

Red Stack

Magazin

DOAG

SOUG
swiss oracle
user group

AOUG
AUSTRIAN ORACLE USER GROUP

inklusive BUSINESS NEWS

AUTOMATISIERUNG



Aus der Praxis

Oracle-Infrastrukturen
mit Chef verwalten



IM INTERVIEW

Marc Linster,
CTO, EnterpriseDB

BUSINESS NEWS

Ethik, Green IT,
Nachhaltigkeit: Das
Business der Zukunft?

Unser Weiterbildungstipp für Sie: **Kompetente Wissensvermittlung im Robotron Schulungszentrum**

Praxis-Workshop: Automatisierung der Datenbank-Administration mittels Ansible

Inhalt des Workshops soll die Einsatzmöglichkeit des Werkzeugs Ansible im Bereich der Oracle Datenbank-Administration sein. Im Rahmen des Workshops gehen wir auf die Konfiguration des Ansible-Oracle-Moduls ein und stellen ausgewählte Anwendungsfälle vor.



15.09.2021
Online



Umfang
4 Stunden



Workshop-Gebühr
275,00 €



Anmeldung unter
robotron.de/ansible



Weitere Kurse und Praxis-Workshops im Robotron Schulungszentrum:

Praxisworkshop Clusterware, ASM und RAC mit Oracle 19c

Unsere nächsten Termine:
09.08.2021, 11.10.2021, 13.12.2021



Praxisworkshop SQL Tuning mit der Oracle Datenbank 19c

Unsere nächsten Termine:
09.08.2021, 25.10.2021, 06.12.2021



Praxisworkshop WildFly

Unsere nächsten Termine:
12.08.2021, 11.10.2021, 25.11.2021

NEU





Klaus-Michael Hatzinger
AOUG-Präsident

Liebe Mitglieder, liebe Leserinnen und Leser,

Es ist seit jeher ein natürliches Bestreben der Menschheit, Arbeitsprozesse in allen Lebensbereichen zu automatisieren. Automatisierung entlastet den Menschen, erhöht die Herstellungsgeschwindigkeit, verbessert die Produktqualität und Produktgleichheit, hebt die Zuverlässigkeit, optimiert die Prozesse und verringert die Fehleranfälligkeit. Letztendlich kann man durch Automatisierung in kürzerer Zeit wesentlich mehr erreichen und das auch noch äußerst kosteneffizient.

Die aktuelle vierte industrielle Revolution, auch Industrie 4.0 genannt, ist durch die Anwendung von Informations- und Kommunikationstechnologien in der Industrie, das Internet der Dinge und durch künstliche Intelligenz gekennzeichnet.

Man kann leicht erkennen, dass jede industrielle Revolution eine erhebliche Beschleunigung mit sich brachte und die Dinge laufend komplexer und vernetzter werden. Bereits automatisierte Bausteine werden zu größeren, wiederum automatisierten Bausteinen zusammengebaut, um noch schneller und effizienter noch mehr zu erreichen. Und das ist letztendlich die Essenz der Automatisierung.

Heute will man sich nicht mehr mit der Installation einer Grundfunktionalität wie etwa einer Datenbank beschäftigen, sondern möchte diese Grundfunktionalität durch Angabe einiger Parameter per Knopfdruck einfach vollautomatisch provisionieren und dann darauf aufbauen.

Die Corona-Pandemie hat der digitalen Transformation seit Anfang 2020 einen ordentlichen Schub verliehen. Nach dem Prinzip „es muss erst etwas passieren, bevor etwas passiert“ beschäftigen sich jetzt mehr denn je Unternehmen mit neuen, nachhaltigen Business-Lösungen, die auch Aspekte der Green IT und Ethik inkludieren. Sind Sie dafür gut aufgestellt?

Ich wünsche Ihnen viele interessante Momente beim Lesen.

Klaus-Michael Hatzinger



QR-Code der
Red Stack inkl.
Business News
Nr. 3/2021!



MUNIQSOFT

TRAINING

ORACLE
Silver Partner

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.

DB Monitoring I	21.06.-25.06.2021	€2.190.-netto
DBA Sommerspecial	05.07.-09.07.2021	€1.690.-netto
APEX Grundlagen	19.07.-23.07.2021	€2.190.-netto
SQL Fortschritt	09.08.-12.08.2021	€1.790.-netto

☎ 089 679090-40
Website: www.munisoft-training.de
Tipps: www.munisoft-training.de/tipps

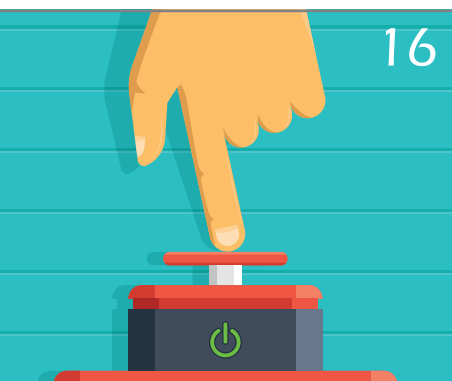
Schulungszentrum
Munisoft Training GmbH
Grünwalder Weg 13a
82008 Unterhaching/München

Mehr Oracle Schulungstermine unter munisoft-training.de

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



Interview mit Marc Linster



Datenbank auf Knopfdruck – Automatisierung mit AWX und ansible-oracle



Der DBA als Schlangenbeschwörer – bye-bye Bash, hi Python!



Ökologischer Lock-in: Worauf Sie bei der Ausgestaltung nachhaltiger Geschäftsmodelle achten müssen

Einleitung

- 3 Editorial
- 5 Timeline
- 6 Aus der Ferne betrachtet: Moderne Zeiten
- 7 „K8s wird ohne Zweifel die wichtigste Automatisierungsplattform der Zukunft sein“
Interview mit Marc Linster

Automatisierung

- 10 Oracle-Infrastrukturen mit Chef verwalten
Bernd Patolla
- 16 Datenbank auf Knopfdruck – Automatisierung mit AWX und ansible-oracle
Dietmar Uhlig
- 26 „Flexibilität as a Service“ Postgres im DevOps-Ökosystem
Jan Karremans & Borys Neselovskyi

Datenbank

- 30 Der DBA als Schlangenbeschwörer – bye-bye Bash, hi Python!
Katharina BÜchse
- 35 Data Exchange with PostgreSQL
Michael Kloker
- 41 Hochverfügbarkeit für die Oracle Datenbank 19c SE2 mit SEHA
Markus Flechtner

Digitalisierung

- 44 „Digitalisierung ist, wenn sich dann tatsächlich auch neue Geschäftsmodelle, neue Geschäftsprozesse und neue Verwaltungsprozesse daraus ergeben“
Interview mit Dr. Jan Mütter

Cloud

- 49 Lockdown oder: Wie ich lernte, die Cloud zu lieben – Teil 1
Dr. Jörg Domaschka, Steffen Moser, Thomas Nau, Simon Volpert
- 53 Safety First – Sicherungskonzepte von Oracle-E-Business-Suite-Umgebungen für OCI
Johannes Michler

Data Analytics

- 57 Smart Data Analytics – Wie Knowledge-Graphen und Artificial Intelligence Daten-Silos in businessrelevantes Wissen umwandeln
Dr. Martin Zablocki
- 61 Ordnung statt Chaos – Metadatenmanagement im Data Lake
Fabian Hardt
- 66 SQL Macros – Game Changing Feature für SQL-Entwickler?
Andrej Pashchenko
- 72 MariaDB-/MySQL-Stolperfallen – und wie komme ich da wieder raus? – Teil 2
Olli Sennhauser

Ethik, Green IT, Nachhaltigkeit: Das Business der Zukunft?

BUSINESS NEWS

- 76 Wege aus dem ökologischen Lock-in: Gestaltung nachhaltiger Geschäftsmodelle
Dr. Thomas Karle & Ana Paula Rivas
- 83 „KI und Machine Learning werden uns helfen, weiterhin im Enterprise Data Management für Unternehmen die beste Wahl zu sein“
Gespräch mit Jochen Wießler
- 85 Die Rolle der Ethik im Datenmanagement
Denise Baidinger
- 87 „Ein respektvoller und wertschätzender Umgang ist ein absolutes Muss und muss von allen gleichermaßen gelebt werden“
Gespräch mit Anne Driescher
- 90 Künstliche Intelligenz im Einsatz für Nachhaltigkeit – Prozessanalyse mit KI
Oliver Fuhrmann
- 92 Interview mit Jens Horstmann, Vorstand Trevisto, über die Zusammenarbeit im KI-Projekt mit SIEMENS und Fraunhofer IIS
Marcos López

Intern

- 93 Neue Mitglieder + Termine
- 94 Impressum + Inserenten

Timeline

9. April 2021

Im DOAG Datenbank Online-Event berichten Amin Farvardin und Ernst Leber über die Installation der PostgreSQL-Datenbank auf der ODA und welche Schritte für die Verlagerung erforderlich sind.

21. April 2021

Der Vorstand der Austrian Oracle User Group trifft sich am 21. April zur Eventplanung der kommenden Monate. In der Vorstandssitzung wird beschlossen, die für den diesjährigen September geplante Anwenderkonferenz aufgrund der immer noch angespannten Corona-Situation nur online stattfinden zu lassen und erst im Frühjahr 2022 wieder auf eine Vor-Ort-Veranstaltung zu setzen. Zudem wird beschlossen, einen technisch orientierten Beirat zu etablieren, der sich regelmäßig trifft, um neue Themen für zukünftige Veranstaltungen zu identifizieren.

22. April 2021

Die DOAG Development Community startet ihr neues Online-Format "DOAG Dev Talks" mit spannenden Diskussionsrunden zu aktuellen Development-Themen. Interessierte sind herzlich eingeladen, sich alle zwei Wochen an der freien Diskussion zu einem Thema mit den jeweiligen Experten zu beteiligen. Zum Start der neuen Reihe berichten die Referenten Niels de Bruijn und Carsten Czarski über das Thema Low-Code und APEX.

26. und 27. April 2021

In einem Berliner Expertenseminar zum Thema „Verwaltung von PostgreSQL-Datenbanken“, das sich insbesondere an Oracle-Datenbank-Administratoren richtet, stellt Christian Pfundtner online die PostgreSQL-Datenbank vor und vergleicht sie mit den entsprechenden Funktionen und Strukturen von Oracle.

29. und 30. April 2021

Das Berliner Expertenseminar mit Marco Patzwahl steht als Online-Veranstaltung auf dem Programm. Der zweitägige Kurs erklärt für DBAs und Entwickler Methoden, um in einer Oracle-Datenbank (Version 11.2 bis 21.2) aufzuräumen und Objekte zu reorganisieren.

4. und 5. Mai 2021

Die APEX connect 2021 [online] heißt seine rund 320 Teilnehmer/innen zum großen Wiedersehen der Community willkommen – bequem und sicher von Zuhause oder dem Büro aus. Auf dem Programm stehen zahlreiche Vorträge und Workshops zu den Themen APEX, JavaScript und PL/SQL. Das Online-Konferenzerlebnis bietet wertvollen Wissensaustausch und innovative Networking-Möglichkeiten. Studenten und Auszubildende, die Mitglied der DOAG sind, nehmen kostenfrei teil. Als beliebtes Highlight erweist sich das Tool Gather.town, mit dem die Teilnehmer mit ihrem eigenen Avatar durch das virtuelle Abbild des Phantasialands spazieren und sich durch Einsatz von Mikrofon und Webcam mit anderen austauschen. Eine moderierte Quizshow und eine virtuelle APEX Bar sorgen am Abend für gute Unterhaltung. Fried Saacke konstatiert: „Als Veranstalter freuen wir uns, dass die neue Konferenzplattform der DOAG in Kombination mit der interaktiven Welt Gather.town von der Community so gut angenommen wird und dass zumindest virtuell wie-

der echte Begegnungen stattfinden. Dies ist Ansporn für uns, die Plattform mit unseren Partnern weiterzuentwickeln und zu nutzen.“

6. Mai 2021

Die DOAG Development Community lädt im Rahmen des Online-Formats "DOAG Dev Talks" diesmal zur Diskussion zum Thema „Feedback/Nachlese der APEX Connect 2021“ ein.

11. Mai 2021

Im Berliner Expertenseminar vermittelt Lothar Flatz als Exadata-Experte der ersten Stunde die Grundprinzipien der Exadata-Datenbank und übt mit den Teilnehmern das Auswerten von SQL-Monitor-Berichten.

14. Mai 2021

Die DOAG Datenbank WebSession mit dem Titel „Wir schaffen das: Ein Leitfaden zur Migration von Oracle nach PostgreSQL“ findet statt. Referent Nabil Rütten zeigt, welche Punkte man beachten sollte, wenn eine Migration einer Oracle-Datenbank nach PostgreSQL geplant ist.

17. und 18. Mai 2021

Die DOAG 2021 Datenbank mit Exaday empfängt die Community am 17. und 18. Mai zum virtuellen Wiedersehen! Die rund 200 Teilnehmer erwartet ein breitgefächertes Online-Konferenzprogramm mit den Streams Datenbank und Exadata. Studenten und Auszubildende können als Mitglied der DOAG kostenfrei teilnehmen. Das Tool Gather.town, mit dem sich die Teilnehmer mit ihrem eigenen Avatar durch das virtuelle Abbild des gewohnten Veranstaltungsraums Van der Valk Airporthotel bewegen und sich mit anderen Teilnehmern an bestimmten Orten treffen und unterhalten, hat auch bei diesem Event seine erfolgreiche Premiere.

19. Mai 2021

Bei der DOAG DAC WebSession "Oracle Machine Learning: Overview, What's New? Customers and ML Apps" stellt Oracle-Referent Charlie Berger bestehende und neue OML-Funktionen vor.

1. und 2. Juni 2021

Im Berliner Expertenseminar zum Thema „Einführung in das Thema Postgres“ vermittelt Marco Patzwahl allen, die planen, eine PostgreSQL-Datenbank einzusetzen, viel praktisches Wissen und Einzelheiten von der Installation und Architektur der PostgreSQL bis hin zum PostgreSQL Backup & Recovery unter Windows und Linux.

14. und 15. Juni 2021

In einem weiteren Berliner Expertenseminar erläutert Jürgen Sieben, wie in APEX schnell und komfortabel entwickelt werden kann, ohne die Anforderungen an eine saubere Anwendungsstruktur zu verletzen.

23. und 24. Juni 2021

Im Berliner Expertenseminar vermittelt Randolph Eberle-Geist den Teilnehmern, wie Oracle "Parallel Execution" funktioniert, was seine Bestandteile sind und welche Diagnosemöglichkeiten für eine systematische Performance-Analyse geboten werden.



Aus der Ferne betrachtet: Moderne Zeiten



*Wolfgang Taschner
Rentner und ehemaliger Chefredakteur
des Red Stack Magazin*

Beim Thema „Automatisierung“ fällt mir spontan der Film „Modern Times“ mit Charlie Chaplin ein. Er spielt einen Fabrikarbeiter, der am Fließband steht und dort testweise von einer Maschine automatisch mit Material versorgt werden soll, um Pausenzeit einzusparen. Das Ganze gerät vollkommen außer Kontrolle und der Protagonist landet zwischenzeitlich in der Psychiatrie. Ganz so folgenschwer wie zu Beginn des letzten Jahrhunderts sollte heutzutage die Automatisierung einer IT-Infrastruktur oder Datenbank nicht mehr ausfallen ...

Als ich vor mehreren Jahren als Chefredakteur des Red Stack Magazin die Idee hatte, diesen Gastkommentar einzurichten, wollte ich bestimmte Themen „Aus der Ferne betrachtet“ behandeln – losgelöst von unternehmerischen Interessen und Zwängen. Meine beiden Freunde Dr. Dietmar Neugebauer und Günther Stürner haben seitdem regelmäßig sehr gute Texte verfasst.

Nachdem ich angefragt wurde, den Gastkommentar zum Schwerpunkt „Automatisierung“ zu schreiben, habe ich mir im ersten Augenblick einen Button gewünscht, über den sich dieser Text automatisch erstellen lässt. Doch so weit sind wir noch nicht und ich greife auf mein bewährtes Vorgehen zurück: Erst mal raus in die Natur und beim Spaziergehen meine Gedan-

ken treiben lassen. Zwei Nachmittage später hatte ich ein klares Bild im Kopf und der Text war schnell geschrieben.

Ich komme zum Thema: In den Oracle-Produkten sind etliche Automatisierungsfunktionen enthalten. Hinzu kommen zahlreiche Tools von Drittanbietern zur Optimierung der Datenbank sowie der IT-Infrastruktur. Doch Vorsicht vor deren unüberlegtem Einsatz. Jede Automatisierung sollte vorher genau überlegt sein, um nicht in die Charlie-Chaplin-Falle zu geraten. Zum Glück gibt es sehr gute Experten, die auf Veranstaltungen der DOAG oder in Beiträgen in den DOAG-Medien ihr Fachwissen mit anderen Anwendern teilen. An dieser Stelle auch mein nimmermüder Appell an Oracle, den Kunden-Support entsprechend zu verbessern, um bei Problemen schnell und kompetent Lösungen anbieten zu können.

Für das fortgeschrittene Alter beschäftigen meine Frau und mich ganz andere Automatisierungs-Themen: etwa ein Fahrzeug, das uns automatisch an ein gewünschtes Ziel bringt; eine Drohne, die uns die bestellten Einkäufe vor der Haustür ablegt; oder ein Roboter, der uns bei der Hausarbeit unterstützt und im Krankheitsfall pflegt. Mal sehen, auf welchem Medium wir uns in eventuell hohem Alter den Charlie-Chaplin-Film einmal ansehen werden.



„K8s wird ohne Zweifel die wichtigste Automatisierungsplattform der Zukunft sein“

Martin Meyer, Redaktionsleiter des Red Stack Magazin, und Borys Neselovskyi, DOAG Themenverantwortlicher für Middleware, unterhielten sich mit Marc Linster, EnterpriseDB, über Automatisierung.

Womit beschäftigen Sie sich in Ihrem Unternehmen EnterpriseDB?
EDB konzentriert sich auf die kommerzielle Anwendung von PostgreSQL, der führenden Open-Source-relationalen Datenbank. Mein Team, das „Office of the CTO“, und ich arbeiten mit unseren großen Kunden und den wichtigsten Partnern an strategischen Projekten, wie zum Beispiel der Migration zur Cloud, der Restrukturierung monolithischer Datenbanken, Umsetzung von Microservice-Architekturen oder dem Entwurf einer strategischen Migrationsfabrik (migration factory), damit ein Kunde Tausende Datenbanken von Oracle zu PostgreSQL bringen kann.

Das Feedback aus diesen Projekten hilft uns bei der Definition unserer strategischen Produktpläne. Auf diese Art und Wei-

se stellen wir sicher, dass unsere Produktlinie sich immer an Kundenbedürfnissen orientiert.

Wieso gibt es Automatisierung?

Im Datenbankbereich sehen wir zwei Arten der Automatisierung. Auf der einen Seite automatisiert man „Infrastruktur als Code“, das heißt die Konfiguration der Netzwerke, der Speichersysteme, CPU, Memory und die Softwarekonfiguration. Dies passiert durch Tools wie zum Beispiel Ansible oder Terraform. Diese sorgen dafür, dass man im Entwicklungs-, Test-, und Produktionsbereich die gleichen Konfigurationen hat und dass man sehr schnell neue Datenbankinstanzen erstellen kann.

Auf der anderen Seite haben wir die Versionskontrolle für die Datenbankschemata, also die DDL. Dafür benutzt man Tools wie beispielsweise Liquibase oder Flywheel, meistens zusammen mit GitHub.

Welche Rolle spielt Automatisierung in DevOps?

Beide Arten der Automatisierung sind sehr wichtig im DevOps-Bereich, in dem man immer wieder sehr schnell die gleichen Umgebungen erstellen will. DevOps ist oft gleichbedeutend mit „Agile“. Ohne durchgehende Automatisierung ist es unmöglich, in 30 Minuten in der Cloud einen neuen Datenbankserver mit High Availability und Monitoring zu konfigurieren. Mit Automatisierung ist das kein Problem.

Die Automatisierung von DevOps-Prozessen hilft bei der Vermeidung von menschlichen Fehlern, die im Rahmen der Installation und Konfiguration von komplexen Umgebungen auftreten. Die Routinen, die im Rahmen der Automatisierung entwickelt werden, dienen oft auch als Dokumentation der Infrastruktur. Standardisierung und Homogenisierung von komplexen Systemen ist ohne Automatisierung kaum möglich.

Wie können zwei normalerweise stark isolierte Unternehmensbereiche Dev und Ops im Rahmen von DevOps zusammenarbeiten?

Die meisten erfolgreichen DevOps-Transformationen bringen die Entwicklung und den Betrieb unter einen Hut. Sie befolgen die Rat-

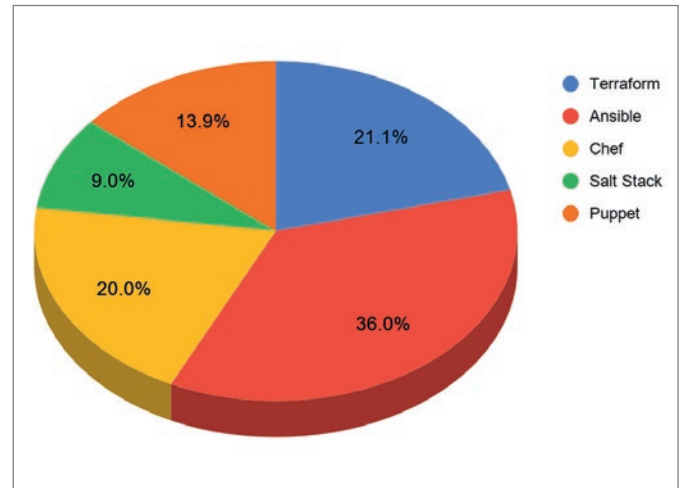


Abbildung: EDB-Umfrageergebnisse zur Popularität von Automatisierungswerkzeugen

schläge des CTO von Amazon „You built it, you run it“. Entwickler, Architekten und Administratoren sollen eng zusammenarbeiten, um alle Anforderungen an ein System, das nach dem DevOps-Prinzip aufgebaut werden soll, zu erfüllen. Alle Komponenten können in den Code überführt werden. Dafür sollen zum Beispiel die Netzwerk- oder Sicherheitsanforderungen von zuständigen Abteilungen klar definiert sein. Wenn das Konzept sauber definiert ist und von allen Beteiligten akzeptiert wird, kann der DevOps-Prozess implementiert werden. Das ist so ziemlich die einzige Möglichkeit, wenn man integrierte DevOps und CI/CD einführen will.

Wie werden die Sicherheitsaspekte bei der Anwendung von DevOps-Praktiken bei der Automatisierung berücksichtigt?

Mittlerweile führen immer mehr Kunden DevSecOps ein und integrieren Qualitäts- und Sicherheitsverifikation in den CI/CD-Prozess. Auf diese Art und Weise wird die Software inkrementell verifiziert und kontinuierlich getestet. Die wichtigsten Sicherheitsprinzipien, die auf dem Drei-A-Prinzip basieren – Authentifizierung, Autorisierung und Auditing – lassen sich gut in DevOps-Mechanismen integrieren. Die Verwaltung und Verteilung von TLS-Zertifikaten ist in vielen Unternehmen bereits voll automatisiert.

Was sind die populärsten Automatisierungsansätze?

EDB hat im Juni 2020 eine Umfrage auf der EDB Website durchgeführt. Mehr als 700 Antworten zeigen uns, dass Ansible sehr populär ist, gefolgt von Terraform und Chef (siehe Abbildung).

EnterpriseDB hat die Postgres-Provisionierung mithilfe von Ansible und Terraform komplett automatisiert. Die Ansible-Routinen sind im GitHub Repository abgelegt und stehen allen Interessenten frei zur Verfügung.

Woran muss man denken, bevor man anfängt zu automatisieren?

Zuerst sollte man sich überlegen, was man automatisieren will. Für uns bei EDB bedeutet das, dass man sich zuerst die richtigen Referenzarchitekturen aussucht. Im Datenbankbereich sind TPS, OLAP vs. OLTP, RPO (recovery point objective), RTO (recovery time objective) und GRO (geographic recovery objective) ausschlaggebend, wenn wir Referenzarchitekturen emp-



Zur Person: Marc Linster

Marc Linster hat an der Universität Kaiserslautern Informatik studiert. Nach fünf Jahren bei der Gesellschaft für Mathematik und Datenverarbeitung (heute ein Teil der Fraunhofer Gesellschaft) und einem Doktorat in der Informatik, hat er mehrere Jahre bei Digital Equipment in der angewandten Forschung und der IT gearbeitet. Am Anfang des Internetzeitalters war Marc als Berater im Integrations- und Supplychainbereich tätig, bevor er bei Polycom (heute Poly) für das technische Redesign des Video as a Service verantwortlich war. Seit 2013 ist Marc bei EDB leitend im Service und Produktbereich. Seit 2020 konzentriert Marc sich als CTO auf die Arbeit mit strategischen Kunden und Partnern.

fehlen. Die meisten Unternehmen definieren drei bis fünf Standard-Architekturen, automatisieren diese und integrieren die Automatisierung mit ihrem ITIL CMDB (Configuration Management Database). Auf diese Art und Weise können interne Kunden schnell und einfach einen neuen Server anfragen, der Server wird konfiguriert und sofort in die CMDB eingetragen.

Was sind anschließend die Tasks, die man automatisieren sollte?
Wir unterscheiden zwischen der Automatisierung der Softwareinstallation (Infrastruktur als Code) und der automatischen Datenbankversionskontrolle. Beide Aspekte sind wichtig. Wenn man Datenbanken betrachtet, ist es wichtig, alle Installations-beziehungsweise Konfigurationsprozesse zu automatisieren. Dazu zählen die Konfiguration des Betriebssystems, Installation und Konfiguration der Datenbank-Software, Bereitstellung der Hochverfügbarkeit, Einrichtung des Backup-Recovery-Konzepts. Und ganz wichtig ist die Konfiguration der Überwachung der Datenbank-Infrastruktur.

Ist das Vorgehen im Datenbankbereich anders als im Anwendungsbereich?

Im Anwendungsbereich gibt es keinen statischen Zustand, man kann die Anwendung einfach durch eine neue Kopie er-

setzen und wieder neu starten. Im Datenbankbereich muss man sowohl die Version der Datenbanksoftware (z. B. PostgreSQL 12.3) und die Version des Datenbankschemas verwalten.

*Welchen Ausblick in die Zukunft können Sie geben?
Was erwarten Sie an zukünftigen Entwicklungen?*

Kubernetes (K8s) ist ein sehr wichtiger Schritt für DevOps und CI/CD. K8s wird ohne Zweifel die wichtigste Automatisierungsplattform der Zukunft sein. Containerisierung und Microservices, zusammen mit fortgeschrittenen K8s-Operatoren (Level 3 oder höher), sind der nächste Schritt in diesem Bereich. Aktuell werden meistens relativ einfache Anwendungen wie Webserver oder Webanwendungen in Kubernetes betrieben. Diese profitieren von nativen Kubernetes-Eigenschaften wie Autoskalierung und Selbstheilung. Aber immer mehr komplexe Anwendungen, zu denen auch Datenbanken zählen, werden in Kubernetes-Clustern aufgebaut. Das größte Problem beim Betrieb von Datenbanken in Containern war die Storage-Anbindung. Diese Hürde wurde mittlerweile überwunden und es steht nichts mehr in Wege, auch große und IO-intensive Datenbanken in Kubernetes zu betreiben.

Zum Glück leben wir
im 21. Jahrhundert.



Entdecken Sie unsere Remote-Services für

- Platform Audits
- License Reviews
- Workshops



Oracle-Infrastrukturen mit Chef verwalten

Bernd Patolla, In&Out

Für die Automatisierung von Oracle-Datenbank-Infrastrukturen und -Datenbanken existieren diverse Tools. In diesem Artikel wird die Automatisierung einer Oracle-SIHA-Infrastruktur vorgestellt. Der Schwerpunkt liegt dabei auch auf der Integration in die vorhandenen Umsysteme wie **Oracle Cloud Control**, **Backup-Tools wie IBM Spectrum Protect und Nagios / CheckMK Monitoring**.

Einführung in Chef

Chef ist eigentlich ein Unternehmen, das 2008 gegründet wurde und das Produkt Chef Infra entwickelt hat. Heute stellt sich das Chef-Universum wie in *Abbildung 1* dar.

Für die Automatisierung von Oracle-Infrastrukturen wird die Chef Infrastructure Automation verwendet. Diese Infrastruktur besteht aus den drei Komponenten (*siehe Abbildung 2*).

- Chef Workstation
- Chef Server
- Clients

Auf der Workstation werden die Cookbooks mithilfe des Tools kitchen entwi-

ckelt, getestet und auf den Chef-Server hochgeladen. Auch wird auf der Workstation die Konfiguration der Server-Infrastruktur mit den Tools knife und chef verwaltet. Empfehlenswert ist das Ablegen der Konfigurations-Informationen in einem separaten Software-Repository wie zum Beispiel GIT. So kann die Konfiguration der Systeme historisiert werden. Der Chef-Server speichert die Cookbooks und die Konfigurations-Informationen der Server-Landschaft / der Clients. Die Clients holen sich die aktuelle Konfiguration vom Chef-Server und anschließend die notwendigen (und definierten) Cookbooks. Diese Cookbooks werden dann in einem Chef-Lauf ausgeführt. Nach einem Chef-Lauf meldet der Client sowohl den

Status als auch seine Konfigurations-Informationen zurück an den Server.

Chef Client / Node

Ein Chef-Node repräsentiert einen Server, eine virtuelle Maschine, einen Container oder eine „Cloud-Maschine“ die unter der Kontrolle von Chef ist. Auf dem Node ist der Chef-Client installiert. Je nach Konfiguration läuft dieser in gewissen Zeitabständen und führt die definierten Cookbooks mitsamt der Konfiguration aus. Die aktuelle Konfiguration des Nodes wird vor dem Ausführen der Cookbooks mit dem Tool ohai erfasst und am Schluss an den Chef-Server geschickt. Die Konfigura-

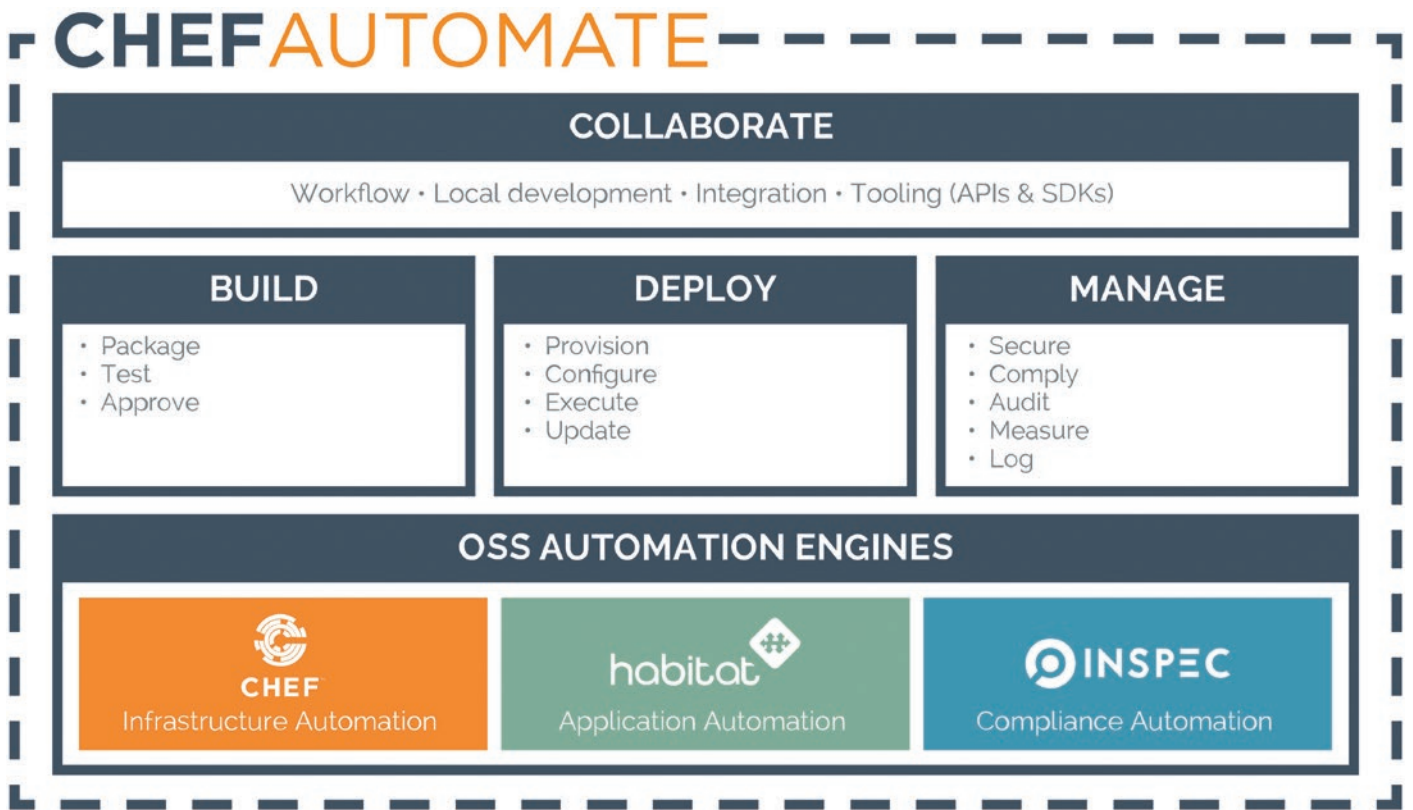


Abbildung 1: Chef Automate Universum (Quelle: https://docs.chef.io/platform_overview)

tion eines „node“ besteht aus verschiedenen Teilen:

- Eine Liste von abzuarbeitenden und versionierten „cookbooks“ (run-list)
- Zuweisung einer „role“ zu dem Node
- Zuweisung eines „environment“ zu dem Node
- Definition von „attributes“

Cookbook

Ein Cookbook beschreibt, wie etwas ausgeführt wird und welcher Endzustand vorherrschen soll. Cookbooks besitzen neben dem Code Metadaten wie eine Versions-Nummer und ein Readme. In

den Metadaten werden Abhängigkeiten von anderen Cookbooks, teilweise bis auf die Versions-Nummer, beschrieben. Daneben besitzt ein Cookbook verschiedene Komponenten:

- Recipes
- Attributes
- File Distribution
- Templates
- Ressourcen und/oder Libraries

Recipes

Ein Recipe ist grundsätzlich in der Sprache Ruby geschrieben. Eine Beschreibung von Ruby ist unter <https://www.ruby-lang.org/en/documentation/> zu finden. Mit der neuesten Chef-Version gibt es inzwischen die Möglichkeit, YAML-Code zu schreiben. Ein Recipe besteht aus einer Reihenfolge von Ressourcen und kann andere Recipes aufrufen. Ein Recipe definiert so punktuell den gewünschten Zustand des Systems. Recipes werden im Subverzeichnis „recipes“ eines Cookbook gespeichert.

File Distribution

File Distribution

Ein File in einem Cookbook kann durch eine Ressource auf den Node kopiert werden. Die in der Ressource definierten Files werden in dem Verzeichnis „files“ vorgehalten.

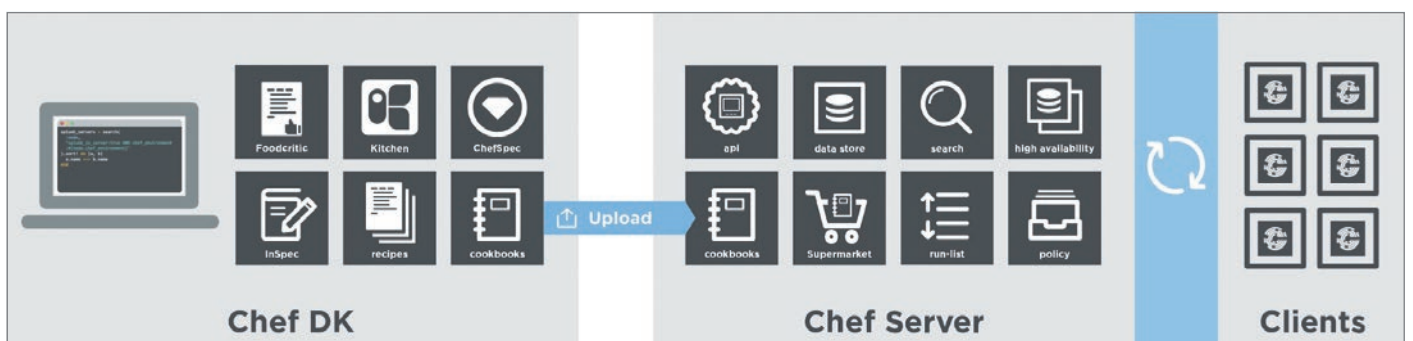


Abbildung 2: Chef-Infrastruktur-Übersicht (Quelle: https://docs.chef.io/platform_overview)

Templates

Templates grenzen sich gegenüber Files durch Platzhalter ab, die während der Laufzeit eines Cookbook mit dynamischen Werten gefüllt werden. In Templates kann auch Ruby-Code zum Erzeugen des entsprechenden Files auf dem Node benutzt werden.

Ressourcen / Libraries

Mit Ressourcen oder Library-Funktionen kann die Funktion von Chef erweitert werden. In einer Ressource können andere Chef-Ressourcen aufgerufen oder Ruby-Code verwendet werden.

Attributes

Attribute können in verschiedenen Stufen definiert werden:

- auf dem Node, um diesen zu beschreiben (z.B. Hostname, IP-Adresse, OS, ...),
- über das CLI oder in einem JSON-File beim Aufruf vom Chef-Client,
- in einem Cookbook,
- in einer Rolle,
- in einer Umgebung / Environment und
- in einem Policy-File.

Attribute sind aufsteigend priorisiert:

- **Default:** Ein Default-Attribut wird bei jedem Chef-Lauf neu gesetzt und hat die niedrigste Priorität in der Hierarchie.
- **Force_default:** Ein als Force_default definiertes Attribut in einem Cookbook hat eine höhere Priorität als ein Default-Attribut einer Rolle oder einer Umgebung.
- **Normal:** Ein normales Attribut bleibt in der Node-Konfiguration konsistent gespeichert.
- **Override:** Ein Override-Attribut wird beim Starten eines Chef-Laufs automatisch resettet. Es wird oft in einem Recipe oder einem Attribute File definiert. Override-Attribute sollten nur dann benutzt werden, wenn unbedingt notwendig.
- **Force_Override:** Wird benutzt, um sicherzustellen, dass ein Override-Attribut übersteuert wird.
- **Automatic:** Automatic-Attribute werden durch das Tool ohai gesetzt. Ein automatisches Attribut kann nicht geändert werden.

Attribute können an verschiedenen Stellen definiert werden:

- in einem JSON-File auf dem Node; es wird durch die Option „-j“ dem Chef-Client bekannt gegeben,
- auf dem Node durch das Tool ohai, das vor jedem Chef-Client-Lauf ausgeführt wird,
- in einem Attribut-File in einem Cookbook,
- in einem Recipe eines Cookbook,
- in Environment-Definitionen,
- iRollen-Definitionen oder
- in einem Policy-File.

Weitere Informationen zu Attributen und deren Hierarchie können der Webseite <https://docs.chef.io/attributes/> entnommen werden.

Roles

Rollen beschreiben eine Reihenfolge von Cookbooks und oder Recipes. Auch können in Rollen Attribute überschrieben werden. Rollen eignen sich beispielsweise für die funktionale Einteilung von Servern. So kann man zum Beispiel für alle Oracle-DB-Server eine Rolle oracle-db erstellen. In einer Rolle können auch für diese Rolle gültige Attribute definiert werden.

Environments

In Environments definiert man normalerweise die unterschiedlichen Funktionsstufen wie Entwicklung, Integration und Produktion einer Umgebung. Diesen Environments können dann Cookbooks mitsamt Versionen zugewiesen werden. So kann etwa eine neuere Version eines Cookbook zuerst in dem Environment „Entwicklung“ eingesetzt, anschließend in das Environment „Integration“ deployt und zum Schluss erst in der „Produktion“ aktiviert werden.

Oracle in Chef

Im Folgenden wird die Installation einer Oracle-19c-Datenbank mittels Chef Infra auf einem Red-Hat-Enterprise-Linux-7-System vorgestellt. Dabei findet eine Integration in die Um Systeme Oracle Cloud Control 13.3, Tivoli Spectrum Protect 8.1.9 und CheckMK (basierend auf Nagios) statt. Die Datenfiles inklusive Control-Files, Online-Redo-Log-Files, Archived-Redo-Log-Files sind in zwei ASM-Diskgruppen abgelegt, nur das spfile, pfile und Password File sind im Filesystem im ORACLE_HOME vorhanden. Der Ablauf der Installation als Überblick über die einzelnen Schritte ist im *Listing 1* aufgeführt.

Die Notation ‘::<name>’ bedeutet in Chef, dass ein anderes Recipe im gleichen

```
# Basis
include_recipe '::oracle-users'
include_recipe '::lvm-fs'
include_recipe '::dep-pkg-install'
include_recipe '::kernel-params'

# Grid / ASM Installation und Integration
include_recipe '::asm-sw-install'
include_recipe '::crs-setup'
include_recipe '::create-asm-instance'

# Cloud Control Agent Installation und Integration
include_recipe '::emcli-install'
include_recipe '::emcc-agent-install'
include_recipe '::emcli-reg-lsnr-asm'

# DB Software Installation
include_recipe '::db-sw-install'
include_recipe '::systemctl-odb'
clear_orainventory 'clear OraInventory' do
  action :clear
end

# Backup Integration
include_recipe '::tdpo-install'
```

Listing 1: Übersicht über die Oracle-Installation

Cookbook aufgerufen wird. Im Recipe `::oracle-users` werden zuerst die OS-User und Gruppen für die Datenbank als auch die Grid-Infrastruktur angelegt. Sowohl für Gruppen als auch für User existieren in Chef entsprechende Ressourcen (*siehe Listing 2*).

Daneben werden noch die User-Limits, sudo-Regeln und E-Mail-Forwarding für die beiden User oracle und grid eingestellt. Im Recipe `::lvm-fs` werden eine Volume Group mit Logical Volumes und Filesysteme für die Mount-Points `/u01` (späteres Oracle-Software-Verzeichnis), `/u10` für das Diag-Verzeichnis und `/u80` für Dumps, ... angelegt. Die Verzeichnis-Namen sind natürlich anpassbar und als Attribute definiert. Das Recipe `::dep-pkg-install` installiert die notwendigen OS-Packages für eine erfolgreiche Oracle-Installation. Die entsprechenden Packages sind als Attribute-Array definiert und können so dynamisch angepasst und erweitert werden. Auch wird das Kernel-Package `oracleasm` konfiguriert. `::kernel-params` installiert ein Tuned-Profil für Oracle inklusive HugePages. Alle Kernel-Parameter und -Werte sind auch als Attribute definiert. Wichtig ist hier, dass nur bei Änderungen des Tuned-Profiles dieses auch aktiviert wird: Der Tuned resettet temporär alle Werte auf die Defaults zurück, wenn ein Profil aktiviert wird. Wenn genau in diesem Augenblick in der Datenbank beispielsweise eine neue Session eröffnet wird, kann das zu Fehlern führen (durch nicht vorhandene HugePages, Shared-Memory-Segmente oder fehlende Semaphore). Die bisherigen Schritte sind teilweise auch in den Preinstallation-RPMs von Oracle enthalten. Die Preinstallation-RPMs stehen allerdings nur Oracle Linux Support Customers zur Verfügung.

GRID-Software

Anschließend fängt die eigentliche Software-Installation an. Dazu wird als Erstes ein Response-File auf dem Node erzeugt (*siehe Listing 3*).

Im Cookbook ist ein Template mit einem Platzhalter `hostname` vorhanden, das nun unter `$ORACLE_BASE` auf dem Filesystem abgelegt wird. Anschließend wird mithilfe einer `bash`-Ressource das (Golden-)Image-File der Grid-Software ausgepackt und das Skript `gridSetup.sh` unter anderem mit dem oben erzeugten Response-

```
%w(oinstall dba asmdba asmoper asmadmin).each do |group_name|
  group group_name do
    gid      node['odb']['group'][group_name]
    action :create
    only_if { node['etc']['group'][group_name.to_s].nil? }
  end
end
user 'oracle' do
  uid      node['odb']['user']['oracle']['uid']
  gid      node['odb']['user']['oracle']['gid']
  shell    node['odb']['user']['shell']
  comment  'oracle admin user'
  manage_home true
  password node['odb']['user']['pwd']
  action   :create
end
```

Listing 2: Oracle-User und Gruppen erstellen

```
template "#{node['odb']['asm']['ora_base']}/asm_swonly_193.rsp" do
  source 'asm_swonly_193.rsp.erb'
  owner  'grid'
  group  'oinstall'
  mode   '0640'
  variables(hostname: node['hostname'])
end
```

Listing 3: Erzeugen eines Response-Files

File für die Installation der Software aufgerufen. Das Package `cvuqdisk` wird aus dem Grid-Home-Verzeichnis installiert.

Im nächsten Recipe `::crs-setup` werden mit dem `roothas`-Perl-Skript (resp. Shell-Skript) die Cluster Ready Services (CRS) konfiguriert.

Jetzt kann die ASM-Instanz erzeugt werden (im Recipe `::create-asm-instance`). Dazu werden zuerst ein `init+ASM.ora`-File, das Password-File und das Audit Dump Directory erzeugt. Anschließend wird mit dem `srvctl` Utility die ASM-Instanz zu den CRS hinzugefügt und dann gestartet. Um unnötig viele Audit-Trace-Files zu vermeiden, wird das Check-Interval in den CRS auf 60 Sekunden hochgesetzt. Für das Management und Monitoring mittels Cloud Control werden der User `asmsnmp` erzeugt und die notwendigen Rechte zugewiesen. Ebenso wird zum Schluss noch ein `spfile` im Filesystem erzeugt. Damit ist die Installation und Konfiguration der Grid-Software und der CRS abgeschlossen.

Installation Cloud Control Agent

Für die Installation und Konfiguration des Oracle Cloud Control Agent wird das Cloud Control CLI benötigt. Beide Software-Kom-

ponenten stellt der Cloud Control Server über eine Web-Schnittstelle zur Verfügung. Das CLI steht über die URL https://cc-server:port/public_lib_download/emcli/kit/emcliadvancedkit.jar zum Download mittels `wget` zur Verfügung. Es wird noch ein JRE oder JDK zum Auspacken und für die Runtime benötigt. Anschließend muss noch die Verbindung zum Cloud Control Server in einem Aufruf `'emcli setup'` definiert werden. Je nach vorhandenen Security-Richtlinien können dabei die Optionen `-autologin`, `-certrans=yes`, `-licans=yes` und `-trustall` benutzt werden. Damit sind die Vorbereitungen für die Installation des Agent geschaffen. Ist auf dem Linux-System der `firewalld` aktiv (Abfrage durch `'firewall-cmd -state'`), so sollten die benutzten Ports freigeschaltet werden. Auch der Agent (resp. ein Shell-Skript für die Installation des Agent) kann vom Cloud Control Server über die URL <https://cc-server:port/install/getAgentImage> direkt heruntergeladen werden. Für die Installation des Agent wird dynamisch ein Response-File angelegt. Das heruntergeladene File wird mit den Parametern `RSPFILE_LOC` und `AGENT_BASE_DIR` aufgerufen. Nach der Installation und Konfiguration werden dem Agent noch die Default-Targets definiert (*siehe Listing 4*).

Nach der Konfiguration können nun die CRS-Services (als Type „has“), die ASM-Ins-

```
# add internal targets to agent
bash 'add_internal_targets' do
  user      'oracle'
  group     'oinstall'
  environment node['odb']['dba']['dba_env.to_s']
  code <<-EOC
    cd #{node['odb']['emcc']['orahome']}/agent_#{node['odb']['emcc']['version']}.0.0.0/bin
    ./emctl config agent addinternaltargets
  EOC
end
```

Listing 4: Definition der Default Targets

```
# register the has service
bash 'EM CC register has service' do
  user      'oracle'
  group     'oinstall'
  environment 'JAVA_HOME' => '/usr/java/latest'
  code <<-EOC2
    cd #{node['odb']['emcli']['home']}
    HAS_CNT=`./emcli get_targets -targets=has_#{node['fqdn']}:has -noheader | wc -l`
    [[ ${HAS_CNT} -eq 0 ]] && ./emcli add_target \
      -name="has_#{node['fqdn']}" \
      -type="has" \
      -host=#{node['fqdn']} \
      -properties="OracleHome:#{node['odb']['asm']['ora_home']};eonsPort:2016"
    exit 0
  EOC2
end
```

Listing 5: Registrierung der CRS im Cloud Control

tanz und der Datenbank-Listener im Cloud Control registriert werden. Dazu wird als Erstes mit „get_targets“ überprüft, ob die Komponente schon registriert sein sollte (von einem vorherigen Chef-Run). Als Beispiel werden im *Listing 5* die CRS registriert.

Oracle-Datenbank-Software

Auch bei der Installation der Oracle-Datenbank-Software wird ein Response-File für den runInstaller erzeugt und anschließend der runInstaller in einer bash-Ressource aufgerufen (mit einigen Parametern). Dann werden die nicht gewünschten Optionen (wie OLAP, RAT, ...) aus dem Oracle Binary entfernt und die benötigten (lizenzierten) (wie z.B. Unified Auditing, Partitioning) hinzugefügt.

Als Abschluss wird noch die Integration der zukünftigen Datenbanken in den System-Start-Prozess systemd erstellt.

Backup Software

Im Folgenden wird die Integration, Installation und Konfiguration der Backup-Soft-

ware Tivoli Spectrum Protect for Oracle beschrieben. Entsprechende Automatisierung-Hilfsmittel und Vorgehensweisen können auch für andere Backup-Software benutzt werden. Von einem (lokalen) Software-Repository wird die Backup-Software heruntergeladen. Die silent-Installation benötigt ein Response-File sowie eine einfache User-Interaktion (*siehe Listing 6*).

Falls auf dem Backup-Server noch ein Passwort für den DB-Server definiert wurde, muss auf dem Client ein sogenanntes Passwort-File erstellt werden. Dies bedingt bei Tivoli Spectrum Protect die interaktive Eingabe des Passworts. Eine einfache Ein- und Ausgabe-Umleitung wie im *Listing 6* ist hier nicht mehr ausreichend. Daher wird für die Automatisierung auf

das Tool expect zurückgegriffen. Dieses sucht in der Ausgabe nach definierten Strings (oder regulären Ausdrücken) und liefert dann die definierte Antwort zurück. In Chef ist das Vorgehen im *Listing 7* aufgeführt.

Nagios-/CheckMK-Integration

Als Basis für das Monitoring der Oracle-Datenbanken wird das Skript check_oracle_health von Gerhard Lauser verwendet. Allerdings wurde das Monitoring um eine dynamische Komponente erweitert: In der zu beobachtenden Datenbank werden in einer Tabelle die Monitoring-Checks und in einer zweiten die ange-

```
# install agent
bash 'install tdpo software' do
  code <<-EOC
    #{tdpo_bin_file} -i silent -f /var/tmp/tdpo_response.txt <<EOF
    2
    EOF
  EOC
end
```

Listing 6: Installation Backup Client Software

```

# create password file
expect_script 'create tdpo password file' do
  user 'root'
  group 'root'
  code <<-EOC
    spawn /bin/tdpoconf PASSWord -tdpo_opt=/opt/tech/scripts/oracle/rman_scripts/tdpo.opt
    expect {
      -regexp "Please enter current password:" {
        exp_send "#{node['hostname']}-bkp-ora\r"
        exp_continue
      }
      -regexp "Please enter new password:" {
        exp_send "#{node['hostname']}-bkp-ora\r"
        exp_continue
      }
      -regexp "Please reenter new password for verification:" {
        exp_send "#{node['hostname']}-bkp-ora\r"
        exp_continue
      }
    }
  EOC
  not_if { File.exist?('/home/oracle/scripts/rman_scripts/TSM.IDX') }
end

```

Listing 7: Konfiguration des TDPO-Password-Files

```

# create sql script:
template "#{node['odb']['check_mk']['script_dir']}/create_nagios.#{sid}.sql" do
  source 'check_mk.sql.erb'
  owner 'oracle'
  group 'oinstall'
  mode '0640'
  action :create
  not_if File.exist?("#{node['odb']['check_mk']['script_dir']}/create_nagios.#{sid}.sql")
end

```

Listing 8: Erzeugen des SQL-Skripts für das Monitoring

passten Thresholds gespeichert. Dementsprechend wird beim Erzeugen der Datenbank ein SQL-Skript zum Erstellen des Monitoring-Users, der Tabellen und das Befüllen derselben mit den Default-Checks und Thresholds ausgeführt. Das Skript ist als Template vorhanden (*siehe Listing 8*), das anschließend mittels sql-plus ausgeführt wird.

Damit wäre dann die Installation der Grid- und Datenbank-Software mit Integration in Cloud Control, der Backup-Software und CheckMK/Nagios automatisiert. Diese Mechanismen können nun auch für weitere Upgrades der Software benutzt werden, wie etwa für ein Out-Of-Place-Patching.

Genauso kann auch das Erstellen einer Datenbank-Instanz automatisiert werden. Inzwischen wurde auch ein Cookbook für das automatisierte In-Place-Patching erstellt, das natürlich auch vorab Applikationen stoppen und Blackouts sowohl im

Cloud Control als auch in anderen Monitoring-Applikationen setzen kann.

Zusammenfassung

Oracle-Infrastrukturen können mit Chef automatisiert betrieben werden. Dabei werden die Installation und Konfiguration sowohl der Datenbank als auch einer GRID-Infrastruktur unterstützt. Es können auch Umsysteme wie Cloud Control, Backup-Systeme sowie weitere Monitoring-Systeme automatisch eingebunden werden. Manuelle und fehlerträchtige Tasks können auf ein Minimum reduziert werden.

Über den Autor

Bernd Patolla ist bei der Firma In&Out AG in der Schweiz als Senior IT Consultant

tätig. Die Schwerpunkte bilden Oracle-Datenbanken, Server- und Storage-Infrastrukturen. Seit mehreren Jahren arbeitet Bernd Patolla in diversen Projekten im Bereich Automatisierungen im Chef-Umfeld.



Bernd Patolla
bernd.patolla@inout.ch



Datenbank auf Knopfdruck – Automatisierung mit AWX und ansible-oracle

Motivation: Eine Datenbank auf Knopfdruck bereitstellen? In der Oracle Cloud gehört das vom ersten Tag an zum Standardrepertoire und ist ein wichtiges Verkaufsargument. In der historisch gewachsenen On-Premises-Welt sind auch die Skripte über die Jahre gewachsen, die dem DBA die Routine-Arbeiten erleichtern. Bekanntermaßen sind die ausschließliche Bedienung der Kommandozeile und das Jonglieren mit Konfigurationsdateien und Parametern nicht jedermanns Sache. Auch gibt es neben dem fachlich-technischen Kern der Skripte weitere Aspekte, die es bei der Arbeit im Team und in mehr als trivialen Infrastrukturen zu beachten gilt: Zugangskontrolle, Parallelisierung, Job-Steuerung, Protokollierung.

Die Kombination aus AWX und dem Paket `ansible-oracle` verspricht auf der einen Seite, mit mäßigem Aufwand das Ziel einer einfachen Bedienung zu erreichen, und auf der anderen Seite, Complianceanforderungen und Nachvollziehbarkeit von Systemänderungen abzudecken. Wir zeigen in diesem Beitrag nicht alles, was möglich ist, sondern das, was wir bei Robotron im praktischen Einsatz verwenden.

Robotron in seinen Rollen als Software-Hersteller einerseits und SaaS-Anbieter für die eigene Software andererseits [1] betreibt viele Datenbanken für nicht ganz so viele verschiedene Applikationen. Die Anforderungen je Datenbank sind deshalb schon speziell, über die Menge ergibt sich jedoch ein Wiederholungseffekt.

ansible-oracle

`ansible-oracle` ist auf GitHub verfügbar [2]. Wer es nicht per „git clone“, sondern als Zip-Datei herunterlädt, muss noch die `ansible-oracle-modules` mitnehmen [3]. Das gesamte Bündel ist bislang weder als Collection noch in der Ansible Galaxy verfügbar, aber dennoch sehr nützlich. `ansible-oracle` enthält mehrere Rollen, um einen Datenbankserver für die Oracle-Installation vorzubereiten, eine Grid Infrastructure oder Datenbank zu installieren, zu patchen oder zu deinstallieren sowie verschiedenste Aktionen mit oder in der DB auszuführen (Tablespaces, Redo, User, Grants managen, um nur einige zu nennen). Die Module aus `ansible-oracle-modules` sind auch ohne die

Rollen aus `ansible-oracle` nutzbar und erlauben es, ganz spezifische eigene Rollen zu schreiben, um zum Beispiel beliebige SQL-Anweisungen auszuführen.

Selbst wenn man sich nicht mit den Internen des Pakets auseinandersetzen will, muss man doch die wichtigsten Datenstrukturen kennen und verinnerlichen, die in den Rollen verwendet werden. Kommentierte Beispiele finden sich in `ansible-oracle` im Unterverzeichnis `group_vars`. Es bietet sich an, genau diese Datenstrukturen (siehe *Abbildung 1*) auch in eigenen Rollen zu verwenden, um sich aufwendige Umwandlungen zu ersparen und die Rollen in eigenen Playbooks kombinieren zu können.

`db_homes_installed` ist die Auflistung der Oracle Homes, die auf der jeweiligen

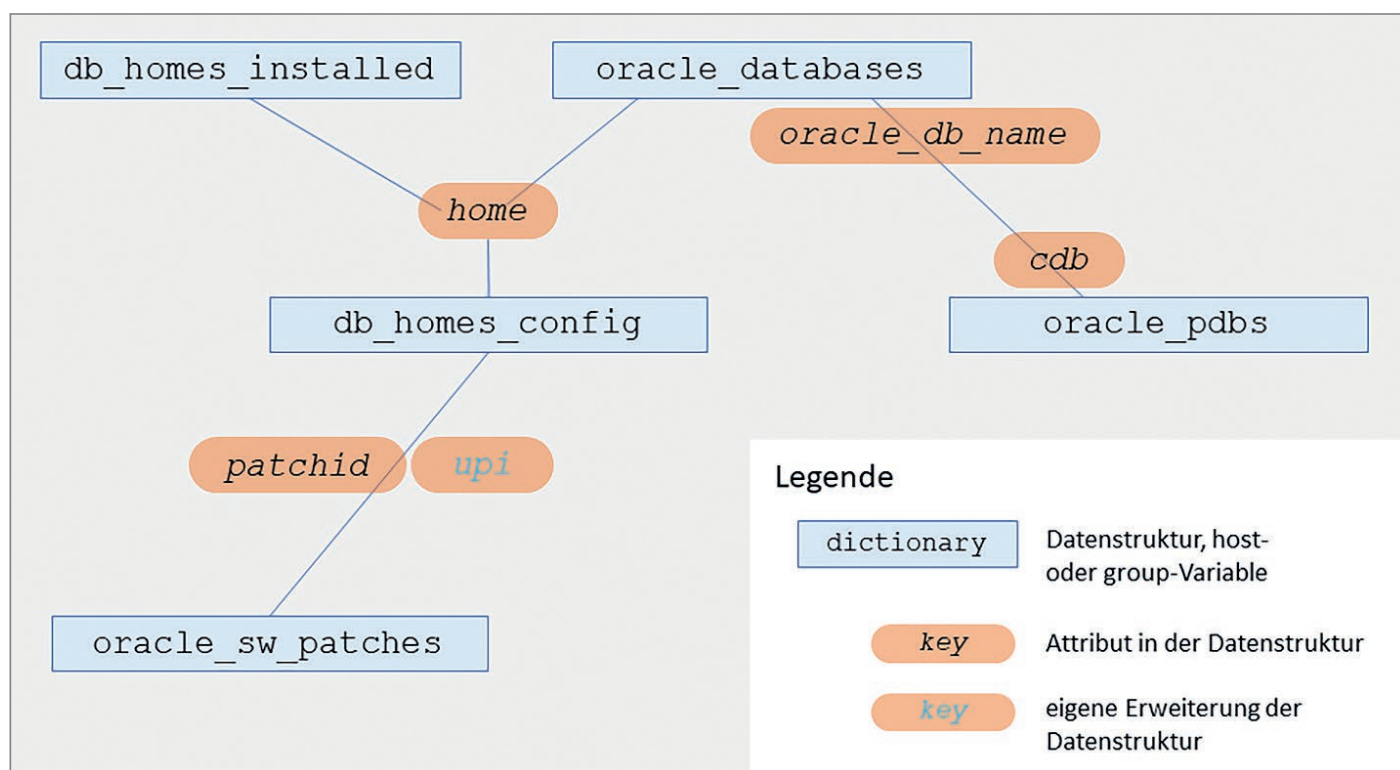


Abbildung 1: Wesentliche Datenstrukturen in `ansible-oracle` (Quelle: Dietmar Uhlig)

Maschine installiert sind. Diese Datenstruktur (Dictionary) enthält nur den symbolischen Namen des Oracle Home und wenige spezifische Parameter. Details der Home-Konfigurationen sind zum Zweck der Wiederverwendbarkeit in `db_homes_config` ausgelagert. In `oracle_sw_patches` stehen Angaben zu den Zip-Dateien der Interim Patches und RU, die man bei Oracle herunterlädt. Die PDBs in `oracle_pdb`s befinden sich auf derselben Gliederungsebene der Datenstrukturen wie die CDBs und die DB Homes (*siehe Listing 1*). Diese Modellierung, im Gegensatz zu tief strukturierten Dictionaries, entspricht dem Ansatz, der in relationalen Datenbanken gewählt wird.

Die Datenstrukturen `db_homes_installed`, `oracle_databases` und `oracle_pdb`s sind in unserer gewachsenen DB-Landschaft host-spezifisch. Per Konvention verfassen wir die Inventories im yml-Format und tragen die drei genannten Dictionaries auch dort ein (*siehe Listing 1*).

Ebenfalls per Konvention sind `db_homes_config` und `oracle_sw_patches` große Datenstrukturen, die zentral als Gruppenvariable geführt werden, vergleichbar mit Katalogen.

AWX

AWX ist mehr als nur eine grafische Oberfläche für Ansible. Es bietet zusätzlich zu den Funktionen der Ansible-Kommandozeilenwerkzeuge unter anderem:

- ein Berechtigungskonzept,
- die Verwaltung von Credentials,
- eine Job-Steuerung,
- einen Workflow-Editor, um Playbooks parallel oder nacheinander und abhängig vom Erfolg des Vorgängers auszuführen
- und noch eine angenehme Kleinigkeit: die E-Mail-Benachrichtigung zu Beginn und/oder Ende eines Jobs.

Alle Funktionen von AWX können auch über ein REST API angesprochen werden. Im heutigen Artikel beschränken wir uns allerdings auf die grafische Oberfläche. Grundlagen der Bedienung findet man im Ansible Tower User Guide [4].

Während die kommerzielle Variante, Ansible Tower, auch direkt auf Red Hat Linux installiert werden kann, wird AWX nur als Docker-Images angeboten. Diese lassen sich jedoch recht einfach installieren, wenn man während der Installation einen direkten Zugriff auf den Docker Hub bekommt. Da sich die Details der Installation von Zeit zu Zeit ändern, wird hier lediglich auf die Installationsanleitung [5] verwiesen.

Arbeitet man nur mit Ansible, kann man eigene und fremde Rollen, Module, Playbooks usw. beliebig auf dem Ansible-Controller verteilen. In der Ansible-Konfigurationsdatei hält man die Struktur über Suchpfade zusammen. Spätestens beim Einsatz von AWX sollte man sich befleißigen, alle notwendigen Dateien in einer Versions-

```
---
all:
  hosts:
    dbsrvk1840:
      oracle_install_version_gi: "19.3.0.0"
      gi_patches: { 19.3.0.0: "{{ gi_patchlists['19.7.0.0'] }}" }
      db_homes_installed:
        - { home: "12.2.0.1.200414-se", apply_patches: True, state: present }
        - { home: "19.7.0.0-ee",      apply_patches: True, state: present }
      oracle_databases:
        - oracle_db_name: zlr17sc
          home: "12.2.0.1.200414-se"
          state: present
          is_container: True
          init_parameters: "{{ init_parameters['12.2.0.1-EMS-SE2-CDB'] }}"
          profiles: "{{ db_profiles['GENERIC'] }}"
        - oracle_db_name: ksr19ec
          home: "19.7.0.0-ee"
          state: present
          is_container: True
          init_parameters: "{{ init_parameters['19.0.0.0-EMS-EE-CDB'] }}"
          profiles: "{{ db_profiles['GENERIC'] }}"
      oracle_pdb:
        - { cdb: zlr17sc, pdb_name: zlr17s, state: present }
        - { cdb: ksr19ec, pdb_name: hir19e, state: present }
        - { cdb: ksr19ec, pdb_name: kcr19e, state: present }

      rds_common:
        children:
          rds_db:
            hosts:
              dbsrvk1840:
                ...
```

Listing 1: Inventory mit Hostvariablen

```
C:\_wa\ansible\pb_projects\ora_db_mgmt\roles>svn proplist -v
Properties on '.':
  svn:externals
    http://svn.int/svn/prj/tags/ansible/rdsintra/stable rdsintra
    http://svn.int/svn/prj/tags/ansible/ansible-oracle/stable ansible-oracle
```

Listing 2: SVN-Externals in der Playbook-Zusammenstellung

verwaltung abzulegen und so zu strukturieren, dass sich das Konstrukt (in AWX: „Projekt“) auf einem beliebigen Rechner (in AWX: im Docker-Container) herunterladen lässt und ohne weitere Anpassungen funktionsfähig ist. Wir verwenden für die Versionierung Subversion. Die Fremdprodukte und die eigene Entwicklung werden als Subversion-Externals „zusammengesteckt“ (siehe Abbildung 2 und Listing 2) [6]. Eine lokale Konfigurationsdatei enthält die Pfade zu den Unterverzeichnissen (siehe Abbildung 2 und Listing 3). AWX erwartet die Playbooks im Project Root Folder. Die Playbook-Dateien dort verweisen per `import_playbook` [7] auf die Playbooks in dem eingebundenen SVN-Projekt (jeweils ein Einzeiler, siehe Listing 4).

Beispiel

Anlegen einer CDB mit Ansible (Kommandozeile)

Für die weiteren Betrachtungen gehen wir davon aus, dass auf jedem DB-Server mehrere Oracle Homes installiert sein können und unter jedem DB-Home in der Regel mehrere Datenbanken laufen. Weiterhin nehmen wir an, dass es nicht egal ist, auf welchem Server eine Datenbank installiert wird, weil jeder Server einem Geschäftsbereich, einer Kritikalität usw. zugeordnet ist. Außerdem soll noch vorausgesetzt werden, dass die Software, also das DB Home und gegebenenfalls die GI, bereits installiert ist.

Zum Anlegen einer Datenbank müssen mindestens Servername, Oracle Home und DB-Name angegeben werden. Wer keine Configuration Management Database [8] betreibt, aus der dynamisch das Inventory generiert wird, sondern das Inventory in Dateiform pflegt, kann die Oracle Homes auf einem DB-Server mit einem Einzeiler herausfinden (siehe Listing 5).

Auslagerung von Standardwerten

ansible-oracle bietet die Möglichkeit, verschiedenste Optionen zum Anlegen einer

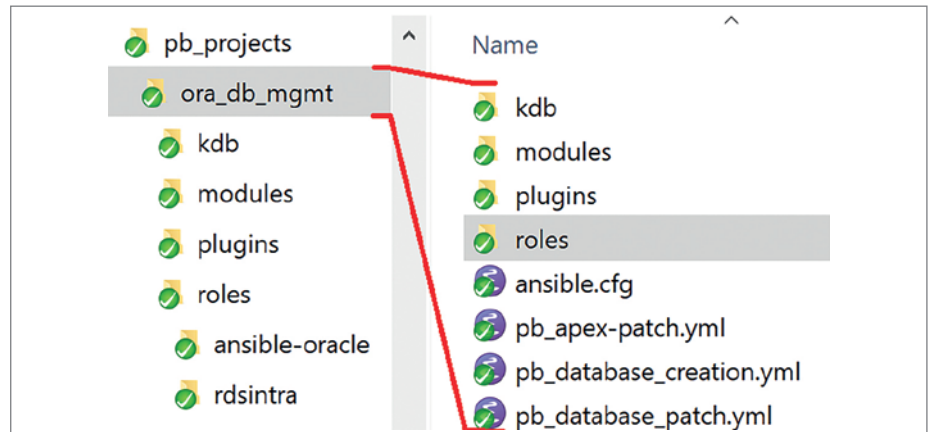


Abbildung 2: Verzeichnisstruktur der Playbook-Zusammenstellung für AWX (Quelle: Dietmar Uhlig)

```
[defaults]
library          = ./modules
roles_path       = ./roles/rdsintra/roles:./roles/ansible-oracle/roles
lookup_plugins  = ./plugins/lookup

display_skipped_hosts = False

callback_whitelist = logstash

# External Log Configuration is done in AWX
#[callback_logstash]
```

Listing 3: Lokale Ansible-Konfigurationsdatei

```
- import_playbook: roles/rdsintra/pb_database_creation.yml
```

Listing 4: Playbook im Project Root – nur ein Verweis auf das eingebundene SVN-Projekt

```
$ ansible -i inv_global.yml dbrsvk1840 -m debug \
-a msg="{{ db_homes_installed |
  json_query('[?state==\`present\`].home') }}"
dbrsvk1840| SUCCESS => {
  "msg": [
    "12.2.0.1.200414-ee",
    "12.2.0.1.200414-se",
    "19.7.0.0-ee",
    "19.8.0.0-se"
  ]
}
```

Listing 5: Abfrage der DB-Homes-Namen im Inventory

Datenbank zu spezifizieren. Meist werden einige Optionen durch die Applikation vor-

gegeben, zu der die Datenbank gehört (Beispiel: Zeichensatz, DB-Optionen wie

Oracle Text, Initialisierungsparameter), und andere durch unternehmensinterne Festlegungen (Beispiel: Storage Type). Man verwendet also im eigenen Unternehmen oft nur eine kleine Anzahl von Variablensätzen. Es wäre vorteilhaft, auf den passenden Variablensatz (z.B. „Applikation EDM“) zu verweisen und die Hostvariablen im Inventory kurz und übersichtlich zu halten. Außerdem hilft so eine Auslagerung, Wildwuchs zu vermeiden, weil Datenbanken mit gleichen Parametern angelegt werden.

Solange solche Einstellungen oder Variablen nicht Teil von Dictionaries sind, lassen sie sich als Gruppenvariable definieren und allein durch die Zuordnung eines Host zur passenden Gruppe setzen [9]. Die oben genannten Beispiele sind hingegen Attribute in Listen von Dictionaries. Dort muss man das Einsetzen von Standardwerten in Ansible „ausprogrammieren“. Dahinter stecken zwar meist einfache Jinja2-Konstrukte, man muss sich dennoch Gedanken darüber machen, ob man die Variable in allen oder nur in bestimmten Zweigen eines Dictionary mit einem bestimmten Wert vorbelegen will.

Zur Verdeutlichung sei hier ein experimentelles Beispiel angegeben. Um den Storage Type nicht pro Datenbank angeben zu müssen, sondern aus einem Standardwert zu lesen, könnte eine neue Variable `effective_storage_type` folgendermaßen angegeben werden:

```
effective_storage_type: >-
  {{ dbh.storage_type | default(storage_type) | default("FS") }}
```

Dann müssten alle Verwendungen von `dbh.storage_type` durch `effective_storage_type` ersetzt werden. Diese Lösung greift tief in die Rollen von `ansible-oracle` ein und birgt die Gefahr, dass die eigenen Änderungen inkompatibel werden mit dem Original. Alternativ lässt man die Datenstrukturen in den Hostvariablen vollständig und verweist an der Stelle auf ein Dictionary mit Standardwerten, die man über die Gruppenzugehörigkeit des Host festlegen kann (siehe Listing 6).

Umsetzung in AWX

Beim Anlegen eines Templates legt der AWX-Administrator fest, welche Werte beim Starten dieses Templates als Job erfragt werden und welche vom Operator nicht änderbar sind. Oft werden Inventory, Credentials und Job-Tags erfragt (siehe Abbildung 3, gelbe Markierung). Es ist

ganz praktisch, wenn die Job-Tags die Abschnitte im Playbook widerspiegeln. Gibt man die Liste der Tags bereits im Template vor, kann man beim Aufruf per Klick einzelne Abschnitte des Playbooks abwählen (siehe Abbildung 4). Nebenbemerkung: Leider müsste man so ein Tag manuell wieder eintippen, wenn man es aus Versehen aus der Liste entfernt hat. Meist wird stattdessen der Job-Dialog abgebrochen und neu aufgerufen.

Survey

Noch heikler ist es, das Feld „Extra Variables“ beim Job-Aufruf vervollständigen zu lassen. Dort muss zusätzlich die yml- oder json-Syntax eingehalten werden. Besser ist es für so einen Fall, die Extra-Variablen in einem Survey abzufragen [10]. Ein Survey lässt Text- und Zahlenfelder sowie Wertelisten in verschiedenen Varianten als optionale oder Pflichtangaben zu. Die Gestaltung eines Survey ist leider nicht sehr mächtig: Wertelisten können nicht dynamisch oder extern gefüllt werden, Eingabefelder lassen sich nicht abhängig voneinander schalten, komplexe Daten (z.B. mehrspaltige Listen) sind nicht

```
oracle_databases:
  - oracle_db_name: orcl
    storage_type: ora_db_defaults['RDSINTRA']['storage_type']
```

Listing 6: Beispiel für das Auslagern eines Standardwertes

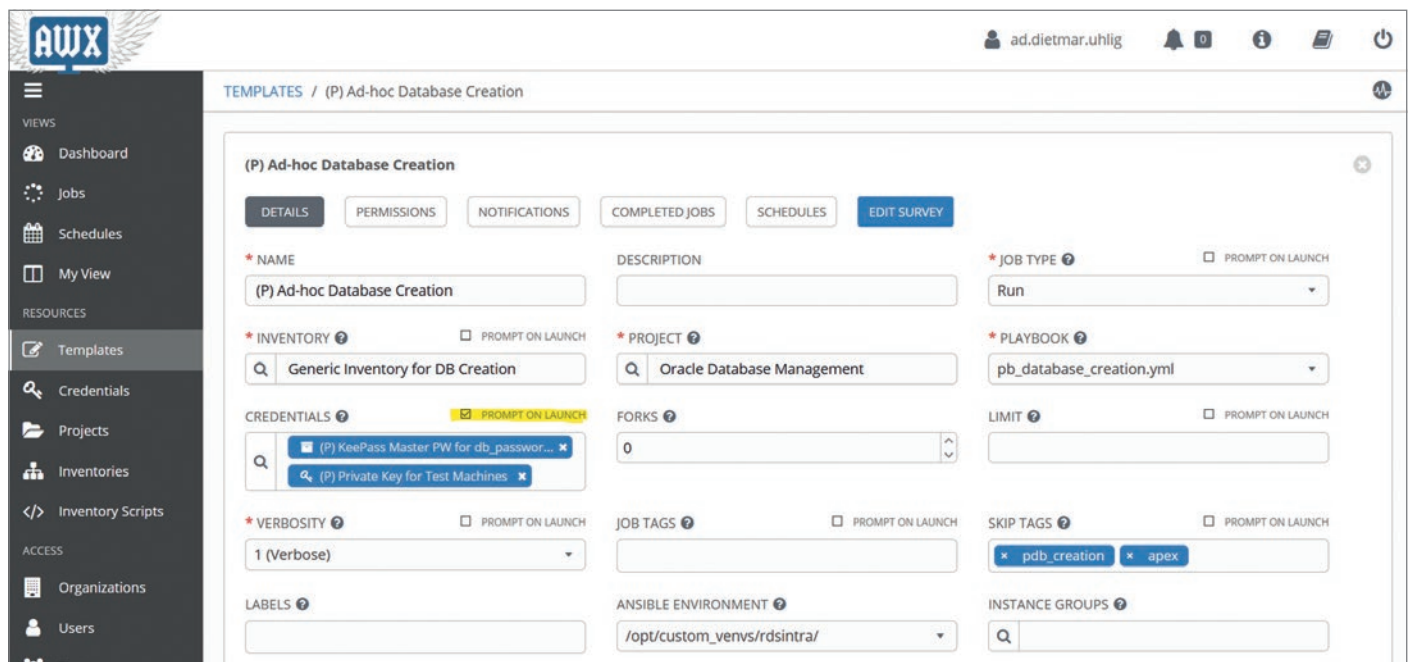


Abbildung 3: Definition eines Templates (Quelle: Dietmar Uhlig)

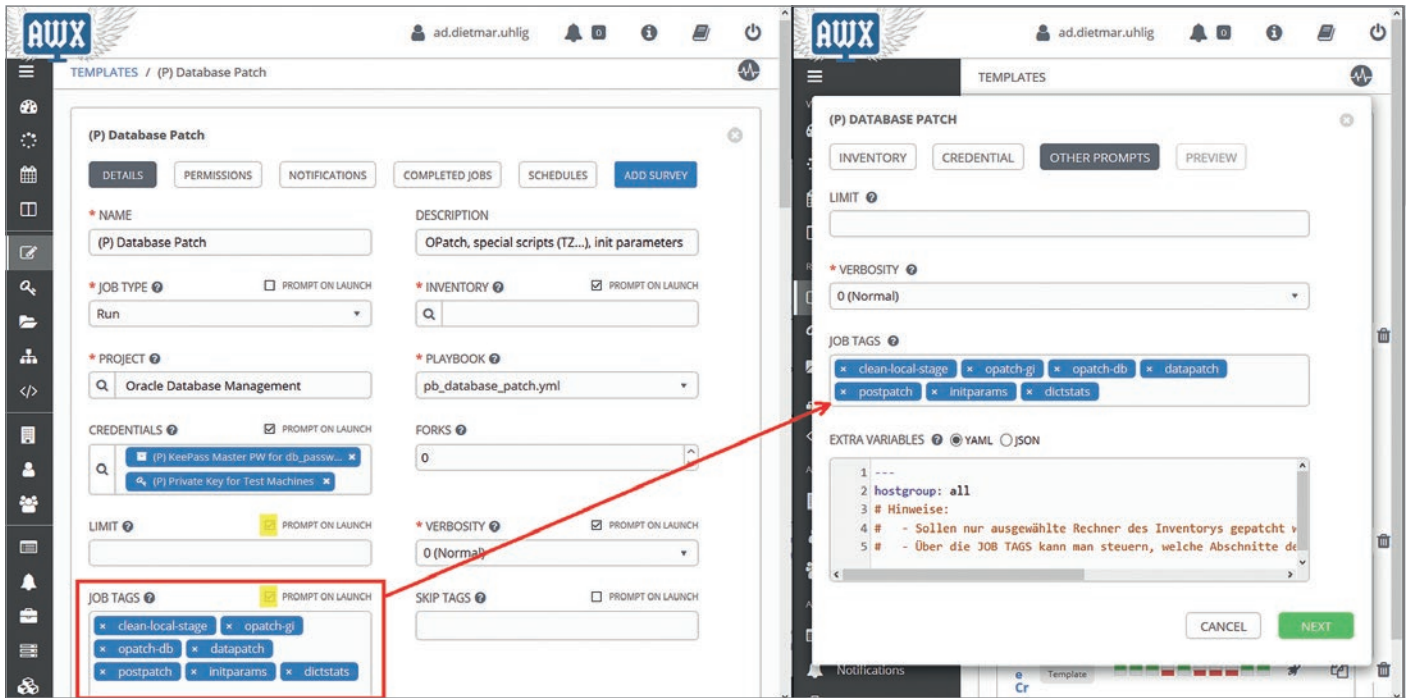


Abbildung 4: Definition eines Templates und Verwendung von Job-Tags (Quelle: Dietmar Uhlig)

vorgesehen. Diese Einschränkungen helfen andererseits, sich auf die wesentlichen Anwendungsfälle zu konzentrieren. Der Anwendungsfall hier ist das Anlegen einer CDB ohne PDBs. *Abbildung 5* zeigt den

Dialog zum Erstellen des Survey. In der rechten Hälfte des Dialogs sieht man die Anordnung der bereits definierten Felder und in der linken Hälfte die Definition eines Felds (hier: des dritten).

Hostname

Normalerweise ist der Rechnername in Ansible fester Bestandteil des Inventory. Es gibt mehrere Möglichkeiten, den Rechnernamen als Variable übergeben

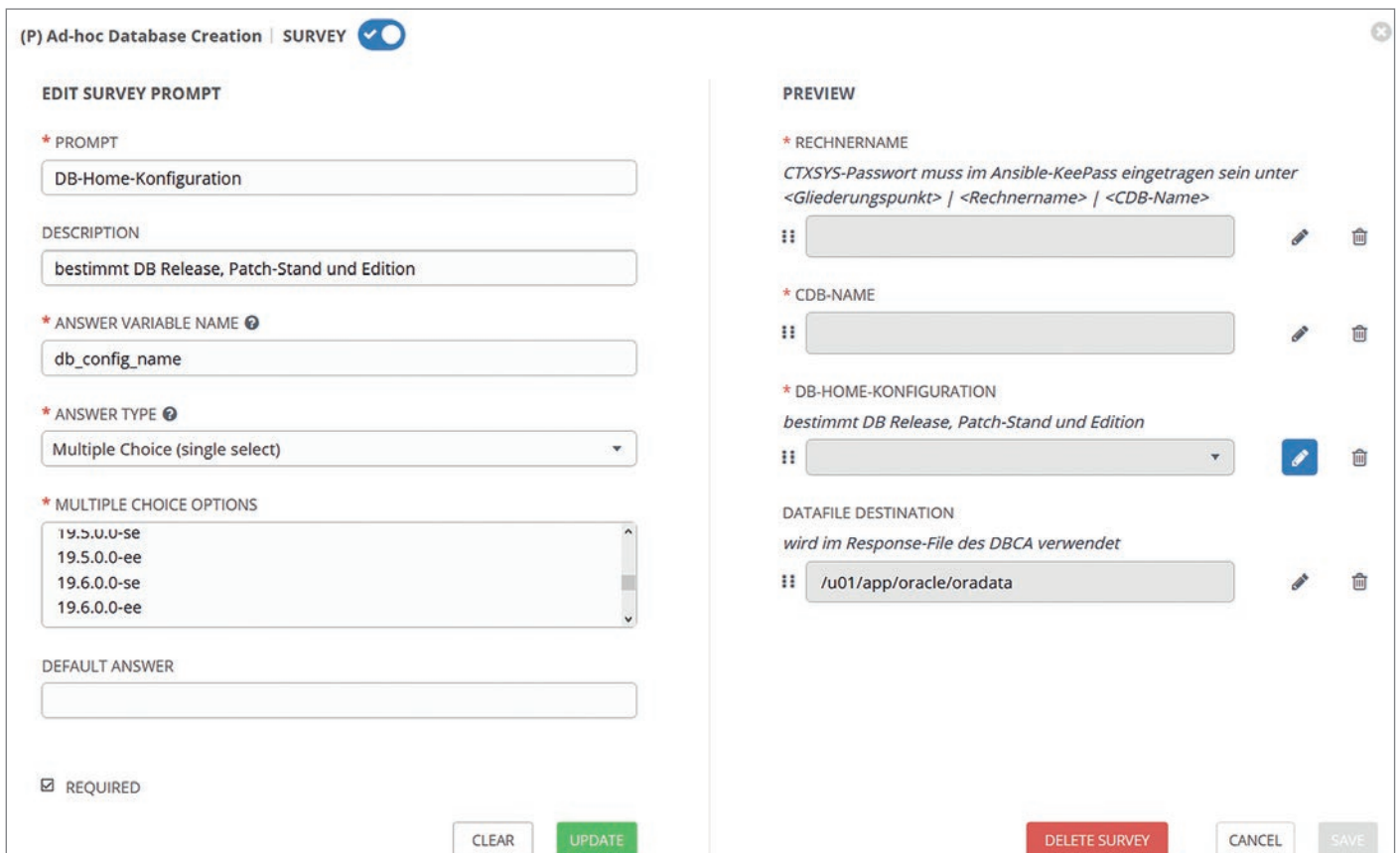


Abbildung 5: Definition einer Drop-Down-Liste im Survey (Quelle: Dietmar Uhlig)

zu können. Eine Variante besteht darin, alle verfügbaren Rechner in einem Inventory aufzulisten und den gewünschten Rechner als limit-Parameter zu verwenden. In unserem Fall wird ein Platzhalter „DB_MACHINE“ in Kombination mit der Variablen `ansible_host` verwendet (siehe Listing 7). Das hat den Vorteil, dass man nicht durch einen ungeschickt übergebenen Wert das Playbook auf zu viele Hosts anwenden kann, aber den Nachteil, dass man Rechner hinter einem Jump Host nicht ohne Weiteres erreicht.

Optionale Variablen

Die Übergabe der Variablen `db_config_name` und `cdb_name` vom Survey an Playbook und Inventory ist trivial. Die Angabe der Datafile Destination sei als lokale Besonderheit und plakatives Beispiel für

Folgendes gestattet: Wenn eine optionale Textvariable ohne Mindestlänge definiert und dann beim Job-Start nicht ausgefüllt wird, übergibt AWX die Variable mit einer leeren Zeichenkette an das Playbook [11]. Wenn man weiterhin die entsprechende Variable aus dem Playbook (einer Rolle...) nur überschreiben will, falls sie im Survey ausgefüllt wurde, muss man noch einen Zwischenschritt einlegen (siehe Listing 8). Dort wird die Eingabe aus dem Survey erst in `datafile_dest_survey` aufgefangen, bevor eventuell `datafile_dest` überschrieben wird.

Listener und GI

Zum Anlegen einer Datenbank muss man den Listener kennen. Hier wird vereinfachend davon ausgegangen, dass entweder eine GI software-only installiert ist oder dass der Listener im DB Home in-

stalliert ist und LISTENER heißt. Mit diesen Einschränkungen lassen sich entweder die Variablen `oracle_home_gi_so` und `oracle_install_version_gi` leicht automatisch ermitteln (siehe Listing 9) oder das Dictionary `listener_installed` festlegen.

Jobstart

Die restlichen Vorarbeiten beschränken sich darauf zu kontrollieren, ob der Oracle-Nutzer und das Oracle-Home auf der gewünschten Zielmaschine existieren, und ein paar Verzeichnisse anzulegen, die von der `ansible-oracle`-Rolle `oradb-manage-db` benötigt werden. Alle diese Vorarbeiten werden in einer Rolle (`oradb-createdb-awx-helper` [12]) gebündelt. Letztendlich sieht das Playbook sehr übersichtlich aus (siehe Listing 10, ohne Tags und Robotron-spezifische Rollen).

```
---
all:
  children:
    rds_db:
      hosts:
        DB_MACHINE:
          ansible_host: "{{ computer_name }}"
          db_homes_installed:
            - home: "{{ db_config_name }}"
              apply_patches: True
              state: present
          oracle_databases:
            - oracle_db_name: "{{ cdb_name }}"
              home: "{{ db_config_name }}"
              state: present
              is_container: True
              storage_type: FS
              oracle_db_type: SI
              oracle_database_type: MULTIPURPOSE
              redolog_size: 200M
              redolog_groups: 3
              archive_log: False
              flashback: False
              force_logging: False
              oracle_db_mem_totalmb: "1024"
              init_parameters: >-
                {{ init_parameters[db_combi[db_config_name]['initparam']] }}
              profiles: "{{ db_profiles['GENERIC'] }}"
              postinstall: >-
                {{ db_postinstall[db_combi[db_config_name]['postinst']] }}
              dbpasswords: "{{ lookup('keepass', '/' + computer_name, '*') }}"
          ...
```

Listing 7: Inventory mit Platzhaltern für das Survey im AWX-Template

```
- name: Optionally set datafile_dest
  set_fact:
    datafile_dest: "{{ datafile_dest_survey }}"
  when: datafile_dest_survey is defined and datafile_dest_survey | length > 0
```

Listing 8: Bedingtes Überschreiben mit einer optionalen Survey-Variablen

```

- name: Check for GI
  stat: path=/etc/oracle/olr.loc
  register: olrloc
  tags: gi_facts

- block:

  - name: Read olr.loc
    slurp: src=/etc/oracle/olr.loc
    register: olrloc_raw

  - name: Read GI home
    set_fact:
      oracle_home_gi_so: >-
        {{ olrloc_raw.content | default('') | b64decode
          | regex_search('^crs_home=.*$', multiline=true)
          | regex_replace('^crs_home=(.*)$', '\1') }}

  - name: Ask for GI version
    shell: "{{ oracle_home_gi_so }}/bin/crsctl query has releaseversion"
    environment:
      ORACLE_HOME: "{{ oracle_home_gi_so }}"
    register: crsctl_rv
    changed_when: false

  - name: Get GI version
    set_fact:
      oracle_install_version_gi: >-
        {%- if gi_version_raw | regex_search('^12\.1') -%}
        12.1.0.2
        {%- elif gi_version_raw | regex_search('^12\.2') -%}
        12.2.0.1
        {%- elif gi_version_raw | regex_search('^19') -%}
        19.3.0.0
        {%- endif -%}
    vars:
      gi_version_raw: >-
        {{ crsctl_rv.stdout | regex_replace('^\.*\[.*?\]\.*$', '\1') }}

when: olrloc.stat.exists
tags: gi_facts

```

Listing 9: Ermittlung zweier GI-Variablen, die zur DB-Installation benötigt werden

```

---
- name: Helper Tasks
  hosts: "{{ hostgroup }}"
  become: yes
  become_user: "{{ oracle_user }}"
  roles:
    - oradb-createdb-awx-helper

- name: Database/CDB Creation
  hosts: "{{ hostgroup }}"
  become: yes
  roles:
    - oradb-manage-db

- name: Datapatch and Other Postpatch Scripts (as Timezone Patch)
  hosts: "{{ hostgroup }}"
  become: yes
  roles:
    - oradb-datapatch
    - oradb-postpatch

...

```

Listing 10: Playbook zum Anlegen der Datenbank (verkürzt)

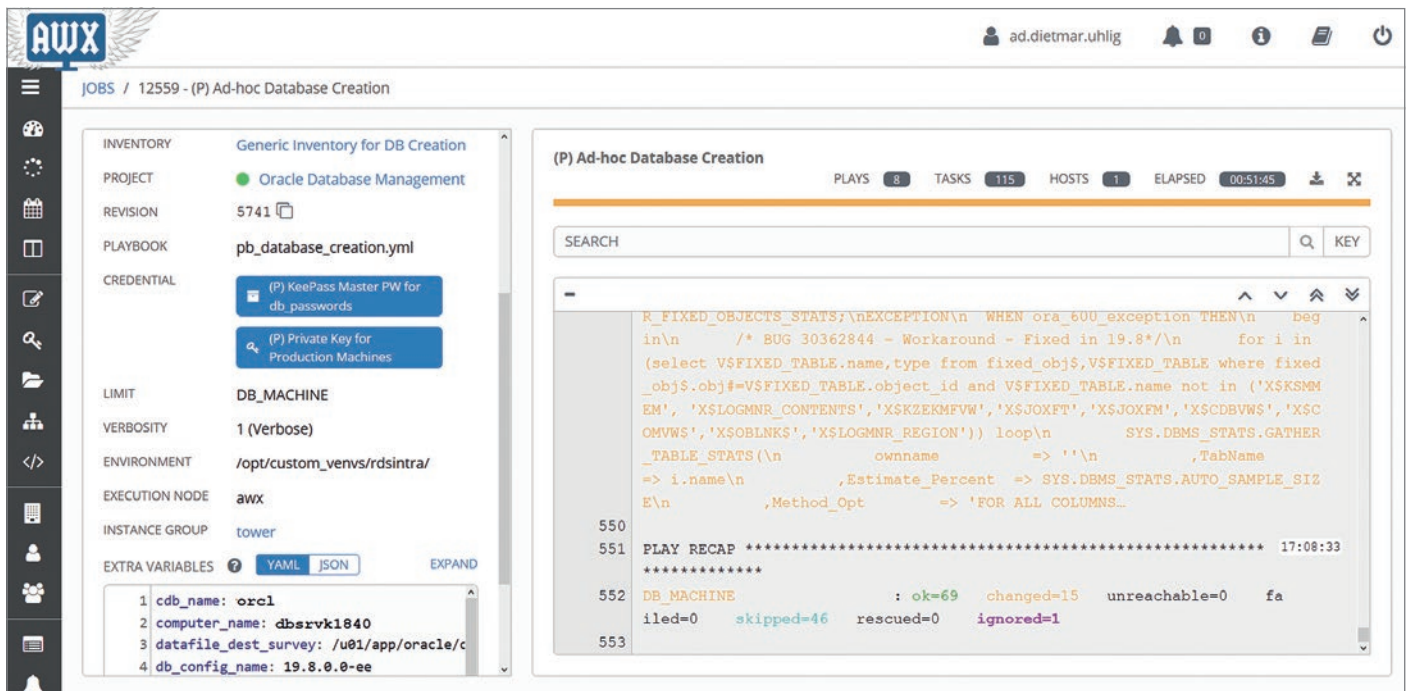


Abbildung 6: AWX-Job, Zusammenfassung (Quelle: Dietmar Uhlig)

Abbildung 6 zeigt die Zusammenfassung des erfolgreichen AWX-Jobs. Links unten im Bild sieht man die in Extra Variables übernommenen Survey-Felder. Auf der rechten Seite werden die Ausschriften von Ansible angezeigt. Bei den Credentials taucht ein Master Password für eine KeePass-Datenbank auf.

Passwörter

Es ist sicherlich aufgefallen, dass das SYS- und das SYSTEM-Passwort, die zum Anlegen einer Datenbank ebenso gebraucht werden wie der Datenbankname, bisher mit keinem Wort erwähnt wurden. Die zugehörigen Felder ließen sich leicht im Survey ergänzen. Im heutigen Beispiel werden jedoch alle Datenbank-Passwörter in KeePass-Datenbanken verwaltet. Das ist wahrscheinlich genauso sicher wie Ansible Vault, aber anwenderfreundlicher.

Eine KeePass-Datenbank ist eine Datei, die wie alle Playbooks, Rollen und Invento-

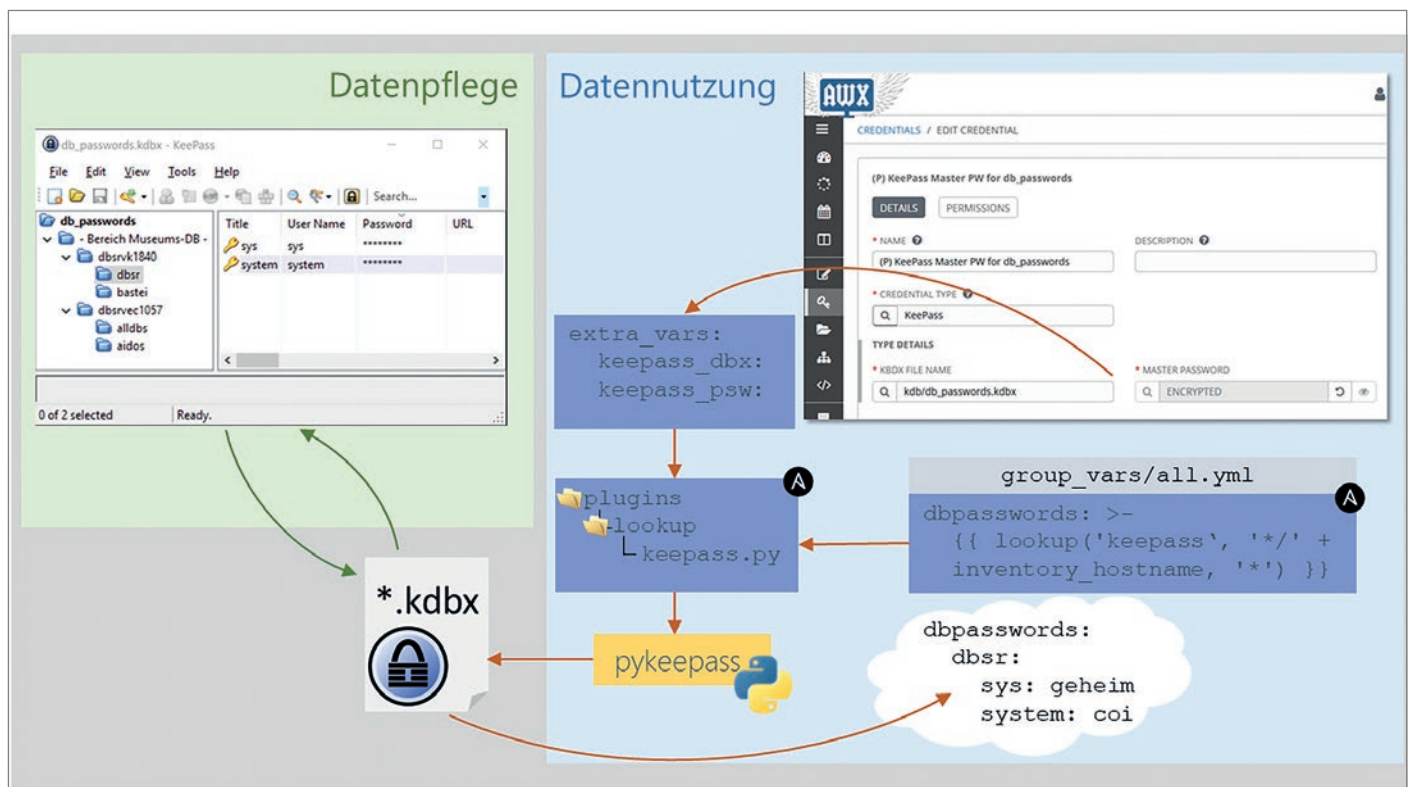


Abbildung 7: Prinzip der Verwendung einer KeePass-Datenbank (Quelle: Dietmar Uhlig)

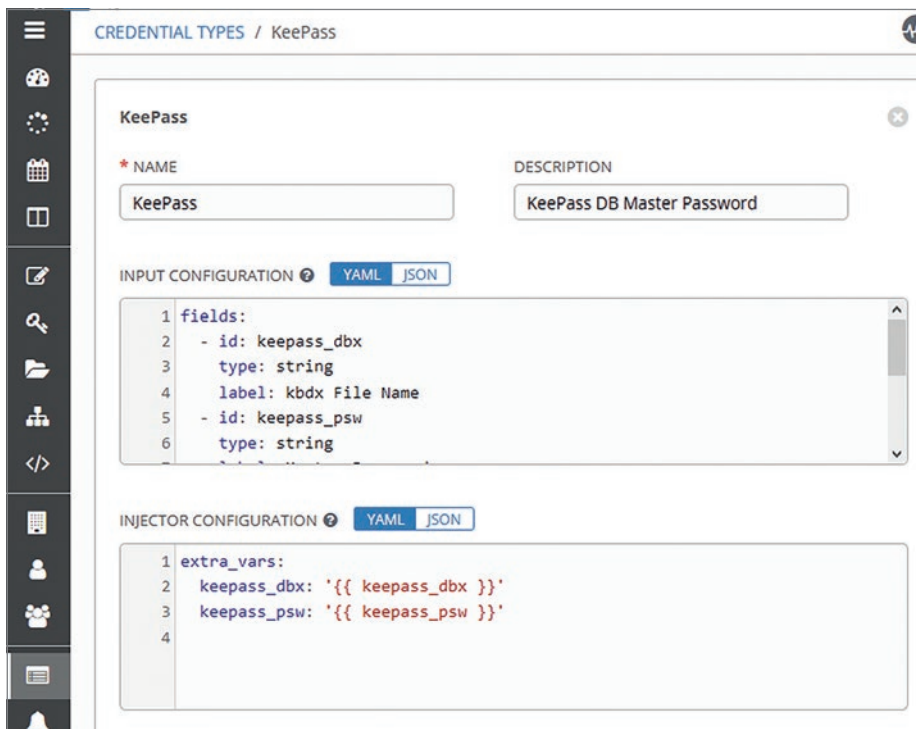


Abbildung 8: Credential Type für KeePass-Master-Passwort (Quelle: Dietmar Uhlig)

ries in Subversion abgelegt wird. KeePass kann zum Glück sehr gut damit umgehen, wenn man nach dem Eintragen eines neuen Passworts merkt, dass sich die Datei in der Zwischenzeit in der Versionsverwaltung geändert hat und jetzt zwei Dateien zusammengeführt werden müssen. Die KeePass-Datenbank muss so strukturiert werden, dass sich alle Hostnamen auf einer Ebene befinden und alle CDB-Namen direkt darunter (siehe Abbildung 7). Das Plug-in `keepass.py` ist ein eigener Fork [13] des Ansible-KeePass-Lookup-Plug-ins.

Um das KeePass-Master-Passwort in AWX abzulegen, definiert man einen eigenen Credential-Typ (siehe Abbildung 8) [14]. Listing 11 gibt die vollständige Input Configuration wieder. Die Verwendung sieht man in Abbildung 7. Der Dateiname

```
fields:
  - id: keepass_dbx
    type: string
    label: kbdx File Name
  - id: keepass_psw
    type: string
    label: Master Password
    secret: true
required:
  - keepass_dbx
  - keepass_psw
```

Listing 11: Input Configuration des Credential Types KeePass

der KeePass-Datenbank enthält den Pfad relativ zum Playbook. AWX übergibt den Dateinamen und das Master-Passwort als Extra Variables an Ansible. In den globalen Gruppenvariablen (`all.yml`) wird das Lookup-Plug-in so aufgerufen, dass es in der zweiten Hierarchieebene nach dem Hostnamen sucht, wie er im Inventory steht, und den darunterliegenden Baum mit Nutzernamen und Passwörtern der Variablen `dbpasswords` zuweist. Genau in dieser Form erwarten viele Rollen aus `ansible-oracle` die Datenbank-Passwörter.

Fazit

Mittlerweile unterstützen AWX und `ansible-oracle` die Routinetätigkeiten der hausinternen Datenbankadministratoren schon sehr gut. Immer wieder gibt es Verbesserungsvorschläge zur Benutzerführung und Erweiterungswünsche bezüglich des Funktionsumfangs der Playbooks. Das hier gezeigte Beispiel lässt mit Absicht noch genügend Raum für eigene Ideen. Doch auch in Ansible und AWX ist nach wie vor viel Bewegung. Es kostet zwar einige Kraft, auf der Höhe der Zeit zu bleiben, aber die Mühe lohnt sich: Die Automatisierung verschafft dem DBA Freiraum, sich weiterzubilden oder spezielle Probleme gründlicher zu untersuchen.

Quellen

- [1] Robotron kann und macht noch viel mehr. Schauen Sie einfach vorbei auf www.robotron.de
- [2] <https://github.com/oravirt/ansible-oracle>
- [3] <https://github.com/oravirt/ansible-oracle-modules>
- [4] <https://docs.ansible.com/ansible-tower/latest/html/userguide/index.html>
- [5] <https://github.com/ansible/awx/blob/devel/INSTALL.md>
- [6] <http://svnbook.red-bean.com/de/1.7/svn.advanced.externals.html>
- [7] https://docs.ansible.com/ansible/latest/collections/ansible/builtin/import_playbook_module.html
- [8] https://de.wikipedia.org/wiki/Configuration_Management_Database
- [9] https://docs.ansible.com/ansible/latest/user_guide/intro_inventory.html#how-variables-are-merged
- [10] https://docs.ansible.com/ansible-tower/latest/html/userguide/job_templates.html#surveys
- [11] https://docs.ansible.com/ansible-tower/latest/html/userguide/job_templates.html#optional-survey-questions
- [12] Die Rolle wird hier nicht abgedruckt. Wesentliche Teile finden sich in Listing 8 und Listing 9.
- [13] https://github.com/duhlig/ansible-keepass/tree/pr_get_unpw_dict
- [14] https://docs.ansible.com/ansible-tower/latest/html/userguide/credential_types.html#create-a-new-credential-type

Über den Autor

Dipl.-Informatiker Dietmar Uhlig arbeitet seit über 20 Jahren bei Robotron. Seit fast 3 Jahren befasst er sich mit Ansible und AWX und macht die Automatisierung zunächst für die hauseigenen DBAs nutzbar, bevor er Kunden berät.



Dietmar Uhlig
dietmar.uhlig@robotron.de



„Flexibilität as a Service“ Postgres im DevOps-Ökosystem

Jan Karremans & Borys Neselovskyi, EnterpriseDB

Schon in den 90er Jahren des vorigen Jahrhunderts arbeitete Dr. Michael Stonebraker an der Berkeley Universität an der Entwicklung einer Datenbank namens Post-Ingres. Zu dieser Zeit war von DevOps, Infrastruktur as a Code oder agiler Entwicklung noch keine Rede. Die Idee von einer schlanken Datenbank, die wie ein Baukasten funktioniert und durch aktivierte Eigenschaften und Funktionen erweitert werden kann, sorgte damals noch für Kopfschütteln. Mit einer Erweiterungsschnittstelle wollte Stonebraker die Datenbank so flexibel und variabel gestalten, dass sie in jedes Anforderungsschema passt. Eine revolutionäre Idee! Denn während Anwendungen heute fast immer modular aufgebaut sind, waren Datenbanken damals als Monolithen eher starr und in sich geschlossen gestaltet. Post-Ingres war also ein wichtiger Schritt auf dem Weg zur modularen Datenbanklösung. Von Anfang an modular aufgebaut, ist sie auch heute noch das Gegenteil eines Monolithen.

Postgres – von Anwendern für Anwender

Aus Post-Ingres entwickelte sich die heutige Postgres-Datenbank. Postgres wird mittlerweile als Open-Source-Lösung von einer aktiven Community gebaut und getrieben. Eine Datenbank von Anwendern für Anwender. Und die haben ein Auge darauf, dass Installation, Konfiguration

und Wartung sogar von Anfängern umgesetzt werden können.

Seit mehreren Jahren geht Postgres zudem den Weg in Richtung DevOps. Und damit kommen wir zum eigentlichen Thema dieses Artikels: „Flexibilität as a Service“. Wie können wir eine Postgres-Datenbank automatisiert aufsetzen und betreiben? Und wie handeln wir in Hochverfügbarkeitsszenarien? Wir zeigen Ih-

nen die wichtigsten Schritte und stellen hilfreiche Tools vor.

Flexibles Datenmanagement

Die Postgres-Datenbank ist mit ihren Modulen und ihrem Fokus auf einen einfachen Betrieb nahezu prädestiniert für den Aufbau eines Datenbankmanagementsystems, bei dem es darum geht, sich immer wieder neu an die Anfor-

derungen des Markts anzupassen. Als Open-Source-Datenbank passt sie zudem perfekt in die DevOps-Welt. Das macht sie bei Endanwendern, Entwicklern und Datenbankspezialisten gleichermaßen beliebt. Spannend sind aber vor allem die Möglichkeiten für die Automatisierung der Datenbankinstallation.

Postgres-Installation automatisieren – so gehts

Installation und Betrieb

Wer eine Postgres-Datenbank zum ersten Mal installiert, wundert sich, wie einfach und schnell eine komplette Datenbankumgebung bereitgestellt werden kann. Die Installationsmedien für Postgres umfassen nur einige Megabytes. Die Provisionierung und Konfiguration eines Postgres Cluster ist in wenigen Schritten erledigt.

Und tatsächlich geht der Betrieb relativ leicht von der Hand: Immer mehr Unternehmen betreiben ihre IT-Landschaften nach DevOps-Prinzipien. Entwickler und DBAs beschreiben Betriebsaufgaben in binärem Code und entwickeln Routinen für die automatisierte Bereitstellung und das Konfigurationsmanagement von Systemkomponenten.

Unternehmenskritische Daten und Hochverfügbarkeit

Datenbanken nehmen eine zentrale Rolle in IT-Landschaften ein, weil sie unternehmenskritische Daten aufbewahren. Daher müssen sie auch ausfallsicher gestaltet werden. Für Postgres ist dies kein großes Hindernis: Mithilfe von Streaming Replication sorgen die Datenbanksysteme für eine hohe Verfügbarkeit. Dafür werden Daten der Master-Datenbank auf eine oder mehrere Standby-Datenbanken repliziert. Fällt die Master-Datenbank aus, übernimmt die Standby-Datenbank die Masterrolle.

Doch das ist noch nicht alles: Hochverfügbare Installationen bestehen aus mehreren Servern. Diese können in einem Rechenzentrum oder geografisch getrennt aufgebaut werden. Der Automatisierungsprozess sollte in der Lage sein, hochverfügbare Systeme zu provisionieren. Jedes Datenbanksystem muss dazu gesichert und überwacht werden: Diese Aufgaben sollen nach Möglichkeit in Code umgewandelt werden.

Das klingt jetzt schon komplizierter: Das Erstellen einer Datenbank, die hochverfügbar ausgelegt werden muss, dazu noch die Konfiguration der Sicherung und die Einrichtung der Überwachung?

Und das komplett automatisiert? Gibt es Tools, die die Umsetzung solcher Anforderungen unterstützen?

Sehr gut für die Automatisierung der Installation von hochverfügbaren Datenbank-Umgebungen eignet sich Ansible. Das Framework stellt eine Plattform zur Verfügung, mit der Prozesse automatisiert und die Infrastruktur konfiguriert werden können. Mit Ansible können Administratoren außerdem Software verteilen, Ad-hoc-Befehle ausführen und die Konfiguration nachträglich verändern.

Automatisierung mit dem Ansible Playbook

Wie funktioniert die Automatisierung mit Ansible? Ansible führt sozusagen ein Drehbuch: Die Installations- und Konfigurationstätigkeiten werden in einem „Playbook“ festgehalten. Darin können alle Tasks, von der kleinen Systemänderung bis zur kompletten Installation eines Servers, inklusive Konfiguration der Anwendung beschrieben werden.

Allerdings kann es sehr aufwendig sein, ein Ansible Playbook für die Postgres-Installation zu erstellen. Hier hilft ein Blick in die Community: Auch Kollegen von uns haben bereits mehrere Playbooks entwickelt und im GitHub Reposi-

Role Name	Description	Last Commit
autotuning	Autotuning: configure vm.overcommit_ratio	last month
init_dbserver	Enable PG service	19 days ago
install_dbserver	Apply playbook changes in the README files	2 months ago
manage_dbserver	Adapted db template reference to allow for value substitution	25 days ago
manage_pgouncer	Adding default value for wal_keep_segments/wal_keep_size in init_dbse...	20 days ago
manage_pgpool2	Adding ability to create database user with the deployment and pcp us...	20 days ago
setup_barman	Preserve ownership of known_hosts file by becoming postgres/barman user.	14 days ago
setup_barmanserver	Adding default value for wal_keep_segments/wal_keep_size in init_dbse...	20 days ago
setup_efm	PGDATABASE value was missing in the pcp_attach_all.sh and pg_pcp_heal...	14 days ago
setup_pemagent	Adding default value for wal_keep_segments/wal_keep_size in init_dbse...	20 days ago
setup_pemserver	Update README.md	20 days ago
setup_pgouncer	Adding default value for wal_keep_segments/wal_keep_size in init_dbse...	20 days ago
setup_pgpool2	Making sure backend_application_nameX taking the PGAPPNAME used in st...	13 days ago
setup_replication	Adding default value for wal_keep_segments/wal_keep_size in init_dbse...	20 days ago
setup_repo	Apply playbook changes in the README files	2 months ago

Abbildung 1: Ansible-Rollen im GitHub Repository von EnterpriseDB (Quelle: Enterprise DB)

tory der Firma EnterpriseDB abgelegt [1]. Solche Routinen können Administratoren als Grundlage für die automatisierte Bereitstellung einer Postgres-Umgebung nutzen. Die Nutzung des Ansible Repository der EnterpriseDB ist kostenlos. Das Repository kann über git mit einem einfachen Kommando heruntergeladen werden:

```
git clone git@github.com:EnterpriseDB/edb-ansible.git
```

In dem Repository sind einzelne Tasks als Rollen definiert. *Abbildung 1* bietet einen Überblick über die Rollen, die im Repository verfügbar sind.

Jede Rolle beschreibt genau einen Installations- oder Konfigurationsprozess. Um die Installation einer neuen Postgres-Umgebung inklusive Binaries kümmert sich die Rolle `install_dbserver`. Will man eine Postgres-Datenbank bereitstellen, kann die Rolle `init_dbserver` verwendet werden. Die Konfiguration der Datenbanksicherung kann mithilfe der Rollen `setup_barman` und `setup_barmanserver` erledigt werden. Die Postgres-Hochverfügbarkeit, die mit dem EDB Failover Manager realisiert wird und auf der Replikation zwischen Master und Standby basiert, kann über die Rollen `setup_replication` und `setup_efm` umgesetzt werden. Der Postgres Enterprise Manager (kurz PEM) bietet eine Plattform zur Überwachung und Administration von großen Postgres-Landschaften. Die Rolle `setup_pemserver` sorgt dafür, dass der PEM automatisiert installiert wird. Damit der PEM Agent auf jedem Datenbankserver läuft, kann er über die Rolle `setup_pemagent` entsprechend konfiguriert werden.

Die erweiterte Konfiguration eines Postgres-Servers, die sich mit Dingen wie dem Connection Pooling oder dem Load Balancing zwischen Datenbank und Endnutzer beschäftigt, kann mit den Rollen `manage_pgpool2` oder `manage_pgpool2` umgesetzt werden. Im besagten Playbook werden die Rollen zusammengefasst und „en bloc“ auf einer Reihe zuvor definierter Hosts ausgeführt. Abgespeichert wird das Playbook als YAML-Datei.

Die Datei hätten wir. Und was geschieht jetzt damit? Während der Ausführung eines Playbooks verbindet sich Ansible über Secure Shell mit den Servern, die verwaltet werden sollen, und führt die In-

```
- hosts: all
  name: Postgres deployment playbook
  become: yes
  gather_facts: yes

# When using collections
# collections:
# - edb_devops.edb_postgres

pre_tasks:
- name: Initialize the user defined variables
  set_fact:
    pg_version: 13
    pg_type: "PG"
    yum_username: ""
    yum_password: ""
    disable_logging: false

roles:
- role: setup_repo
- role: install_dbserver
- role: init_dbserver
- role: autotuning
- role: setup_replication
- role: setup_efm
- role: setup_pemserver
- role: setup_pemagent
```

Listing 1: Playbook mit Routinen zur Installation und Konfiguration einer Postgres-Umgebung

stallations- und Konfigurationsaufgaben der Reihe nach aus. Das nachfolgende Playbook zeigt die Routinen zur Installation und Konfiguration einer Postgres-Umgebung (siehe Listing 1).

Das Playbook führt folgende Schritte aus:

1. Installation und Konfiguration eines Postgres-Advanced-Servers, der mit einer Master- und einer Standby-Datenbank ausgestattet ist.

2. Installation und Konfiguration des EDB Failover Manager, der für den automatisierten Failover zuständig ist, falls die Master-Datenbank ausfällt.
3. Installation und Konfiguration des Postgres Enterprise Manager, der für die Überwachung der Postgres-Umgebung zuständig ist.

Abbildung 2 zeigt, wie eine mit Ansible installierte Postgres-Umgebung in der Praxis aussehen könnte.

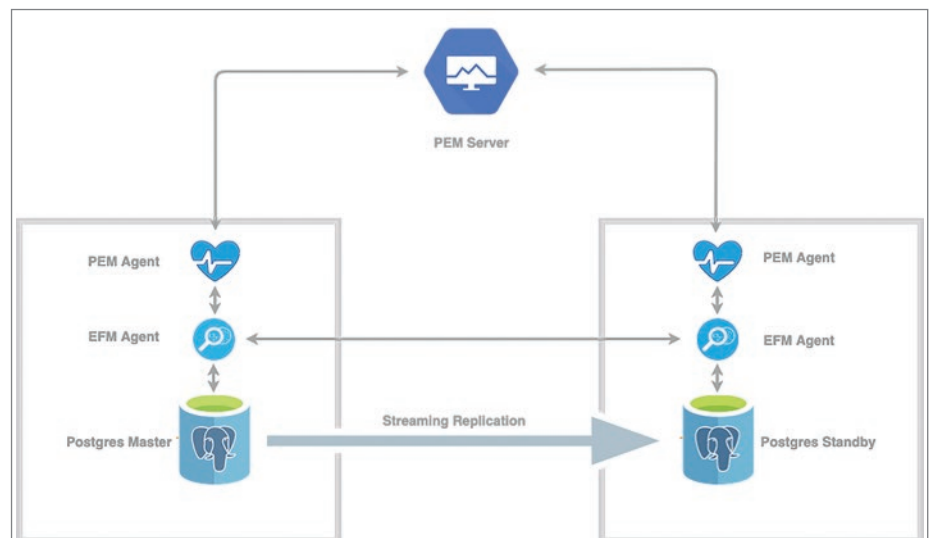


Abbildung 2: Blue Print einer mit Ansible installierten Postgres-Umgebung (Quelle: Enterprise DB)

Ansible setzt voraus, dass die initiale Serverinstallation bereits erfolgt ist. Tools, die für die Bereitstellung von Servern und Netzwerkkomponenten zuständig sind, lassen sich mit Ansible gut kombinieren. So können komplette Postgres-Landschaften sowohl in Kundenrechenzentren als auch in Cloud-Umgebungen eingerichtet werden.

EDB Postgres Deployment Tool

Ein weiteres Tool, das die Provisionierung von kompletten Postgres-Umgebungen in der PublicCloud ermöglicht, ist das Postgres Deployment Tool von EnterpriseDB (EDB). Das Werkzeug wurde von Mitarbeitern des Unternehmens entwickelt und steht auf dem EDB GitHub Repository offen zum Download bereit [2]. Dieses Tool nutzt Ansible und Terraform [3] und soll vor allem dabei helfen, hochverfügbare Umgebungen in der Cloud bereitzustellen. Das Deployment Tool unterstützt gängige Cloud-Anbieter wie AWS, Azure und Google Cloud.

Warum passt Postgres so gut in Cloud-native Umgebungen?

Viele Unternehmen automatisieren die Installation und Konfiguration von Datenbankumgebungen heute mithilfe von Container-Technologie. Dank ihrer Leichtgewichtigkeit passt Postgres hervorragend zu Microservices und wird oft in Container-Umgebungen ausgeführt. Hier dient Postgres meist als Backend für Webanwendungen, analytische Applikationen oder IoT-Landschaften.

Container verbrauchen im Vergleich zu virtuellen Maschinen weniger Ressourcen und können über Orchestrierungsplattformen wie Kubernetes [4] oder OpenShift [5] leicht verwaltet werden. Fällt ein Container aus, stellt die Verwaltungsebene eine Ersatzinstanz bereit. Mit Kubernetes können Anwendungen einfach skaliert werden, indem weitere Instanzen gestartet werden, sobald sich die Arbeitslast erhöht.

EDB hat unter dem Label Cloud Native Postgres einen Kubernetes Operator und Container Images entwickelt [6]. Der Operator kümmert sich um die automatische Installation einer Postgres-Umgebung in Kubernetes. Seine wichtigsten Eigenschaften sind:

- die Realisierung der Hochverfügbarkeit über Mechanismen des Kubernetes Cluster
- eine „Selbsteheilungsfunktion“ durch Failover und automatische Wiederherstellung von Replikaten
- die Unterstützung von synchroner und asynchroner Replikation
- das Kapazitätsmanagement mit Scale-up/-down-Funktionen
- geplante Umschaltungen für Wartungsarbeiten
- die Definition von Services für Lese- und Schreibzugriff
- Rolling Updates für Minor-PostgreSQL-Versionen und Operator Upgrades
- TLS-Verschlüsselung und Client-Zertifikat-Authentifizierung
- die kontinuierliche Sicherung und Point-in-Time-Recovery der Datenbanken

Mithilfe des Kubernetes Operator kann ein Postgres Kubernetes Cluster in Rechenzentren installiert werden. Aber auch in der Cloud: Der Operator unterstützt die Provisionierung einer Postgres-Umgebung in den Kubernetes Services der gängigsten Cloud-Anbieter. Damit lässt sich ein Postgres Kubernetes Cluster sehr bequem in der Google Kubernetes Engine (GKE), im Azure Kubernetes Service (AKS) oder auch im Amazon Elastic Kubernetes Service (EKS) provisionieren und betreiben. Und für den ein oder anderen gut zu wissen: Der EnterpriseDB Kubernetes Operator ist auch für den Betrieb auf der Plattform OpenShift von Red Hat zertifiziert.

Fazit

Mit Automatisierungswerkzeugen von EDB kann fast jede Postgres-Architektur in allen Serverformen installiert werden: physikalisch, virtualisiert und in Containern. Ob in der Cloud oder in eigenem Rechenzentrum: Der Betrieb von Postgres-Landschaften kann damit sehr flexibel gestaltet werden. Insbesondere die Ansible-Routinen und die Cloud-Native-Plattform bieten eine solide Grundlage für den Aufbau der Postgres-Landschaft, die den Anforderungen an eine moderne Architektur gerecht wird. Der Begriff „Flexibilität as a Service“ passt insofern sehr gut zu Postgres und spiegelt die besondere Flexibilität beim Aufbau und Betrieb der Datenbank wider.

Quellen:

- [1] <https://github.com/EnterpriseDB/edb-ansible>
- [2] <https://github.com/EnterpriseDB/postgres-deployment>
- [3] <https://www.terraform.io/>
- [4] <https://de.wikipedia.org/wiki/Kubernetes>
- [5] <https://de.wikipedia.org/wiki/OpenShift>
- [6] <https://www.enterprisedb.com/products/postgresql-on-kubernetes-ha-clusters-k8s-containers-scalable>



Jan Karremans
jan.karremans@enterprisedb.com



Borys Neselevskiy
borys.neselevskiy@enterprisedb.com



Der DBA als Schlangenbeschwörer – bye-bye Bash, hi Python!

Katharina BÜchse, EQUIcon Software Jena

Eigentlich hatte die Bezeichnung für Python ursprünglich gar nichts mit Schlangen zu tun, sondern bezog sich auf die britische Komikertruppe Monty Python [1]. Wie sich jedoch am Logo unschwer erkennen lässt, hat sich diese Interpretation gewandelt. Und so wird der DBA, der sich, getrieben von der Suche nach einer guten Programmiersprache zur Automatisierung seiner Aufgaben, mit Python beschäftigt, unweigerlich zum Schlangendompteur. Da diese Tätigkeit allerdings nicht ganz ungefährlich ist, DBAs in Zeiten des Fachkräftemangels aber gebraucht werden, liefert Ihnen dieser Artikel einige Tipps, damit Sie bei der Arbeit mit den riesigen Würgeschlangen nicht erdrosselt werden.

Auch wenn unter den ITlern das Thema Datenbanken oft als „in die Jahre gekommen“ gilt, so arbeitet man (und natürlich auch frau und alle anderen) als Datenbankadministrator natürlich auf der Höhe der Zeit und ist an der Automatisierung möglichst vieler seiner Aufgaben interessiert. Dabei fällt die Wahl der „Programmiersprache“ in der Regel auf die beziehungsweise das Nächstliegende: Shell-Skripte, meist in Kombination mit Crontab. Das mag unter anderem daran liegen, dass Oracle eine gewisse Vorliebe zu Linux nicht verleugnen

kann. Zudem hat man das schließlich auch schon immer so gemacht, irgendwie mussten doch auch vor 30 Jahren schon Backups angestoßen oder Datenbanken repliziert werden. Und selbst, wenn Sie dieses „vor 30 Jahren“ gar nicht miterlebt haben, weil Sie damals noch Sandburgen gebaut und mit Bauklötzern gespielt haben, so begegnen Ihnen Shell-Skripte trotzdem immer wieder. Schließlich gibt es genügend Kollegen, die nicht müde werden zu erklären, wie bereits im vorigen Jahrtausend Datenbankaufgaben erledigt wurden.

Shell-Skripte – die Herausforderung

Nun ist es so, dass Shell-Skripte für die zunehmende Komplexität der Aufgaben möglicherweise nicht hundertprozentig gut geeignet sind. Um höhere Komplexität kommen Sie dennoch nicht herum, außer Sie haben Lust, ganze Farmen an Skripten zu verwalten, in denen die Skripte alle ungefähr dasselbe tun und am besten noch mit Copy & Paste entstanden sind. (Wenn das auf Sie zutreffen sollte, brauchen Sie die-

sen Artikel nicht weiter zu lesen.) Ich selbst finde echte Bauernhofstätigkeiten durchaus spannend, bei Skripten liegt mir (und allen meinen Kollegen) der geringere Verwaltungsaufwand aber mehr am Herzen. Das bringt allerdings einige Herausforderungen mit sich, weil man dann nämlich versucht, mit einer reinen Skriptsprache Programme zu schreiben. Im Endeffekt führt das dazu, dass Sie sich geraume Zeit auf einschlägigen Seiten wie Stack Overflow, Tutorialspoint oder YouTube aufhalten oder den lieben Kollegen Löcher in den Bauch fragen. Doch selbst der gewiefteste Shell-Guru wird möglicherweise nicht bei jedem Skript-Aufruf schlüssig erklären können, wieso das Skript nicht das tut, was es soll. Das mag daran liegen, dass Sie im 25. Aufruf von Funktion „xyz“ vergessen haben, um die Übergabeparameter Anführungsstriche zu setzen. Oder Sie haben in irgendeiner Funktion eine Variable nicht als „local“ definiert, finden es jedoch aufgrund des Wiedererkennungswerts gut, wenn Variablen mit demselben Inhalt auch gleich heißen...

Mein Weg zu Python

Da mich persönlich diese Erfahrungen sehr frustrierten und ich zudem das Gefühl hatte, dass ich mehr Zeit damit verbringe, die Irrwege des Shell-Skripts zu begreifen, anstatt wirklich produktiv tätig zu werden, zog ich eines Tages einen Schlussstrich. Selbst mein Onkel stieg schon als kleiner Bub auf die Kanzel, um zu verkünden, dass alles anders werden müsse. Genau das dachte ich mir auch. Ich brauchte eine (Skript-) Sprache, die einfach zu erlernen, weit verbreitet (und somit hoffentlich auch gut dokumentiert) und zudem möglichst vielseitig war. Meine Wahl fiel ziemlich schnell auf Python, trotz einiger Gegenargumente unter der Kollegenschaft. Da ich nun passenderweise auch in einer eher kleinen, aber durchaus breit aufgestellten Firma arbeite, in der man auf dem Gang regelmäßig Leute trifft, die sich mit völlig anderen Dingen beschäftigen als man selbst, war auch schnell ein Guru für die von mir angestrebte Alternative gefunden.

Aller Anfang ist ... aufregend!

Die größte Herausforderung ist wohl, wie so oft, wenn man etwas Neues beginnt,

dass man sich in komplett neue Dinge einarbeiten und Zeit damit verbringen muss, Grundlegendes zu lernen, bevor man überhaupt nur anfangen kann, über konkrete Lösungen der eigenen Probleme nachzudenken. Hier hilft ganz klar die Konzentration auf das „große Ziel“ und die Freude über die üblicherweise anfangs sehr steile Lernkurve. Ich jedenfalls hatte Motivation und Neugier genug, um ganz viel Neues zu lernen, neue Software zu installieren und mit wachem Geist (und Mund-Nasen-Schutz) „meinem“ neuen Guru zu lauschen. Dessen Denkstrukturen waren gefühlt komplett anders als meine, er lieferte mir jedoch quasi mit jedem Satz wertvolle Stichworte, womit ich mich beschäftigen sollte. Diese Stichworte reichten oft schon aus – schließlich gibt es im Netz schier unfassbar viele Möglichkeiten, um Unterstützung beim Lernen zu bekommen. Die YouTube-Tutorials von Corey Schafer [2] kann ich jedem ans Herz legen, der neu in die Thematik einsteigt oder sein Python-Grundlagenwissen auffrischen möchte. Zudem existieren jede Menge schicke Entwicklungsumgebungen, die einem das Leben als Programmierer leichter machen. Ich selbst entschied mich für Visual Studio Code (VSCode), unter anderem weil es das „remote“-Entwickeln auf einem Server mittels SSH-Verbindung unterstützt und weil man damit direkt auf Jupyter Notebooks zugreifen und diese bearbeiten kann. Jupyter Notebooks wiederum waren genau das Richtige für mich, weil sie zum Ausprobieren und Testen von Code-Schnipseln hervorragend geeignet sind. Bezüglich VSCode mag der eine oder andere vielleicht Zahnschmerzen oder Bauchweh bekommen, weil die eigene Arbeit von Microsoft getrackt wird. Aber auch dafür gibt es Abhilfe: VSCodium. (Letzteres war übrigens ein wertvoller Hinweis eines Zuhörers während meines DOAG-Vortrags, für den ich sehr dankbar bin!)

Python und Oracle: Achtung, Stolpersteine!

War bei der allgemeinen Einarbeitung in Python meine Begeisterung noch ungebremst und groß, weil so vieles möglich und das Licht am Ende des Bash-Wirrwarr-Tunnels unglaublich nahe schien, so wurde die Freude doch etwas ausge-

bremst, als ich endlich so weit war, mich meinen eigentlichen DBA-Aufgaben zuzuwenden. Immerhin schreibt man ja leider nicht schicke Automatisierungsskripte nur um deren Schönheit willen, sondern weil man seine DBA-Aufgaben an den Rechner abtreten will (oder muss). Im Klartext bedeutete das, dass ich nun also dieses von mir mit rosaroter Brille betrachtete Python mit dem aufgrund der langen Jahre des Zusammenseins nicht mehr als ganz so glänzend wahrgenommenen Oracle verheiraten musste. Dabei ist zu bedenken, dass heutzutage selbst bei der Entwicklung von Programmiersprachen auf das Gewicht geachtet wird, damit diese ja nicht zu dick werden. Der reine Python-Befehlskatalog ist also relativ schlank gehalten, er lässt sich hingegen leicht erweitern: Objekte und Methoden für spezielle Aufgaben sind in sogenannte Module ausgelagert, die man bei Bedarf einfach importiert.

Das Modul, das für eine echte Python-Oracle-Vermählung notwendig ist, heißt „cx_Oracle“. Dessen Einrichtung verlangt dem einen oder anderen möglicherweise einen langen Atem ab und es könnte die Frage aufkommen, ob das nicht vielleicht doch „irgendwie einfacher“ geht. Tatsächlich gibt es unter Umständen einen kürzeren Feldweg, aber Sie wissen ja, wie das mit Feldwegen so ist: Kaum ist das Wetter ein bisschen schlechter, sind die Klamotten dahin. Deswegen lernen Sie jetzt erst einmal die mit Stolpersteinen zugespaltene, aber saubere Variante namens cx_Oracle kennen.

Stolperstein Nr. 1: Die Python-Version

Der Stolpersteinhaufen zeigt seine ersten Ausläufer bereits in der Tatsache, dass cx_Oracle 8 (derzeit aktuelle Version) nur für die Version Python 3 verfügbar ist. Das verwundert zunächst einmal wenig, schließlich ist Python 3 die neuere Version im Vergleich zu Python 2, deren Support schon im letzten Jahr ausgelaufen ist (das „end-of-life“ wurde für den 01.01.2020 festgelegt). Nun ist jedoch in Oracle Linux 6 und in Oracle Linux 7 Python 2 die Standard-Version. Erst ab OL 7.7 ist Python 3 als „python3“ überhaupt vorinstalliert, bei älteren Versionen muss es nachinstalliert werden. Das ist mit einem einfachen yum-

install-Befehl erledigt, solange man keine großen Ansprüche an die konkrete Version (beispielsweise „mindestens Python 3.6“) hegt. Falls doch, kann man entweder „global“ auf dem Server mittels EPEL-Repository und yum eine etwas neuere Version installieren. Oder man verwaltet unterschiedliche Python-Versionen für einen konkreten Benutzer mit dem Tool pyenv, das etwa unter [3] sehr schön vorgestellt wird. Wenn man nicht mit pyenv arbeitet, kann man sich noch überlegen, ob man für python3 einen Alias einführt oder doch jedes Mal „python3“ eintippt...

Stolperstein Nr. 2: Modulinstallation mit pip/pip3

Ein kleineres Stolpersteinchen, das auch schnell zur Seite geräumt ist, besteht darin, dass der Otto-Normalverbraucher-Python-Entwickler eher nicht mit Oracle arbeitet (wieso eigentlich?) und somit das Modul cx_Oracle erst mit dem Paketmanager pip (bzw. pip3) nachinstalliert werden muss.

Stolperstein Nr. 3: Client-Einrichtung und Umgebungsvariablen

Hat man das erledigt, kann man tatsächlich das Modul cx_Oracle in sein Python-Skript einbinden, allerdings ... in Zeiten von Microservices und Containerisierung wird man sein Skript vielleicht nicht unbedingt direkt auf seinem Datenbankserver

ausführen. Das Python-Modul selbst baut keine Verbindung zu Oracle-Datenbanken auf, dazu benötigt man schon einen Client. Dieser Stolperstein könnte also je nach gewählter Infrastruktur kleinere oder größere Schmerzen bereiten. Die gute Nachricht ist, dass ein Instant Client ausreichend ist. Die nicht so gute Nachricht lautet, dass auf OS-Ebene der LD_LIBRARY_PATH entsprechend gesetzt sein muss. Das kann man global tun (vielleicht keine so gute Idee, könnte zu ungeahnten Komplikationen führen) oder man bastelt ein Shell-Skript um sein Python-Skript (wollten wir nicht gerade von Shell-Skripten wegkommen?) oder man setzt mit patchelf den „runpath“ für alle Dateien im Instant-Client-Verzeichnis (das wird von Oracle jedoch nicht unterstützt). Am besten freundet man sich wohl mit dem Gedanken an, dass man niemals vollkommen ohne Shell-Skripte arbeiten wird, aber die Zeilenanzahl dieser Skripte von nun an auf ein Minimum reduziert wird und die Skripte somit schön und übersichtlich sein werden.

Stolperstein Nr. 4: Dokumentation und Strukturierung der Modul- Elemente

Sind nun alle technischen Voraussetzungen geschaffen, besteht der letzte Schritt darin herauszufinden, welche Objekte und Methoden man überhaupt aus dem cx_Oracle-Modul benötigt, um beispielsweise eine Verbindung zur Oracle-Datenbank

aufzubauen. Als frisch gebackener Python-Einsteiger ist man in diesem Moment vielleicht geneigt, sein neu gewonnenes Wissen anzuwenden und einfach mal die help()-Methode für das Modul aufzurufen, um sich einen Überblick über die im Modul vorhandenen Objekte und Methoden zu verschaffen. Das ist jedoch keine gute Idee, weil help(), angewandt auf cx_Oracle oder cx_Oracle.Connection, bis auf bloße Namen nichts weiter anzeigt. (Versuchen Sie dies zum Beispiel einmal mit „docker.client“ vom Modul Docker, dann merken Sie einen Unterschied.) Und so googelt man sich wieder durch das Internet und findet bei GitHub den kompletten Quellcode in c [4] und auf „readthedocs“ eine ausführliche Dokumentation [5], die allerdings vielleicht nicht hundertprozentig intuitiv ist. Wenn Sie wissen wollen, welche Parameter cx_Oracle.Connection() unterstützt, dann findet sich das unter „Module interface“; die Liste der zu den cx_Oracle.Connection()-Objekten gehörenden Methoden ist dagegen direkt bei „Connections“ aufgeführt. Für den Einstieg gibt es durchaus hilfreiche Beispielfragen, sodass man schnell selbst loslegen kann. Damit Sie jetzt nicht gleich das Heft aus der Hand legen und die Internetseite aufrufen müssen, gibt es zur Veranschaulichung auch hier ein Beispiel.

Ein Beispiel mit cx_Oracle

Jedem/Jederschon mal mittels „echter“ Programmiersprache eine Verbindung zu einer Datenbank aufgebaut, eine Anfrage verschickt und das Ergebnis zurückgeholt hat, wird das nun folgende Beispiel ganz logisch erscheinen. Der Aufbau, wie in Listing 1 dargestellt, ist im Grunde auch ganz einfach:

1. Mittels cx_Oracle.Connection wird die Verbindung aufgebaut. Dazu werden dem Connection-Objekt sämtliche für die Verbindung notwendigen Daten übergeben. Möchte man sich mit sysdba-Rechten anmelden, ergänzt man die Parameterliste noch um „mode=cx_Oracle.SYSDBA“.
2. Als Nächstes wird für die frisch angelegte Verbindung ein Cursor erstellt. Cursor sind notwendig, weil Datenbanken mit Mengen hantieren, Programmierer aber lieber einzelne Elemente „anfassen“ und verarbeiten. Ein Cursor kümmert sich um genau diese „Übersetzung“.

```
import cx_Oracle
import sys
try:
    # 1) Verbindung zur Datenbank erstellen
    myConnection = cx_Oracle.Connection(
        user='c##katha',
        password='***',
        dsn='10.35.78.172:1526/spistby')
    # 2) Cursor anlegen
    myCursor = myConnection.cursor()
    # 3) Anfrage ausführen
    myQueryResult = myCursor.execute(
        'select name, open_mode from v$pdb')
    # 4) Anfrageergebnisse abholen
    for row in myQueryResult:
        print(row)
except cx_Oracle.DatabaseError as myDatabaseError:
    # 5) Fehlerbehandlung
    print(myDatabaseError, file=sys.stderr)
```

Listing 1: Python-Skript mit lesender SQL-Abfrage

3. Der Cursor wiederum besitzt eine Methode namens „execute“, der man eine Anfrage als String übergeben kann. Diese Anfrage wird ausgeführt und das Ergebnis als Liste zurückgegeben (auf die über „myQueryResult“ zugegriffen werden kann).
4. Um sich das Ergebnis anzeigen zu lassen, wird die Liste mittels for-Schleife durchlaufen und jedes einzelne Element ausgegeben.
5. Mit try-except werden Fehler abgefangen. Das Modul cx_Oracle bietet ein eigenes Objekt cx_Oracle.DatabaseError, das man mit einem einfachen Print-Befehl ausgeben lassen kann. Man erhält auf diese Weise tatsächlich die ORA-Meldungen, die von der Datenbank zurückgegeben wurden.

Das so erhaltene Skript kann direkt in einem Jupyter Notebook ausgeführt oder als Python-Skript mit der Endung „.py“ abgespeichert und mit „python3 <Skriptname.py>“ aufgerufen werden.

Der kürzere Feldweg

Hat man erst einmal alle Stolpersteine zur Seite geräumt, ist das Zusammenspiel von Python und Oracle-Datenbank ein ziemlich sorgenfreies Kinderspiel. Allein die Tatsache, sich um die Formatierung der Ergebnisse keine Gedanken mehr machen zu müssen, sondern sofort jedes Element direkt ansprechen zu können, klingt doch verlockend, oder?

Es mag jedoch Situationen geben, in denen dieses spezielle Modul, das so viel kann und das somit in seiner Bedienung auch in gewisser Form erlernt sein will, wie „mit Kanonen auf Spatzen geschossen“ wirkt. Angenommen, Sie haben kein Problem damit, weiterhin SQL*Plus oder ein anderes Konsolen-Tool zu verwenden, um die reine Datenbank-Kommunikation durchzuführen. Meist bestehen Ihre Befehle auch nur aus einem Einzeiler oder wenigen, kurzen Anfragen und Sie loggen Ihr Vorgehen lokal pro (SQL*Plus-)Aufruf.

Unter diesen Umständen hilft Ihnen definitiv „subprocess.run()“ aus dem Modul „subprocess“. Dieser Methode übergeben Sie den Befehl, der vom Betriebssystem ausgeführt werden soll, in Form einer Liste. Vielleicht würden Sie sich an dieser Stelle der Einfachheit halber lieber einen String wünschen, wie es bei os.system() tatsächlich auch der Fall ist, die Liste hat jedoch einen entscheidenden Vorteil: Ihre Elemente werden nicht durch die Shell interpretiert, sondern direkt ausgeführt. Somit werden etwaige Leerzeichen innerhalb der Elemente auch nicht als Trennzeichen „übersetzt“.

Im Endeffekt bedeutet die Verwendung von subprocess.run() nichts anderes, als dass Sie die Kommunikation mit der Datenbank Ihren altbewährten Methoden überlassen, die Sie wahrscheinlich eh schon eingerichtet haben: Sei es das SQL*Plus, das schon längst installiert ist, oder das kurze Bash-Skript, das wiederum SQL*Plus zusammen mit einem SQL-Skript aufruft.

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

Das Tolle ist nun, dass Sie die Liste, die Sie an `subprocess.run()` übergeben, mit Python-Mitteln zusammenstellen können: Hier kann Ihr Programmiererherz höher schlagen, Sie können die kompliziertesten Algorithmen erstellen. Sie können auch mit Python-Mitteln die Bash- und SQL-Skripte vor deren Ausführung anpassen. Aber zum Schluss machen Sie das, was Sie schon immer getan haben: den eigentlichen Aufruf an das Betriebssystem übergeben. Vielleicht klingt das jetzt in Ihren Ohren erst einmal komplizierter oder aufwendiger als die Verwendung von `cx_Oracle`, weil Sie ja trotzdem einen Client (eben z.B. SQL*Plus) brauchen und plötzlich auch noch Gummistiefel anziehen müssen, um bestimmte Schlecht-Wetter-Risiken selbst abzufangen. Wenn Sie jedoch unterschiedliche Tätigkeiten ausführen, beispielsweise auch noch RMAN-Backups verwalten oder andere (Konsolen-)Tools einsetzen, für die es möglicherweise gar kein Python-Modul gibt, dann ist `subprocess.run()` eine herrliche Sache. Es ist äußerst vielseitig und kommt dem gemeinen Datenbankadministrator, der es gewohnt ist, auf der Konsole zu arbeiten, sehr entgegen. Aber vergessen Sie bitte nicht die damit verbundenen Risiken: Sie müssen Fehler, die Sie abfangen wollen, vorher selbst definieren und können nicht wie in *Listing 1* einfach „mal so eben“ auf `cx_Oracle.DatabaseError` zurückgreifen.

Ein Beispiel mit `subprocess.run()`

Um Sie nicht in trockener Theorie stehen zu lassen und damit Sie sehen, dass `subprocess.run()` im Datenbankumfeld tatsächlich wunderbar eingesetzt werden kann, gibt es mit *Listing 2* an dieser Stelle noch einen Skript-Ausschnitt „aus dem wahren Leben“.

In diesem Ausschnitt wird eine Archive-log-Datei innerhalb einer Hochverfügbarkeitsinfrastruktur (mit Standard Edition 2) mittels `asm-copy` von der Primärseite zur Standby-Seite kopiert. Es gibt ein Objekt „`primary_side`“ und ein Objekt „`standby_side`“, die sämtliche Informationen zur jeweiligen „Seite“ enthalten. Zudem wurde der OMF-Name des Archive-log, der in „`archive_log_name`“ enthalten ist, in einer gekürzten Variante (ohne OMF-generierten Anteil) in „`archive_log_name_short`“ abgelegt. Zu Beginn des gesamten Skripts muss

```
<...>
ssh_prefix      = ["ssh", "-t", primary_side.host_name]
asmcmd_prefix  = ["ORACLE_HOME=" + primary_side.oracle_home,
                  "ORACLE_SID=" + primary_side.oracle_sid,
                  primary_side.oracle_home + "/bin/asmcmd"]
asmcopy_command = ["cp",
                   primary_side.file_path + "/" + archive_log_name,
                   standby_side.connection_string + ":" + \
                   standby_side.file_path + "/" + archive_log_name_short]
command_result = subprocess.run(ssh_prefix + asmcmd_prefix + \
                                asmcopy_command, capture_output=True)
<...>
```

Listing 2: Kopieren von ArchiveLogs mit `subprocess.run()`

das Modul „`subprocess`“ mittels `Import`-Befehl eingebunden werden. Für den eigentlichen Kopierbefehl werden nun, wie in *Listing 2* sichtbar, drei Listen erstellt, die sämtliche für den Kopiervorgang benötigten Befehls-elemente enthalten. Sie werden beim eigentlichen Aufruf von `subprocess.run()` in der letzten abgebildeten Zeile zu einer Liste zusammengefügt, die dann mit `subprocess.run()` als ein zusammenhängender Befehl an das Betriebssystem übergeben wird. Durch Setzen von „`capture_output=True`“ kann in einem nächsten Schritt auf die Ausgabe, abgelegt in „`command_result.stdout`“ und „`command_result.stderr`“, zugegriffen und diese kontrolliert werden.

Zusammenfassung

Python ist eine der beliebtesten Programmiersprachen, die aber leider, trotz ihrer zusätzlichen Persönlichkeit als Skriptsprache, noch nicht so richtig in der Welt der Datenbankadministration angekommen ist. Wenn Sie jedoch ernsthaft an der Automatisierung komplexerer Sachverhalte interessiert sind, dann sollten Sie der Schlange eine Chance geben. Lassen Sie sich nicht von einem sicherlich etwas holprigen Start abschrecken. Sie werden es nicht bereuen, sich mit einer Sprache beschäftigt zu haben, die viel mehr kann als nur „Skript“ und die sich, gerade für den Anfang, erst einmal relativ einfach als zusätzliches Hilfsmittel in Ihre vorhandene Arbeitslandschaft mit aufnehmen lässt. Dabei kann Ihnen „`subprocess.run()`“ helfen – wenn Sie es dann doch genauer wissen wollen, wagen Sie den nächsten Schritt und probieren Sie `cx_Oracle` aus. Je mehr Leute dieses Modul verwenden, desto besser wird es gepflegt und desto einfacher wird die

Schlangendressur auch für (alle anderen) Oracle-Datenbankadministratoren.

Quellen

- [1] [https://de.wikipedia.org/wiki/Python_\(Programmiersprache\)](https://de.wikipedia.org/wiki/Python_(Programmiersprache))
- [2] https://www.youtube.com/channel/UC-CezlgC97PvUuR4_gbFUs5g
- [3] <https://realpython.com/intro-to-pyenv/>
- [4] https://github.com/oracle/python-cx_Oracle
- [5] https://cx-oracle.readthedocs.io/en/latest/user_guide/introduction.html

Über die Autorin

Katharinas Freude an Datenbanken wurde vor 15 Jahren während ihres Mathematikstudiums entfacht und ließ sie seitdem nicht mehr los. Bis ihr Weg allerdings der Oracle-Spur folgte, sollten noch weitere 10 Jahre vergehen. Nun arbeitet sie als Oracle-Datenbankadministratorin bei der Equicon Software GmbH Jena und beschäftigt sich, neben der Automatisierung als solcher, auch mit Themen wie Hochverfügbarkeit mit SE2, Oracle-XE-Datenbanken auf Docker und Multitenant.



Katharina Büchse
buechse@equicon.de



Data Exchange with PostgreSQL – Teil 1

Michael Kloker, Boehringer Ingelheim / IT RDM

Aufgrund der Lizenzpolitik von Oracle und des allgemeinen Kostendrucks in der IT überlegen momentan viele Firmen, ihre Anzahl an Oracle-Datenbanken deutlich zu reduzieren und auf alternative Datenbanken umzusteigen. Was sich zwischenzeitlich immer stärker als Oracle-Alternative etabliert hat, ist PostgreSQL. Hat man Applikationen im Einsatz, die auch PostgreSQL unterstützen, können diese migriert werden. Aber ist die Datenbankmigration das Einzige, was berücksichtigt werden muss? Die wenigsten Datenbanken sind Stand-alone-Datenbanken, die allermeisten haben Schnittstellen zu anderen Datenbanken. Wie geht man mit diesen Schnittstellen um? In einer homogenen Oracle-Umgebung lässt sich der Datenaustausch einfach mit DB Links realisieren, aber wie kann er über Technologiegrenzen hinweg aussehen (*siehe Abbildung 1*)? Im Zuge dieser Entwicklung habe ich mich mit der Schnittstellenthematik beschäftigt und untersucht, wie PostgreSQL in eine bereits bestehende Datenbankumgebung aus Oracle-Datenbanken und MSSQL-Server passt und welche Möglichkeiten des Datenaustausches es zwischen den verschiedenen Datenbanktechnologien gibt (*siehe Abbildung 2*).

Im diesem ersten Teil des Artikels „Data Exchange with PostgreSQL“ gehe ich etwas allgemeiner auf die PostgreSQL Foreign Data Wrapper ein, die als Basis für den Datenaustausch in PostgreSQL dienen. Wir schauen uns den Postgres Foreign Data Wrapper an, mit dem der Datenaustausch zwischen verschiedenen PostgreSQL-Datenbanken möglich ist, und den Dblink Foreign Data Wrapper, um Funktionen und Prozeduren auf einer Remote-PostgreSQL-Datenbank auszuführen.

Im zweiten Teil werde ich auf den Oracle Foreign Data Wrapper eingehen, mit dem der Zugriff von PostgreSQL auf Oracle-Datenbanken möglich ist, sowie auf das Oracle Gateway for ODBC für den Zugriff von Oracle auf PostgreSQL.

Im dritten und letzten Teil werden wir den Table Data Stream Foreign Data Wrapper kennen lernen, mit dem man von PostgreSQL auf MSSQL-Server zugreifen kann, und Linked Server mit psqLODBC für den Zugriff von MSSQL-Server auf PostgreSQL. Außerdem werden wir im dritten Teil PostgREST anschauen, eine Möglichkeit, schnell und einfach REST APIs für PostgreSQL bereitzustellen.

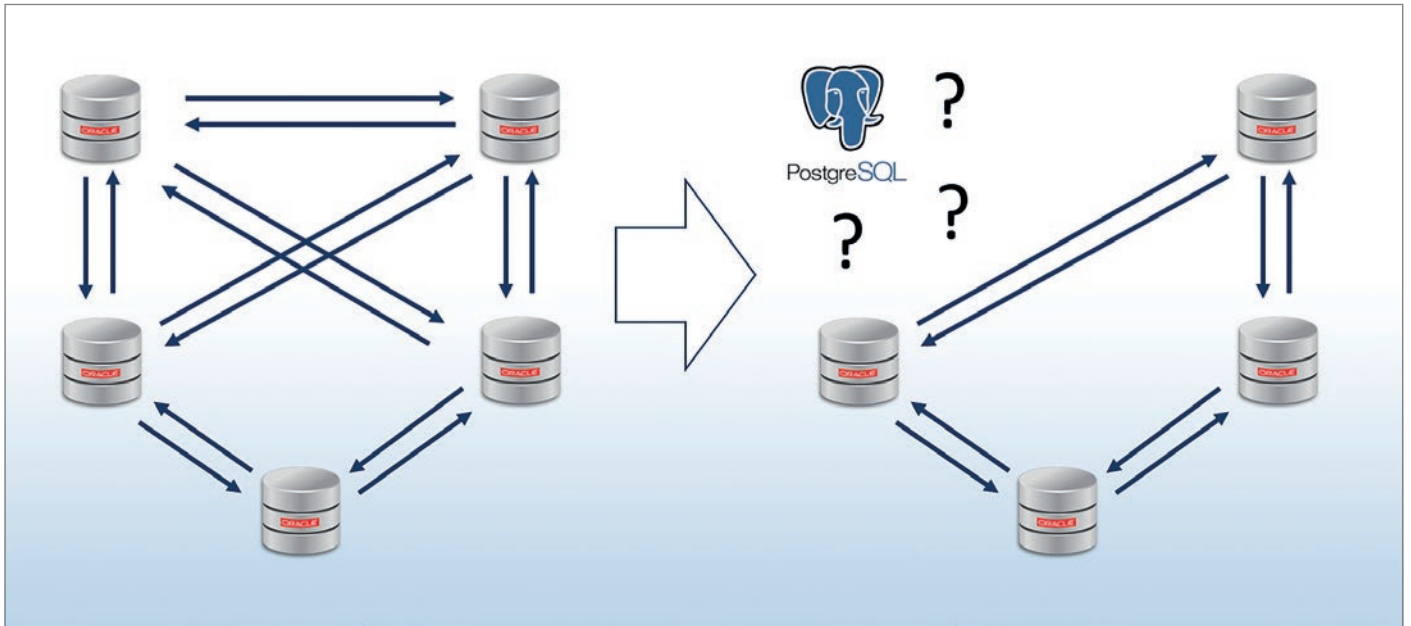


Abbildung 1: Datenaustausch zwischen Oracle-Umgebung und PostgreSQL? (Quelle: © Oracle und PostgreSQL)

Allgemeines zu Postgres

PostgreSQL, im Folgenden nur noch als Postgres bezeichnet, ist, nachdem es neu installiert wurde, vom Funktionsumfang her noch eingeschränkt. Einen Datenbanklink zu einer anderen Postgres-DB zu erstellen, wie es von Oracle her bekannt ist, geht zum Beispiel nicht. Allerdings bietet Postgres die Möglichkeit, sogenannte Extensions zu installieren und damit den Funktionsumfang zu erweitern. Um auf weitere Datenquellen zuzugreifen, gibt es

die Foreign Data Wrapper (FDW) Extensions. Diese ermöglichen Postgres den Zugriff auf eine Vielzahl verschiedener Datenquellen, so auch den Zugriff auf andere Postgres-DBs, auf Files, Oracle-DBs, NoSQL DBs oder Hadoop, um nur einige zu nennen. Eine Liste der verfügbaren FDW findet man im Postgres-Wiki (https://wiki.postgresql.org/wiki/Foreign_data_wrappers).

Die Installation einer FDW Extension erfolgt immer in zwei Teilen. Im ersten Teil erfolgt die Installation auf Betriebssystemebene, dabei werden Files im Postgres-

Extension-Verzeichnis erstellt. Im zweiten Teil erfolgt die Konfiguration innerhalb der Postgres-Datenbank. Dabei wird zuerst die Extension erstellt, danach ein Foreign-Server-Objekt mit den Verbindungsinformationen und ein User-Mapping mit den Authentifizierungsinformationen für die Zieldatenquelle.

Im Vergleich zu Oracle werden bei Postgres zwei Objekte benötigt, um einen „Datenbank-Link“ zu erstellen, der Foreign Server und das User-Mapping. Bei Oracle werden alle benötigten Informati-

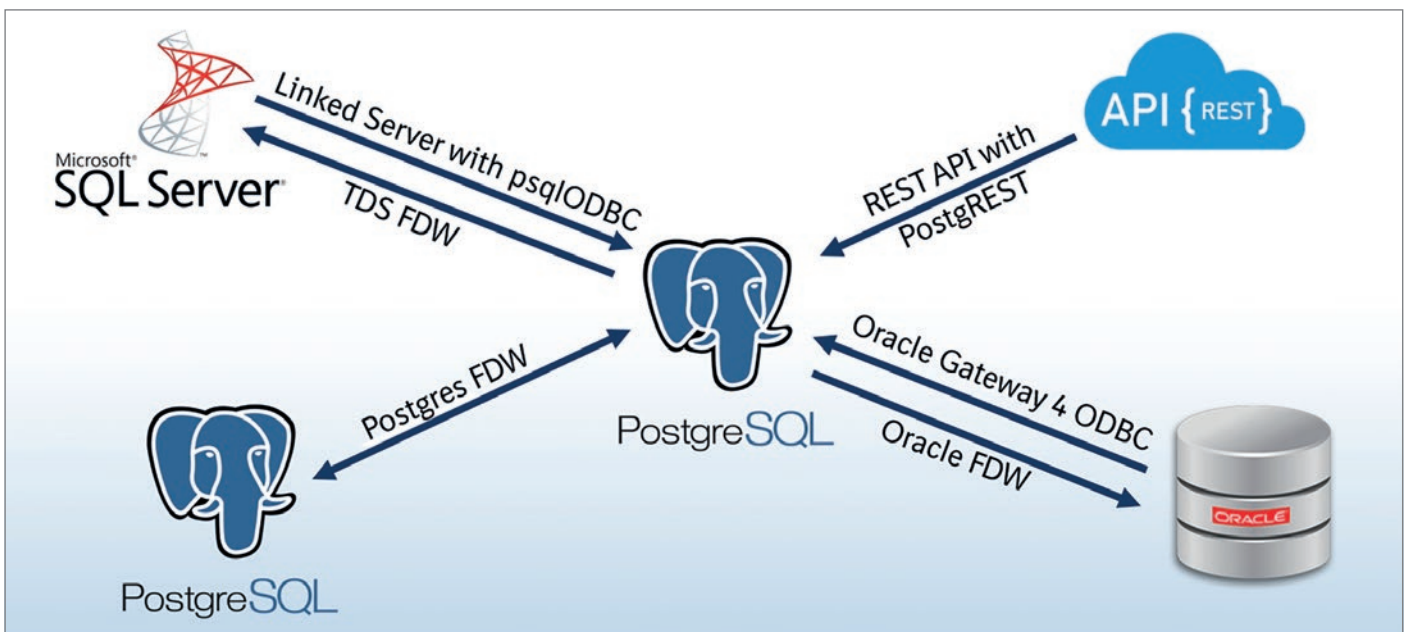


Abbildung 2: Einbindung von PostgreSQL in eine Datenbankumgebung aus Oracle-Datenbanken und MSSQL-Server ? (Quelle: © Oracle, Microsoft und PostgreSQL)

onen (Zieldatenbank, User, Passwort) im CREATE DATABASE LINK Statement mitgegeben. Wieso trennt Postgres Verbindungs- und Authentifizierungsinformationen? Greift man auf Datenquellen zu, die keine Authentifizierung benötigen, entfällt das User-Mapping, wie etwa beim File FDW, mit dem auf Dateien zugegriffen wird, vergleichbar mit external Tables unter Oracle. Greift man aber auf Datenquellen zu, die eine Authentifizierung erfordern, wird das User-Mapping benötigt.

Zum Schluss werden Foreign Tables erstellt. Beim Erstellen der Foreign Tables werden die Spalten der Remote-Tabellen auf Spalten einer Postgres Foreign Table gemappt, soll heißen, die Foreign Table ist eine Art View, mit der Tabellendefinition der Remote-Tabelle, die aber keine Daten beinhaltet. Mit diesen Informationen kann der Postgres Optimizer Statements auf die Remote-Tabelle entsprechend optimieren. Auf die Foreign Tables werden nun, wie gewohnt, DML Statements ausgeführt.

Dieses Vorgehen aus CREATE EXTENSION, CREATE FOREIGN SERVER, CREATE USER MAPPING und CREATE FOREIGN TABLE ist für alle FDW, den Postgres FDW, den Oracle FDW und den Table Datastream (TDS) FDW, die wir in dieser Artikelserie besprechen, gleich.

Wenn man aus der Oracle- oder MS-SQL-Server-Welt kommt, hört sich dies zunächst kompliziert an, ist es aber nicht. Wir werden uns im Anschluss die FDW genauer anschauen und sie Schritt für Schritt installieren und konfigurieren.

Der Postgres FDW: Zugriff auf Remote-Postgres-Tabellen

Möchte man von einer Postgres-DB auf eine andere Postgres-DB zugreifen, egal ob die Remote-Postgres-DB im gleichen Postgres Cluster läuft oder zu einem anderen gehört, muss der Postgres FDW installiert werden. Mit diesem können dann SELECT-, INSERT-, UPDATE- und DELETE-Statements auf Remote-Tabellen ausgeführt werden.

Wie oben bereits erwähnt, muss der Postgres FDW zuerst auf dem Betriebssystem installiert werden. Dazu installieren wir das Postgres Contrib Package (siehe Listing 1). Dieses Paket beinhaltet eine Reihe von häufig verwendeten Post-

```
# yum install rh-postgresql12-postgresql-contrib.x86_64
```

Listing 1: Installation des Postgres Contrib Package

```
pgdb1=# create extension postgres_fdw schema public;
pgdb1=# \dx
```

List of installed extensions			
Name	Version	Schema	Description
postgres_fdw	1.0	public	foreign-data wrapper for remote PostgreSQL servers

Listing 2: Erstellen der Postgres FDW Extension

```
pgdb1=# CREATE SERVER pgdb2
        FOREIGN DATA WRAPPER postgres_fdw
        OPTIONS (host 'pghost', port '5432', dbname 'pgdb2');
pgdb1=# \des
```

List of foreign servers		
Name	Owner	Foreign-data wrapper
pgdb2	admin1	postgres_fdw

Listing 3: Erstellen des Foreign-Server-Objekts für den Postgres FDW

```
pgdb1=# CREATE USER MAPPING for admin1
        SERVER pgdb2
        OPTIONS (user 'remoteuser', password '*****');
pgdb1=# \deu
```

List of user mappings	
Server	User name
pgdb2	admin1

Listing 4: Erstellen des User-Mappings für den Postgres FDW

gres Extensions, darunter auch den Postgres FDW.

Anschließend wird in der Datenbank mit CREATE EXTENSION die Postgres FDW Extension erstellt (siehe Listing 2). Die Extension wird im Beispiel im Schema public erstellt, damit können auch andere User in der DB die Extension nutzen. Mit dem psql Shortcut \dx können die installierten Extensions abgefragt werden (siehe Listing 2).

Ist die Extension installiert, wird im nächsten Schritt das Foreign-Server-Objekt mit den Verbindungsinformationen für die Remote-DB (Postgres Host, Port und DB Name) erstellt (siehe Listing 3). Beim Erstellen des Foreign Server wird die oben erstellte postgres_fdw Extension referenziert. Mit dem Shortcut \des werden Foreign-Server-Objekte angezeigt, darun-

ter auch der gerade erstellte Foreign Server pgdb2 (siehe Listing 3).

Als Nächstes wird das User-Mapping mit Username und Passwort erstellt, mit dem die Authentifizierung gegen die Remote-Postgres-DB durchgeführt wird. Das User-Mapping referenziert das oben erstellte Foreign-Server-Objekt pgdb2. Das User-Mapping wird hier im Beispiel im Schema admin1 erstellt. Dadurch kann nur der User admin1 das User-Mapping nutzen und auf die Remote-Tabelle zugreifen. Würde man es im public-Schema erstellen, könnten andere DB User das User-Mapping auch nutzen und auf die Foreign Table zugreifen, was eventuell nicht erwünscht ist. Mit dem psql Shortcut \deu lassen sich die User-Mappings anzeigen (siehe Listing 4).

Nun wird die Foreign Table erstellt (siehe Listing 5). Dabei werden die Spalten mit

den jeweiligen Datentypen der Remote-Tabelle auf Spalten mit den gleichen Datentypen der Foreign Table gemappt. Im unteren Beispiel hat die Remote-Tabelle 2 Spalten, Spalte 1 vom Datentyp INTEGER und Spalte 2 vom Datentyp VARCHAR. Die Spaltennamen der Foreign Table können von den Spaltennamen der Remote-Tabelle abweichen. Unter Options wird das Schema angegeben, in dem sich die Remote-Tabelle befindet, sowie der originale Tabellenname in der Remote-DB. Mit dem psql Shortcut \det werden die Foreign Tables angezeigt.

Nun können beliebige SELECT-, INSERT-, UPDATE- und DELETE-Statements auf der Foreign Table ft_pgdb2_pt1 ausgeführt werden (siehe Listing 6).

Möchte man auf alle oder mehrere Tabellen einer Remote-DB zugreifen, muss man nicht jede Foreign Table einzeln erstellen. Eine größere Anzahl von Foreign Tables lässt sich mit dem Befehl IMPORT FOREIGN SCHEMA ... erstellen. Dabei kann man alle Tabellen oder mit Filtern nur bestimmte Tabellen eines Remote-Schemas importieren oder auch bestimmte Tabellen ausschließen.

Der Dblink FDW: Ausführen von Remote-Funktionen/-Prozeduren in Postgres

Um Funktionen und Prozeduren in einer Remote-Postgres-DB auszuführen, wird der Dblink FDW benötigt. Der Dblink FDW ist, wie der Postgres FDW, im Contrib Package enthalten, das wir bereits kennengelernt und installiert haben (siehe Listing 1). Daher können wir direkt mit der Konfiguration in der Datenbank beginnen. Das Vorgehen ist gleich wie beim Postgres FDW. Zuerst wird die Dblink FDW Extension, dann das Foreign-Server-Objekt und anschließend das User-Mapping in der DB erstellt. Da wir mit dem Dblink FDW auf keine Tabellen zugreifen, entfällt das Erstellen von Foreign Tables.

Listing 7 zeigt, wie die Dblink Extension in der Datenbank erstellt wird. Mit dem psql Shortcut \dx werden die installierten Extensions angezeigt und man sieht, ob die Installation erfolgreich war. Beim Dblink FDW ist zu beachten, dass die Extension nur dblink heißt, bei der Erstellung des Foreign Server aber dblink_fdw angegeben werden muss. Mit dem Short-

```
pgdb1=# CREATE FOREIGN TABLE
        ft_pgdb2_pt1 (id int, text varchar)
        SERVER pgdb2
        OPTIONS (schema_name 'admin2', table_name 'pt1');
pgdb1=# \det
        List of foreign tables
 Schema |      Table      | Server
-----+-----+-----
 admin1 | ft_pgdb2_pt1   | pgdb2
```

Listing 5: Erstellen der Foreign Table

```
pgdb1=# select * from ft_pgdb2_pt1;
 id |      text
----+-----
  1 | Host pghost
  2 | Postgres DB pgdb2
  3 | Schema admin2
  4 | Table pt1

pgdb1=# insert into ft_pgdb2_pt1
        values ( 10, 'Greetings from pgdb1' );
pgdb1=# update ft_pgdb2_pt1
        set text='Update from pgdb1' where id=10;
pgdb1=# delete from ft_pgdb2_pt1 where id=10;
```

Listing 6: SELECT / INSERT / UPDATE / DELETE Statements auf eine Foreign Table

```
pgdb1=# create extension dblink schema public;
pgdb1=# \dx
        List of installed extensions
 Name | Version | Schema |      Description
-----+-----+-----+-----
 dblink | 1.2 | public | connect to other PostgreSQL
        | | | databases from within a
        | | | database
pgdb1=# \dew
        List of foreign-data wrappers
 Name | Owner | Handler | Validator
-----+-----+-----+-----
 dblink_fdw | admin1 | - | dblink_fdw_validator
```

Listing 7: Create dblink_fdw Extension

```
pgdb1=# CREATE SERVER pgdb2_dblink
        FOREIGN DATA WRAPPER dblink_fdw
        OPTIONS (host 'pghost', port '5432', dbname 'pgdb2');
pgdb1=# \des
        List of foreign servers
 Name | Owner | Foreign-data wrapper
-----+-----+-----
 pgdb2_dblink | admin1 | dblink_fdw
```

Listing 8: Create Foreign Server für die Dblink Extension

cut \dew können die installierten FDWs mit ihren Namen angezeigt werden (siehe Listing 7).

Im Listing 8 wird das Foreign-Server-Objekt mit der Referenz auf den Dblink

FDW und den Verbindungsinformationen (Postgres Host, Port und DB Name) für die Remote-Postgres-Datenbank erstellt.

In Listing 9 wird das User-Mapping mit den Authentifizierungsinformationen im

```
pgdb1=# CREATE USER MAPPING for admin1
        SERVER pgdb2_dblink
        OPTIONS (user 'remoteuser', password '*****');
```

Listing 9: Create User Mapping für die Dblink Extension

```
pgdb1=# SELECT dblink_connect('pgdb2con', 'pgdb2_dblink');
```

Listing 10: Aufbauen einer Session zur Remote-DB

```
pgdb1=# SELECT * FROM dblink
        ('pgdb2con','select f_sum(8, 7)')
        as t(sum int);

 sum
-----
  15
```

Listing 11: Aufruf einer Remote-Postgres-Funktion, Beispiel 1

```
pgdb1=# SELECT * FROM dblink(
        'pgdb2con',
        'select f_write_pt1(10, '#execute remote function')')
        as t(result boolean);

 result
-----
      t
```

Listing 12: Aufruf einer Remote-Postgres-Funktion, Beispiel 2

```
pgdb1=# SELECT * FROM dblink_exec(
        'pgdb2con',
        'call p_write_pt1(11, 'execute remote procedure')'
        );
 dblink_exec
-----
      CALL
```

Listing 13: Aufruf einer Remote-Postgres-Prozedur

```
pgdb1=# SELECT dblink_disconnect('pgdb2con');
```

Listing 14: Beenden einer dblink-Session

admin1-Schema erstellt, der oben erstellte Foreign Server wird dabei referenziert.

Damit sind die Vorbereitungen abgeschlossen und man kann nun Funktionen und Prozeduren auf der Remote-DB ausführen. Mit der Funktion `dblink_connect` (siehe Listing 10) wird eine Session zur Remote-DB aufgebaut. Der erste Parameter ist der Verbindungsname, der zweite Pa-

rameter ist der Name des Foreign Server, der oben erstellt wurde.

Nun wird mit der `DB_LINK`-Funktion eine Funktion in der Remote-DB aufgerufen. Parameter 1 ist der Name der Datenbankverbindung, die oben erstellt wurde, Parameter 2 ist der Aufruf der Remote-Funktion mit den entsprechenden Parametern. Als Ergebnis wird eine Tabelle mit dem Rückgabewert ausgegeben.

Listing 11 zeigt den Aufruf einer Funktion `F_SUM`, der 2 Zahlen übergeben werden, die auf der Remote-DB addiert werden. Das Ergebnis wird als Tabelle `t` mit der Spalte `sum` ausgegeben.

Listing 12 zeigt den Aufruf einer Funktion `F_WRITE_PT1`, der eine Zahl und ein String übergeben werden, die auf der Remote-DB in die Tabelle `pt1` geschrieben werden. Als Ergebnis wird zurückgegeben, ob der Insert erfolgreich war oder nicht.

Mit Postgres 11 wurden Prozeduren eingeführt. Listing 13 zeigt, wie mit der Funktion `DBLINK_EXEC` eine Prozedur in einer Remote-Postgres-DB aufgerufen wird. Die Prozedur `P_WRITE_PT1` schreibt die beiden ihr übergebenen Parameter in die Tabelle `pt1` in der Remote-DB. Einen Rückgabeparameter gibt es hier nicht.

Mit `dblink_disconnect` (siehe Listing 14) wird die Session zur Remote-DB beendet.

Zusammenfassung

In diesem ersten Teil von „Data Exchange with Postgres“ sind wir allgemein auf die Foreign Data Wrapper eingegangen, die in Postgres als Basis für den Datenaustausch mit anderen Datenquellen dienen. Wir haben gesehen, wie mit dem Postgres FDW Daten zwischen verschiedenen Postgres-Datenbanken ausgetauscht und wie mit dem Dblink Foreign Data Wrapper Remote-Funktionen und -Prozeduren aufgerufen werden. Damit unterstützt Postgres die Funktionalität, die wir auch von Oracle her kennen und erwarten.

Ausblick

Im zweiten Teil von „Data Exchange with PostgreSQL“ betrachten wir den Oracle FDW, für den Zugriff von Postgres auf Oracle, sowie das Oracle Gateway for ODBC, mit dem der Zugriff von Oracle auf Postgres möglich ist. Im dritten Teil werden wir den Table Data Stream FDW, für den Zugriff von Postgres auf MSSQL-Server, sowie Linked Server mit `psqlODBC` für den Zugriff von MSSQL-Server auf Postgres kennen lernen. Zusätzlich werfen wir noch einen Blick auf PostgREST, eine Möglichkeit, schnell und einfach REST-Schnittstellen für den Zugriff auf Postgres bereitzustellen.

Quellen

- https://wiki.postgresql.org/wiki/Foreign_data_wrappers#Big_Data_Wrappers
- <https://www.postgresql.org/docs/9.5/postgres-fdw.html>
- <https://www.postgresql.org/docs/9.1/sql-createextension.html>
- <https://www.postgresql.org/docs/9.1/sql-createserver.html>
- <https://www.postgresql.org/docs/10/sql-createusermapping.html>
- <https://www.postgresql.org/docs/9.4/sql-createserver.html>
- <https://www.postgresql.org/docs/9.5/sql-createforeigntable.html>
- <https://www.postgresql.org/docs/9.5/sql-importforeignschema.html>
- https://www.percona.com/blog/2018/08/21/foreign-data-wrappers-postgresql-postgres_fdw/
- <https://help.aiven.io/en/articles/977358-postgresql-dblink-extension-use-example>

Über den Autor

Michael Kloker arbeitet seit 20 Jahren bei Boehringer Ingelheim in der IT. In dieser Zeit hat er an einer Vielzahl unterschiedlicher Themen gearbeitet und verfügt über ein sehr breites und tiefes Wissen zu Betriebssystemen, Datenbanktechnologien, Virtualisierungslösungen, Applikationsservern, Verzeichnisdiensten, Authentifizierungsmethoden, Applikations- und Prozesswissen. Heute kümmert sich Michael Kloker um die Implementierung, den störungsfreien Betrieb, Systemupdates und die Weiterentwicklung bestehender Computersysteme für die pharmazeutische Entwicklung sowie

den Datenaustausch und die Schnittstellenthematik unterschiedlicher Computer- und Datenbanksysteme.



Michael Kloker

michael.kloker@boehringer-ingelheim.com

DOAG Datenbank Kolumne: Wie gebe ich Feedback und warum?

Andreas Buckenhofer, Mitglied der Delegiertenversammlung Data Analytics,
Stellv. Themenverantwortung Soft Skill

Für den DBA sind nicht nur technische Themen von Belang. Soziale Kompetenzen werden gerne vernachlässigt.

Wie lange liegt die letzte Schulung zu Soft Skills zurück? Kommunikation findet immer statt. Egal ob am Arbeitsplatz zu Hause oder im Büro.

Viele Unternehmen haben in den letzten Jahren eine Feedback-Kultur etabliert. Im Homeoffice sind schnell Ausreden gefunden, wie etwa die Aussage "Virtuell geht das nicht". Rückmeldung zu geben, bleibt nach wie vor wichtig und funktioniert auch virtuell mit eingeschalteter Kamera.

Bei Feedback denkt man zuerst an Negatives ("konstruktives Feedback"). Feedback darf und soll auch positiv sein. Welches Lob oder welche Anerkennung wollten Sie schon immer mal loswerden? Sprechen Sie einen Kollegen oder eine Kollegin an. Feedback ist nicht nur für Führungskräfte relevant: Rückmeldungen werden sowohl Hierarchie-übergreifend gegeben (z.B. Vorgesetzter an Mitarbeiter oder umgekehrt) als auch auf der gleichen Hierarchieebene.

Zur Vorbereitung eines Feedbackgesprächs eignet sich die nachfolgende WWW-Formel. Diese Struktur gilt ebenfalls für positives Feedback. Wichtig ist, dass Sie aus der Ich-Perspektive wiedergeben und keine persönlichen Wertungen, Anschuldigungen oder gar Vermutungen äußern. Suchen Sie zeitnah das Gespräch und schieben Sie es nicht auf.

- **Wahrnehmung**
Beschreiben Sie konkret, was Sie erlebt haben. Geben Sie Ihre Wahrnehmung so neutral wie möglich und ohne Bezug auf vermutete Charaktereigenschaften wieder ("Video-Kamera").
Beispiel: Mir ist aufgefallen, dass Du Dich in den vergangenen Projektmeetings nicht äusserst.
- **Wirkung**
Beschreiben Sie aus Ihrer subjektiven Perspektive, was die Wahrnehmung in Ihnen ausgelöst hat.
Beispiel: Ich fühle mich mit den Themen und Herausforderungen im Meeting alleingelassen.
- **Wunsch / Erwartung (optional)**
Beschreiben Sie klar und deutlich, was Sie erreichen möchten.
Beispiel: Ich wünsche mir, dass Du Dich im Meeting einbringst. Du hast viele Ideen und ein breites technisches Wissen, von dem alle profitieren.

Wer sich weiterentwickeln bzw. seine eigene Wirkung auf andere erfahren möchte, fordert aktiv Rückmeldungen ein: Warum nicht einmal auf einen Kollegen/eine Kollegin zugehen und Feedback einfordern?



Hochverfügbarkeit für die Oracle Datenbank 19c SE2 mit SEHA

Markus Flechtner, Trivadis Germany

Hochverfügbarkeit für Oracle-Datenbanken verbindet man meist mit den Begriffen Enterprise Edition, Real Application Cluster und Data Guard – und mit hohen Preisen. Aber auch für die Standard Edition 2 der Oracle-Datenbank kann man ohne Zusatzprodukte die Verfügbarkeit der Datenbank deutlich erhöhen.

Die „Maximum Availability Architecture“ (MAA) ist sicher der Gold-Standard für Hochverfügbarkeit mit Oracle-Datenbanken. In den MAA-Varianten fährt Oracle alles auf, was Rang und Namen – und einen hohen Preis – hat. Es gibt jedoch auch für die Standard Edition 2 (SE2) der Datenbank Möglichkeiten, die Datenbank hochverfügbar zu machen: Dazu zählen Programme von Drittherstellern, wie zum Beispiel DBVisit Standby oder TVD-Standby, mit denen Data-Guard-ähnliche Funktionalität bereitgestellt wird, und RAC, das bis zur Version 18c der Oracle-Datenbank kostenfrei bei der SE2 dabei war. Zwar war dieses RAC etwas

eingeschränkt (max. zwei Sockets im Cluster, max. acht Threads pro Instanz), aber für viele Kunden reichte das aus.

Etwas überraschend hat dann Oracle im April 2019 RAC aus der Feature-Liste der Datenbank-Version 19c gestrichen. Der Unmut bei den SE2-Nutzern war groß, denn üblicherweise geschehen solche Änderungen bei Oracle etwas langfristiger. Zudem haben die von Oracle angebotenen Alternativen, wie der Wechsel in die Oracle Cloud oder der Wechsel zur Enterprise Edition, nicht wirklich Zuspruch gefunden. Im März 2020 kam dann jedoch die Nachricht aus Redwood Shores „Standard Edition 2

– We heard you“ und Oracle gab bekannt, dass es bald ein Feature namens „Standard Edition 2 High Availability“ (SEHA) geben würde [1].

Was ist „SEHA“?

SEHA ist eine Failover-Datenbank-Lösung, die auf Oracle Grid Infrastructure beruht. Die Datenbank läuft auf einem Knoten im Cluster und wenn dieser Knoten ausfällt, dann wird die Datenbank automatisch auf einem anderen Knoten im Cluster gestartet. Dabei schreibt Oracle vor,

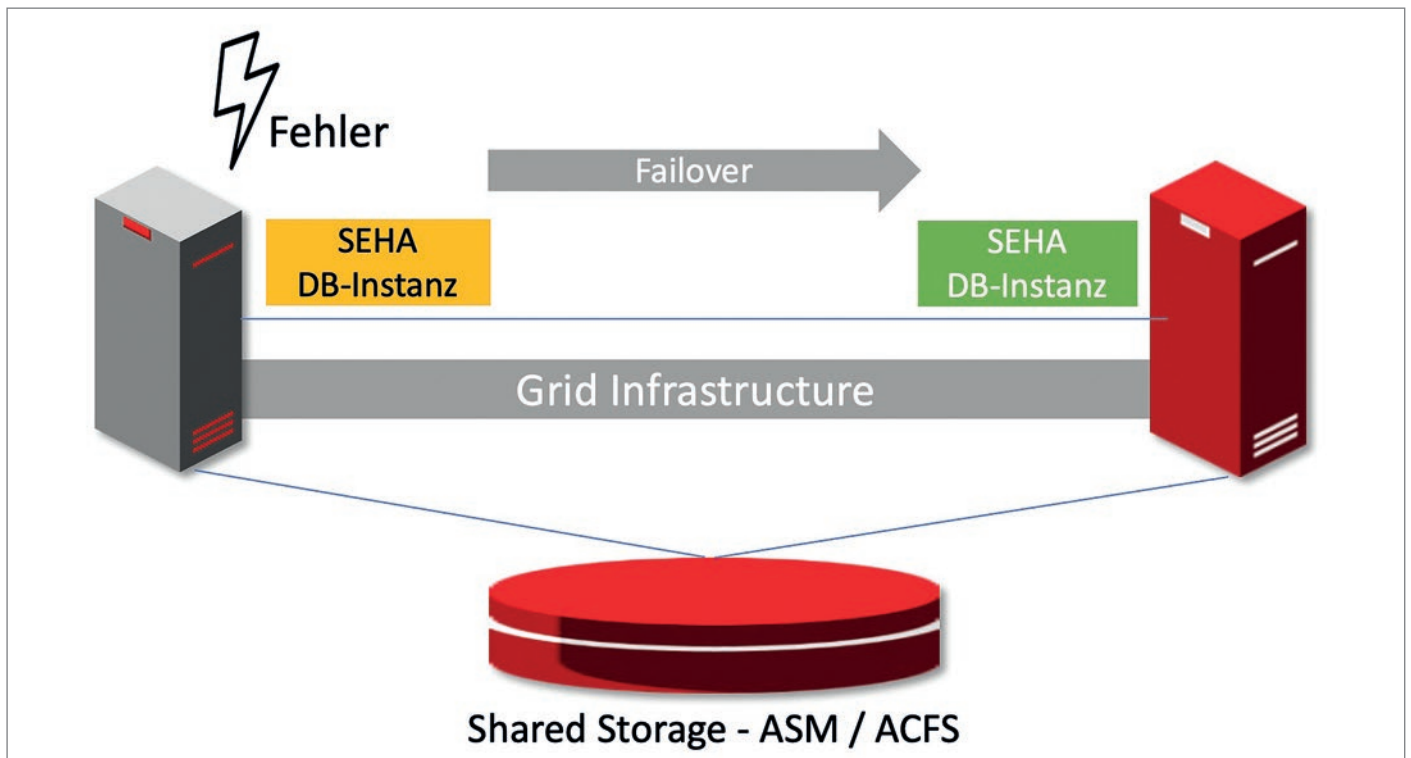


Abbildung 1: Überblick über Standard Edition High Availability (SEHA) (Quelle: Markus Flechtner)

dass ASM oder ACFS als Storage-Lösung für die Datenbank genutzt werden muss (siehe Abbildung 1). SEHA wurde mit dem Release-Update (RU) 19.7 (April 2019) veröffentlicht. Da es in diesem RU aber noch einige „Issues“ mit SEHA gibt, empfiehlt es sich, ein höheres Release-Update zu nutzen und es für die Grid-Infrastructure-Software und die RDBMS-Software zu installieren.

SEHA ist verfügbar für Linux x86-64bit, Solaris Sparc und Microsoft Windows. Für Windows-Nutzer kann SEHA ein Ersatz für das abgekündigte Oracle Fail Safe sein. Eine separate Lizenz ist für SEHA nicht erforderlich, das heißt, es reicht die SE2-Lizenz und es gilt auch die Einschränkung der SE2, dass ein Server im Cluster maximal zwei Sockets haben darf. SEHA fällt unter die „10-Tages-Regel“, das bedeutet grob gesagt, wenn der zweite Knoten im Cluster nur maximal zehn angebrochene Tage im Jahr für eine aktive Datenbank genutzt wird, dann muss der Knoten nicht lizenziert werden. Die Einzelheiten dazu sind im Oracle-Dokument „Licensing Data Recovery Environments“ [2] nachzulesen. Nutzt man den Cluster jedoch so, dass auf beiden Knoten Datenbanken laufen, die von dem jeweils anderen Knoten abgesichert werden, dann sind natürlich beide Knoten zu lizenzieren.

Wie richtet man SEHA ein?

Die Datenbank wird wie eine normale SE2-Datenbank angelegt. Danach muss man dafür sorgen, dass Password-File und spfile von allen Cluster-Knoten aus erreichbar sind, und sie zum Beispiel im ASM ablegen, sofern das nicht bereits beim Anlegen der Datenbank geschehen ist. Nutzt man den dbca, dann wird das spfile beim Anlegen der Datenbank im ASM abgelegt und man muss nur noch das Password-File ins ASM schieben (siehe Listing 1).

Wichtig ist auch, dass der Parameter LOCAL_LISTENER nicht gesetzt ist, damit sich die Datenbank nicht beim falschen Listener registriert.

Sobald dies alles eingerichtet ist, muss man der Datenbank die gewünschten Clus-

ter-Knoten als mögliche Server angeben (siehe Listing 2).

„Switchover“ und Failover mit SEHA

Ein „Switchover“ bei SEHA ist ein sogenannter Relocate, bei dem die Datenbank-Instanz auf dem Ursprungsknoten gestoppt (SHUTDOWN IMMEDIATE) und dann auf einem anderen Knoten neu gestartet wird. Im Gegensatz zum „RAC-One-Node“ ist also immer nur eine Instanz aktiv und die Anwendungen müssen sich nach einem derartigen Relocate neu mit der Datenbank verbinden. Der Befehl für den Relocate ist in Listing 3 dargestellt.

```
$ asmcmd pwcopu /u00/app/oracle/dbs/orapwSEHADB \
+DATA/SEHADB/orapwSEHADB
$ srvctl modify database -db SEHADB -pwfile \ +DATA/SEHADB/orapwSEHADB
```

Listing 1: Beispiel: Ablegen des Password-Files im ASM

```
$ srvctl modify database -db SEHADB -node node1,node2
```

Listing 2: Einrichten der SEHA-Datenbank

```
srvctl relocate database -db <DBNAME> -node <ZIELKNOTEN>
```

Listing 3: *srvctl relocate*

```
SEHA_APP.REDSTACK.DOAG =  
(DESCRIPTION =  
  (ADDRESS= (PROTOCOL=TCP) (HOST=sehacluster) (PORT=1521))  
  (CONNECT_DATA =  
    (SERVER = DEDICATED)  
    (SERVICE_NAME = SEHA_APP_SVC.redstack.doag)  
  )  
)
```

Listing 4: *Einfacher Connect-String für eine SEHA-Datenbank*

Am besten testet man eine SEHA-Konfiguration, indem man einmal auf den zweiten Server wechselt und dann wieder auf den Ausgangsserver zurückgeht.

Ein Failover mit SEHA ist einfach. Wenn ein Cluster-Knoten ausfällt, dann erkennt Grid Infrastructure dies und startet die Instanz automatisch auf dem anderen Knoten. Wenn nur die Datenbankinstanz ausgefallen ist, dann wird zuerst probiert, die Instanz lokal neu zu starten, bevor auf den anderen Knoten gewechselt wird. In jedem Fall müssen sich die Anwendungen aber neu anmelden.

TNS-Connect-String für SEHA

Der Connect-String für eine SEHA-Datenbank unterscheidet sich nicht vom Connect-String einer RAC-Datenbank. Als Host ist der SCAN-Name des Clusters einzutragen und als Service der jeweilige Applikationsservice (siehe Listing 4).

Wie in [5] gezeigt wird, funktionieren auch Fast-Application-Notification (FAN) und Fast-Connection-Failover (FCF) mit SEHA. Hingegen ist Application Continuity (AC) nicht für SEHA verfügbar, denn AC erfordert die Enterprise Edition und eine zusätzliche Lizenz für Active DataGuard oder RAC.

Wer andersherum hofft, mit SEHA seinen mit eigenen Skripten realisierten Failover-Cluster für eine Enterprise-Edition-Datenbank ersetzen zu können, wird leider enttäuscht: „PRCD-2088: failed to configure the single instance database EEHACDB with multiple nodes because it is not a Standard Edition database“ kommt in diesem Fall als Fehlermeldung.

Es ist zu wünschen, dass Oracle die mit SEHA eingebaute einfache Failover-Fähigkeit zukünftig auch für die Enterprise Edition ermöglicht.

Fazit

Mit der Standard Edition High Availability kann man ohne Zusatzlizenzen oder -produkte die Verfügbarkeit von Standard-Edition-2-Datenbanken mit einfachen Mitteln und ohne zusätzliche Kosten deutlich erhöhen. Man erreicht damit zwar nicht die Verfügbarkeit der Enterprise Edition, für viele Anwendungen reichen diese Möglichkeiten jedoch aus.

Quellen

- [1] „Standard Edition 2 – We Heard You! Announcing: Standard Edition High Availability“ - <https://blogs.oracle.com/maa/standard-edition-2-announcing-standard-edition-high-availability>
- [2] Licensing Data Recovery Environments: <http://www.oracle.com/us/corporate/pricing/data-recovery-licensing-070587.pdf>
- [3] Vortrag Markus Michalewicz (Oracle) – „Standard Edition High Availability The Why, What & How“
- [4] Vortrag „Hochverfügbarkeit für die Oracle Datenbank SE2“ – DOAG-Konferenz 2020
- [5] Mathias Zarick: FAN events and FCF for SEHA – <https://mathiaszarick.wordpress.com/2021/02/28/fan-events-and-fcf-for-seha/>

Dokumentation

- Oracle Database 19c - Database Installation Guide - Chapter 10 „Installing Standard Edition High Availability“

Über den Autor

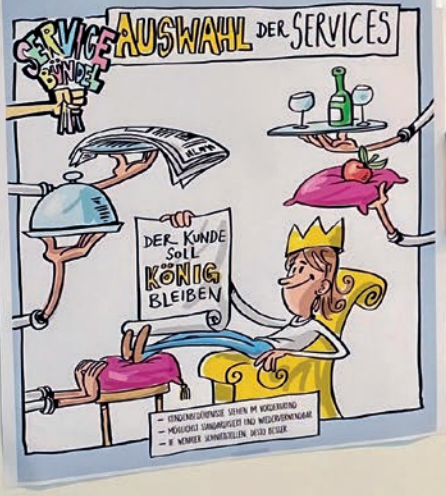
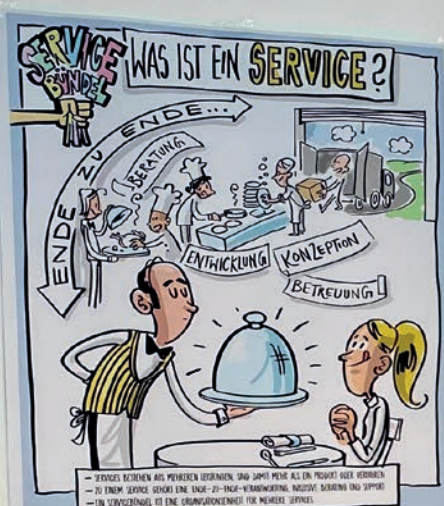
Markus Flechtner begann seine Tätigkeit im Oracle-Umfeld in den 1990ern als Software-Entwickler in den Bereichen PL/SQL, Oracle-Forms und Oracle-Reports. Später wechselte der Diplom-Mathematiker in den Oracle Support und spezialisierte sich auf die Tätigkeit als Datenbankadministrator. Seit 2008 ist er als Consultant bei Trivadis. Seine Schwerpunkte sind Oracle-Hochverfügbarkeit sowie Upgrade- beziehungsweise Migrationsprojekte.

Als Hauptreferent ist er bei Trivadis verantwortlich für die Kurse „Oracle Database New Features für DBAs“, „Oracle Multitenant“ und „PostgreSQL für Oracle-DBAs“, weiterhin ist er Referent für den Kurs „Oracle Real Application Clusters“.

Bei der Deutschen Oracle-Anwendergruppe (DOAG) ist er seit 2015 Delegierter der jährlichen Delegiertenversammlung und seit 2019 verantwortlich für den Themenbereich „Open-Source-Datenbanken“. Er ist Mitautor des Buches „Der Oracle DBA: Handbuch für die Administration der Oracle Database 12c“ und wurde 2020 zum Oracle ACE ernannt.



Markus Flechtner
Markus.flechtner@trivadis.com



Jan Mütter ist seit 2 Jahren einer von 2 Hauptverantwortlichen für die organisatorische Transformation von IT.NRW. Auf den Postern im Hintergrund sind die wesentlichen Designgrundsätze der neuen Organisation als Comics dargestellt. (Quelle: IT.NRW)



„Digitalisierung ist, wenn sich dann tatsächlich auch neue Geschäftsmodelle, neue Geschäftsprozesse und neue Verwaltungsprozesse daraus ergeben“

Martin Meyer, Redaktionsleiter des Red Stack Magazin, und Armin Wildenberg, DOAG Vorstand Data Analytics, unterhielten sich mit Dr. Jan Mütter, Geschäftsbereichsleitung „IT-Lösungen“ bei IT.NRW, über Digitalisierung und digitale Transformation.

Was ist Ihre berufliche Verbindung zum Thema Digitalisierung? Mit welchen Fragen und Aufgaben beschäftigen Sie sich?

Ich bin beim Landesbetrieb Information und Technik Nordrhein-Westfalen Leiter des Geschäftsbereichs IT-Lösungen, also letztendlich das, was man das Softwarehaus nennt. Der Landesbetrieb IT.NRW ist der IT-Provider für die Landesverwaltung und die Landesregierung in Nordrhein-Westfalen und damit der zentrale IT-Dienstleister – nicht der einzige, aber vielleicht der zentrale Dienstleister. In meinem Geschäftsbereich IT-Lösungen geht es im Moment ganz massiv um die Umsetzung des E-Government-Gesetzes des Landes Nordrhein-Westfalen, also die Digitalisierung der Verwaltungsarbeit in NRW, aber auch die Umsetzung des Online-Zugangsgesetzes. Dies ist ein Bundesgesetz, das regelt, dass bis Ende nächsten Jahres jeder Bürger und jedes Unternehmen alle Verwaltungsleistungen komplett online erreichen und in Anspruch nehmen können muss. Hierzu werden sehr viele Lösungen entwickelt.

Seit sehr vielen Jahren stellen wir – beziehungsweise mein dafür zuständiger Geschäftsbereich – zwischen 400 und 600 Fachverfahren für die Landesverwaltung zur Verfügung. Wir unterstützen die Verwaltungsarbeit mit IT-Technik und damit treiben wir im weitesten Sinne auch Digitalisierung voran. Beispielprojekte sind etwa auch das neue Insolvenz-Portal, über das die Verwaltung von Insolvenzen auch online abgewickelt werden soll. Das Landesamt für Besoldung und Versorgung, das für die Abwicklung der Bezügezahlungen für inzwischen etwa 680.000 Landesbedienstete und Pensionäre zuständig ist, unterstützen wir bei der Zahlbarmachung der Bezüge. Mein Geschäftsbereich macht also nichts anderes als Digitalisierung.

Was sind die Gründe für eine digitale Transformation? Wer braucht Digitalisierung?

Ich komme ja aus dem Bereich der öffentlichen Verwaltung. Es gibt zum einen Landes- und Bundesgesetze, in denen steht, dass wir Digitalisierung umsetzen sollen. Der richtige Einsatz von Digitalisierung optimiert und vereinfacht Verwaltungsprozesse. Wir als IT.NRW sind ja auch statistisches Landesamt und wissen in dieser Funktion recht genau, wie unsere Bevölkerungsprognose in Nordrhein-Westfalen aussieht. Somit wis-

sen wir, dass wir in den nächsten fünf Jahren ein Drittel des Personals in der Verwaltung aus Altersgründen verlieren werden. Hier kommt auch eine neue Generation, die es gewohnt ist, digitaler zu arbeiten. Gleichzeitig nimmt die Menge der Aufgaben immer mehr zu, die Verarbeitung von Informationen und Daten wird dabei mehr und mehr eine Rolle spielen. Digitalisierung ist deshalb auch ein wichtiges Mittel, um sicherzustellen, dass auch in fünf Jahren die öffentliche Verwaltung noch handlungsfähig ist. Denn hier stehen wir vor einem großen Transformationsprozess. Ein konkretes aktuelles Beispiel ist hier etwa die Auszahlung der Corona-Hilfen. Diese müssen irgendwie zu den Menschen gebracht werden. Hier sind sowohl die Sorgen als auch die Summen riesig und die Zahl der Empfänger ist auch riesig, daher muss hier etwas getan werden. Unserer Meinung nach wird dies ohne Digitalisierung und Automatisieren gar nicht mehr funktionieren können und die öffentliche Verwaltung wird nicht mehr in der Lage sein, ihre Aufgaben zukünftig noch mit traditionellen Arbeitsprozessen erledigen zu können.

Hinzu kommt ein allgemeiner Erwartungsdruck von außen: Jeder hat privat natürlich sein Smartphone, mit dem alles Mögliche online erledigt wird. Daher erwartet man natürlich auch von der öffentlichen Verwaltung, dass sie sich so modern organisiert, dass man mit ihr auch digital kommunizieren kann oder ein Anliegen online vortragen kann.

Als nachgeordneter Landesbetrieb des Ministeriums für Wirtschaft, Innovation, Digitalisierung und Energie (MWIDE) arbeiten wir derzeit überwiegend für Minister Andreas Pinkwart, der sich das Voranschreiten der Digitalisierung auf seine Fahnen geschrieben hat. Zu seinem Bereich gehört auch der Beauftragte der Landesregierung für Informationstechnik (CIO). Diese Rolle wird seit September letzten Jahres durch Prof. Dr. Andreas Meyer-Falcke wahrgenommen. IT.NRW ist dem CIO fachlich unterstellt.

Wir sind praktisch sein Rechenzentrum, sein IT-Provider, der die Leistung für eine Digitalisierung von Verwaltungsarbeiten erbringt. Elektronische Akten, elektronische Laufmappen oder aber auch die Errichtung von Scanzentren für Papierpost zur Überführung in den digitalen Prozess kann ich hier als bedeutende Dinge nennen, mit denen wir uns zurzeit beschäftigen.

Was bedeutet Digitalisierung? Wo findet sie statt?

Für mich ist ein Punkt wichtig. Nämlich der Unterschied zwischen Elektronifizierung und der tatsächlichen Digitalisierung. Im Moment beschäftigen wir uns intensiv mit Elektronifizierung, also die Überführung von Papierakten in elektronische Akten. Wenn wir diese elektronischen Akten dann aber genau auf die gleiche Art und Weise weiterbearbeiten, wie wir das vorher getan haben, ist eigentlich noch nicht viel gewonnen. Das ist für mich wie der Wechsel von der Vinyl-Schallplatte zur CD. Digitalisierung ist, wenn sich dann tatsächlich auch neue Geschäftsmodelle, neue Geschäftsprozesse und neue Verwaltungsprozesse daraus ergeben. Das ist dann für mich die Königsdisziplin. Dazu brauchen wir allerdings erst mal die Basiskomponenten, um überhaupt in die elektronische Welt hineinzukommen. Ich habe vor ein paar Jahren auf einer Konferenz in Südtirol erfahren, dass es in Italien ein Gesetz gibt, das es der Verwaltung verbietet, Bürger nach Daten zu fragen, die sie an anderer Stelle schon erfasst hat. Der Auftrag der dortigen Verwaltung besteht darin, sich selbst um die Datenbeschaffung zu kümmern, bevor sie die Bürger mit Anfragen behelligt. Davon sind wir aus Datenschutzgründen noch weit entfernt, aber das ist es, was ich mir darunter vorstellen. Ich habe selbst drei Kinder und wurde immer wieder regelmäßig nach deren Alter für einen Kindergeldberechtigungsnachweis gefragt, obwohl anhand der der Verwaltung bekannten Geburtsdaten leicht feststellbar gewesen wäre, dass meine Kinder noch minderjährig sein müssten. Aus Papierakten PDFs zu machen, ist für mich noch Elektronifizierung und noch keine wirkliche Digitalisierung – diese findet erst statt, wenn wir die Akten durch digitale Prozesse ersetzen können.

Welche Herausforderungen können sich aus der Digitalisierung ergeben? Welche Fragen sind zu lösen?

Für uns als öffentlicher IT-Dienstleister ist es natürlich immer sehr wichtig, dass die Daten, die uns als Verwaltung anvertraut werden, sicher sind, verfügbar sein müssen und wir sicherstellen können, dass sie authentisch und nicht verändert worden sind. Die Daten dürfen nur von denjenigen Personen eingesehen werden können, die tatsächlich dazu auch Zugang haben sollen. Dies betrifft also alles, was mit dem Thema IT-Sicherheit, Datenschutz und Datensicherheit zu tun hat. An uns als öffentlichen IT-Dienstleister werden natürlich noch höhere Ansprüche gestellt werden. Das finde ich aber auch völlig in Ordnung. Die Argumentation ist bisher immer, dass der Umgang mit den Papierakten zukünftig in der digitalen Welt mindestens genauso gut sein würde. Dies reicht allerdings nicht wirklich und damit kann man Datenschützer auch nicht wirklich beruhigen. Vor vielen Jahren hatte ich so ein typisches Erlebnis im Gebäude einer Behörde, als alle Mitarbeiter im Sommer ihre Fenster geöffnet hatten, als dann plötzlich viele Akten vom Wind erfasst nach draußen flogen und viele Menschen hektisch auf den Parkplatz eilten, um die Akten alle wieder einzusammeln. Die Tatsache, dass alle Dokumente wiedergefunden wurden, machte die Situation dabei nicht besser.

Welche Berufe sind von der Digitalisierung betroffen? Welche Berufe könnten in Zukunft verschwinden?

Von Digitalisierung betroffen sind meiner Meinung nach viele klassische Verwaltungsaufgaben. Hier kann sehr viel vollständig

automatisiert und digitalisiert werden. Dabei können sich Berufsbilder dramatisch verändern und schließlich könnten Berufe sogar tatsächlich verschwinden. Die eigentliche Sachbearbeitung durch Menschen fände dann nur noch für die Fälle statt, die wirklich komplex oder schwierig zu bearbeiten sind oder bei denen es einen Ermessensspielraum bei Entscheidungen gibt, wo die Maschine nicht weiterweiß. In einigen Bereichen werden die Einflüsse möglicherweise nicht so groß sein, aber selbst im medizinischen Bereich, etwa in der Pflegedokumentation, wird die Digitalisierung eher zunehmen.

Wo steht Deutschland bei der Digitalisierung? Gibt es regionale Unterschiede?

Ich kann nur aus der Sicht der öffentlichen Verwaltung sprechen. Meines Erachtens ist Deutschland nicht hoffnungslos abgeschlagen, aber es steht auch nicht an der Spitze der Digitalisierung. Es gibt regionale Unterschiede und Nordrhein-Westfalen könnte im bundesdeutschen Ländervergleich bestimmt noch zulegen. Allerdings ist Nordrhein-Westfalen hier aufgrund seiner Größe und der digitalen Historie auch eine ganz besondere Herausforderung.

Einige Länder sind da inzwischen schon weiter. Allerdings ist Datenschutz aus guten Gründen ein riesiges Thema in Deutschland, da die Erfahrungen in der Vergangenheit nicht immer positive waren. Manchmal gibt es für unseren Föderalismus ja auch Gründe, warum er so ist, wie er ist. Wenn man nur auf die Digitalisierung schaut, werden manche Dinge dadurch nicht einfacher.

Ein Beispiel dafür, wie sich Datenschutzbestimmungen auswirken können, sind die Vorgaben des Zensusgesetzes, das prinzipiell für die ganze EU gültig ist. Hierin steht, dass die alle



Zur Person: Dr. Jan Mütter

Dr. Jan Mütter (55 Jahre, Physiker) ist seit mehr als 20 Jahren bei IT.NRW beziehungsweise der Vorgängerorganisation LDS NRW beschäftigt. Seit über 10 Jahren verantwortet er den Projekt-Bereich, und seit 8 Jahren den Geschäftsbereich IT-Lösungen, d.h. Lösungs-Projekte, Softwareprojekte, etc. Jan Mütter beschäftigt sich seit fast 20 Jahren mit agilen Projektmanagementmethoden, sowie ihrer Einführung in Softwareentwicklungsteams. Er ist nach den Regeln der IPMA nach Level A und B zertifizierter Projektdirektor, sowie nach Prince2 und MSP practitioner.

zehn Jahre erfassten Daten, die ausschließlich zum Zwecke der Volkszählung erhoben werden dürfen, nicht weiterverwendet werden können, um die Register der Meldeämter zu bereinigen. Das sogenannte Rückspiel-Verbot führt bei uns zum Beispiel dazu, dass die von uns ermittelten amtlichen Einwohnerzahlen von den Einwohnerzahlen der Städte und Gemeinden zum Teil beachtlich abweichen. In den Niederlanden hingegen gibt es dagegen ein zentrales Register, das binnen einer Woche mit SQL-Unterstützung Daten zur Volkszählung aktualisiert. Hier machen sich auch unterschiedliche Ansätze mit und ohne Föderalismus und Dezentralisierung bemerkbar.

Was ist der Unterschied zwischen Digitalisierung und Automatisierung?

Digitalisierung bedeutet, dass Geschäftsprozesse oder Verwaltungsprozesse wirklich komplett digital ablaufen und man auch an jeder Stelle nachsehen kann, wo man steht. Digitalisierte Prozesse bieten die Möglichkeit, den Fortschritt oder Bearbeitungsstand jederzeit verfolgen zu können. Wo befindet er sich jetzt gerade? Wer bearbeitet etwas? Gibt es noch offene Fragen? Diese Fragen könnten beantwortet werden. Automatisierung bedeutet nur, dass etwas automatisch abläuft, aber dies ginge auch ohne IT-Technik.

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

Wir nennen dies bei uns im Land Nordrhein-Westfalen nicht nur Digitalisierung, sondern wirklich „digitale Transformation“. Wir verändern die öffentliche Verwaltung, weg von der Art, wie sie seit 150 Jahren nach Max Weber funktioniert hat, in eine komplett neue Verwaltung. Viele von den Prozessen, die heute ablaufen, laufen ja heute so ab, weil es damals mit der Technik nicht anders ging. Jetzt haben wir jedoch auf einmal technische Möglichkeiten, mit denen es anders gehen könnte. Diese müssen wir jetzt aufgreifen und nutzen, dann funktioniert Verwaltung auch digital. Auch dazu ein Beispiel: Bisher ist es immer so, dass man bei der Geburt

eines Kindes Kindergeld beantragen kann. Der Bürger wird zum Bittsteller, über dessen Anliegen befunden wird.

Wenn man über Digitalisierung redet, stellt sich die Frage: Wie sähe hier der optimale Prozess aus? Der Anspruch auf Kindergeld begründet sich eigentlich schon allein aus der Existenz des Kindes. Dieser beginnt mit dem Geburtsdatum und dauert bis zum achtzehnten Lebensjahr des Kindes. Lediglich danach müsste man, beispielsweise bei Aufnahme eines Studiums oder einer Ausbildung, überprüfen beziehungsweise nachweisen, ob eine Verlängerung des Kindergeldanspruchs infrage kommt. Der Automatismus würde hier 18 Jahre lang eine Antragstellung von Millionen Eltern einsparen.

Eigentlich gibt es auch überhaupt keinen Grund dafür, dass man Kindergeld beantragen muss, da ein gesetzlicher Anspruch besteht, dass mindestens 18 Jahre lang Kindergeld gezahlt wird. Dem Meldeamt müsste nicht nur das Geburtsdatum eines neugeborenen Kindes, sondern auch gleichzeitig noch die Kontonummer der Eltern mitgeteilt werden. Dann würde automatisch einfach 18 Jahre Kindergeld überwiesen und nach 18 Jahren wird damit wieder aufgehört. Damit hätten wir praktisch schon etwa 90% des Standardgeschäfts abgebildet. Digitalisierung bedeutet auch, den Sinn und Zweck der heutigen Prozesse selbst infrage zu stellen und nicht nur durch neue technische Möglichkeiten abzubilden.

Eine digitale Transformation würde dazu führen, dass Bürgerinnen und Bürger bei Bestehen eines gesetzlichen Anspruchs an dieser Stelle nicht mehr nur als (Online)-Antragsteller oder als Bittsteller betrachtet werden, sondern bei einem zeitgemäßen Verständnis von öffentlicher Verwaltung auch das Selbstverständnis der Institution verändern.

Mit der Digitalisierung wird das Angebot der öffentlichen Verwaltung für den Bürger online und damit sprechzeiten- und ortsunabhängig und asynchron. Dies ist der Grundgedanke des Online-Zugangsgesetzes: Bis Ende des Jahres 2022 müssen bis zu 5000 Verwaltungsleistungen für Bürger und Unternehmen vollständig online abgewickelt werden können.

Fundstück: die Gründung der Primavera Community

DOAG Online

In der Frühjahrsausgabe 2011 berichtete die Business News über die Gründungsveranstaltung mit 35 Unternehmen.

Vor über 10 Jahren, am 15. September 2010, war es so weit: Nach einem vorangegangenen Kick-off-Meeting gründete sich in Köln die Primavera Community innerhalb der DOAG'schen Community-Familien. Über 60 Teilnehmer aus sieben Branchen und 35 Unternehmen schlossen sich zum Interessenverbund des professionellen Projekter-

minplanungs-Tools von Oracle zusammen. Vorträge, Diskussionen, Networking und Wissensaustausch der Primavera-Nutzer standen dabei im Fokus der Mitglieder, siehe Business News, *Frühjahr 2011*, S. 26.

Pandemie-bedingt ist es – wie bei vielen anderen auch – etwas still um diese DOAG Business Solutions Community geworden,

aber sie lebt! Denn wie uns der DOAG Themenverantwortliche Primavera Sebastian Hunke gerade mitteilte: "Wir wollen sehr gerne unsere Re-Launch-Veranstaltung, die für Juni 2020 geplant war, so bald wie möglich persönlich nachholen, vielleicht im Frühjahr 2022 – und vor allem face-to-face!" Wir freuen uns darauf und sind gespannt!



KI als größte Herausforderung in den nächsten Jahren

Die Corona-Pandemie hat in den letzten 12 Monaten das Leben sowohl im privaten als auch beruflichen Umfeld deutlich und vielleicht nachhaltig verändert.

Die sozialen Kontakte haben durch Homeoffice, virtuelle Meetings und Onlineschulungen und -konferenzen die physischen Treffen und Präsenzveranstaltungen fast vollständig abgelöst. Unternehmen aller Branchen haben sich in ihren Abläufen und Prozessen den veränderten Gegebenheiten angepasst. Doch wie sehen die Unternehmen die Auswirkungen auf die Zukunft?

Bei der aktuellen Zukunftsstudie 2021 geben nach einem Zwischenstand (Quelle: 2021, Prof. Tim Bruysten, richtwert GmbH, www.richtwert.eu) 84 – 85 % der befragten Unternehmen aus diversen Branchen und unterschiedlicher Größe an, dass der Markt sich unwiederbringlich verändert habe. Eine Teilnahme an der laufenden Studie ist unter www.zukunftsstudie.com zurzeit noch möglich.

Dabei werden auf die Frage nach den Disruptionen mit der größten Relevanz neben regulatorischen Anforderungen die Veränderungen durch Künstliche Intelligenz (KI) als die maßgebliche Disruption genannt. In der Gesamtbetrachtung steht KI auf der ersten Position (siehe Abbildung 1), innerhalb der Konzerne nach der Regulatorik auf Rang zwei (siehe Abbildung 2).

Schlagworte wie Data Scientist, Robotics und Machine Learning werden demzufolge weiter in den Sprachgebrauch der Unternehmen übergehen und die Automatisierung von Prozessen, verbunden mit einer intelligenten Analyse, wird die Gestaltung der Unternehmen prägen.

Dieser Entwicklung wird auch die DOAG für ihre Mitglieder Rechnung tragen und im kommenden Sommer und Herbst mit Beiträgen in den Magazinen und bei den Online-Events sowie neuen Formaten die Thematik KI aufnehmen.

Armin Wildenberg
Leiter Data Analytics Community

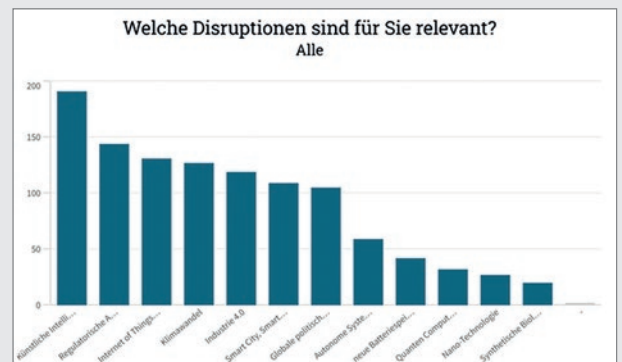


Abbildung 1 (Quelle: 2021, Prof. Tim Bruysten, richtwert GmbH, www.richtwert.eu)

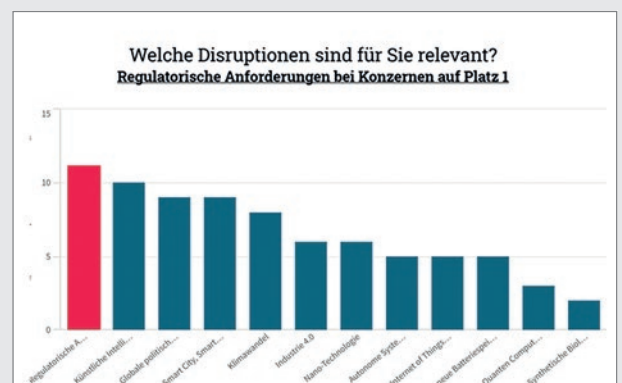


Abbildung 2 (Quelle: 2021, Prof. Tim Bruysten, richtwert GmbH, www.richtwert.eu)



Lockdown oder: Wie ich lernte, die Cloud zu lieben – Teil 1

Dr. Jörg Domaschka, Institut für Organisation und Management von Informationssystemen
Steffen Moser, School of Advanced Professional Studies (SAPS)
Thomas Nau, Kommunikations- und Informationszentrum (kiz)
Simon Volpert, Institut für Organisation und Management von Informationssystemen
Alle Autoren arbeiten an der Universität Ulm.

Die COVID-19-Pandemie hat im vergangenen Jahr oft schmerzhaft gezeigt, wie grundlegend wichtig eine stabile, durchgehende, sichere und zeitgemäße Versorgung mit IT-Services im täglichen Leben ist. Auch die Hochschulen bilden hier keine Ausnahme. Dabei spielt „Die Cloud“, obwohl schon lange als nächster Schritt der IT-Evolution gehandelt, in diesem Bereich oft nach wie vor nur eine Nebenrolle. Dies ist aus mehreren Gründen verständlich: Zum einen existieren vielfältige unterschiedliche Definitionen und Ansichten über das, was eine Cloud ist beziehungsweise ausmacht. Zum anderen ist die IT oft sehr eng mit lokalen Prozessen und Anforderungen verwoben, die eine On-Premises-Lösung – das scheinbare Gegenteil einer Cloud – geradezu erzwingen.

Diese Serie aus drei Artikeln fasst die Geschehnisse in den Wochen vor und nach der Anordnung des Lockdowns an der Universität Ulm im März 2020 zusammen. Ziel ist es, den Lesern die insbesondere technischen Voraussetzungen und Entscheidungen aufzuzeigen, die es letztendlich erlaubt haben, die Arbeitsfähigkeit von Lehre, Verwaltung und Forschung an der Universität Ulm in sehr hohem Maße zu gewährleisten. Die Serie zeigt das Spannungsfeld auf und fasst die Entstehungsgeschichte unserer Cloud-Lösung für unser Videokonferenzsystem sowie die gemachten Erfahrungen in der Hoffnung zusammen, anderen als Anregung für eigene interne Diskussionen zu dienen.

Teil 1: Hintergründe

Der vorliegende Artikel fokussiert sich auf die Darstellung der organisatorischen und technischen Hintergründe. Vor allem wird die Video-Konferenz-Software BBB vorgestellt, deren Einführung an der Universität Ulm diese Serie beleuchtet. Im Anschluss diskutieren wir Probleme auf Netzwerk-Ebene, die im Betrieb auftreten, aber aufgrund der Konzeption von BBB als Browser-only-Lösung nicht immer gelöst werden konnten.

Teil 2 der Serie fokussiert sich auf die Skalierung von BBB und die Automatisierung des Betriebs. Teil 3 beschäftigt sich dann mit dem Konfigurations-Management.

Was ist die Cloud?

Die Antworten auf die Frage, was denn eine Cloud ist, fallen vielfältig aus. Während die einen an das Hosting virtueller Maschinen denken, ist es für andere die Magie hinter der Synchronisation von Daten zwischen Handy und Tablet. Oder auch die Möglichkeit, einfach und von überall gemeinsam an Dokumenten zu arbeiten. Allen Antworten gemein ist die Unabhängigkeit vom Standort durch die Nutzung des Internets für den Zugriff auf Daten und/oder Rechenleistung beziehungsweise Anwendungen.

Das amerikanische National Institute of Standards and Technology (NIST) [1] definiert sie wie folgt:

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models.

Oder in anderen Worten, es geht um den einfachen, universellen und bequemen Zugriff auf Ressourcen über das Internet.

Diese sind nach Bedarf, also on demand ohne händische Interaktion, verfügbar, lassen sich skalieren und sind „elastisch“. Elastizität macht einen der großen Vorteile einer echten Cloud-Lösung aus. Es bedeutet, dass die Größe und Menge der Ressourcen nach Bedarf wachsen, aber auch schrumpfen kann. Technisch einfach umsetzbar ist elastisches Verhalten in der Regel aber nur für zustandslose Anwendungen beziehungsweise solche, bei denen sich der Zustand gut von der eigentlichen Anwendung trennen und speichern lässt.

Auch die Dienstleistungs- und Einsatzmodelle von Cloud-Lösungen variieren. In unserem Szenario kam die im Red Stack Magazin 06/2019 [2] unter dem Titel „Die eigene Cloud – Wann eine On-Premises-Lösung Sinn macht“ beschriebene bwCloud zum Einsatz. bwCloud ist eine auf OpenStack basierende, private IaaS-Cloud, betrieben im Zusammenschluss mehrerer Baden-Württembergischer Universitäten.

Der Tag X

Das Schreckgespenst des Lockdowns Mitte März 2020 vor Augen, waren in kurzer Zeit auch schwerwiegende technische Entscheidungen mit weitreichenden Konsequenzen zu treffen. Die Anforderungen für den Start des Sommersemesters am 1. April 2020 ließen sich wie folgt in vier Bereiche unterteilen:

- Bereitstellung einer Plattform für ein Semester, das zu 100% online stattfindet. Diese musste die Erstellung, das Bearbeiten und die Verteilung von Videoinhalten ebenso unterstützen wie virtuelle Vorlesungen und Online-Konferenzen.
- Einige Hundert Beschäftigte sollten in kürzester Zeit im Homeoffice eine gewohnte Arbeitsumgebung vorfinden. Dies schloss die Möglichkeit der telefonischen Erreichbarkeit unter der Büronummer ebenso ein wie ein Angebot für Videokonferenzen.
- Eine Storage-Cloud musste das kollaborative Arbeiten an Dokumenten und deren Austausch unterstützen.
- Sehr viele Mitarbeiter/innen und Studierende benötigten sowohl technische als auch organisatorische Unterstützung beim Sprung in die Wolken.

Als besondere technische Herausforderung, nicht zuletzt wegen gegensätzlicher Anforderungen, kristallisierte sich sehr schnell der Bedarf nach einer skalierenden Plattform für die Online-Lehre heraus. Während das seit mehreren Jahren eingesetzte Moodle bereits die notwendigen Anforderungen erfüllte, waren neben der Bereitstellung einer Streaming-Lösung für die Verbreitung asynchroner Lehrinhalte insbesondere der Aufbau und Betrieb einer Plattform zur synchronen Lehre und Koordination zwischen Studierenden und Lehrenden, aber auch für studentische Kleingruppen, ein Arbeitsschwerpunkt. Diese Video-Conferencing-Anwendung steht im Rahmen der Artikel-Serie im Mittelpunkt.

Für eine kommerzielle Lösung sprachen vor allem sowohl die schnelle Verfügbarkeit in der Breite als auch ein geringerer zu erwartender Support-Aufwand. Allerdings gab es zum damaligen Zeitpunkt keine derartige Lösung, die gleichzeitig in die Lehr- und Lernplattform der Universität integrierbar war. Letzteres war beim Open-Source-Projekt *BigBlueButton* (BBB) gegeben, für das jedoch im Kommunikations- und Informationszentrum (kiz) der Universität kein ausreichendes Know-how verfügbar war.

Ausschlaggebend für den letztendlichen Erfolg bei der Umsetzung, nicht nur für BBB, war die Bildung von Task-Forces über die Einrichtungsgrenzen an der Universität hinweg. Auch bei der Bereitstellung der Video-Conferencing-Lösung brachten Mitglieder verschiedener Einrichtungen ihre Expertise in unterschiedlichen Bereichen ein. Hierzu zählen bei BBB vor allem die Erfahrung im Betrieb und der Konfiguration, Kenntnisse im Bereich der Software-Automatisierung und Wissen über etablierte Betriebs-, Management- und Kommunikationsstrukturen.

Hintergrund zu BBB

Bei BigBlueButton (BBB) handelt es sich um eine Client-Server-basierte Video-Konferenz-Software. Die Software steht unter der GNU Lesser General Public License (LGPL) und gilt damit als Open-Source-Software. Ihre Entwicklung wird von Blindside Networks Inc. mit Sitz im kanadischen Ottawa koordiniert. Typischerweise ist BBB als eine On-Premises-

Lösung gedacht, die von der nutzenden Einrichtung selbst auf einer Server-Infrastruktur gehostet wird.

BBB wird an der Universität Ulm seit 2012 von der Einrichtung für berufs begleitende wissenschaftliche Weiterbildung, der *School of Advanced Professional Studies (SAPS)*, eingesetzt. Die SAPS setzt seit ihrem Bestehen auf einen hohen E-Learning-Anteil in ihren weiterbildenden Masterstudiengängen, sodass eine gute Integration des Studiums in den Berufsalltag möglich wird. BBB ist hierbei eines der Werkzeuge, die als Bestandteil des sogenannten „Virtuellen Schreibtischs in der Cloud“ den berufs begleitend Lernenden zur synchronen Kommunikation bereitgestellt wird.

BigBlueButton-Architektur

Architektonisch besteht BBB aus einem Server-Teil und einem Client, wobei Letzterer seit Version 2.2.0 (erschieden im November 2019) in einem HTML5-fähigen Webbrowser abläuft. Entsprechend werden auf Seiten der Nutzer weder Plugins noch Browser-Extensions benötigt. Auch auf lokal installierte Client-Anwendungen wird, im Unterschied zu vielen anderen Videokonferenzsystemen, verzichtet. Ebenso existieren keine dedizierten Apps für mobile Endgeräte auf Android- oder iOS-Basis. Auch hier wird grundsätzlich auf eine Nutzung der dort vorhandenen Browser gesetzt.

Der Server-Teil von BBB kann als Beispiel für die gelebte Unix-Philosophie dienen. Es handelt sich um einen Verbund aus einer Vielzahl von Modulen für die unterschiedlichen Aufgaben in einem Videokonferenzszenario. Neben Eigenentwicklungen des BBB-Teams seien zwei wichtige Kernmodule angeführt, die zur Anwendung kommen: die VoIP-Telekommunikationslösung FreeSWITCH sowie der Medienserver Kurento. FreeSWITCH ist unter den BigBlueButton-Modulen in der Regel die Komponente, die am meisten Rechenzeit benötigt. Kurento dagegen verursacht in aller Regel die größte Menge an Netzwerkverkehr.

FreeSWITCH dient dem Kodieren und Dekodieren, dem Mischen und Verteilen der Audiosignale der bidirektional per Audio kommunizierenden Konferenzteilnehmer. Ebenso bindet FreeSWITCH Teilnehmer ein, die sich per Telefon in Video-

Konferenzen einwählen. Kurento kommt dagegen beim Verteilen von Videoströmen, Screen-Sharings und dem Audio-Strom für die nur-zuhörenden Teilnehmer zur Anwendung. Beide Dienste stellen weiche Echtzeitanforderungen, bei deren Nichteinhaltung die Dienstgüte leidet. Insbesondere führt eine Überlast bei FreeSWITCH sofort zu Qualitätsverlusten, die sich zuallererst durch kurze Unterbrechungen und Verzerrungen im Audiosignal bemerkbar machen. Entsprechend müssen im Falle von virtualisierten Hostings Überbuchungen vermieden oder zumindest deren Effekte im Blick behalten werden.

Echtzeitkommunikation in BBB

Die Medienkommunikation zwischen BBB-Client und BBB-Server funktioniert sowohl für Audio als auch für jedes gesendete und empfangene Video, über getrennte WebRTC-Sitzungen. Während das Audiosignal auf dem Server gemischt wird, werden die Video- und Screen-Sharing-Ströme grundsätzlich nur verteilt. Entsprechend entstehen für jeden Teilnehmer maximal zwei Audio-Ströme (Senden des eigenen Stroms und Empfangen des gemischten Stroms), während pro Teilnehmer maximal N+1 Video-Ströme entstehen (Senden des eigenen Stroms, Empfangen von N-1-Video-Strömen anderer Teilnehmer, Empfangen eines Screensharing-Stromes) [3]. Die benötigte Rechenlast hängt damit in grober Näherung quadratisch von der Anzahl der Teilnehmenden einer Konferenz ab. Weitere parallel stattfindende Konferenzen oder Breakout-Sitzungen einer Konferenz tragen in der Regel linear zur Serverlast bei.

WebRTC-Verbindungen [4] können sowohl über UDP (präferiert) als auch über TCP aufgebaut werden. Es ist jedoch zum Verbindungsaufbau unbedingt nötig, dass sich Client und Server über die Art der Verbindung einigen. Dies umfasst neben der Wahl des Transportprotokolls, UDP vs. TCP, und der Wahl der Audiobeziehungsweise Video-Codecs auch die Feststellung, ob eine direkte Verbindung zwischen beiden Partnern möglich ist und welche IP-Adressen beziehungsweise Ports dabei verwendet werden sollen. Hierbei kommen sogenannte STUN-Server [5] zur Feststellung der eigenen

Remote-IP-Adresse(n) zum Einsatz. Weiterhin werden TURN-Server verwendet, wenn keine direkte bidirektionale UDP-Konnektivität zwischen Client und Server möglich ist. Die TURN-Server fungieren dann als Proxy und bieten in Richtung Client an, WebRTC über TCP abzuwickeln. Sowohl STUN als auch TURN werden beispielsweise von Coturn [6] realisiert.

Der Ablauf der Aushandlung ist dabei über das Internet Connectivity Establishment (ICE)-Protokoll vorgegeben, das in der Regel in SDP beziehungsweise SIP eingebettet ist (der geneigte Leser sei hierzu an die RFCs 8839 und 8445 verwiesen). In BBB findet die Verhandlung mittels SDP/SIP über einen https-/WebSocket-Kanal in einem Node.js-Prozess statt. Im Anschluss daran etablieren Client und Server mehrere WebRTC-Verbindungen.

4 Der Einfluss von Netzwerkaspekten im Betrieb

Die Bedeutung der Netzeigenschaften zwischen Client und Server, im Sinne einer Gesamtsicht, steht insbesondere in einem weichen Echtzeitsystem der Relevanz von Server- und Client-Eigenschaften in nichts nach. Mit der Entscheidung, BBB als hochskalierten Dienst on premises in der bwCloud zu betreiben, war klar, dass die Funktionsfähigkeit des Universitäts-Netzwerks, aber auch seine Anbindung an das Internet, ein kritisches Erfolgskriterium ist.

Zu Beginn der Pandemielage wurde deswegen die Netzinfrastruktur auf der Seite der Cloud-Infrastruktur um den Faktor 4 erweitert, ebenso wurden Peering- und Transitkapazitäten zwischen Hochschulnetz und den kommerziellen Netzzugangsanbietern in Deutschland erheblich ausgebaut. Dass dies notwendig war, zeigte sich bei den Lastbeobachtungen. Ebenso wurden die virtuellen Maschinen, in denen BBB ausgeführt wurde, auf bwCloud-Seite mit 10 Gbit/s Netzwerk-Bandbreite ausgestattet.

Während die Server-Seite durch den Kapazitätswachstum den Anforderungen gewachsen war, begegneten wir über den First-, Second- und Third-Level-Support Netzengpässen an anderen Stellen deutlich: Vereinzelt führen lokal überlastete Carrier-Grade-NAT-Komponenten bei den Zugangsprovidern zu einem Versagen des STUN-Protokolls infolge von Timeouts. Um solche cgNAT-Komponenten möglichst zu umgehen, wurde die Ent-

scheidung getroffen, BBB als Dual-Stack-Service anzubieten. Leider zeigte sich besonders FreeSWITCH hier unflexibel: Wird die Management-Verbindung (https/Websocket) beispielsweise via IPv6 aufgebaut, so wird auch für die WebRTC-Kommunikation nur ein IPv6-Stack angeboten – analog für IPv4. Etwaige Firewall-Probleme, die erst bei den WebRTC-Aushandlungen festgestellt werden, können also durch die Verwendung des jeweils anderen IP-Stacks nicht umschifft werden. Kurento schlägt sich hier flexibler und geht bei den WebRTC-Aushandlungen mit beiden IP-Stacks ins Rennen.

Die Hauptursache für schlechte Nutzererlebnisse bei Videokonferenzen stellen jedoch mit großem Abstand die häuslichen Netzinfrastrukturen auf Benutzerseite dar. Sind sie bei Lehrenden anzutreffen, sind die Auswirkungen durchaus erheblich, da der Uplink-Datenstrom für das Audiosignal dann bei allen Teilnehmern einer Videokonferenz beeinträchtigt wird.

In der Praxis erweisen sich hier die Störungen des häuslichen WLANs als die häufigste Ursache. Eine mangelnde Funkausleuchtung des Arbeitsbereichs, eine nutzer verursachte Netzüberlastung, etwa durch parallel laufende Cloud-Backups sowie eine Kanalüberlastung durch störende Nachbar-WLANs galt es in ausführlichen Supportsitzungen zu ermitteln und zu beseitigen. Hier zeigt sich der Nachteil der rein browserbasiert arbeitenden Anwendung. Die gerade für den Nutzersupport wichtigen historischen Verbindungsstatistiken sind nur aufwendig zu gewinnen und bedürfen der aktiven Mitarbeit des Nutzers. Speed-Tests, die über eine längere Zeit Latenz, die Paketverlustrate und den Latenz-Jitter messen, erwiesen sich im Second- und Third-Level-Support als vorteilhaft. Als besonders hartnäckig erwiesen sich die in einigen Notebooks implementierten Funktionen zur Ortung des Geräts. Diese erfolgt über das Abtasten der SSID-Broadcasts der WLAN-Access-Points in der Umgebung. Insbesondere Webbrowser, die abhängig von parallel geöffneten Webseiten gerne die Standortbestimmung anfordern, können so zu periodischem Packet-Loss beitragen, der sich dann in einem Echtzeitkommunikationssystem in wiederkehrenden Audio-Rucklern zeigt. Probleme dieser Art sind nicht neu, bleiben jedoch bei der Nutzung von asynchronen Streaming-Diensten dank der Pufferung verborgen. Bei Videokonfe-

renz-Werkzeugen, die mit einem lokalen Client arbeiten, könnte gegebenenfalls der Client darauf Einfluss nehmen, dass das Betriebssystem solche Standort-Abfragen temporär unterlässt.

In einigen Fällen erwiesen sich WLAN-Repeater und Power-Line-Communication-Lösungen als problematisch, Letztere interferieren gerne mit Super-Vectoring, das auf den in Deutschland weit verbreiteten VDSL2-Anschlüssen gefahren wird und eine Bandbreite von 35 MHz aufweist. Das DSL-Spektrum überlappt in weiten Teilen mit dem der Power-Line-Adapter.

Zusammenfassung

Die Wahl von BBB als Online-Konferenz-Werkzeug war richtig und wichtig. Während im April 2020 viele kommerzielle Anbieter für Video-Conferencing aufgrund der gestiegenen Nachfrage in die Knie gingen, konnte an der Universität Ulm das Semester weitestgehend ohne technische Störungen ablaufen. Es war uns trotz des geringen zeitlichen Vorlaufs gelungen, zum Beginn des Sommersemesters 2020 Lehrenden und Studierenden ein stabiles BBB-Cluster zur Verfügung zu stellen. Die konkrete Lösung und deren technische Umsetzung werden in den folgenden Teilen dieser Artikel-Serie beleuchtet.

Durch die Wahl von BBB konnten auch kurzfristig Personen von der Online-Lehre überzeugt werden, die kommerziellen Anbietern aus Datenschutzgründen skeptisch gegenüberstehen. Gleichzeitig kann durch den Aufbau umfassender FAQs und einer Knowledge Base inzwischen der überwiegende Teil an Support-Anfragen bereits vor oder spätestens vom First-Level-Support an beantwortet werden. Dies gilt umso mehr, als das Debuggen von Netzwerk-Problemen auf Nutzer-Seite im Einzelfall sehr aufwendig ist. Entsprechend kann der reine Betrieb auf technischer Seite relativ personalarm durchgeführt werden.

Referenzen

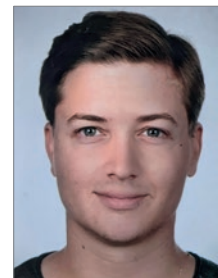
- [1] <https://csrc.nist.gov/publications/detail/sp/800-145/final>
- [2] https://www.doag.org/formes/pub-files/11888839/06_2019-Red_Stack_Magazin-WEB.pdf
- [3] Seit einigen Versionen unterstützt BBB auch mehrere Video-Kameras pro Teil-

nehmer, sodass diese Betrachtung so nicht vollständig korrekt ist. In der Praxis haben jedoch nur sehr wenige Teilnehmer mehr als einen Video-Strom.

- [4] <https://webrtc.org/>
- [5] https://de.wikipedia.org/wiki/Session_Traversal_Utility_for_NAT
- [6] <https://github.com/coturn/coturn>



Thomas Nau
thomas.nau@uni-ulm.de



Simon Volpert
simon.volpert@uni-ulm.de



Jörg Domaschka
joerg.domaschka@uni-ulm.de



Steffen Moser
steffen.moser@uni-ulm.de



Safety First – Sicherungskonzepte von Oracle-E-Business-Suite-Umgebungen für Oracle Cloud Infrastructure

Johannes Michler, PROMATIS Gruppe, Ettlingen (TechnologieRegion Karlsruhe)

Die Oracle Cloud Infrastructure (OCI) wird von vielen unserer Kunden für den Betrieb ihrer Oracle E-Business Suite verwendet. Besonders, wenn nicht nur Entwicklungs- und Testsysteme in der OCI betrieben werden, wird hier ein solides Konzept zur Sicherung (und Wiederherstellung) der Umgebung für den Katastrophenfall benötigt – egal ob dieser durch Benutzer- oder Systemfehler ausgelöst wird. Sehen wir uns die möglichen Optionen hierfür einmal genauer an.

Grundkonzepte und Terminologie – RPO und RTO

Die beiden wichtigsten Begriffe beim Entwurf einer Backup-Strategie sind zweifellos Recovery Point Objective (RPO) und

Recovery Time Objective (RTO). RPO ist die akzeptierte Datenmenge, deren Verlust im Falle einer Katastrophe in Kauf genommen wird. Beispielsweise bedeutet ein RPO von 30 Minuten, dass man nach Eintritt einer Katastrophe in keiner

Situation mehr als 30 Minuten an Transaktionen verlieren möchte. RTO bezeichnet die Zeit, die benötigt wird, um die Instanz nach einer Katastrophe wieder betriebsbereit zu machen. Die Oracle-Dokumentation der Datenbank – insbe-

sondere „High Availability Overview“ – bietet weitere Einzelheiten dazu (vgl. [1]).

OCI-Isolationsgrade

Es gibt drei mit der Oracle Cloud Infrastructure verbundene Isolationsgrade, die zum Schutz vor Ausfällen beitragen, siehe dazu die OCI-Dokumentation:

- **Region:** Eine Region ist ein stark isolierter Teil der Oracle-Cloud-Infrastruktur, der sich in einem geografischen Gebiet befindet, zum Beispiel EU-Frankfurt oder US-West (Phoenix). Die regionale Trennung bietet beispielsweise sogar Schutz vor großen Naturkatastrophen.
- **Verfügbarkeitsdomäne (Availability Domain; abgekürzt mit AD):** Die meisten Regionen sind in drei Verfügbarkeitsdomänen unterteilt. Dabei handelt es sich um isolierte Datenzentren, wodurch ein gleichzeitiger Ausfall sehr unwahrscheinlich ist. Da Verfügbarkeitsdomänen die Infrastruktur (wie Stromversorgung oder Kühlung) oder das interne Netzwerk der Verfügbarkeitsdomänen nicht gemeinsam nutzen, ist es unwahrscheinlich, dass ein Ausfall in einer Verfügbarkeitsdomäne innerhalb einer Region die Verfügbarkeit anderer Domänen innerhalb derselben Region beeinträchtigt.
- **Fehlerdomäne:** Dies ist eine Partition innerhalb eines Rechenzentrums. In dem Recheninstanzen in verschiedenen Fehlerdomänen platziert werden, hat der Ausfall einer physischen Maschine in Fehlerdomäne 1 keine Auswirkungen auf eine physische Maschine (und deren virtuelle Maschinen) in Fehlerdomäne 2.

Block, Object und File Storage

Die Oracle Cloud-Infrastruktur bietet drei persistente Speichertypen mit unterschiedlichen Vor- und Nachteilen. Alle drei Typen sind auch für Backups relevant:

Block Storage ist ein Speichertyp, der an eine Recheninstanz entweder als Boot- oder als zusätzliches Volume angeschlossen ist. In der Regel erfolgt die Anbindung über iSCSI. Obwohl die Performance sehr hoch ist, befindet sich der Block Storage in nur einer Verfügbarkeitsdomain (Domain, in der die Recheninstanz läuft). Block-Volu-

mes können bequem zu definierten Zeiten durch Backup-Richtlinien gesichert werden, wobei diese Backups auch regionenübergreifend in Object Storage abgelegt werden können. Alle Daten von Oracle-E-Business-Suite-Instanzen (Apps- und DB-Tier) befinden sich normalerweise auf Block-Volumes (siehe [2]).

Object Storage kann als „Web-Service“ betrachtet werden, mit dem besonders große Objekte gespeichert und abgerufen werden können. Wie in [3] genannt, handelt es sich um einen Service „pro Region“; Object Storage ist dabei durch die automatische Speicherung zahlreicher Kopien über mehrere Verfügbarkeitsdomains hinweg äußerst robust.

File Storage bietet einen NFS-Einhängpunkt, der an mehrere Recheninstanzen gemountet werden kann, letztendlich auch über Verfügbarkeitsdomains hinweg (oder sogar Regionen, wenn diese verbunden sind). Während der Dienst ein „pro AD“-Dienst ist, kann sich der File Storage in AD1 befinden und eine Recheninstanz in AD2, was einen bequemen Zugriff auf den Speicher ermöglicht (vgl. [4]). Der Dienst bietet Snapshot-Funktionen und speichert mehrere (dauerhafte) Kopien aller Daten (innerhalb eines AD).

Ziele dieser Diskussion

Für den weiteren Umfang dieses Artikels werde ich Strategien beschreiben, die dabei helfen, ein RPO und RTO (für Produktionssysteme) von jeweils etwa 30 Minuten zu erreichen. Die Strategie sollte in der Lage sein, Ausfälle einer Verfügbarkeitsdomain zu bewältigen, muss aber nicht „regionale Ausfälle“ abdecken.

In diesem Beitrag werde ich drei verschiedene Szenarien behandeln, da für diese unterschiedliche Verfahren geeignet sind:

- Umgang mit Entwicklungsinstanzen
- Umgang mit Conference Room Pilot oder anderen Testinstanzen
- Umgang mit Produktivinstanzen

Entwicklungsinstanzen

Bei Entwicklungsinstanzen ist ein komplettes Backup der Instanz in der Regel nicht notwendig. Im Falle einer Katastro-

phe ist man normalerweise in der Lage, eine neue Entwicklungsinstanz als neue Kopie des Produktionssystems einfach zu erstellen. Wenn alle Entwickler nach „gängigen Best Practices“ arbeiten, befindet sich ihr gesamter Quellcode in einem Quellcode-Verwaltungssystem und kann wieder problemlos auf dem neuen Entwicklungssystem installiert werden.

In der Realität ist dies jedoch nicht immer der Fall. Insbesondere bei der PL/SQL- und APEX-Entwicklung ist es eine weit verbreitete Praxis, direkt in der Datenbank zu entwickeln und nur gelegentlich den Quellcode in ein Versionskontrollsystem zu übernehmen – oft erst beim Übergang von der Entwicklung zum Testen.

Um dieses Szenario zu bewältigen, ist es ratsam, auch „einige Backups“ von Entwicklungssystemen durchzuführen. In der Regel ist es jedoch ausreichend, die folgenden Inhalte zu sichern:

- (Benutzerdefinierte) Datenbankobjekte wie Pakete, Prozeduren oder Views: Diese können leicht durch expdp mit einem Filter auf Objektnamen im Muster „XX%“ gesichert werden, der beispielsweise um 8 Uhr morgens, mittags und um 16 Uhr nachmittags ausgeführt wird. Das Ergebnis kann auf File Storage in einer anderen Verfügbarkeitsdomain gesichert und von dort nach Erstellung einer neuen Entwicklungsinstanz wiederhergestellt werden.
- APEX-Anwendungen können mit sqlcl gesichert werden (siehe [5]). Dies kann wiederum mehrmals täglich ausgeführt und das Ergebnis auf File Storage geschrieben werden.
- XML Publisher Reports werden in der Tabelle xdo_lobs gespeichert. Diese Tabelle kann regelmäßig unter Verwendung von expdp gesichert werden.

Wir packen alle drei Exporte in ein Shell-Skript, das alte Daten nach vier Wochen bereinigt und so ein Sicherheitsnetz für Entwickler bietet, die ihren Quellcode nicht sofort in das GIT-Repository einchecken. Das bedeutet ein RPO von 4 Stunden (nur für den Quellcode) und ein RTO von 2 bis 12 Stunden (Erstellung einer neuen P2T-Kopie und Wiederherstellung des aktuellsten Quellcodes). Die Kosten sind minimal – ca. 1 GB werden im File Storage gespeichert.

CRP- und Testinstanzen

CRP- oder Test-Instanzen können vor allem bei der Umsetzung eines neuen Oracle-E-Business-Suite-Implementierungsprojekts eine entscheidende Rolle spielen. Oftmals ist es gar nicht so einfach, diese durch neue Kopien der Produktion zu ersetzen, besonders während der heißen Phasen, beispielsweise im Rahmen eines Benutzerakzeptanztests.

Ich habe gute Erfahrungen mit Block-Volume-Backups gemacht, die entweder manuell oder zu festen Zeiten (teils inkrementell) unter Verwendung benutzerdefinierter oder vorgefertigter Backup-Richtlinien erstellt wurden. Solche Backups können leicht vom Apps- und DB-Block-Volume sowie vom Boot-Volume erstellt werden (vgl. [6]). Die Backups sind „konsistente Momentaufnahmen“ der zugrunde liegenden Volumes zum Zeitpunkt ihrer Erstellung. Es ist einfach, aus den Backups wieder Boot- und Daten-Volumes zu erstellen und dann eine neue Datenbank- und Apps-Tier-VM-Instanz basierend auf diesem Backup zu starten. Dies kann innerhalb weniger Minuten geschehen.

In der Regel füge ich eine Richtlinie wie folgt an:

- Durchführung eines vollständigen Backups jeden Sonntagmorgen um 2 Uhr, das 2 Wochen behalten wird.
- Durchführung eines täglichen inkrementellen Backups um 4 Uhr, das 5 Tage behalten wird.

Das bedeutet ein RPO von bis zu einem Tag und ein RTO von ca. 30 Minuten. Die Kosten hängen von der Größe der Instanz ab, bei einer mittelgroßen Instanz liegen sie im Bereich von ca. 25 bis 50 Euro pro Monat.

Produktivumgebung

In der Produktion ist es oft nicht tragbar, im Katastrophenfall einen ganzen Arbeitstag zu verlieren. Die klassische Art, dies zu bewältigen, basiert auf der Verwendung regelmäßiger RMAN-Gesamt-/Inkrement-Backups in Kombination mit einem Backup der aktuellen Archive Logs. Damit lässt sich ein RPO von ca. 30 Minuten recht einfach erreichen. Der Nachteil dieser Methode ist die RTO: Wenn die Datenbank sehr groß ist, wird das letzte komplette Backup wiederhergestellt, gefolgt von einem inkrementellen Backup, woraufhin anschließend alle aktuellen Archive Logs hinzugezogen werden. Bei Instanzen mit vielen

Terabyte an Daten kann der Restore-und-Recovery-Vorgang durchaus einen ganzen Tag in Anspruch nehmen.

Es gibt jedoch einen bequemen „Turbomodus“ für diesen Ansatz: Beginnend mit dem Oracle-Datenbank-Release 12c kann die Wiederherstellung auf der Basis von „3rd Party Storage Snapshots“ durchgeführt werden. Dies ermöglicht die Kombination eines Snapshot-Ansatzes – wie im vorigen Abschnitt gezeigt – mit der Anwendung der Archive Logs nur vom aktuellen Tag an.

Die wichtigsten Schritte, um dies einzurichten, sind:

- Einrichten eines File Storage in einer anderen Verfügbarkeitsdomain und Einhängen in die Datenbank-Tier.
- Einrichten dieses Einhängepunkts als zweites/optionales Ziel der Archive Logs.
- Sicherstellen, dass mindestens alle 30 Minuten ein Archive Log unter Verwendung des Parameters `archive_lag_target` erstellt wird.
- Optional: Definieren des obigen Einhängepunkts als sekundären Kontrolldatei-Speicherort.

Dieser Ansatz ergibt ein RPO von 30 Minuten mit einem RTO von 30 bis 90 Mi-

The screenshot shows the Oracle Cloud console interface for configuring a Backup Policy. The main heading is "EBS_Backup". Below the heading are buttons for "Edit Name", "Duplicate", "Add Tags", and "Delete". The "Backup Policy Information" section shows the OCID as "...flltq", the Compartment as "jmichler(root)/ebs_hol_compartment", the creation time as "Tue, Jan 12, 2021, 08:08:43 UTC", and the Cross Region Copy Target as "None". The "Schedules" section contains a table with two entries:

Schedule Type	Backup Type	Start Time	Retention Time
Daily	Incremental	01:00, Regional Time	A week
Weekly	Full	Sunday, 01:00, Regional Time	A week

Abbildung 1: Definition einer BLOCK Storage Backup Policy (Quelle: PROMATIS)

nuten (abhängig von der Menge der zu „apply-enden“ Archive Logs). Die Kosten entsprechen den zuvor beschriebenen 1-Tages-Speicher-Snapshot-Kosten zusätzlich des File Storage für die Archive Logs von einem Tag.

Hybrider Ansatz

Dem einen oder anderen Datenbankadministrator ist das beschriebene Verfahren eines Recovery auf Basis von „3rd Party Storage Snapshots“ zu „modern“ und noch zu wenig etabliert. So ist auch eine Kombination von klassischen RMAN-Sicherungen mit Storage-Snapshots möglich. Einer unserer Kunden hat mit folgender Vorgehensweise gute Erfahrungen erzielt:

- Nächtlich um 1 Uhr erfolgt ein Snapshot der Datenbank (Software und Datenfiles); jeweils sonntags „voll“ (5 TB) und ansonsten inkrementell (20-50 GB).
- Täglich um 21 Uhr erfolgt ein RMAN-Backup auf OCI Object Storage; samstags als „Level 0“-Voll-Backup (ca. 3 Stunden; 600GB) und ansonsten inkrementell (10 Minuten; 10-30 GB).
- Stündlich erfolgt ein Backup der Archive-Logs auf OCI Object Storage (wenige Sekunden).

Im Falle eines Desasters wird dann zunächst der neueste Storage-Snapshot in seinen alten Zustand gebracht und auf diesem (bereits komplett konfigurierten) das neueste Control-File-Autobackup aus dem OCI Object Storage wiederhergestellt. Dann kann mit folgendem Befehl ein Recovery und anschließendes „alter database open resetlogs“ erfolgen (wobei RMAN automatisch gegebenenfalls hilfreiche inkrementelle Backups anwendet und dann Archive-Logs wiederherstellt und anwendet):

```
recover database UNTIL SCN
6232192258870 SNAPSHOT TIME "TO_
DATE('20.01.2021 01:00:52', 'DD.
MM.YYYY HH24:MI:SS')";
```

Sollte das Recovery basierend auf dem Snapshot aus irgendeinem Grund scheitern, so steht auch ein RMAN Full/Incremental Backup zur Verfügung, das dann (mit deutlich erhöhter Laufzeit) repariert werden kann.

Das Verfahren kombiniert daher die niedrigen Kosten bei geringer RTO mit bewährten RMAN-Backups.

Apps-Tier für die Produktion

Bisher haben wir hauptsächlich die Datenbank-Ebene einer Oracle-E-Business-Suite-Instanz abgedeckt. Für die Apps-Tier könnte – neben den oben beschriebenen, in der Regel nächtlichen, Snapshots – einer der folgenden Ansätze verwendet werden:

- Platzieren des \$INST_TOP auf eine NFS-Freigabe auf dem File Storage oder
- Implementierung eines rsync-Backups von \$INST_TOP auf einem Block-Volumen zu einem entfernten Dateispeicherort oder
- Sicherstellen, dass sich keine wichtigen Daten auf dem Apps-Tier befinden: Häufig sind diese Daten nur temporär und müssen nicht mehr als einmal täglich gesichert werden.

Fazit

Wie oben gezeigt, bietet die Oracle-Cloud-Infrastruktur sehr leistungsstarke, benutzerfreundliche und dazu kostengünstige Möglichkeiten, alle Backup- und Wiederherstellungsanforderungen für Oracle-E-Business-Suite-Instanzen abzudecken.

Kurze RPO- und RTO-Zeiten können viel einfacher erreicht werden als mit herkömmlichen On-Premises-Rechenzentren.

Referenzen

- [1] Oracle High Availability Overview and Best Practices
<https://docs.oracle.com/en/database/oracle/oracle-database/19/haoww/index.html>
- [2] Oracle Cloud Infrastructure Documentation, Overview of Block Volume
<https://docs.oracle.com/en-us/iaas/Content/Block/Concepts/overview.htm>
- [3] Oracle Cloud Infrastructure Documentation, Overview of Object Storage
<https://docs.oracle.com/en-us/iaas/Content/Object/Concepts/objectstorageoverview.htm>
- [4] Oracle Cloud Infrastructure Documentation, Overview of File Storage
<https://docs.oracle.com/en-us/iaas/Content/File/Concepts/filestorageoverview.htm>
- [5] Oracle User's Guide, 1 Working with Oracle SQLcl
<https://docs.oracle.com/en/database/oracle/sql-developer-command-line/19.4/sqcug/working-sqlcl.html#GUID-1343FA2B-BDB4-4645-B4D4-CD7C3E200AC9>
- [6] Oracle Cloud Infrastructure Documentation, Overview of Block Volume Backups
<https://docs.oracle.com/en-us/iaas/Content/Block/Concepts/blockvolumebackups.htm>

Über den Autor

Johannes Michler ist als Senior Principal Consultant, Systemarchitekt und Projektleiter für die PROMATIS Gruppe mit Fokus auf serviceorientierte Architekturen (SOA), Web-Portale sowie BPMN- und BPEL-basierte Prozessautomatisierung tätig. Seit 2014 bekleidet er die Funktion „Senior Vice President – Head of Platforms & Development“; bei der Horus software GmbH ist er Mitglied im Management Board. Seit 2010 ist er für die DOAG als Referent und Autor mit zahlreichen wissenschaftlichen und anwendungsnahen Beiträgen aktiv. Daneben nimmt er regelmäßig als Referent zahlreiche Veranstaltungen der Oracle Community (IOUG & OATUG) wahr.



Johannes Michler
johannes.michler@promatis.de



Smart Data Analytics – Wie Knowledge-Graphen und Artificial Intelligence Daten-Silos in businessrelevantes Wissen umwandeln

Dr. Martin Zablocki, Trivadis

Unternehmen müssen rechtzeitig auf Marktbewegungen und neue Technologieentwicklungen reagieren können, wenn sie langfristig erfolgreich sein wollen. Täglich werden neue Produkte und Dienstleistungen auf globalen Märkten angeboten. In einer noch nie dagewesenen Dynamik fusionieren, kooperieren oder verschwinden Akteure von Märkten. Doch welche Möglichkeiten haben Unternehmen, ihre Prozesse und Entscheidungen an diese neuen Herausforderungen anzupassen? Wie kann der Bedarf an smarten 360-Grad-Sichten auf beispielsweise Produkte oder Märkte schnell und mit einer hohen qualitativen Güte gedeckt werden? In diesem Artikel zeigen wir, wie Trivadis mit ihrer **Smart-Data-Analytics**-Lösung **Knowledge-Graphen** und **Künstliche Intelligenz** kombiniert, um seine Kunden dabei zu unterstützen, die Entwicklung von Märkten, Technologien und Branchen auf der ganzen Welt zu analysieren, um somit komplexe Probleme zu lösen sowie Innovationen zu beschleunigen.

Knowledge-Graphen als Missing Link zwischen den Daten

Eine zentrale Herausforderung für moderne, datengetriebene Lösungen ist die Integration von Daten aus unterschiedlichen Quellen. Informationen über Unternehmen, ihr Produktportfolio, neue Veröffentlichungen und Markteintritte sind über Social-Media-Kanäle, Fachartikel und Publikationen geradezu in Echtzeit weltweit verfügbar. Doch echte Mehrwerte für das

Business entstehen nicht durch das reine Ansammeln dieser Informationen in sogenannten „Daten-Silos“. Erst durch eine einheitliche Veredelung der Daten, die die Semantik der Inhalte berücksichtigt, wird eine smarte Verschmelzung von heterogenen Datenquellen ermöglicht. Eine solche Wissensbasis kann dann zur Beantwortung komplexer Fragen herangezogen werden und als Entscheidungsgrundlage für das Business dienen.

Knowledge-Graphen versprechen, eine zentrale Schlüsselrolle in diesem Prozess

einzunehmen, indem sie ein für das Data-Management bisher fehlendes, jedoch essenzielles Puzzlestück nachliefern: ein intelligentes Wissensmanagement, das beliebige Informationen aus aller Welt organisiert und daraus Wissen entstehen lässt. Das hört sich zunächst etwas abstrakt an, beschreibt aber im Grunde einen hochgradig vernetzten Graphen, der uns in die Lage versetzt, wertvolle 360-Grad-Sichten auf zum Beispiel Kunden, Technologien, Produkte und sogar ganze Wertschöpfungsketten zu modellieren [1].

Im Vergleich zu klassischen Datenbanken ist es hierfür nicht erforderlich, Verbund-Operationen zwischen Tabellen zu definieren; in Knowledge-Graphen sind alle Informationen, die zusammengehören, bereits miteinander verknüpft. Mithilfe geeigneter Anfragesprachen werden zusammenhängende Informationen (z.B. *Gib mir alle Informationen zum Produkt X*) direkt abgerufen, ohne die Logik von Primärschlüsseln oder die Bedeutung von Spalten aus unterschiedlichen Speichersystemen interpretieren zu müssen.

Die Technologien hinter Knowledge-Graphen haben ihre Wurzeln in dem 2001 vorgestellten **Semantic Web** [2]. Doch richtig spannend und marktreif für den Einsatz in Enterprise-Lösungen wurden diese Ideen erst in den letzten Jahren, mit dem intelligenten Zusammenspiel aus Artificial Intelligence (AI) und Wissensmodellierung auf Grundlage von Knowledge-Graphen. So dienen Knowledge-Graphen auf der einen Seite den vielen AI-gestützten Lösungen als wertvolle Wissensquelle, um neue Modelle zu trainieren. Wichtiges Einsatzgebiete hierfür ist beispielsweise das Query Answering, wie wir es in unseren Alltagshilfen Alexa, Siri und Hey Google vorfinden. Für die automatische Beantwortung von Fragen, die in einer natürlichen Sprache gestellt werden, benötigt man ein breites, faktenbasiertes Wissen, wie es beispielsweise in öffentlichen Knowledge-Graphen wie DBPedia [3] und Wikidata [4] bereitgestellt wird. Auf der anderen Seite sind für den Aufbau und die Kuration von qualitativ hochwertigen Knowledge-Graphen AI-gestützte Verfahren nicht mehr wegzudenken.

Was ist ein Knowledge-Graph?

Ein Knowledge-Graph ist eine Wissensdatenbank, in der Informationen über beliebige Entitäten (Menschen, Orte, Dinge, Organisationen) und die Verbindungen zwischen ihnen („kennt“, „wurde gebo-

ren in“, „hat Tochterunternehmen“, „wurde gekauft von“) abgespeichert werden können. Informationen werden hierbei in Form von oftmals dreiwertigen Fakten modelliert, die auch ganz intuitiv als eine kurze Aussage von Menschen gelesen werden können (siehe *Abbildung 1*).

Jede der drei Entitäten kann dabei als ein Knoten in einem Graphen interpretiert werden, der weitere Eigenschaften wie zum Beispiel eine textuelle Beschreibung, die Gültigkeit, Herkunft der Information oder die Schreibweise in anderen Sprachen enthält. Essenziell hierbei ist, dass alle drei Elemente gleichwertig als Knoten behandelt werden können, obwohl ein Knoten in dem obigen Beispiel als Kante interpretiert werden kann. Dies ermöglicht es, in modularisierter Bauweise einen Knowledge-Graphen aus Fakten zu modellieren, die sich aufeinander beziehen, indem die gleichen Entitäten mehrfach wiederverwendet werden.

Für die Modellierung der einzelnen Entitäten ist es dabei essenziell, auf ein Kernwerkzeug von Knowledge-Graphen zurückzugreifen: weltweit gültige Identifikatoren, sogenannte „Internationalized Resource Identifier“ (IRIs). Folgt man den Best-Practice-Regeln [5] vom Erfinder des Semantic Web, so sollten diese IRIs sowohl von Menschen als auch von automatisierten Diensten konsumiert und interpretiert werden können und das unabhängig von der verwendeten Softwarelösung. Am obigen Beispiel illustriert, entspricht der IRI für Leonardo da Vinci dem Weblink <https://www.wikidata.org/wiki/Q762>. Ein Mensch kann diesen Link anklicken und bekommt eine Webseite, generiert aus dem Knowledge-Graphen von Wikidata, die alle Informationen zu Leonardo da Vinci anzeigt. Eine Applikation kann den publizierenden Server hinter dem IRI mithilfe der W3C-standardisierten Anfragesprache SPARQL [6] nach dem Geburtsort von Leonardo da Vinci fragen und bekommt als Antwort wieder einen IRI zurück, in diesem Fall: <https://www.wikidata.org/wiki/Q154184>.

From Data to Business

Der erfolgreiche Einsatz von Knowledge-Graphen verspricht, die Lücke zwischen Daten und geschäftsrelevanten Entscheidungsprozessen zu schließen. In der *Abbildung 2* wird dieser Prozess exemplarisch für einige Datenquellen (links) und mögliche Business Cases (rechts) illustriert, die Trivadis in den vergangenen Jahren für ihre Kunden umgesetzt hat. Die intelligente Verknüpfung (semantische Integration) von Entitäten aus heterogenen, oftmals auch unstrukturierten Datenquellen und deren Harmonisierung steht hierbei im Mittelpunkt der Wissens-Wertschöpfung.

Für den Zugriff auf das Wissen im Knowledge-Graphen möchten wir mehrere erprobte Konzepte vorstellen. Im einfachsten Fall greift eine Anwendung mittels einer Anfragesprache wie SPARQL auf die standardisierten Schnittstellen der darunterliegenden semantischen Graph-Datenbank zu. Hierbei gilt zu beachten, dass die geforderten sicherheitsrelevanten Funktionalitäten direkt von der Graph-Datenbank bereitgestellt werden sollten. In produktiven Szenarien wird immer häufiger auf das von Datawarehouses bekannte Konzept von Marts zurückgegriffen. Darunter versteht man eine neue Sicht auf das Wissen im Knowledge-Graphen. Diese Sicht kann zum Beispiel den Zugriff auf Informationen einschränken oder eine Kompaktierung und Transformation von Wissen vornehmen, die für die nachfolgende Verarbeitung oder Darstellung in einer Applikation optimiert ist.

Für die technische Umsetzung solcher Knowledge Marts stehen zwei unterschiedliche Ansätze bereit:

Bei dem ersten Ansatz handelt es sich um sogenannte „virtuelle Marts“. Diese können in Form von zusätzlichen Metadaten in sogenannten „Ontologien“ unabhängig von den Daten gepflegt werden. Sehr ausdrucksstarke Ontologien wie OWL [7] und RDFS [8] ermöglichen es, ein und dieselben Daten in unterschiedlichen Kontexten zu beleuchten. Der zweite Ansatz umfasst materialisierte Graph Marts. Darunter können sowohl neue Knowledge-Graphen verstanden werden, die nur einen Ausschnitt der Daten vorhalten, als auch grundverschiedene Speicherformate wie zum Beispiel eine



Abbildung 1: Fakten als elementare Bausteine für Knowledge-Graphen (© Trivadis)

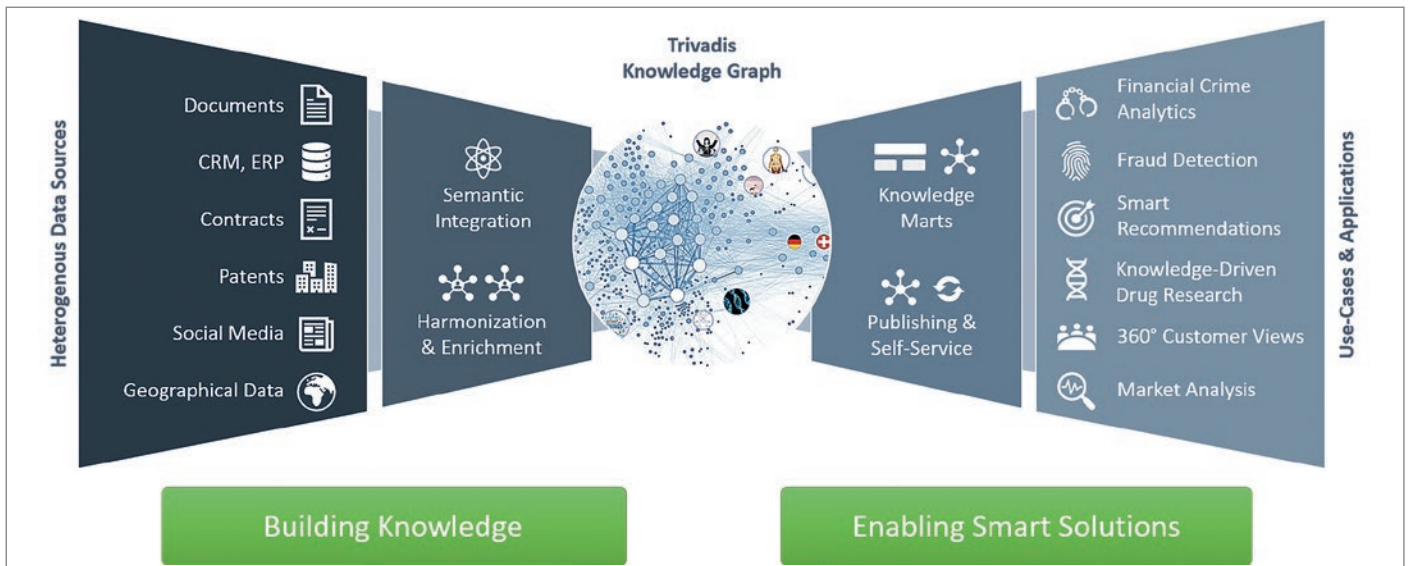


Abbildung 2: Knowledge-Graph fungiert als Datenherz zwischen Datenquellen und smarten Lösungen (© Trivadis)

Datenbank oder ein Property-Graph. Das Ziel ist es hier, das Wissen aus dem Knowledge-Graphen für neue Anwendungen und Anforderungen zu öffnen, die ganz andere Voraussetzungen an ihre Quelldaten festlegen. So wird man zum Beispiel einer Web-Applikation, die eine sehr niedrige Latenz benötigt, vielleicht eher eine voraggregierte NoSQL-Datenbank bereitstellen. Anwender von klassischen Reporting-Apps hingegen, die größtenteils auf tabellarischen Inhalten arbeiten, werden sehr dankbar für eine Tabelle mit Schemainformationen sein, die aus den Graph-Daten abgeleitet wird.

Nach dieser Einführung in die grundlegenden Konzepte von Knowledge-Graphen beschreiben wir im nächsten Abschnitt ein erfolgreich umgesetztes Kundenprojekt für Smart Data Analytics.

Smart Data Analytics für Markt- und Technologiemonitoring

Ein präzises Verständnis der Entwicklung von Märkten und Technologien im Kontext des eigenen Unternehmens bildet eine Schlüsselkomponente für den Erfolg. Gemeinsam mit der Technischen Hochschule Nürnberg und dem Fraunhofer SCS (Nürnberg) haben wir eine Anwendung entwickelt, die hierfür Knowledge-Graphen als Kernelement nutzt. *Abbildung 3* skizziert den wissenschöpfenden Prozess ab der Datenquelle bis hin zur Visualisierung des integrierten, angereicherten und harmonisierten Wissens im Knowledge-Graphen. Als Informationsgrundlage werden zunächst unstrukturierte Quellen

wie Nachrichtenfeeds sowie andere öffentlich verfügbare Textdokumente wie beispielsweise Publikationen, Patente und Artikel automatisch textuell erfasst (Schritt 1) und mit modernen, KI-gestützten Verfahren, zu denen in diesem Fall primär das sogenannte „Entity Discovery“ und „Entity Linking“ gehören (Schritt 2), in einem Knowledge-Graphen abgebildet (Schritt 3). Dieser zentrale Datenkern repräsentiert das Wissen aus den vereinnahmten unstrukturierten Texten in einer strukturierten und auf Fakten basierten Weise. Für das Markt- und Technologiemonitoring reicht es, zwischen Unternehmen, Technologien und Orten zu unterscheiden. Dieses Konzept lässt sich jedoch, je nach Anwendungsfall, auf beliebige weitere Informations-Typen erweitern.

Die Schritte 1 bis 3 werden fortlaufend durchgeführt. Das ermöglicht es, mit Änderungen über die Zeit hinweg zu erfassen, um daraus Indikatoren für Marktbebewegungen abzuleiten (Schritt 4).

Eine Feedbackschleife (vgl. Schritt 4 → 3) sorgt für die kontinuierliche Optimierung der Verfahren und Modelle, indem neue Erkenntnisse sowie menschliche Bewertungen als zusätzliche Datenbasis für die Erkennung von Technologien, Orten und Unternehmen herangezogen werden können.

Aufbauend auf diesem, sozusagen veredelten Wissen werden für das Technologie- und Marktmonitoring unterschiedliche Trendanalyse-Tools, Reports und Applikationen entwickelt (Schritt 5). Um das Reporting zu vereinfachen, wurden den Entwickler*innen Tabellen mit Schemainformationen bereitgestellt, sodass sie ihre gewohnten Skills für

die Erstellung von aussagekräftigen Reports einsetzen könnten.

Typische Fragen, bei denen unsere Lösung hilfreiche Informationen bereitstellen kann, sind:

- Was für neue aufstrebende Technologien gibt es in einem bestimmten Branchensektor, beispielsweise ein neues Material für die Herstellung von Energiespeichern?
- Welche aufsteigenden Unternehmen könnten sich, basierend auf ihrer Produkt-Portfolio-Entwicklung, als neuer Zulieferer, Akquisitionsziel oder potenzieller Kunde eignen?
- Welche Unternehmen stehen in Konkurrenz in bestimmten Märkten und in welche Richtung entwickeln sie sich?

Zu den spannenden Herausforderungen, die wir gemeinsam untersucht haben, zählen vor allem:

- Die Heterogenität der Datenquellen und Formate, die Geschwindigkeit, mit der Informationen ankommen und verarbeitet werden sollen, sowie die Granularität, mit der sie konsumiert werden können (Batch-, API-, Event-basiert).
- Es sollten vorwiegend Open-Source- oder lizenzfreie Softwarekomponenten verwendet werden.

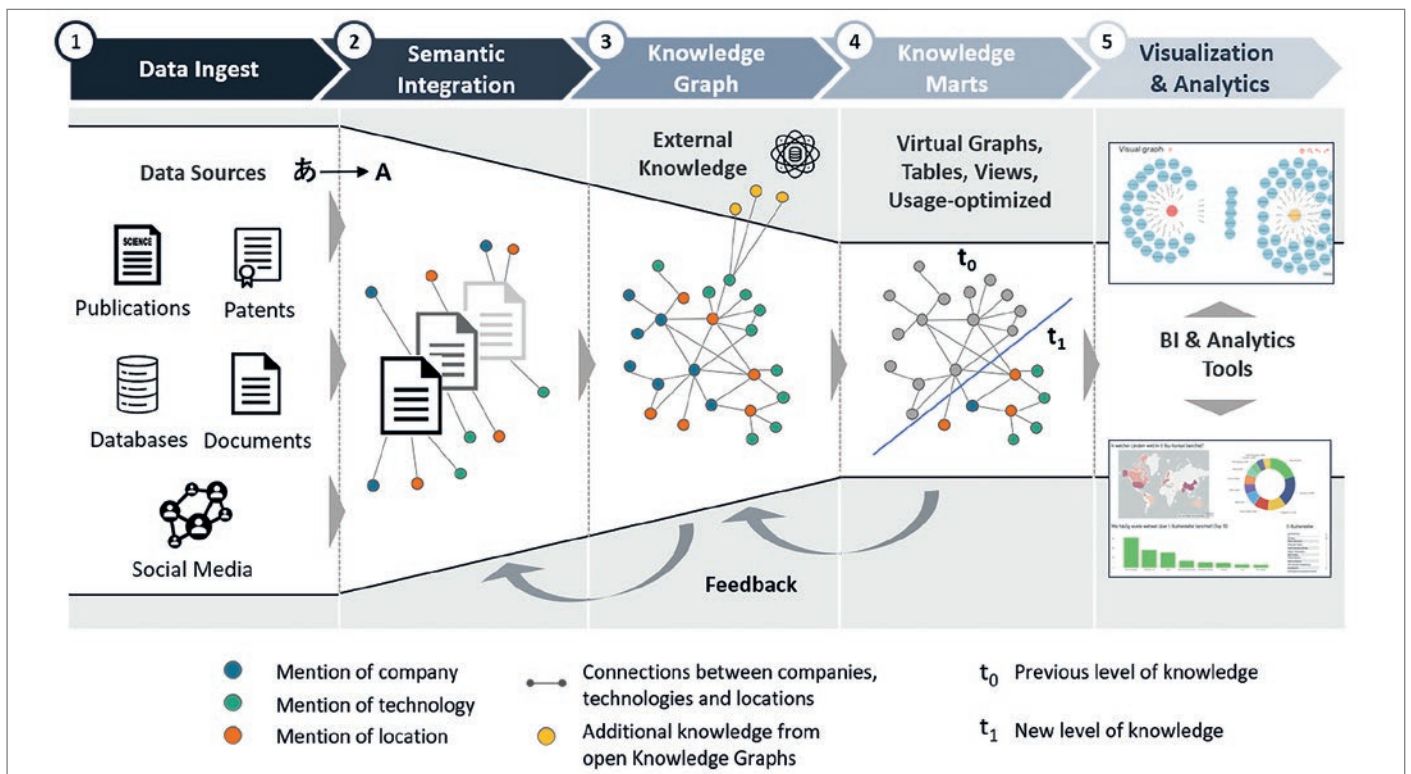


Abbildung 3: Smart Data Analytics für Markt- und Technologiemonitoring (© Trivadis)

• Öffentliche Knowledge-Graphen wie Wikidata.org oder DBpedia.org glänzen zwar mit einem beeindruckend breit aufgestellten Wissen und kommen daher oft in Pilotanwendungen zum Einsatz. Jedoch fehlt es diesen Graphen oftmals an Tiefe, wenn spezielles, domänenspezifisches Vokabular in den Texten verarbeitet werden soll. Als Lösung wurden hier insbesondere „Active-Learning-Verfahren“ betrachtet: Die aus Datenquellen extrahierten Informationen, bei denen sich die KI-gestützten Verfahren nicht sicher sind, werden Expert*innen zur Kontrolle vorgezeigt. Dieses wertvolle Expertenwissen wird jedoch nicht nur dazu verwendet, die angezeigten fehlerhaften Informationen zu korrigieren. Das Wissen fließt ebenfalls in das ursprüngliche KI-Verfahren zurück, so dass die Qualität aller zukünftigen Informationen verbessert wird.

Schlusswort

Mit der Menge und Vielfalt an Daten, die täglich produziert und publiziert wird, steigen Möglichkeiten wie auch Herausforderungen, geschäftsrelevante Prozesse und Entscheidungen zu unterstützen.

Smart Data Analytics ermöglichen es, mit der innovativen Kombination aus Knowledge-Graph-basierter Wissensmodellierung und AI-gestützter Informationsverarbeitung eine komplexe Datenlandschaft in nutzbares Wissen umzuwandeln. Unser vorgestelltes „Markt- und Technologiemonitoring“ gibt einen ersten Einblick in die vielfältigen Einsatzmöglichkeiten von Knowledge-Graphen als zentralen Puzzlestücken in der Datenveredelung.

Quellen

- [1] M. Zablocki, „Knowledge-Graphen: Das essenzielle Puzzlestück“, Trivadis Blog Interview, 12.11.2020, <https://news.trivadis.com/blog/interview-martin-zablocki>
- [2] T. Berners-Lee, J. Hendler, and O. Lassila, „The Semantic Web“ Scientific Am., May 2001, pp. 34–43.
- [3] DBpedia, 24.11.2020, <https://www.dbpedia.org/>
- [4] Wikidata, 24.11.2020, <https://www.wikidata.org>
- [5] T. Berners-Lee, „Linked Data“, 27.07.2006, <https://www.w3.org/DesignIssues/LinkedData.html>
- [6] W3C, „SPARQL 1.1 Query Language“, 21.03.2013, <https://www.w3.org/TR/sparql11-query/>
- [7] W3C, „Web Ontology Language (OWL)“, 11.12.2012, <https://www.w3.org/OWL/>
- [8] W3C, „RDF Schema 1.1“, 25.02.2014, <https://www.w3.org/TR/rdf-schema/>

Über den Autor

Dr. Martin Zablocki beschäftigt sich seit über 10 Jahren mit der Architektur von Datenplattformen und der Entwicklung von Knowledge-Graphen. Er ist promovierter Informatiker mit Schwerpunkt auf der intelligenten Nutzung von Daten. Bei der Trivadis ist er in wechselnden Rollen als Projektleiter, Architekt und Technologieexperte in der Automobil- und Life-Science-Branche unterwegs. In den vergangenen Jahren unterstützte er verschiedene öffentliche Einrichtungen in der Schweiz bei der smarten Veröffentlichung und Verknüpfung ihrer Daten mittels Knowledge-Graphen und Linked Open Data.



Dr. Martin Zablocki
Martin.Zablocki@trivadis.com



Ordnung statt Chaos – Metadatenmanagement im Data Lake

Fabian Hardt, Opitz Consulting Deutschland

Haben auch Sie einen Data Lake in Ihrem Unternehmen aufgebaut und bereits die ersten Quellsysteme angebunden? Durch moderne Frameworks, Streaming und Big Data Tools wird es uns immer leichter gemacht, beispielsweise auch externe Quellen wie Social Media anzubinden. Doch welcher Fachbereich kennt diese Daten eigentlich im Detail? Was verbirgt sich hinter den einzelnen Attributen in der angelieferten Datenstruktur? Schnell wird klar, dass Ihr Data Lake ohne geeignetes Metadatenmanagement im Chaos enden wird. Daher möchte ich Ihnen in diesem Artikel ein paar Grundlagen zu Metadaten sowie dem nötigen Metadatenmanagement im Data Lake näherbringen.

Was sind eigentlich Metadaten?

Metadaten sind „Daten über Daten“. Zugegeben, diese Definition ist erst einmal sehr wenig aussagekräftig. Ein einfaches Beispiel, das Sie jedoch alle kennen werden, ist der Besuch einer Bücherei. Auch in einer Bücherei erfolgt die gezielte Suche nach Büchern in der Regel über Metadaten, wie zum Beispiel den Autor, den Titel,

das Erscheinungsjahr oder aber die ISBN. In diesem Beispiel haben es die Betreiber der Bücherei mittels rudimentärem Metadatenmanagement also möglich gemacht, über eine kleine Datenbank gezielt in ihrem Bücherbestand zu suchen.

Bei Metadaten handelt es sich um strukturierte Daten, die wiederum Daten aller Couleur beschreiben können [1]. Es wird also mit der Erfassung von Metadaten eine Möglichkeit geschaffen, auch se-

mistrukturierte oder unstrukturierte Daten nach bestimmten Kriterien zu filtern, zu gruppieren oder zu suchen. Alles in allem kann man sagen, dass diese unerlässlich sind, um wenig strukturierte Daten im Data Lake wieder auffindbar zu machen. Sie sollten stets so erfasst werden, dass sie maschinell les- und auswertbar sind. Auf diese Weise können sie zu einem späteren Zeitpunkt im Prozess zum Generieren von Code genutzt werden oder aber

zur automatisierten Berechtigungssteuerung durch generierte Zugriffs-Policies.

Verschiedene Arten von Metadaten

In der Praxis begegnen einem verschiedene Arten von Metadaten. Zum einen **technische Metadaten**, die beim Datenbewirtschaftungsprozess einfach und automatisiert erhoben werden können. Dies sind beispielsweise Dateinamen, Dateigrößen, Dateitypen, aber auch das Ladedatum oder die Bezeichnung des Quellsystems [2]. Diese Arten von Metadaten sind sehr statisch, aber auch qualitativ sehr hochwertig, da sie automatisiert erfasst werden und es sich hierbei um Basisattribute handelt, die sich in der Regel nicht häufig ändern. Technische Metadaten werden überwiegend im Zusammenhang mit externen Prozessen gesammelt, also beim Data-Ingest-Prozess. Dieses automatisierte Sammeln von Metadaten wird auch „Metadata Harvesting“ genannt und von immer mehr Toolherstellern unter diesem Begriff beworben.

Eine weitere Art von Metadaten sind die fachlichen Metadaten. Diese sind mit die wichtigsten, aber auch die am schwersten zu erfassenden. Es besteht ein direkter Bezug zu den Nutzdaten im Data Lake. Beispielsweise könnte es sich hierbei um eine Schemadefinition handeln, die Attributen aus Dateien weitergehende Informationen über den Inhalt hinzufügen. Konkret könnten dies Informationen zu Inhalten von Tabellen beziehungsweise Spalten sein oder auch Schemainformationen zu XML- oder JSON-Dateien. Bei fachlichen Metadaten kann es sich ebenso um benötigte Governance-Informationen wie Berechtigungen oder Aufbewahrungsfristen handeln, die für eine automatisierbare Berechtigungssteuerung oder Datenlöschung benötigt werden. Diese fachlichen Metadaten können in der Regel maximal teilautomatisiert erfasst werden, da sie üblicherweise von der zuständigen Fachabteilung beigesteuert werden müssen. Je qualitativ hochwertiger die fachlichen Metadaten gepflegt sind, desto bessere Ergebnisse lassen sich zum Beispiel in Ad-hoc-Reportings erzielen, da Poweruser auch fachbereichsübergreifende Daten sinnvoll miteinander verknüpfen und auswerten können.

Bei der dritten Art handelt es sich um operative Metadaten, diese beziehen

sich in der Regel auf interne Prozesse. Dies sind technische Metadaten, teilweise auch solche, die innerhalb der technischen Systemkomponenten ohne weiteres Zutun erfasst werden. Dies sind zum Beispiel Metadaten-Repositories von ETL Tools, Nutzerzugriffsstatistiken oder Audit-Informationen von Security-Komponenten.

Die vierte und letzte Metadatenart sind **API-Metadaten**. Hierbei handelt es sich um nach außen gerichtete Metadaten, die ein API des Data Lake näher beschreiben. Dies erfolgt in der Regel in Form einer API-Beschreibung, als Swagger/OpenAPI, RAML oder API-Blueprint-Datei. Auf diese Weise werden die Methoden und Parameter des API für einen Entwickler oder Nutzer dieser Data-Lake-Programmierschnittstelle beschrieben. Damit in Zusammenhang stehen dann weitere API-Metadaten, wie Zugriffsstatistiken oder die Herkünfte der API-Zugriffe, die dann eher nach innen gerichtete Metadaten darstellen.

Metadaten-Tools

Zur Umsetzung von gutem Metadatenmanagement im Data Lake empfiehlt sich der Einsatz von geeigneten Tools. Hierbei unterscheidet man drei Arten von Tools:

Zum einen das **Data Dictionary**, einen Metadatenkatalog für technische Metadaten. Data Dictionaries sind fest integrierte Bestandteile von Software-Systemen, ein prominentes Beispiel ist das Oracle Data Dictionary, das Metainformationen zur gesamten Datenbank katalogisiert. Dies können Tabellenstrukturen, Datentypen der Spalten, aber auch Berechtigungsinformationen sein. Diese Informationen werden gerne von Entwicklern genutzt, um generische Frameworks zur Datenbewirtschaftung zu entwickeln. Data Dictionaries sind also fest vorgegebene Systeme und können nicht ausgetauscht werden. Sie können lediglich zur Erhöhung der Datenqualität im Data Lake beitragen.

Etwas weiter gefasst, aber immer noch sehr spezialisiert, sind die **Data-Lake-spezifischen Tools**. Diese sind fest in die Data-Lake-Plattform integrierte Bestandteile und können allgemein auch als „Data-Lake-Management-Plattform“ bezeichnet werden. Es handelt sich hierbei meist

nicht um ein einzelnes Tool, sondern um mehrere spezialisierte. Dies könnten API-Metadaten sein, die in Form eines „Developer-Portals“ zur Verfügung gestellt werden. Des Weiteren kann es sich um Self-Service-Data-Preparation-Werkzeuge handeln, die anhand fachlicher Metadaten Quellcode zur Datenaufbereitung generieren können. Da dieser Code sehr plattformspezifisch sein muss, sind diese Tools oft eng verzahnt mit der eingesetzten Data-Lake-Umgebung. Ein letztes und prominentes Beispiel ist die Autorisierung anhand von Metadaten. Hierbei können Berechtigungen auf fachlichen Entitäten vergeben werden und die Tools „übersetzen“ diese Berechtigungsinformationen in technische Access-Control-Informationen, die sehr technisch geprägt sind. Hierbei ist eine enge Verzahnung zu einem Data Catalog sehr sinnvoll.

Kommen wir nun zum Platzhirschen auf dem Gebiet Metadatenmanagement, dem **Data Catalog**. Dieses Tool ist spezialisiert auf die nötigen Aufgaben rund ums Metadatenmanagement. Hierzu zählen die (teil-)automatisierte Erfassung, die Speicherung sowie die Bereitstellung der Metadaten über diverse Komponenten des Data Lake hinweg. Außerdem kann ein Data Catalog sowohl technische und fachliche als auch operative Metadaten verwalten. Durch diese komponentenübergreifende Sicht auf sämtliche Metadaten ermöglicht ein Data Catalog ein durchgehendes Data Lineage, das eine transparente Verfolgung der Datenströme im Data Lake erlaubt sowie Beziehungen der Daten untereinander visualisieren kann. Dieses Tool bringt idealerweise eine hohe Anzahl von APIs/Connectoren mit, um ein Metadata Harvesting aus diversen Quellen auf einfache Art zu ermöglichen. Doch ein Data Catalog unterstützt nicht nur auf diese technische Art und Weise, sondern auch in organisatorischer Hinsicht. Durch die Bereitstellung eines unternehmensweiten Glossars (Business Glossary) hilft er, ein einheitliches Verständnis der Mitarbeiter zu domänenspezifischen Fachbegriffen zu erlangen und zu schärfen. Doch die wahrscheinlich größten Vorteile lassen sich durch ein unternehmensweites Tagging der Daten erzielen. Hierbei werden den Daten direkt beim Data Ingest fachspezifische Labels zugeordnet. Dieser Schritt entspricht im Wesentlichen der Inven-

tarisierung einer Bibliothek, um das Eingangsbeispiel nochmals aufzugreifen. Auf diese Weise ist es mithilfe solcher Labels möglich, etwa Kundendaten getrennt von Produktdaten quellenübergreifend wiederzufinden und gezielt für Auswertungen verfügbar zu machen. Nicht zuletzt sollte ein Data Catalog jedoch auch kollaborative Funktionen anbieten, zum Beispiel ein Rating, um die Datenqualität zu bewerten, ein Wiki, um weiterführende Informationen zu einer Datenquelle zu erfassen, oder eine Speichermöglichkeit für Suchabfragen, sodass definierte Filter gespeichert und mit Kollegen ausgetauscht werden können. Einige Hersteller solcher Tools bieten mithilfe gängiger Machine-Learning-Verfahren eine Analyse des Nutzerverhaltens an, sodass nutzerbezogen entsprechende Suchergebnisse priorisiert dargestellt werden können. Außerdem werden zum Nutzerverhalten passende Suchabfragen vorgeschlagen. Zu guter Letzt hilft ein Data Catalog aber

auch bei der Erfüllung gesetzlicher Auflagen wie der DSGVO. Er kann helfen, Aufbewahrungs- und Löschrufen präzise einzuhalten, und bietet entsprechende Logging- und Auditingfeatures, um dies entsprechend nachvollziehen zu können.

Praxisbeispiel Data Ingest

Nun etwas konkreter! In folgendem Abschnitt möchte ich darauf eingehen, wie eine automatisierte Metadatenerfassung direkt beim Data Ingest aussehen kann. In *Abbildung 1* ist der Data-Ingest-Prozess auf zwei verschiedenen Abteilungen dargestellt – Finance und Marketing.

Bei den Daten der Finance-Abteilung handelt es sich um klassische Batch-Daten, die in regelmäßigen Abständen per Data Flow in den Data Lake überführt werden. Schematisch dargestellt sind dies relationale Daten aus einem Data Warehouse sowie klassischerweise Ta-

bellenkalkulationsdateien. In der Marketing-Abteilung handelt es sich um selbstberechnete Kennzahlen (hier schematisch NPS – Net Promoter Score) und Social-Media-Daten als Daten-Stream, etwa aus dem Kurznachrichtendienst Twitter. Kommen wir zur eigentlichen Prozessierung. In diesem Schritt sollen zusätzlich Metadaten zu den Eingangsdaten erfasst werden. Hier ist wieder eine Unterscheidung zu machen; rein technische Metadaten lassen sich direkt aus den Daten beziehungsweise der Datenquelle extrahieren. Dies sind zum Beispiel Dateigrößen, Ablageorte, Dateitypen oder auch Tabellenstrukturen, die aus dem Quell Data Dictionary entnommen werden können. Etwas schwieriger ist es hingegen mit fachlichen Metadaten. Welche Entitäten lassen sich einem Kunden zuordnen, welche einem Produkt? Um solche Metadaten korrekt im Data Catalog zuordnen zu können, sind Begleitdateien, beispielsweise im JSON-Format, ein probates Mittel.

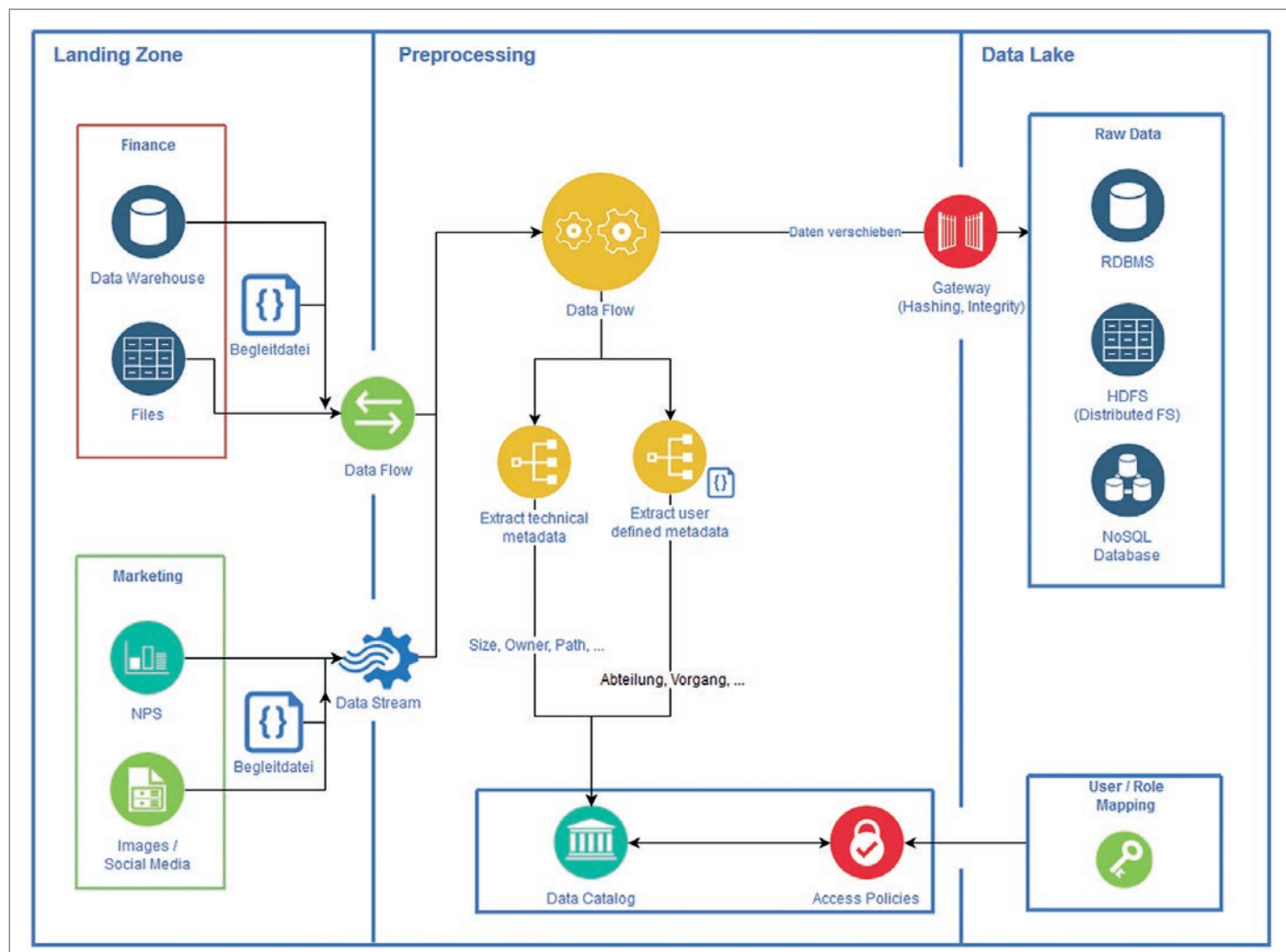


Abbildung 1: Data-Ingest-Prozess (Quelle: Fabian Hardt)

Filename	Filesize	Data owner	Data source	Content reference
Test.jpg	3MB	Moni Tor	Marketing	RK: 20190524M<<HASH>>
Video.mkv	50213MB	Max Smith	Marketing	hdfs://marketing/Video.mkv
Invoice.docx	12MB	Peter Fox	Accounting	hdfs://accounting/Invoice.docx
Invoice.pdf	5MB	Peter Fox	Accounting	RK: 20190523A<<HASH>>

Abbildung 2: Beispiel für eine Metadatenstruktur (Quelle: Fabian Hardt)

Abbildung 2 zeigt eine beispielhafte Metadatenstruktur mit technischen Metadaten (Filename, Filesize, Content reference) sowie fachlichen Metadaten, wie Data owner und Data source. Hier werden konkret ein Datenverantwortlicher sowie die zuständige Abteilung des Quellsystems benannt. In echten Szenarien sollten die fachlichen Metadaten deutlich konkreter sein und beispielsweise bis zur konkreten Entität wie Produkt oder Kunde referenzieren, sodass eine gezielte Suche quellübergreifend ermöglicht werden kann. Dieser Vorgang des Zuordnens wird Tagging genannt und sollte immer mit den Begrifflichkeiten des Business Glossary übereinstimmen.

Diese Begleitfiles, in denen die fachlichen Metadaten geliefert werden, sind vorab mit dem Fachbereich der entsprechenden Datenquelle abzustimmen und als Datenschnittstelle zu definieren. Erst wenn auch diese Begleitfiles geliefert wurden, beginnt eine automatische Verarbeitung der Quelldaten im Data Lake. Dieser Trigger sorgt für eine Verbesserung der Datenqualität, da zumindest die definierten Pflichtattribute gefüllt sein müssen, damit die Daten eingeleitet werden. Als Abschluss des Data-Ingest-Prozesses werden die „Ziel-Metadaten“ erfasst, also technische Zeitstempel zum Ablagezeitpunkt, Ablageorte sowie Aufbewahrungs- und Berechtigungsmetadaten. Diese werden mit den Eingangsmetadaten verknüpft und ermöglichen somit eine saubere Visualisierung des Datenflusses im Data-Lineage-Graphen.

Sind diese Metadaten nun alle erfasst, kann der Data Catalog als zentrale Stelle für diverse weitere automatisierte Prozesse genutzt werden. Dies kann ein Housekeeping-Prozess sein, der Nutzdaten zu gesetzlich vorgeschriebenen Zeitpunkten löscht, beispielsweise nach Ablauf der ma-

ximalen Aufbewahrungsdauer. Abbildung 3 zeigt einen Flow, der in regelmäßigen Abständen den Ablauf von Löschrufen prüft, indem er die maximale Aufbewahrungsdauer mit dem Einleitungsdatum der Daten prüft. Daten wie die maximale Aufbewahrungsdauer oder in speziellen Fällen auch Löschoratorien können durch die Data-Governance-Abteilung direkt in den Data Catalog eingepflegt werden.

Dank des Data Catalog liegen nun alle benötigten Metadaten an einer zentralen Stelle vor, sodass alle Voraussetzungen für eine Automatisierung dieses Löschrufenprozesses vorhanden sind und die eigentliche Datenlöschung vollautomatisiert durchgeführt werden kann.

Ein ähnlich nützlicher und weit verbreiteter Anwendungsfall ist die automatisierte Berechtigungssteuerung im Data Lake. Diese stellt viele Unternehmen vor große Herausforderungen, da die Daten oft in unterschiedlichen Datenspeichersystemen abgelegt sind, die sehr unterschiedliche Berechtigungsmechanismen benötigen. Doch mit einem Data Catalog steht uns nun eine zentrale Stelle zur Verfügung, in der Berechtigungs-Policies auf abstrakter Ebene und auf fachlichen Entitäten basierend erfasst werden können. Dies bedeutet, dass alle Informationen vorliegen, um die technischen Berechtigungen, die in den verschiedenen Datenspeichern benötigt werden, gene-

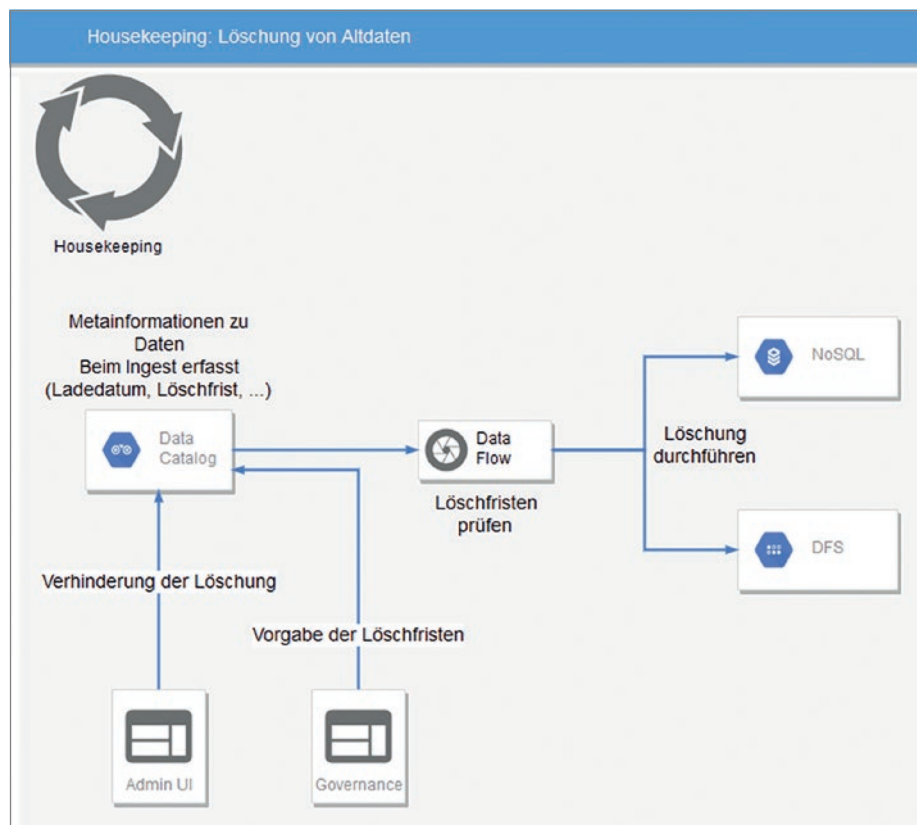


Abbildung 3: Housekeeping-Prozess (Quelle: Fabian Hardt)

rieren zu können. Beim Zugriff auf Daten über Data-Lake-APIs kann zudem auf die Berechtigungsmetadaten des Data Catalog zugegriffen werden, um eine fachliche Autorisierung zu ermöglichen. Alles in allem ist ein Data Catalog also zwingend erforderlich, um komplexere Berechtigungsstrukturen in einem Enterprise Data Lake abzubilden.

Fazit

Zusammenfassend lässt sich sagen, dass Metadaten und in der Folge auch ein Data Catalog in einem Enterprise Data Lake unerlässlich sind. Zum einen lassen sich die Entstehung von Chaos im Data Lake eindämmen sowie die Datenqualität massiv steigern. Der Einsatz eines Data Catalog eröffnet neues Automatisierungspotenzial und kann zur korrekten Einhaltung von Datenschutzanforderungen beitragen. Mit Funktionen wie dem Business Glossary können die Data-Governance-Abteilung un-

terstützt und somit einheitliche Begrifflichkeiten im Unternehmen etabliert werden.

Außerdem unterstützen Data Catalogs bei der automatisierten Erfassung von Metadaten und bieten eine geeignete Kollaborationsplattform für Daten-Analysten, die gezielt Suchfilter speichern und untereinander austauschen können. Durch ein Mindestset an Metadaten, das unternehmensweit definiert werden sollte, kann ein Data Catalog ein echtes Quality Gate direkt beim Data Ingest abbilden und verringert somit die Entstehung eines Datensumpfes.

Quellen

- [1] Gunnar Auth (2004): Metadaten – Grundlagen und Bedeutung im Data Warehousing. In: Gunnar Auth: Prozessorientierte Organisation des Metadatenmanagements für Data-Warehouse-Systeme. BoD, Norderstedt
- [2] <http://www.business-intelligence24.com/data-management/business-intelligence-metadatenmanagement/business-intelligence-metadatenmanagement>

Über den Autor

Fabian Hardt arbeitet als Senior Consultant bei der OPITZ CONSULTING Deutschland GmbH. Er hat mehrjährige Projekterfahrung in BI- und Big-Data-Projekten und beschäftigt sich mit modernen Architekturen für die gestiegenen Anforderungen im Zeitalter der Digitalisierung.



Fabian Hardt
Fabian.Hardt@opitz-consulting.com

ORAWORLD feiert fünften Geburtstag

DOAG Online

Vor nunmehr einem halben Jahrzehnt, im Mai 2016, erschien die erste Ausgabe des ORAWORLD-Magazins.

Zuvor gab es den EOUC-Newsletter, aus dem schließlich vor fünf Jahren die ORAWORLD hervorging. Das eMagazine liefert seinen Lesern seither regelmäßig spannende Storys und wissenswerte Artikel rund um Themen aus dem Oracle-Kosmos.

Die bis heute bestehenden Rubriken "Tech & Nerds" und "Users & Groups" wurden bereits bei der Premiere eingeführt. Mit der dritten Ausgabe kam die Rubrik "Work & Life" hinzu, in der über den technischen Tellerand hinausgeblickt wird und persönliche Geschichten Raum bekommen – wie in der neuen Interviewreihe "What's Your Passion?", in der bekannte Community-Vertreter über ihre Hobbys und Leidenschaften sprechen.

Anlässlich des Geburtstags haben wir noch einmal in der ersten ORAWORLD



gestöbert: Die Premierenausgabe widmete sich dem Thema "Women in IT" und warf dabei – auch anhand einer eigens durchgeführten Umfrage – einen kritischen Blick auf den geringen Frauenanteil innerhalb der Oracle User Groups. Zudem behandelten die Autoren den Status quo von Oracle Forms und gingen der Frage nach, warum sich das Twitter-Hashtag #letsreckthistogether innerhalb der APEX-Community immer größerer Beliebtheit erfreute. Darüber hinaus stellte sich mit der ITOUG die damals noch sehr junge, im Jahr 2014 gegründete italienische User Group vor.

Seither sind 24 Ausgaben des Magazins erschienen, die allesamt zum kostenfreien Download auf der ORAWORLD-Website (www.oraworld.org) zur Verfügung stehen. Die nächste Ausgabe 25 erscheint am 21. Juni.



SQL Macros – Game Changing Feature für SQL-Entwickler?

Andrej Pashchenko, Trivadis Germany

Mit der Version 20c hat Oracle erstmalig SQL Macros vorgestellt – eine grundlegend neue Möglichkeit für die Kapselung der Geschäftslogik. Um generische benutzerdefinierte Erweiterungen zu entwickeln, griff der Entwickler bis jetzt meist auf die eine oder andere Variante der PL/SQL-Funktionen zurück, inklusive der zuletzt in 18c eingeführten polymorphen Tabellenfunktionen, aber auch inklusive bekannter Nachteile der Laufzeitinteraktion zwischen SQL und PL/SQL. Dieser Artikel gibt Einblicke in die völlig neue Art des Zusammenspiels von SQL und PL/SQL in der Datenbank.

Die Rolle von SQL als lingua franca der Datenbanken ist unumstritten. Modernes SQL wird immer mächtiger. Das bedeutet allerdings auch längere und komplexere SQL-Statements. Der Wunsch, diese durch Auslagern der (hoffentlich wiederverwendbaren) Abschnitte zu ver-

einfachen, ist durchaus nachvollziehbar. Macht man das über PL/SQL-Funktionen, sollte man sich jedoch der Auswirkungen bewusst sein. Denn diese führen zur Laufzeit zum bekannten Context-Switch, dessen Einfluss zwar durch das neue Pragma UDF ab 12.1 reduziert, aber nicht kom-

plett weggeräumt werden konnte. Rufen solche PL/SQL-Funktionen auch selbst noch SQL auf, ist der Einfluss noch größer. Dazu kommen in diesem Fall auch noch Konsistenz-Probleme. Denn jede weitere SQL-Abfrage, die nun „rekursiv“ ausgeführt wird, wird den neueren Stand

in der Datenbank sehen können (zumindest bei der standardmäßigen Transaktionsisoliationsstufe READ COMMITTED).

Auch für den Query-Optimizer ist diese Vorgehensweise nachteilig: Beim Optimieren der Hauptabfrage ist er nicht in der Lage, die weiteren – über aufgerufene PL/SQL-Funktionen verteilten – SQL-Abfragen zu sehen und bei der Ausarbeitung des Ausführungsplans in Betracht zu ziehen. Aus Performance-Perspektive wäre ein SQL-Statement ohne PL/SQL zur Laufzeit oft die bessere Alternative.

SQL Macros lösen diesen Interessenkonflikt auf die elegante Art und Weise: Sie lassen in Richtung Entwickler das SQL einfach und gut lesbar aussehen, verstecken aber vom Optimizer keine wichtigen Details.

Funktionsweise der SQL Macros

SQL Macros sind spezielle PL/SQL-Funktionen, die SQL-Fragmente als Text zurückliefern. Diese Funktionen darf man im SQL-Statement überall dort verwenden, wo auch „normale“ Funktionsaufrufe erlaubt sind. Der wesentliche Unterschied ist, dass diese Funktionen nicht zur Ausführungszeit aufgerufen werden, sondern schon während des Parse-Vorgangs vom SQL-Statement. Die von SQL Macros zurückgelieferten SQL-Fragmente werden in die Hauptabfrage integriert und dann wird diese „neue“ Abfrage von der Datenbank geparkt. Die Datenbank sieht somit den kompletten Text der SQL-Anweisung in der ganzen Komplexität, wobei Teile davon nach benutzerdefinierter Geschäftslogik von den SQL-Macro-Funktionen zusammengestellt werden.

Schauen wir uns ein einfaches Beispiel an: Für eine Auswertung berechnen wir wie im *Listing 1* das Verhältnis von Provision zu Gehalt (Spalten COMM zu SAL). Natürlich wird der produktiv eingesetzte Code weitere Aspekte berücksichtigen, wie zum Beispiel eventuelles Teilen durch Null, Umgang mit NULL-Werten etc. Aber wir wollen es einfach halten. Gleichzeitig wollen wir nach diesem berechneten Wert filtern und nur Datensätze selektieren, bei denen die Provision mindestens 90% des Gehalts beträgt. Dazu bräuchten wir entweder eine weitere Abfrage, in der wir auf die neu berechnete Spalte RATIO zugreifen können,

oder wir wiederholen die Berechnung in der WHERE-Klausel wie im *Listing 1*.

Wenn dieser Ausdruck noch an einigen weiteren Stellen im Applikationscode verwendet wird, wäre das beispielsweise ein Kandidat, um ihn in ein SQL Macro auszulagern. Dazu legen wir eine Funktion `get_ratio` (siehe *Listing 2*) an. Die SQL-Macro-Funktionen dürfen als Rückgabe-Datentyp nur Text zurückliefern, also CHAR, VARCHAR2 oder CLOB. Das weitere Unterscheidungsmerkmal, das eine Funktion zu einem SQL Macro macht, ist die neue Klausel `SQL_MACRO`, hinter der in Klammern noch der Typ des Makros folgt. Es

```
SELECT empno, ename, sal, comm
,      round(comm/sal,2) AS ratio
FROM   emp
WHERE  round(comm/sal,2) > 0.9 ;
```

Listing 1: Ursprüngliche Abfrage

```
CREATE OR REPLACE FUNCTION get_ratio RETURN VARCHAR2
SQL_MACRO(SCALAR) IS
BEGIN
  RETURN 'round(comm/sal,2)';
END;
/
SQL> SELECT empno, ename, sal, comm
2      ,      get_ratio ratio
3 FROM   emp
4 WHERE  get_ratio > 1 ;
```

EMPNO	ENAME	SAL	COMM	RATIO
7654	MARTIN	1250	1400	1,12

Listing 2: Skalares SQL Macro und die Verwendung im SQL

```
SQL_ID 31b4113p8xf3y, child number 1
-----
SELECT empno, ename, sal, comm, get_ratio ratio
FROM   emp
WHERE  get_ratio > 0.9
```

Id	Operation	Name
0	SELECT STATEMENT	
* 1	TABLE ACCESS STORAGE FULL	EMP

Predicate Information (identified by operation id):

```
2 - storage(ROUND("COMM"/"SAL",2)>.9)
   filter(ROUND("COMM"/"SAL",2)>.9)
```

Listing 3: Der SQL-Ausdruck als Prädikat sichtbar

gibt zwei Typen: skalare Makros und Tabellenmakros.

Skalare Makros

Skalare Makros liefern SQL-Ausdrücke für einen skalaren Wert zurück und dürfen somit nur in SELECT-, WHERE-, ORDER-BY-, HAVING- und GROUP-BY-Klauseln verwendet werden. Diesen Typ werden wir für unser Beispiel verwenden. Die Funktion selbst enthält keine Logik, sie gibt nur den gewünschten SQL-Ausdruck zurück.

Das angelegte SQL Macro kann dann in der Abfrage anstelle des ursprünglichen SQL-Ausdrucks eingesetzt werden. Die Vereinfachung für den Entwickler ist in diesem konkreten (trivialen) Beispiel marginal. Wie sieht es für die Datenbank aus? *Listing 3* zeigt einen Auszug aus dem Ausführungsplan und im Abschnitt mit Prädikatsinformationen kann man nun unseren SQL-Ausdruck finden. Mit einer

```

CREATE OR REPLACE FUNCTION get_ratio(p_numerator IN NUMBER
                                     , p_denominator IN NUMBER)
RETURN VARCHAR2 SQL_MACRO(SCALAR) IS
BEGIN
    RETURN 'round(p_numerator/p_denominator,2)';
END;
/

SQL> SELECT empno, ename, sal, comm
       2 ,      get_ratio(comm, sal) ratio1
       3 ,      get_ratio(:p_comm, sal) ratio2
       4 ,      get_ratio(1000, sal) ratio3
       5 FROM    emp
       6 WHERE   get_ratio(comm, sal) > 0.9 ;

   EMPNO ENAME          SAL   COMM  RATIO1  RATIO2  RATIO3
-----
   7654 MARTIN          1250   1400    1,12    1,6     0,8

select * from table(dbms_xplan.display_cursor(sql_id => '81nb4fkbnpsd', cursor_child_no=>0, format=>'basic
+projection'));

Plan hash value: 3956160932

-----
| Id | Operation                                | Name                                |
-----
|  0 | SELECT STATEMENT                          |                                     |
|  1 |  RESULT CACHE                              | 9y33d23fkzwx93b5j20rn6cpv0 |
|  2 |    TABLE ACCESS STORAGE FULL             | EMP                                  |
-----

Column Projection Information (identified by operation id):
-----
 1 - "EMPNO"[NUMBER,22], "ENAME"[VARCHAR2,10], "SAL"[NUMBER,22],
    "COMM"[NUMBER,22], ROUND("COMM"/"SAL",2)[22],
    ROUND(:P_COMM/"SAL",2)[22], ROUND(1000/"SAL",2)[22]
 2 - "EMPNO"[NUMBER,22], "ENAME"[VARCHAR2,10], "SAL"[NUMBER,22],
    "COMM"[NUMBER,22]

```

Listing 4: Parameterübergabe

herkömmlichen PL/SQL-Funktion bleibt er von der Datenbank verborgen.

Parameter

Was ist, wenn wir die Berechnung mit anderen Spalten oder gar Variablen/Konstanten durchführen wollten? Wie können dafür auch das SQL Macro verwenden. Wie herkömmliche PL/SQL-Funktionen können SQL Macros auch Parameter akzeptieren. Listing 4 zeigt zunächst die Implementierung der Funktion mit Parametern.

Sehr wichtig ist, wie die Parameter in der Funktion verwendet werden. Auf den ersten Blick wirken die Parameternamen als Teil des Rückgabestrings verwirrend. Kein Konkatenieren, keine Bindevariablen! Hier kommt eine Spezialbehandlung

```

CREATE VIEW v_emp_sales AS
SELECT *
FROM emp
WHERE deptno = (SELECT deptno
                FROM dept
                WHERE dname = 'SALES');

SQL> SELECT empno, ename
       2 FROM v_emp_sales;

   EMPNO ENAME
-----
   7499 ALLEN
   7521 WARD
   7654 MARTIN
   7698 BLAKE
   7844 TURNER
   7900 JAMES

6 rows selected.

```

Listing 5: View für die Mitarbeiter einer Abteilung

```

CREATE OR REPLACE FUNCTION f_emp_sales (p_dname IN VARCHAR2)
RETURN VARCHAR2 SQL_MACRO(TABLE) IS
BEGIN
    RETURN 'SELECT *
          FROM   scott.emp
          WHERE  deptno = (SELECT deptno
                          FROM   scott.dept
                          WHERE  dname = p_dname)';
END;
/

```

Function F_EMP_SALES compiled

```

SQL>
SQL> SELECT empno, ename
       2 FROM   f_emp_sales ('ACCOUNTING');

```

EMPNO	ENAME
7782	CLARK
7839	KING
7934	MILLER

Listing 6: SQL Macro für die Mitarbeiter einer Abteilung

der SQL Macros ins Spiel. Die Parameternamen funktionieren im Rückgabestring wie Platzhalter, die bei der Rückgabe des Strings aus der Funktion durch die Parameterwerte ersetzt werden.

Das kann man mithilfe des Ausführungsplans im *Listing 4* gut erkennen. In der SELECT-Liste wird die Funktion in drei Varianten aufgerufen: mit Spaltennamen als Parameter, mit Bindevariable und mit einer Konstante. Die Ausgabe des Ausführungsplans ist wieder etwas abgekürzt, diesmal schauen wir uns die Spaltenprojektionen an. Dort sehen wir, dass die Parameterwerte genauso übernommen wurden: Auch die Bindevariable kommt so in den SQL-Text. Wichtige Erkenntnis ist, dass die Kontrolle über den Einsatz von Bindevariablen durch die Verwendung von SQL Macros nicht verloren geht und beim Entwickler bleibt.

Tabellenmakros

Die zweite Art der SQL Macros – die Tabellenmakros – sind nur in der FROM-Klausel verwendbar und liefern ein SQL-Fragment zurück, aus dem weiter selektiert werden kann. Tabellenmakros bieten Potenzial für vielfältigere Anwendungsfälle an. Bevor wir die ganze Vielfalt erkennen können, steigen wir wieder mit einem einfacheren Anwendungsfall ein. Es soll eine Hilfsstruktur geschaffen werden, um nur Daten der Beschäftigten aus der Abteilung „SALES“ zu selektieren. Eine Hilfsstruktur braucht

man in diesem Fall nicht zwingend, aber wenn, dann würde man wahrscheinlich eine View anlegen, wie im *Listing 5*.

Wir wissen, dass Tabellenmakros ein SQL-Fragment zurückliefern, aus dem man selektieren kann. Es wäre also möglich, ein SQL Macro anzulegen und die SQL-Abfrage wie in der View **v_emp_sales** (siehe *Listing 5*) zurückzuliefern. Worin besteht der Unterschied, was sind die Vorteile?

Wir legen das SQL Macro (siehe *Listing 6*) so an, dass der Vorteil sofort sichtbar wird. Im Gegensatz zu einer herkömmlichen View können wir das Makro mit einem Parameter versehen! Somit haben wir gerade eine

Art „parametrisierbare“ View erstellt – ein Feature, das ganz oben auf der Wunschliste vieler Oracle-Entwickler stand.

Die letzte Abfrage im *Listing 6* sieht sehr einfach aus. Doch welche Abfrage läuft nach der Ausführung des SQL Macro tatsächlich in der Datenbank? Für Tabellenmakros kann man es mithilfe der Prozedur **DBMS_UTILITY.EXPAND_SQL_TEXT** sehen (siehe *Listing 7*).

Die Tabellenmakros bieten noch weitere Möglichkeiten an, damit der zurückzugebende SQL-Code dynamisch gebaut werden kann. Dem Entwickler stehen zwei spezielle Parametertypen zur Verfügung, die man bereits im Kontext von polymorphen Tabellenfunktionen in der Version 18c gesehen hat: Tabellen- und Spaltenlistenparameter.

Je nach Anwendungsfall können die SQL Macros eine oder mehrere Inputtabellen vorsehen. Das Wort „Tabellen“ ist in diesem Fall ein Sammelbegriff: Views, Materialized Views oder in der WITH-Klausel deklarierte Unterabfragen sind auch erlaubt. Auch das Wort „Input“ muss man an der Stelle richtig interpretieren: Es geht nicht um Datenfluss, denn bei der Ausführung von SQL Macros beim Parsen fließen noch gar keine Daten. Es geht lediglich darum, dass diese „Tabellen“ für den Zusammenbau vom Rückgabe-SQL relevant sind.

Die Tabellenparameter dienen dazu, die Metainformationen über diese Datenstrukturen innerhalb des SQL Macro zur Verfügung zu stellen. Der Datentyp von Tabellenparametern ist ein im Package **DBMS_TF** deklarierter Recordtyp **TABLE_T**.

```

DECLARE
    l_clob CLOB;
BEGIN
    DBMS_UTILITY.expand_sql_text (
        input_sql_text => 'select empno, ename
                          FROM   f_emp_sales (''ACCOUNTING'')',
        output_sql_text => l_clob );
    DBMS_OUTPUT.put_line(l_clob);
END;
/

SELECT "A1"."EMPNO" "EMPNO", "A1"."ENAME" "ENAME"
FROM (SELECT "A2"."EMPNO" "EMPNO", "A2"."ENAME" "ENAME", ...
      FROM (SELECT "A3"."EMPNO" "EMPNO"
            , "A3"."ENAME" "ENAME", ...
            FROM "SCOTT"."EMP" "A3"
            WHERE "A3"."DEPTNO"= (SELECT "A4"."DEPTNO" "DEPTNO"
                                FROM "SCOTT"."DEPT" "A4"
                                WHERE "A4"."DNAME"='ACCOUNTING')
            ) "A2"
      ) "A1"

```

Listing 7: Der tatsächliche SQL-Code bei Tabellenmakros

Über weitere verschachtelte Records hat man Zugriff auf die Informationen zu Spalten der verwendeten Tabelle: Spaltennamen, Datentyp, Länge, Collation etc.

Das hört sich zwar kompliziert an, aber die Übergabe dieser Parameter ist kinderleicht: Die Datenbank übernimmt die Konvertierung zum Datentyp DBMS_TF.TABLE_T und die Befüllung der Metainformationen im Hintergrund. Man muss nur den Tabellennamen oder -alias beim Aufruf des SQL Macro referenzieren.

Mithilfe von Tabellenparametern lassen sich mit SQL Macros sogenannte polymorphe Views bauen. Denn die Geschäftslogik eines Makros wird über die Metadaten zu Tabellen ausformuliert und die Makros können dann mit beliebigen Tabellen ausgeführt werden. Weiter besteht oft die Anforderung, bestimmte Spalten – in der Regel als Listen – als Parameter zu übergeben. Dazu verwendet man Parameter vom Typ DBMS_TF.COLUMNS_T. Beim Aufruf übergibt man diesen mithilfe des Pseudooperators COLUMNS().

Komplexeres Anwendungsbeispiel

Ein weniger triviales Beispiel hilft hier zu verstehen, wie das alles zusammenspielt und welches Potenzial dahintersteckt. Die Aufgabe lautet, die ersten *n* Datensätze aus der Tabelle EMP als Schlüssel-Wert-Pärchen darzustellen. Die Lösungsstrategie kann hier zum einen sein, die relativ neue Top-N-Syntax „FETCH FIRST N ROWS“ zu verwenden, um das Ergebnis auf die gewünschten Datensätze einzuschränken, und zum anderen, den UNPIVOT-Operator zu nutzen, um aus Spalten eines Datensatzes je einen Datensatz im Format Spaltenname-Spaltenwert zu produzieren. Zu beachten ist, dass die Spalten im UNPIVOT alle denselben Datentyp haben müssen. Für unsere Anforderung bedeutet das, dass wir alle Spalten vorher zu Texten konvertieren sollten. Die Abfrage einer möglichen Lösung könnte wie im *Listing 8* aussehen.

Jetzt können wir die Abfrage aus dem *Listing 8* als Grundlage nehmen und ein generisches SQL Macro für die gestellte Aufgabe entwickeln. Die Funktion **UNPIVOT_TOP_N** (siehe *Listing 9*) bekommt einen Tabellenparameter **p_tab**, denn wir wollen, dass das Makro nicht nur für die Tabelle EMP funktioniert. Neben der Einschränkung der An-

```
SQL> SELECT col_name, col_value
2   FROM   (SELECT to_char(empno) AS empno
3           ,      ename
4           ,      to_char(mgr) AS mgr
5           FROM   emp
6           ORDER BY ename
7           FETCH FIRST 2 ROWS ONLY)
8   UNPIVOT INCLUDE NULLS
9   (col_value FOR col_name IN (empno, ename, mgr) );
```

COL_N	COL_VALUE
EMPNO	7876
ENAME	ADAMS
MGR	7788
EMPNO	7499
ENAME	ALLEN
MGR	7698

Listing 8: Top-N als Key-Value

```
SQL> CREATE FUNCTION unpivot_top_n (p_tab IN DBMS_TF.TABLE_T
2                                     , p_limit IN VARCHAR2
3                                     , p_order IN DBMS_TF.COLUMNS_T)
4 RETURN CLOB SQL_MACRO IS
5   v_order_list VARCHAR2(2000);
6   v_sql CLOB;
7   v_cols clob ;
8   v_unpivot clob ;
9   v_str varchar2(200);
10  v_name dbms_id;
11  BEGIN
12    FOR I IN 1..p_tab.column.count LOOP
13      v_name := p_tab.column(i).description.name;
14      CASE p_tab.column(i).description.type
15        WHEN dbms_tf.type_varchar2 THEN
16          v_str := v_name;
17        WHEN dbms_tf.type_number THEN
18          v_str := 'to_char('||v_name||') as '||v_name;
19        WHEN dbms_tf.type_date THEN
20          v_str := 'to_char('||v_name||
21                    ', 'YYYY-MM-DD HH24:MI:SS') as '||v_name;
22      END CASE;
23      v_cols := v_cols || ',' || v_str;
24      v_unpivot := v_unpivot || ',' || v_name;
25    END LOOP;
26    v_cols := SUBSTR(v_cols,2);
27    v_unpivot := SUBSTR(v_unpivot,2);
28
29    SELECT LISTAGG(column_value,',')
30    INTO   v_order_list
31    FROM   TABLE (p_order);
32
33    v_sql := 'SELECT col_name, col_value
34            FROM   ( SELECT '||v_cols||
35                    ' FROM p_tab ORDER BY '||v_order_list||
36                    ' FETCH FIRST p_limit ROWS ONLY)
37            UNPIVOT INCLUDE NULLS
38            (col_value for col_name in ('||v_unpivot||') )';
39    RETURN v_sql;
40  END;
41  /
```

Function UNPIVOT_TOP_N compiled

Listing 9: Top-N Key-Value als SQL Macro

```

266 SELECT *
267 FROM unpivot_top_n(scott.emp
268 , 2
269 , COLUMNS(ename));
270

```

COL_NAME	COL_VALUE
1 EMPNO	7876
2 ENAME	ADAMS
3 JOB	CLERK
4 MGR	7788
5 HIREDATE	1987-05-23 00:00:00
6 SAL	1100
7 COMM	(null)
8 DEPTNO	20
9 EMPNO	7499
10 ENAME	ALLEN
11 JOB	SALESMAN
12 MGR	7698
13 HIREDATE	1981-02-20 00:00:00
14 SAL	1600
15 COMM	300

```

271 SELECT *
272 FROM unpivot_top_n(scott.dept
273 , 1
274 , COLUMNS(loc));
275

```

COL_NAME	COL_VALUE
1 DEPTNO	40
2 DNAME	OPERATIONS
3 LOC	BOSTON

Abbildung 1: Ein SQL Macro ist einsetzbar mit verschiedenen Tabellen (Quelle: Andrej Pashchenko)

zahl der Datensätze **p_limit** sehen wir noch einen Spaltenlisten-Parameter **p_order** vor. Der wird benötigt, um die ORDER-BY-Klausel für die Top-N-Abfrage festzulegen. Denn diese wird nur dann konsequente Ergebnisse liefern, wenn eine eindeutige Sortierung vorgegeben ist. Für das SQL müssen wir zwei Spaltenlisten dynamisch zusammenbauen: eine für die UNPIVOT-Klausel und eine für die SELECT-Liste, bei der wir noch zusätzlich je nach Datentyp alle Spalten zum Text konvertieren. Dafür nutzen wir die Metadaten der Tabelle – reine Fleißarbeit (siehe Zeilen 12 bis 27). Dann noch das Array aus dem Parameter **p_order** als kommaseparierte Liste darstellen und wir haben alle Zutaten, um das endgültige SQL nach dem Rezept aus dem Listing 8 zusammenzubauen (Zeilen 33 bis 38). Achten Sie darauf, wie hier mit Parametern umgegangen wird: Die in der Prozedur dynamisch gebauten Bestandteile werden zwar konkateniert, aber die Funktionsparameter – die Tabelle **p_tab** und die Anzahl **p_limit** – werden einfach referenziert, die Ersetzung mit ihren Werten findet automatisch statt.

Wie das Listing 9 zeigt, haben wir mit recht überschaubarem Aufwand ein richtig generisches Konstrukt schaffen können. Das SQL Macro kann jetzt mit beliebigen Datenquellen eingesetzt werden, wie es in der Abbildung 1 dargestellt ist. Die Anwendungsfälle, die man mit den zur Verfügung

stehenden Mitteln auf eine ähnliche Art und Weise adressieren kann, sind vielfältig. Von einfacheren Tools, wie CSV- oder HTML-Export, über parametrisierte Views, Vereinfachung des Zugangs zu komplexen syntaktischen Strukturen des modernen SQL wie MATCH_RECOGNIZE-Klausel bis hin zur Unterstützung neuer, nicht im Standard-SQL existierender Konstrukte, wie spezieller Joins (Fuzzy Joins, Best-Match Join).

Fazit

Natürlich müssen sich die SQL Macros als ganz neue Technologie erst noch beweisen. Der Ansatz erscheint dennoch vielversprechend. Mächtigkeit und Flexibilität der modernen SQL-Möglichkeiten können ausgenutzt werden – und dies erst noch mit guter Lesbarkeit, Modularisierung des Codes und optimaler Ausführungsperformance.

Oracle hat die Wichtigkeit des neuen Features bestätigt, indem die Tabellenmakros auf Long Term Release 19c zurückportiert wurden. Eine gute Nachricht für diejenigen, die nicht bei Innovation Releases mitmachen und auf den nächsten Long Term Release 23c warten müssen – Sie können mit Tabellenmakros ab Version 19.6 arbeiten!

Wie immer ist bei einem neuen Feature der ein oder andere Bug durchaus zu erwarten. Und es sind auch einige offensichtliche

Hausaufgaben vom Datenbankhersteller zu erledigen, beispielsweise muss das echte SQL-Statement besser sichtbar und nicht nur auf Umwegen herauszufinden sein. Jedoch sind in einigen Projekten des Autors die Schwachstellen der bisherigen Implementierung und die Vorteile durch möglichen Einsatz der SQL Macros an diesen Stellen so offensichtlich, dass man die ersten sorgfältigen Tests nicht mehr in die ferne Zukunft schieben möchte. Vielleicht sieht es bei Ihnen ähnlich aus und wir hören noch aus den Zeiten von 19c die ersten Praxis-Berichte? Viel Spaß mit SQL Macros!

Weitere Informationen

- <https://docs.oracle.com/en/database/oracle/oracle-database/21/nfcon/sql-macros-282902288.html>
- <https://blog.sqlora.com/en/tag/sql-macros/>



Andrej Pashchenko
andrej.pashchenko@trivadis.com



MariaDB-/MySQL-Stolperfallen – und wie komme ich da wieder raus? – Teil 2

Olli Sennhauser, FromDual

Im ersten Teil des Artikels, in der letzten Red-Stack-Ausgabe 2/2021, stellte ich eher allgemeine oder schon länger bekannte Stolperfallen vor. In diesem zweiten Teil beleuchte ich eher MariaDB-/MySQL-spezifische Stolperfallen, die das tägliche Leben des MariaDB-/MySQL-Nutzers prägen können.

Crash

Ich kann das Wort schon bald nicht mehr hören! Es wird heute völlig inflationär verwendet. Praktisch alles, was nicht normales Datenbankverhalten ist, wird heute als „Crash“ bezeichnet. Was wir bis heute unter dem Begriff Crash schon alles angetroffen haben, reicht von „Datenbankverbindung ist verloren gegangen“ über „Datenbankverbindung wurde terminiert“, „Systemstillstand“ (verursacht durch beispielsweise Locks), „Verbindungen wurden abgewiesen“, „Datenbank wurde sauber runtergefahren, maximale Anzahl erlaubter Verbindungen wurde überschritten“ bis hin zu „Datenbankserver wurde durch

den Oom-Killer (Out-of-Memory) terminiert“ (siehe Listing 1). Das sind alles keine Crashes. Ein Crash erfolgt dann, wenn die Datenbank auf einen Bug stößt und sich terminiert oder durch das O/S terminiert wird. Dies erfolgt üblicherweise mit einem

Stack-Trace, der im Error Log festgehalten wird (siehe Listing 3).

Und wie komme ich da wieder raus? Als Erstes sollte man aufhören, wie wild mit irgendwelchen Begriffen um sich zu schmeißen, und die Zeit viel lieber in die Problem-

```
Syslog:
```

```
...  
Out of memory: Kill process 1904 (mysqld) score 39 or sacrifice child  
Killed process 1904 (mysqld) total-vm:2855024kB, anon-rss:449640kB,  
file-rss:0kB  
...
```

```
NO Shutdown sequence in MariaDB / MySQL error log (kill -9)
```

Listing 1: Auszug aus dem Syslog, wenn der Oom-Killer die Datenbank terminiert.


```

MariaDB / MySQL error log:

[Note] bin/mysqld (initiated by: unknown): Normal shutdown
...
[Note] bin/mysqld: Shutdown complete

--

Version: '10.0.19-MariaDB' socket: '/var/lib/mysql/mysql.sock' port: 3306 MariaDB Server
/usr/sbin/mysqld (mysqld 10.0.19-MariaDB) starting as process 4469 ...
...
[Note] Server socket created on IP: '::'.
[Note] bin/mysqld: ready for connections.

```

Listing 2: Normale Shutdown- und Startup-Sequenz der Datenbank bei einem Neustart. Gefunden im Error Log der Datenbank.

analyse investieren. Denn erst dann, wenn wir das Problem verstanden haben, können wir konkret und zielführend etwas dagegen unternehmen. Zur Analyse gehören unter anderem das Auswerten der Server-Logs, der Datenbank-Logs sowie der Applikations-Logs (siehe Listing 2). Auch ein gutes Monitoring kann bei der Analyse des Problems manchmal helfen.

Tabellen ohne Primary Key

In einem relationalen Datenmodell sollte jede Zeile eindeutig identifizierbar sein. Dies geschieht typischerweise durch den eindeutigen Primary Key. Wenn Tabellen keinen Primary Key haben, ist das also schlechtes Design.

Neben diesem mehr oder weniger philosophischen Grund gibt es allerdings auch noch technische Gründe, warum jede Tabelle einen Primary Key haben sollte. Technisch gesehen ist InnoDB eine Art Key-Value Store. Am besten performt InnoDB über den Primary Key. Was aber, wenn keiner da ist? Sowohl Latenz als auch Durchsatz werden dann potenziell schlechter. Auch bei der Master/Slave-Replikation hilft ein Primary Key, Replication-Lags zu vermeiden. Und sowohl Galera Cluster als auch der InnoDB Cluster erfordern zwingend auf jeder Tabelle einen Primary Key.

Und wie komme ich da wieder raus? Auf jeder Tabelle einen Primary Key anlegen. Falls es kein schlaues, eindeutiges Kriterium gibt, kann man immer eine Spalte mit einem AUTO_INCREMENT PRI-

MARY KEY darauf anlegen. Dies gilt auch für Star-Schemas in OLAP-Systemen.

Server has gone away

Als ich zum ersten Mal die Meldung „Server has gone away“ gesehen habe, war ich ganz fürchterlich erschrocken. Ein Blick in das Error Log der Datenbank hat jedoch sofort gezeigt, dass die Datenbank sauber durchgelaufen ist. Also Fehlalarm! Diese Meldung macht so eine völlig untaugliche Aussage. Besser wäre eine Meldung in der Art: „Connection was terminated“, was der Realität in den meisten Fällen wesentlich näherkommt. Der Grund für diese Client-seitige Meldung ist ganz einfach, dass die Verbindung zum Server abgebrochen ist. Was die genaue Ursache hierfür ist, kann der Client nicht beurteilen.

Wie kann das passieren? Die häufigste Ursache für diesen Fehler ist entweder, dass die Verbindung gekillt wurde, beispielsweise vom Admin, wait_timeout überschritten wurde oder aber, dass ein zu großes Datenpaket angefordert wurde. Letzteres kann über die Variable max_allowed_packet gesteuert werden. Der wesentlich seltenere Fall ist, dass die Datenbank gekillt oder gestoppt wurde.

Und wie komme ich da wieder raus? Die Situation, dass der DB-Server terminiert wurde, kann relativ schnell überprüft und ausgeschlossen werden. Danach sollte beobachtet werden, wie oft, nach welcher Idle-Zeit und bei welcher Abfrage dieser Fehler auftrat. Nach der Analyse des Verhaltens kann in Betracht gezogen werden, für die betroffene Connection die Variable max_allowed_packet anzupassen oder aber die wait_timeout zu erhöhen.

```

111208 14:35:05 [ERROR] mysqld got signal 11 ;

This could be because you hit a bug. It is also possible that this binary
or one of the libraries it was linked against is corrupt, improperly built,
or misconfigured. This error can also be caused by malfunctioning hardware.
We will try our best to scrape up some info that will hopefully help diagnose
the problem, but since we have already crashed, something is definitely
wrong
and this may fail.

...

Thread pointer: 0x0
Attempting backtrace. You can use the following information to find out
where mysqld died. If you see no messages after this, something went
terribly wrong...
stack_bottom = (nil) thread_stack 0x48000
./bin/mysqld(my_print_stacktrace+0x2e) [0xa1195e]
./bin/mysqld(handle_segfault+0x3f6) [0x6340a6]
/lib64/libpthread.so.0 [0x3248e0eb10]
./bin/mysqld(ma_checkpoint_execute+0x42d) [0x83bead]
./bin/mysqld(ma_checkpoint_background+0x23f) [0x83c9af]
/lib64/libpthread.so.0 [0x3248e0673d]
/lib64/libc.so.6(clone+0x6d) [0x32482d40cd]

```

Listing 3: Echter Crash mit Stack-Trace des mysqld-Prozesses aus dem Error Log der Datenbank.

Too many connections

Wenn Sie diese Fehlermeldung sehen, wissen Sie, dass der Datenbank-eigene Sicherheitsmechanismus zugeschlagen hat. MariaDB/MySQL lassen per Default nur 151 gleichzeitige Datenbankverbindungen zu. Dieser Schutz dient dazu, die Datenbank gegen Überlast durch Amok laufende Anwendungen zu schützen. Der Wert kann für die gesamte Instanz, pro User oder pro Account festgelegt werden.

Dieser Schutz springt dann an, wenn entweder mehr Last als erwartet auf die Datenbank zukommt oder die Datenbank mit der ihr zugemuteten Last nicht (mehr) klar kommt. Häufig sind dies entweder Locks in der Datenbank oder lange laufende Queries.

Und wie komme ich da wieder raus? Es ist natürlich verlockend, jetzt einfach den Überlastschutz durch das Erhöhen der Variable `max_connections` auszuhebeln, es löst allerdings das grundsätzliche Problem nicht. Eigentlich sollte man in Absprache mit der Anwendung und der die Datenbank umgebende Infrastruktur untersuchen, warum es zu dieser Situation kommen konnte, und erst dann Maßnahmen ergreifen. Ein guter Einstiegspunkt, um herauszufinden, warum das Problem auftritt, ist wieder mal der Befehl `SHOW PROCESSLIST`.

Slave Lag

Unter Slave Lag verstehen wir in einer Master/Slave-Replikation ein signifikan-

tes Zurückfallen des Slaves hinter seinen Master (siehe *Listing 4* und *Abbildung 1*). Dies ist dann problematisch, wenn entweder vom Slave gelesen wird, dass signifikant alte Daten gelesen werden, oder aber im Switchover- oder Failover-Fall, wenn auf einen Slave geschwenkt wird, der zeitlich weit zurückliegt.

Der Grund für das Hinterherhinken des Slaves kann verschiedene Gründe haben: Da auf dem Slave per Default alle Befehle serialisiert abgearbeitet werden, können einzelne lang laufende Statements (DML oder DDL) die nachfolgenden Statements lange aufhalten. Ein weiterer Grund können fehlende Primary Keys (siehe weiter oben) sein. Sie führen zu einer sub-optimalen Ausführung der Queries auf dem Slave. Und zu guter Letzt kann eine zu schwache Dimensionierung des Slaves dazu führen, dass dieser schlicht und einfach mangels Ressourcen nicht hinterherkommt.

Und wie komme ich da wieder raus? Je nach Grund können verschiedene Möglichkeiten das Problem lösen. Ein Slave sollte immer mindestens genauso stark wie sein Master dimensioniert sein. Neuere MariaDB/MySQL Releases arbeiten optimaler in verschiedenen der oben genannten Szenarien. Also würde ein Upgrade auf einen aktuellen Release das Problem möglicherweise entspannen. Primary Keys auf allen Tabellen anzulegen, ist eine weitere Möglichkeit, das Problem zu entschärfen. Schließlich kann man den Slave auch so konfigurieren, dass er weniger sicher arbeitet als sein Master, was

```
SQL> SHOW SLAVE STATUS\G
***** 1. row *****
...
Slave_IO_Running: Yes
Slave_SQL_Running: Yes
...
Seconds_Behind_Master: 999
```

Listing 4: SHOW SLAVE STATUS zeigt das Hinterherhängen des Slaves.

zu einer verbesserten Performance führt. Und als letzte Möglichkeit kann ein Slave auch parallel replizieren, was jedoch wieder andere Probleme mit sich bringt.

Sending data

Wer sich schon mal die Ausgabe von `SHOW PROCESSLIST` angesehen hat, hat sicher bemerkt, dass ab und zu für längere Zeit Queries im State „Sending data“ hängen geblieben sind (*siehe Listing 5*). Was hat es damit auf sich? Wenn man das Internet konsultiert, findet man manchmal die falsche Aussage: „Sending data“ bedeutet, dass Daten über das Netz gesendet werden. Diese Aussage stand früher sogar mal so in der MySQL-Dokumentation. Richtig hingegen ist, dass die Storage Engine arbeitet, Daten zusammensucht und sie an den Executor im SQL-Layer „hoch“ schickt, der diese Daten dann weiterverarbeitet (*siehe Abbildung 2*). Also ein klassischer Fall für ein lang laufendes Query und das daraus folgende Query Tuning.

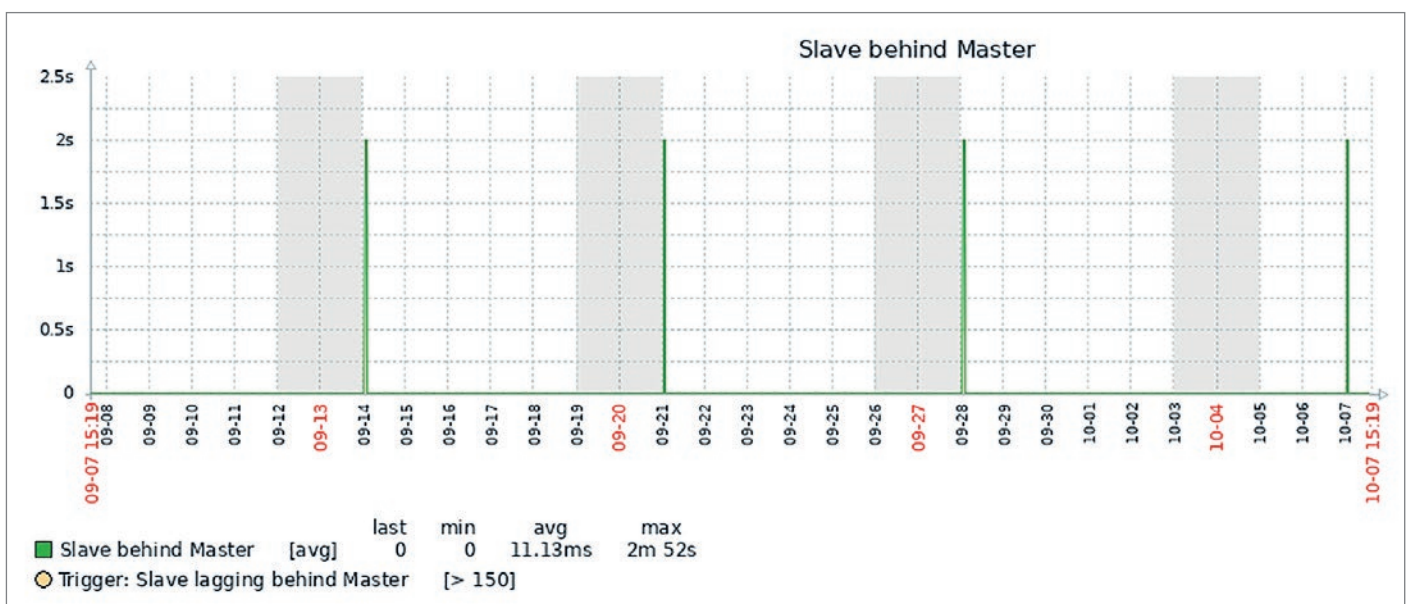


Abbildung 1: Jeden Montagmorgen hängt der Slave bis knapp 3 Minuten hinter seinem Master hinterher. Grund: noch unbekannt.

```
SQL> SHOW PROCESSLIST;
+-----+-----+-----+-----+-----+-----+-----+
| Id      | User | Host      | db   | Time | State      | Info                                     |
+-----+-----+-----+-----+-----+-----+-----+
| 577558 | app  | 10.0.0.4 | test | 42  | Sending data | SELECT * FROM very_big_table          |
+-----+-----+-----+-----+-----+-----+-----+
```

Listing 5: SHOW PROCESSLIST zeigt lange laufende Queries im State Sending data.

Was kann man dagegen tun? Da es sich um ein klassisches Query-Tuning-Problem handelt, als Erstes das gesamte Query sicherstellen und wegspeichern. Allenfalls noch gleich mit dem Befehl SHOW EXPLAIN FOR <thread_id>; den aktuellen Query Execution Plan mit wegsichern. Anschließend gilt es, das Query und den Query Execution Plan zu analysieren und zu verstehen, ob das Query oder die das Query ausführende Anwendung noch Optimierungspotenzial haben oder ob an den Daten oder dem Datenmodell Änderungen vorgenommen werden müssen, um die Abfrage schneller zu machen.

Query Tuning

Query Tuning ist ein Thema für sich und würde einen ganz eigenen Artikel benötigen. Daher hier nur kurz zusammengefasst die wichtigsten einfachen Probleme, die wir in der Praxis sehen. Fundamental wichtig für das optimale Ausführen von Queries sind Indizes. Diese müssen vorhanden und richtig gesetzt sein, sodass der Optimizer sie zu Hilfe nehmen kann.

Als einfache Faustregel gilt: Alle Spalten, die in der WHERE Clause verwendet werden, sind gute Kandidaten für Indizes.

Was kann man jetzt bei Queries falsch machen?

```
WHERE column LIKE '%...%'
```

In diesem Beispiel kann der Index auf die Spalte „column“ nicht genutzt werden aufgrund des %-Zeichens links im Abfragestring. Entweder wird das %-Zeichen links entfernt oder andere Technologien wie Volltextsuche oder gar ein Volltext-Indexer wie Solr werden genutzt, um die Abfrage zu beschleunigen.

```
WHERE UNIX_TIMESTAMP(Date)
< 12454232423;
```

Ein weiteres Problem tritt dann auf, wenn die indexierte Spalte in eine Funktion gepackt wird. Auch in diesem Fall kann der Optimizer den Index nicht verwenden. Der einfachste Weg aus dieser Situation

ist es, das Query so umzuschreiben, dass die Funktion auf der rechten Seite, um die Konstante herum, liegt.

```
WHERE b = 5 AND c = 2;
```

Um das Problem für diese Abfrage zu verstehen, muss man die Datenstruktur, genauer die Indizes auf der Tabelle, kennen. In diesem Beispiel liegt der Index auf den Spalten (a, b, c). Der Optimizer kann den Index in diesem Fall nicht verwenden, weil ihm in der WHERE Clause die Spalte a fehlt. Entweder muss das Query so umgeschrieben werden, dass a ebenfalls in der WHERE Clause vorkommt, oder aber es muss ein neuer Index nur auf die Spalten (b, c) angelegt werden.

Fazit

Es gibt bei MariaDB/MySQL zahlreiche Stolperfallen, die man kennen sollte, um aus ihnen auch wieder herauszufinden. In beiden Artikeln wurden die häufigsten dieser Stolperfallen aufgezeigt und Lösungsansätze dafür beschrieben, wie man mit dem Problem umgehen kann. Datenbanken sind recht komplizierte Gebilde, aber das macht sie ja auch so spannend...

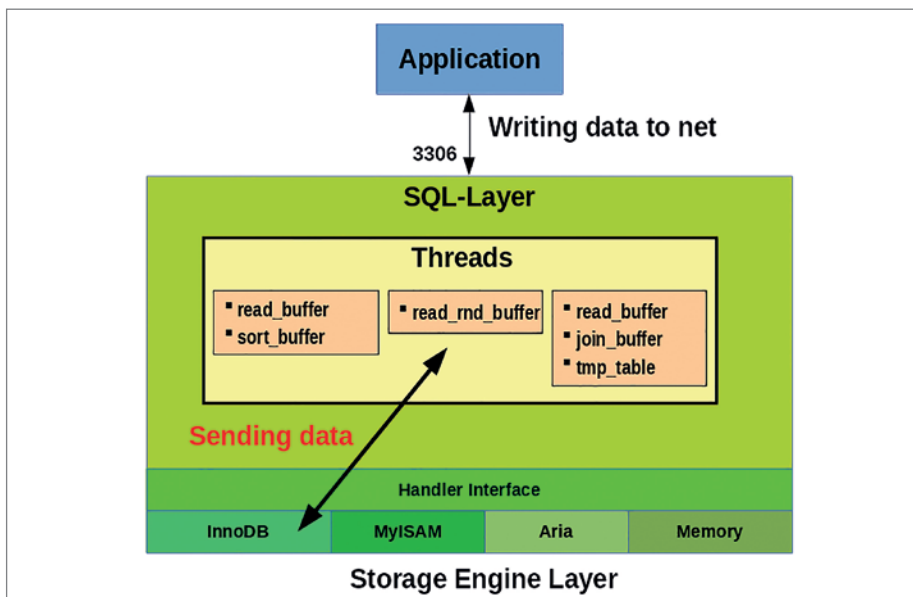


Abbildung 2: Sending data ist die Kommunikation zwischen dem Executor und der Storage Engine.



Olli Sennhauser
oli.sennhauser@fromdual.com

BUSINESS INSIGHTS

NEWS

03/2021



Wege aus dem ökologischen Lock-in: Gestaltung nachhaltiger Geschäftsmodelle

Dr. Thomas Karle & Ana Paula Rivas, Horus software GmbH

Ökologie und Ökonomie wurden lange Zeit als separate Themen betrachtet, die auf den ersten Blick wenig miteinander zu tun haben. Jedoch lässt sich an vielen aktuellen Beispielen beobachten, dass Unternehmen aufgrund von Umweltproblemen vor große wirtschaftliche Probleme gestellt und ohne große Vorwarnung dazu gezwungen werden, ihr Geschäftsmodell massiv anzupassen. Auf der anderen Seite können sich Unternehmen durch die Umsetzung von Nachhaltigkeit in ihrem Geschäftsmodell auch signifikante Wettbewerbsvorteile verschaffen. Im vorliegenden Beitrag wird ein Ansatz vorgestellt, mit dem Unternehmen ihr Geschäftsmodell auf transparente Weise bezüglich ökologischer Nachhaltigkeit umbauen können. Der Ansatz umfasst ein Vorgehen, das neben Steuerungsmechanismen für das Management auch konkrete Konzepte zur Transformation einzelner Geschäftsprozesse und mögliche Zertifizierungen beschreibt.

Ethik in der Unternehmenspraxis

Die industrielle Revolution kann als ein komplexer Prozess verstanden werden, bei dem durch die Einführung von technischen und organisatorischen Innovationen in der Produktionskette die Beschleunigung und das Wachstum der industriellen Produktion möglich war, was zu einer modernen, urbanisierten Gesellschaft und langfristig zu einem kontinuierlichen Wirtschaftswachstum geführt hat. Es ist unbestreitbar, dass die Industrialisierung beträchtliche Fortschritte in Bezug auf das gesellschaftliche Wohlergehen und die Lebensqualität hervorgebracht hat, jedoch können auch die durch sie verursachten aktuellen Probleme nicht ignoriert werden. Auf Basis des Data Catalog der Weltbankgruppe [1] lässt sich leicht feststellen, dass sich im Lauf dieses Prozesses eine Störung des Mensch-Natur-Verhältnisses entwickelt hat. Diese Störung wird durch die vielen aktuellen Umweltprobleme wie Klimawandel, abgeholzte Wälder, Knappheit von Süßwasser, Luftverschmutzung und das Artensterben aufgrund eines rücksichtslosen Umgangs mit den Ressourcen deutlich sichtbar.

Seit ihren Anfängen basiert die industrielle Revolution auf einem linearen Modell von Produktion und Verbrauch, besser bekannt als „take-make-dispose“: Dabei werden Ressourcen aus der Natur entnommen, gefolgt von der Herstellung von Produkten, dem Verkauf und dem Verbrauch, um schließlich in Mülldeponien oder Verbrennungsanlagen entsorgt zu werden. Dieses

Wirtschaftsmodell basierte auf zwei Annahmen: Ressourcen sind unbegrenzt vorhanden und Ressourcen sind jederzeit leicht verfügbar. Diese beiden Grundannahmen der industriellen Revolution sind im aktuellen globalen Kontext jedoch nicht mehr gültig. Vielmehr ist es notwendig, den bestehenden Widerspruch zwischen den Ideen des Wirtschaftswachstums und denen des Umweltbewusstseins aufzulösen.

In den letzten Jahren scheint der Themenkranz Ökologie, Nachhaltigkeit und grüne Prozesse zu einem Trend geworden zu sein. Diese Themen beschäftigen sich nicht nur mit ethischen Fragen, sondern beziehen auch wirtschaftliche Aspekte mit ein. Viele Unternehmen sind sich inzwischen der Instabilität des herkömmlichen Wirtschaftsmodells bewusst und erkennen wirtschaftliche Risiken, die aus Umweltproblemen entstehen können. Es ist notwendig, über Veränderungen in der Wirtschaft nachzudenken, die auch nachhaltige Aspekte mit einbeziehen, um einerseits positive Entwicklungen für Umwelt und Gesellschaft voranzutreiben und andererseits wirtschaftlichen Schäden vorzubeugen. Nachhaltige Geschäftsmodelle sind mehr als nur eine mögliche Alternative zu herkömmlichen Geschäftsmodellen. Die Bewertung der Nachhaltigkeit von Unternehmensprozessen kann als Chance für Innovation, Verbesserung oder gar das Generieren völlig neuer Geschäftsmodelle genutzt werden, um Wettbewerbsvorteile zu erlangen und einen greifbaren wirtschaftlichen Wert zu schaffen.

Nachhaltige Prozesse umsetzen heißt die Zukunft bauen

Veränderungen, die in einer Organisation stattfinden, sind untrennbar mit den Veränderungen in den entsprechenden Geschäftsprozessen derselben Organisation verbunden [2, 3]. Die Kompetenzen eines Unternehmens sind nicht nur in seinen Produkten zu erkennen, sondern sie zeigen sich auch in seinen Prozessen. In vielen Unternehmen finden im Rahmen der Prozessgestaltung oft jeweils separate Projekte und Abstimmungen zu den Aspekten Zielsetzung, Strategie und grobe Organisationsstruktur in der Unternehmensführung sowie zu Detailabläufen mit den jeweiligen Geschäftsregeln in den Fachbereichen statt. Dementsprechend werden diese Aspekte oft separat gestaltet, ohne dass diese im Zusammenhang behandelt werden. Betrachtet und analysiert man jedoch einen Prozess im Gesamtkontext unter Berücksichtigung aller relevanten Aspekte, so erhält man eine ganzheitliche Betrachtung innerhalb eines Unternehmens, die sowohl bei der Gestaltung und der Verbesserung als auch bei der späteren Nutzung verwendet werden kann. Hierzu ist eine kollaborative Plattform erforderlich, die eine gemeinschaftliche und zusammenhängende Bearbeitung dieser Aspekte ermöglicht.

Zusammenfassend wurden in [4] zum Thema Innovationsmanagement für Green BPM bereits grobe Ansätze beim Geschäftsprozessmanagement vorgestellt, um die Transformation von aktuell konventionell laufenden Geschäftsprozessen hin zu nachhaltigen

Geschäftsprozessen (Green Processes) zu unterstützen. Nachhaltigkeit wurde hierbei als ein umzusetzendes Ziel mit entsprechenden Detailspekten betrachtet, die es ausgehend vom bisher klassisch umgesetzten Geschäftsprozess einzubauen und zu erweitern gilt.

Das Hauptziel von „grünem“ BPM und dem Umbau zu nachhaltigen Vorgehensweisen besteht dann darin, die durch die Geschäftsprozesse verursachten negativen Auswirkungen auf die Umwelt zu reduzieren, ohne dabei die wirtschaftlichen Ziele einer Organisation zu vernachlässigen. In diesem Artikel wird nun darüber hinaus noch ein konkretes Konzept in Form eines Katalogs aus Green Services und einer dazugehörigen Vorgehensweise zu deren Umsetzung im Unternehmen vorgestellt. Um die Umweltziele bezüglich Nachhaltigkeit zu unterstützen, ist es hierbei notwendig, Umwelt und Organisation als Gesamtsystem zu verstehen. Der Katalog enthält bereits vorgefertigte Services und Anforderungen in Form von Beispielen für Nachhaltigkeitsziele für Unternehmen, mit denen eine unternehmensspezifische Definition konkreter Prozessumbaumaßnahmen zur Erreichung dieser Ziele sowie deren transparente Überprüfung bezüglich der Erfüllung in einem Unternehmen erfolgen kann.

Green Services als Wettbewerbsvorteil

Ein Geschäftsprozess kann als eine Reihe von Aktivitäten betrachtet werden, der im Rah-

Ökologischer Lock-in
 Selbst wohlgemeinte Nachhaltigkeitsinitiativen scheitern oft nicht am Willen der Verantwortlichen, sondern an ihrer Umsetzung im Unternehmen. Die Gründe hierfür können vielfältig sein: mangelnde Kapitalausstattung, fehlendes Wissen, überalterte Produktionsmittel, ungeeigneter Produktmix und in ganz vielen Fällen auch starre Geschäftsprozesse, die sich Nachhaltigkeitsaspekten schlichtweg verschließen. Eine solche Umsetzungskrise lässt sich treffend als „ökologischer Lock-in“ bezeichnen. Wege aus einem derartigen Lock-in lassen sich nur finden, indem alle für die Umsetzung relevanten Aspekte, beginnend mit dem Geschäftsmodell über die Geschäftsprozesse bis hin zu den IT-Systemen, zur Disposition gestellt werden. Insofern wird die Unterstützung des Executive Management, ja selbst der Eigentümer, zum kritischen Erfolgsfaktor für mehr Nachhaltigkeit im Unternehmen.

men seiner Durchführung Werte schafft. Diese Wertschöpfung ist der Kern eines jeden Geschäftsmodells. In einem nachhaltigen Geschäftsmodell wird diese Wertschöpfung auf Basis von nachhaltigen Dienstleistungen, sogenannten Green Services oder auch Eco Services, erbracht. Hammerl et al. [5] stellen diesbezüglich einen systematischen Ansatz für Produkte und Dienstleistungen vor. *Abbildung 1* zeigt die grundlegende Idee und die Struktur des Ansatzes von [5] und ergänzt diesen um konkrete Services auf Basis des Konzepts der Ökoeffizienz, das vom World Business Council for Sustainable Development (WBCSD) [6] eingeführt wurde.

Nach dem zugrunde liegenden Ansatz von [5] können Dienstleistungen in primäre

(immaterielle) und sekundäre (mit einem Produkt verbundene) Dienstleistungen unterteilt werden. Primäre Dienstleistungen stellen reine Dienstleistungen dar und sind immaterieller Natur. Sekundäre Dienstleistungen werden in produktergänzende und produktersetzennde Dienstleistungen unterschieden. Zu den ergänzenden Produktdienstleistungen können Dienstleistungen gehören, die den Lebenszyklus eines Produkts verlängern, sowie Dienstleistungen zur Produktrücknahme. Produktersetzende Dienstleistungen können in nutzungsorientierte und ergebnisorientierte Dienstleistungen unterteilt werden. Das entworfene Green-Services-Modell in *Abbildung 1* basiert auf dem Ansatz von Hammerl. So kann dieser Katalog als Basis ge-

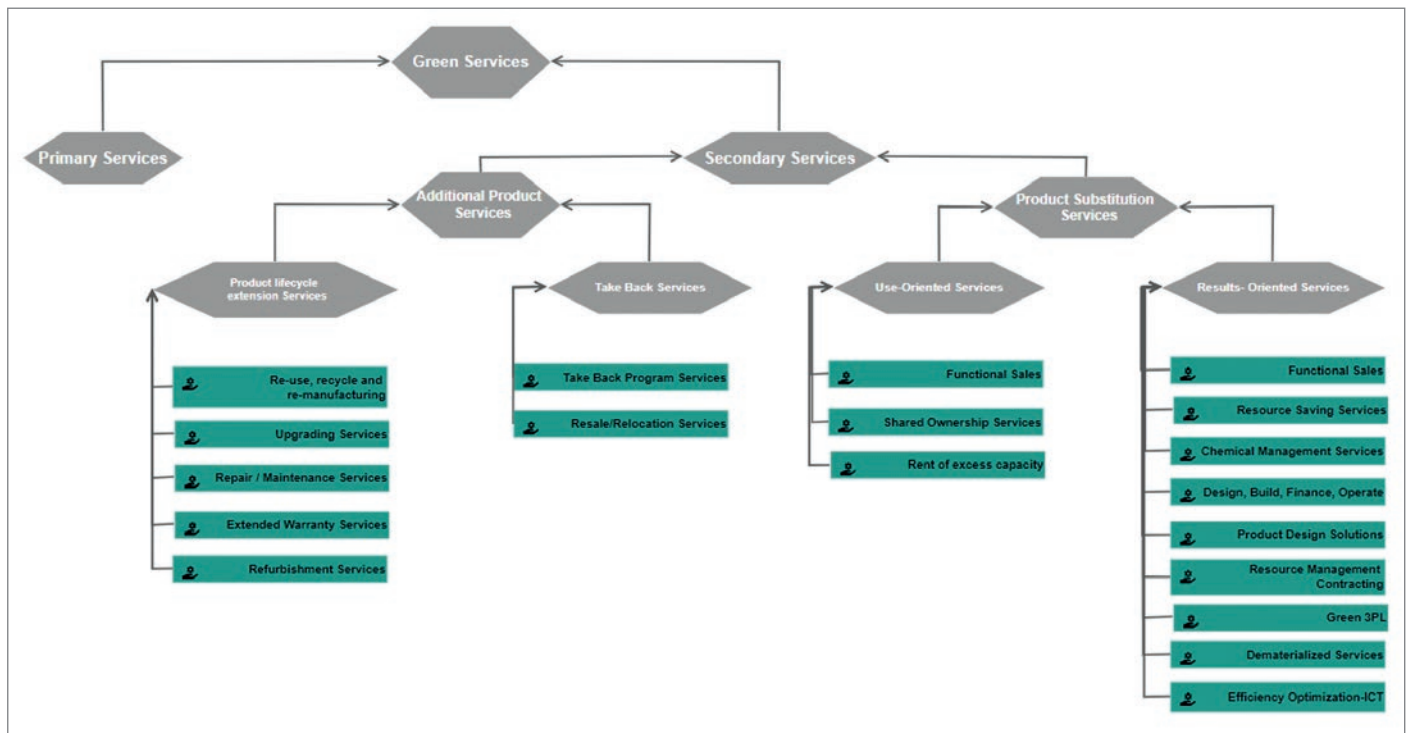


Abbildung 1: Green Services [5,6,7]

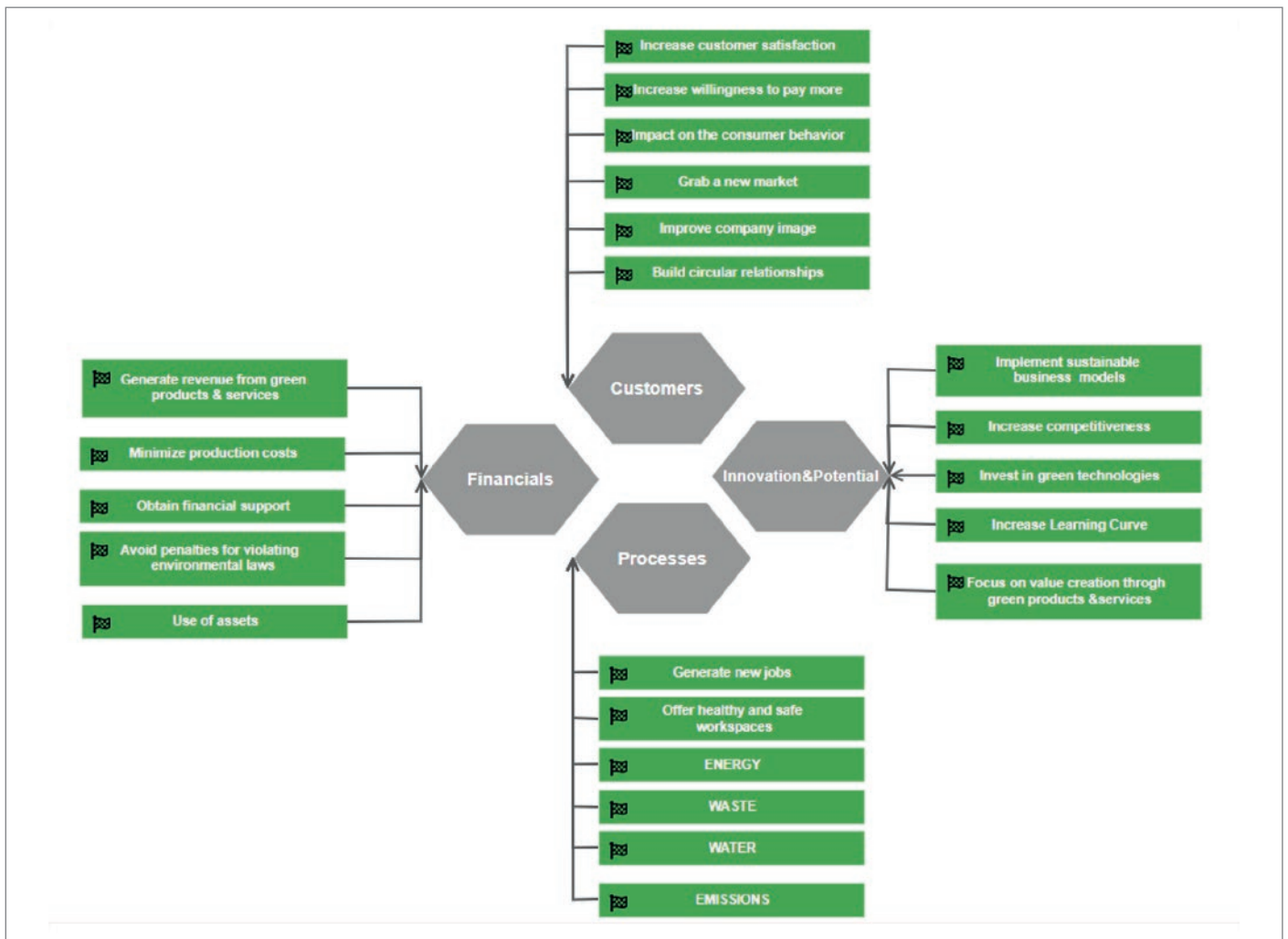


Abbildung 2: Green Business Goals [7]

nutzt werden, um ein neues, nachhaltiges Geschäftsmodell zu definieren.

Balanced Scorecard: Grüne Ziele, Strategien und Risiken

Für Butler, Henderson & Raiborn [8] sind die größten Beschränkungen für die Einbeziehung von Nachhaltigkeit in die finanziellen Ziele und Strategien eines Unternehmens die fehlende Messbarkeit der Auswirkungen des aktuellen Stands bezüglich Nachhaltigkeit auf den Shareholder Value, die fehlenden Möglichkeiten, die positiven Auswirkungen der Umsetzung von Nachhaltigkeit auf die finanzielle Performance zu dokumentieren und das Fehlen einer strukturierten Entscheidungsgrundlage, die auch Umweltaspekte berücksichtigt. Hier kann jedoch eine entsprechende Nutzung der Balanced Scorecard (BSC) dazu beitragen, diese Barrieren zu überwinden, indem sie einen Rahmen für die Darstellung und Ermittlung von qualitativen Kennzahlen zu den Abläufen eines Unternehmens bereitstellt und diese mit Unterneh-

menszielen und -strategien zur Umsetzung und Einhaltung von Nachhaltigkeitsprinzipien verknüpft. Das BSC-Konzept entstand in den 90er-Jahren und wurde von Kaplan und Norton [9] als neue Methode zur Messung der Leistung von Unternehmen vorgeschlagen, um von der einseitigen Ausrichtung auf Buchhaltung und finanzielle Ziele wegzukommen. Sie führten diese neue Methodik ein, um finanzielle und nicht-finanzielle Ziele innerhalb eines Performance-Management-Systems zu kombinieren. Hierbei wurden alle relevanten Aspekte für ein ausgewogenes Gesamtbild des Unternehmens identifiziert, jeweils Ziele für diese gesetzt und Wege gefunden, um den jeweiligen Fortschritt bezüglich deren Erreichung zu messen.

Hier ist vor allem die dynamische Natur dieses Instruments hervorzuheben, welche die Ursache-Wirkungs-Beziehungen zwischen den Perspektiven berücksichtigt und dem Management eine umfassende und leicht verständliche Sicht auf den aktuellen Stand der Umsetzung des Geschäftsmodells

bietet. Durch die Einbindung von qualitativen Kennzahlen zu Nachhaltigkeit in dieses Instrument ist es möglich, auch eine Bewertung von Umweltaspekten vorzunehmen und aufgrund der Ursache-Wirkungs-Beziehungen zu und zwischen den vier Standard-Perspektiven Finanzen, interne Prozesse, Innovation & Potenzial und Kunden die Zusammenhänge und Abhängigkeiten erkennen und darstellen zu können. Hierbei werden die folgenden zugrunde liegenden Instrumente genutzt:

- **Green Business Goals:** Im Zuge der geplanten Anpassung des Geschäftsmodells eines Unternehmens für einen Umbau auf Basis von Green Services werden zunächst die diesbezüglich zu erreichenden Ziele definiert. Die Ziele sollten im Einklang mit der Mission des Unternehmens stehen und aus dieser abgeleitet werden können. Bei der Zieldefinition für Green Business Goals ist zu berücksichtigen, dass die Umweltaspekte auf die gleiche Stufe wie die finanziellen Aspek-

te gestellt werden. Dies ermöglicht eine Steuerung hin zu einer Reduzierung negativer Umweltauswirkungen, die im Rahmen der Geschäftsprozesse entstehen, ohne die wirtschaftlichen Ziele einer Organisation zu vernachlässigen. Die Green Business Goals werden zunächst für die interne Prozessperspektive definiert. Weitere Green Business Goals können dann jedoch auch für die weiteren Perspektiven festgelegt und zugeordnet werden. *Abbildung 2* zeigt Beispiele für Green Business Goals, orientiert an der Grundstruktur der Balanced Scorecard.

- Green Business Strategies:** Die Green Business Goals sind der Ausgangspunkt für die Definition von jeweils zuzuordnenden Green Business Strategies, die im Unternehmen angewendet werden sollen, um die Nachhaltigkeitsziele zu erreichen. Auch hier beginnt die Definition bei der Perspektive der internen Prozesse.

- Risiken beim „Going Green“:** Auch die Risiken müssen betrachtet und bewertet werden. Beim „Going Green“ sind hier neben den üblichen Unternehmensrisiken insbesondere die möglichen Gefahren durch die Umstellung des Geschäftsmodells zu berücksichtigen, welche die Leistungserbringung einschränken oder das Unternehmen destabilisieren könnten. Risiken sind den zuvor festgelegten Zielen des Unternehmens zugeordnet und können auch die zugehörigen Strategien negativ beeinflussen.

Key Environmental Indicators

Das Orientieren und Steuern anhand von Leistungsindikatoren ist ein wesentlicher Bestandteil des hier vorgestellten Konzepts. Die Messung von Leistungsindikatoren für Nachhaltigkeit kann entlang der gesamten Wertschöpfungskette erfolgen, um den Erreichungsgrad der Green Business Goals zu bewerten. Vom Brocke et al. [10] bekräftigen, dass diese Nachhaltigkeitsindikatoren, sogenannte Key Environmental Indicators, auf

Aktivitäts-, Prozess- und Unternehmensebene existieren müssen. Darüber hinaus müssen sie in Verbindung mit den klassischen Metriken, die sich auf die Effizienz des Unternehmens sowie die Kosten und die Qualität der Ergebnisse beziehen, analysiert werden. Die Berücksichtigung von Nachhaltigkeitsindikatoren bedeutet für viele Unternehmen einen Paradigmenwechsel in der Unternehmens- und der Managementstruktur.

Bei den Kriterien für die Bewertung von Nachhaltigkeitsindikatoren gibt es zwei Kategorien, die von der International Standard Organisation [11] bereitgestellt werden: Management-Leistungsindikatoren und operative Leistungsindikatoren. Das hier vorgestellte Konzept konzentriert sich auf die Untersuchung von Leistungsindikatoren, die mit konkreten Prozessaktivitäten verbunden werden können, um Verbesserungen bezüglich der Nachhaltigkeit bei den internen Prozessen bewerten und steuern zu können. *Abbildung 3* zeigt eine Struktur mit konkreten Beispielen für Nachhaltigkeitsindikatoren.

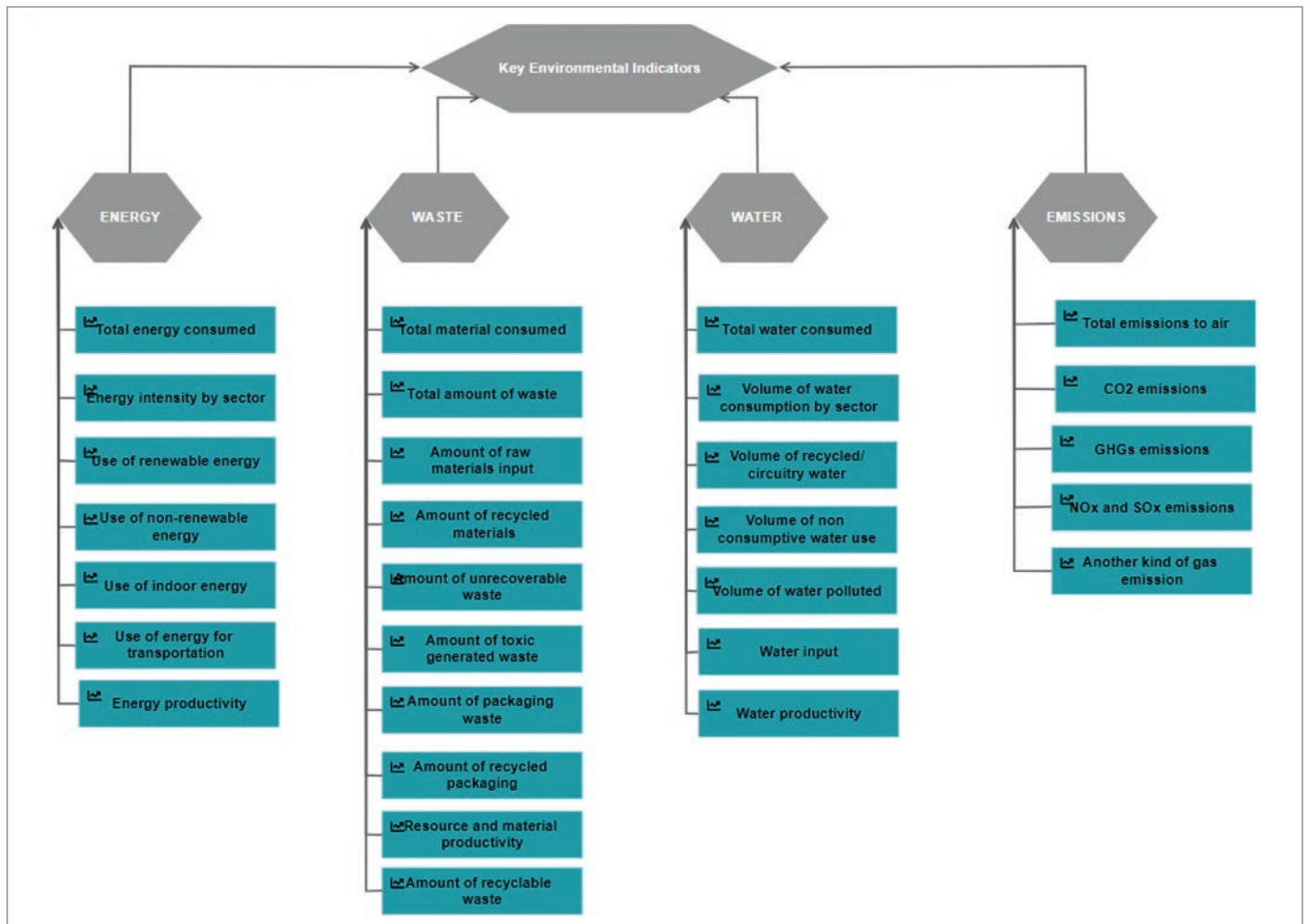


Abbildung 3: Key Environmental Indicators [7]

Nachhaltige Prozesse

Das Geschäftsprozessmanagement spielt neben den bisher vorgestellten – eher strategisch orientierten – Aspekten eine zentrale Rolle bei der Transformation in ein „grünes“ Geschäftsmodell. Die Darstellung in Form von Prozessmodellen – idealerweise auf verschiedenen Abstraktionsebenen – unterstützt das Management, den Prozessverantwortlichen und den Nachhaltigkeitsbeauftragten dabei, Schnittstellen oder Abhängigkeiten zwischen einzelnen Prozessen auf den unterschiedlichen Ebenen zu erkennen und weitere Strategien zur Verbesserung der Nachhaltigkeitsaspekte in den Geschäftsprozessen zu entwickeln. Somit wird ein Überblick über das operative Geschäft geschaffen, der auch die Beziehungen, Wechselwirkungen und Einflüsse bezüglich Umweltaspekten im Rahmen der konkret laufenden Geschäftsprozesse berücksichtigt, um notwendige Maßnahmen und Modifikationen durchzuführen.

Die Idee, „grüne“ Prozessmodelle zu etablieren, geht auf den Beitrag der Autoren Leymann et al. [12] zurück. Hier werden Prozessmuster eingeführt, die für die Umsetzung von Nachhaltigkeit in Geschäftsprozessen angewendet werden können, sogenannte Green Business Process Pattern. Daraus wurden drei grundlegende Prozessmuster als Konzept aufgegriffen und in das im Rahmen dieses Ansatzes vorgestellte Vorgehen eingebunden. Ein weiteres viertes Konzept ergibt sich aus der Kreislaufwirtschaft. Für die Transformation der Geschäftsprozesse für ein Green Business Model wird der Einsatz der folgenden grundlegenden Konzepte empfohlen [12]:

- **Green Compensation:** Dieses Prozessmuster kann als mögliche Lösung für Unternehmen genutzt werden, bei denen es schwierig ist, bestehende Prozesse neu zu gestalten oder umzustrukturieren. In einem solchen Szenario können Kompensationsaktivitäten – beispielsweise der Kauf von entsprechenden Zertifikaten zur Finanzierung von Klimaschutzprojekten – eruiert werden, um die durch die konventionellen Abläufe verursachten Umweltauswirkungen (zumindest teilweise) auszugleichen.
- **Green Variant:** Das nächste Konzept ist der Einsatz einer Green Variant. Bei der Umstellung auf einen nachhaltigen Prozess können aus der finanziellen und/oder der Kundenperspektive Nachteile entstehen. Ein Unternehmen ist möglicherweise nicht

in der Lage, seine Prozessabläufe zu ändern, ohne dass zusätzliche Kosten entstehen oder sich der Komfort aus Kundensicht verschlechtert. Als Konsequenz könnte diese Änderung eine Erhöhung der Produktpreise und eine geringere Akzeptanz am Markt bedeuten. Bei der Green Variant wird ein zusätzlicher Weg umgesetzt, indem ein alternatives „grünes Produkt“ angeboten wird, das dann über eine spezielle nachhaltigere Geschäftsprozessvariante abgewickelt wird, die weniger negative Umweltauswirkungen mit sich bringt als die ursprüngliche Variante, gegebenenfalls jedoch mehr Kosten verursacht.

- **Green Process Improvement:** Dieses Konzept ist die Anwendung von BPM-Praktiken auf die Umsetzung von Nachhaltigkeitsaspekten. Hierbei werden potenziell zu optimierende Prozesse und einzelne Aktivitäten auf der Grundlage von Leistungs- und Umweltkennzahlen sowie Zielen und Strategien identifiziert. Der Identifizierung von Prozessen folgt eine Minimierung der Schritte zum Erreichen eines Ziels oder die Vermeidung von Aktivitäten, die einen hohen Schaden für die Umwelt bedeuten.
- **Circular Procedure:** Kreislaufverfahren stellen einen Ansatz dar, der darauf abzielt, den Nutzen von Produkten, Komponenten und Materialien über den gesamten Lebenszyklus zu maximieren. Die Hauptidee ist die Wiederverwendung von Materialien und Komponenten. Kreislaufverfahren sind auf eine effiziente Nutzung von Ressourcen ausgerichtet. Über den gesamten Lebenszyklus eines Produkts hinweg werden in den dazugehörigen Geschäftsprozessen – soweit wie möglich – Rückführungen von Komponenten und Materialien sowie die Reduzierung oder Beseitigung von Abfällen angestrebt. Die Kreislauffähigkeit kann anhand des Grads der Rückgewinnung von Materialien, Wasser und Energie sowie der Verlängerung des Produktlebenszyklus bewertet werden.

Für die Umsetzung von Nachhaltigkeitszielen in den Geschäftsprozessen können oft auch Kombinationen der aufgeführten Konzepte zu einer guten Lösung führen.

Zertifizierungen für nachhaltige Geschäftsprozesse

Die Umsetzung von Nachhaltigkeit in Unternehmen kann zertifiziert werden. Bei-



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 Ettlingen
+49 7243 2179-0
www.promatis.de

Ettlingen | Hamburg | Berlin | Münster
Wien | Zürich | Denver

spielsweise können sich Unternehmen nach ISO 14001 [13] zertifizieren lassen. Eine weitere Möglichkeit stellt die Zertifizierung als B Corporation [14] dar. Das Hauptziel eines Geschäftsmodells einer B Corporation ist es, soziale, wirtschaftliche und ökologische Veränderungen mit positiven Auswirkungen zu fördern und gleichzeitig Wachstum, Gewinne und Beschäftigung zu generieren. Um eine B Corporation zu werden, muss sich ein Unternehmen zu einer kontinuierlichen Verbesserung verpflichten und dies durch regelmäßige Neuzertifizierungen nachweisen. Seit der Einführung im Jahr 2006 gibt es mittlerweile über 4.000 zertifizierte B Corporations in mehr als 60 Ländern.

B Lab Germany wurde 2020 als Non-Profit-Organisation und unternehmerische Gemeinschaft gegründet. Das B Lab Germany richtet sich in erster Linie an verantwortungsbewusste Unternehmer aus allen Branchen, Gründer und deren Mitarbeiter, aber auch an Vertreter aus Organisationen, Medien, Wissenschaft und Politik. Hierbei wird ein kollaborativer Prozess mit dem Ziel unterstützt, eine verantwortungsvolle, ökologisch nachhaltige und sozial gerechte Wirtschaft als neuen Status quo in Deutschland, Europa und der ganzen Welt zu etablieren.

Fazit

Mit dem vorliegenden Beitrag wurde ein Ansatz vorgestellt, mit dem Unternehmen ihr Geschäftsmodell auf transparente Weise bezüglich ökologischer Nachhaltigkeit umbauen können. Auf Basis eines Katalogs mit Green Services können nachhaltige Aspekte in die Prozesse eines Unternehmens integriert werden. Im Rahmen des Vorgehens für eine solche Transformation werden zunächst die zu erreichenden Nachhaltigkeitsziele definiert. Im Anschluss daran werden Strategien zur Erreichung der gesetzten Ziele entwickelt. Darüber hinaus werden auch Risiken für eine solche Transformation betrachtet und mit den Zielen und Strategien in Bezug gesetzt. Zur Messung des Grads der Erreichung der Nachhaltigkeitsziele werden spezielle Key Performance Indicators (KPIs) definiert und in einer hierzu angepassten Balanced Scorecard zusammengestellt. Dieser Teil des Vorgehens stellt die Steuerungsmechanismen für das Management eines Unternehmens bereit, das einen solchen Umbau plant. Ein besonderes Augenmerk liegt bei diesem Ansatz auf der gemeinsamen Betrachtung der ökonomischen und der ökologischen Aspekte.

Für die konkrete Anpassung von Geschäftsprozessen wurden vier kombinierbare Basiskonzepte (Green Compensation, Green Variant, Green Process Improvement und Circular Procedure) vorgestellt, mit denen existierende Geschäftsprozesse analysiert und sukzessiv umgebaut werden können. Die zuvor beschriebene Balanced Scorecard mit den ökonomischen und ökologischen Aspekten ist hier ein wichtiges Steuerungselement, um zu beurteilen, welche Green Services im Rahmen der Transformation der Geschäftsprozesse berücksichtigt werden sollen. Dieses Konzept dient als Grundlage für die Gestaltung und Transformation von nachhaltigen Geschäftsprozessen, stellt jedoch keinen starren Rahmen dar, sondern ermöglicht das spezifische Anpassen an das jeweilige Unternehmen.

Literatur

- [1] The World Bank Group – Data Catalog: <https://data.worldbank.org/indicator>
- [2] F. Schönthaler, G. Vossen, A. Oberweis und T. Karle, Business Processes for Business Communities: Modeling Languages, Methods, Tools, Springer: <https://doi.org/10.1007/978-3-642-24791-0>, 2012.
- [3] G. Vossen, F. Schönthaler und S. Dillon: The Web at Graduation and Beyond: Business Impacts and Developments, Springer, 2017.
- [4] T. Karle: Innovationsmanagement für Green BPM, in: DOAG Business News 05/2020 – Nachhaltiges Innovations- und Produktmanagement.
- [5] B. Hammerl et al: Nachhaltige Produkte und Dienstleistungen: Leitfaden zur Entwicklung zukunftsfähiger Geschäftsfelder, Vienna, Austria, Federal Ministry for Transport, Innovation and Technology, 2003.
- [6] WBCS – Eco-efficiency: <https://www.wbcd.org/Projects/Education/Resources/Eco-efficiency-Learning-Module>
- [7] A.P. Rivas: Going Green through Horus Business Modeler Software: Transformation of a Conventional Business into a Sustainable Model, Bachelor Thesis, Karlsruhe, Hochschule Karlsruhe – Technik und Wirtschaft, 2020.
- [8] J. Butler, S. Henderson, C. Raiborn: Sustainability and the Balanced Scorecard: Integrating Green Measures into Business Reporting. Management Accounting Quarterly. Vol. 12, No. 2, 2011.
- [9] R.S. Kaplan, D.P. Norton: The Balanced Scorecard: Translating Strategy into Action, Harvard Business Press, Boston, USA, 1996.
- [10] J. vom Brocke, S. Seidel, J. Recker: Green Business Process Management: Towards the Sustainable Enterprise. Berlin, Germany: Springer Verlag GmbH, 2012.
- [11] International Organization for Standardization: Environmental Management – Environmental Performance Evaluation – Guidelines, 2013: <https://www.iso.org/obp/ui/#iso:std:iso:14031:ed-2:v1:en>
- [12] F. Leymann, A. Nowak, D. Scheicher, D. Schumm, S. Wagner: Green Business Process Patterns, IAAS Institute of Architecture of Application Systems, University of Stuttgart, 2011.
- [13] ISO 14001: <https://www.iso.org/iso-14001-environmental-management.html>
- [14] B Corporation: <https://bcorporation.eu/about-b-lab/country-partner/germany>



Ana Paula Rivas
ana.rivas@horus.biz

Ana Paula Rivas ist Consultant Digitized Processes bei der Horus software GmbH mit Schwerpunkt Green Business Process Management. Bereits während ihres Studiums an der Hochschule Karlsruhe Technik und Wirtschaft hat sie das Thema der Gestaltung nachhaltiger Geschäftsmodelle vertieft und dies in einer entsprechenden Bachelor Thesis bearbeitet.



Dr. Thomas Karle
thomas.karle@doag.org

Dr. Thomas Karle ist COO und Strategieberater der Horus software GmbH, der Product Company der PROMATIS Unternehmensgruppe. Hier ist er in die Entwicklung von Methoden und Produkten für geschäftsprozessorientierte Ansätze zur Implementierung von Unternehmenssoftware-Lösungen und zu sonstigen Business-Transformationen wie der Umsetzung von mehr Nachhaltigkeit in Unternehmen eingebunden. Darüber hinaus ist er Vorstand Business Solutions der DOAG.



„KI und Machine Learning werden uns helfen, weiterhin im Enterprise Data Management für Unternehmen die beste Wahl zu sein“

Gespräch mit Jochen Wießler, ORACLE

Jochen Wießler, Vice President ERPM Oracle Deutschland, verantwortet als ausgewiesener Experte für Business-Applikationen seit 2020 das hiesige Geschäft für die Bereiche Enterprise Resource Planning (ERP), Supply Chain Management (SCM) und Enterprise Performance Management (EPM). Was er in seiner neuen Position als Vizepräsident anstoßen möchte, welche Schwerpunkte er setzen will und wie sich die Digitalisierung in Deutschland auf das Oracle-Geschäft auswirkt, darüber sprach er mit Marcos López, Redaktionsleitung Business News.

Herr Wießler, was haben Sie sich bezüglich Ihres Aufgabenfeldes für 2021 vorgenommen?
Oracle ist in allen drei Bereichen (ERP, SCM, EPM) sehr gut aufgestellt und wir wollen unseren Wachstumskurs auf jeden Fall fortsetzen. Das vergangene Jahr 2020 war sehr erfolgreich für den Bereich ERPM in Deutschland. Diesen Trend wollen wir wei-

ter beibehalten und den Erfolg im Jahr 2021 übertreffen. Die große Stärke von Oracle ist unsere ganzheitliche Plattform von der Infrastruktur bis hin zur Applikation. Alles vollkommen Cloud-basiert und integriert. Kürzlich wurde festgestellt, dass Oracle die am schnellsten wachsende Cloud-Firma der Welt ist. Diesen Sog wollen wir nutzen und

gerade im SAP-Heimatland Deutschland unsere Marktanteile weiter ausbauen.

Welche Veränderungen hat die Coronavirus-Pandemie auch für Ihre Arbeit gebracht?

Für uns als Cloud-Anbieter bietet die Pandemie auch Chancen. Nie war es für viele Unternehmen dringender, ihr Geschäft und

auch ihre Liquidität trotz aller Unabwägbarkeiten zu planen, ihre Lieferketten stabil aufrechtzuerhalten und auch ihre Prozesse zu digitalisieren. Dabei setzen viele Unternehmen auf Agilität und Cloud-basierte Anwendungen. Sie stellen den Status quo infrage und damit oft auch die üblichen Hersteller. Das ist unsere Chance – und die nutzen wir gerade.

Wie sehen Sie die Oracle-Business-Software-Lösungen im deutschen Markt aufgestellt? Es gibt derzeit mit der Oracle E-Business Suite, den Oracle Cloud Applications und Oracle NetSuite gleich drei unterschiedliche ERP-Systeme von Oracle. Wie findet ein Kunde die für ihn am besten geeignete Lösung?

Kunden, die ein neues ERP-System einführen möchten, sollten sich immer von ihren Kernanforderungen leiten lassen. Da gibt es keine „one size fits all“-Lösung. Oracle deckt alle Bedürfnisse der Kunden ab. Bei „on-premises“-orientierten Kunden ist die Oracle E-Business Suite das Produkt der Wahl, in mittelständisch orientierten Unternehmen, die eine vollständige ERP Suite zu geringen Kosten suchen, bietet sich Oracle NetSuite an. Für alle anderen Bereiche ist der Kunde mit Oracle Fusion ERP bestens aufgestellt. Mehr als 220 verfügbare best practices unterstützen auch die etwas komplexeren Anforderungen.

In der Vergangenheit war Oracle im ERP-Umfeld nicht unbedingt als ein intensiver Anbieter bei deutschen Unternehmen erkennbar. Der Großteil der hiesigen Kunden sind Tochterunternehmen ausländischer Konzerne. Gibt es aktive lokale Anstrengungen, den deutschen Markt aktiver zu bearbeiten?

Aus meiner Sicht ist Oracle Cloud ERP der „hidden champion“ im deutschen Markt. Weltweit gehört Oracle bereits zu den führenden Cloud-ERP-Anbietern mit bereits über 7.500 Kunden. Die Lösung ist von allen führenden Analysten als hochmodern und marktführend anerkannt. Nur in Deutschland tun wir uns zugegebenermaßen noch etwas schwer. In vielen Unternehmen ist SAP über lange Jahre gewachsen, und obwohl viele Unternehmen wegen der Kosten und der Komplexität beim Umstieg auf S/4 HANA verärgert sind, trauen sie sich nicht so ganz, das „Alte“ über Bord zu werfen. Aber wir machen bereits große Fortschritte in den Bereichen Enterprise Performance Management und auch im Bereich Supply Chain Management. Hier haben wir marktführende Lösungen und ge-

winnen weiter das Vertrauen vieler deutscher Unternehmen. Das wird uns mittelfristig auch für das ERP-Geschäft helfen.

Wie sehen Sie die Bereitschaft deutscher Unternehmen, die ERP-Systeme in Cloud-Lösungen zu bringen?

Das ist unterschiedlich. Je größer und traditionell verwurzelt die Unternehmen mit der SAP sind, desto geringer nehmen wir die Bereitschaft wahr, wirklich den Schritt in die Cloud zu gehen. Wir sehen vielmehr in den Firmen aus dem Service- und Projekt-orientierten Geschäft eine Zielgruppe, die Cloud-affin und an einem schnellen Time-to-Market interessiert ist. Außerdem sehen wir die große Zahl von Unternehmen, die sogenannten „Grown Start-Ups“, die mit einem digitalisierten Geschäftsmodell groß geworden sind, als eine Zielgruppe, für die sich unser Angebot äußerst attraktiv darstellt. Diese Unternehmen wollen schnelle Einführungen, auf Standards basierend, hin zu einer skalierbaren Cloud-Plattform, die sie bei ihrem Wachstum unterstützt. Die zunehmende Digitalisierung der Prozesse und die Anforderungen von sich stetig ändernden Märkten/Geschäftsmodellen werden die Verbreitung von Cloud-basierten ERP-Systemen weiter beschleunigen.

An welchen Lösungen sitzen die Entwicklerteams aktuell?

Die Cloud bringt es mit sich, dass Innovationen nicht mehr mit einem „Major-Release“ und einer großen Ankündigung ausgerollt werden. Innovation passiert bei uns ständig, in der Regel mit Verbesserungen und Weiterentwicklungen alle drei Monate. Unsere Kunden können so jederzeit und kontinuierlich die Innovationen konsumieren. Ohne ins Detail zu gehen, aber Oracle ist immer dann stark, wenn es um Daten geht. Ich behaupte, dass kein anderes Unternehmen die Fähigkeit besitzt, Unternehmensdaten aller Art so effizient mit einer ganzheitlichen Cloud-Plattform zu verwalten und zu verarbeiten wie Oracle – von der Infrastruktur bis hin zu den diversen Applikationen. KI und Machine Learning bilden hier wichtige Eckpfeiler, die uns helfen werden, weiterhin im Enterprise Data Management für Unternehmen die beste Wahl zu sein.

Welchen Einfluss hat der aktuelle Stand der digitalen Transformation in Deutschland auf die Implementierung neuer ERP-/SCM-Lösungen? Wie oben erwähnt wird die zunehmende Digitalisierung von Prozessen und Ge-

schäftsmodellen die Implementierung Cloud-basierter Unternehmenslösungen weiter beschleunigen. In vielen Unternehmen starten Projekte zur Modernisierung der Finanzprozesse begleitet von dem Wunsch, eine agilere Plattform zu haben, um präziser zu planen oder strategisch modellieren zu können. Hier sind wir mit Oracle Fusion ERP und Enterprise Performance Management perfekt aufgestellt. Ähnlich sieht es bei den Lieferketten aus, die gerade in diesen Zeiten so wichtig für viele Unternehmen sind. Seien es Order-to-Cash-Prozesse oder End-to-End-Logistiklösungen, wir bieten Unternehmen leistungsstarke Supply-Chain-Management-Lösungen in der Cloud, die sich auch in SAP-Umgebungen integrieren lassen. Oracle ist hier bereits der Partner vieler, agiler und wachstumsorientierter Unternehmen.

Herr Wießler, herzlichen Dank für das Interview.



Jochen Wießler
jochen.wiessler@oracle.com

Jochen Wießler, Vice President ERPM Oracle Deutschland, ist seit 2020 bei Oracle beschäftigt. Er verfügt über lange Erfahrung im Geschäft von Business-Applikationen. Vorherige Stationen waren Führungspositionen bei Microsoft (unter anderem als Dynamics Lead) und auch bei SAP (Leiter Mittelstand und Partner). Die letzten drei Jahre verbrachte Jochen Wießler als Geschäftsführer DACH bei dem ERP-Anbieter Unit4. In seiner jetzigen Funktion verantwortet er in Deutschland das Geschäft für die Bereiche Enterprise Resource Planning, Supply Chain Management und Enterprise Performance Management.



Die Rolle der Ethik im Datenmanagement

Denise Baidinger

*Unternehmungen aller Größenordnung rufen aus, datengetrieben zu arbeiten, ihre Prozesse zu digitalisieren, zu automatisieren und schon bald KI einzusetzen. Neueste technologische Formen der Datennutzung und -teilung stehen für den Chief Digital und Information Officer bereit, um damit die nächste Ära einzuläuten. Datenstrategen konzipieren fleißig, Formen der Daten-Governance werden erprobt, Metadatenkataloge eingeführt, Datenplattformen designt und an der Datenqualität geschraubt, um somit mittelfristig den Data Scientists jegliche Grundlage zu schaffen, mit komplexen Algorithmen neueste Erkenntnisse aus den Daten ziehen zu können. Innerhalb dieser Grundlagenarbeit sowie innerhalb der analytischen Leistung wird viel investiert – eine Frage bleibt jedoch häufig auf der Strecke: **Wie weit wollen wir gehen? Wann entsteht die Gefahr, mittels Datennutzung Kunden oder Mitarbeiter zu diskriminieren und gegebenenfalls auszuschließen? Welches Wertegerüst spannt den Rahmen? Wir alle tragen die Innovationskraft in uns – aber wollen wir uns auch die Frage nach der Verantwortung und den ethischen Grenzen stellen? Halt! Erst eine klar ausgeprägte Datenethik und eine gelebte Datenkultur spannen das Spielfeld, auf dem sich die Experten austoben sollten und dürfen.***

Regulatorik versus Ethik

Zu unterscheiden sind hierbei die regulatorischen Anforderungen von der Etablierung des ethischen Wertegerüsts. Alle Unternehmen, die mit personenbezogenen Daten aus Europa arbeiten, haben die Datenschutzgrundverordnung einzuhalten – das Gesetz schreibt es vor. Das ethische Wertegerüst beschreibt im Gegensatz dazu das selbstgewählte Verständnis, welchen Wert Daten für die Unternehmen besitzen und

welche ethischen Rahmenbedingungen und Kontrollinstanzen vorliegen. Folgende Fragen können dabei gestellt werden: Welche unserer Daten wollen wir nutzen und/oder teilen? Wie werden die Daten korreliert? Welche Algorithmen finden wir vor? Wie werden die Erkenntnisse genutzt und weiterverarbeitet? Das Ziel der ethischen Grunddiskussion ist es, jegliche Art der Diskriminierung gegenüber Kunden, Mitarbeitern oder Geschäftspartnern aus-

zuschließen und jeden Menschen in Bezug auf seine Selbstbestimmung und Freiheit zu schützen. Mit der wertorientierten Datennutzung wird stets das Ziel verfolgt, eine etwaige Überwachung, Ausbeutung oder einen Eingriff in die Privatsphäre beziehungsweise in die Autonomie zu verhindern [1] [2]. Unternehmenswerte sind dabei immer unternehmensspezifisch, die ethischen Grenzen können daher unterschiedlich ausfallen.

Die Einführung der Datenethik im Unternehmen

Mein Rat ist, die ethische Diskussion von Beginn an mitzudenken – Unternehmen konzipieren ihre Datenstrategie; die Datenethik ist ein unumstrittener Teil! Sie ist ein neuer Faktor, der einen großen Einfluss auf die Wettbewerbsfähigkeit haben kann. Nachhaltigkeit und Ethik sind meiner Meinung nach die neuen Säulen der unternehmerischen Reputation und beeinflussen damit mittel- und langfristigen Unternehmenswert deutlich. Somit kann festgehalten werden, dass die Datenethik als neue Disziplin innerhalb der Wirtschaft, aber auch Wissenschaft, eingeführt werden muss, um die Diskussionen umfangreich und ziel führend zu führen. Die große Frage ist nun, wie wir uns unternehmensintern möglichen ethischen Grenzen annähern. Dazu ist eine inter- und sogar transdisziplinäre Zusammenarbeit von verschiedensten Kompetenzträgern notwendig, die heutzutage nur selten in der Wirtschaft anzutreffen sind. Die Zeit der Geisteswissenschaftlicher ist meiner Meinung nach gekommen – Philosophen, Theologen, Historiker sind herzlich willkommen und sogar aufgefordert, die neuen Sichtweisen gegenüber Entscheidern aus der Wirtschaft vorzustellen, sie zu sensibilisieren und damit neue Entscheidungswege zu eröffnen. Diese erhalten durch einen aktiven Auftrag seitens des Vorstandes oder des Aufsichtsrats Legitimation, auch wirklich flächendeckend Gehör zu finden. **Datenethik ist keine Dekoration für Marketingzwecke, sondern ein ernstes Feld**, die durch Technologie eröffneten Möglichkeiten nicht unwissentlich zu missbrauchen. Damit ist Datenethik ein wichtiger Teil, um unternehmerische Verantwortung zu übernehmen und nachhaltig zu agieren. Nachdem ethische Vorgaben gemacht sind, geht es darum, diese auch wirklich innerhalb der Forschungsarbeit zu integrieren und die Einhaltung nachzuhalten. Das Prinzip „Ethic by Design“ fordert dabei, das ethische Wertegerüst als Grundlage der Forschungsaktivitäten von Beginn an zu berücksichtigen, umzusetzen und zu testen; zusätzlich die gesamte Belegschaft innerhalb der Unternehmen davon zu überzeugen, dass dieses Wertegerüst die Grundlage jeglicher Handlung ist. Das World Economic Forum veröffentlichte im Dezember 2020 in Zusammenarbeit mit Deloitte und dem Markkula Center for Applied Ethics at Santa Clara University einen Report zum Thema „Ethics by Design – an organizational approach to responsible use of technology“ [3], der Empfehlungen zum praktischen Ein-

satz gibt. Dabei spielen Change-Methodiken und Schulungsprogramme eine entscheidende Rolle, um Mitarbeiter abzuholen, ihr Interesse und ihre Aufmerksamkeit auf ethische Werte zu richten und sie dafür zu gewinnen, daran mitzuwirken und sogar Begeisterung zu empfinden. Auch das Open Data Institute publizierte im Jahr 2019 ein Data Ethics Canvas [4], eine Art Werkzeugkoffer, der dabei hilft, ethische Herausforderungen zu identifizieren und zu behandeln. Dies sind zwei prominente Beispiele, die aufzeigen, dass die Bedeutung der Datenethik innerhalb der Gesellschaft zunimmt; die Anzahl der wissenschaftlichen und nichtwissenschaftlichen Veröffentlichungen dazu steigt insgesamt deutlich.

Nachhaltige Umsetzung der Datenethik

Zur nachhaltigen Umsetzung der Datenethik ist die Gründung eines Ethikrates im Unternehmen damit unausweichlich. Die Bestimmung und kontinuierliche Nachkontrolle der Datenethik-Prinzipien gibt den Scientists und Analysten langfristig eine Orientierung dahingehend, in welchen Datenräumen und nach welchen Regeln geforscht wird, Daten gesammelt und genutzt werden. Der Ethikrat dient damit als Kontrollinstanz, um die Arbeit mit Daten in der Unternehmung, die Formen der Datenteilung im Partnerökosystem und die Datennutzung (inkl. Daten-Monetarisierung) zu überblicken, zu steuern und immer wieder mit den definierten ethischen Werten abzugleichen. Entscheidern, einerseits innerhalb der IT-, andererseits innerhalb der Business-Abteilungen, ist sogar zu raten, eine ausführliche und präzise Datenethik als Rahmenbedingung für ihr Aufgabenspektrum einzufordern oder diese selbst mitzugestalten, um jederzeit sicherzustellen, dass ein Datenmissbrauch ihrerseits ausgeschlossen ist. Viele Unternehmen denken langsam darüber nach, ein Gremium zu ethischen Fragestellungen zu gründen; oftmals werden die Rufe seitens der Aufsichtsräte laut. Sollten wir, die damit operativ tätig sind und im Alltag viele Entscheidungen treffen, nicht gegenüber dem Topmanagement eine definierte Ethik als Arbeitsgrundlage fordern? Sobald die Ethik unser wirtschaftliches Handeln in technologischen Zeiten beeinflusst, kann die Nachhaltigkeit ihr Potenzial entfalten. Vielleicht erleben wir eine Zeit, in der wir technologische Innovationen aufgrund von ethischen Bedenken bewusst ablehnen oder neu designen. Das beschreibe ich als techno-

logische Reife, das Prinzip „Ethic by Design“ in Forschungsprojekten anzuwenden und damit technologische Potenziale ganz bewusst und voller Überzeugung zugunsten der Gesellschaft, der persönlichen Freiheit und Selbstbestimmung ungenutzt zu lassen. Als fortschrittliche Gesellschaft müssen wir uns der ethischen Grenzen bewusst sein und die Konsequenzen verstehen.

Dieser Artikel ist angelehnt an meinen Beitrag im Herausgeberband „Marketing Analytics“, Springer Verlag, Erscheinungsdatum: voraussichtlich Sommer 2021 [5].

Quellen

- [1] Gutachten der Datenethik-Kommission der Bundesregierung. Oktober 2019, Berlin
- [2] Data Ethics Framework, UK Department for Digital, Culture, Media & Sport, August 2018, London
- [3] White Paper „Ethics by Design: An organizational approach to responsible use of technology“, World Economic Forum, 2020, Cologne
- [4] The Data Ethics Canvas, Open Data, 2019, London
- [5] Herausgeberband „Marketing Analytics“, Springer Verlag, Erscheinungsdatum: voraussichtlich Sommer 2021.



Denise Baidinger
Mail@denise-baidinger.de

Denise Baidinger beschäftigt sich seit ihrem Masterstudium mit Fragestellungen der Nachhaltigkeit und Ethik, die unsere technologiegetriebene Wirtschaft stark beeinflussen. In den Jahren 2012 und 2013 forschte sie innerhalb des Netzwerkes EIT Climate-KIC mit dem Ziel, Systeminnovationen für den Übergang zu einer kohlenstoffarmen Wirtschaft in Europa zu entwickeln. Von 2013 bis 2017 konzipierte Denise Baidinger als Senior Consultant Modelle zur Customer und Employee Experience. Seit 2018 arbeitet sie für die Deutsche Bahn AG als Manager IT-Strategie und initiierte in den Jahren 2018 und 2019 die Einführung einer konzernweiten Daten-Governance. Mit ihren digitalen Beiträgen zu der Thematik „Human Centricity – Der Mensch ist im Mittelpunkt“ sensibilisiert Denise Baidinger dafür, die persönlichen menschlichen Bedürfnisse in den Mittelpunkt zu setzen.



„Ein respektvoller und wertschätzender Umgang ist ein absolutes Muss und muss von allen gleichermaßen gelebt werden.“

Gespräch mit Anne Driescher, MT AG

Anne Driescher ist Fachbereichsleiterin Recruiting & Staffing bei der MT AG, einem mittelständischen IT-Unternehmen mit über 270 Mitarbeitern an vier Standorten. Die MT AG entwickelt moderne IT-Lösungen, die die Anpassungsfähigkeit der Kunden an aktuelle Marktbedingungen erhöhen. Hier verantwortet sie in einem Team die Personalgewinnung und das Freelancermanagement sowie die Projektbesetzung, aber auch die Mitarbeiterbindung, das Feel Good Management (FGM), das Personalmarketing und das Employer Branding. Wie gutes Recruiting & Staffing im Jahr 2021 gelingen kann und wie FGM dazu beiträgt, moderne Arbeitswelten Wirklichkeit werden zu lassen, erzählt Anne Driescher im Interview mit Marcos López, Redaktionsleitung Business News.

Frau Driescher, wie sind Sie zu Ihrem Job gekommen? Bruder Zufall, Schwester Glück oder Bewerbung auf den Traumjob?

Es war wohl ein bisschen von allem. Während meines Bachelor-Studiums der Wirtschaft arbeitete ich bei Univativ in Düsseldorf, einer

studentischen Unternehmensberatung. Einer der dort tätigen Account Manager empfahl mir die MT AG, ich solle mich einfach nach meinem Studium initiativ bewerben. Gesagt, getan und Glück gehabt – im November 2011 durfte ich im Staffing/Freelan-

cermanagement starten. Seitdem bin ich hier und ausgesprochen zufrieden.

Zusätzlich habe ich ein nebenberufliches Master-Studium in Wirtschaftspsychologie an der PFH in Göttingen absolviert. Die Kosten für das Master-Studium übernahm die

MT AG. Dies hat mir zusätzlich wissenschaftliches, wertvolles Wissen vermittelt, unter anderem aus den Bereichen biologische Psychologie, Entwicklungspsychologie, Meditation, Coaching, Organisationspsychologie, Konfliktmanagement, Führung und Personalpsychologie. Das hat mir wichtige Impulse für meine Tätigkeit bei der MT AG und unsere Ausrichtung im Recruiting gegeben. Anschließend durfte ich die Fachbereichsleitung für das Recruiting & Staffing übernehmen, und ich bin sehr stolz auf mein mittlerweile siebenköpfiges Team.

In Ihrem Team ist auch eine Feel Good Managerin beschäftigt. Wie sieht ihre Arbeit aus?

Zunächst sei gesagt, dass die MT AG und ich uns einig darüber waren, was unter Feel Good Management (FGM) zu verstehen ist und wie dieses in der MT AG umgesetzt werden sollte. Auf diese Weise konnten wir das FGM gemeinsam entwickeln und in einer Art iterativen Prozess optimieren. Man muss wissen: Das FGM als Stelle und Aufgabenfeld ist generell noch sehr frisch in der Unternehmenswelt und es wird sich in den kommenden Jahren weiter verändern und entwickeln.

Das FGM bei der MT AG beschreibt die individuelle Betreuung unserer Kolleginnen und Kollegen mit dem Ziel, die Zufriedenheit, das Engagement, die Motivation und auch die Identifikation unserer Mitarbeiter mit der MT AG zu erhöhen, ihre Gesundheit zu fördern und die Fluktuation zu senken.

Unsere Mitarbeiter sollen die Möglichkeit haben, sich an eine unabhängige Vertrauensperson zu wenden, der sie ihre Wünsche und Bedürfnisse, aber auch Sorgen und Herausforderungen mitteilen können. Unsere Feel Good Managerin nimmt diese Anfragen sehr ernst und bemüht sich darum, jedem Anliegen gerecht zu werden. Oft sind es Kleinigkeiten, Kollegen wünschen sich zum Beispiel bestimmte Kosmetikartikel in den sanitären Räumen, haben Verbesserungsvorschläge zur Büroausstattung oder Fragen zum Angebot der MT AG im Rahmen der Kinderbetreuung. Aber auch, wenn es mal Unstimmigkeiten zwischen Kollegen oder mit Vorgesetzten gibt, können diese vertraulich mit ihr besprochen werden. Die Themen sind somit sowohl privater als auch beruflicher Natur. Letztendlich soll unser Feel Good Management das Wir-Gefühl stärken.

Was sollte betriebliches Feel Good Management beinhalten? Und was nicht?

Das FGM berücksichtigt Themen wie das betriebliche Gesundheitsmanagement, die

Organisation und Durchführung von Mitarbeiter-Events, Benefits und Geschenke für Mitarbeiter, Spendenaktionen für Umwelt und Soziales und vieles andere mehr. Darüber hinaus unterstützt die Feel Good Managerin im Onboarding und nimmt an Feedback- und Austrittsgesprächen teil. Unser FGM beginnt somit mit der Anstellung eines MT-AG-Mitarbeiters. Unsere Feel Good Managerin stellt sich bei jedem neuen Kollegen persönlich vor und händigt ihm eine liebevoll gestaltete Willkommenstüte aus. Mitarbeiter sollen sich vom ersten Tag an bei uns willkommen fühlen und wissen, dass sie sich mit Ideen, Verbesserungsvorschlägen und Fragen vertraulich an sie wenden können. Auch persönliche Dinge dürfen besprochen werden – oft hilft es, einfach mal mit jemandem zu reden. Gleichzeitig unterstützt sie bei der Organisation und Durchführung von MT-AG-Veranstaltungen und unternehmensweiten Aktionen, der Auswahl und Gestaltung von Mitarbeiter-Goodies und kleinen Aufmerksamkeiten zur Hochzeit, Geburtstagen, Geburt eines Kindes und so weiter.

Das betriebliche Gesundheitsmanagement und entsprechende Maßnahmen wie die Zusammenarbeit und Aktionen mit Krankenkassen, die Gestaltung und Durchführung von Gesundheitsprogrammen und Workshops zu Themen wie Ernährung, Bewegung und psychischem Wohlbefinden sind ebenfalls ein wichtiger Bestandteil unseres FGM.

Vor allem in der Corona-Zeit, Stichwort: mobiles Arbeiten, ist es wichtig, den engen Kontakt zu den Mitarbeitern zu pflegen. Unsere Feel Good Managerin veranstaltet daher zwei Mal pro Woche ein „Käffchen“ per MS Teams, bei dem sich Kollegen austauschen und miteinander in lockerer Atmosphäre quatschen können. Zudem ist sie die (nicht-fachliche) Ansprechpartnerin für unsere (dualen) Studenten und veranstaltet regelmäßige Treffen für unseren „Nachwuchs“, wie Pizza- oder Waffelessen, und sie hilft bei der Erstellung von Haus- oder Bachelorarbeiten.

Betriebliches FGM sollte sich aus meiner Sicht **nicht** mit rechtlichen Aspekten befassen, also keine juristische Hilfestellung oder Beratung bieten. Unsere Feel Good Managerin ist auch keine ausgebildete Psychologin. Sie führt weder Therapien noch Coachings durch! In solchen Fällen würde sie dann gegebenenfalls an Experten weiterleiten. Auch wenn es um rein fachliche Themen geht, kann das FGM nur bedingt weiterhelfen, die Anfragen aber durchaus an die verantwortlichen Personen weiterleiten.

Wichtig aus meiner Sicht ist, dass das FGM eine „unabhängige“ Instanz in einer Organisation darstellt, die zwar in engem Austausch mit anderen Bereichen steht, jedoch nicht direkt an den Vorstand „berichtet“.

Hat die MT AG ein Leitmotiv für den Umgang mit dem Rohstoff „Mitarbeiter“? Wie sorgt man dafür, dass alle 270 Mitarbeiter an vier Standorten (neben Ratingen auch Frankfurt am Main, Köln und München) von den Benefits des FGM profitieren?

Ich denke der Schlüssel ist die individuelle Betreuung. Dadurch, dass wir (noch) nicht allzu groß sind und zudem ein IT-Dienstleister, der bestens mit digitalen Mitteln und Werkzeugen vertraut ist, sind wir gut darin, uns virtuelle Lösungen einfallen zu lassen. Selbstverständlich kann man nie überall zugleich sein, und es gibt in all unseren Niederlassungen auch lokale Unterschiede.

Wichtig allerdings ist, dass man grundsätzlich alle Mitarbeiter gleichbehandelt und gleichermaßen ernst nimmt. Ein respektvoller und wertschätzender Umgang ist ein absolutes Muss und muss **von allen** gleichermaßen gelebt werden. Es geht also vielmehr um die Unternehmenskultur als Ganzes. Ich denke, dass der kollegiale Umgang aller die Basis für ein gutes Miteinander ist und jeder ist ein Teil davon.

Unser FGM ist dabei lediglich das Ergebnis einer großartigen Unternehmenskultur der gesamten MT AG. Allein die Tatsache, dass unser Vorstand eine solche Stelle bewilligt hat, zeigt aus meiner Sicht die Einstellung der Unternehmensleitung. Dieses positive Menschenbild wird top down weitergegeben und in der gesamten MT AG gelebt. Kurz: Unsere wertschätzende und offene Unternehmenskultur sorgt dafür, dass jeder von den Benefits der MT AG profitieren kann und stets nach gerechten und fairen Lösungen für alle gesucht wird. Fühlt sich dennoch jemand benachteiligt, so kann er sich an seinen Vorgesetzten oder unsere Feel Good Managerin wenden. In der Regel lässt sich immer eine Lösung finden.

Was könnte man noch besser machen?

In der Standardisierung von Prozessen gibt es meiner Meinung nach Nachholbedarf. Es ist super, dass jeder absolut selbstständig arbeiten und sich frei entfalten kann. Dies sorgt aber auch für einen hohen Abstimmungsaufwand und Reibungsverluste. Deswegen haben wir unseren Onboarding-Prozess vereinheitlicht, was gerade

für unser Back-Office und für uns im Recruiting eine große Erleichterung bedeutet. Wir könnten als MT AG auch selbstbewusster in unserer Kommunikation werden. Die MT AG ist meiner Meinung nach ein großartiges Unternehmen, auf das wir stolz sein können und das aufgrund seiner Werteorientierung den Mitarbeitern viele Möglichkeiten bietet. Dies sollten wir mutig und auch deutlicher kommunizieren.

Warum sollte sich ein zukünftiger Bewerber für die MT AG und nicht für eine andere Consulting-Firma oder einen anderen IT-Dienstleister entscheiden?

Wegen der großen Freiheiten, die er bei uns hat. Bei uns kann man sich kreativ einbringen und, wenn man die entsprechende Eigeninitiative mitbringt, richtig was bewegen. Man erhält dabei viel Unterstützung und hat große Chancen. Das macht einfach Spaß! Engagement, Motivation und Leistung werden gesehen und gewürdigt. Weiterbildungen, Trainings und fachliche Entwicklung werden (finanziell) positiv bewertet und gefördert.

Wir sind groß genug, um unseren Mitarbeitern umfassende Möglichkeiten und vielfältige Projekte mit modernen Technologien in einem professionellen Umfeld zu ermöglichen. Gleichzeitig sind wir inhabergeführt und legen Wert auf eine familiäre Wohlfühl-Atmosphäre, in der man respektvoll und wertschätzend miteinander umgeht. Dabei hat die Familie bei der MT AG einen besonderen Stellenwert. Wir unterstützen Eltern mit Kindern, wo es geht, damit sie es schaffen, Privatleben und Beruf – vor allem in der aktuellen Pandemie – unter einen Hut zu bekommen: mit Teilzeit-Arbeit, Sabbaticals, unbezahltem Urlaub, Kita-Zuschüssen etc.

Ebenso erwähnen sollte man, was bei uns mit „flachen Hierarchien“ gemeint ist: Wir haben nicht nur eine Duz-Kultur bis zum Vorstand, sondern es kann und soll tatsächlich jeder Mitarbeiter jederzeit an unseren Vorstand herantreten und Ideen und Vorschläge direkt mit dem oberen Management besprechen. Ich finde es beeindruckend, dass man teilweise das Gefühl hat, dass Hierarchiestufen gar nicht existieren. Ein Werkstudent ist genauso wichtig und wird ebenso ernst genommen und gefördert wie ein Senior Berater, Principal oder Bereichsleiter. Was zählt, ist, was man tut, erreicht und welches Engagement und welche Leistung man zeigt.

Welche Auswirkungen hat die Corona-Pandemie auf Ihre Arbeit? Was hat sich verändert?

Da wir als IT-Dienstleister bestens mit den digitalen Medien vertraut sind, war die Umstellung, von zu Hause aus zu arbeiten, nicht dramatisch und hat völlig reibungslos funktioniert. Die MT AG hat sich schnell und ohne großen Aufwand auf die neue Arbeitssituation einstellen können. Die höchst flexiblen Arbeitszeiten (auch Arbeit in Randzeiten ist möglich) sind ein Ausdruck unserer starken Ergebnisorientierung. Solange das Arbeitsergebnis stimmt, kann ich auch spät abends arbeiten, vorausgesetzt, man hält die Ruhepausen ein. Eltern können Urlaub nehmen, auch längeren unbezahlten Urlaub, sie können Überstunden abbauen. Anfangs haben Eltern ihre Kinder mit ins Büro gebracht. Als es den großen Lockdown gab, wurden allerdings alle ins Home-Office geschickt. Bis heute gilt bei uns: „Mobiles Arbeiten first.“ Ins Büro kommen wir nur, wenn es dringend notwendig ist, um die Ansteckungsgefahr auf ein Minimum zu reduzieren.

Des Weiteren starteten wir im April 2020 unter den Beschäftigten eine erste Umfrage zur Krisenkommunikation des Unternehmens. Wir erhielten sehr positives Feedback, und seitdem gibt es bei uns mindestens alle drei Monate Umfragen mit wechselnden Themen, die sich meist auf die aktuelle Arbeitssituation beziehen. So bleiben wir auch in diesen Zeiten in engem Kontakt mit unseren Kolleg*innen.

Bei der Umstellung von Präsenz- auf Remote-Arbeit via digitale Tools: Was hat gut geklappt? Was hat Ihnen eventuell nicht gefallen?

Es hat alles gut funktioniert. Wir nutzen verschiedenste Tools zur unternehmensweiten Kommunikation, unter anderen MS Teams und Yammer und für große Video-Sessions Zoom. Mitarbeiter konnten zudem Monitore, Tastaturen und andere Arbeitsmittel mit nach Hause nehmen, wurden mit Headsets ausgestattet und konnten, nach Bedarf, ergonomische Schreibtischstühle auf Firmenkosten bestellen etc. So wurde von Anfang an die Arbeitsfähigkeit am Arbeitsplatz zu Hause sichergestellt.

Wie gehen Sie und Ihre Firma mit dem Thema Nachhaltigkeit um? Spielt das eine Rolle?

Das spielt auf jeden Fall eine Rolle. Nachhaltigkeit sollte ganzheitlich begriffen werden. Aus unserer Sicht geht es im Kern darum, Verantwortung für das eigene Handeln zu übernehmen. Unsere Fürsorgepflicht gegen-

über unseren Mitarbeitern nehmen wir sehr ernst und wir sind uns auch unserer sozialen und gesellschaftlichen Verantwortung bewusst. Von der MT AG sind nicht nur unsere 270 Mitarbeiter abhängig, sondern auch deren Familien. Aus diesem Grund haben wir, wie oben erwähnt, zu Beginn der Pandemie unseren Mitarbeitern sofort die Arbeit von zu Hause aus ermöglicht. Zudem gibt es zahlreiche weitere Schutz-Maßnahmen, die wir als MT AG ergriffen haben. Erst kürzlich haben wir beispielsweise allen Mitarbeitern 30 FFP2-Masken für sie und ihre Familien zukommen lassen. Zusätzlich unterstützen wir verschiedene soziale Projekte, teilweise initiiert von unseren Mitarbeitern, und wir spenden für zahlreiche Umwelt-Projekte, siehe dazu die Website der MT AG.

Hat die Corona-Pandemie die Digitalisierung Ihrer Firma und Ihrer Kunden befördert?

Wir sind ein IT-Dienstleister, Digitalisierung ist unser Ding. Unsere Kunden hingegen scheint die Pandemie bezüglich Digitalisierung regelrecht beflügelt zu haben. Die Remote-Arbeit ist nun auch bei ihnen angekommen und wird vollständig akzeptiert. Gab es früher noch Unsicherheiten oder Vorbehalte, sind diese nun verschwunden. Das ist aus unserer Sicht eine sehr erfreuliche Entwicklung.

Frau Driescher, herzlichen Dank für das Interview.



Anne Driescher

Anne.Driescher@mt-ag.com

Als 32-jährige Fachbereichsleiterin Recruiting & Staffing beschäftigt sich die in Bochum lebende Anne Driescher vor allem damit, wie man die besten Mitarbeiter für die MT AG gewinnen und langfristig binden kann. Das Recruiting und seine Randthemen sind ein extrem dynamisches, spannendes und vielseitiges Feld, in das sie auch ihre Expertise aus dem Studium der Wirtschaft (Bachelor of Science) und der Wirtschaftspsychologie (Master of Science) einbringen kann. Sie ist gerne sportlich aktiv und verbringt viel Zeit in der Natur beim Joggen, Reiten oder Bouldern. Am liebsten ist sie in den Bergen: Hochgebirgstouren und Klettersteige in den Alpen sind für sie absolute Highlights im Urlaub.



Künstliche Intelligenz im Einsatz für Nachhaltigkeit – Prozessanalyse mit KI

Oliver Fuhrmann, MBA, Trevisto AG

KI-Lösungen für transparentere Produktionsprozesse und nachhaltigeres Wirtschaften – das ist die Aufgabe, die sich der Nürnberger IT-Dienstleister Trevisto zusammen mit dem Fraunhofer Institut für Integrierte Schaltungen IIS und Siemens Smart Infrastructure in einem Forschungsprojekt zur Digitalisierung der industriellen Produktion durch intelligente Prozessanalyse mithilfe künstlicher Intelligenz (KI) stellt.

Im Siemens-Gerätewerk in Amberg wollen die Partner im Projektvorhaben „AI-Nalyze“ durch den Einsatz von künstlicher Intelligenz mehr Transparenz in den Produktionsprozessen erzeugen und Steuerungsbedarfe sowie Verbesserungspotenziale identifizieren. Dabei soll eine automatische Aufnahme und Analyse von Prozessmodellen durch den Einsatz von KI erfolgen. Prozessschritte werden in Echtzeit identifiziert, und es werden produktionsrelevante Ereignisse abgeleitet – durch die Daten der in der Smart Factory existierenden (Automatisierungs-)Technologien.

Im Siemens-Gerätewerk in Amberg werden Produkte für industrielle Schalttechnik wie SIRIUS gefertigt. Diese Produkte bieten Maschinen und Anlagen Schutz gegen Überlast aus dem Netz. Im Falle eines Falls trennen Leistungsschalter das Equipment schnell vom Netz und verhindern so kostspielige Schäden. Darüber hinaus werden am Standort noch Leistungsschalter für industrielle Anwendungen, Infrastruktur und Gebäude entwickelt.

Die Analyse von Produktionsprozessen von Leistungsschaltern steht im Vordergrund des Projekts. Leistungsschalter schüt-

zen elektrische Einrichtungen zuverlässig bei Kurzschluss, Erdschluss oder Überlastfehlern. Leistungsschalter trennen je Variante bei Kurzschluss, Überstrom, Phasenausfall oder Phasenunsymmetrie den Verbraucher vom Netz. Das Auslöseverhalten des Leistungsschalters wird von vielen mechanischen und elektrischen Toleranzen der Teile und von Prozesswerten der Produktion beeinflusst. Um die von der Produktnorm vorgegebene Auslösezeit einzuhalten, wird der Leistungsschalter am Ende der Produktion justiert. Die Vorgaben des Justierprozesses werden über die Grenzstromprüfung über-

prüft und gegebenenfalls aktualisiert. In diesem Prozess können hohe Aufwände für die Justierung der Leistungsschalter entstehen. Erkenntnisse aus Daten, wie beispielsweise Daten zu Fertigungstoleranzen, Chargendaten oder Widerstandsdaten der Schweißung, ermöglichen es, relevante Faktoren für die Qualität der Produktion der Leistungsschalter zu identifizieren. Aufgabe ist das Erstellen eines neuen Algorithmus zum Abgleich der Leistungsschalter mit dem Ziel, Ausschüsse und Nacharbeitskosten nach der Grenzstromprüfung zu reduzieren. Mit diesem Ansatz können Predictive-Quality-Methoden entwickelt werden, die durch geringere Prüf- und Nacharbeitsaufwände zum einen Kosten einsparen und zum anderen dazu beitragen, Ressourcen-schonender und somit nachhaltiger zu produzieren.

Bewährtes Vorgehen führt zu vielversprechenden Ergebnissen

Im ersten Schritt des Entwicklungszyklus müssen Machine-Learning-Modelle mithilfe von für den Anwendungsfall relevanten Datensätzen aus dem Produktions- und Prüfprozess von Leistungsschaltern trainiert werden. Nach einer Datenbereinigung wird zunächst eine statistische Analyse der Daten durchgeführt, um die Abhängigkeiten und Muster zu finden. Lineare Korrelationen zwischen Zielvariable und Eingabevariablen konnten bereits aus einem ersten Datenausgang ermittelt werden. Zu den zu analysierenden Messwerten des Einzelgerätes zählen der Widerstand der Heizentwicklung, Kraft und Weg des Schaltschlusses sowie weitere Einflussfaktoren. Ziel der sich daran anschließenden Feature Selection ist es, die Dimension des Suchraums und den damit verbundenen Informationsverlust möglichst klein zu halten. Durch einen kleineren Suchraum kann das Modell einfacher, schneller und genauer die zugrunde liegenden Muster lernen und diese später generalisieren. Der Informationsverlust soll klein gehalten werden, damit das Modell trotz der Dimensionsreduktion alle für die Aufgabe nötigen Merkmale aus den Daten beziehen kann. Trevisto analysiert dafür Daten wie Grenzstrom und weitere Prüfdaten der Anlagen und ermittelt, welche Einflüsse das Auslöseverhalten des Leistungsschalters entscheidend beeinflussen und im Justieralgorithmus berücksichtigt werden müssen. Dazu werden möglichst viele Informationen in eine für den Algorithmus nutzbare Form gebracht. Bei der Modellbewertung werden parallel angelegte

Modelle miteinander verglichen, sodass das jeweils beste Modell für den entsprechenden Anwendungsfall ausgewählt werden kann. Abschließend folgen die Optimierung der Modellparameter und die Modellvalidierung. Die entwickelten Modelle und Algorithmen können schnell in der Produktion eingesetzt werden. Die Forschungsergebnisse sollen zeigen, dass durch den Einsatz von KI sowohl Effizienzsteigerungen als auch optimale Qualitätssicherung möglich sind. Die automatisierte Optimierung des Algorithmus über die Ergebnisse aus der Grenzstromprüfung ermöglicht eine laufende Anpassung an sich ändernde Bedingungen.

Gute Zusammenarbeit zwischen KI- und Fachexperten

Erfolgsfaktor für jedes KI-Projekt ist die gute Zusammenarbeit zwischen KI- und Fachexperten, denn bei der Identifikation der Parameter, die sich positiv auf die Nacharbeitskosten und Ausschüsse auswirken, hilft ein intensiver Austausch. Hier zeigt sich die Stärke in der Zusammenarbeit von KI-Experten mit Ingenieuren: Die Optimierung der einzelnen Bauteile können Fachexperten ableiten, wenn das Machine-Learning-Modell mit einer hohen Sicherheit sagt, welche Messwerte für die Vorhersage der Gesamtqualität relevant waren.

Fördermittelgeber und Förderprogramm BAYERN DIGITAL

Die Durchführung des Projekts wird gefördert im Rahmen des Programms Informations- und Kommunikationstechnologie durch das bayerische Staatsministerium für Wirtschaft, Landesentwicklung und Energie.

Ziel des Förderprogramms ist es, den Einsatz der Informations- und Kommunikationstechnik in neuen Produkten und Anwendungen zu beschleunigen und so die Wettbewerbsposition bayerischer Unternehmen weiter zu verbessern sowie neue Arbeitsplätze zu schaffen. Das Förderprogramm unterstützt den Aufbau von neuen und den Ausbau von bestehenden Kooperationen, in denen sich Unternehmen und Forschungseinrichtungen gemeinsam der anwendungsorientierten Forschung und Entwicklung widmen. Es werden im Themenfeld Digitalisierung Forschungs- und Entwicklungsvorhaben bayerischer Partner aus den unterschiedlichsten Anwendungsbranchen gefördert.

Im Rahmen des großen Maßnahmenpakets BAYERN DIGITAL werden von 2015 bis 2022 6 Milliarden Euro in die Digitalisierung Bayerns investiert.

Die Partner

Die Fraunhofer-Gesellschaft mit Sitz in Deutschland ist die weltweit führende Organisation für anwendungsorientierte Forschung. Unter ihrem Dach arbeiten 74 Institute und Forschungseinrichtungen an Standorten in ganz Deutschland. Rund 28.000 Mitarbeiterinnen und Mitarbeiter erzielen das jährliche Forschungsvolumen von 2,8 Milliarden Euro. Das Fraunhofer-Institut für Integrierte Schaltungen IIS mit Hauptsitz in Erlangen betreibt internationale Spitzenforschung für mikroelektronische und informationstechnische Systemlösungen und Dienstleistungen.

Die Siemens AG (Berlin und München) ist ein führender internationaler Technologiekonzern, der seit mehr als 165 Jahren für technische Leistungsfähigkeit, Innovation, Qualität, Zuverlässigkeit und Internationalität steht. Das Unternehmen ist in mehr als 200 Ländern aktiv und zwar schwerpunktmäßig auf den Gebieten Elektrifizierung, Automatisierung und Digitalisierung. Siemens ist weltweit einer der größten Hersteller energieeffizienter, ressourcenschonender Technologien, sowie einer der führenden Anbieter effizienter Energieerzeugungs- und Energieübertragungslösungen, Pionier bei Infrastrukturlösungen, ebenso bei Antriebs- und Softwarelösungen für die Industrie. Darüber hinaus ist das Unternehmen ein führender Anbieter bildgebender medizinischer Geräte wie Computertomographen und Magnetresonanztomographen sowie in der Labordiagnostik und klinischer IT.



Oliver Fuhrmann

oliver.fuhrmann@trevisto.de

Oliver Fuhrmann verantwortet bei der Trevisto die Ressorts Business Development und Marketing. Er ist seit über 20 Jahre in der ITK-Branche tätig und hat zahlreiche Projekte in der Industrie, im Handel und im Dienstleistungssektor geleitet. Seine Schwerpunkte liegen in den Bereichen Strategie- und Organisationsentwicklung, digitale Transformation von Geschäftsmodellen und Prozessen, Innovation und Service Development.

Interview mit Jens Horstmann, Vorstand Trevisto, über die Zusammenarbeit im KI-Projekt mit SIEMENS und Fraunhofer IIS

Marcos López, Redaktionsleitung Business News

Herr Horstmann, wie kam es zu diesem „Triumvirat der Zusammenarbeit“ von Siemens, Fraunhofer IIS und Trevisto?

Wir pflegen mit dem Fraunhofer IIS einen intensiven Austausch und haben eine langjährige Forschungspartnerschaft etabliert, die sich in einer Reihe von gemeinsamen Forschungsprojekten ausdrückt. Das vereinte Interesse an den Themen rund um Industrie 4.0 führte schließlich zur Smart Factory nach Amberg, in der die Siemens AG zusammen mit den Projektpartnern Analytics und Prognose zur Erforschung von KI-Lösungen betreibt. In unserer Zusammenarbeit geht es zunächst vornehmlich um KI-Lösungen zum prädiktiven Qualitätsmanagement in der Produktion.

Wie viele Fachexperten sind an diesem Projekt beteiligt? Geben Sie uns bitte einen Einblick...

Aus dem KI-Team der Trevisto sind in der Spitze bis zu vier Data Scientists beteiligt, hinzu kommen zwei Kollegen vom Fraunhofer Institut sowie je nach Spezifikation der Use Cases mehrere Siemens-Verantwortliche aus unterschiedlichen Fabrik-Prozessen.

Welche Rolle spielte Trevisto dabei? Was lief hervorragend, wo besteht Optimierungsbedarf?

Die zentrale Aufgabe der Trevisto ist das Aufsetzen der Systeme rund um KI, Zwischenspeichern der Daten und natürlich das Aufsetzen, Parametrisieren und Trainieren des KI-Frameworks. Also eigentlich der Kern des Projektes. Die Zulieferung der Daten seitens Siemens hat hervorragend funktioniert. Dies ist oft ein Knackpunkt in solchen Forschungsprojekten. Auch der Aufbau der KI-Infrastruktur läuft bisher sehr gut. Dabei haben wir noch keinen Optimierungsbedarf festgestellt. Insgesamt liegen wir sogar vor dem Zeitplan.

Wie sehen Sie das? Ist man sich in Deutschland über die Bedeutung von KI und IoT in der

gesellschaftlichen Entwicklung bewusst und einig?

Ich bin bisher der Meinung, dass hier zu wenig „probiert“ wird. Man ist sich der Möglichkeiten bewusst, aber leider traut man sich nicht, Geld in die Hand zu nehmen und zumindest die Möglichkeiten in sogenannten Proof of Concepts zu ermitteln. Deutschland vertraut hier mehr auf Hardware-Verbesserungen als auf Software-Optimierungen. Das ist auf der einen Seite nachvollziehbar, da wir gerade in Maschinenbau und Produktion viele Ingenieure im Einsatz haben, die eher Hardware-nah ausgebildet worden sind. In der Gesellschaft kommt KI eher durch die Hintertür und erleichtert uns meist das Leben, da ist die Akzeptanz eher hoch. Ich habe zumindest noch niemanden kennengelernt, der die Bildoptimierung im Mobiltelefon abschalten will, weil er KI ablehnt.

Was machen wir in Deutschland bezüglich KI gut, was machen andere Länder besser?

Bisher zeigt sich hier ein indifferentes Bild in der deutschen IT-Landschaft: Wir sehen wichtige Börsengänge deutscher IT-Unternehmen in diesem und im letzten Jahr, gleichzeitig droht der industrielle Mittelstand als Anwender den Anschluss zu verlieren. Hier gibt es deutlichen Nachholbedarf in der Umsetzung von Industrie-4.0-Konzepten. Im Bereich von Startups haben wir tolle Ideen, aber mit der konsequenten Finanzierung hapert es ziemlich. Dies ist in Amerika einfacher, hier wird Wachstum wichtiger gesehen als früher Gewinn. In China ist die staatliche Förderung für KI deutlich ausgeprägter.

Was nehmen Sie aus der Kooperation mit den Global Playern für zukünftige Projekte mit?

Bei einem so großen Player wie Siemens spielen globale Fragestellungen des nachhaltigen

Wirtschaftens eine große Rolle, in technischer Hinsicht zukünftig das Thema Cloud. Wir bei der Trevisto werden bei der Verbreitung von KI „nicht locker lassen“. Dies bedeutet für uns, stets am Ball zu bleiben und sich abzeichnende Entwicklungen im KI-Umfeld aufzunehmen und weiter voranzutreiben. Hier geht es uns insbesondere auch darum, KI-Anwendungen für den Mittelstand tauglich zu machen. Wir bringen die Entwicklung eines technologischen Stacks und fachlichen Vorgehensmodells für den schnellen, transparenten und kostengünstigen Einsatz von künstlicher Intelligenz im Mittelstand voran. KI hat sich seit einigen Jahren rasant weiterentwickelt. Bei KI denken viele nur an den Einsatz von Chatbots in Callcentern oder Assistenzsystemen in Autos. Neue und verbesserte Einsatzmöglichkeiten erweitern in den kommenden Jahren den Anwendungshorizont und machen sie branchenübergreifend zur Schlüsseltechnologie.

Herr Horstmann, herzlichen Dank für das Interview.



Jens Horstmann

jens.horstmann@trevisto.de

Jens Horstmann gründete gemeinsam mit Jürgen Engler 2009 die Trevisto AG. Zuvor war er unter anderem bei Oracle Deutschland tätig. Die Trevisto AG ist ein IT-Dienstleister mit Hauptsitz in Nürnberg. Mit seiner Expertise bietet das Unternehmen seinen Kunden Mehrwerte in den Bereichen Big Data, Business Intelligence, Analytics und Künstliche Intelligenz.

Wir begrüßen unsere neuen Mitglieder

Persönliche Mitglieder

- Alexandra Marcinova
- Aris Brauchart
- Lucie Weber
- Barbara Kadar
- Matthias Fillon
- Raphael Kodindo

Firmenmitglieder DOAG

- BWI GmbH,
Joachim Theurer
- Complion AG,
Stephanie Engelhard



Termine

Juni

06

14. - 15.06.2021

**Berliner Expertenseminar
mit Jürgen Sieben**
Online

16.06.2021

**DOAG Data Analytics Websession:
SQL Multi Data Lake Setup**
Online

17.06.2021

**DOAG Dev Talk zum Thema
"JavaScript in APEX?"**
Online

23. - 24.06.2021

**Berliner Expertenseminar
mit Randolph Eberle-Geist**
Online

Juli

07

01.07.2021

**DOAG Dev Talk zum Thema
"PL/SQL vs. JavaScript"**
Online

09.07.2021

**DOAG Datenbank Websession: Daten-
bankhärtung mit der Standard Edition 2**
Online

13. - 14.07.2021

**Berliner Expertenseminar
mit Dani Schnider**
Online

15.07.2021

**DOAG Dev Talk zum Thema
"Datenbankprogrammierung"**
Online

21.07.2021

**DOAG Data Analytics Websession:
Intelligente Prozessanalyse mit KI I**
Online

29.07.2021

**DOAG Dev Talk zum Thema
"Barrierefreiheit"**
Online

August

08

12.08.2021

**DOAG Dev Talk zum Thema
"1 Jahr Home-Office"**
Online

13.08.2021

DOAG Datenbank Websession
Online

18.08.2021

**DOAG Data Analytics Websession:
Intelligente Prozessanalyse mit KI II**
Online

26.08.2021

**DOAG Dev Talk zum Thema
"Wieviel Datenschutz geht?"**
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: Christian Luda

Redaktionsleitung Red Stack Magazin:

Martin Meyer

Redaktionsleitung Business News:

Marcos López

Kontakt: redaktion@doag.org

Weitere Redakteure (in alphabetischer

Reihenfolge): Denise Baidinger, Katharina

Büchse, Jörg Domaschka, Anne Driescher,

Markus Flechtner, Oliver Fuhrmann,

Fabian Hardt, Klaus-Michael Hatzinger,

Jens Horstmann, Thomas Karle,

Jan Karremans, Michael Kloker,

Lajos Lange, Marc Linster, Johannes Michler,

Steffen Moser, Jan Mütter, Borys Neselovskyi,

Thomas Nau, Andrej Pashchenko,

Bernd Patolla, Ana Paula Rivas,

Oliver Sennhauser, Wolfgang Taschner,

Dietmar Uhlig, Simon Volpert,

Jochen Wiefßler, Armin Wildenberg,

Martin Zablocki

Titel, Gestaltung und Satz:

Alexander Kermas

DOAG Dienstleistungen GmbH

(Anschrift s.o.)

Fotonachweis:

Titel: © mast3r | <http://stock.adobe.com>

S. 06: © maxicam | www.123rf.com

S. 07: © elenabsl | www.123rf.com

S. 10: © Elnur Amikishiyev | www.123rf.com

S. 16: © freepik | www.freepik.com

S. 26: © lankogal | <http://stock.adobe.com>

S. 30: © Python Logo | www.wikipedia.org

S. 35: © pratyaksa | www.123rf.com

S. 41: © Konstantin Mironov |

www.123rf.com

S. 48: © leedsn | www.123rf.com

S. 49: © lzflzf | www.123rf.com

S. 53: © rawpixel | www.123rf.com

S. 57: © Eric Eric | www.123rf.com

S. 61: © Oleksandra Naumenko |

www.123rf.com

S. 65: © Starline | www.freepik.com

S. 66: © wrightstudio | www.123rf.com

S. 72: © loganban | www.123rf.com

Titel S. 76: © OpenClipart-Vectors |

www.pixabay.com

S. 83: © Kheng Ho Toh | www.123rf.com

S. 85: © freepik | www.freepik.com

S. 87: © Anton Yankovyi | www.123rf.com

S. 90: © mohamed_hassan |

www.pixabay.com

S. 93: © gmast3r | www.123rf.com

Anzeigen:

DOAG Dienstleistungen GmbH

(verantwortlich, Anschrift s.o.)

Kontakt: sponsoring@doag.org

Mediadaten und Preise unter:

www.doag.org/go/mediadaten

Druck:

WIRmachenDRUCK GmbH,

www.wir-machen-druck.de

Alle Rechte vorbehalten. Jegliche Vervielfältigung oder Weiterverbreitung in jedem Medium als Ganzes oder in Teilen bedarf der schriftlichen Zustimmung des Verlags.

Die Informationen und Angaben in dieser Publikation wurden nach bestem Wissen und Gewissen recherchiert. Die Nutzung dieser Informationen und Angaben geschieht allein auf eigene Verantwortung. Eine Haftung für die Richtigkeit der Informationen und Angaben, insbesondere für die Anwendbarkeit im Einzelfall, wird nicht übernommen. Meinungen stellen die Ansichten der jeweiligen Autoren dar und geben nicht notwendigerweise die Ansicht der Herausgeber wieder.

Inserentenverzeichnis

B4Bmedia.net AG
<https://e-3.de>

U 3

MuniQsoft Consulting GmbH
www.muniqsoft-consulting.de

S. 33

Robotron
Datenbank-Software GmbH
www.robotron.de

U 2

dbi services sa
www.dbi-services.com

S. 9

MuniQsoft Training GmbH
www.muniqsoft-training.de

S. 3

DOAG e.V.
www.doag.org

U 4

PROMATIS software GmbH
www.promatis.de

S. 81

Alles, was die SAP-Community wissen muss,
finden Sie monatlich im E-3 Magazin.

Ihr Wissensvorsprung im Web, social media
sowie PDF und Print: e-3.de/abo

Wer nichts
weiß,
muss alles
glauben!

Marie von Ebner-Eschenbach



Ohhhhh! **Must Have**

Jetzt das **E-3 Magazin** abonnieren mit
dem Promo Code „rs21“
und kostenfrei fünf Ausgaben erhalten,
keine automatische Verlängerung.

 e-3.de/abo

www.e-3.de



2021 DOAG

Konferenz + Ausstellung

16. - 18. November
ONLINE

EARLY BIRD
BIS ZUM
30. SEPT.

2021.doag.org

Eventpartner: **AOUG**
AUSTRIAN ORACLE USER GROUP

SOUG
swiss oracle
user group

iJUG
Verbund

