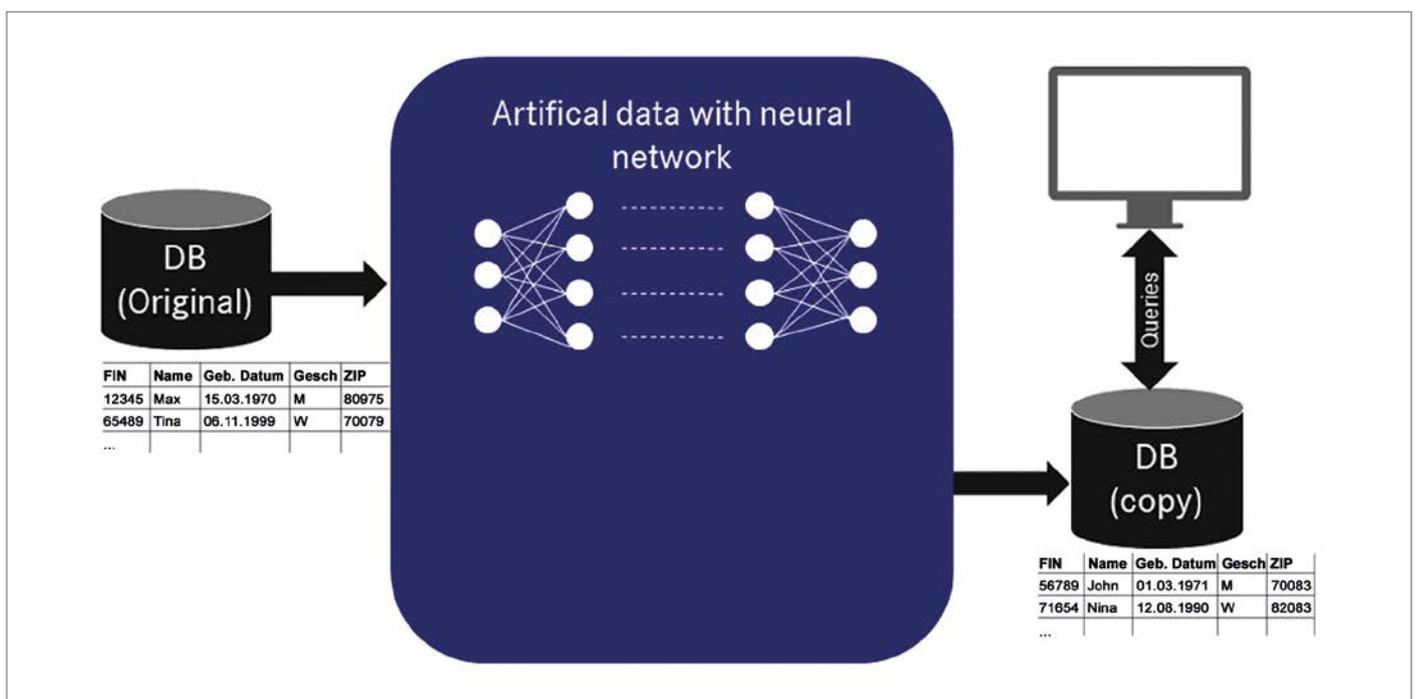


Abbildung 3: Gesamtansatz zum Erstellen einer synthetischen Datenkopie (Quelle: Andreas Buckenhofer)

conclusions can be made to (the identical) individuals.

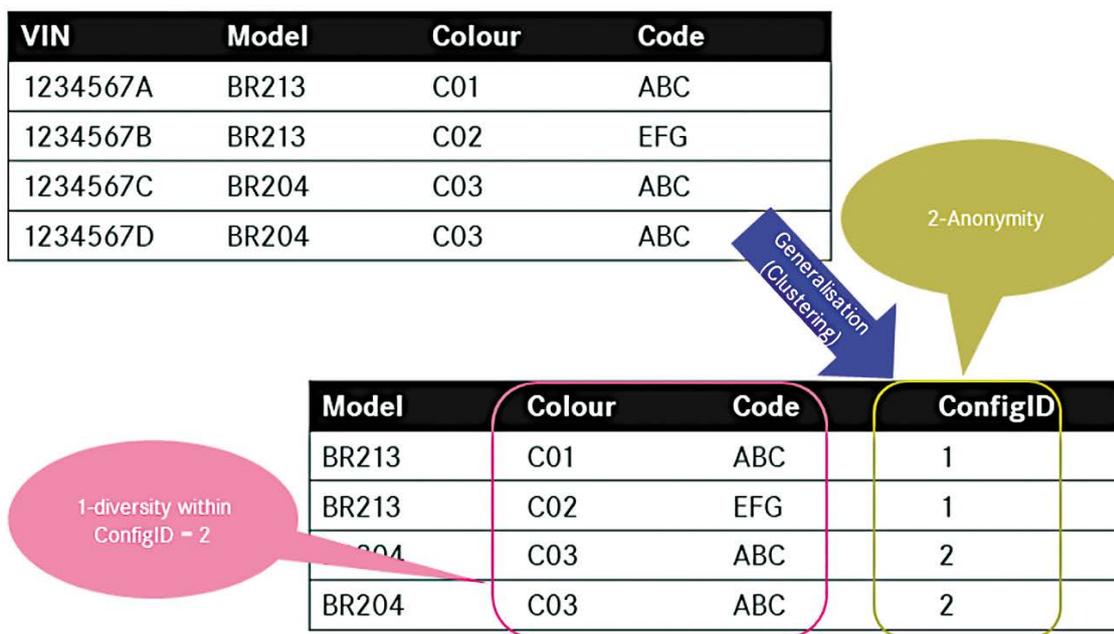


Abbildung 4: Originaldaten in einer Tabelle (Quelle: Andreas Buckenhofer)

tering): Das Ergebnis enthält nur noch aggregierte Daten. Das Diagramm in *Abbildung 4* zeigt die Originaldaten in einer Tabelle, einschließlich detaillierter Daten wie Fahrzeugidentifizierungsnummern (VINs). Die resultierende Tabelle enthält gruppierte Daten ohne Details. Das Ergebnis wird im Beispiel als 2-Anonymität bezeichnet, da die Anzahl der Mindestzeilen innerhalb eines Clusters zwei beträgt. ConfigID 1 wird in zwei Zeilen mit unterschiedlichen Daten angezeigt. Es ist nicht möglich, auf Einzelpersonen zu schließen. ConfigID 2 hat dagegen nur identische Werte. Das Ergebnis heißt 1-Diversity. Daher sollte ConfigID 2 aus dem resultierenden Datensatz entfernt werden, da Rückschlüsse auf (identische) Personen gezogen werden können.

Abschließende Bemerkungen

Anonymisierung ist eine komplexe Aufgabe. Anwendungsfälle und ihre Anforderungen sind sehr unterschiedlich. Das Anonymisierungsergebnis kann sehr schnell für Analytics-Zwecke unzureichend sein oder, wie Paul Ohm 2010 schrieb:

Data can be either useful or perfectly anonymous but never both.

Die Anonymisierung ist nicht nur ein einmaliger Schritt. Die Kombination mit neuen Daten kann zu unbekanntem Re-Identifizierungsrisiken führen. Der Anonymisierungsprozess und dessen Ergebnis müssen kontinuierlich überprüft werden.

Vertrauen wird bei der Arbeit mit Daten immer wichtiger – Ethik und Datenschutz müssen beachtet werden. Der gesamte Datenlebenszyklus muss den Datenschutz gewährleisten, von der Erstellung und Speicherung über die Auswertung und gemeinsame Nutzung bis hin zur Archivierung und Löschung von Daten. **Jeder ist für den vertrauensvollen Umgang mit personenbezogenen Daten verantwortlich.**

Über den Autor

Andreas Buckenhofer arbeitet bei Daimler TSS in der BU „Vehicle Platforms“ und verfügt über mehr als 20 Jahre Erfahrung in Data-Warehouse-Projekten. Seine praktischen Erfahrungen gibt er gerne in internen Vorträgen und als Sprecher auf internationalen Konferenzen weiter. An der Dualen Hochschule Baden-Württemberg hält er regelmäßig eine Vorlesung über Data Warehousing und Big Data. Er ist aktives Mitglied in der Datenbank

Community der DOAG und wurde von Oracle zum ACE Associate ernannt.

Quellen

[1] <https://www.youtube.com/watch?v=xHATN-hKakFc&t=1s>



Andreas Buckenhofer
andreas.buckenhofer@gmail.com



Agile Entwicklung ohne automatisiertes Testen = Schiffbruch

Jan Ott, Trivadis

In diesem Beitrag werde ich erläutern, wie sich der Wechsel von der traditionellen zur agilen Software-Entwicklung auf das Testen eines Projekts auswirkt. Sie werden etwas über agile Tests erfahren sowie darüber, welche Herausforderungen Tester in einer agilen Umgebung zu bewältigen haben und welche Vorteile daraus gezogen werden können. Dadurch erhalten Sie Einblick in Möglichkeiten, wie in Ihrem Unternehmen beziehungsweise bei Ihrem Produkt sowohl Produktivität als auch Kundenzufriedenheit gesteigert werden können.

Agil – Warum wirkt sich das auf die Software-Entwicklung aus?

Software wurde in der Vergangenheit traditionell in Zyklen von 3-6 Monaten entwickelt. Dabei nahm das Testen am Ende des Zyklus mehrere Wochen in Anspruch. Dies führte dazu, dass der Kunde lange auf Neuerungen oder Fehlerbehebungen warten musste. Heute ist dies nicht mehr zeitgemäß. Daher wurden die Zyklen im agilen Umfeld auf 2-3 Wochen gekürzt. Nach so einem Zyklus soll ein Software-Paket verfügbar sein, das ausgeliefert werden kann. Ob dies dann bis in die Produktion ausgeliefert wird, kommt auf die Gegebenheiten des Projekts an.

Es gibt sicher noch mehr Gründe, wieso Software-Entwicklung heute agil gemacht wird. Doch die Kürzung der Zyklen ist der wichtigste Grund, weshalb traditionelles Testen von Hand im agilen Umfeld zu Schiffbruch führt.

Was kann schiefgehen – ein Beispiel

Ein größeres Projekt mit 80+ Mitarbeiter/innen in mehreren Teams lief bei einem Kunden über einige Jahre. Unser Team hatte 10-12 Mitarbeiter*innen, davon 1-2 Kolleg*innen, die nur für das Testen zuständig waren. Eigentlich gute Voraussetzungen – oder?

Auf dem Papier war alles dokumentiert. Es gab Tests für die verschiedensten Testkategorien:

- Unit-Tests
- Regressions-Tests
- Integrations-Tests
- Abnahme-Tests
- ...

Es gab auch Continuous Integration/Continuous Delivery, abgekürzt CI/CD, also die automatisierte Integration der verschiedenen Teile der Software zu einem Paket sowie dessen automatisierte Auslieferung.

Leider hat es trotz all dieser Dokumenten immer wieder schwerwiegende Fehler in der Produktion gegeben. Auch traten bereits korrigierte Fehler zu einem späteren Zeitpunkt wieder auf, Termine wurden nicht eingehalten. Daher wurden die Dauer und der Aufwand für die Tests erhöht. In den Test-Zyklen war das ganze Team am Testen oder am Korrigieren der Fehler. Unsere Leistung verschlechterte sich, während unsere Fehlerquote zunahm. Das Ergebnis war ein sehr unzufriedener interner Auftraggeber.

Es lag aber nicht daran, dass die Kollegen/innen nicht gute Arbeit leisteten oder sich nicht genug einsetzten.

Was lief schief?

Die drei wichtigsten Punkte zeigten sich deutlich:

1. Es wurde nicht wirklich agil entwickelt. Eigentlich wurde das Wasserfallmo-

dell einfach in ein agiles Mäntelchen (Scrum) gehüllt. Ein Beispiel: Ein Zyklus dauerte zwar 3 Wochen. In Wirklichkeit waren es aber doch wieder 3-4 Monate, da nach den 3 Wochen nicht ein getestetes und ausrollbares Paket abgeliefert wurde. Stattdessen wurden in den 3 Wochen einfach nur geplante Tasks abgearbeitet.

2. Das Testen wurde bis auf ein paar Unit-Tests auf die Test-Zyklen am Ende des dritten Monats verschoben. Dies führte dazu, dass wohl viele Fehler gefunden wurden, dies aber erst spät. Durch all die in der Testphase erfolgten Korrekturen wurde alles wieder instabil.
3. Es wurde vorwiegend von Hand getestet.

Die Erfolgskriterien des Testens

Automatisiert

Tests müssen möglichst automatisiert ablaufen.

Was ist die Konsequenz, wenn die Tests von Hand gemacht werden? Es ist aus Zeitgründen nicht möglich, nach jeder Fehlerkorrektur wieder alles zu testen, daher werden Tests weggelassen. Dies führt natürlich zu einer reduzierten Testabdeckung, was im oben erwähnten Projekt dazu geführt hat, dass Fehler übersehen wurden.

Fortlaufend

Tests müssen fortlaufend ausgeführt werden.

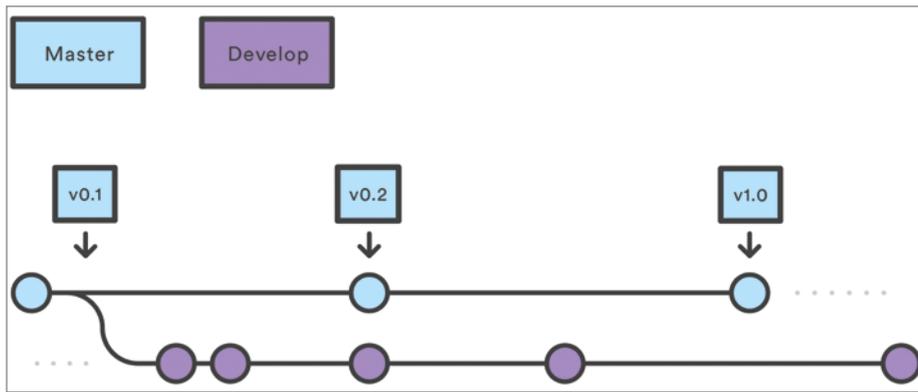


Abbildung 1: Einfacher Baum (Quelle: © Atlassian.com)

In meinem Projekt war das nicht der Fall. Die Tests wurden nur zum Teil während der Entwicklung ausgeführt. Somit kam es in der Test-Phase zu vielen Fehlern. Diese mussten dann in kurzer Zeit korrigiert werden, wodurch die Software/der Code in der Testumgebung zusätzlich instabil wurde. Das verunmöglichte ein sauberes Testen.

Integriert

Tests müssen Teil eines Integrations-/Deployment-Prozesses sein.

Da in meinem Projekt die Tests von Hand ausgeführt wurden, waren sie nicht Teil des Integrations-Prozesses. Es wurde während des Baus des Software-Pakets nur darauf geachtet, dass die Teile kompiliert und ausgerollt werden konnten. Wenn das erfüllt war, dann wurde das Paket als „bereit“ erklärt.

Schnell

Tests müssen schnell ausgeführt werden können.

Auch hier lag das Problem bei der manuellen Ausführung der Tests. Die Rückmeldungen ließen auf sich warten, was den ganzen Prozess verzögerte. Dennoch waren die Tester immer voll ausgelastet mit Vorbereiten der Tests, Dokumentation und anderen Aufgaben.

Versionierung

Tests und Code müssen lückenlos versioniert werden. Das umfasst den Code, die Tests, die Testdaten, die Resultate der Tests, den Code der Tools für die Automation, die Skripte und so weiter.

Leider gab es in meinem Projekt auch hier Lücken. Bei uns war es zum Beispiel die Komponente „Automations-Tool“, die nicht in ein Versionierungs-Tool eingeeckelt werden konnte. Wir hätten dafür im Tool ein

neues Projekt eröffnen müssen, womit die Prozesse bei der Weiterentwicklung hätten angepasst werden können. Wenn aber ein Fehler in der Produktion aufgetreten wäre, hätte das alte Projekt noch zur Verfügung gestanden. Dies wurde nicht gemacht, was zu Problemen führte. Es muss daher alles versioniert werden, am besten im Versionierungs-Tool. Ist dies nicht möglich, muss es mittels anderer Optionen erreicht werden.

Wichtige Schritte

Abgeleitet von der Applikations-Architektur und den zur Entwicklung verwendeten Tools muss festgelegt werden, auf welche Art und Weise CI/CD umgesetzt werden kann.

Die durch Architektur und Toolset gegebenen Rahmenbedingungen haben

Auswirkungen sowohl auf den Ablauf des Testens als auch auf das Test-Toolset.

Auch allfällig weitere Vorgaben seitens des Auftraggebers müssen berücksichtigt werden.

Versionierung

Fangen wir mit der Versionierung an. Da stellen sich unter anderem die Fragen:

- Wie versioniere ich?
- Was muss versioniert werden?
- Brauche ich einfache oder komplexere Versionierungs-Bäume? (siehe Abbildung 1 und 2)

Dazu gibt es gute Artikel, hier eine Auswahl:

- A successful Git branching model von Vincent Driessen <https://nvie.com/posts/a-successful-git-branching-model/>
- Atlassian – von hier sind auch die beiden Grafiken <https://www.atlassian.com/git/tutorials/comparing-workflows/gitflow-workflow>
- GitHub <https://guides.github.com/introduction/flow/>

Integration - fortlaufend, automatisiert

Diese Punkte laufen ineinander. Auch hier sind wieder Fragen zu beantworten:

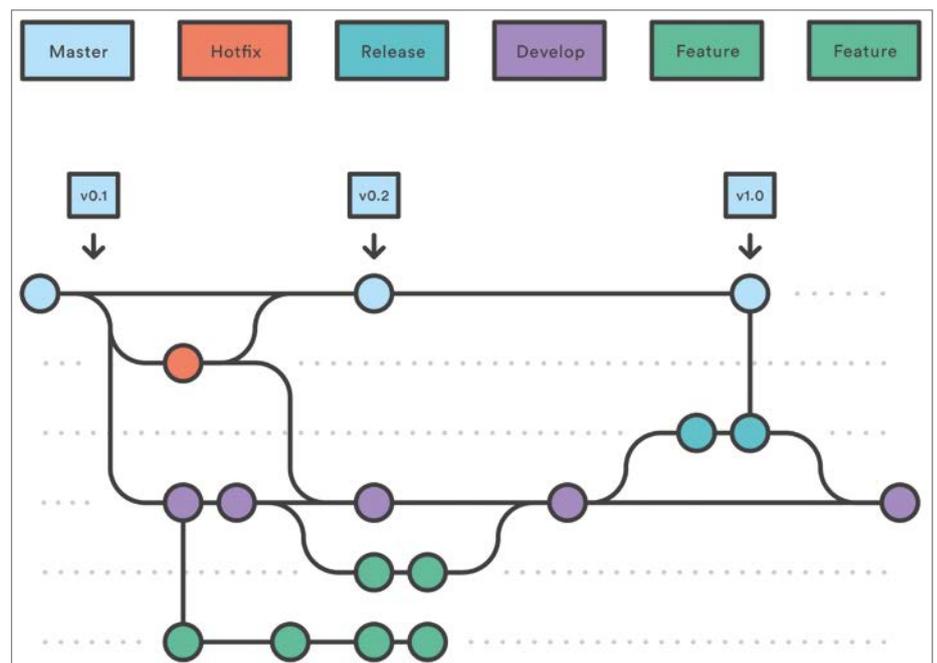


Abbildung 2: Komplexerer Baum (Quelle: © Atlassian.com)

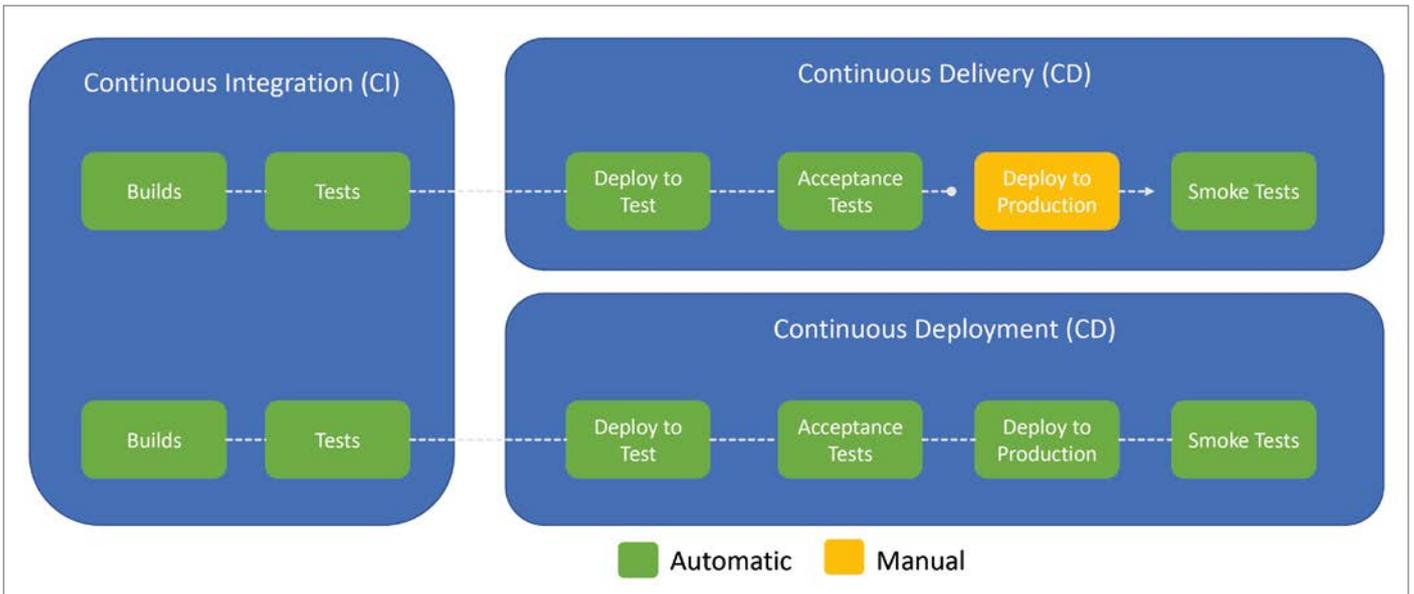


Abbildung 3: Beispiel eines CI/CD-Ablaufs (Quelle: Jan Ott)

- Was muss automatisiert werden?
- Welche Tools unterstützen meinen Prozess?
- Wie/Was kann ich versionieren?

Nehmen wir das Beispiel Jenkins. Mit Groovy können seit Version 2 die Pipelines als Skripte erstellt werden. Das bedeutet, dass diese in einem Versionierungstool eingetragt werden können. Mittels Jenkins werden die verschiedenen Teile der Integration oder des Deployments in einen Ablauf gebracht, der dann bei Bedarf immer und immer wieder ausgeführt werden kann.

Ablauf

Der Ablauf als Ganzes wird Continuous Integration und Continuous Delivery/Deployment genannt, abgekürzt CI/CD. Dabei geht es darum, dass bei bestimmten Ereignissen automatisch ein Prozess ausgelöst wird. Ein solches Ereignis kann zum Beispiel ein Check-in in das Versionierungstool sein. Dieser Prozess baut das Paket, testet es, checkt weitere Artefakte ein und rollt das Paket auf Development oder eine andere Umgebung aus.

Der Ablauf von Continuous Integration und Continuous Delivery/Deployment muss entworfen, dokumentiert und dann umgesetzt werden (siehe Abbildung 3).

Continuous Integration beinhaltet den Bau des auszuliefernden Pakets und verschiedene Tests. Das bereitgestellte Pa-

ket wird dann entweder mit Continuous Delivery oder Continuous Deployment ausgeliefert.

Wir haben uns in den meisten Projekten für Continuous Delivery entschieden, da wir den Prozess des Deployments in die Produktion manuell starten wollten. Dieser manuelle Schritt war uns wichtig, um eine zusätzliche Sicherheit zu haben, dass nicht automatisch ein Paket in der Produktion appliziert wird.

Es gibt heute Projekte, bei denen automatisch in die Produktion ausgerollt wird. Das geht manchmal sogar so weit, dass mehrmals täglich ausgerollt wird. So geht jede Änderung, jede Fehlerbehebung nach erfolgreichen Tests direkt in die Produktion.

Im beschriebenen Projekt wurde der Ablauf in einer Grafik festgehalten (siehe Abbildung 4). Der Prozess wurde dabei nicht durch ein Check-in ausgelöst, sondern von Hand. Zudem wurde eine

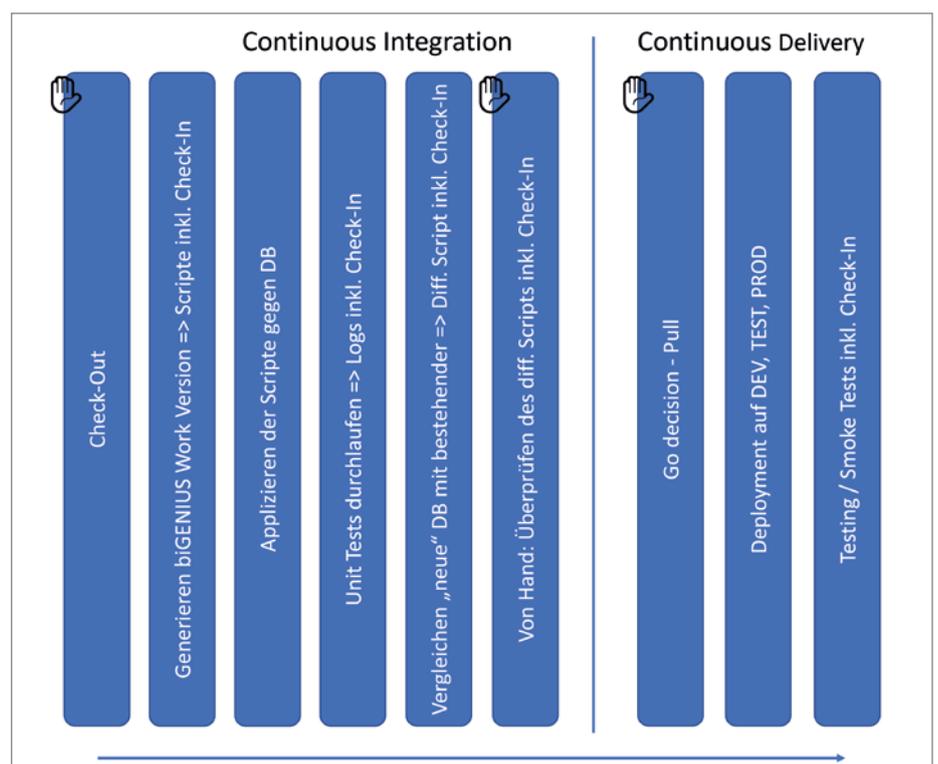


Abbildung 4: CI/CD-Beispiel (Quelle: Jan Ott)

weitere Prüfung von Hand eingeführt, nämlich die, ob die Differenz-Skripte (SQL Statements, um die Datenbank von Stand A auf Stand B zu migrieren) so belassen werden können. Das Tool, das wir verwendeten, hat hier gute Rückmeldung gegeben: So wurde zum Beispiel gemeldet, ob es zu Datenverlust kommen könnte. Einen weiteren Eingriff von Hand gab es, um das Paket auf die verschiedenen Umgebungen auszurollen. Diese Punkte sind mit den gelben Händen markiert.

Für jeden dieser Schritte mussten wir die entsprechenden Skripte schreiben, die in eine Jenkins-Pipeline integriert wurden.

Schnell

Dieser Punkt ist noch offen. Geschwindigkeit ist hier wichtig. Wir hatten uns daher entschieden, zweigleisig zu fahren.

Bei einem Prozess wollten wir schnell sein, denn langsame Testprozesse werden von Entwicklern als Zeitverschwendung empfunden und gerne ignoriert.

Zum Beispiel soll eine Funktion geändert werden. Der Entwickler codiert die Änderung und checkt sie ein. Zudem werden entsprechende Unit-Tests implementiert und eingecheckt. Die Jenkins-Pipeline wird vom Entwickler angestoßen und die Unit-Tests werden ausgeführt. Dies kann mehrmals, schnell hintereinander, immer mit kleinen Änderungen angestoßen werden. Ist alles so, wie es sein sollte, wird alles eingecheckt. Hier ist Geschwindigkeit wichtig.

Der andere Prozess darf länger laufen, da hier niemand direkt auf das Resultat wartet, zum Beispiel jede Nacht. Es können dabei auch Regressions-, Performance- und weitere Tests durchlaufen werden.

Testdaten

Bevor wir zum Abschluss kommen, noch ein paar Gedanken zu Testdaten.

Ich benutze meist zwei oder mehrere Sätze von Daten. Der erste Satz ist klein und sehr durchdacht. Ich bin ein Verfechter von kleinen, guten Datensätzen für Unit-, Regressions- und sogar erste Integrations-Tests. Die Entwickler kennen die Daten und die Tests sind schnell. Es treten auch seltene „unbeabsichtigte“ Fehler auf.

Die anderen Sätze sind größer. Meist wurden diese generiert, abgeleitet aus produktiven Daten oder sind produktive

Daten, die anonymisiert wurden. Es werden nie unveränderte produktive Daten verwendet. Die Ausnahme ist eine Kopie der Produktion für Abnahmetests vom Business oder für Load-Tests. Doch dies muss mit der Rechtsabteilung geprüft werden, denn die Gesetze lassen hier wenig Spielraum zu. So kann es zu empfindlichen Bußen kommen.

Gedanken zum Abschluss

Testen ist ein Thema, seit es Software-Entwicklung gibt. Leider wird es noch immer vielfach sehr stiefmütterlich behandelt. In den meisten Unternehmen wird auch heute noch mehr schlecht als recht getestet.

Dies hat leider zur Folge, dass wir in der agilen Welt Schiffbruch erleiden. Was früher noch überspielt beziehungsweise umgangen werden konnte, tritt im agilen Umfeld nun umso mehr ans Tageslicht. Fehler, die die automatisierten Tests aufzeigen würden, bleiben im Dunkeln.

Es muss deshalb umfassender getestet werden. Wenn das manuell erledigt wird, dauert es zu lange. Nur mit automatisiertem Testen kann eine vernünftige Geschwindigkeit erreicht werden. Das Testergebnis muss schnell, wenn nicht umgehend, zur Verfügung stehen.

Die Kosten sind dabei nicht zu vernachlässigen. Nehmen wir wieder das Beispielprojekt. In unserem Team waren 1-2 Personen vollzeitlich mit dem Testen beschäftigt. Sie haben während der Sprints versucht zu testen, wurden aber, da sie nur von Hand testeten, vom Entwicklerteam buchstäblich überrollt mit all den Änderungen. Die Tester wurden dann während der drei Test-Wochen von den anderen 8-10 Personen des Teams zu 50-80% unterstützt. Das Resultat war ein Super-GAU.

Bauen wir jedoch einen CI/CD-Ablauf mit den entsprechenden Tests und Testdaten und setzen die gleichen Ressourcen ein, ist das Ergebnis ein anderes. Wenn das Projekt schon am Laufen ist, wird es einige Zeit brauchen, bis Ergebnisse sichtbar werden. Doch mit der Zeit wird die Abdeckung immer besser. Es werden mehr und mehr Fehler frühzeitig gefunden. Die Entwickler können so sicher sein, dass ihre Änderungen umfassend getestet werden. Nicht, dass der Code vollständig fehlerfrei ist – wir wissen ja alle, fehlerfreien Code gibt es nicht –, doch kann ich mir sicher

sein, dass wichtige Zahlen getestet werden, die Performance stabil ist, die Teile gut zusammenarbeiten und noch mehr.

Zudem hat sich in den letzten Jahren in der Tool-Welt einiges getan. Auch in der neuen Cloud-Welt gibt es entsprechende integrierte Toolsets, die angeboten werden. Dadurch sinkt der Aufwand.

Mein Tipp: Bei einem neuen Projekt gleich von Anfang an CI/CD- und Test-Aufwand einplanen. So ist es am einfachsten aufzubauen. Bei einem bestehenden Projekt sofort beginnen. Jeder Test, der automatisiert durchläuft, ist ein Schritt in die richtige Richtung.

Am Anfang eines Projektes will man möglichst schnell etwas zeigen, Resultate liefern. Doch je mehr Code, Daten und Objekte vorhanden sind, desto komplexer wird es und die Geschwindigkeit sinkt. Der Aufwand, CI/CD dann einzubauen, wird größer und es wird schwieriger zu argumentieren. Wirft man jedoch das Steuer nicht herum, wird man Schiffbruch erleiden. Garantiert!

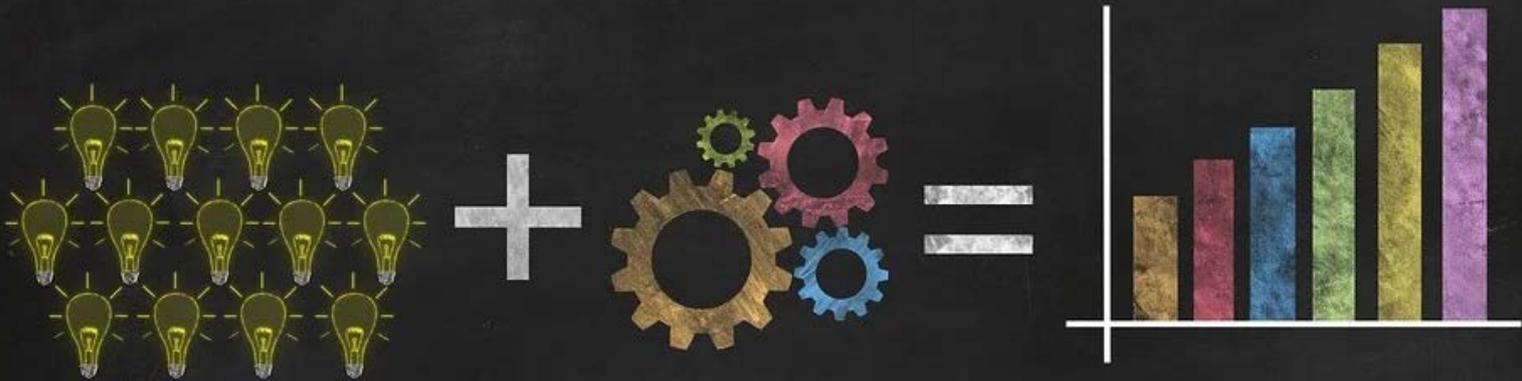
Daher empfehle ich, CI/CD einzuführen und Test für Test auszubauen. So wie man ein Haus Stein für Stein baut oder einen Marathon Schritt für Schritt läuft.

Über den Autor

Jan Ott arbeitet seit über 25 Jahren in der Softwareentwicklung. Der Schwerpunkt seiner Tätigkeit liegt im Bereich Data Warehousing mit Oracle- und Big-Data-Technologien. Seit vielen Jahren gibt er sein Wissen im Oracle- und Big-Data-Umfeld weiter, sowohl als Trainer in Kursen als auch als Speaker auf nationalen und internationalen Konferenzen.



Jan Ott
jan.ott@trivadis.com



Requirements Engineering – wieso, weshalb, warum?

Tobias Schweiker

Anforderungen für Anwendungen? Wir wissen doch, was die Anwenderin oder der Anwender braucht. Und wenn das Ergebnis diese nicht zufrieden stellt? Anforderungsermittlung und Requirements Engineering wird leider immer noch in vielen Projekten vernachlässigt. Damit eine Anwendung oder ein Produkt hinsichtlich Funktionalität, Kosten und Zeit zur Zufriedenheit des Kunden erstellt werden kann, ist gutes Requirements Engineering unerlässlich.

Dieser Artikel zeigt die Benefits eines guten Requirements Engineerings auf und gibt Tips und Argumente für einen frühzeitigen und erfolgreichen Einsatz von Anforderungserhebung, -Analyse und Verwaltung.

„Habe ich den Sachverhalt für die Ergänzung der Anforderung richtig dargestellt?“ Diese Frage bei einer Besprechung über eine Konzeption löste Erstaunen aus. „Aber das ist doch selbstverständlich“, bekam ich als Antwort von einem Teilnehmer aus dem Fachbereich des Unternehmens, mit dem Unterton „das braucht man doch nicht zu erwähnen“. Ist das aber für eine Software-Entwicklerin oder einen Software-Entwickler, die oder der sich mit der fachlichen Materie nicht so auskennt, auch selbstverständlich? Oder

kann es nicht eher passieren, dass das für den Fachbereich Selbstverständliche dann im Endprodukt fehlt? In solchen Situationen kann gutes Requirements Engineering zum Erfolg eines Projektes beitragen.

Wieso es Requirements Engineering braucht.

Ich kenne noch die Zeiten, als es abgegrenzte Verantwortungsbereiche gab: Die Fachabteilungen nutzen das Produkt und

gaben ihre Änderungswünsche an Konzeptionäre weiter. Diese hatten Fachwissen und erstellten mit diesem Fach- und IT-Konzepten. Die Fachkonzepte wurden von den Fachabteilungen abgenickt, die IT-Konzepte an die Programmierer „über den Zaun geworfen“. Nach kurzer Zeit als Entwickler habe ich mich gefragt „Verstehst du auch, was du liest?“ [1] Spätestens als ich einmal feststellen musste, dass der Inhalt eines IT-Konzepts nicht zum Titel passte, war mir klar, dass da etwas mit der Kommunikation schief gegangen war.

Verschiedene Untersuchungen über die Gründe und Ursachen für Erfolg und Misserfolg von Projekten (z.B. von Forbes/Gartner oder der Standish Group [2] und [3]) haben ergeben, dass Kommunikation der Schlüsselfaktor für den Erfolg ist.

Weitere Haupterfolgsfaktoren für erfolgreiche Projekte sind:

- Einbindung der Endbenutzer und -benutzerinnen
- Transparenz
- Klare Anforderungen
- Verbindlichkeit
- Professioneller Umgang mit Risiken und Unsicherheiten
- Unterstützung durch das obere Management

Die Hauptpunkte, die zum Scheitern der Projekte führen sind:

- fehlende Zuarbeit durch Benutzerinnen und Benutzer
- unvollständige/unklare Anforderungen
- häufige Anforderungsänderungen

Gutes Requirements Engineering kann die Erfolgsfaktoren fördern und die Misserfolgsfaktoren beziehungsweise deren Risiken minimieren.

Was ist nun Requirements Engineering? Eine allgemein anerkannte Definition ist die des IREB (International Requirements Engineering Board):

Requirements Engineering ist ein systematischer und disziplinierter Ansatz zur Spezifikation und zum Management von Anforderungen mit den folgenden Zielen:

1. Die relevanten Anforderungen zu kennen, Konsens unter den Stakeholdern über die Anforderungen herzustellen, die Anforderungen konform zu vorgegebenen Standards zu dokumentieren und die Anforderungen systematisch zu managen.
2. Die Wünsche und Bedürfnisse der Stakeholder zu verstehen und zu dokumentieren.
3. Die Anforderungen zu spezifizieren und zu managen, um das Risiko zu minimieren, dass das System nicht den Wünschen und Bedürfnissen der Stakeholder entspricht.

Zentrales Objekt sind also Anforderungen. Eine Anforderung ist (laut [5]):

1. Eine Bedingung oder Fähigkeit, die von einem Benutzer (Person oder System) zur Lösung eines Problems oder zur Erreichung eines Ziels benötigt wird.
2. Eine Bedingung oder Fähigkeit, die ein System oder Teilsystem erfüllen oder besitzen muss, um einen Vertrag, eine Norm, eine Spezifikation oder andere, formell vorgegebene Dokumente zu erfüllen
3. Eine dokumentierte Repräsentation einer Bedingung oder Eigenschaft gemäß (1) oder (2).

Damit ergeben sich folgende Tätigkeiten im Requirements Engineering:

- Erheben und Ermitteln der Anforderungen
- Dokumentieren und Verwalten der Anforderungen
- Prüfen, ob Anforderungen zusammenpassen, sich nicht widersprechen und vollständig sind. Unvollständige und fehlende Anforderungen müssen nach erhoben werden.

Weshalb Requirements Engineering (RE) alle Beteiligten unterstützen kann

RE vermittelt zwischen den Welten (Fachabteilung, IT, Management) und verbessert damit die **Kommunikation**. Verschiedene Beteiligte an Projekten denken unterschiedlich und sprechen oft verschiedene „Sprachen“. Das Requirements Engineering übersetzt bzw. definiert Begrifflichkeiten und kann auf fehlendes Begriffswissen anderer Beteiligter hinweisen. So trägt es zum Verständnis komplexer Sachverhalte bei. Glossare enthalten Begriffsdefinitionen, Begriffsmodelle zusätzlich Beziehungen zwischen den Begriffen.

RE holt alle relevanten Personen (Stakeholder) ins Projekt-Boot und **bindet** damit unter anderem die **Endbenutzerinnen und -benutzer** ein. Zu Beginn eines Projekts wird überlegt, wer in irgendeiner Weise etwas zum erwarteten Produkt beizutragen hat. Dies sind natürlich in erster Linie die Anwenderinnen und Anwender, aber auch viele andere Personen oder Personengruppen. Der (Software-)Betrieb hat seine Vorstellungen und vielleicht Vorgaben, wie das Produkt in die Systemlandschaft eingebunden werden soll. Die Entwickler/innen haben sich Entwicklungsrichtlinien gegeben, die schon zu Beginn berücksichtigt werden sollten, aber auch zum Erfolg des Projekts beitragen können. Darüber hinaus hat auch das Management, sowohl des Kunden als auch des Lieferanten, Vorstellungen und Ideen zum Produkt. Zu beachten sind weiterhin gesetzliche Rahmenbedingungen und Normen. Durch RE entfällt somit das Problem *fehlender Zuarbeit durch Benutzer*.

RE gewährleistet die Nachverfolgbarkeit der Anforderungen und schafft damit **Transparenz**. Anforderungen und Ände-

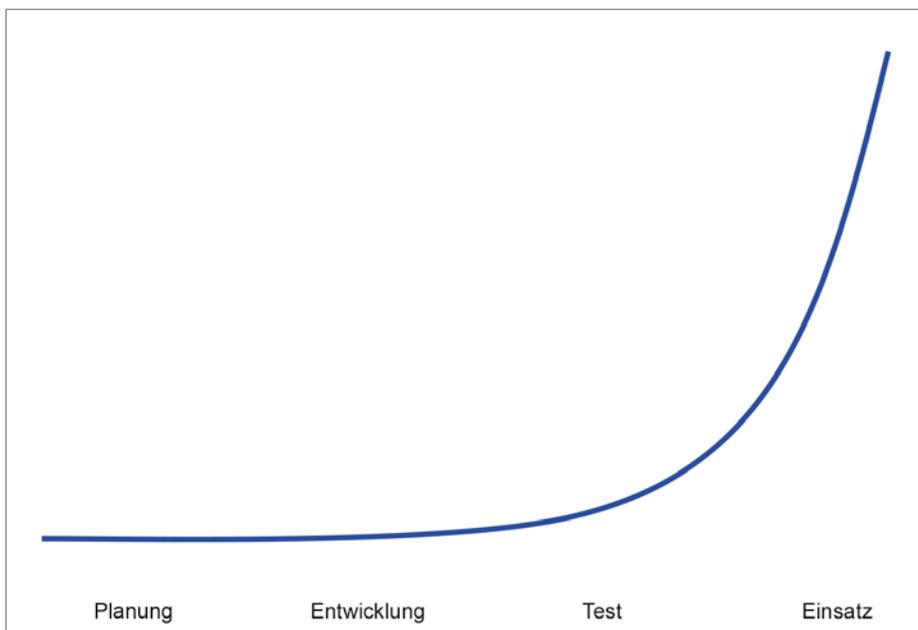


Abbildung 1: Zehnerregel der Fehlerkosten (Quelle: Tobias Schweiker)

ungen (Change Requests) werden strukturiert dokumentiert. Damit können sich alle Projektbeteiligten, insbesondere die Projektleitung, schnell und einfach über den aktuellen Projektstand informieren. Die Nachverfolgbarkeit der Änderungen minimiert auch die Risiken *häufiger Anforderungsänderungen*.

RE sorgt für **klare Anforderungen**, z.B. indem Satzschablonen definiert werden, anhand derer die Anforderungen beschrieben werden. Auch Modelle des Produkts können für verständliche Anforderungen sorgen. In graphischen Diagrammen, mit denen die Modelle meist erstellt werden, können viele Sachverhalte verständlicher dargestellt werden als nur mit Text. Nebenläufigkeiten sind in Diagrammen leichter zu erkennen. Solche Modelle können mit BPMN, UML oder anderen Notationen erstellt werden. Mit solchen Mitteln hilft RE der IT, die Anforderungen zu verstehen und gibt den Stakeholdern die Gewissheit, dass ihre Anforderungen richtig umgesetzt werden. Oft kann man damit auch aus den Wünschen der Anwender die tatsächlichen Anforderungen herleiten und damit sicherstellen, dass das implementiert wird, was den Anwenderinnen und Anwendern nutzt. Und es gibt keine *unvollständigen oder unklaren Anforderungen*.

RE macht konkrete Vorgaben zum Test und sorgt damit für **Verbindlichkeit**. Mit der Dokumentation der Anforderungen hat man auch gleich die Grundlage für Testfälle. Es gab sogar eine RE-Methode, bei der Testfälle zur Konzeption der Anforderungen geschrieben wurden und anhand derer implementiert wurde.

RE nimmt auch Randbedingungen verschiedenster Art in den Blick und verwaltet sie wie die Anforderungen strukturiert. Die Randbedingungen können technischer Natur, beispielsweise Schnittstellen zu anderen Systemen sein oder zu berücksichtigende gesetzliche Vorgaben oder Industrienormen. Mit der strukturierten Vorgehensweise wird die Risikoanalyse und das Risikomanagement erleichtert und der **professionelle Umgang mit Risiken und Unsicherheiten unterstützt**.

RE hilft, die Entwicklung zu strukturieren und zu planen. Es gibt verschiedene Programme zur Verwaltung von Anforderungen. Damit kann man Anforderungen gruppieren und in Beziehung zueinander setzen. So lässt sich z.B. erkennen, ob und welche Anforderungen voneinander

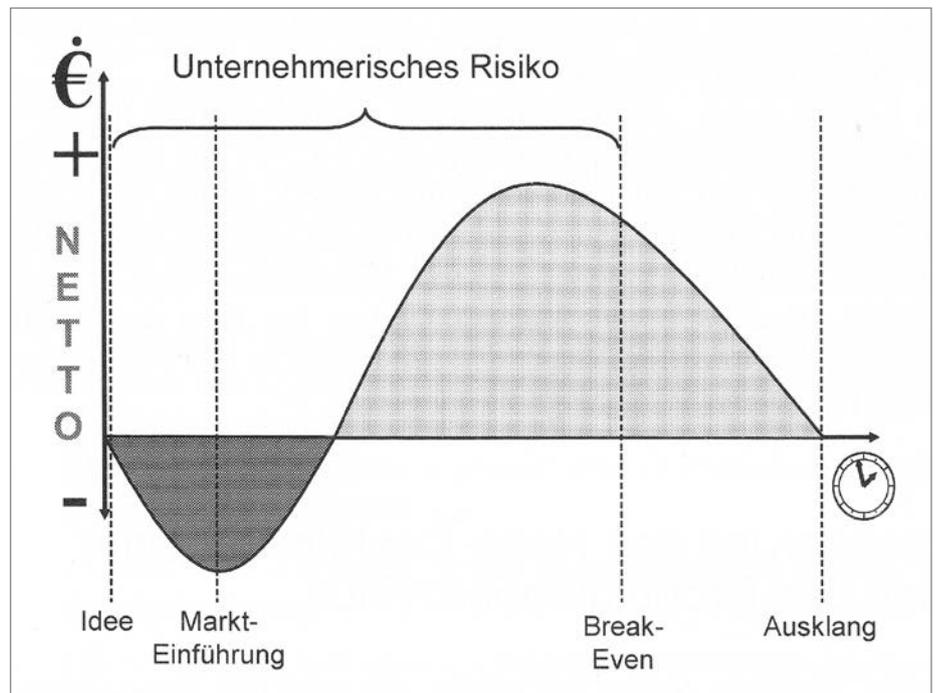


Abbildung 2: Produktlebenszyklus und Netto-Cashflow (Quelle: © Colin Hood, Rupert Wiebel [6])

abhängen und deshalb gemeinsam implementiert werden sollten.

Warum Requirements Engineering Kosten sparen und Gewinn erhöhen kann.

Kosten sparen? „Requirements Engineering kostet doch Arbeitszeit und damit Geld. Und wir können dann ja gar nicht gleich mit der Implementierung anfangen.“ Das ist zwar richtig, aber nur ein Teil der Wahrheit - leider der, der oft dazu führt, dass Requirements Engineering vernachlässigt wird. Schaut man sich aber die gesamte Projektlaufzeit oder den Lebenszyklus eines Produkts an, wird man feststellen, dass man mit Requirements Engineering tatsächlich Kosten sparen und sogar mehr Gewinne erwirtschaften kann.

Erfahrungen aus Projekt- und Qualitätsmanagement haben gezeigt, dass sich die Kosten für die Fehlerbehebungen in jeder Phase um den Faktor 10 steigen (siehe *Abbildung 1*).

Das kann jeder nachvollziehen: Korrekturen in der Planung sind schnell erledigt. Fällt während der Entwicklung eine Unstimmigkeit oder eine fehlende Anforderung oder ein offener Punkt (wie soll das jetzt gemacht werden?) auf, muss dies erst neu geplant und dann als Software-Änderung in den Prozess eingeführt werden.

Wenn ein Fehler erst bei der Anwenderin oder dem Anwender entdeckt wird, muss dies erst wieder neu in die Planung einbezogen werden. Dann muss die Korrektur an der richtigen Stelle, unter Berücksichtigung evtl. betroffener Funktionen und ohne diese zu verändern, programmiert werden. Daran schließen sich dann umfangreiche Tests an, auch um sicherzustellen, dass sich am Gesamtsystem außer der Korrektur nichts geändert hat. Wenn das System, wie es meist der Fall ist, bereits für neue Anforderungen weiterentwickelt wird, muss die Korrektur auch in der Versionsverwaltung richtig berücksichtigt werden. Meist sind dann mehrere Entwicklungszweige mit entsprechendem Verwaltungsaufwand nötig. Stillstandskosten und der Ärger der Anwendenden und Kunden kommen noch dazu.

Je früher also ein Fehler oder eine fehlende oder unvollständige Anforderung entdeckt wird, desto geringer sind die Kosten, diesen Mangel zu beheben. Es gibt in [6] auf den Seiten 28-30 eine ausführliche Beispielrechnung, wie hoch die Kosteneinsparungen sein können.

Man kann mit Requirements Engineering aber auch den Ertrag und damit den Gewinn erhöhen. Die *Abbildung 2* zeigt einen typischen Verlauf von Kosten und Ertrag für einen Produktlebenszyklus. Zunächst kostet die Entwicklung, erst später wird mehr eingenommen als investiert.

Mit Requirements Engineering können die Entwicklungszeit und die Entwicklungskosten reduziert werden. Die Anforderungen sind früher konsolidiert, und in der Entwicklung tauchen weniger Fragen und Unklarheiten auf. Man braucht weniger Entwicklungszyklen zur Fertigstellung. Damit kann das Produkt früher verkauft werden und die Einnahmen decken noch früher die Entwicklungskosten (siehe Abbildung 3).

Fazit

Gutes Requirements Engineering unterstützt alle Projektbeteiligten bei ihren Aufgaben im Projekt und verbessert die Kommunikation zwischen ihnen. Es hilft, die Anforderungen eindeutig zu dokumentieren, das Projekt zu strukturieren und Risiken zu erkennen. Damit können Kosten gespart und der Gewinn erhöht werden. Probieren Sie's aus.

Quellen

- [1] Die Bibel (Lutherübersetzung 2017), Apostelgeschichte 8 Vers 30
- [2] Projektmanagement Blog: CHAOS Report: Oft zitiert, aber was steckt dahinter? (<https://pm-blog.com/2010/01/29/chaos-report-viel-zitiert-aber-was-steckt-dahinter/>), abgerufen am 29.10.2020
- [3] Chaos Report der Standish Group, zitiert aus <https://de.wikipedia.org/wiki/Chaos-Studie>
- [4] Klaus Pohl, Chris Rupp (3. Auflage, 2011): Basiswissen Requirements Engineering – Aus- und Weiterbildung nach IREB-Standard zum Certified Professional for Requirements Engineering Foundation Level. dpunkt.verlag.

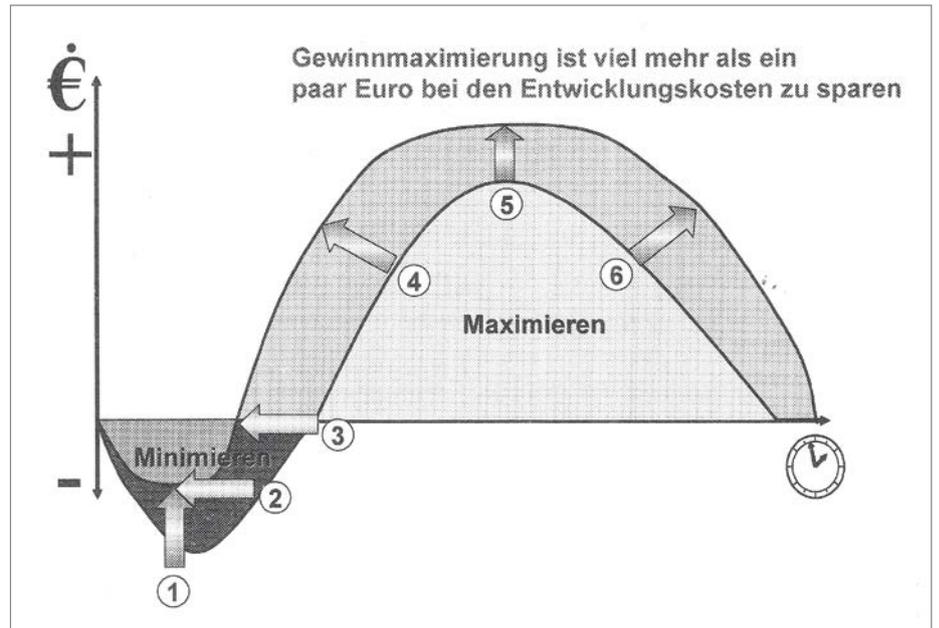


Abbildung 3: Gewinn erhöhen (Quelle: © Colin Hood, Rupert Wiebel [6])

- [5] IREB; Pohl, K.; Rupp, C.: Basiswissen Requirements Engineering: Aus- und Weiterbildung zum Certified Professional for Requirements Engineering – Foundation Level nach IREB-Standard. dpunkt, Heidelberg, 2009
- [6] Colin Hood, Rupert Wiebel (2005): Optimieren von Requirements Management und Engineering, Springer Verlag Berlin Heidelberg

der Industrie konnte er vielfältige Erfahrungen mit dem Requirements Engineering sammeln. In der Development Community der DOAG engagiert er sich als Themenverantwortlicher Requirements Engineering.

Über den Autor

Tobias Schweiker ist Datenbank-Anwendungsentwickler und zertifizierter Requirements Engineer (Certified Professional for Requirements Engineering (CPRE) Advanced Level). In vielen Projekten für verschiedene Kunden in der Versicherungsbranche und



Tobias Schweiker
Tobias@Schweiker-RE.de

JAVA 15 im Detail

DOAG Online

Im September 2020 wurde das derzeit aktuelle Java 15 veröffentlicht. Sollten Sie die Inhalte des Updates verpasst haben oder noch einmal einen genaueren Blick darauf werfen wollen, haben Sie nun die Gelegenheit.

Aufgrund des seit einschließlich Java 11 geltenden halbjährigen Update-Zyklus erscheint alle sechs Monate ein neues Update des beliebten JDK. Der Veröffentlichungs-

termin für das nächste Update, Java 16, ist für März 2021 angesetzt. In seinem Artikel ab Seite 12 in der **Java aktuell 1/21** widmet sich **Falk Sippach**, embarc und JUG

Darmstadt, den relevanten Punkten des letzten Updates aus Entwicklersicht (siehe: www.doag.org/de/home/news/java-aktuell-012021-werkzeuge-tools-frameworks/detail/).



MariaDB-/MySQL-Stolperfallen – und wie komme ich da wieder raus? – Teil 1

Olli Sennhauser, FromDual

Dem Wunsch der Konferenz-Teilnehmer folgend, nicht immer nur „New Features“ vorzustellen, sondern auch mal „War Stories“ aus der Praxis zu liefern, habe ich mir vorgenommen, die häufigsten und mühsamsten Stolperfallen zu beleuchten, in die man als unbedarfter Admin oder Entwickler für MariaDB oder MySQL-Datenbanken laufen kann.

Viele der weiter unten beschriebenen Probleme treten auch bei anderen RDBMS wie Oracle oder PostgreSQL auf. Daher können die folgenden Tipps auch auf diesen Systemen angewandt werden.

Im ersten Teil des Artikels stelle ich einige allgemeine oder schon länger bekannte Stolperfallen vor. Diese können auch für nicht MariaDB-/MySQL-Nutzer von Interesse sein. Im zweiten Teil, in der nächsten Red-Stack-Ausgabe 3/2021, werde ich dann mehr auf MariaDB-/MySQL-spezifische Stolperfallen eingehen.

MySQL ≠ MariaDB

Eines der ersten Probleme, das wir bei Kunden feststellen, ist, dass die meisten Nutzer immer noch nicht zwischen MariaDB und MySQL unterscheiden. Diese beiden Produkte haben sich in den letzten 10 Jahren dermaßen auseinanderentwickelt, dass man heute von einem Fork sprechen muss. Zwar sind die meisten Ei-

genschaften immer noch gleich oder zumindest sehr ähnlich, aber im Detail, dort wo es darauf ankommt, unterscheiden sich die beiden Produkte bereits signifikant. Am ehesten sieht man die Unterschiede in den beiden Optimizern, die unterschiedliche Möglichkeiten haben und somit unterschiedliche Resultate mit unterschiedlichen Query-Laufzeiten liefern können, sowie im Bereich Hochverfüg-

barkeit, wo sich inzwischen verschiedene HA-Technologien etabliert haben. Auch neue Features sind nicht mehr auf beiden Systemen gleich verfügbar.

Ab MariaDB Version 10.3 und MySQL Version 8.0 muss man ernsthaft anfangen zu differenzieren.

Und wie komme ich da wieder raus? Zuerst muss man ein Bewusstsein in der Firma schaffen. Dann sollte man eine Ent-

scheidung treffen: Will ich eine Single-Vendor- oder eine Multi-Vendor-Strategie fahren – und falls Single-Vendor-Strategie: Für welchen Anbieter soll ich mich entscheiden? Ein weiteres Entscheidungskriterium ist: Welches der beiden Produkte unterstützt der Softwarehersteller denn überhaupt? Wenn das gewählte RDBMS nur zweite Wahl ist oder „auch“ unterstützt wird, ist potenzieller Ärger in naher oder ferner Zukunft vorprogrammiert.

Wenn man eigene, DBMS-nahe Software schreibt, kommt man früher oder später nicht darum herum, im Code, je nach eingesetztem Produkt, die einzelnen Codepassagen anzupassen. Das kann dann wie in *Listing 1* aussehen.

KISS

KISS ist eines meiner Lieblingsmantras. Das Konzept stammt aus den 60er Jahren und ist von der U.S. Navy. Wobei es auch schon wesentlich früher Gedanken in diese Richtung gab. KISS ist ein Akronym für „Keep it stupid, simple“. Je nach Weltauffassung werden auch minimal andere Interpretationen des Akronyms geboten. Die Idee dahinter ist: Einfache Lösungen sind typischerweise robuster und gehen weniger schnell kaputt als komplizierte Lösungen (*siehe Abbildung 1*). Gerade in der IT neigen wir dazu, Lösungen unnötig kompliziert zu machen. Dies ist einerseits der Technikverliebtheit der involvierten Personen zu verdanken, andererseits haben natürlich äußere Beeinflusser auch kein Interesse an einfachen Lösungen, da dies ja den Nutzen der äußeren Beeinflusser dramatisch reduziert.

Und wie komme ich da wieder raus? Bei komplizierten Lösungen sollte man sich Gedanken dazu machen, ob es auch einfacher geht und ob wirklich alle Anforderungen zwingend und nötig sind, die eine Lösung kompliziert machen.

MyISAM Tabellen-Locking

Etwas aus der Mottenkiste gegriffen, aber leider immer noch aktuell. Zuerst jedoch etwas Hintergrundinformation dazu: MariaDB und MySQL bieten verschiedene sogenannte Storage Engines zum Speichern der Daten an. Das sind „Module“, die den Zugriff auf die Tabellen und somit die Daten durchführen. Bei der Erstellung der Ta-

belle kann ich entscheiden, welche Storage Engine verwendet werden soll. Die MyISAM Storage Engine war Default bis circa 2013. Dann wurde der Default auf InnoDB geändert. Das ist inzwischen 7 Jahre her! Wir treffen allerdings immer noch Nutzer an, die vereinzelt oder flächendeckend MyISAM-Tabellen einsetzen. Sei es aus Unwissenheit oder weil man den Schritt in die Zukunft noch nicht in Angriff genommen hat.

Warum sollte man MyISAM überhaupt meiden? Als Erstes sollte man wissen, dass diese Storage Engine inoffiziell von den Herstellern abgekündigt wurde und irgendwann aus dem Angebot herausfallen wird. Außerdem gibt es auch technische Gründe, aus denen MyISAM nicht so optimal ist. Als Hauptkriterium sehe ich für mich, dass MyISAM nicht crash-safe ist, das heißt, wenn die Hardware plötzlich den Strom verliert oder die Datenbank plötzlich oder unsauber gestoppt wird, können Datenverluste und/oder Tabellenkorruptionen auftreten. Und diese Situation sehen wir bei Firmen, die relativ kritische Systeme, wie zum Beispiel Webshops, betreiben. Das krassste Beispiel habe ich bei einem großen, internationalen Industriekonzern erlebt, der sein ganzes Rechnungswesen in MyISAM-Tabellen gespeichert hatte. Glücklicherweise ließ er sich davon überzeugen, dass MyISAM nicht die optimale Wahl war.

Ein weiteres Problem von MyISAM-Tabellen ist das Table-Level-Locking. Bei jedem Schreib- oder Lesezugriff wird (etwas vereinfacht ausgedrückt) immer die ganze Tabelle gesperrt. Entweder wird gelesen oder es wird geschrieben, beides gleichzeitig geht nicht (außer beim Lesen). Moderne Anwendungen haben jedoch oft die Situation, dass massiv parallel von Daten gelesen und auf die gleichen Daten geschrieben wird, und dies gleichzeitig. Da aber MyISAM-Tabellen nur eine Operation auf einmal zulassen, führt dies sehr schnell zu Staus auf der Tabelle. Der Durchsatz bricht zusammen oder wird gar nicht erst erreicht.

Das ist allerdings noch nicht das Ende des Problems. Bei heftigen Lastspitzen führt dies dazu, dass immer mehr Datenbank-Verbindungen aufgebaut werden, bis ein vorher definiertes Limit erreicht wird, was dann zu Verbindungsabbrüchen, Applikations-Fehlern und schließlich einem Ausfall der Applikation führen kann (*siehe Listing 2*).

Und wie komme ich da wieder raus? MariaDB und MySQL bieten eine weitere Storage Engine namens InnoDB an. Diese ist weit besser für massiv parallelen Zugriff geeignet, da sie mit Transaktionen arbeitet und die sogenannten ACID-Eigenschaften erfüllt. Eine Konvertierung ist technisch ganz einfach mit dem Befehl `ALTER TABLE ... ENGINE = InnoDB;` zu be-

```
// MySQL 5.7 ff.
if ( ($aInstanceInfo['branch'] == 'MySQL')
    && ($aInstanceInfo['mr_version'] >= '050700') )
{
    // MySQL 5.7.0 - 5.7.99
    if ( ($aInstanceInfo['mr_version'] >= '050700')
        && ($aInstanceInfo['mr_version'] <= '050799'))
    {
        ...
    }
    // MySQL 8.0 DOES support this feature but has different tables!
    elseif ( ($aInstanceInfo['mr_version'] >= '080000') ) {
        ...
    }
}
// MariaDB and MySQL 5.6 and older
else {

    // MariaDB 10.5 and newer
    if ( $aInstanceInfo['mr_version'] >= '100500') {
        ...
    }
    else {
        ...
    }
}
```

Listing 1: Code-Anpassungen

The complete Open Source High-Availability and Scale-Out Architecture stack

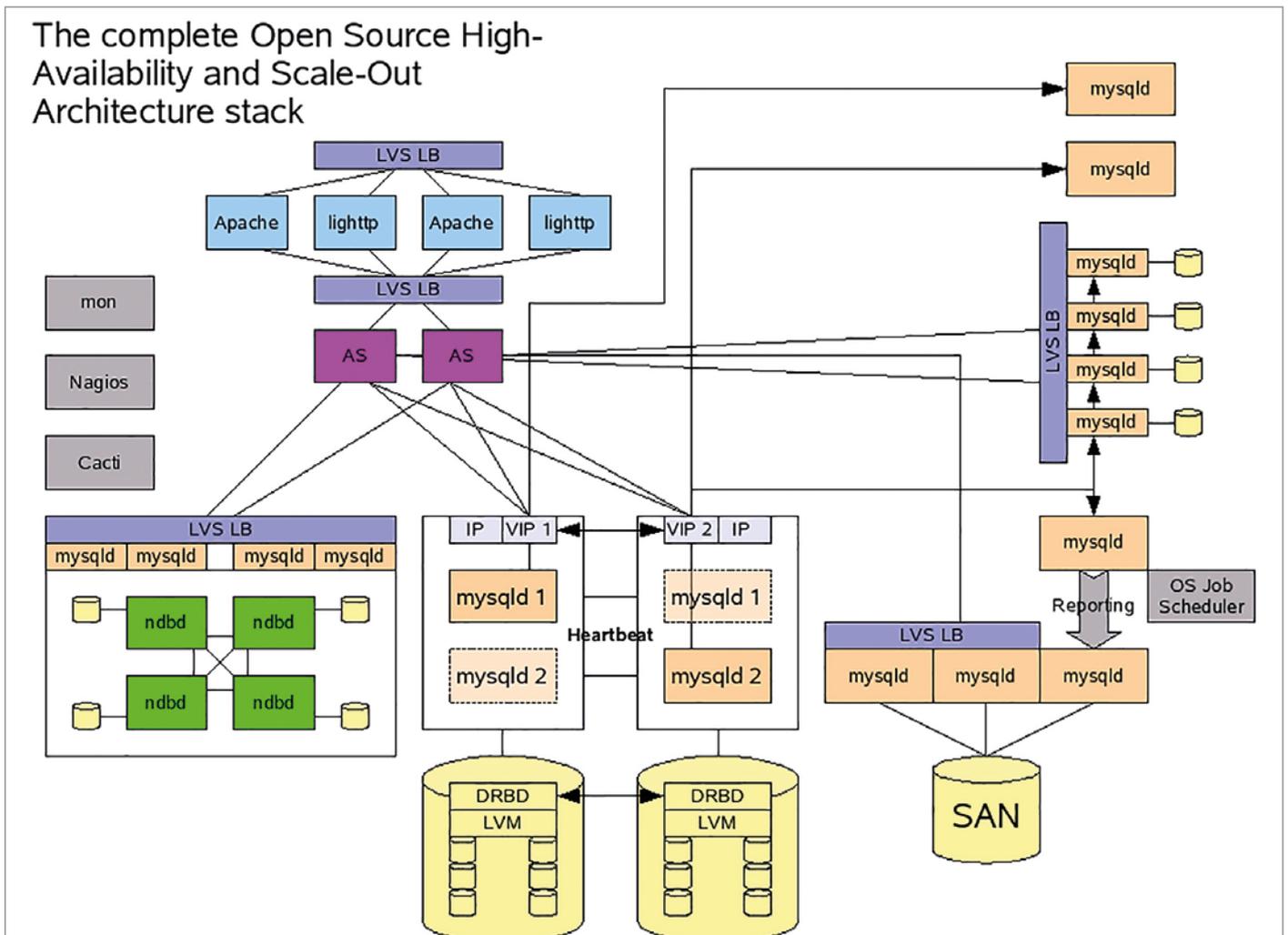


Abbildung 1: Hier ging es vor allem darum, alle möglichen Architekturen zu integrieren, und nicht darum, eine möglichst einfache Lösung zu bauen (Quelle: Olli Sennhauser)

werkstelligen. Jedoch liegt, wie so oft, der Teufel im Detail und die Konvertierung sollte daher vorher ausführlich getestet werden, um Nebenwirkungen rechtzeitig zu erkennen und zu eliminieren.

Große ibdata1-Datei

Auch bereits etwas in die Jahre gekommen ist das Problem der großen ibdata-Dateien. Zuerst zum Verständnis, wozu diese Dateien überhaupt gebraucht werden: Die ibdata-Dateien sind Dateien des sogenannten System Tablespace. Das ist ein Ort, an dem die InnoDB Storage Engine ihre eigenen Meta-Daten zum Verwalten der Tabellen speichert, ähnlich wie die Registry unter Windows. Das ist der Stand von heute. Früher wurden jedoch im System Tablespace auch Userdaten abgespeichert. Ältere Oracle-DBAs mögen sich vielleicht sogar noch daran erinnern, dass das bei Oracle

vor langer, langer Zeit auch mal der Fall war? Heute hingegen werden in MariaDB/MySQL die Userdaten ebenfalls in eigenen Tablespace abgelegt. Warum also ist der System Tablespace immer noch groß?

Dafür gibt es verschiedene Gründe: Es können immer noch alte Userdaten darin liegen, die noch nicht herausmigriert wurden. Im Weiteren schrumpft der System Tablespace nie mehr! Einmal aufgebläht, bleibt er für immer so groß. Das heißt, man schleppt vielleicht ganz viel Luft darin mit herum. Falls all diese Gründe nicht zutreffen, gibt es noch zwei weitere Möglichkeiten: InnoDB braucht für das Zurückrollen von Transaktionen sogenannte UNDO Records, die in einem Rollback-Segment liegen, welche wiederum in UNDO Logs liegen. Per Default liegen diese UNDO Logs im System Tablespace. Wenn jetzt eine große Transaktion läuft, wird diese UNDO Log vergrößert, bis die Transaktion beendet ist. Dies bläht allerdings auch den System Tablespace auf, da das UNDO Log

darin liegt. Nach der Transaktion schrumpft zwar das UNDO Log wieder, nicht aber der System Tablespace. Das gleiche Phänomen haben wir bei temporären Tabellen vom Typ InnoDB. Und wie kommen wir da wieder raus? Der System Tablespace schrumpft NIE mehr, außer wird „dumpen“ die gesamte Datenbank mittels mariadb-dump/mysql-dump und laden sie in eine neu erstellte, leere Datenbankinstanz, was bei sehr großen Datenbanken sehr zeitaufwändig ist. Damit der System Tablespace nicht noch einmal oder noch weiter anschwillt, kann man eigene UNDO Tablespaces und Temp Tablespaces definieren.

Noisy Neighbours

Ja, ja, die lauten Nachbarn! Das Phänomen ist mir das erste Mal circa 2005 begegnet. Oder besser, es ist mir dort das erste Mal bewusst geworden und seither habe

```

Time State Info
6 Sending data select UserIndex, LastUsedShopLanguage from webshop_user2 wh
6 Waiting for table level lock update webshop_user2 set DataChange=1, LastDataChange=NOW()
6 Waiting for table level lock select UserIndex from webshop_user2 where UserIndex=317249 a
5 Waiting for table level lock update webshop_user2 set DataChange=1, LastDataChange=NOW()
5 Waiting for table level lock select UserIndex from webshop_user2 where UserIndex=1631890
5 Waiting for table level lock select UserIndex from webshop_user2 where UserIndex=695182 a
5 Waiting for table level lock select UserIndex from webshop_user2 where UserIndex=1583223
5 Waiting for table level lock select UserIndex from webshop_user2 where UserIndex=969880 a
5 Waiting for table level lock select Password, UserIndex, SubShop, MainSubShop, State, Adv
3 Waiting for table level lock select Password, UserIndex, SubShop, MainSubShop, State, Adv
3 Waiting for table level lock select UserIndex from webshop_user2 where UserIndex=1661429
3 Waiting for table level lock select UserIndex from webshop_user2 where UserIndex=654008 a
2 Waiting for table level lock select UserIndex from webshop_user2 where UserIndex=795808 a
2 Waiting for table level lock select UserIndex from webshop_user2 where UserIndex=1654024
2 Waiting for table level lock select Password, UserIndex, SubShop, MainSubShop, State, Adv
2 Waiting for table level lock select UNIX_TIMESTAMP>LastDataImport) from webshop_user2 whe
    
```

Listing 2: Stau bei MyISAM-Tabellen

ich es nie mehr vergessen. Dieses Phänomen kann immer dann auftreten, wenn sich mehrere Anwendungen oder Systeme Hardware-Ressourcen teilen müssen. Angefangen beim SAN über das Netzwerk hin zu CPU-Taktzyklen oder RAM. Am häufigsten tritt es bei SAN („geschartes“ I/O) oder Virtualisierung („gescharte“ CPU oder RAM) auf. Wie manifestiert es sich? Durch plötzlich auftretende Performance-Einbrüche, ohne dass „man“ was am System geändert hätte (siehe Abbildung 2 und 3).

Das Problem liegt hier beim „man“ und beim „Sharen“. „Man“ heißt, wir (DBAs) haben nichts geändert, das schließt allerdings nicht aus, dass nicht jemand anderes etwas geändert hat. Und da die Ressourcen „geschart“ sind, kann diese Änderung der anderen die Performance unseres Systems massiv einbrechen lassen. Das Problem liegt oft darin, dass zwi-

schen „uns“ und „denen“ massive virtuelle Mauern bestehen und wir gar nicht so genau wissen, was auf den Systemen so alles passiert.

Diese Probleme sind recht schwer zu erkennen und noch schwieriger ist es herauszufinden, wer am Problem „schuld“ ist.

Und wie komme ich da wieder raus? Als Erstes, und da wird jetzt der eine oder andere IT-Verantwortliche einen roten Kopf bekommen: keine „gescharte“ Ressourcen für kritische Systeme! Und ja, ich weiß auch, dass das teuer ist. Und wenn doch „sharen“, dann: NICHT überprovisionieren (und schon ist der TCO im Eimer) und schauen, dass alle darüber informiert sind, wann was auf dem Gesamtsystem läuft. Falls das auch nicht möglich ist: Messen, messen und nochmal messen, damit der Schuldige anhand von Symptomen isoliert werden kann.

BLOB (und TEXT) in der Datenbank speichern

Es ist verlockend, BLOBs (Binary Large Objects) (siehe Abbildung 4) und die technisch damit verwandten TEXT-Datentypen (sehr lange Strings) in der Datenbank zu speichern. Tun Sie es nicht! Sie machen Ihrem Admin damit früher oder später nur Ärger und verlangsamen das System auf unterschiedliche Weise. Ich habe es früher auch mal cool gefunden, als ein großer Anbieter von Datenbanksystemen damit geprahlt hat, dass man jetzt sogar Videos in der Datenbank speichern könne. Das ist nur Quatsch!

Folgende Probleme können dabei auftreten: Die Query-Latenzen können größer werden und somit der Datenbank-Durchsatz geringer, weil die DB damit beschäftigt ist, BLOBs hin und her zu

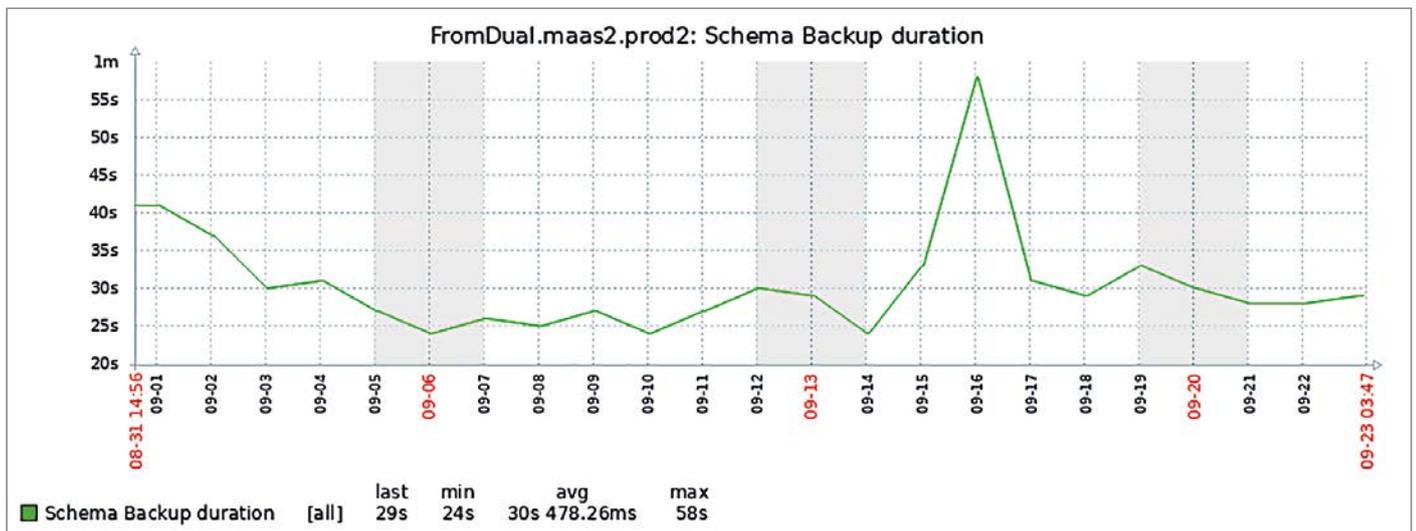


Abbildung 2: Backup dauert plötzlich doppelt so lange wie üblich.

schieben. Die Backup-Größe wird ansteigen und die Backup-Dauer sich verlängern. Von den Restore-Zeiten sprechen wir gar nicht erst. DDL-Operationen werden unendlich lange dauern. Und zu guter Letzt: Der Buffer Pool (Cache, RAM) wird mit diesen Daten geflutet.

Und wie komme ich da wieder raus? BLOBs gar nicht erst in der Datenbank ablegen, sondern in einer Technologie, die für diese Art von Daten geschaffen wurde: zum Beispiel Storage-File!

Volle Platte

Was immer wieder passiert und Datenbanken gar nicht mögen, ist eine volle Platte. Wenn die Platte vollläuft, bleibt die Datenbank mit großer Wahrscheinlichkeit sofort stehen. Zumindest für Schreibvorgänge. Im guten Fall läuft die Datenbank einwandfrei weiter, sobald wieder Platz verfügbar ist. Im schlechten Fall crasht die Datenbank und/oder korrumpiert die Daten.

Und was kann ich dagegen tun? Als Erstes einmal hilft Vorbeugen! Das heißt ein gutes Monitoring zu implementieren, das sieht, dass die Platte vollläuft und rechtzeitig alarmiert. Im Weiteren: Knausern mit Diskplatz ist eine ganz schlechte Idee! Lassen Sie Ihrer Datenbank immer genügend Reserve auf der Platte, um auch mal mit unvorhergesehenen Ereignissen klarzukommen. Wir wollen ja heute alle so agil sein! Das bedeutet jedoch, dass wir auch genügend Reserve brauchen, um die Agilität abzufangen.

Und wie komme ich da wieder raus? Als Erstes schauen, ob irgendetwas gelöscht werden kann, was sicher nicht mehr gebraucht wird, wie zum Beispiel alte Logfiles oder andere Leichen. Dann schauen, ob etwas gelöscht oder wegkopiert werden kann, was nicht jetzt gerade gebraucht wird. Typische Kandidaten sind hier die Binary Logs. Das sind häufig auch genau diejenigen Dateien, die die Platte gefüllt haben. Als Alternative dazu kann auch die Platte vergrößert werden. Das funktioniert mit modernen Technologien wie SAN, LVM, Virtualisierung und modernen Filesystemen (ext4, xfs) eigentlich recht problemlos online.

Locking

Locking ist ein notwendiges Mittel der Datenbank, um Datenkonsistenz zu gewähr-

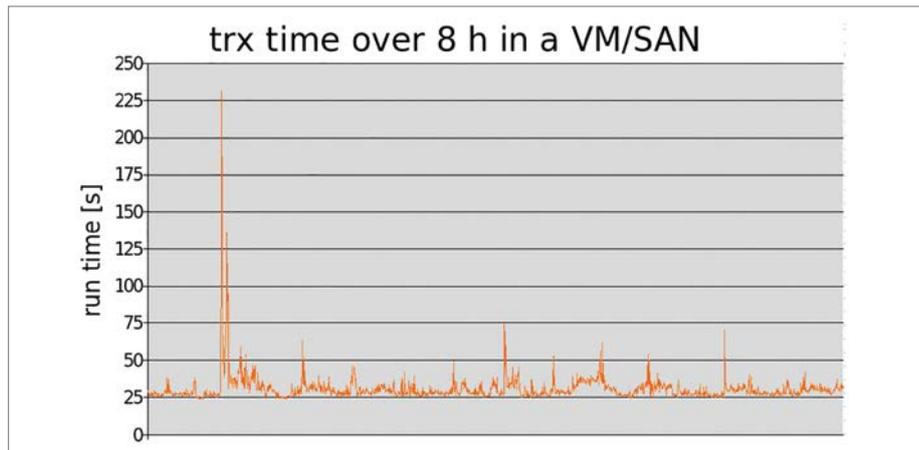


Abbildung 3: Job, der immer genau das gleiche tut und ca. 25 Sekunden dauert, läuft plötzlich bis zu 9-mal länger!

leisten. Ausgelöst wird Locking durch die Anwendung und nicht durch die DB selbst oder gar den DBA. Wenn Sie also Locking-Probleme haben, fragen Sie Ihre Entwickler oder Ihre Anwendung, was sie tut, und hackt nicht auf dem DBA oder der Datenbank herum! Ihr DBA hilft Ihnen aber sicher gerne bei der Lösung des Problems, wenn Sie ihn nett fragen.

Es gibt verschiedene Arten und Ebenen von Locks, die je nach verwendetem DBMS unterschiedlich gestaltet sind: explizite Locks, die durch die Anwendung bewusst gesetzt werden (LOCK TABLE...), und implizite Locks, die durch das RDBMS benötigt werden. Die MyISAM Storage Engine macht Locks auf Tabellen-Ebene für Lese- und Schreibzugriffe. Bei jedem DML- oder SELECT-Befehl wird ein Metadata Lock gesetzt, der Veränderungen der Tabellenstruktur während des DML-Befehls verhindert. Die InnoDB Storage Engine setzt Locks auf einzelne Rows, damit diese nicht durch mehrere Verbindungen gleich-

zeitig geändert werden. Galera Cluster hat auch noch zusätzliche eigene Locking-Mechanismen, um die Konsistenz im ganzen Cluster sicherzustellen.

Und wie gehe ich damit um? Sowohl der Entwickler als auch der Admin sollte lernen, welche Arten von Locks es gibt, um abschätzen zu können, welche Auswirkungen ein bestimmter Befehl auf die Datenbank hat. Explizite Locks und MyISAM-Tabellen sollten vermieden werden. Vorsicht ist auch bei lange laufenden Queries angebracht, weil diese unter Umständen Locks lange halten. Und Entwickler sollten sich angewöhnen, lieber viele kleine und schnelle Transaktionen zu verwenden als wenige große und lange dauernde Transaktionen.

In einem zweiten Teil in der nächsten Red-Stack-Ausgabe 03/2021 werde ich eher MariaDB-/MySQL-spezifische Stolperfallen beleuchten, die das tägliche Leben des MariaDB-/MySQL-Nutzers prägen können.

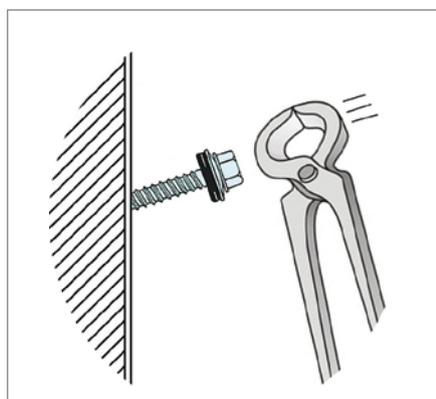


Abbildung 4: BLOB: Das richtige Werkzeug für den richtigen Einsatzzweck

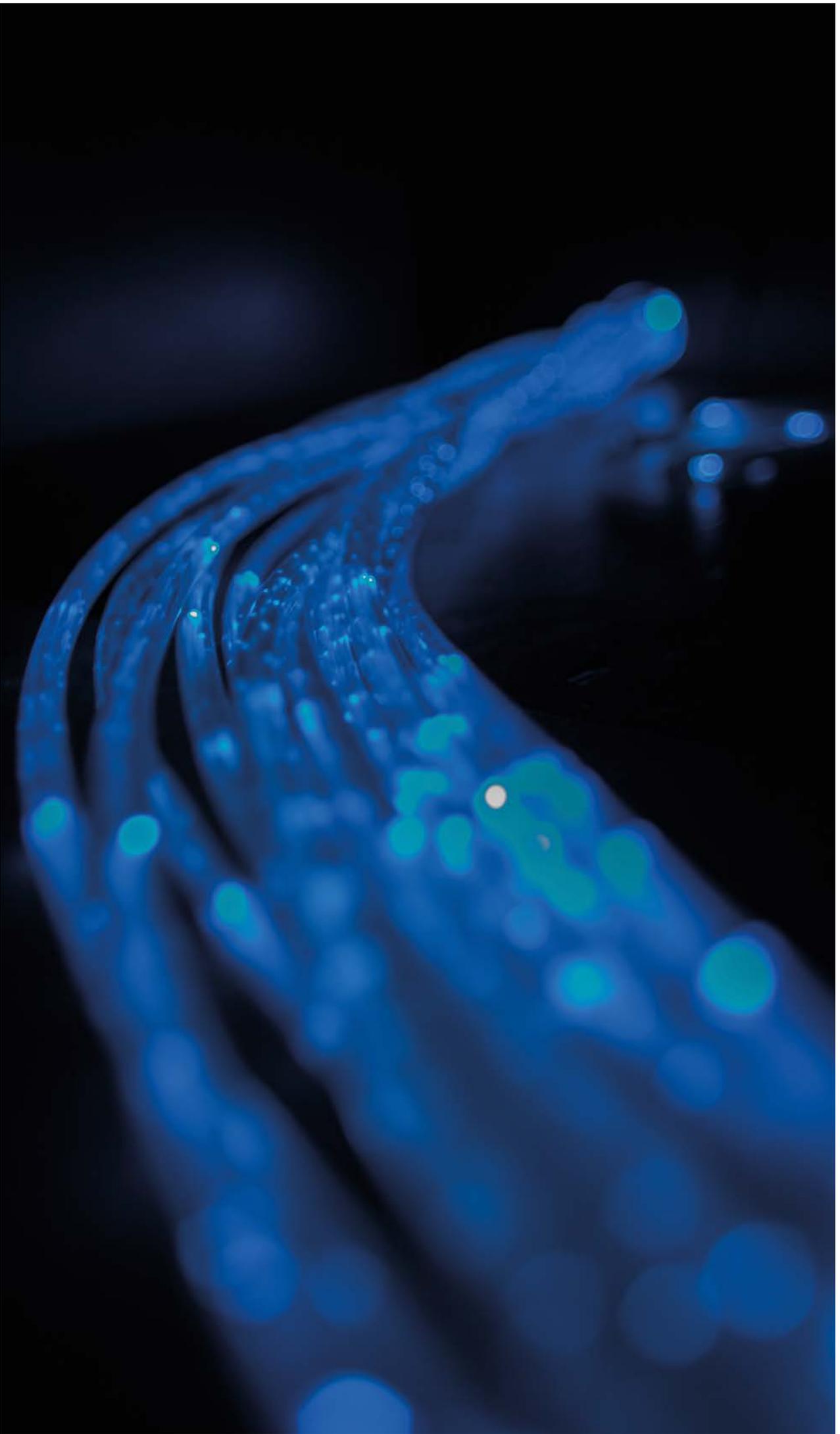


Olli Sennhauser
oli.sennhauser@fromdual.com

BUSINESS

NEWS

02/2021



Die Bedürfnisse des Kunden identifizieren – Ein Data-Science-Projektbericht: Von der Datenanreicherung bis zur gezielten Kundenansprache

Jürgen Günter, Andre Stapper, Karolina Weber, Hermann Wilde, MT AG

Wie gut kenne ich eigentlich die Bedürfnisse meiner Kunden? Habe ich genug Informationen über sie und kann ich daraus Wissen gewinnen und dieses nutzbringend im Unternehmen einsetzen? Diese Fragen sollte man sich regelmäßig stellen, denn sie sind ein entscheidender Erfolgsfaktor. Wenn nicht genügend kundenspezifische Informationen in der eigenen Datenlandschaft vorhanden sind, kann aus Vertriebs- und Marketingsicht nur eine allgemeine Kundenansprache erfolgen. Doch um den Umsatz zu steigern, müssen wir wissen, was der Kunde wirklich braucht – und dafür benötigen wir mehr Informationen über ihn. Wie wir diese und weitere Fragestellungen bei einem unserer Kunden aus dem Bereich Lebensmittelgroßhandel beantwortet haben, möchten wir in diesem Data-Science-Projektbericht mit Ihnen teilen.

Der Kunde

Als Lebensmittelgroßhändler verkauft unser Kunde Artikel für die Profiküche wie Großküchen sowie Lebensmittel an Gastronomie und Hotellerie, also Restaurants, Hotels, Kantinen und Mensen. Er konnte sein Geschäft in den letzten Jahrzehnten in ganz Europa ausweiten und besitzt allein in Deutschland mehrere Stützpunkte, von denen aus er seine Endkunden mit einer eigenen Logistik beliefert. Durch mehrere Eigenmarken und zum Konzern gehörende Metzgereien hebt er sich von der Konkurrenz ab. Zudem bietet er maßgeschneiderte Lösungen an, um seine Kunden im B2B-Bereich zu betreuen. Für unseren Kunden war dies das Startprojekt in die Themengebiete Data Science und Big Data.

Bisherige Situation und Ziel des Projektes

Die bisherigen Vertriebs- und Marketingaktivitäten waren eher unspezifisch und

allgemein gehalten. Es gab mehrere Sortimentskataloge, die alle zu den persönlichen Kundengesprächen mitgenommen werden mussten und an sämtliche Kunden verteilt wurden. Die Endkunden wurden weder über den Webshop noch auf postalischem Wege durch Flyer und Broschüren individuell angesprochen. Zudem verfügte der unternehmensinterne Vertrieb über keine adäquaten Instrumente für die Bestandskundenbetreuung und Neukundenakquise, die ihm alle relevanten Kundeninformationen auf einen Blick bereitstellen konnten. Daraus abgeleitet bestand das Ziel unseres Projektes darin, die Kunden auf Marketing- und Vertriebs-ebene individueller anzusprechen, um damit den Umsatz und den wirtschaftlichen Erfolg des Unternehmens zu steigern. Dafür war es zwingend notwendig, mehr über die eigenen Endkunden und deren Präferenzen zu erfahren und dieses Wissen allen Fachabteilungen zur Verfügung zu stellen. Eine abschließende

Erfolgskontrolle der Maßnahmen sollte die Wirksamkeit der neuen Services sicherstellen.

Herausforderung

Über die Endkunden waren jedoch nur sehr wenige Informationen im Data Warehouse (DWH) des Unternehmens bekannt. Neben elementaren Stammdaten wie Name und Anschrift waren kaum weiterführende Informationen verfügbar. Weder in den kundenbezogenen Tabellen des DWH noch in den anderen operativen Systemen waren Angaben über die Branche der Kunden hinterlegt. So war auf Basis der Daten nicht auswertbar, ob ein bestimmter Gastronomiekunde ein italienisches, griechisches oder spanisches Restaurant, ein Kiosk oder gegebenenfalls eine Raststätte ist. Die einzelnen „Küchenvarianten“ haben natürlich unterschiedliche Anforderungen und Bedürfnisse. Hotels greifen eher zum deutschen Rumpsteak, Raststätten dagegen zu

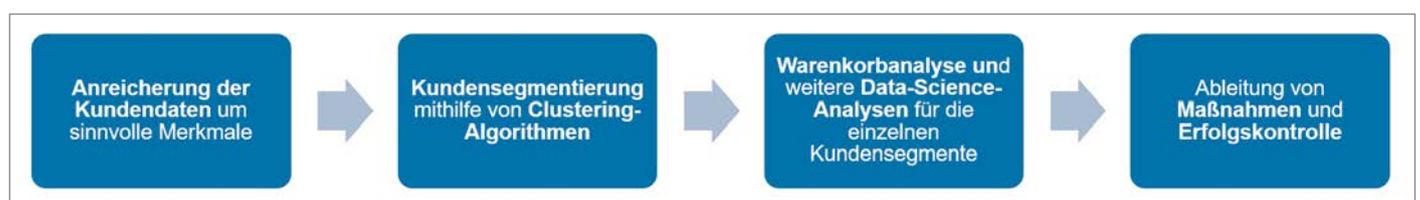


Abbildung 1: Projektschritte (© MT AG)



Abbildung 2: Schaubild der Datenanreicherung (© MT AG, Logos von Facebook, Google, Nokia Here)

Pommes und Bockwurst. Solche Angaben sind für die gezielte Ansprache essenziell und mussten in unserem Fall zuerst beschafft werden.

Lösungsansatz und Vorgehen

Der erste Schritt im Projekt war deshalb eine **Datenanreicherung** der Kundeninformationen mithilfe von externen Diensten (siehe Abbildung 1).

Anschließend teilten wir unter Einsatz von Clustering-Algorithmen aus dem Machine-Learning-Bereich (Schritt 2) die Gastronomiekunden anhand ihrer Stammdaten, weiterer Kundenmerkmale und ihres Kaufverhaltens in unterschiedliche Gruppen ein. Ein vereinfachendes Beispiel einer Gruppe sah wie folgt aus: spanische Restaurants mit einem hohen Umsatz bei Meeresprodukten und Gemüse, die mittelpreisige Gerichte anbieten und online häufig und sehr gut bewertet waren. In der abschließenden Cluster-Bildung benutzten wir weitere Merkmale, die im späteren Abschnitt **Kundensegmentierung** (ebenfalls Schritt 2) vorgestellt werden. Darauf aufbauend führ-

ten wir für jede dieser Gruppen eine **Warenkorbanalyse** (Schritt 3) durch, bei der die häufig gemeinsam gekauften Produkte identifiziert wurden. Damit konnten wir zum Beispiel den spanischen Restaurants mediterrane Warenkörbe anbieten, bestehend aus Produktvorschlägen und -kombinationen, die für sie von Interesse waren. Aus diesen Erkenntnissen und zusätzlichen Data-Science-Analysen, die wir im weiteren Verlauf dieses Berichtes noch vorstellen, entwickelten wir verbesserte und neue **Produkte** (Schritt 4). Im Folgenden stellen wir die Projektschritte im Detail vor.

Möglichkeiten der Datenanreicherung

An zusätzliche Daten kann man über unterschiedliche Ansätze kommen:

- Informationen über Analyse der Stammdaten ableiten: Zum Beispiel wird der Kunde „Pizzeria Italia“ voraussichtlich ein italienisches Restaurant sein.
- Daten bei einem Adresshändler unter Berücksichtigung der DSGVO dazukaufen.

Es ist von Fall zu Fall zu bewerten, welcher Anbieter infrage kommt. In unserem Fall spielten dafür Daten im Bereich Gastronomie und Hotellerie eine Rolle.

- Einsatz von Web-Crawling für weitere Kundenmerkmale.
- Nutzung externer Dienste mittels Web APIs.

Um eine optimierte Datengrundlage für eine Kundensegmentierung zu schaffen, entschieden wir uns nach der Prüfung der möglichen Varianten gemeinsam mit unserem Kunden für die Nutzung von Web APIs. Dabei waren die Aspekte des Aufwandes und der Datenverfügbarkeit sowie juristische Gründe entscheidend. Durch die Nutzung von Web APIs von später noch vorgestellten Anbietern konnten wir gezielt weitere Kundenmerkmale abfragen.

Auswahl der externen Dienste

Die wichtigste Frage bei der Auswahl der externen Dienste ist sicherlich, welche Informationen für den jeweiligen Anwendungsfall benötigt werden. Für Kunden aus dem Gastronomiebereich identifizierten wir die Branche, den Küchenstil, die durchschnittliche Bewertungshöhe und die Anzahl der Online-Bewertungen sowie die Preiskategorie als wichtige Merkmale. Bei der Auswahl evaluierten wir zehn Anbieter hinsichtlich der dazu verfügbaren Datenbasis, der Kosten und der rechtlichen Nutzung. Gemeinsam entschieden wir uns

```

if (longitude != '-' and latitude != '-'):
    searchtext = urllib.parse.quote(data[''+column][i])
    url = 'https://places.cit.api.here.com/places/v1/discover/search?in=' \
        +latitude+', '+longitude+', r=500&q='+searchtext+ \
        'app_id='+here_app_id+'&app_code='+here_app_code
    response = requests.get(url, verify=False)
    
```

Listing 1: API-Call des Kartendienstes NOKIA HERE (© Copyright MT AG).

Name	Status	Kategorie	Küchenstil	Preiskategorie
Restaurant Lecker GmbH	Gefunden	Restaurant	Italienisch	Mittel

Tabelle 1: Ergebnis der API-Calls im Idealfall (© MT AG)

für zwei kostenlose Anbieter – den Kartendienst *Nokia Here* [1] und das *Facebook Graph API* [2] – sowie das kostenpflichtige *Google Places API* [3].

Auf den Webseiten von Google, Facebook und Nokia Here finden sich diverse Informationen zu sogenannten Point-of-Interests (siehe *Abbildung 2*). Über das fiktive Restaurant *Lecker* lässt sich beispielsweise ermitteln: Es ist ein mittelpreisiges, italienisches Restaurant, das über 102 Bewertungen verfügt, die im Schnitt bei einem Wert von 4,4 von 5 Sternen liegen. Über eigens entwickelte Prozesse in Python und eine anschließende Anbindung in das ETL-Tool (*Extract, Transform, Load*) des Kunden konnten diese Daten unter Berücksichtigung der DSGVO und der Nutzungsbedingungen der externen Dienste angezogen und abgespeichert werden.

Ablauf des API-Calls

Die grundsätzliche Ansprache von Web APIs ist nicht sehr komplex, da dies ein standardisiertes Vorgehen (zum Beispiel REST API) ist. Das API wird mit bestimmten Suchparametern angesprochen und man erhält eine Antwort zurück. Um Daten zu einem Unternehmen zu erhalten, führt man eine Suchanfrage mit einer Kombination des Kundennamens und der -adresse sowie der Accountdaten des jeweiligen Dienstes aus. Ein solcher Aufruf ist in *Listing 1* abgebildet.

Im Idealfall werden die Daten als Ergebnis der API-Anfrage wie in *Tabelle 1* zurückgeliefert.

Plan vs. Wirklichkeit

So die Theorie. Aber wie im echten Leben wich die Realität auch in diesem Projekt davon ab. Wir wurden mit unterschiedlichen Hürden konfrontiert, die die Datenanreicherung komplizierter gestalteten, als es zunächst den Anschein hatte (siehe *Abbildung 3*).

Der erste Aufwandstreiber war die mangelnde Qualität der Stammdaten aus dem



Abbildung 3: Plan vs. Wirklichkeit (© MT AG)

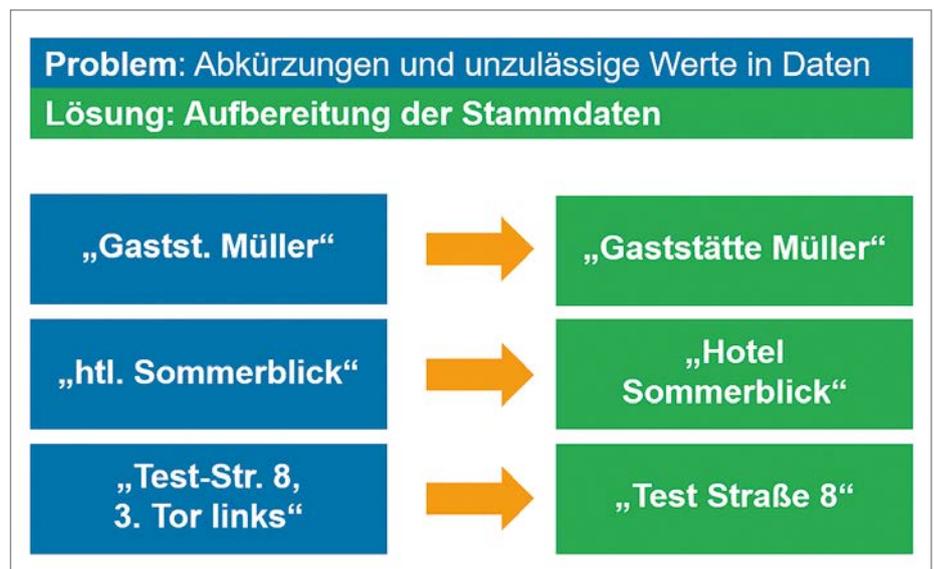


Abbildung 4: Aufbereitung der Stammdaten für die API-Calls (© MT AG)

DWH, weshalb wir diese vor dem Aufruf des API-Calls zuerst aufbereiten mussten (siehe *Abbildung 4*). Zum einen wurden im DWH Abkürzungen im Namen und in der Adresse der Kunden verwendet. Zum anderen kamen auch unzulässige Werte in den Daten vor. Hierbei handelte es sich häufig um Zusatzangaben für die Lieferung, zum Beispiel „3. Tor links“.

Des Weiteren kam es vor, dass mehrere Namensbezeichnungen und Adressen für

denselben Kunden im DWH gespeichert waren. Das war darauf zurückzuführen, dass das System historisch gewachsen war und es bisher kein standardisiertes Verfahren beim Anlegen von Kunden gegeben hatte. Wie in *Tabelle 2* dargestellt, waren die Angaben zum Namen der Lokalität, zum Inhaber und gegebenenfalls zur Kette in unterschiedlichen Datenfeldern verteilt. Mit Hilfe einer Abfrage-Logik konnten diese Daten in eine wahrscheinliche Reihenfolge gebracht und systematisch für den API-Call benutzt werden.

Als dritte Hürde stellte sich die fachliche Validierung der API-Antworten heraus. Neben Mehrfachangaben für Kunden im System bekam man für viele Suchanfragen mehrere Einträge als Antwort zurück. Dies war vor allem bei Lokalitäten in der Nähe von Sehenswürdigkeiten und Ausflugszielen zu beachten. Aber wie konnten wir sicherstellen, dass wir wirklich unse-

Name1	Name2	Name3
Name des Inhabers	Name des Restaurants	-
Name des Restaurants	Name des Inhabers	-
Name der Kette	-	Name des Restaurants
-	Name des Inhabers	Name der Kette

Tabelle 2: Keine eindeutigen Kundennamen im System (© MT AG)

ren gesuchten Kunden über den API-Call gefunden hatten? Eine solche Prüfung der wahrscheinlichsten Antwort erfolgte mithilfe der sogenannten Levenshtein-Metrik (siehe Tabelle 3). Hohe Werte bedeuten dabei eine große Ähnlichkeit von zwei Zeichenketten – in unserem Fall dem Suchbegriff und der Antwort aus dem API-Call.

Ergebnis der Datenanreicherung

Lohnt sich der Aufwand für die Aufbereitungsschritte? Ganz eindeutig ja! Denn die Trefferquote für die API-Calls konnte durch die Optimierung signifikant verbessert werden, wie in Abbildung 5 dargestellt. Wurden zu Beginn mit der einfachen Variante der API-Calls lediglich knapp 58% der Gastronomiekunden gefunden, so konnte in den weiteren Iterationen durch die Optimierungsschritte und die Konsolidierung mehrerer externer Dienste die Trefferquote auf über 92% gesteigert werden. Bei den rest-

Suchbegriff	API-Antwort	Levenshtein-Metrik
Restaurant zur Seeschänke	Restaurant Seeschänke	84 %
	See Auerbach	19 %

Tabelle 3: Validierung API-Antworten (© MT AG)

lichen 8% handelte es sich um fehlerhafte Stammdaten, die durch einen manuellen Prozess abgeglichen werden mussten.

Das hieß: Nach der Datenanreicherung wussten wir bei 92% der Gastronomiekunden, welcher Branche und welchem Küchenstil sie angehören. Wir konnten die Kunden zum Beispiel in „positiv bewertete italienische Restaurants“ und „internationale Hotels“ unterteilen. Wir wussten auch, wie viele Bewertungen der Kunde bekommen hatte, und dementsprechend, wie bekannt er war. Zudem konnten wir nun das Preisniveau des Kunden bestimmen. All diese Informationen waren nach der Da-

tenanreicherung in der eigenen Datenlandschaft vorhanden und standen den jeweiligen Fachabteilungen zur Verfügung. Zuvor waren diese Informationen unbekannt! Des Weiteren waren die Kundendaten nun mit ausreichend Merkmalen angereichert, um eine Kundensegmentierung und weitere Analysen durchführen zu können.

Kundensegmentierung

Ziel einer Kundensegmentierung ist es, relevante Gruppen – für den entsprechenden Fachbereich und Verwendungszweck – innerhalb seiner Kunden zu identifizieren, um anschließend Rückschlüsse auf



Abbildung 5: Verbesserung der Trefferquote (© MT AG)



Abbildung 6: Prozess der Kundensegmentierung (© MT AG)

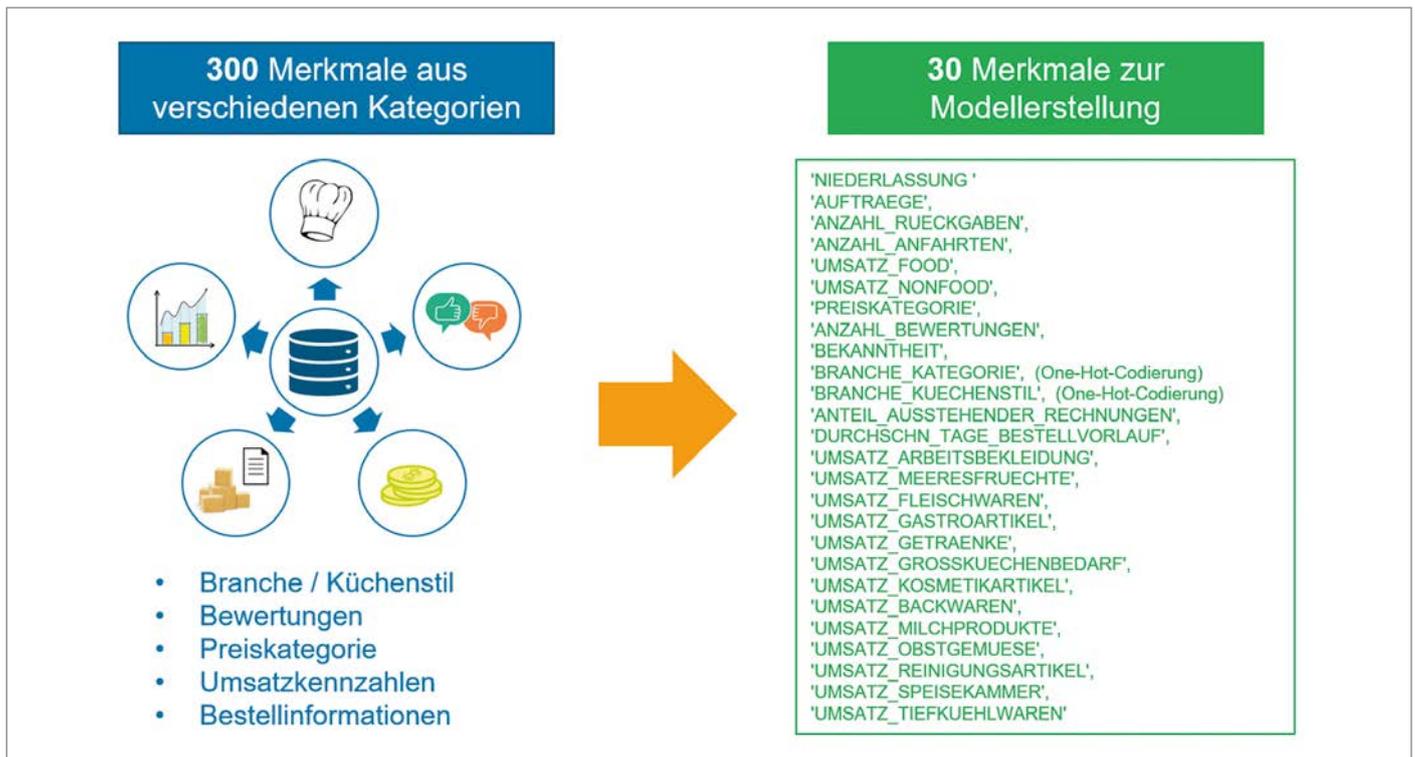


Abbildung 7: Auswahl der relevanten Merkmale (© MT AG)

Business Actions ziehen zu können. Dafür werden die Kunden anhand bestimmter Eigenschaften und Merkmale aufgeteilt. Kunden einer Gruppe sind sich in diesen Eigenschaften besonders ähnlich. Der Prozess der Kundensegmentierung besteht aus drei Teilschritten und ist in *Abbildung 6* dargestellt.

Auswahl der relevanten Informationen

Zuerst wählten wir Merkmale über die Kunden aus, die bei der Segmentierung berücksichtigt werden sollten. Hierbei war besonders auf den Verwendungszweck der Kundengruppen zu achten. Für das Ziel, die Vertriebs- und Marketingaktionen beim Lebensmittelgroßhändler zu optimieren, waren verschiedene Informationen notwendig. Zum einen sammelten wir Umsatzkennzahlen und Bestellinformationen aus dem internen Datenbestand. Zum anderen verwendeten wir die zuvor angereicherten Informationen über Küchenstil, Online-Bewertungen und Preiskategorie. Aus den 300 abgeleiteten Merkmalen der genannten Kategorien identifizierten wir 30 Merkmale zur Modellerstellung (*siehe Abbildung 7*). Denn eine größere Anzahl von Merkmalen verbessert nicht unbedingt das Ergebnis der Kundensegmentierung. Qualitativ hochwertige und aussagekräftige Merkmale sind wichtiger als

die Quantität der Informationen! Merkmale, die fachlich nicht zum Ziel der Kundensegmentierung beitragen, „verwirren“ den Algorithmus eher und verfälschen das Ergebnis [4].

Erstellen von Segmenten mit Machine-Learning-Algorithmen

Im zweiten Schritt führten wir mithilfe von maschinellem Lernen die eigentliche Segmentierung der Kunden durch. Dafür

verwendeten wir sogenannte Clustering-Algorithmen. Diese können die Kunden anhand der zuvor ausgewählten Merkmale trennen und in Cluster beziehungsweise Gruppen aufteilen (*siehe Abbildung 8*). Sie berechnen die Ähnlichkeit der Kunden mithilfe des Abstands der einzelnen Datenpunkte. Sowohl in diesem als auch in anderen Projekten haben wir sehr gute Erfahrungen mit den Algorithmen *K-Means* [5] und *DBSCAN* [6] gemacht.

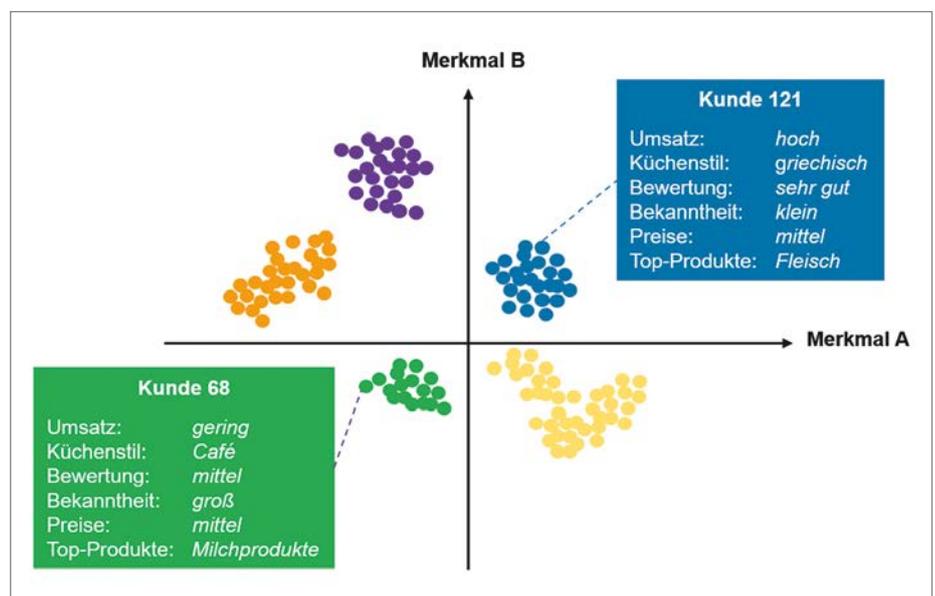


Abbildung 8: Bildung von Kundensegmenten durch Clustering-Algorithmen (© MT AG)

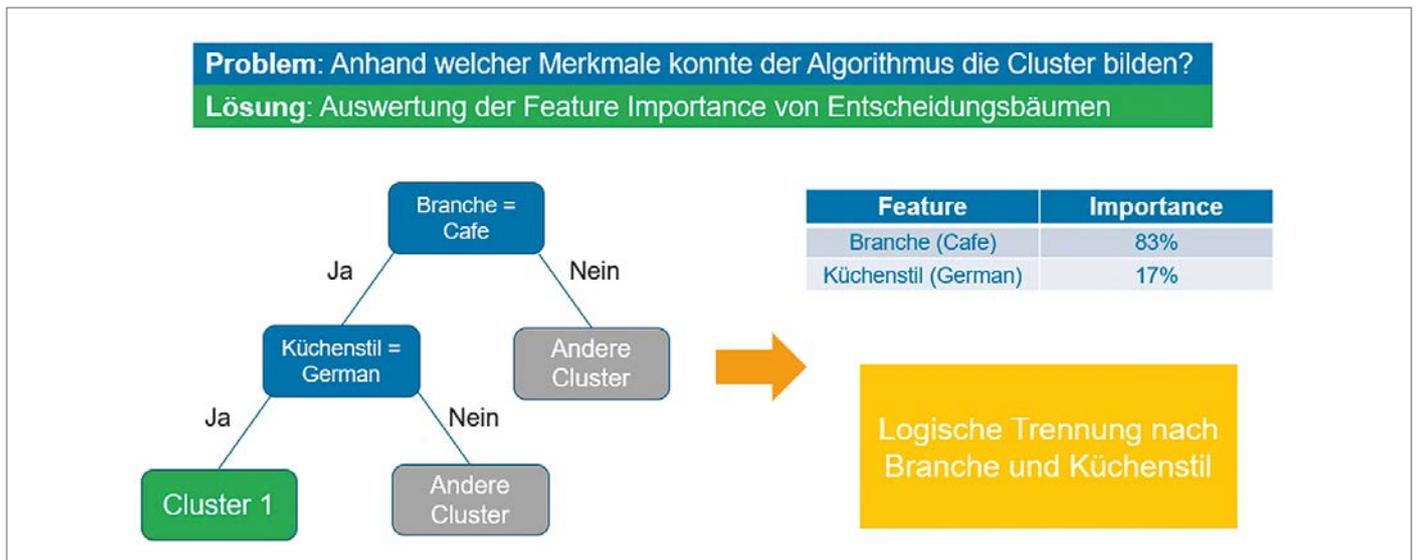


Abbildung 9: Fachliche Auswertung der Kundensegmente mit Entscheidungsbäumen (© MT AG)

Der Weg zu optimalen Kundensegmenten erforderte mehrere Iterationen und Optimierungsschritte in der Modellerstellung. Beispielsweise musste die optimale Anzahl von Clustern gefunden werden. Dafür wurden Verfahren wie die Elbow-Methode [7] eingesetzt. Des Weiteren sollte auch die technische Güte der Modelle anhand von Metriken ermittelt werden. Eine dieser Metriken war der sogenannte Silhouetten-Koeffizient [8]. Er eignete sich zur Bewertung, wie klar die einzelnen Cluster voneinander getrennt wurden.

Fachliche Interpretation der Kundensegmente

Nach der technischen Erstellung der einzelnen Cluster mussten diese auch fachlich interpretiert werden. Wie konnten sich die Cluster fachlich benennen und beschreiben lassen? Und welche Eigenschaften hatten die Kunden innerhalb der Cluster? Um diese Fragen zu beantworten, wurden bestimmte Machine-Learning-Verfahren eingesetzt. Mithilfe von Entscheidungsbäumen und der *Feature Importance* [9] konnten die

wichtigsten Merkmale bestimmt werden, die für die Einteilung in die Cluster entscheidend waren. Ein Beispiel dazu wird in *Abbildung 9* gezeigt.

Die Auswertung bestätigte uns dabei, dass eine pragmatische, logische Trennung der Kunden nach Branche und Küchenstil zu sinnvollen Kundensegmenten führte und gleichzeitig die fachliche Interpretierbarkeit jener Segmente gewährleistete. Das Ergebnis waren 32 Kundensegmente, wie zum Beispiel „mittelpreisige griechische

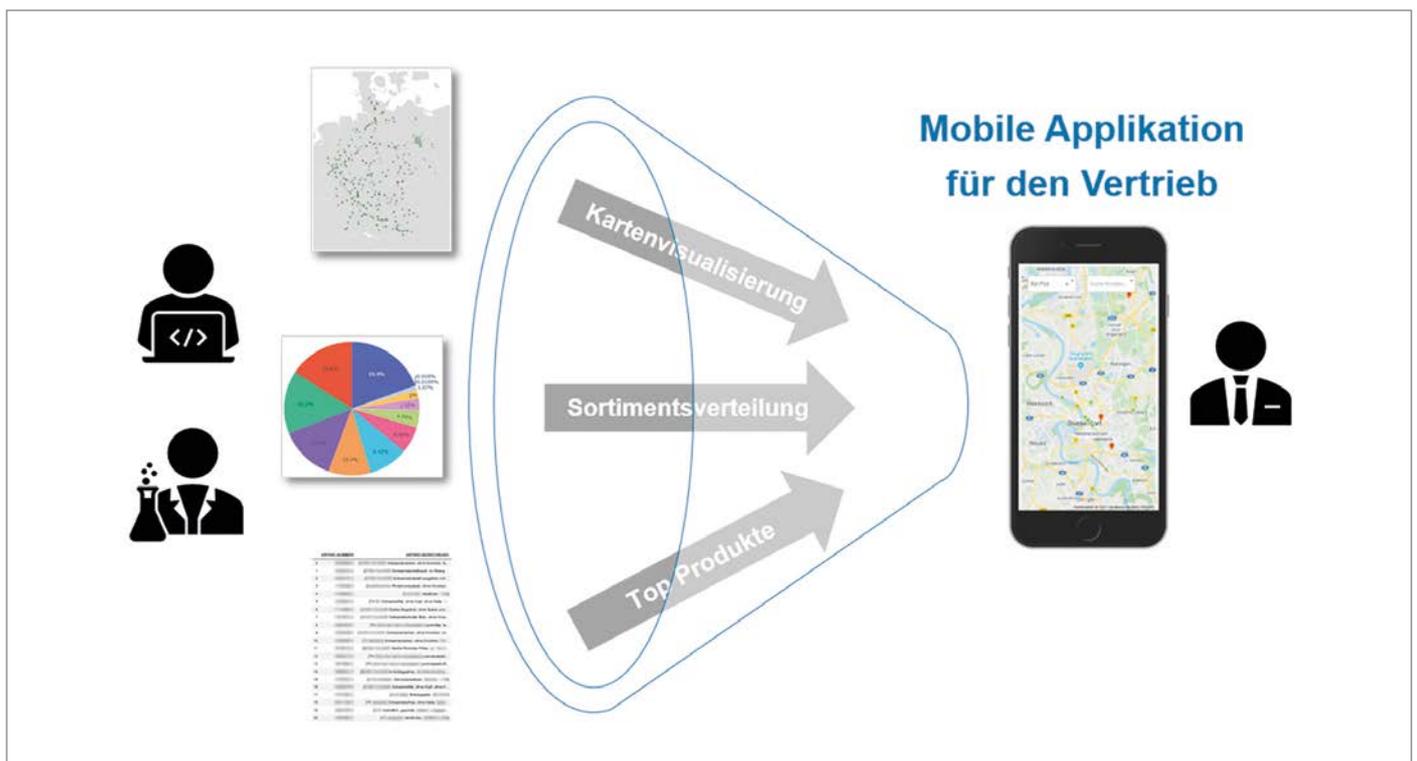


Abbildung 10: Mobile Applikation für den Vertrieb (© MT AG, Kartendaten © 2021: Google, GeoBasis-DE/BKG)



Abbildung 11: Abgeleitete Maßnahmen, betrieblicher Nutzen (© MT AG unter Verwendung von Anelka, Pixabay [11], Kartendaten © 2021: Google, GeoBasis-DE/BKG)

Restaurants mit sehr guten Online-Bewertungen und einem hohen Umsatz bei Fleischprodukten“.

Warenkorbanalyse und Auswertungen der Kundensegmente

Wie konnten wir nun aus den Kundensegmenten nützliches Wissen ziehen?

Die neu erzeugten Kundensegmente bildeten eine optimale Basis für weiterführende Analysen. Die Warenkorbanalyse war ein solches Verfahren. Dabei wurden mithilfe von Algorithmen wie beispielsweise *Apriori* oder dem *FP-Growth* [10] sogenannte Assoziationsregeln erstellt, um das Kaufverhalten der Kunden besser zu verstehen. Dieses Wissen ermöglichte es uns, den Warenbestand und die Produktpräsentation zu optimieren. Die Assoziationsregeln lieferten interessante Erkenntnisse, wie zum Beispiel: „Wenn ein Kunde Pommes kauft, kauft er zu 80% Tomatensuppe dazu.“ Solche Regeln können dann auf Webseiten und in Online-Shops genutzt werden. Ruft nun ein Kunde einen Pommes-Artikel auf, so kann als Produktvorschlag der Artikel „Tomatensuppe“ angezeigt werden. Dies wird als *Next-Best-Offer* bezeichnet. Auch automatisierte Preisanpassungen oder die Erstellung von Produktbündeln sind aufbauend auf solchen Regeln möglich.

Zudem konnten die einzelnen Kundensegmente durch weitere Analysen ausge-

wertet werden. Dabei konnten wir einige Auffälligkeiten entdecken, die offensichtlich oder auch überraschend waren: Zum einen konnten wir durch den Vergleich der Kundensegmente die griechischen Restaurants als besonders umsatzstarke Kundengruppe identifizieren. Zum anderen analysierten wir die gekauften Produktgruppen innerhalb eines Kundensegments. Eine der vielen spannenden Erkenntnisse war unter anderem, dass griechische Restaurants mit guten Online-Bewertungen besonders viel Ouzo kaufen. Des Weiteren waren manche Kunden erkennbar, die sich im Vergleich zum Rest eines Kundensegments auffällig verhielten. So gab es viele Bars und Pubs, die gar keine Getränke bestellten. Das ist eventuell darauf zurückzuführen, dass sie einen separaten Getränkehändler hatten oder noch nicht vom Getränkeangebot des Großhändlers wussten. Diese zusätzlichen Informationen konnten Vertriebler bei ihrer täglichen Arbeit und Vorbereitung auf Kundengespräche unterstützen.

Aus Erkenntnissen neue Produkte und Services ableiten

Wie ließen sich diese zusätzlichen Informationen nutzen, um unternehmerischen Mehrwert zu erzielen? Dazu wurden mehrere Services beziehungsweise Produkte optimiert oder neu entwickelt. Als ein gänzlich neues Produkt entstand eine mobile Appli-

kation zur Unterstützung der Vertriebsmitarbeiter (siehe Abbildung 10). Sie bündelt die verschiedensten Informationen zu einem Kunden beziehungsweise einem Kundensegment. Auf einer Kartendarstellung wird visualisiert, wo die jeweiligen Kundensegmente vertreten sind. Zusammen mit den meistgekauften Produkten und umsatzstarken Produktsortimenten für das jeweilige Kundensegment können das Kaufverhalten des jeweiligen Kunden analysiert und entsprechende Produktvorschläge angeboten werden. Wenn ein griechisches Restaurant bisher nur Bedarfsartikel für eine Großküche, aber noch kaum Fleischprodukte gekauft hat, können ihm die Top-Produktvorschläge aus dem griechischen Kundensegment angeboten werden.

Die Erkenntnisse und Empfehlungen zu dem Kunden werden dem Vertriebler in Form von übersichtlichen Grafiken und Auswertungen in der App zur Verfügung gestellt, damit er sich nicht mehr mit mühseliger Beschaffung von Informationen herumschlagen muss. Ein weiterer Vorteil besteht darin, dass der Vertriebler diese Informationen jederzeit auf seinem mobilen Gerät abrufen und so in Kunden- und Verkaufsgesprächen nutzen kann. Umständliches Zusammensuchen von Informationen und hektisches Wälzen von Produktkatalogen in den Gesprächen sind nicht mehr notwendig.

Eine weitere Business Action war die Optimierung des Webshops. Anhand der neuen Informationen aus den Analysen wird das Design des Webshops je nach Segmentzugehörigkeit des Kunden dynamisch angepasst. Das heißt, Kunden mit einem italienischen Restaurant erhalten ein Design mit italienischem Flair und zudem Produktangebote, die speziell auf die Bedürfnisse der italienischen Küche ausgerichtet sind.

Darüber hinaus wurden die allgemeinen Produktkataloge überarbeitet. Anhand der Analysen entstanden neue, auf die einzelnen Kundensegmente ausgerichtete Werbekataloge. Sie enthalten ausgewählte Produktkombinationen und Angebote, die die Bedürfnisse der verschiedenen Küchenrichtungen und die Präferenzen der Kunden gezielt ansprechen. Dadurch müssen die Vertriebsmitarbeiter nicht länger mehr als zwei Dutzend unterschiedliche Kataloge zum Kundengespräch mitnehmen und können sich mithilfe der spezialisierten neuen Kataloge und der App auf eine bessere Beratung des Kunden fokussieren.

Fazit

Mithilfe der Analysen und der daraus entstandenen Produkte und Services wurde in diesem Projekt erheblicher betrieblicher Nutzen für unseren Kunden erzielt (siehe *Abbildung 11*). Zum einen lieferten die angereicherten Kundeninformationen in Kombination mit den darauf aufbauenden Auswertungen neues und zentrales Wissen für die verschiedenen Fachabteilungen. Zum anderen ermöglichten die Vertriebs-App, der Webshop und die angepassten Produktkataloge eine verbesserte Kundenansprache bei Vertriebs- und Marketingaktivitäten. Diese neuen und verbesserten Services kamen sowohl bei den Vertriebsmitarbeitern als auch bei den Kunden positiv an. Zudem ist es ab jetzt möglich, all diese Informationen zu nutzen, sich bei der Neukundensuche auf lukrative Kundensegmente zu fokussieren und diese gezielt mit passgenauen Produkten anzusprechen.

Für weitere Projektinformationen zum Thema Kundensegmentierung finden Sie den zugehörigen Video-Vortrag auf der Homepage der MT AG:

<https://www.mt-ag.com/portfolio/downloadbereich/mt-videos/>

Quellen

- [1] <https://developer.here.com/>
- [2] <https://developers.facebook.com/docs/graph-api/>
- [3] <https://developers.google.com/places/web-service/overview>
- [4] Géron, Aurélion (2019), Hands-On Machine Learning with Scikit-Learn, Keras & Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems (S. 26 f), O'Reilly Media, Inc., Sebastopol, Canada
- [5] <https://scikit-learn.org/stable/modules/clustering.html#k-means> (Stand: 05.01.2021)
- [6] <https://scikit-learn.org/stable/modules/clustering.html#dbscan> (Stand: 05.01.2021)
- [7] <https://www.scikit-yb.org/en/latest/api/cluster/elbow.html> (Stand: 05.01.2021)
- [8] https://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html (Stand: 05.01.2021)
- [9] https://scikit-learn.org/stable/auto_examples/ensemble/plot_forest_importances.html (Stand: 07.01.2021)
- [10] Sebastian Raschka, http://rasbt.github.io/mlxtend/user_guide/frequent_patterns/fpgrowth/#frequent-item-sets-via-the-fp-growth-algorithm (Stand 07.01.2021)
- [11] <https://pixabay.com/de/photos/oliven%C3%B6l-tomaten-kr%C3%A4uter-gew%C3%BCrze-346997/> (Stand: 11.01.2021)



Jürgen Günter
juergen.guenter@mt-ag.com

Die Autoren sind ein eingespieltes Projektteam aus dem Analytics-Bereich der MT AG. Gemeinsam konnten sie bereits mehrere interdisziplinäre Data-Science- und Big-Data-Projekte in unterschiedlichen Branchen umsetzen. Dabei bündeln sie die Expertise über verschiedene Machine-Learning-Frameworks beziehungsweise Plattformen wie Jupyter mit Python, Cloudera ML, Azure ML, Databricks, Amazon SageMaker, RapidMiner und KNIME. Zudem deckt das Team den gesamten Analyseprozess ab – von der Datenintegration über die Modellerstellung bis hin zur Bereitstellung und Visualisierung. Die agile Vorgehensweise des Teams nach Scrum unterstützt den Erfolg in Data-Science-Projekten.



Hermann Wilde
Hermann.Wilde@mt-ag.com



Andre Stapper
Andre.Stapper@mt-ag.com



Karolina Weber
Karolina.Weber@mt-ag.com

Optimierung und Automatisierung maschinellen Lernens



Malena Reiners, NRW.BANK

Der Einsatz von KI-Technologien in deutschen Unternehmen nimmt laut einer neuen Studie des IDG (International Data Group) [1] weiter stark zu. Viele Entscheidungsträger sehen im maschinellen Lernen (ML) eine Technologie der Zukunft – trotzdem wird ihr Einsatz in sieben von zehn Führungsetagen weder diskutiert, geschweige denn geplant. Als Hauptgrund für die fehlenden Investitionen nannten 70 Prozent der befragten Geschäftsführenden den Faktor Zeitmangel. Gefordert seien vielmehr Impulse aus der KI-Forschung: eine klare Nachvollziehbarkeit sowie reduzierte Daten- und Zeitintensität. Sogenannte Ausdünnungs-Methoden (Pruning) und daraus resultierende Verfahren zum automatisierten maschinellen Lernen (AutoML) können mögliche Antworten auf diese Forderungen sein. Was darunter zu verstehen ist und wieso diese von Vorteil sind, wird im Folgenden dargestellt.

Herausforderungen des maschinellen Lernens

Der Vorteil von selbstlernenden Algorithmen beziehungsweise von maschinellem Lernen gegenüber traditioneller Programmierung ist, dass keine Regeln und Bedingungen vorgegeben werden müssen. Stattdessen lernt der Algorithmus, selbstständig mit den Eingabedaten umzugehen. Bis jedoch ein künstliches neuronales Netz (KNN) mit einem solchen eigenständigen Training beginnen kann, ist bereits im Vorfeld eine große Anzahl von Entscheidungen für das Design der Netzwerkarchitektur zu treffen. Selbst wenn der Typ der Netzarchitektur, zum Beispiel ein Faltungsnetz zur Bilderkennung, bereits feststeht, gibt es zahlreiche Variationen der verschiedenen Bausteine des ML-Verfahrens. Die sogenannten Hyperparameter einer Netzwerkarchitektur definieren nicht nur deren Tiefe (Anzahl der Schichten), Breite (Anzahl der Neuronen) und Komplexität (Anzahl der Verbindungen), sondern auch die verschiedenen verwendeten Methoden wie Dropout oder Pooling. Diese können je nach Designart des Programmierenden die Performance des KNN stark beeinflussen. Die Hyperparameter werden auch als die nicht-lernbaren Parameter des Netzwerkes bezeichnet, da sie vor Beginn des Trainings festgelegt werden müssen. Ein grundlegendes Verständnis dafür, wie künstliche neuronale Netze lernen, ist Voraussetzung, um die Hyperparameter optimal anpassen zu können. Und selbst Experten kommen um ein oft mühseliges, da-

ten- und zeitintensives „trial and error“ bei der Optimierung nicht herum.

Komplexität von ML-Architekturen

Das Vorbild künstlicher neuronaler Netze ist das biologische Gehirn, bestehend aus bis zu 100 Milliarden Nervenzellen (Neuronen), die wiederum mit über einer Trillion Synapsen (Verbindungen) miteinander verbunden sind [2]. Daher ist es nicht verwunderlich, dass ML-Architekturen aktuell häufig möglichst komplex gestaltet sind, also viele Neuronen und Verbindungen besitzen. In der aktuellen Forschung gibt es Experimente und Anwendungen von KI-Modellen mit mehreren Hundert Schichten. Dabei führt jede Extra-Schicht jedoch zu einer längeren Rechenzeit und einer geringeren Nachvollziehbarkeit der internen Prozesse. Durch im-

mer mehr nutzbare Ressourcen galt – vor allem im Bilderkennungsbereich – lange Zeit: je komplexer die Architektur, desto genauer die Vorhersagen des KNN. Mit einer steigenden Komplexität, hier als Anzahl der zu lernenden Verbindungen zwischen den Neuronen definiert, bedarf es allerdings auch einer steigenden Anzahl von Daten und Berechnungen. Dadurch nimmt auch die Dauer des Trainings und der Bedarf nach immer größer werdenden Rechenressourcen zu. Bei steigenden Kosten für die enormen Rechenressourcen und der „Je mehr desto besser“-Mentalität stellt sich jedoch auch schnell die Frage der technischen Limitierungen.

Ausdünnen der ML-Architekturen

Wann endet der Trend der immer größer werdenden KI-Modelle? Gilt wirklich: je kom-

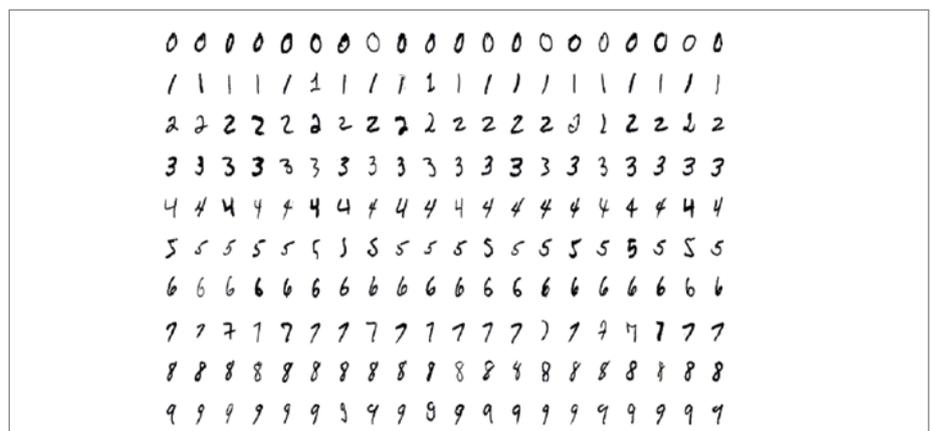


Abbildung 1: Ein Auszug aus dem MNIST-Datensatz (Quelle: Fan / Zhou / Jiheng 2010)

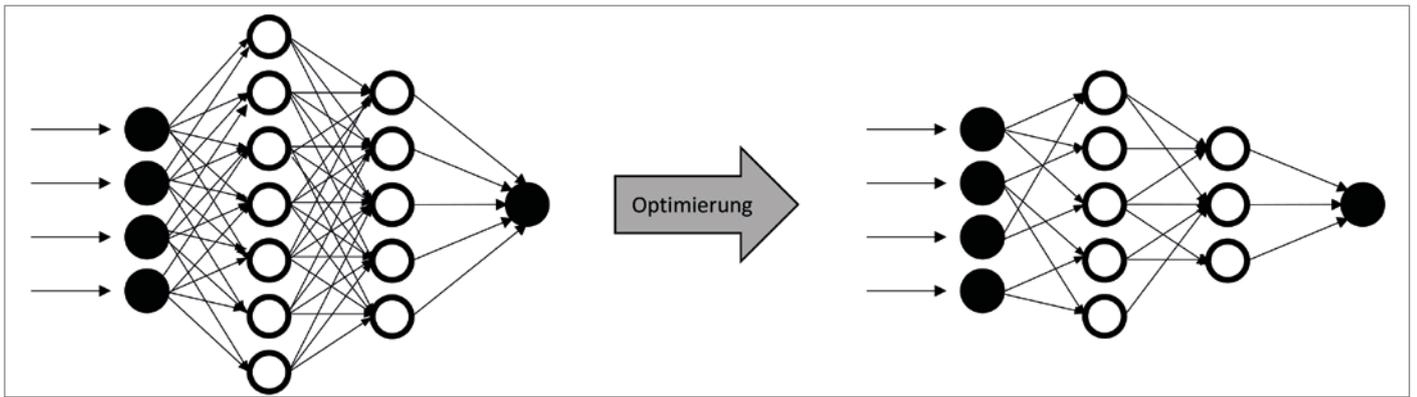


Abbildung 2: Optimierung der Netzwerkarchitektur. Durch eine Reduktion der Verbindungen zwischen den Neuronen wird das Netzwerk ausgedünnt. Die Performance wird dadurch nicht wesentlich eingeschränkt (Quelle: Eigene Darstellung)

plexer, desto besser? Schon einfachere Netzwerkarchitekturen, wie sie beispielsweise für den „Hello World“-Datensatz des maschinellen Trainings, den MNIST-Datensatz [3], entworfen wurden, überschreiten die Anzahl von 60.000 zu lernenden Verbindungen (lernbaren Parametern). Dieser MNIST-Datensatz besteht aus handgeschriebenen, 28x28 Pixel großen Schwarz-Weiß-Bildern der Ziffern 0-9. In *Abbildung 1* ist ein Auszug dieses Datensatzes dargestellt, wobei jede Ziffer zu genau einem Bild des Datensatzes gehört. Nachdem das KNN mit Daten zu jeder der zehn Klassen trainiert worden ist (Trainingsdatensatz), kann es die Vorhersage für ein Bild, das nicht im Training genutzt wurde, treffen (Testdatensatz). Die Genauigkeit der Vorhersage ist der Prozentsatz, wie viele vorher noch nicht gesehene Bilder richtig klassifiziert wurden. Die Fehlerrate beschreibt den Prozentsatz falsch vorhergesagter Bilder des Testdatensatzes. Beide Begriffe werden häufig als Performance des KNN zusammengefasst.

Diese Performance gilt es auch bei der Verwendung von Ausdünnungs-Verfahren (Pruning) zur Reduzierung der Komplexität für KNN beizubehalten. Dies wird erreicht, indem alle Verbindungen zwischen den Neuronen, deren Wert unterhalb eines vorher festgelegten Grenzwertes liegt, nach jedem Update-Schritt im Lernalgorithmus gekappt werden. Gibt es keine Verbindungen mehr zu oder von einem Neuron, fällt dieses ebenfalls weg. Somit wird darin investiert, wichtige beziehungsweise entscheidungsrelevante Verbindungen zu behalten und zu stärken und unwichtige beziehungsweise entscheidungsirrelevante Verbindungen zu entfernen. Dieses Verfahren ist in *Abbildung 2* skizziert.

Dass überhaupt Werte der Verbindungen durch Updateschritte unter einen Grenzwert

nahe null fallen, wird durch die Verwendung sogenannter Regularisierungsmethoden gewährleistet. Bei diesen werden große Werte der Verbindungen in der Zielfunktion des Lernprozesses neben den falschen Vorhersagen ebenfalls „bestraft“. Das Modell lernt somit, sich auf die wichtigen Verbindungen zu konzentrieren. Resultat der Pruning-Strategie ist ein ausgedünntes Netzwerk, das eine nahezu gleich gute Performance liefert. In gewisser Weise wird dabei der Hyperparameter, der für die Anzahl der Verbindungen zwischen den Neuronen steht, variabel gehalten und während des Trainings ebenfalls gelernt.

Forschungsergebnisse zeigen, dass die Schichten eines KNN, das auf den MNIST-Datensatz trainiert wurde, um mehr als 96 Prozent in ihrer Komplexität reduziert werden können, und das mit nur minima-

len Einbußen der Performance [4], dargestellt in *Abbildung 3*. Kurz: Die Komplexität der KNN ist nicht immer das Entscheidende. Vielmehr geht es darum, die relevanten Verbindungen zwischen den Neuronen zu finden und das KNN darauf zu beschränken. Dies kann ausreichend sein und spart enorme Rechenressourcen, Trainingsaufwand und den Bedarf großer Datensätze.

Automatisiertes maschinelles Lernen

Das zuvor beschriebene Ausdünnen ist ein möglicher Bestandteil des sogenannten AutoML-Prozesses. Das Ziel dieses Prozesses ist eine durchgängige End-to-End-Anwendung von künstlichen Intelligenzen, die selbstständig auch alle weiteren Hyperparameter innerhalb des Lernprozesses bestimmt. So wird auch Nicht-Experten die Nutzung von

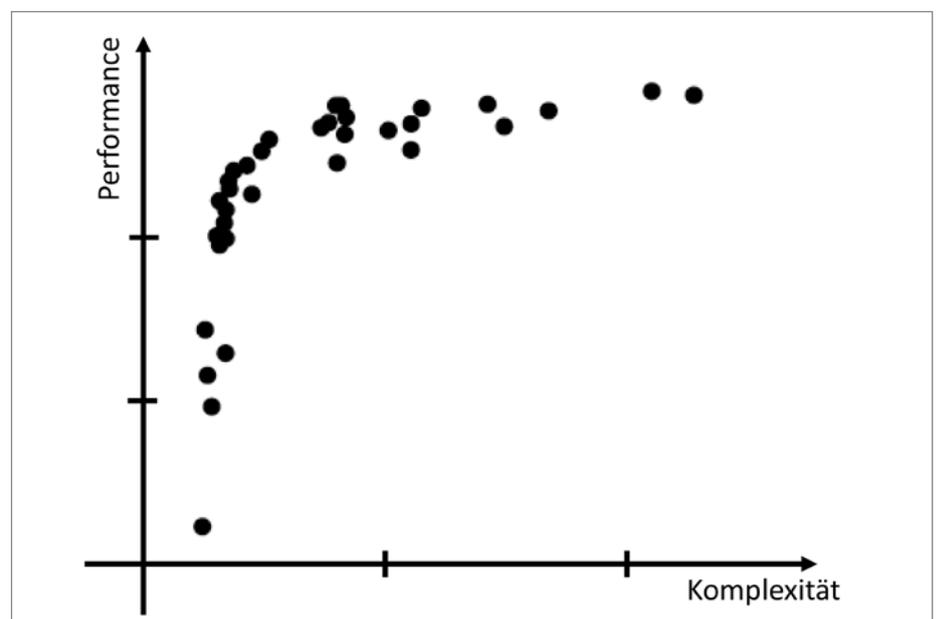


Abbildung 3: Die Performance des KNN bleibt bei sinkender Komplexität vorerst auf einem gleichen Niveau, bis das Netzwerk zu wenige Verbindungen aufweist (Quelle: Eigene Darstellung)

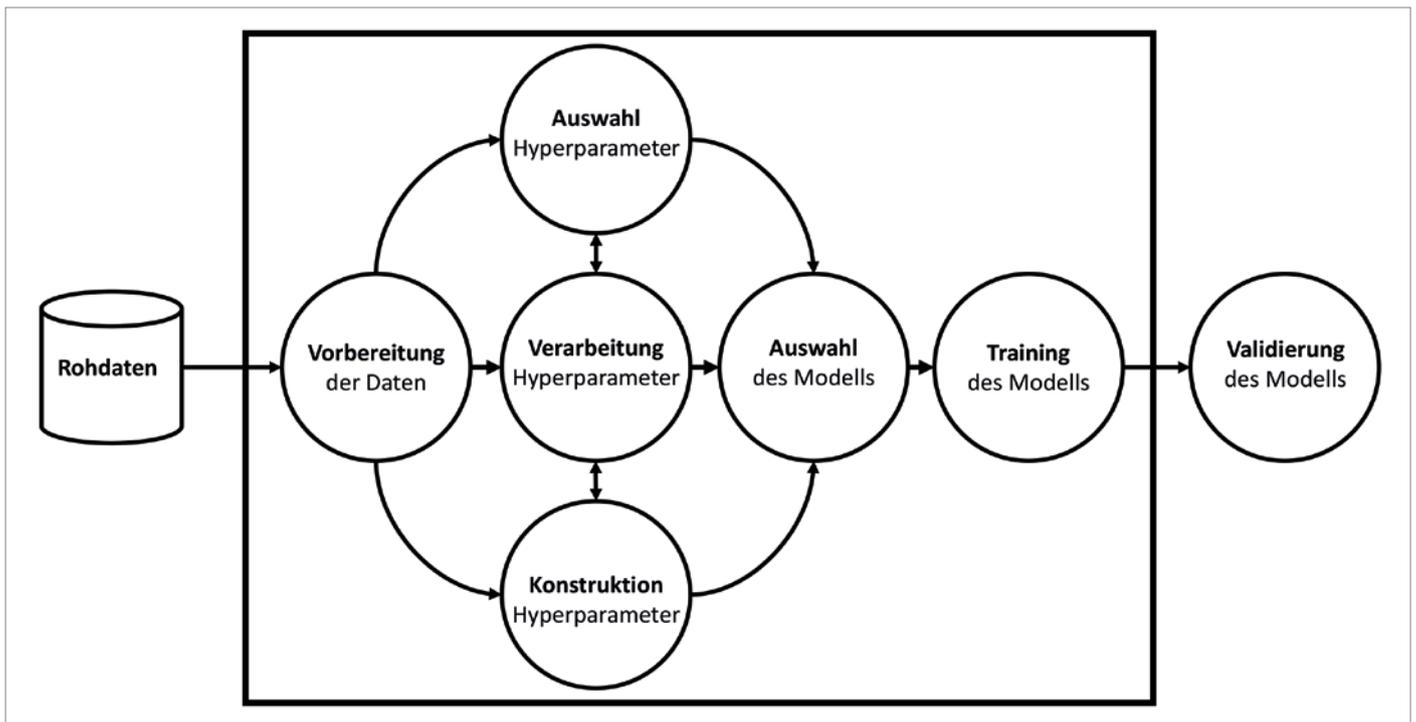


Abbildung 4: AutoML als Zukunftstechnologie (Quelle: Eigene Darstellung)

ganzheitlichen KI-Modellen und Techniken ermöglicht. Es deckt die komplette Pipeline vom Rohdatensatz bis zum einsatzfähigen maschinellen Lernmodell ab. In *Abbildung 4* ist schematisch dargestellt, welche Prozessschritte automatisiert erfolgen werden (alles innerhalb des Kastens) [7]. Dieser hohe Automatisierungsgrad hat den Vorteil, dass mögliche einfachere Lösungen schneller generiert werden können. Diese sind oft besser als von Hand entworfene Modelle.

Fazit

Während mit dem Pruning-Verfahren vorerst ein Hyperparameter (die Komplexität) mit in den Lernprozess integriert wird, könnte in einem vollständigen AutoML-Verfahren jeder einzelne Hyperparameter, der vor Beginn des Lernprozesses bestimmt werden muss, automatisiert ermittelt und im Trainingsprozess angepasst werden. Die Architekturen des KNN würden somit in allen Komponenten variabel gehalten. Mit dem *MorphNet* von Google wird erstmalig eine Architektur vorgeschlagen, die zusätzlich zum Ausdünnen der Schichten diese auch selbstständig erweitern kann [8].

Die Forschungen in diesem Umfeld laufen auf Hochtouren, sodass mit minimalem Aufwand an Rechenressourcen das bestmögliche Ergebnis in Form von geringer Komplexität und hoher Vorhersagegenauigkeit erzielt werden kann.

Quellen

- [1] Lufthansa Industry Solutions, (2020), Folgestudie 2020: Die neue IDG-Studie „Machine Learning 2020“, <https://www.lufthansa-industry-solutions.com/de-de/studien/idg-studie-machine-learning-2020>
- [2] Frank Bäumer, (2015), Welt in Zahlen – Das Gehirn <https://www.daserste.de/information/wissen-kultur/w-wie-wissen/sendung/2009/welt-in-zahlen-das-gehirn-100.html>
- [3] Fan, Guangzhe & Wang, Zhou & Wang, Jiheng, (2010), CW-SSIM kernel based random forest for image classification. Proceedings of SPIE - The International Society for Optical Engineering. 10.1117/12.863528.
- [4] Reiners, Malena and Klamroth, Kathrin and Stiglmayr, Michael, (2020), Efficient and Sparse Neural Networks by Pruning Weights in a Multiobjective Learning Approach, <https://arxiv.org/abs/2008.13590>
- [5] Lechao Xiao, Yasaman Bahri, Jascha Sohl-Dickstein, Samuel S. Schoenholz, Jeffrey Pennington, (2018), Dynamical Isometry and a Mean Field Theory of CNNs: How to Train 10,000-Layer Vanilla Convolutional Neural Networks <https://arxiv.org/pdf/1806.05393.pdf>
- [6] Fabian Niehaus, (2019), Neuronale Netze: Potenziale und Hürden <https://www.cologne-intelligence.de/blog/detail/show/neuronale-netze-potenziale-und-huerden/>
- [7] Randal S. Olson, Nathan Bartley, Ryan J. Urbanowicz, Jason H. Moore, (2016), Eval-

uation of a Tree-based Pipeline Optimization Tool for Automating Data Science, <https://arxiv.org/abs/1603.06212>

- [8] Ariel Gordon, Elad Eban, Ofir Nachum, Bo Chen, Hao Wu, Tien-Ju Yang, Edward Choi, (2018), MorphNet: Fast & Simple Resource-Constrained Structure Learning of Deep Networks, <https://arxiv.org/abs/1711.06798>



Malena Maria Reiners
malena.reiners@nrwbank.de

Malena Reiners schloss im September 2019 ihr Masterstudium der Mathematik mit Schwerpunkt Optimierung an der Bergischen Universität ab. Bereits im Masterstudium, aber auch danach als Angestellte an der Universität, beschäftigte sie sich intensiv mit dem Thema Optimierung von KI-Modellen. Sie durfte für zehn Monate in Zusammenarbeit mit der Bergischen Innovationsplattform für Künstliche Intelligenz (BIT) an Forschungsarbeiten zu aktuellen Optimierungsverfahren für Faltungsnetzwerke teilnehmen; dies führte zum Einreichen und Veröffentlichung eines eigenen Papers im September 2020. Seit Oktober 2020 ist sie als Projektmanagerin in der IT-Abteilung der NRW.BANK beschäftigt und arbeitet auch dort an der Optimierung aktueller Automatisierungsverfahren.



Digitale Evolution in Zeiten von VUCA: Agilität, das Gebot der Stunde

Dr. Frank Schönthaler und Marion Rosemann, PROMATIS Gruppe, Ettlingen (TechnologieRegion Karlsruhe)

Covid-19 lässt uns alle die VUCA-Welt spüren: Exponentiell steigende Fallzahlen, täglich neue, teils widersprüchliche Nachrichten und Verhaltensregeln, radikal reduzierte Planbarkeit. „Fahren auf Sicht“ lautet die Devise. Zum Zeitpunkt dieses Beitrags ist noch unklar, wohin die Reise uns führt. So erfährt das VUCA-Paradigma durch die Corona-Krise wieder völlig neue Relevanz und rückt vermehrt in den Fokus der Aufmerksamkeit. Es stellt sich die Frage, welche Strategie in Zeiten von Unbeständigkeit und Komplexität ratsam ist. Gerade mit der digitalen Evolution können Unternehmen diese Unsicherheiten überwinden.

Digitale Evolution versus Disruption

Innovationen und Neuerungen sind natürlich wichtig, doch sollte man darüber nicht vergessen, dass auch in der Technologie Erfahrung ein hohes Gut ist. Harvard-Professor Clayton Christensen gilt als Erfinder der Theorie der disruptiven Innovation (engl. disruptive = störend, unterbrechend). Digitale Disruption beschreibt laut Christensen einen Prozess, bei dem eine etablierte Tech-

nologie oder ein Service von einer Innovation verdrängt wird. Gefestigte Unternehmen können jedoch mit diesem Konzept eines Neubeginns schnell scheitern, denn es bedeutet ein Zurück auf null und damit einen erhöhten Kapitalbedarf. Die digitale Evolution hingegen verfolgt einen Ansatz, der wesentlich risikoärmer durchzuführen ist, und mit ihrer Dynamik ist sie ohne Zweifel ein wichtiger Treiber für die Veränderungen in der VUCA-Welt.

Herkunft und Herausforderung der VUCA-Welt

Mit dem Zerfall der Sowjetunion Anfang der Neunzigerjahre erlangte das Akronym VUCA seine erstmalige Bedeutung. Gebildet aus den englischsprachigen Begriffen volatility (Unbeständigkeit), uncertainty (Unsicherheit), complexity (Komplexität) und ambiguity (Mehrdeutigkeit) beschrieb es sehr treffend den Zustand der mit dem Ende des Kalten Krieges entstandenen multila-

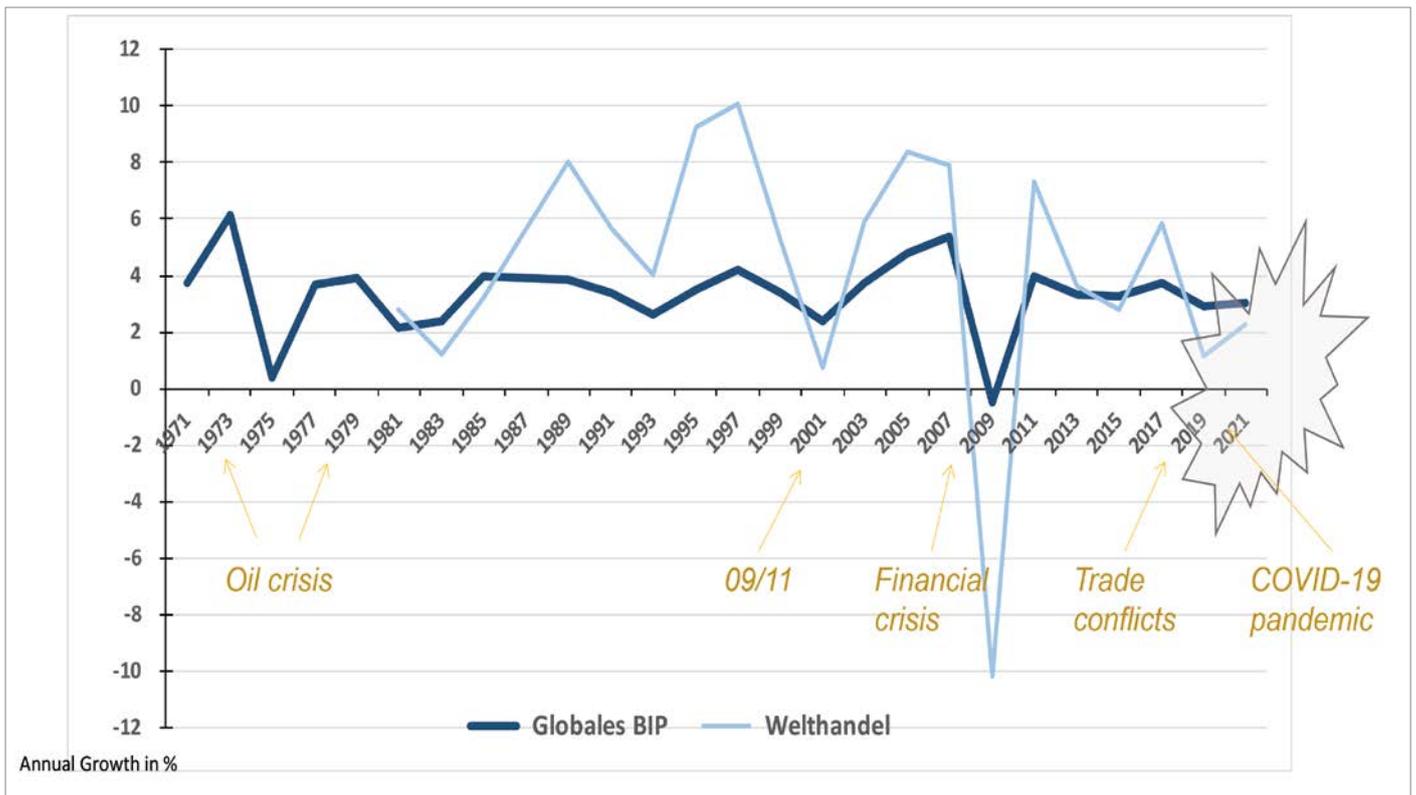


Abbildung 1: Langfristige Entwicklung des globalen BIP und Welthandels (Quelle: OECD Economic Outlook No. 106 – Nov. 2019 [from 2019 prognoses])

teralen Welt (siehe [BeL14, MKK16]). Auch die Betrachtung unbeständiger und unsicherer Krisenzeiten (siehe Abbildung 1) lässt auf das hochexplosive Potenzial zukünftiger Krisen schließen. Fragen nach disruptiver Digitalisierung, Flüchtlingskrisen, Verschärfung der Energiekrise, nahende Klimakatastrophe sowie die Folgen von BREXIT und fortschreitender Deglobalisierung beschäftigen den Menschen maßgeblich in seinem Lebens- und Arbeitsumfeld.

Dies führt unweigerlich zu Erfolgsfaktoren, die für das Meistern der VUCA-Welt elementar sind. Auch hierfür wird VUCA gerne als Akronym genutzt: vision (Zukunftsvision), understanding (Verstehen), clarity (Klarheit) und agility (Agilität). Übertragen auf die strategische Unternehmensführung setzt Erfolg eine Zukunftsvision voraus, die als Leitbild und Motivation für die Aktivitäten im Unternehmen dient. Zudem müssen die Zusammenhänge verstanden werden, die Ursachen und die Regeln, die für die Veränderungen verantwortlich sind, deren Konsequenzen und die Risiken, die mit den Veränderungen verbunden sind. Klarheit steht für transparente Prozesse und Regelwerke sowie eindeutig formulierte Ziele, Strategien und Pläne. Und schließlich ist für erfolgreiche Unternehmen in der VUCA-Welt Agilität unabdingbar, die Fähigkeit,

sich schnell und effizient an die Veränderungen anzupassen (vgl. [RSN19]). Aber was genau ist Agilität und wie kann ein Unternehmen agil werden?

Agilität als Schlüsselfaktor der Digitalisierung

Agilität ist kein neues Thema, sondern existiert bereits seit fast 70 Jahren in unterschiedlichen Facetten und Ausprägungen. Es erhält aber durch die Digitalisierung eine größere Bedeutung und eine hohe Aktualität. Agilität ist quasi das Zusammenspiel von Geschwindigkeit, Flexibilität und Fokus und beschreibt die Fähigkeit eines Unternehmens, sich kontinuierlich an seine komplexe, turbulente und unsichere Umwelt anzupassen, indem es diese Veränderungen möglichst rechtzeitig antizipiert und sein Geschäftsmodell und seine Arbeitsprozesse darauf ausrichtet.

In einem Beitrag der GfO (Gesellschaft für Organisation e.V.) wird im Zusammenhang mit der Transformation zum agilen Unternehmen eine interessante Beobachtung behandelt: Das Mehr an Anpassungsfähigkeit zulasten der Stabilität geht oftmals mit einem Verlust der Identität und einer Gefährdung der sozialen Integration der Organisation einher. Daraus leiten die Autoren die Empfehlung ab, Agilität und Stabilität nicht

als getrennte Phänomene zu betrachten, sondern als zwei sich wechselseitig bedingende Spannungspole, die in Abbildung 2 (rechts) um die Performance der Organisation als drittem Spannungspol ergänzt worden ist. Das Management der Spannungspole ist dann eine essenzielle Aufgabe der Unternehmensführung (siehe [GLL18]).

Das Prinzip der Agilität muss allerdings über das gesamte Unternehmen hinweg gelebt werden. Es reicht nicht, punktuell agile Projekte durchzuführen und ansonsten starre Unternehmensprozesse zu leben. Vielmehr muss das Prinzip der Agilität fest in der Unternehmensarchitektur verankert werden (siehe Abbildung 2 links). So können agile Unternehmen in einer immer komplexeren Welt schneller auf Umbrüche reagieren und gleichzeitig qualitativ hochwertige Ergebnisse produzieren. Auch deshalb wird agiles Denken und Handeln in Zukunft ein unerlässlicher Erfolgsfaktor sein.

Das Geheimnis agiler Unternehmen

Die Darstellung der Unternehmensarchitektur in Abbildung 2 orientiert sich an den Vorschlägen des BITKOM (siehe [BIT11]). So lautet die Anforderung an die Unternehmens-IT, die Komplexität des wirtschaftlichen Handelns widerzuspiegeln sowie die laufenden Veränderungen und Herausforderungen des

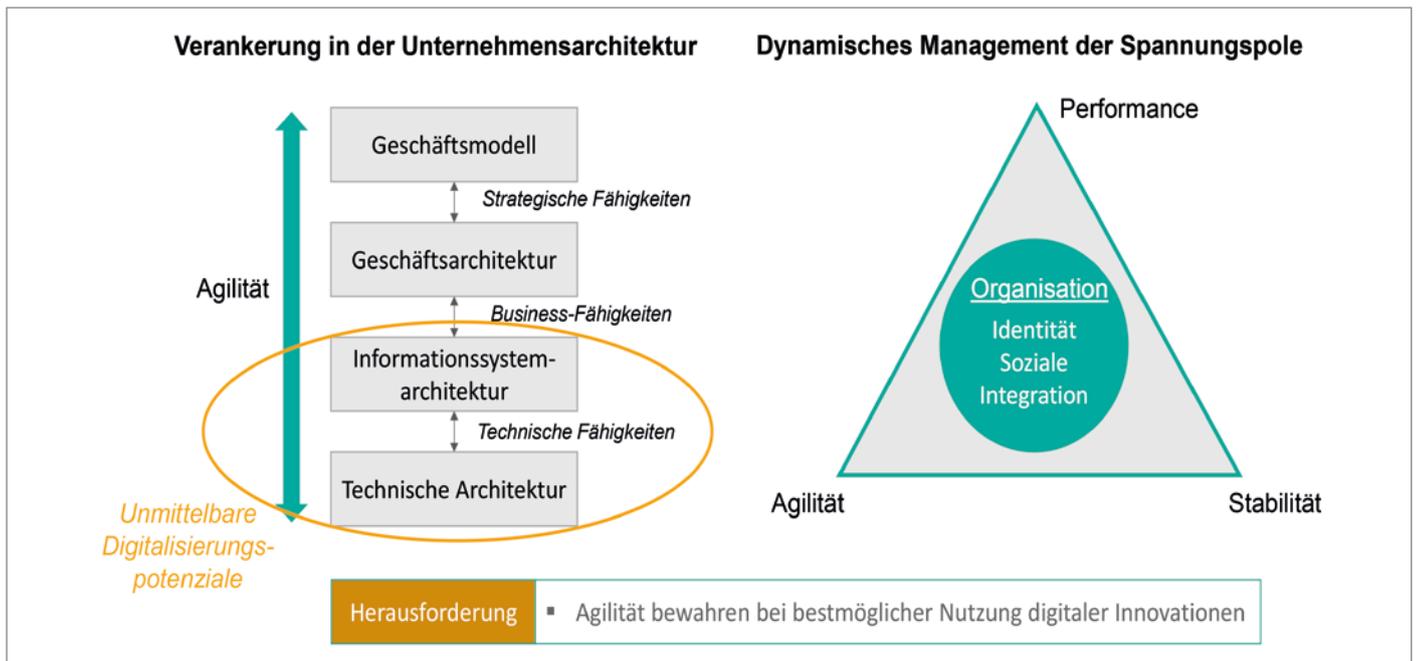


Abbildung 2: Führung in agilen Unternehmen (Quelle: PROMATIS)

Markts zu berücksichtigen. Das gedankliche Konzept und das Handwerkszeug dafür liefert Enterprise Architecture Management (EAM). EAM bietet mit seinem ganzheitlichen Ansatz, der Strukturierung in Architekturebenen, den zugehörigen Umsetzungsmethoden und der Governance-Grundlagen für eine strategische Ausprägung der gesamten Geschäftsprozess-IT-Landschaft im Sinne der Business-Anforderungen und bereitet den Boden für lokale Verbesserungen und Optimierungen. Ausgehend vom Geschäftsmodell als oberster Ebene, auf der der Geschäftskontext, Ziele, Strategien und Performance-Kennzahlen beschrieben sind, folgt die Geschäftsarchitektur, die die strategischen Fähigkeiten zur Umsetzung des Geschäftsmodells bereitstellt. Die Informationssystemarchitektur in Verbindung mit der technischen Architektur sorgt für die Business-Fähigkeiten zur Umsetzung der Unternehmensstrategien; gerade hier sind unmittelbare Digitalisierungspotenziale gegeben. Nun stellt sich allerdings die Frage nach den direkten Handlungs- beziehungsweise den strategischen Digitalisierungsfeldern.

Strategische Digitalisierungsfelder

Bevor ein Unternehmen in die digitale Welt startet, sollte Klarheit über die Digitalisierungsfelder herrschen, damit Orientierung vorhanden ist und die Ressourcen richtig priorisiert werden. Alles andere führt zu Missverständnissen, zu Verzettelung und am Ende zu ausufernder Diskussion ohne Ergebnis. Ein ef-

fizienter Diskussions- und Ordnungsrahmen der Digitalisierungsfelder ist die Digitalisierungsmatrix nach [Stö20]. Diese besteht aus zwei Koordinaten: Auf der Abszisse befinden sich das aktuelle und das neue Geschäftsmodell. Digitalisierungslösungen können das bestehende Geschäftsmodell verstärken oder ein völlig neues initiieren. Auf der Ordinate sind der Markt- und der Produktivitätsfokus dargestellt. So muss jede Digitalisierungsmaßnahme zu einem Fortschritt in die eine oder andere Richtung führen. Diese Digitalisierungsmatrix bildet damit die Grundlage für die Erarbeitung und die Entscheidung über notwendige Schwerpunkte der Zukunft. Entlang den Dimensionen der Digitalisierungsmatrix können nun vier digitale Normstrategien dargestellt werden, in denen die Digitalisierungsfelder liegen. Die erste ist die Professionalisierung durch Digitalisierung. Innerhalb des bestehenden Geschäftsmodells wird primär an der Produktivität gearbeitet, das heißt an Abläufen, Funktionen, Durchlaufzeiten und Führungsprozessen. Die zweite wird als Perfektionierung durch Digitalisierung bezeichnet. Auch hier steht die Produktivität im Vordergrund, allerdings erfolgt dies im Rahmen eines neuen Geschäftsmodells. Normstrategie drei ist die Marktentwicklung durch Digitalisierung. Im Rahmen des bestehenden Geschäftsmodells werden neue Märkte, Kunden, Produkte sowie Dienstleistungen forciert. Die vierte wird als Innovation durch Digitalisierung beschrieben. Kennzeichen sind Marktlösungen durch ein völlig

neues Geschäftsmodell. Dies sind die radikal neuen Ansätze, die oftmals zu Veränderungen von Unternehmen und Branchen führen und als Substitution beziehungsweise Disruption betitelt werden.

Agilität in Digitalisierungsprojekten

Wie aber gilt es, die Marschroute für Digitalisierungsprogramme, bestehend aus mehreren optimal aufeinander abgestimmten Projekten, zu definieren? Viele Unternehmen möchten sich agil transformieren – aber nicht in letzter Konsequenz. Häufig ist hier ein opportunistisches Vorgehen anzutreffen, das sich in nicht zielführenden Verhaltensweisen äußert. Es wird beispielsweise versucht, Marktführer oder Mitbewerber – die „glitzernen Sterne“ – zu kopieren oder zu imitieren. Ähnliche Stolpersteine sind ausschließlich zufallsgesteuerte Wissensakquisition oder unstrukturiertes, wenig zielgerichtetes Experimentieren mit Digitaltechnologien. Eine selektive Wahrnehmung, das „Digital Cherry Picking“, beschreibt in dem Zusammenhang das Ausblenden von Beweisen oder die Täuschung durch unvollständige Belege. Sich das jeweils Beste aus verschiedenen Konzepten herauszuholen, während ein bedeutsamer Anteil verwandter Fälle oder Daten ignoriert wird, die dieser Position widersprechen könnten, wird sich als Fallstrick erweisen. Auch bei der „Moving-Target-Problematik“ sollten alle Alarmglocken eines jeden Projektleiters läuten, dies führt ebenfalls zu keiner gewünschten Erfolgsstrategie. Wandel um des Wandels

willen allein – ohne Konsistenz und klare Strategie – ist folglich überflüssig. Stattdessen gilt es, die Herausforderung zu meistern, die Agilität bei vorgegebenen Ressourcen-, Kosten- und Zeitbudgets zu bewahren. Wenn also von digitaler Evolution die Rede ist, wird oft die technologische Evolution, die Bereitstellung neuer Technologien, in den Mittelpunkt gerückt. Hierzu bedarf es zunächst der Entwicklung einer digitalen Vision. Es findet ein gezielter Wissensaufbau in der Nutzung digitaler Innovationen statt. Im Idealfall wird mit Modellen und Prototypen zur Identifikation und Priorisierung von Evolutionswellen gearbeitet. Systematisch und anhand von umfassenden Unternehmensmodellen, die die Prozesse innerhalb des Unternehmens optimiert darstellen, muss Wert auf Transparenz und Offenheit gelegt werden. Der Geschäftsprozess-Spezialist spricht hier von einem evolutionären Prozess, der unterschiedliche Stadien, die sogenannten Wellen, durchläuft. Der Treiber ist die Innovation, dem die Implementierung folgt, doch entscheidend für eine langfristig ausgerichtete Digitalisierung sind das permanente Monitoring und die daraus folgenden Prozessverbesserungen. Für eine erfolgreiche Umsetzung des digitalen Evolutionsprogramms ist das Thema Governance ebenfalls ein wesentlicher Baustein, um der Komplexität der Digitalisierung gerecht zu werden. Dabei besteht das Ziel darin, eine nachhaltige Wertschöpfung zu gewährleisten, die sowohl effektiv als auch effizient ist. Ergänzend für eine erfolgreiche digitale Evolution sollte ein Ansatz der Planbarkeit gelebt werden. Wir sprechen hier von „Solution-driven Requirements Discovery“ in agilen Umsetzungsprojekten. Diese lösungsorientierte Ermittlung von Anforderungen spiegelt sich auch im umfassenden Vorgehensmodell IQPM™ (vgl. [IQPM20]) wider, in dem ein Vorgehen für eine zielgerichtete Projektarbeit definiert ist. IQPM steht für Integrated Quality Process Model, betont also bereits im Namen die Wichtigkeit der Qualität des Entwicklungsprozesses und der Prozessergebnisse. IQPM definiert Projektaktivitäten, Ergebnisobjekte, Rollen sowie kausale und zeitliche Abhängigkeiten; zudem Anforderungen an die im Projekt zu verwendenden Methoden und Tools. Elementare Bestandteile sind Vorgaben für das Projekt- und Qualitätsmanagement sowie für die Steuerung des organisatorischen Veränderungsprozesses. Bewährte Best Practice Templates sorgen dafür, dass IQPM den Anforderungen eines Geschäftsprozess-Reengineering ebenso gerecht wird wie einem agil durchgeführten Digitalisierungsvorhaben.

Erfolgsfaktoren in agilen Digitalisierungsprojekten

Aufbauend auf der zuvor beschriebenen strategischen Herangehensweise bedarf es allerdings folgender Erfolgsfaktoren, die aus Praxiserfahrungen von Unternehmen abgeleitet wurden:

- **Agiles Vorgehensmodell**
 - Klar strukturiertes und definiertes Vorgehensmodell
 - Methodenexpertise (vgl. Scrum Master)
 - Effektive Teamausbildung

In der Konsequenz ergibt sich ein Nutzen, der vor allem in einer hohen Team-Performance und Ergebnisqualität, Agilität und Vorhersehbarkeit sowie Team-Zufriedenheit durch Methodensicherheit liegt.
- **Erfahrenes Projektmanagement**
 - Erfahrung in agilen Digitalisierungsprojekten
 - Ausgeprägte Kommunikations- und Lösungskompetenz
 - Leadership, gepaart mit Innovationsfreude

Als Resultat lassen sich Agilität, verbunden mit Termin- und Budgettreue – die Basis für einen effektiven Veränderungsprozess –, sowie ein synergetisches Zusammenwirken von Projektteam und Auftraggeber erzielen.
- **Motiviertes Team**
 - Agile Kompetenz und Neugier
 - Fähigkeit zur interdisziplinären Teamarbeit
 - Infrastruktur für verteiltes, asynchrones Arbeiten

Auch hier ergibt sich eine hohe Team-Performance und Ergebnisqualität, ein Teamspirit, der für Ergebnisoptimierung sorgt, sowie eine optimale Nutzung aller Fähigkeiten im Team.
- **Innovation**
 - Bereitschaft und Fähigkeit zur digitalen Innovation
 - Komplementarität von Business-, Prozess- und Technologie-Exzellenz
 - Wirksames Change Management

Der Schlüssel der Innovation liegt demnach klar auf der Hand und drückt sich in Nachhaltigkeit der digitalen Innovation, der optimalen Ausschöpfung der Innovationspotenziale sowie in der Schnelligkeit und Wirtschaftlichkeit in der Umsetzung aus.



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.

PROMATIS Gruppe
Pforzheimer Str. 160
76275 Ettligen
+49 7243 2179-0
www.promatis.de

Ettligen | Hamburg | Berlin | Münster
Wien | Zürich | Denver

Agile Business-Plattform – Sprungbrett in die Digitalisierung

Neben der Definition der jeweils optimalen Strategie ist die Fähigkeit zur zuverlässigen Umsetzung der Strategie entscheidend. Denn eines ist vollkommen sicher: Eine Strategie ist nur so gut wie ihre Umsetzung! Aber wie muss eine optimale Plattform aussehen, um dort agile Digitalisierungsprojekte durchzuführen? Oracle bietet hierfür eine komplette Suite an Software-as-a-Service-Lösungen (SaaS) und Plattform-as-a-Service-Lösungen (PaaS) an, die miteinander kombiniert werden. In *Abbildung 3* ist eine idealisierte Business-Plattform mit den Oracle Public Cloud Services dargestellt, ergänzt um die Horus Public Cloud (vgl. [SVO11, VSD17]), die den Bereich der Prozessmodellierung abbildet.

Weitere Treiber für Agilität

Wagt man einen Blick auf weitere Potenziale in Oracle-Digitalisierungslösungen, so lassen sich Trends beobachten, die durch eine Intensivierung des Einsatzes moderner digitaler Technologien getrieben sind:

- **Autonomous Computing**

Dieser Lösungsansatz bietet eine autonome Datenverarbeitung beziehungsweise die Fähigkeit eines Computers, sich selbst automatisch durch adaptive Technologien zu verwalten. Dank intelligenter Automati-

sierung von Routineaufgaben erhalten die Business- und IT-Experten mehr Freiräume für die nicht automatisierbaren Aufgaben bei gleichzeitig verbesserter Ergebnisqualität. Auch die Vermeidung von Fehlern sowie laufende Prozessverbesserungen führen schließlich zu mehr Sicherheit und einer besseren Performance.

- **Machine Learning**

Durch den zunehmenden Einsatz von Machine-Learning-Verfahren wird ein wesentlich höherer Automatisierungsgrad erreicht, indem ganze Bündel von Aufgaben im Planungsprozess komplett in die Software verlagert werden. Die Nutzung erstreckt sich von allen Cloud-Applikationen bis hin zur Infrastruktur. Verschiedenste Algorithmen, die in der Lage sind, selbstständig aus Daten zu lernen und hieraus Wissen zu generieren sowie intelligente Prozesse zu realisieren und Entscheidungen zu treffen, etablieren somit neuartige Services über die gesamte Wertschöpfungskette hinweg.

- **Oracle Fusion Analytics Warehouse**

Oracle Fusion Analytics Warehouse bietet vorgefertigte Datenmodelle und Analysen für Oracle-Cloud-Anwendungen, die auf Autonomous Data Warehouse und Oracle Analytics basieren. Dies ermög-

licht die automatische Generierung eines Analytics Warehouse mit jederzeit aktuellen Daten aus den Oracle Fusion Cloud Applications (verfügbar für ERP und Human Capital Management; geplant für Customer Experience, Supply Chain Management und NetSuite). Unternehmen profitieren von vorgefertigten Best Practice KPIs, Berichten und Dashboards sowie von der Datenintegration aus beliebigen externen Datenquellen.

Vom Geschäftsprozess in die Smart Solutions

Zurück zu unserer idealisierten Business-Plattform, mag es nicht verwundern, dass ein großes Potenzial in Smart Solutions zu finden ist, denn sie sind die ideale Basis, um den Weg des digitalen Wandels erfolgreich zu beschreiten. Sie vereinheitlichen die IT-Infrastruktur und machen die gesamte Organisation mit allen Fachabteilungen schneller, flexibler und intelligenter. Aber was zeichnet Smart Solutions im Detail aus? Es kommen neben Autonomous Computing ebenfalls Big Data, Internet of Things (IoT) und Artificial Intelligence (AI) zum Einsatz. *Abbildung 4* illustriert die konsequente Umsetzung dieses Konzepts: So verschmelzen intelligente Daten und relevante Geschäftsanwendungen aus allen Unternehmensbereichen. Erst die sinnvolle Auswahl und Konfiguration solcher Smart Applications stellt eine fundierte

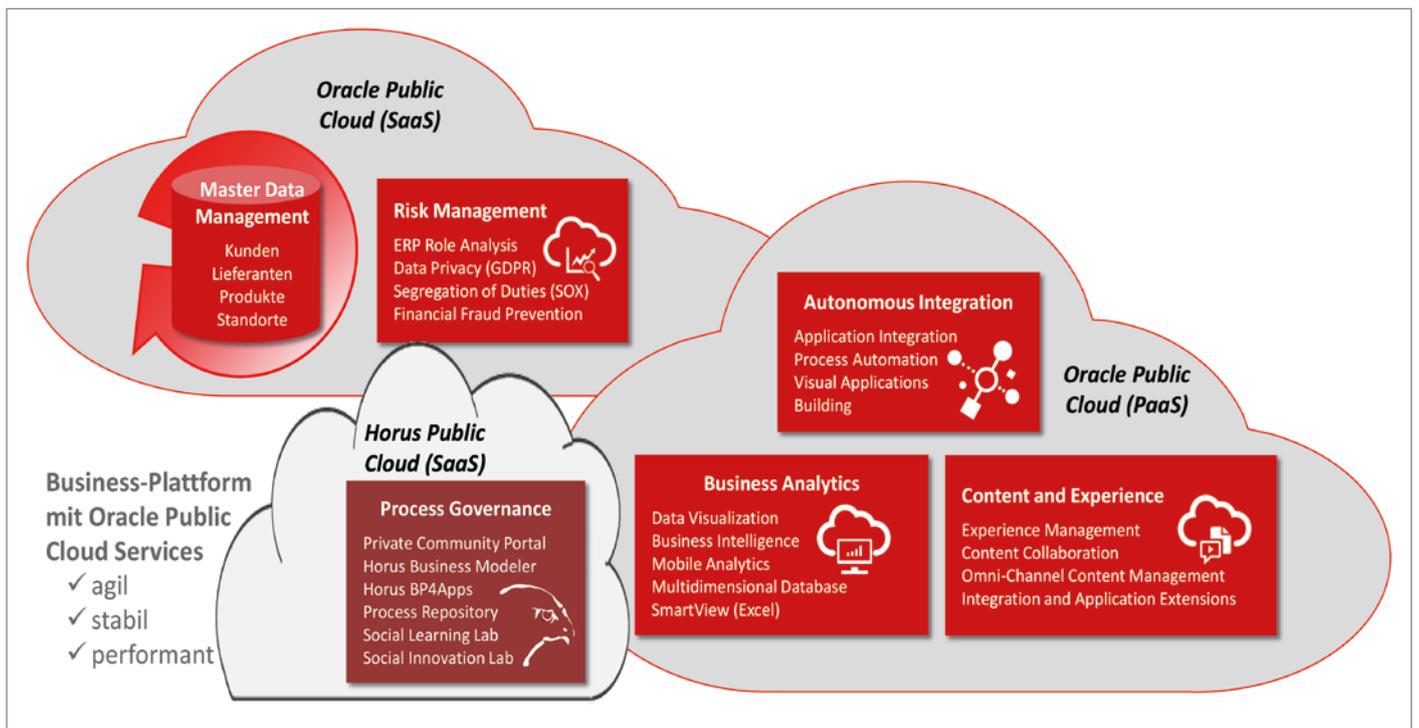


Abbildung 3: Idealisierte Business-Plattform mit Oracle-Produkten (Quelle: Oracle/PROMATIS)

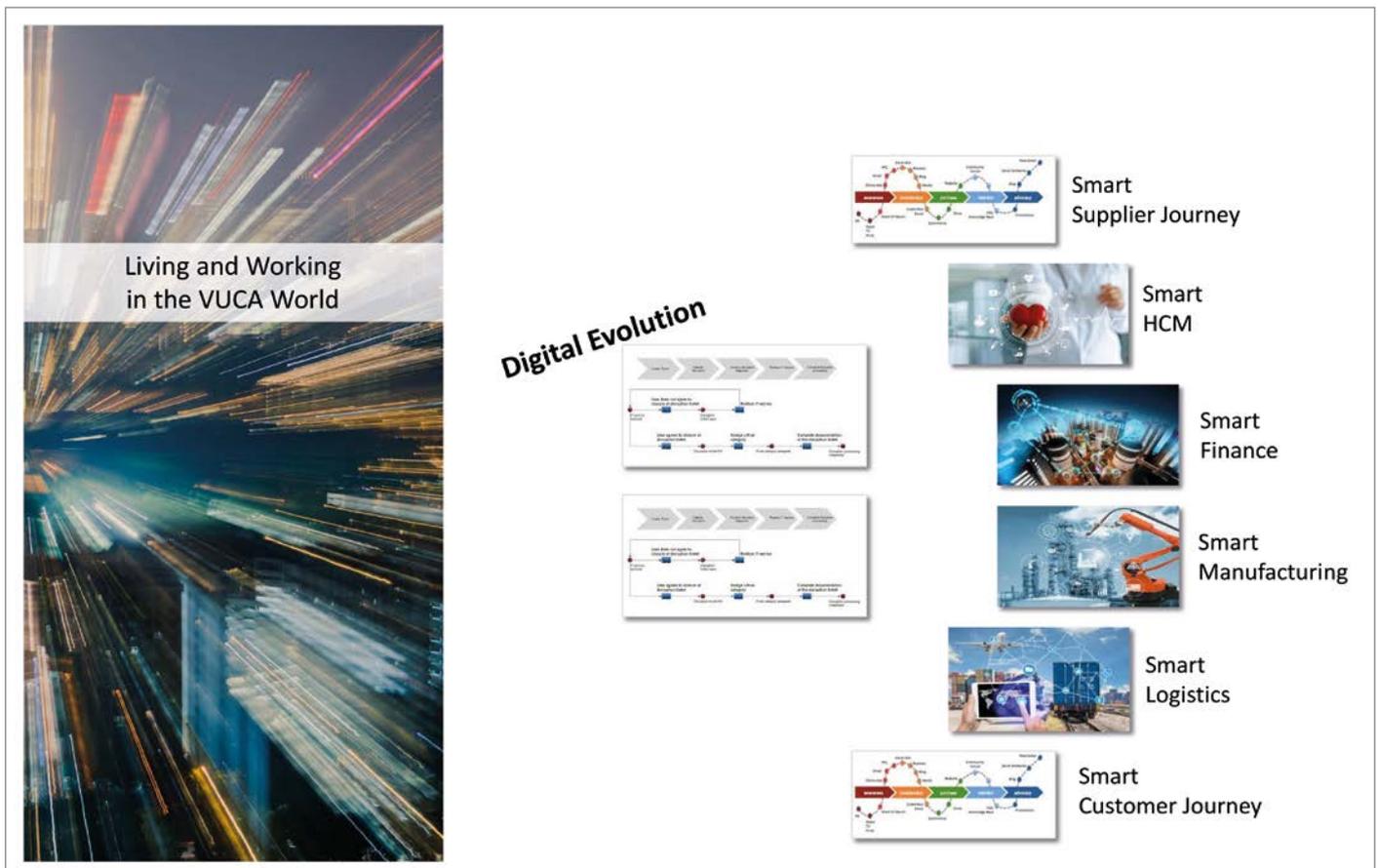


Abbildung 4: Digitale Evolution durch den Einsatz von Smart Solutions (Quelle: PROMATIS)

Unterstützung der Geschäftsprozesse dar und verdient im Kontext von Agilität die Bezeichnung Smart Solutions.

Fazit und Ausblick

Gravierende Veränderungen sind Kennzeichen und Treiber für die VUCA-Welt. Wie schon zu Beginn erwähnt, ist zum Zeitpunkt dieses Beitrags noch unklar, wohin uns die Covid-19-Reise führt. So ist die aktuelle Corona-Krise ein Musterbeispiel für eine ungewisse und tiefgreifende Veränderung, die kaum jemand in dieser Intensität für wahrscheinlich hielt. Jetzt im Stillstand zu verharren, wäre ein großer Fehler. Gerade in Zeiten massiven Umbruchs gilt es, in knapp bemessenen Zeitfenstern Maßnahmen zu ergreifen, um die schlimmsten Auswirkungen abzuwenden und aus der Krisenzeit nach Möglichkeit auch eine Zeit für Chancen hervorzubringen. Die Ausführungen in diesem Beitrag zeigen, dass es durchaus Perspektiven, praktikable Konzepte und leistungsfähige Planungsinstrumente gibt, um diesen Herausforderungen wirksam zu begegnen. Agilität ist und bleibt das Gebot der Stunde, denn gerade agile Unternehmen reagieren in komplexer

werdenden Welten schneller auf Umbrüche und den digitalen Wandel.

Referenzen

[BeL14] N. Bennett, G.J. Lemoine (2014): What VUCA really Means for You. In Harvard Business Review, Jan.-Feb. 2014.

[BIT11] BITKOM (Hrsg.) (2011): Enterprise Architecture Management – neue Disziplin für die ganzheitliche Unternehmensentwicklung. BITKOM, Berlin. <https://www.bitkom.org/Bitkom/Publikationen/Leitfaden-EAM-Enterprise-Architecture-Management.html>

[GLL18] H. Gergs, A. Lakeit, B. Linke (2018): Agilität braucht Stabilität: Was Unternehmen von Kampfflugzeugen, James Bon und Moses lernen können. In zfo 05/2018, Schäffer-Poeschel Verlag, 314-319.

[IQPM20] IQPM™ - Projekt-Vorgehensmodell der PROMATIS Gruppe, Ettlingen

[RSN19] D.K. Rigby, J. Sutherland, A. Noble (2019): Das agile Unternehmen. In Harvard Business Manager, Jan. 2019, 33-42.

[Stö20] R. Stöger (2020): Die strategischen Digitalisierungsfelder. Klarheit über die Schwerpunkte der Digitalisierung gewinnen. In zfo 05/2020, Schäffer-Poeschel Verlag, 109.

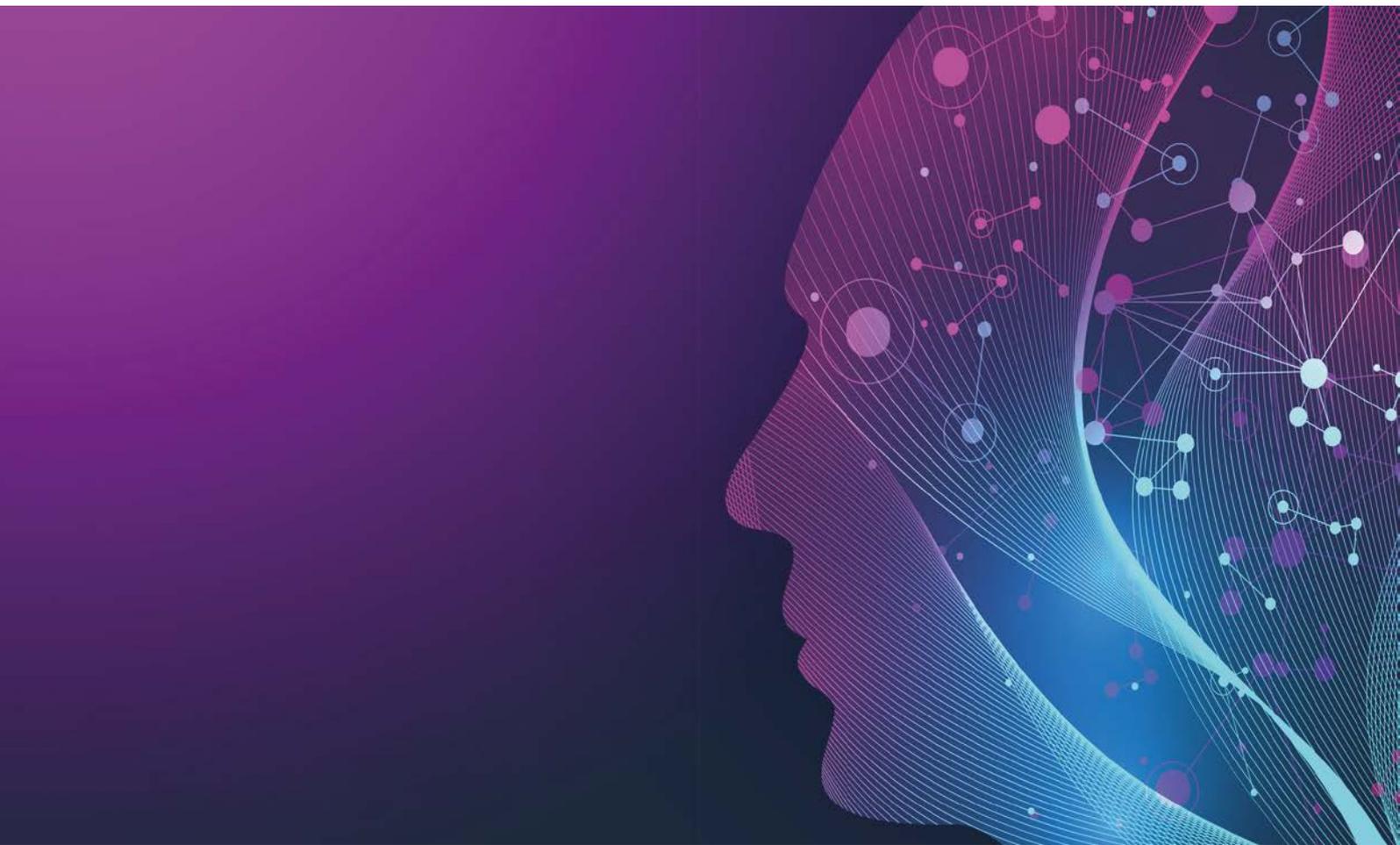
[SVO11] F. Schönthaler, G. Vossen, A. Oberweis, T. Karle (2011): Geschäftsprozesse für Business Communities: Modellierungssprachen, Methoden, Werkzeuge. Oldenbourg Wissenschaftsverlag, München.

[VSD17] G. Vossen, F. Schönthaler, S. Dillon (2017): The Web at Graduation and Beyond: Business Impacts and Developments. Springer International Publishing, Cham, CH.



Dr. Frank Schönthaler
 Geschäftsführender Gesellschafter (CEO)
 PROMATIS Gruppe
 frank.schoenthaler@promatis.de

Dr. Frank Schönthaler verantwortet als geschäftsführender Gesellschafter der PROMATIS-Gruppe die Unternehmens- und Produktstrategie sowie das operative Kerngeschäft. Als Executive Consultant begleitet er internationale Kunden in ihrer digitalen Evolution.



Data Science, KI, maschinelles Lernen: Was kann das alles (nicht)?

Maximilien Kintz, Fraunhofer IAO

Der Hype um künstliche Intelligenz (KI), maschinelles Lernen, neuronale Netze, Data Science und Co. wird nicht kleiner – Verwirrung, unrealistische Erwartungen oder unbegründete Ängste bleiben bei vielen bestehen. In diesem Artikel wird versucht, einen Überblick über diese und verwandte Themen sowie aktuelle Stärken, Grenzen und offene Forschungsfragen zu schaffen.

Definitionen

Fangen wir mit dem Oberbegriff **Künstliche Intelligenz** an. Eingeführt in den 1950er Jahren [1], hat er bis heute keine genau etablierte Definition. Er wird generell benutzt, um Softwaresysteme zu beschreiben, die nichttriviale Aufgaben erledigen, für die man typischerweise menschliche In-

telligenz bräuchte. Wie die Systeme gebaut sind oder wie diese Aufgaben aussehen können, ist meistens offengelassen.

Immer häufiger wird aber der verwandte Begriff **Maschinelles Lernen** (ML), eigentlich eine Untermenge von KI, als Synonym verwendet. Klassische Softwareprogramme basieren auf von Menschen definierten

Reihen von Anweisungen und Regeln, um Aufgaben zu lösen (zum Beispiel „wenn der Kunde das Produkt A gekauft hat, soll man ihm auch Produkt B und C, aber nicht D vorschlagen“). Bei Verfahren des maschinellen Lernens versucht man, ähnliche Ziele zu erreichen, ohne die Regeln vorgeben zu müssen. Stattdessen erhält der Computer eine (große)

Menge an sogenannten Trainingsdaten (zum Beispiel die Kaufhistorie von vielen Kunden), aus denen er selbst lernen soll, wie die Regeln sind (zum Beispiel welche Produkte vorgeschlagen werden sollten). Für diese Trainingsdaten muss typischerweise das korrekte oder erwartete Ergebnis bekannt sein.

Wenn die Aufgabe die Klassifikation von Bildern ist, muss also zu jedem Bild aus dem Trainingsdatensatz die korrekte Klasse (zum Beispiel Hund oder Katze) mitgegeben werden. Man spricht in diesem Fall von überwachtem Lernen. Es wird versucht, nach diversen Verfahren die Regeln für eine Klassifikation (abgebildet in einem „Modell“) zu erraten. Diese Modelle sind aber typischerweise für Menschen nicht direkt lesbar oder nachvollziehbar, nur deren Ergebnisse (zum Beispiel Produktempfehlungen für neue Kunden) können gesichtet werden. Man spricht dann von Black-Box-Verfahren.

Neuronale Netze sind eine bestimmte Sorte von Black-Box-Verfahren des maschinellen Lernens, die in den letzten Jahren vor allem bei der Analyse von Bildern und neuerdings auch von Text und Sprache wichtige Durchbrüche hatten. Heutzutage lassen sich mit relativ wenig Aufwand und hoher Effizienz zum Beispiel Bilder mit neu-

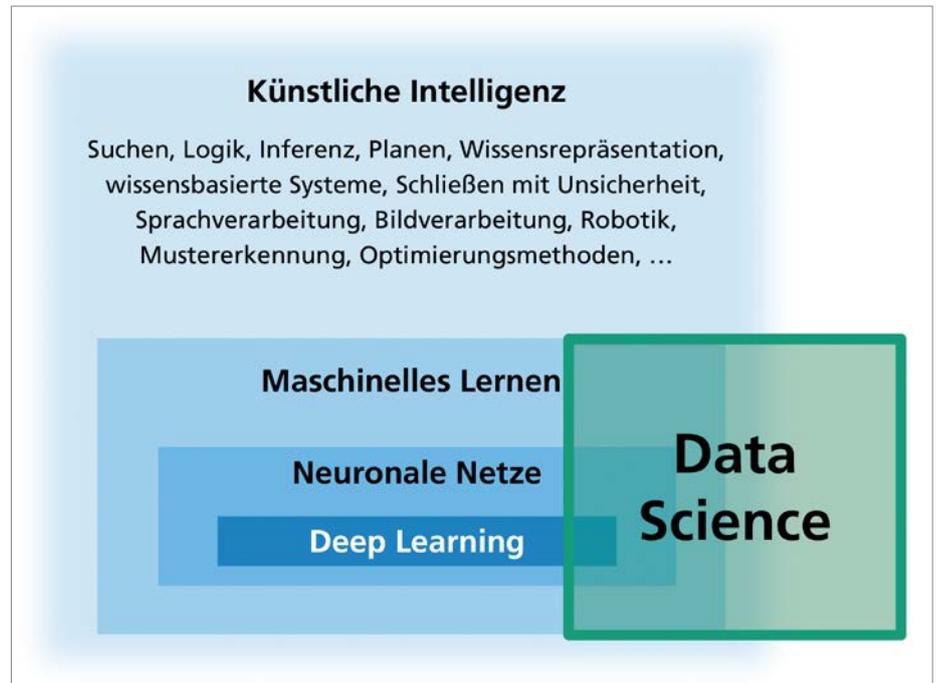


Abbildung 1: Einordnung der Begriffe Künstliche Intelligenz, maschinelles Lernen, neuronale Netze, Deep Learning und Data Science (Quelle: Fraunhofer IAO)

ronalen Netzen klassifizieren – obwohl dieselbe Aufgabe mit regelbasierten Verfahren besonders schwierig ist. Wenn man von **Deep Learning** spricht, redet man vom

Trainieren von besonders „großen“ oder „komplexen“ (man sagt gerne auch „tiefen“) neuronalen Netzen, wobei in den meisten Artikeln zum Thema nicht genauer spezifiziert ist, wie sich diese Tiefe definiert.

Und **Data Science**? Der Begriff wurde ursprünglich in den 1960er Jahren als Synonym für Informatik (Computer Science) verwendet. Ende der 1990er Jahre wurde er nach einem Vorschlag von Chikio Hayashi [2] in seiner noch heutigen Definition genutzt: eine Form von Datenanalysen (Data Mining) und Statistik-Projekten, die sich nicht nur mit den Analyse-Methoden, sondern auch mit der Nutzung der Ergebnisse befasst. Projekte des maschinellen Lernens sind immer auch Data-Science-Projekte. Ein Data-Science-Projekt muss aber nicht unbedingt ein KI- oder ML-Projekt sein.

In den 1990er Jahren wurde ein Standard-Prozessmodell für Data-Science-Projekte entwickelt, das immer noch sehr gut zu ML-Projekten passt. Der sogenannte **CRISP-DM** (Cross Industry Standard Process for Data Mining) besteht aus sechs Phasen und wird als Kreis dargestellt (siehe Abbildung 2): Mit jeder Durchführung wird etwas Neues erlernt, das in einer weiteren Projektdurchführung angewendet werden kann oder soll. Nach CRISP-DM fängt jedes Projekt mit *Business Understanding* an. Hier geht es darum, ein Verständnis für das zu lösende Problem oder die zu beantwortende Frage zu schaffen. Ohne konkretes Ziel braucht man auch kein Data-

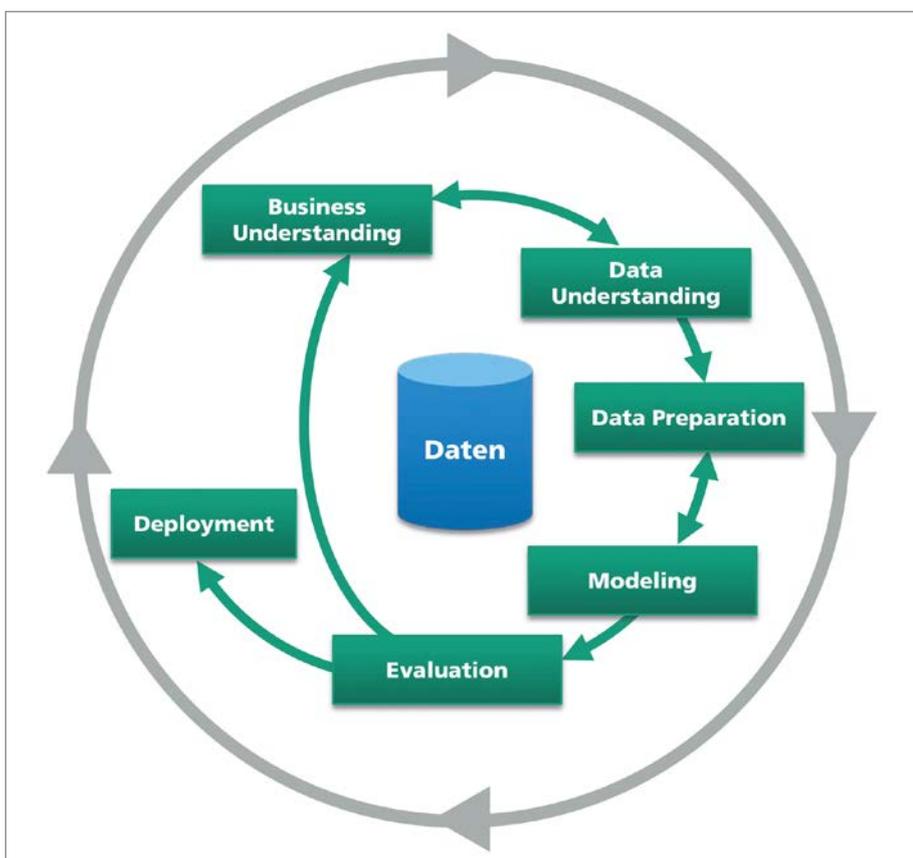


Abbildung 2: Das CRISP-DM-Modell (Quelle: Eigene Darstellung nach C. Shearer [3])

Science-Projekt! In einer zweiten Phase, *Data Understanding*, werden die benötigten und gegebenenfalls verfügbaren Daten gesichtet und mit Verfahren der Statistik analysiert. Hier sollen Experten evaluieren, ob diese sich für die Beantwortung der Fragen aus *Business Understanding* eignen oder ob zuerst weitere Daten gesammelt werden sollen. Falls die Daten geeignet sind, geht es weiter mit *Data Preparation*. Das ist typischerweise die zeitintensivste Aufgabe, die bis zu 80% des Aufwands eines Data-Science- oder ML-Projekts in Anspruch nimmt. Hier werden die zur Verfügung stehenden Daten bereinigt und qualifiziert, oft mit hohem manuellen Aufwand.

Der nächste Schritt ist für Data Scientists der spannendste: Im *Modeling* werden die passenden Algorithmen ausgewählt, deren Parameter justiert und die Modelle an sich berechnet. Deren Güte muss dann begutachtet werden, das passiert im Schritt *Evaluation*. Aus dieser *Evaluation* lernt man normalerweise einiges über die zu beantwortende Frage – unabhängig davon, wie gut die Modelle doch waren. Deshalb kann es hier weiter mit neuem Business Understanding gehen. Wenn die Modelle nach der Evaluation als gut (oder gut genug) bewertet werden, können sie eingesetzt werden, das ist das *Deployment*.

Grenzen

Obwohl sich viele (und jeden Tag ein bisschen mehr) Aufgaben heutzutage sehr gut mit KI und ML lösen lassen (Analyse von Texten, Vorhersagen, Klassifikation von Dokumenten, Erkennung von Auffälligkeiten wie Betrugsfällen und so weiter), darf man nicht vergessen, dass diese Ansätze keine magischen Werkzeuge sind, die perfekt alle Aufgaben auf Anhieb lösen können. Im Folgenden werden wichtige Grenzen von KI- und ML-Verfahren benannt:

- Viele Modelle, wie tiefe neuronale Netze, sind für Menschen nicht direkt verständlich („Black Box“). Das bedeutet, dass ein Anwender nicht ohne Weiteres erkennen kann, warum eine KI auf einem Bild eine Katze erkennt (oder nicht) oder warum ein Schadensfall möglicherweise als Betrugsversuch klassifiziert wird oder nicht.
- Eine KI hat weder Bewusstsein noch ein Verständnis der Aufgaben, die sie erledigt. Die Software manipuliert nur Zahlen, ohne über mögliche Konsequenzen für die Menschen nachzudenken. Damit dies nicht in die falsche Richtung geht, müssen KI-Experten mit Sorgfalt die Modelle trainieren und evaluieren.

- Das Trainieren von maschinellen Lernmodellen benötigt ausreichend viele und gute Daten. Die Modelle können nicht ohne Weiteres schlauer als ihre Datenbasis werden. Dies kann zu schlechten Ergebnissen führen, weil zum Beispiel in den Trainingsdaten nicht korrekt hinterlegt ist, welche Fälle Betrugsversuche waren und welche nicht. Oder weil wichtige Attribute oder Aspekte der Daten vergessen wurden und andere nicht, obwohl man diese hätte löschen sollen (man denke hier an berühmte Beispiele von KI-Modellen, die etwa auf Basis von ethnischer Herkunft Entscheidungen getroffen haben, obwohl dieses Kriterium in der Praxis nicht genutzt werden sollte).

Lösungsansätze und aktuelle Forschungsthemen

Die KI-Community arbeitet natürlich daran, diese bekannten Schwächen und Grenzen der Verfahren zu minimieren oder damit umzugehen.

Die Blackbox-Problematik wird mit Ansätzen der erklärbaren KI (**XAI**) angegangen. So versucht man zum Beispiel, mit Ansätzen wie LIME, LRP oder SHAP die Zonen eines Bilds, die einen großen Einfluss auf die Klassifizierung in einer bestimmten Kategorie (zum Beispiel als Hund oder Katze) haben, grafisch zu markieren. Analog können auch die Wörter oder Sätze in einem Text hervorgehoben werden, die besonders zu dessen Klassifikation beitragen (zum Beispiel welche Teile einer E-Mail besonders darauf hinweisen, dass es sich um eine Spam-Mail handelt).

Die Komplexität der Justierung der Modelle wird mit dem sogenannten automatisierten maschinellen Lernen (AutoML) re-

duziert. **AutoML** versucht von sich selbst aus, durch viele Versuche und Kombinationen, automatisch für bestimmte Daten die besten Modelle zu erzeugen. Da die Anzahl an möglichen Parameterkombinationen für das maschinelle Lernen enorm ist, lässt sich diese Aufgabe nicht einfach mit Durchprobieren lösen – die in der Praxis dafür aufgewendete Zeit ist nicht vertretbar.

Um den Bedarf an großen Mengen von qualitativ hochwertigen Trainingsdaten zu reduzieren, werden mehrere Lösungswege untersucht und erprobt. Mit dem halbüberwachten Lernen oder mit dem aktiven Lernen müssen zwar viele Datensätzen verfügbar sein, aber nur wenige davon müssen annotiert, das heißt von Hand markiert oder klassifiziert werden. Beim **halb-überwachten** Lernen sind die Algorithmen dann in der Lage, bei den unmarkierten Datensätzen die richtige Klassifikation oder das richtige Label zu erraten (indem einige Annahmen über die Datensätze getroffen werden, zum Beispiel „ähnliche Datensätze müssen auch ähnliche Labels haben“). Beim **aktiven Lernen** werden Experten während der Lernphase – also während die Modelle berechnet werden – ab und zu von den Modellen explizit gefragt, einige wenige zusätzliche Trainingsdaten zu markieren. Die Algorithmen sind in der Lage zu erkennen, für welche Datenpunkte aus dem Trainingsdatensatz es besonders wichtig ist, diese Information zu haben. Parallel dazu wird auch mit synthetisch erzeugten Trainingsdaten gearbeitet. Diese werden also von Software-Programmen generiert, die versuchen, die Komplexität der realen Welt abzubilden. Die Erfahrung zeigt, dass dieser Ansatz vielversprechend ist, aber auch hier in Kombination

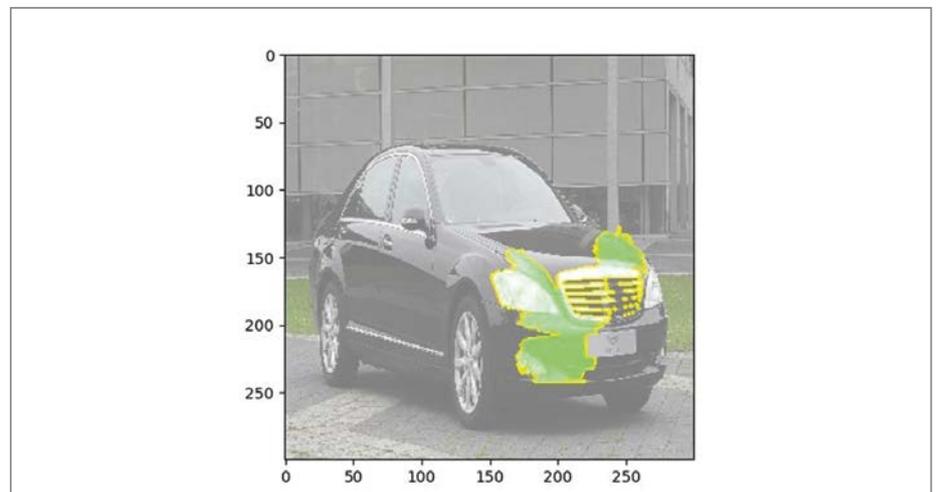


Abbildung 3: Woran erkennt eine KI, dass das Auto auf dem Bild ein Mercedes-Benz-Modell ist? (Quelle: Fraunhofer-IAO-Prototyp unter Verwendung der Software-Bibliothek Lime [4])

mit der Nutzung einer kleinen Menge an echten Datensätzen am besten funktioniert.

Fazit

Mit Data Science, KI und maschinellem Lernen lassen sich Aufgaben erledigen, die mit klassischer Programmierung sehr schwer oder unmöglich wären. Jedoch darf man keine Magie von diesen Ansätzen erwarten, und deren Grenzen hängen sehr von der Quantität und Qualität der zur Verfügung stehenden Daten ab. Aktuelle, neue Entwicklungen wie XAI, halb-überwachtes und aktives Lernen oder AutoML sollen helfen, Grenzen der Verfahren (unter anderem schlechte Nachvollziehbarkeit, Aufwand beim Aufsetzen der Trainingsdaten oder Justierung der Modelle) in naher Zukunft zu verringern. Aktuelle Software-Bibliotheken existieren bereits und werden derzeit immer häufiger in KI-Projekten eingesetzt.

Quellen

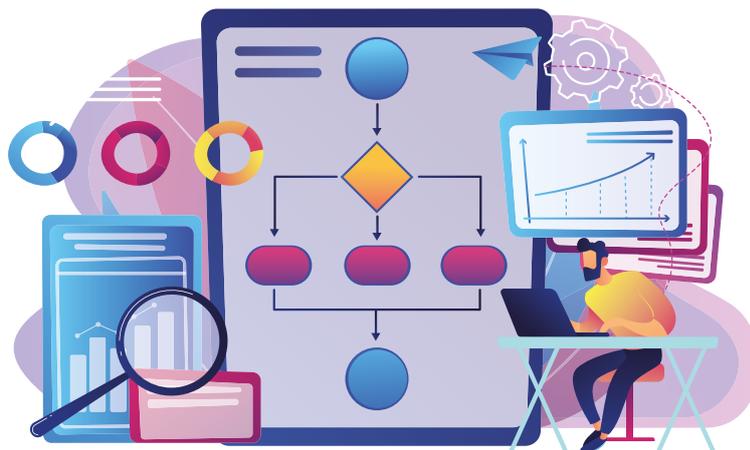
- [1] John McCarthy, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon, A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence (1955), <https://web.archive.org/web/20080930164306http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>
- [2] Hayashi C. (1998) What is Data Science? Fundamental Concepts and a Heuristic Example. In: Hayashi C., Yajima K., Bock HH., Ohsumi N., Tanaka Y., Baba Y. (eds) Data Science, Classification, and Related Methods. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Tokyo. https://link.springer.com/chapter/10.1007/978-4-431-65950-1_3
- [3] Shearer C., The CRISP-DM model: the new blueprint for data mining, J Data Warehousing (2000); 5:13—22
- [4] <https://github.com/marcotcr/lime>



Maximilien Kintz

maximilien.kintz@iao.fraunhofer.de

Dr.-Ing. Maximilien Kintz hat in Nantes (Frankreich) und Stuttgart Ingenieurwissenschaften und Informatik studiert. Er forscht seit 2009 am IAT der Universität Stuttgart und am Fraunhofer IAO, wo er das Team „Angewandte Künstliche Intelligenz“ leitet. Im Team werden Lösungen untersucht und entwickelt, die Unternehmen und Organisationen bei der Optimierung und Automatisierung von Arbeitsabläufen unterstützen. Schwerpunkte der Arbeit von Herrn Kintz sind die Analyse von Textdokumenten, Dashboards und die Überwachung von Prozessen.



Process Mining: Daten für ganzheitliches Geschäftsprozessmanagement

Janna Meyer, MEHRWERK & Claus Engel, PROMATIS software

Prozesse sind die Maschinerie eines jeden Unternehmens. Eine gut geölte und gewartete Maschine ist die Grundlage für hochwertigen Output. Wartet man die Maschine rechtzeitig, bleibt der Output hochwertig, wartet man sie zu spät, sind die Konsequenzen zu tragen. Den richtigen Zeitpunkt und das richtige Maschinenteil für die Wartung zu finden, ist der Schlüssel zum Erfolg. Dasselbe lässt sich über Prozesse sagen. Prozesse, die heute gut funktionieren, können morgen veraltet sein, zu unflexibel oder zu kostspielig. Prozesse, die sich heute schon durch unnötig hohe Komplexität auszeichnen, werden in der Zukunft zum Bremsklotz des Unternehmens. Sie blockieren Wachstumspotenziale und führen zu Ungemach bei allen Beteiligten. Um Maschinen in Schuss zu halten, setzen Unternehmen heute auf Predictive Maintenance. Warum sollte man für Prozesse nicht auf etwas Ähnliches setzen? Gibt es vielleicht so etwas wie Predictive Maintenance für Prozesse? Nun, vielleicht gibt es den Begriff nicht, aber es gibt die Technologie – Process Mining.

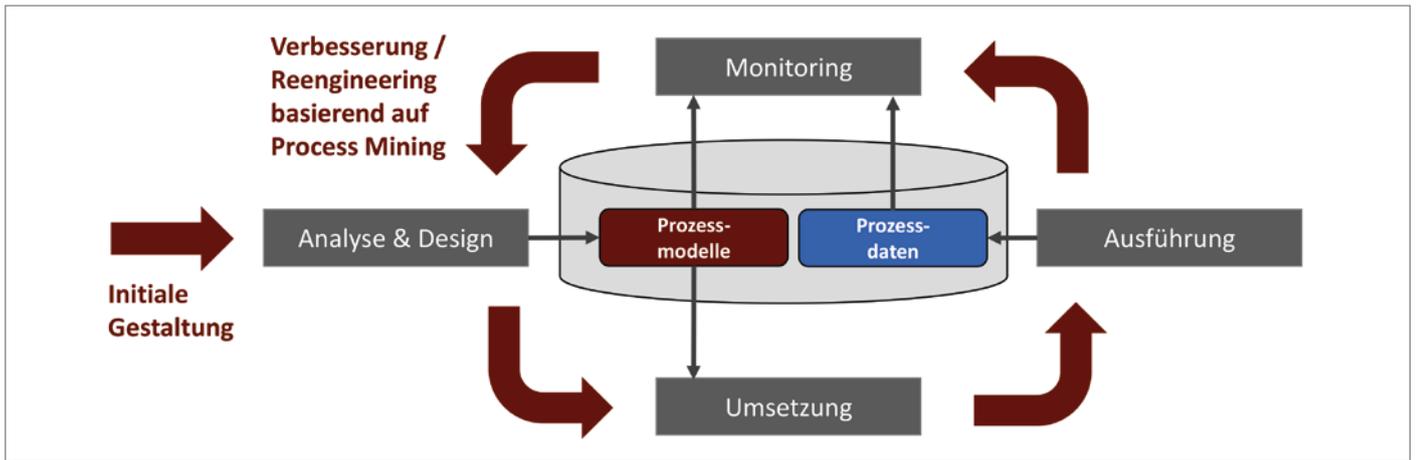


Abbildung 1: Geschäftsprozessmanagementzyklus, unterstützt durch Process Mining [vgl. 2, 3]

Warum Prozessdaten nützlich sind

Lassen sich Schwierigkeiten in Prozessen feststellen, werden die internen Beteiligten oder gar externe Berater, die klassischen Interviewmethoden im Gepäck, mit Zettel und Stift bewaffnet durch die Firma geschickt, um das Ausmaß der Probleme zu manifestieren. Es folgen extensive Prozessbeobachtungen, manuelle Dokumentationen, Diskussionsrunden und Ergebnisinterpretationen [1 S.7]. Die entstandenen Lösungsvorschläge können von hoher Relevanz sein – die richtige Maßnahme, die den Kern des Problems effektiv und effizient behebt. Oft sind sie aber auch eine Reihe von Flickerlösungen zur Symptombehandlung – Prozessoptimierung mit dem Gießkannenprinzip. Was beide Ergebnisse gemein haben, ist Folgendes:

1. Sie sind, trotz aller Anstrengung nach Objektivität, **subjektiv**, da die Problemanalyse auf Erfahrung statt auf Daten basiert.

2. Die zugrunde liegende Analyse ist **nicht beliebig wiederholbar**, da der Faktor Mensch in jedem Interview anders reagieren wird und die resultierenden Aufwände nicht vertretbar sind.
3. Die Subjektivität und fehlende Wiederholbarkeit machen die Lösungsvorschläge **angreifbar**.
4. Die Informationsbeschaffung ist oft **aufwendiger**, als die eigentlichen Verbesserungsmaßnahmen abzuleiten.

Doch es gibt eine Technologie, durch die das Geschäftsprozessmanagement sich dieser Nachteile entledigen kann. Eine Technologie, die auf Daten zurückgreift, und zwar genau auf die Daten, die die Prozesse selbst in den IT-Systemen des Unternehmens hinterlassen haben. Diese Technologie nennt sich Process Mining. Wie Process Mining den Geschäftsprozessmanagementzyklus schließt, dadurch aus einer reaktiven Unter-

nehmenskultur eine aktive wird und sich in Folge Process Literacy (Prozesskompetenz) in den Organisationen ausbreitet, wird in diesem Artikel geschildert.

Die Datenlücke im Geschäftsprozessmanagementzyklus

Geschäftsprozessmanagement wird in der Regel zyklisch durchgeführt (siehe Abbildung 1). In der Analyse- und Designphase werden auf Basis von Anforderungen dazu passende Prozessmodelle entworfen. Die Prozesse der entstandenen Modelle werden durch Konfiguration und Implementierung anschließend systemtechnisch umgesetzt. Sobald die IT-Systeme die neuen Prozesse unterstützen, beginnt die Ausführungsphase. Im Rahmen des Monitorings stehen die Prozesse unter Beobachtung, um Änderungsbedarfe zu erkennen. Auf Basis der Erkenntnisse aus dem Monitoring können sich kleine Verbesserungen oder auch um-

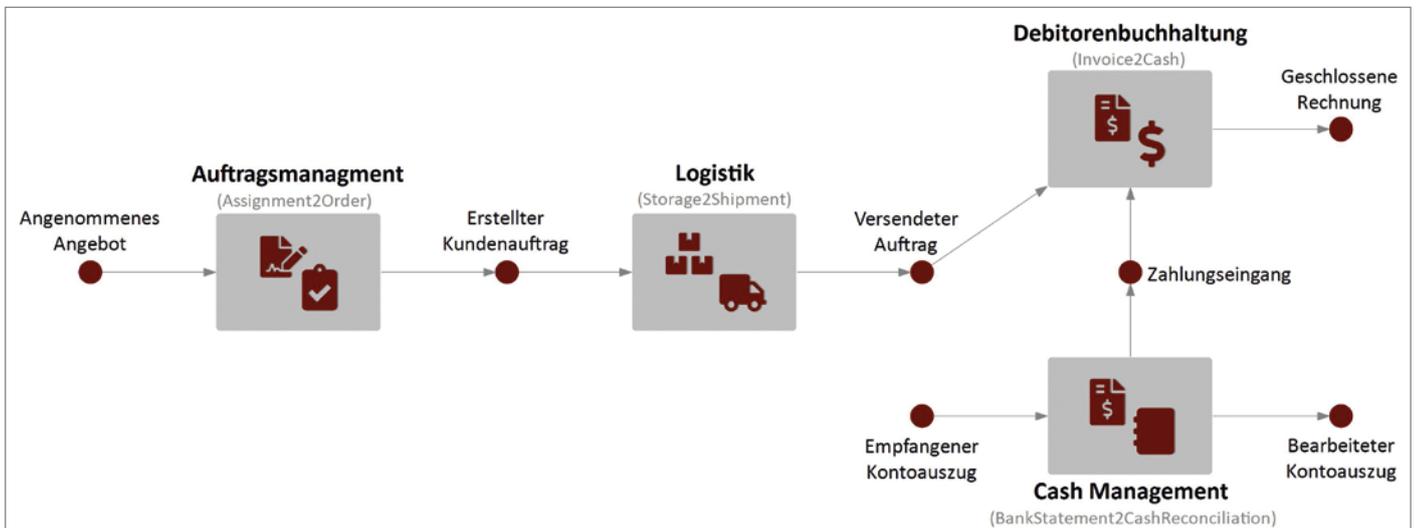


Abbildung 2: Order2Cash-Prozess [vgl. 1, 3]

fassende Reengineering-Anforderungen an den Prozess ergeben, die dann den Lebenszyklus wieder von vorne anstoßen. Prozessmodelle stellen die zentrale Basis für die Umsetzung von Geschäftsprozessen durch Konfiguration von Standardsoftware und/oder der Implementierung kundenspezifischer Zusatzfunktionen dar. Sie werden zum Verständnis, als Diskussionsgrundlage, zu Dokumentationszwecken, zur Verifikation, Performanceanalyse etc. eingesetzt [4]. Bis vor Kurzem gab es laut [4] nur wenig Verknüpfung zwischen dem eigentlichen Prozessentwurf und den Daten der Prozessausführung, was dazu führt, dass in den meisten Organisationen weder die Ausführungs- noch die Monitoring-Phase inklusive der erforderlichen Diagnose systematisch und kontinuierlich unterstützt werden. Dies hat zur Folge,

1. dass nur schwerwiegende Probleme oder größere äußere Änderungen eine neue Iteration des BPM-Lebenszyklus anstoßen und
2. dass Informationen aus den vergangenen und laufenden Prozessen nicht aktiv in die Umgestaltung der Modelle eingeschlossen werden.

Process Mining ermöglicht, diese Lücke im BPM-Lebenszyklus wirklich zu schließen (siehe *Abbildung 1*) und damit in eine kontinuierliche Verbesserung einzutreten. Dabei sind die Analysen stets datenbasiert und objektiv sowie nach einmaliger Implementierung zu beliebigen Zeitpunkten ohne großen Aufwand wiederholbar.

Was also ist Process Mining?

Process Mining ist eine Technologie, die einen datengetriebenen Blick auf Prozesse im Allgemeinen und Geschäftsprozesse im Speziellen schafft. Sie ermöglicht vollständige Transparenz der Prozessrealität auf Basis von Ereignisdaten aus IT-Systemen. Diese Transparenz gilt in Fachkreisen als Voraussetzung für die digitale Transformation [1 S.2]. Sie erlaubt das Ableiten und Anpassen von Prozessmodellen, die Durchführung von Abweichungs- und Fehlerursachenanalysen sowie eine kontinuierliche Prozessüberwachung bis hin zur operativen Prozessunterstützung.

Anwendungsfall Order-to-Cash-Prozess

Als konkreter Anwendungsfall wird in diesem Beitrag der Einsatz von Process-Mi-

ning-basiertem Geschäftsprozessmanagement anhand des Order-to-Cash (O2C) auf Basis einer Oracle E-Business Suite beschrieben. *Abbildung 2* zeigt den Ablauf eines typischen O2C-Prozesses [2]. Prozessmodelle werden idealerweise hierarchisch aufgebaut, d.h. jeweils relevante Aspekte werden auf verschiedenen Abstraktionsebenen dargestellt [2,5]. Auf einer hohen Abstraktionsebene besteht der O2C-Prozess aus folgenden Schritten: Der Prozess startet mit einer Beauftragung des Kunden in Form eines angenommenen Angebots. Im Auftragsmanagement wird dann ein Auftrag im System erstellt. Beim darauffolgenden Teilprozess Logistik wird die Lieferung entsprechender Produkte im Lager zusammengestellt und der Versand an den Kunden in die Wege geleitet. Auf Basis der Lieferung wird die Fakturierung durch die Debitorenbuchhaltung angestoßen. Die Zahlungen des Kunden werden über das Einlesen von Kontoauszügen im Cash Management im System erzeugt. Durch die Zuordnung korrespondierender Zahlungen zu Rechnungen werden diese dann am Ende des Prozesses geschlossen. Stornierungen geschehen in den Detailprozessen und führen durch die Erstellung von Gutschriften ebenfalls zu geschlossenen Rechnungen.

Was Daten über Prozesse wissen – Process Discovery

Gute Prozessmodelle zu erstellen, ist mehr Wissenschaft denn Kunst [4]. Handgemachte Modelle tendieren neben starker Vereinfachung aus Verständnisgründen zur Subjektivität [4]. Selbst wenn Modelle 80% der als repräsentativ angenommenen Fälle abdecken, zeigt sich typischerweise, dass die 20% der nicht abgedeckten Fälle 80% der Probleme verursachen. Diese 20% nicht modellierter Prozessfälle lassen sich nur durch einen datengetriebenen Ansatz wie Process Mining effizient erfassen und darstellen. Zu sehen ist dies im „Spaghettidiagramm“ in *Abbildung 3*, das die Prozessrealität zum Prozessmodell in *Abbildung 2* aufzeigt. An dieser Stelle sei nochmal darauf hingewiesen, dass das gezeigte Diagramm automatisiert auf Basis der echten Daten erzeugt wurde. Dieses „Process Discovery“, das „Entdecken der Prozesse“, ermöglicht, die gesamte Komplexität der Unternehmensprozesse zu erfassen und jede Prozessschleife, jede Abweichung und jeden Sonderprozess nicht nur zu erkennen, sondern auch einem

Prozesskontext zuzuordnen, beispielsweise bestimmten Unternehmensorganisationen, Kunden, Materialien oder Ländern. So zeigt ein erster Blick in eine Process-Mining-Applikation wie MEHRWERK ProcessMining, dass für den Beispielprozess, Order-to-Cash (O2C), aus einer Oracle E-Business Suite, das vorher verdeutlichte Prozessmodell eine starke Vereinfachung der Realität darstellt. In Wirklichkeit finden sich viele unerwartete Prozessvarianten, wie in *Abbildung 3* deutlich wird. Dabei ist es wichtig, die Anzahl der Prozessvarianten zu ihrer Häufigkeit in Vergleich zu setzen: Im O2C-Beispiel wickeln die häufigsten 10 Prozessvarianten immerhin 48% der Kundenaufträge ab. Solche ersten Aussagen ermöglichen das reale und schnelle Einschätzen der Prozessstandardisierung und geben kurzfristige Ziele für die Prozessoptimierung her. So lassen sich beispielsweise recht einfach neue und messbare Prozesszielvorgaben festlegen, wie z.B. „im nächsten Jahr decken unsere Top-10-Prozesse 60% statt heute 48% unserer Kundenaufträge ab“. Spannend ist auch die zeitliche Perspektive der Prozesse. Werden Prozesse innerhalb gewisser Zeitvorgaben abgewickelt? Falls nicht, wo entstehen Zeitverzögerungen? Solche Fragen sind schnell beantwortet, wenn man einen Blick auf die Prozessgrafik mit entsprechenden Prozessdurchlaufzeiten wirft – so wie in *Abbildung 5*. Die angezeigten Cases (im Process Mining der Begriff für eine Prozessinstanz, hier ein einzelner Kundenauftrag und dessen Prozessabfolge) haben alle eine Durchlaufzeit von der Erstellung des Kundenauftrags bis zur Aktivität von mehr als 130 Tagen. Da dies im Order-to-Cash-Prozess ungewöhnlich ist und auf Problematiken wie verspätete Kundenzahlungen hinweist, lohnt sich der Blick auf den Prozessabschnitt „Invoice to Cash“ von der Rechnung bis zur Zahlung. Auf einen Blick werden so weitere Optimierungspotenziale gehoben, wie der strategische Ansatz, das Kundenzahlungsverhalten zu verbessern, sowie der operative Ansatz, das Oracle System und die Mitarbeiter in die Lage zu versetzen, die Eingangszahlung schneller zuzuordnen. Deutlich werden hier vor allem die verstrichenen Zahlungsfristen sowie lange Zeiten bis zur Zuordnung der Eingangszahlung.

Wichtig dabei ist, dass Process Mining nicht nur den Fluss des Prozesses objektiv und automatisch zugänglich macht, sondern gleichzeitig auch den Kontext des Prozesses mit einbezieht. So ist für die ver-

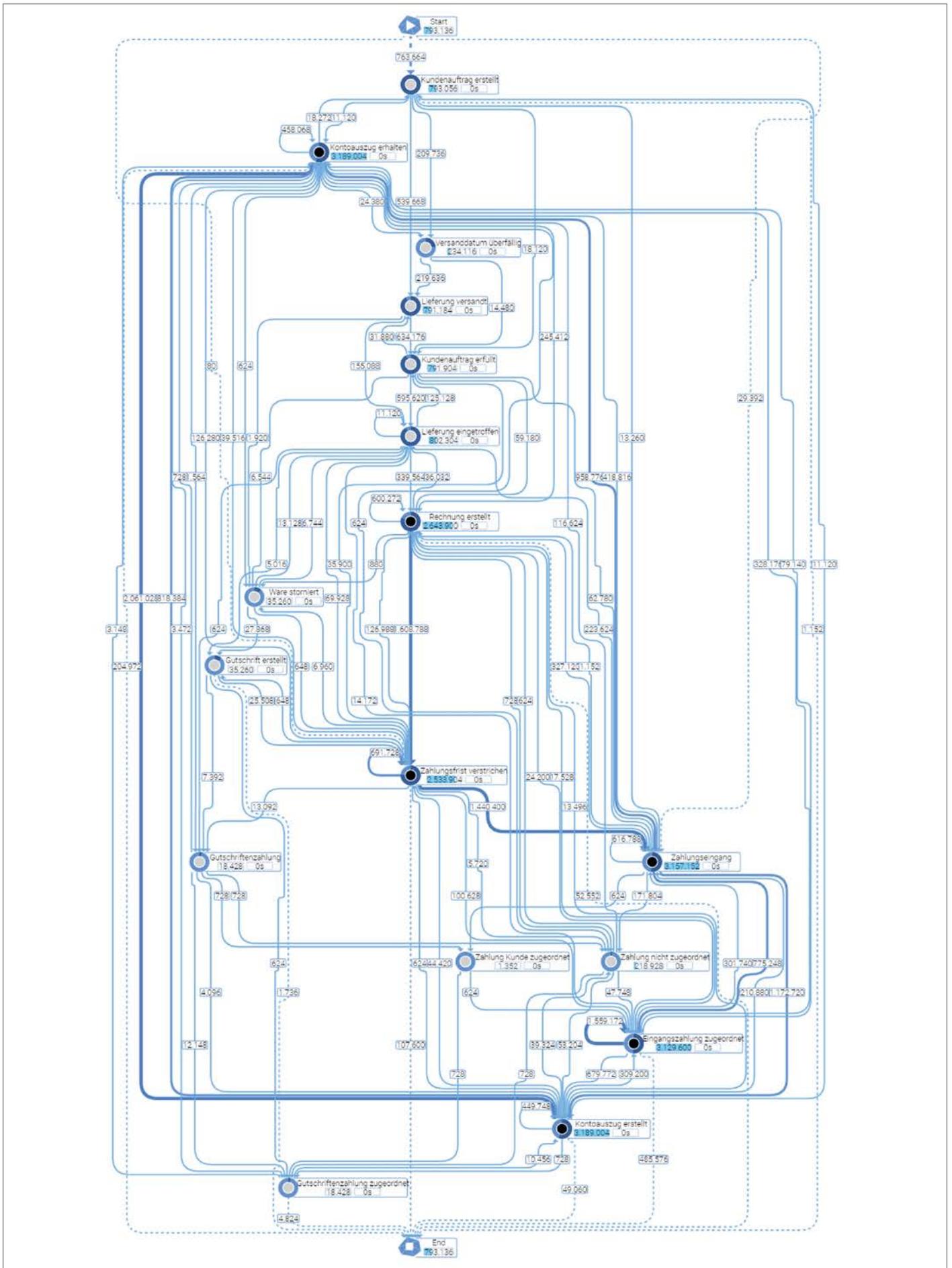


Abbildung 3: Alle Prozessvarianten des Beispiels O2C-Prozesses (© MEHRWERK GmbH)

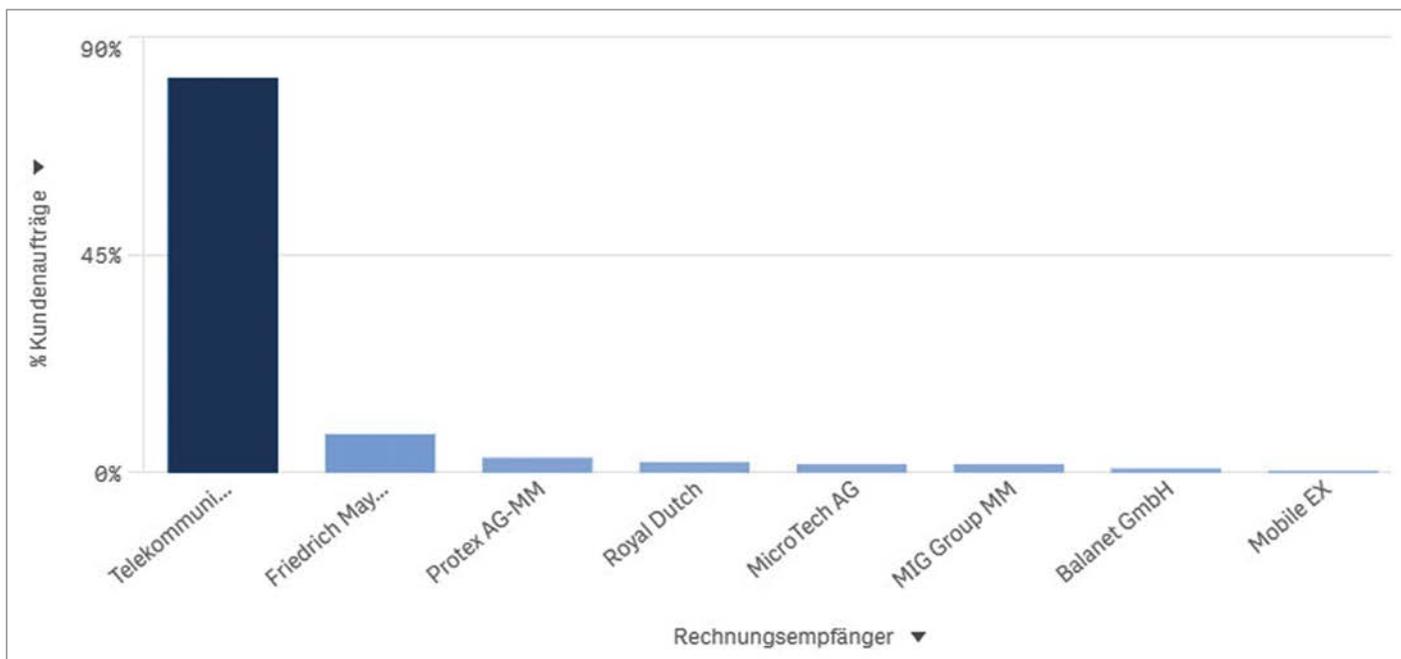


Abbildung 4: Kunden im Zusammenhang mit zu langen Prozessdurchlaufzeiten (> 130 Tagen) (© MEHRWERK GmbH)

späteten Kundenzahlungen die logisch folgende Frage nach den betroffenen Kunden schnell beantwortet, wenn man *Abbildung 4* zurate zieht. Die angezeigten Kunden stehen im Verhältnis zu den lang laufenden Prozessinstanzen und besonders ein Kunde sticht hervor, auf den die Bemühungen zur Verbesserung des Zahlungsverhaltens in einer ersten Optimierungsrunde beschränkt werden können. Eine solche Ursachenanalyse ist kennzeichnend für Process Mining und viele Tools unterstützen eine automatisierte „Root Cause Analysis“ inzwischen im Standard. Diese Ursachenanalyse und die neue Transparenz erlauben im Geschäftsprozessmanagement, die blinden Flecken auf der Prozesskarte durch Prozessdaten zu füllen, und unterstützen vor allem in den Phasen Prozessdesign und Prozessanforderungen des BPM-Lebenszyklus‘.

Modell vs. Realität – Abweichungen entdecken mit Conformance Checking

Ist die Prozesstransparenz einmal gegeben, entstehen Fragen wie „welche Abweichungen treten auf“, „welche kritischen Abweichungen sind häufig“ oder „welche Gründe gibt es für bestimmte Abweichungen“. Diese Fragen lassen sich durch „Conformance Checking“ beantworten, eine Teildisziplin des Process Mining, die prüft, ob die Prozessrealität aus den Ereignisdaten mit den erstellten Modellen konform ist, und die Abweichungen aufzeigt [6]. Diese Funkti-

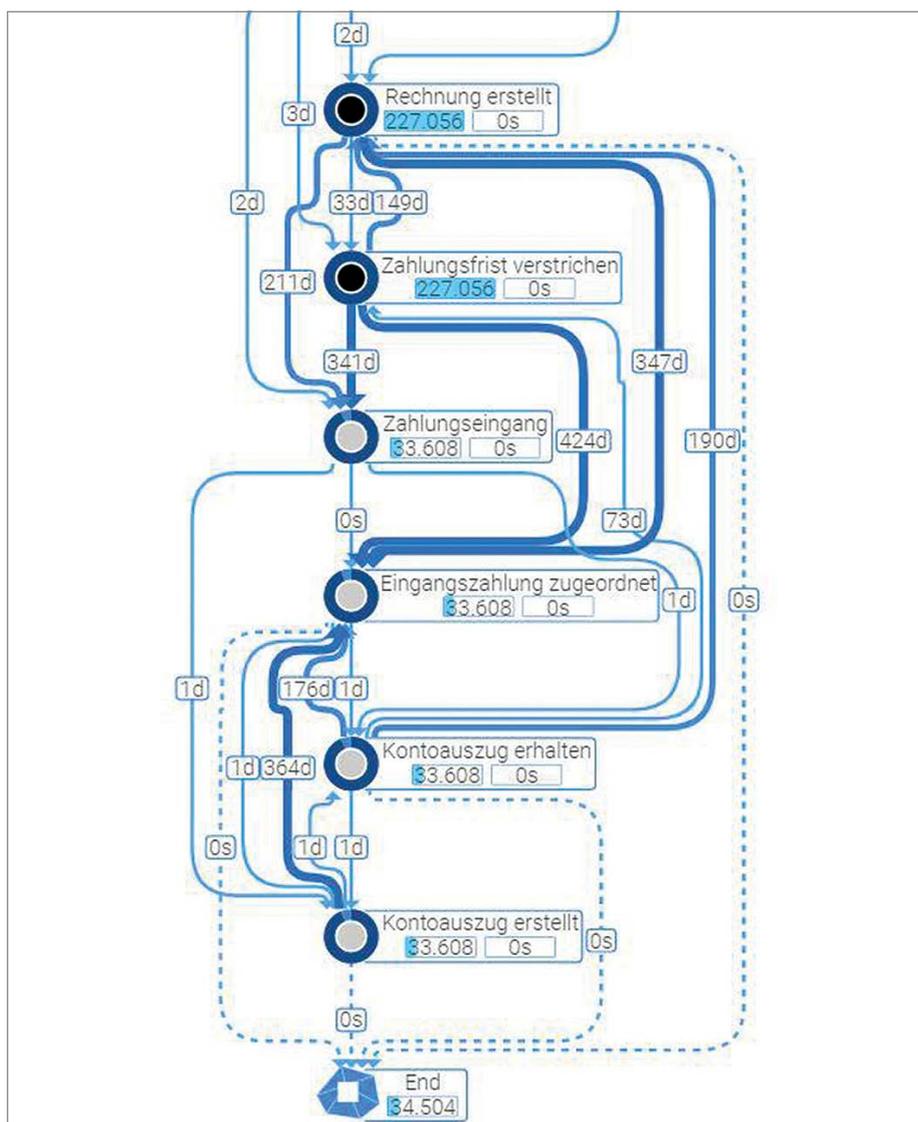


Abbildung 5: Prozessausschnitt - Zeitverzug im O2C-Prozess (© MEHRWERK GmbH)

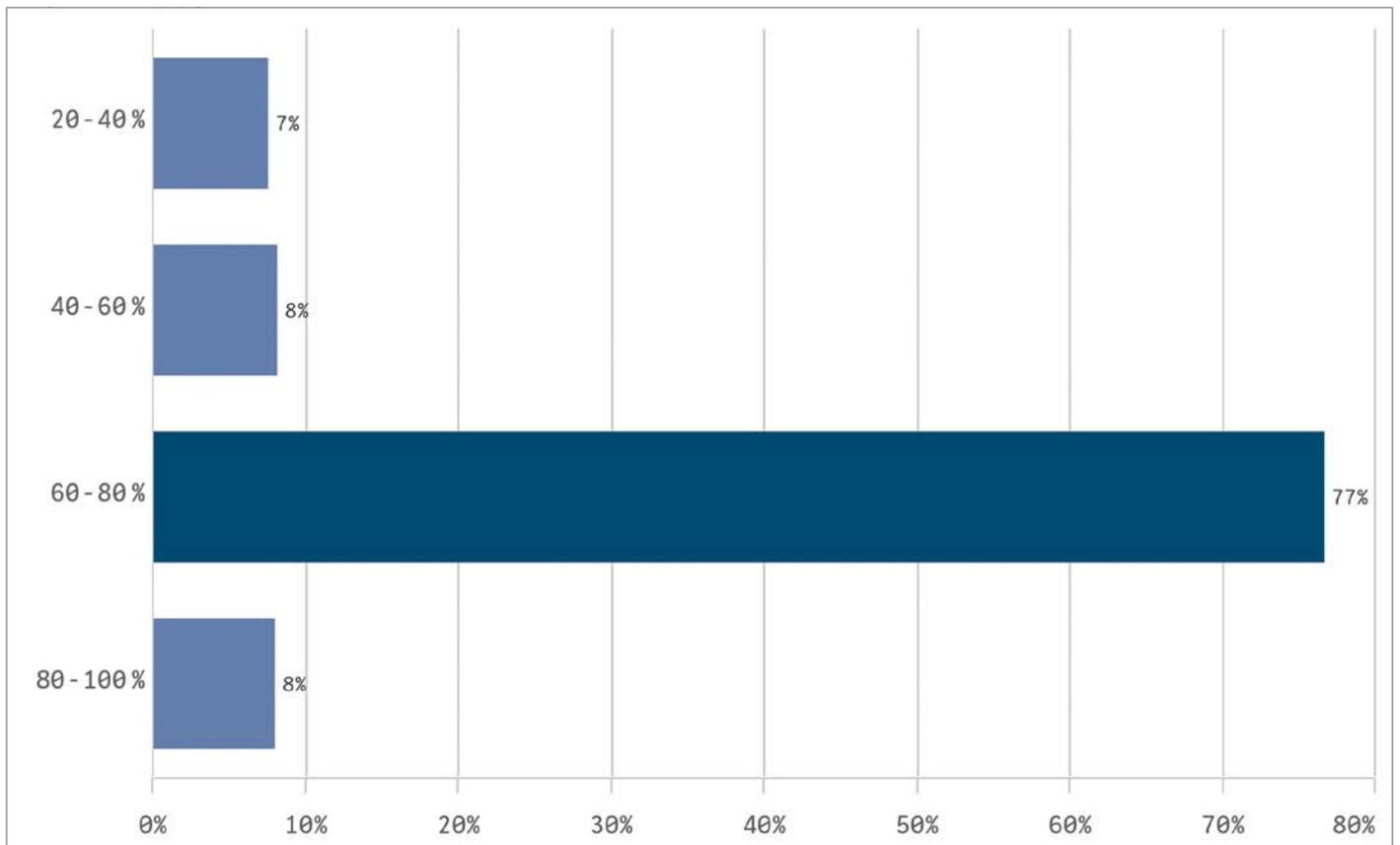


Abbildung 6: Verteilung der Happy Path Fitness im O2C-Beispiel (© MEHRWERK GmbH)

onalität ist im Geschäftsprozessalltag äußerst dienlich, um die Modellqualität und -Reife zu prüfen und zeitgleich die Ausführung der Prozesse im Blick zu halten. Im O2C-Beispiel beträgt die sogenannte Happy Path Fitness 74,26%, was bedeutet, dass die Prozessrealität dem Prozessmodell zu 74,26% entspricht. Die Happy Path Fitness ist dabei eine Metrik, die die Schwere der Abweichungen zwischen realen Prozesse vom Prozessmodell beziffert. In *Abbildung 6* wird für den Beispielprozess deutlich, dass nur 8% der Kundenaufträge eine Prozess Conformance von mehr als 80% erreichen. 77% erreichen immerhin 60-80%. Diese Information deutet auf signifikante Prozessabweichungen hin. So wird in *Abbildung 7* deutlich, dass für Warenstornierungen, auf die eine Gutschrift folgt, laut Prozessmodell eine Rechnung erstellt werden müsste, dies in der Realität aber nicht umgesetzt wird. Somit sind entweder die Mitarbeiter nicht korrekt für den Prozess geschult oder das Prozessmodell ist nicht mehr aktuell. Beides kann behoben werden und dient der Förderung der Prozesskompetenz, neudeutsch „Process Literacy“, im Unternehmen. Spannend ist auch ein Blick auf abweichende Aktivitäten, jene Aktivitäten bzw. Statusän-

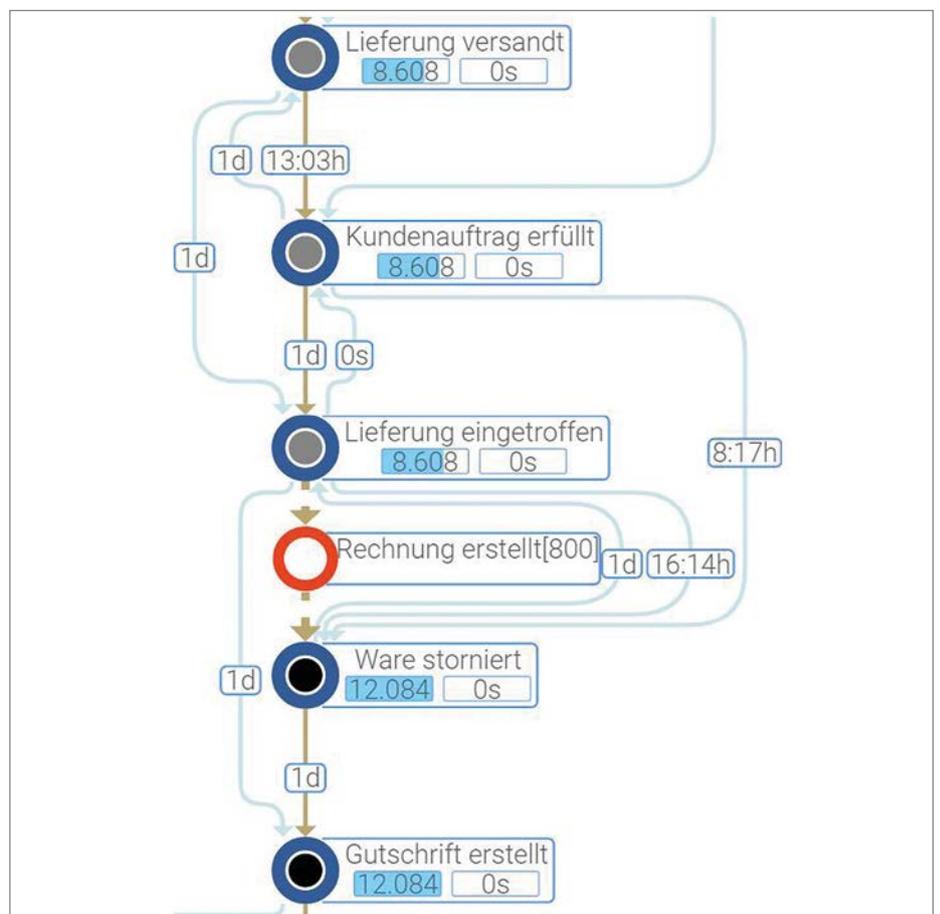


Abbildung 7: Abweichung zwischen Modell und Realität (© MEHRWERK GmbH)

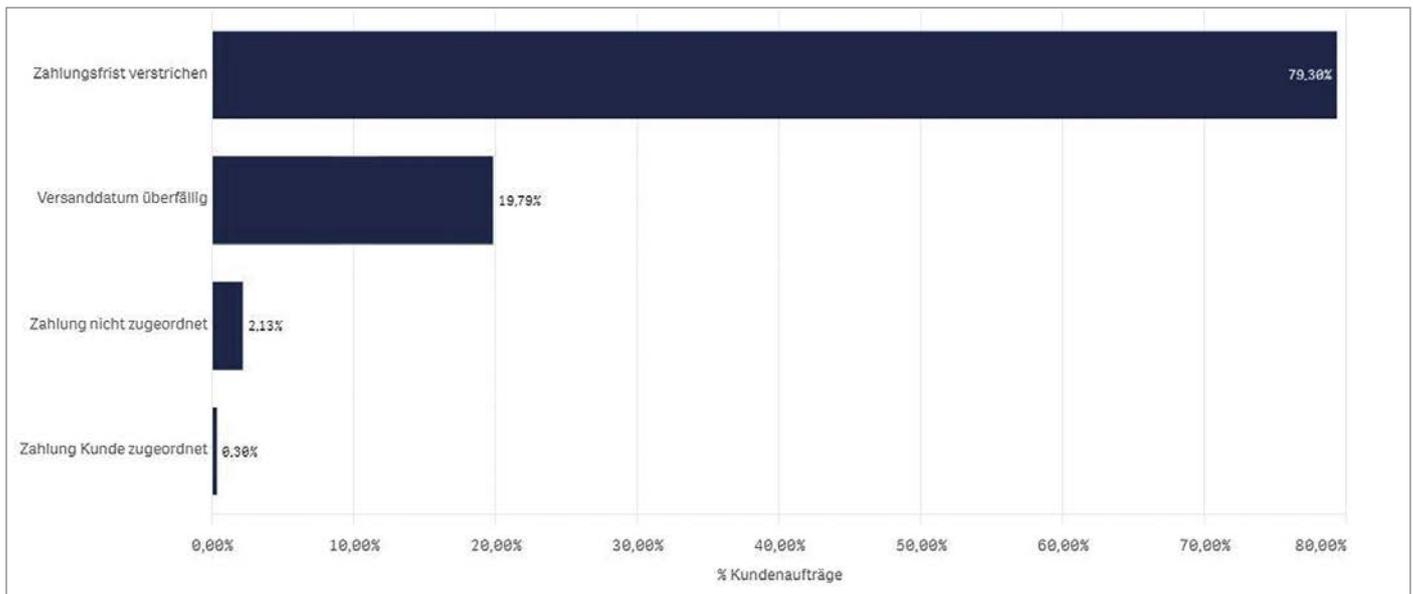


Abbildung 8: Die häufigsten Prozessabweichungen im O2C-Beispiel (© MEHRWERK GmbH)

derungen, die im Prozessmodell nicht erfasst werden, jedoch in der Realität einen hohen Einfluss haben. *Abbildung 8* zeigt für den O2C-Prozess zwei recht schmerzhafte Abweichungen, nämlich das häufige Verstreichen der Zahlungsfrist sowie das Überschreiten des Versanddatums. Beides ist höchst unerwünscht im Prozess, wird aber typischerweise in Unternehmen nicht in direkter Kombination mit Prozessmodellen erfasst und geprüft, da es sich nicht um „Aktionen“ in dem Sinne, sondern vielmehr um den Prozessstatus handelt. Oft bleiben solche Auswertungen isoliert im Unternehmensreporting stehen, obwohl sie im Rahmen eines Prozesses auftreten, der sie, wenn entsprechend „designed“, ebenso gut verhindern kann –entweder durch vorbeugende Workflows im IT-System oder durch frühzeitiges Gegensteuern im laufenden Prozess. Das Process Mining geht noch ein Stück weiter, verbindet die Prozess- mit der Datensicht und bietet in solchen Fällen an, die Problemursachen mittels automatisierter Ursachenanalyse direkt zu identifizieren und über integrierte Prozessüberwachung sowie integriertes Maßnahmenmanagement vorzuwarnen, wenn Fälle einen unerwünschten Prozessverlauf einschlagen.

Daten steuern Prozesse – Operational Support

Dazu werden offene Prozesse kontinuierlich überwacht und vorhergesagt, sodass potenziellen Schwierigkeiten wie beispielsweise in *Abbildung 9* frühzeitig entgegen gewirkt werden kann. Für den abgebildeten

Kundenauftrag wurde bereits eine Rechnung erstellt (blau), auf die mit hoher Wahrscheinlichkeit die roten Aktivitäten folgen. Da die Zahlungsfrist als wahrscheinliches Ereignis aufgeführt wird, lohnt es sich, dem betroffenen Kunden eine Erinnerung zu-

kommen zu lassen, bevor es überhaupt so weit kommt. Manche Process Mining Tools unterstützen dabei sogar, solche Maßnahmen direkt an entsprechende Mitarbeiter oder Bots zu übergeben. So verbessern sich auf Dauer wichtige Prozesskennzahlen, wie

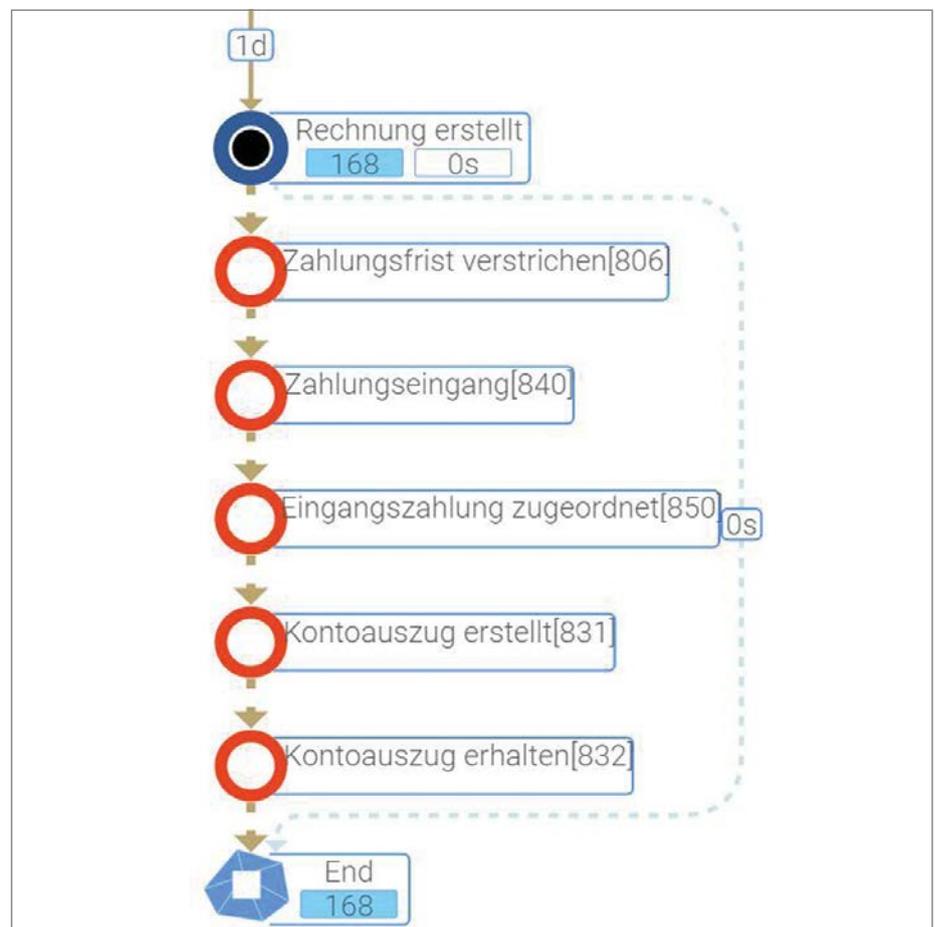


Abbildung 9: Prozessüberwachung und -vorhersage für gezielte präventive Gegenmaßnahmen (© MEHRWERK GmbH)

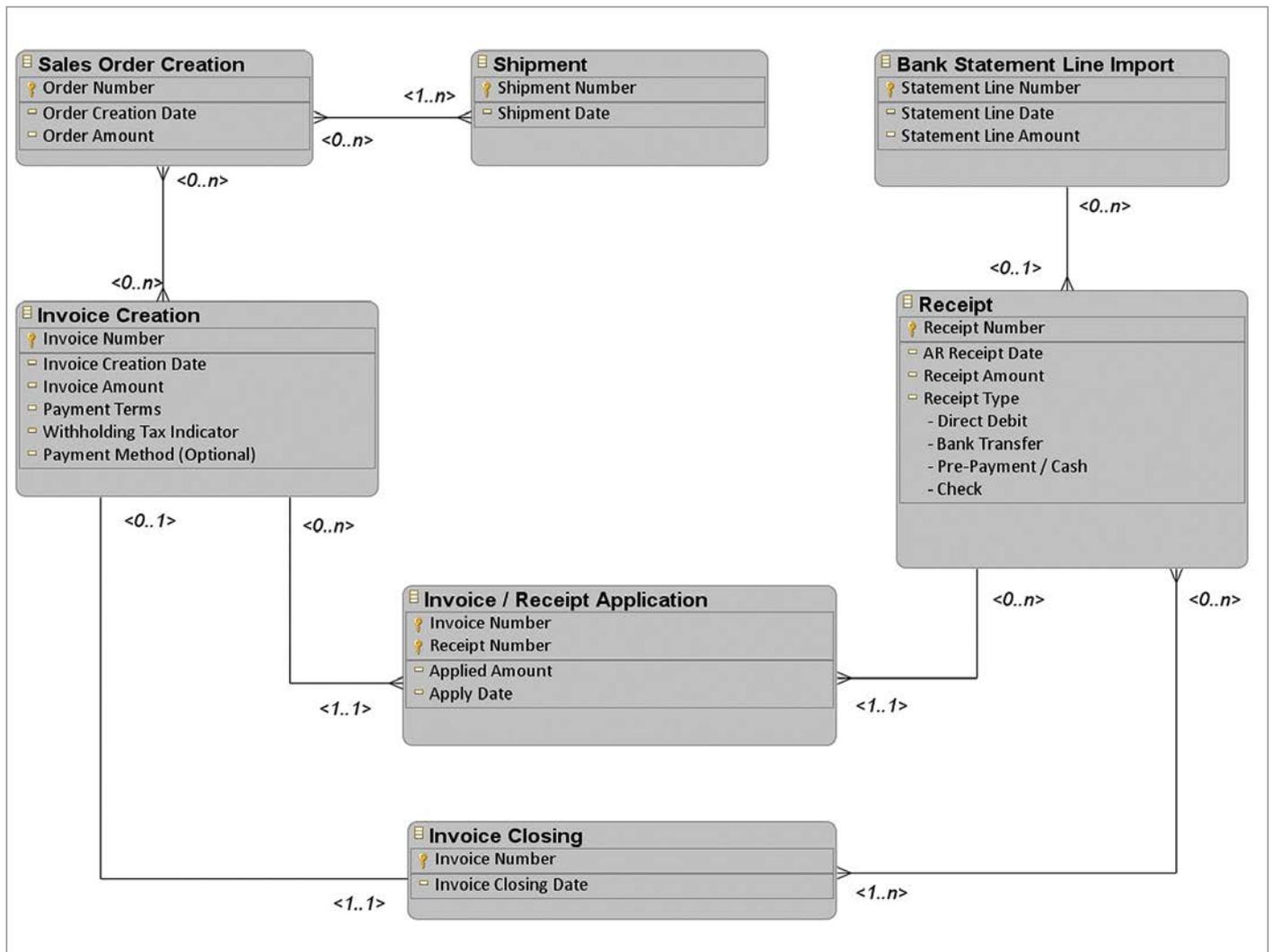


Abbildung 10: Datenmodell für O2C-prozessrelevante Ereignisse in einer Oracle EBS [vgl. Oracle eTRM]

im Order-to-Cash z.B. die Days Sales Outstanding.

Prozesse, Daten und Menschen an Bord

Oft betrachten Unternehmen ihre Prozesse im Geschäftsprozessmanagement als reine Kette von Aktivitäten. Aber Prozesse beinhalten mehr als nur das: So sind vor allem die ausführenden Ressourcen oder einzuhaltenden Zeitvorgaben nicht zu vernachlässigen. Anstatt Ressourcenverhalten oder Zeitvorgaben abzuschätzen, können mit Process Mining historische Ereignisdaten zur Hand genommen werden. So lassen sich Prozessmodelle weiter aufwerten. Wo im Geschäftsprozessmanagement bisher das Futter für kontinuierliche datengetriebene Überwachung und Diagnose fehlte, schließt Process Mining die Lücke. Und vor allem schließt es die Lücke nicht nur für die Prozessbeauftragten, sondern bringt den Prozess (je nach Plattform) zu den involvierten Mitarbeitern. Self-Service

und rollenbasierte Dashboards erlauben entsprechend der Datenkompetenz maßgeschneiderte Analysen für alle Teilnehmer, wodurch sich das Verständnis für den eigenen Prozess verbessert und Process Literacy Einzug in das ganze Unternehmen hält.

Analysen auf Knopfdruck

Um Process-Mining-Projekte erfolgreich auszurollen und die wiederholbare, kontinuierliche und ganzheitliche Analyse von Geschäftsprozessen zu ermöglichen, müssen die Prozessdaten aufbereitet und bereitgestellt werden. Dieser Teil eines Projektes kostet aus Erfahrung bis zu 80% des Gesamtaufwandes. Einmal implementiert, können die Daten über Extraktionsroutinen in beliebigen zeitlichen Abständen wiederholt zur Verfügung gestellt werden. Für den in diesem Artikel beschriebenen Anwendungsfall wurde eine Datenextraktion aus einer Instanz der Oracle E-Busi-

ness Suite (EBS) im Release 12.2. durchgeführt. Hinter dem umfangreichen User Interface der Oracle EBS verbergen sich Tausende von Tabellen und Views, aus denen die prozessrelevanten Informationen herausgefiltert werden müssen. Die Electronic Technical Reference Manuals (eTRM) von Oracle enthalten alle technischen Informationen des EBS-Datenmodells und können bei der Ermittlung und Aufbereitung der relevanten Daten genutzt werden. Um den O2C-Prozess vollständig abzubilden, werden zunächst die benötigten Events im Datenmodell der EBS identifiziert. Dann wird eine Datenextraktion initial konzipiert. *Abbildung 10* zeigt hierzu beispielhaft eine zu erzeugende Extraktionsdatenstruktur für relevante O2C-Events der Oracle E-Business Suite. Einige Marktanbieter stellen für verschiedene Anwendungsfälle vordefinierte Extraktionsroutinen für Oracle-basierte Systemlandschaften bereit. Diese decken

Standard-Objekte und -Felder ab, müssen jedoch bei Custom-Objekten und -Feldern, die kundenspezifische, prozessrelevante Events beinhalten, angepasst werden. In kontinuierlichen Process-Mining-Projekten werden die Daten dann inkrementell aus dem IT-System extrahiert, sodass wiederholbare und schnelle Prozessanalysen garantiert werden. Auf Knopfdruck, genau dann, wenn man es braucht.

Das Fazit: Prozesse erzeugen Daten und Daten steuern Prozesse

Die Erfahrung zeigt, dass, sobald die Process-Mining-Ergebnisse allgemein akzeptiert sind, weniger Zeit für die Diskussion darüber aufgewendet wird, was passiert, sondern sich der Schwerpunkt der Diskussion auf die Frage verschiebt, warum die Dinge passieren [1 S.8]. Ist der Fokus einmal auf dem „Warum“, kann kontinuierliches Prozessmanagement verstärkt werden, da das „Was“ nicht mehr aufwendig beschafft und infrage gestellt werden muss – die Daten neu zu laden, genügt. Die Verzahnung von Process Mining mit Analysetools aus der Business Intelligence führt zum Aufbau eines komplexen Prozessverständnisses, indem Prozesse und ihre Kennzahlen im Transaktionskontext betrachtet und mit klassischen Kennzahlen angereichert werden. Dieses ganzheitliche Bild ermöglicht das kollaborative und frühzeitige Erkennen von Abweichungen oder Prozessveränderungen und erhöht so die Reaktionsgeschwindigkeit, wobei zusätzlich aufgezeigt wird, an welchem Punkt im Prozess eingegriffen werden sollte. Was uns zurück zur Eingangsfrage führt – Gibt es so etwas wie Predictive Maintenance

für Prozesse? –, die wir mit Ja beantworten können.

Literaturverzeichnis

- [1] L. Reinkemeyer (ed.), Process Mining in Action, Springer Nature Switzerland AG, 2020.
<https://doi.org/10.1007/978-3-030-40172-6>.
- [2] Schönthaler F., Vossen G., Oberweis A., Karle T., Geschäftsprozesse für Business Communities – Modellierungssprachen, Methoden, Werkzeuge, Oldenbourg Verlag, 2011.
- [3] Hammer M., Champy, J. A., Reengineering the Corporation – A Manifesto for Business Revolution, Harper Collins Publishers (überarbeitete Ausgabe), 2003.
- [4] van der Aalst, W., Process Mining-Data Science in Action, 2nd ed. Springer-Verlag Berlin Heidelberg, 2016.
- [5] Karle T., Wissensbasierte kollaborative Einführung komplexer ERP-Prozesse, Know-Tech 2013 – Kongress für Wissensmanagement und Social Media, Hanau, 2013.
- [6] Dixit, P. M., et al., Enabling interactive process analysis with process mining and visual analytics, In Proceedings of the 10th International Conference on Health Informatics (2017), E. L. van den Broek, A. L. N. Fred, and H. Gamboa, Eds., vol. 5, pp. 573-584.



Janna Meyer

janna.meyer@mehrwerk.net

Janna Meyer arbeitet als Process Mining Specialist bei der MEHRWERK GmbH. Neben ihrer Verantwortung für verschiedene Produktentwicklungen von MPM ProcessMining – wie zum Beispiel Conformance Checking, automatisierte Ursachenanalysen oder MPM eXecution – leitet sie ganzheitlich Process Mining Implementierungsprojekte bei internationalen Marktführern, um deren Kernprozesse transparent und damit messbar zu verbessern. Ihre Leidenschaft ist Technologie wie MPM ProcessMining erfolgreich nutzbar zu machen, so dass Organisationen diese als Bereicherung für ihre Zusammenarbeit erleben.



Claus Engel

claus.engel@promatis.de

Claus Engel ist Vice President Operations SCM / Logistics bei der PROMATIS software GmbH und als Projektleiter in verschiedenen geschäftsprozessorientierten Implementierungen Oracle-basierter Unternehmenssoftware aktiv. Hierbei wird als Technologie auch Process Mining eingesetzt und die dazu erforderliche Extraktion der Prozessdaten aus den Oracle-Applikationen erfolgt mit dem im Beitrag beschriebenen Vorgehen.

Querverweis: Erfolgreich von zuhause arbeiten und Kollegenkontakt pflegen

DOAG Online

Nicht nur seit der Corona-Pandemie und den damit verbundenen Lockdowns arbeiten etliche Menschen von Zuhause aus.

Home Office ist ein Luxus, den viele Arbeitnehmer unter normalen Gegebenheiten sehr zu schätzen wissen. Die Ersparnis des Arbeitsweges und der damit verbundenen Zeit ist nur einer der vielen Vorteile des Remote-Arbeitens.

In Zeiten von Corona, wenn dieser Luxus zum Alltag wird, fehlen mittlerweile vielen Menschen die sozialen Kontakte zu ihren Kollegen.

Sven Peters gibt Ihnen in der **Java aktuell 5/20 ab Seite 63** wertvolle Tipps und Tricks,

wie Sie erfolgreich von Zuhause aus arbeiten und gleichzeitig die soziale Verbindung im Team trotz räumlicher Trennung stärken (siehe: <https://www.doag.org/de/home/news/java-aktuell-052020-microservices/detail/>).



Compliance und Corona: Neue Normen und Technologien?

Alexey Zakharov, Boston SoftDesign/DBI

Das Bankensystem wurde nicht immer nur als Instrument der Verwaltung von Finanzen und Investitionen angesehen, sondern auch als Plattform für betrügerische Handlungen: illegaler Wucher, illegale Überweisungen, die Finanzierung krimineller Aktivitäten und vieles andere mehr. Da Finanzbetrug jedes Jahr mit neuen Systemen, Ideen und Technologien ausgestattet wird, sollen die Maßnahmen zur Abwehr – Finanzkontrolle oder Compliance – immer einen Schritt voraus sein. Das Hauptziel der Compliance ist, die finanziellen Verluste und den damit verbundenen Reputationsverlust einer Bank oder einer anderen Finanzinstitution zu kontrollieren und im besten Falle auszuschließen. In diesem Artikel wird eine Reihe von Compliance-Technologie-Tools begutachtet und versucht, die wichtigsten Trends in der technologischen Sicherheit der nächsten Jahre zu ermitteln.

In diesem Artikel möchte ich über die Trends der Automatisierung von grundlegenden Aufgaben der finanziellen Compliance sprechen. Aber welche grundlegenden Compliance-Verfahren müssen zuerst automatisiert werden? Die erste Aufgabe der Compliance ist es, die Kunden zu identifizieren, vollständige Informationen über sie zu erhalten und diese Informationen mit verschiedenen Listen von „nicht korrekten“ Kunden abzugleichen. Dies wird KYC genannt, „Know your Client“. Obwohl dieser offizielle Begriff erst vor Kurzem eingeführt wurde, existierte dieses Verfahren eigentlich schon immer: Zu allen Zeiten wollten Geldverleiher, Wechselser und dann Banken, Börsen und andere Finanzinstitute nichts mit Kunden zu tun haben, die einen schlechten Ruf hatten. Zuerst tauschten die Vertreter der Banken die Listen der Kunden, dann wurde viel Arbeit in verschiedene Listen und Quellen gesteckt, die mittlerweile in der EU, den USA und vielen anderen Ländern üblich sind.

Vertrauensbildung durch KYC

Meistens beginnt KYC für den Kunden mit dem sogenannten „Onboarding“. Dabei werden alle möglichen Informationen, Passdaten, Informationen zum Status des Kontos, Staatsangehörigkeit und so weiter gesammelt und analysiert. Der wichtigste Teil von KYC ist die Überprüfung von allen Kunden mittels Sanktions- und PEP-Listen („Politically exposed person“). Auch beinhaltet KYC die ständige Überwachung der Kunden, wiederholte Überprüfungen sowie die Überprüfung der Begünstigten und in einigen Fällen sogar die Überprüfung der „Kunden der Kunden“. Das Ziel von KYC ist es zu wissen, mit welchem Kunden die Bank arbeitet. Als Ergebnis dieses Verfahrens wird jedem Kunden ein Risiko zugewiesen, das später das Vertrauen beeinflusst, das dem Kunden zugeschrieben wird.

KYC und die nachfolgende „Due Diligence“ sind der erste Schritt im Kampf gegen Geldwäsche und die effektivste Operation, um Terrorismusfinanzierung und Steuerhinterziehung zu behindern. Die Due-Diligence-Prüfung ist eine sorgfältige Prüfung und Analyse eines Unternehmens, insbesondere im Hinblick auf seine wirtschaftlichen, rechtlichen, steuerlichen und finanziellen Verhältnisse, die durch einen potenziellen Käufer eines Unternehmens vorgenommen wird [7].

Heute arbeiten die KYC-Prozesse mit Gesichtserkennung und anderen biometrischen

Technologien sowie Social-Media-Analysen; moderne Server- und Softwareprodukte ermöglichen die Verarbeitung großer Mengen von Informationen nahezu in Echtzeit. Das alles hilft der Compliance und zeigt uns die neuen Möglichkeiten, die Bankenindustrie sicherer zu machen.

Pandemie ermöglicht Ausnahmen von der Regel

Allerdings ist ein Großteil dieser Technologien unter dem Corona-Lockdown nicht verfügbar, ebenso wie einige andere Regeln und Verfahren von KYC nicht zur Verwendung kommen. Banken aus der ganzen Welt bieten sogar neuen Kunden die Möglichkeit an, ein neues Konto für „soziale Zahlungen“ online zu eröffnen. Damit sind Zahlungen mit sozialem Motiv gemeint, wie zum Beispiel Überweisungen an Verwandte, Spenden an gemeinnützige Stiftungen und so weiter. Es ist erwähnenswert, dass die Banken einiger Länder, zum Beispiel Russland, diese Möglichkeit abgelehnt haben. Ja, es gibt viele Streitigkeiten um diese Situation: Denn ohne die Formalitäten im Zusammenhang mit der Überprüfung der Dokumente des Kunden ist die Wahrscheinlichkeit groß, ein Mitglied einer kriminellen Organisation zu akzeptieren – und dies bereitet den Banken verständlicherweise Kopfschmerzen. Normalerweise wird das KYC-Verfahren immer für neue Kunden durchgeführt. Jetzt, aufgrund der Pandemie, bietet sich neuen Kunden somit die Möglichkeit, ein Konto zu eröffnen, ohne die Bank betreten zu müssen und ohne einige Compliance-Prüfungen zu absolvieren, die obligatorisch wären. Muss eine Bank nach der Pandemie die KYC-Prüfungen für die neuen Kunden nachträglich durchführen? Und wird es nicht zu spät oder gar sinnlos sein? Doch damit enden die Compliance-Aufgaben nicht.

AML-Verfahren gegen Geldwäsche

Nach der Identifizierung des Kunden warten andere Aufgaben auf das Finanzinstitut, etwa die ständige Kontrolle der Transaktionen, die vom Kunden durchgeführt werden. Der Komplex der entsprechenden Maßnahmen und Verfahren wird abgekürzt als AML (Anti-Money Laundering) bezeichnet. Der Ausdruck „Geldwäsche“ steht für eine Reihe von Maßnahmen, um Geld aus „illegalen Machenschaften“ (Bestechung, Betrug und so weiter) „sauber“ zu machen. Geldwäsche ist eine facettenreiche Art des Betrugs: In schöner Regelmäßigkeit findet man neue

Algorithmen und Möglichkeiten, die „dubiosen“ Ursprünge von Geld zu verbergen. Da Geldwäsche eine ganz bestimmte Abfolge bestimmter Schritte hat, ist in den meisten Fällen das effektivste Mittel, um festzustellen, ob eine Reihe von Operationen einen Präzedenzfall für Geldwäsche darstellt, der sogenannte Szenario-Mechanismus. Zum Beispiel können Transaktionen wie regelmäßige Überweisungen der gleichen Menge oder einer großen Summe (zum Beispiel mehr als zehntausend Dollar) in „cash“ einen Versuch darstellen, das Geld zu „waschen“. „Best practice“ von AML umfasst auch die KYC-Verfahren: Während der Ausführung der Szenarien müssen die Bankensysteme parallel die wiederholte Überprüfung der oben besprochenen Listen durchführen. AML ist die komplexeste Phase der Compliance, die die meiste Zeit und das meiste Geld kostet. Die KYC- und AML-Verfahren sind mit den Gesetzen und den Normen internationaler Organisationen, insbesondere der FATF (Financial Action Task Force on Money Laundering), der finanzbehördlichen Eingreiftruppe für Geldwäsche, geregelt.

Hat die Pandemiekrise die AML-Verfahren beeinflusst? Definitiv ja, da KYC, wie ich oben gesagt habe, und die Bekämpfung der Geldwäsche einen Teil der Funktionalität beziehungsweise ihrer Wirksamkeit verloren haben. Banken der ganzen Welt haben Lockerungen für Überweisungen von Konto zu Konto eingeführt: Die Grenzen der Überweisungen sind erhöht, Ausnahmen für Überweisungen in einigen Ländern genehmigt (Russland und Aserbaidschan) und so weiter. Natürlich machen die Banken das, um die Zahlungen zu vereinfachen; das ist in dieser schwierigen Pandemie-Zeit notwendig. Die Finanzinstitute geraten jetzt zwischen zwei Fronten: Ignorieren sie die Überweisungen zwischen Kunden, die bereit sind, sich in schwierigen Zeiten einander mit Finanzen zu helfen, so übersieht die Bank möglicherweise kriminelle Überweisungen. Wie im Fall mit KYC sorgt dieser Schritt unter Compliance-Experten für heftige Kontroversen. Wie ich bereits bemerkt habe, haben Lockerungen im KYC-Bereich die Chancen erhöht, einen „kriminellen Kunden“ zu verpassen; die Buchungen eines solchen Kunden können unter anderem einfach verloren gehen.

FATCA und CRS

Ein weiterer wichtiger Schritt der Compliance ist die Bekämpfung von Steuerhin-

terziehung. In vielen Finanzinstituten enthalten schon KYC und Listen-Screening die Maßnahmen zur Identifizierung von Steuerhinterziehern. Es gibt jedoch separate Vorschriften, auf die wir achten sollten: das amerikanische Gesetz FATCA (Foreign Account Tax Compliance Act) und sein europäischer Bruder, der CRS-Standard (Common Reporting Standard). Beide Standards verpflichten Länder, Daten über Einwohner auszutauschen, um Präzedenzfälle von Steuerhinterziehung auszuschließen. Angesichts der neuen Verbindung zu CRS wird es immer schwieriger, diesen Standard zu erfüllen: Viele Länder, CRS-Teilnehmer, haben nicht die technischen Fähigkeiten, um die Daten von allen Steuerzahlern zu sammeln und zu analysieren. Als Beispiel dient hier Aserbaidschan: In vielen Banken wird die Überwachung und CRS-Berichterstattung noch manuell durchgeführt.

Datenbanktechnologien – das Fundament der Compliance

Es ist kein Geheimnis, dass die Banken die wichtigsten Nutzer von Innovationen der Hardware/Software geworden sind, seitdem es die Informationstechnologie gibt. Die Automatisierung trifft dabei alle Aspekte der Arbeit der Banken und Compliance-Aufgaben sind keine Ausnahme davon.

Die Haupttechnologien von Compliance, die den Banken helfen, „nicht saubere“ Transaktionen und Zahlungen schnell zu definieren und verdächtige Zusammenhänge zwischen Konten zu finden, sind Datenbank-Technologien. SQL bietet hier zum Beispiel fast grenzenlose Möglichkeiten. Auch heute, mit der Entwicklung von spezialisierten Systemen und Tools, geben viele Banken die SQL-basierten Compliance-Systeme nicht auf. De facto ist das der einfachste Weg, um eine große Anzahl von Transaktionen zu verarbeiten: Eine Tranche erscheint in der Datenbank, dann kann man diese sofort analysieren und bei Bedarf stoppen. Mit der Entwicklung der Informationstechnologie werden die Compliance-Tools jedoch mit neuen, flexiblen Tools ergänzt.

Weitere Optionen bieten spezielle Software-Komplexe führender IT-Vendors, die auf die multilaterale Analyse und Kontrolle aller Aktivitäten der Kunden orientiert sind, wie zum Beispiel Financial Crime and Compliance Management (FCCM) von Oracle. Produkte wie FCCM enthalten die Module, die alle Compliance-Verfahren automatisie-

ren, sich komplett in bestehende Infrastrukturen integrieren, eine hohe Leistung zeigen und oft bereits vorkonfigurierte Funktionen wie Szenario-Sets oder fertige Vorlagen für FATCA- und CRS-Berichte enthalten.

Die Entstehung neuer betrügerischer Konzepte (zum Beispiel Diebstahl von Kontodaten, illegale Geschäfte, Geldwäsche) sowie der Anstieg der Finanzierung terroristischer Aktivitäten stellen die Technologie vor neue Herausforderungen. Noch vor 15 Jahren war es genug, nur einmal pro Tag (normalerweise in der Nacht) für alle Transaktionen im Bankensystem die AML-Prüfungen durchzuführen und einfach einen Report zu machen. Vielleicht gab es verdächtige Konten, Zahlungen mit Unsummen oder eine seltsame Häufigkeit von Zahlungen – das alles sind Anzeichen von Betrug. Aber heute, in der instabilen und schnellen Welt, es ist wichtig, Online-Screenings für alle Transaktionen auszuführen, das heißt, all diese Prüfungen so schnell wie möglich online zu machen: Nur diese Methodik kann potenziell gefährliche Transaktionen rechtzeitig stoppen.

Geopolitische Situation als Treiber der Compliance-Prozesse

Aber was genau zwingt Compliance-Technologien und die Vendors neuer Tools und Systeme, die Geschwindigkeit der Kontrolle beständig zu erhöhen? In erster Linie ist es die geopolitische Situation. Diese setzt Trends, nicht nur in den Krisenherden, sondern auch in den angrenzenden Ländern. Im Zusammenhang mit dem Iran zum Beispiel muss man die besondere Aufmerksamkeit auf die Länder Zentralasiens und des Kaukasus lenken. Die Länder dieser Regionen sind die Vermittler zwischen Europa und dem Nahen Osten, nicht nur in kultureller und wirtschaftlicher Hinsicht, sondern auch im Kontext von Operationen im Zusammenhang mit Geldwäsche und Terrorismusfinanzierung. Diese Situation zwingt die Länder des Nahen Ostens, die neuen Compliance-Standards einzuführen – und die ganze IT-Welt, diese Standards zu automatisieren. Damit ist das Niveau der Vorbereitung von Compliance-Offizieren (Bankmitarbeiter, die für Compliance-Prüfungen verantwortlich sind) der Länder des Kaukasus und des Nahen Ostens um ein Vielfaches höher als das Niveau ihrer Kollegen aus Europa. Aus diesem Grund werden die Tausende von Tranchen aus europäischen Ländern während AML-Screenings in Europa als „sauber“, aber in den Banken der östlichen Regionen als verdächtig markiert. Die Situ-

ation im Nahen Osten zwingt die IT-Spezialisten der Banken, Compliance-Prozesse zu beschleunigen, die Integration in alle Bankensysteme zu verbessern und nach neuen Wegen zu suchen, um verdächtige Tranchen schneller und hochwertiger abzufangen.

Fazit

In den neuen Realitäten von Covid und Isolierung spielen nicht nur die Technologien der Sicherheit und Remote-Arbeit eine wichtige Rolle, sondern auch die Technologien der Finanzkontrolle. Daher empfehlen wir den Finanzinstituten dringend, die Compliance-Technologien zu überdenken und auf die stabilen Vendor-Lösungen zu setzen.

Quellen:

- [1] <https://wirtschaftslexikon.gabler.de/definition/due-diligence-35668>



Alexey Zakharov

azaharov@bostonsd.ru

Alexey Zakharov ist Diplom-Mathematiker und Entwicklungsleiter von Boston SoftDesign (Europa- und USA-orientiert) sowie DBI (in Russland und ehemaligen GUS-Staaten tätig). Er ist seit mehr als 17 Jahren in der IT beschäftigt – vier davon bei Oracle – und hat Erfahrungen als Berater für Budgetierung, Reporting-Systeme, Oracle FCCM, AML, KYC, BI und Middleware.

Wir begrüßen unsere neuen Mitglieder

Persönliche Mitglieder

- David Becker
- Oliver Ziegler
- Lucy Murr
- Dennis Horn
- Lee Haines
- Gert-Jan Wijman
- Olaf Eggers
- Jochen Czwalina
- Ulrike Becker
- Markus Fest
- Andrea Obermeier
- Frank Naumann
- Stefan Krause
- Marta Termin
- Florian Schadow
- Christian Kohrs
- Sven Kasemann
- Mona Schiele
- Nora Zeitoun
- Anita Kaniut
- Aron Ferguson

Firmenmitglieder DOAG

- NOWEDA Apothekergenossenschaft eG, Detlef Baublies
- Stadt Nürnberg, Albert Rösch



Termine

März 03

12.03.2021

**DOAG Datenbank Online-Event:
Autoupgrade von Datenbanken mittels
Autoupgrade**
Online

16. - 17.03.2021

JavaLand 2021
Online

19.03.2021

Java aktuell 2/2021
Erscheinungstermin am Kiosk

24.03.2021

**DOAG Data Analytics Online-Event –
Knowledge Graph**
Online

April 04

09.04.2021

DOAG Datenbank Online-Event
Online

21.04.2021

**DOAG Data Analytics Online-Event –
HOL: Oracle Machine Learning**
Online

Mai 05

12.05.2021

**DOAG Datenbank Online-Event:
Wir schaffen das: Ein Leitfaden zur
Migration von Oracle nach PostgreSQL**
Online

04. - 05.05.2021

APEX connect 2021
Online

17. - 18.05.2021

DOAG Datenbank 2021
Online

19.05.2021

**DOAG Data Analytics Online-Event –
Grundlagen Machine Learning I + II**
Online

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: Björn Bröhl. Die DOAG Deutsche ORACLE-Anwendergruppe e.V. informiert kompetent über alle Oracle-Themen, setzt sich für die Interessen der Mitglieder ein und führt 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): Mario Beck, Clemens Bleile, Andreas Buckenhofer, Klaus Engel, Michael Fischer, Jürgen Günter, Dr.-Ing. Maximilien Kintz, Markus Klenke, Andreas Kother, Stefan Kühnlein, Janna Meyer, Nisanth Muthukirushnasamy, Borys Neselovskyi, Jan Ott, Malena Maria Reiners, Malena Reiners, Marion Rosemann, Dr. Frank Schönthaler, Tobias Schweiker, Olli Sennhauser, Andre Stapper, Günther Stürner, Carsten Thalheimer, Karolina Weber, Carsten Wiesbaum, Hermann Wilde, Alexey Zakharov.

Titel, Gestaltung und Satz:

Alexander Kermas
DOAG Dienstleistungen GmbH
(Anschrift s.o.)

S. 89: © Visual Generation |
<http://stock.adobe.com>
S. 98: © ar130405 | www.pixabay.com
S. 101: © gmast3r | www.123rf.com

Fotonachweis:

Titel: © frannyanne | www.123rf.com
S. 10: © ismagilov | www.123rf.com
S. 14: © ammentorp | www.123rf.com
S. 18: © Imam Fathoni | www.123rf.com
S. 25: © geralt | www.pixabay.com
S. 29: © marygor123 | www.123rf.com
S. 35: © Akhararat Wathanasing |
www.123rf.com
S. 39: © pratyaksa | www.123rf.com
S. 44: © tomertu | www.123rf.com
S. 49: © gopixa | www.123rf.com
S. 54: © Tithi Luadthong | www.123rf.com
S. 59: © ar130405 | www.pixabay.com
S. 63: © loganban | www.123rf.com
Titel S. 68: © STERN_Photo | www.pixabay.com
S. 80: © geralt | www.pixabay.com
S. 86: © berya113 | www.123rf.com

Anzeigen:

Julia Bartzik,
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

dbi services sa www.dbi-services.com	S. 23	MuniQsoft Consulting GmbH www.muniqsoft-consulting.de	S. 43	PROMATIS software GmbH www.promatis.de	S. 83
DOAG e.V. www.doag.org	U 2, U 3, U 4	MuniQsoft Training GmbH www.muniqsoft-training.de	S. 3		

The graphic features a central blue circle containing the text '2021 DOAG Konferenz + Ausstellung'. This circle is surrounded by several concentric blue circles and intersected by thick, colorful diagonal bars in shades of teal, pink, and purple. Small white circles are scattered along the blue lines, resembling nodes in a network or data points. The overall design is modern and tech-oriented.

2021
DOAG
Konferenz + Ausstellung

*Save
the
Date*

16. - 19. November in Nürnberg und online

2021.doag.org

Eventpartner:

AOUG
AUSTRIAN ORACLE USER GROUP

SOUG
swiss oracle
user group

iJUG
Verbund



APEX connect
by DOAG

4. - 5. Mai 2021
online

apex.doag.org

Save the Date

