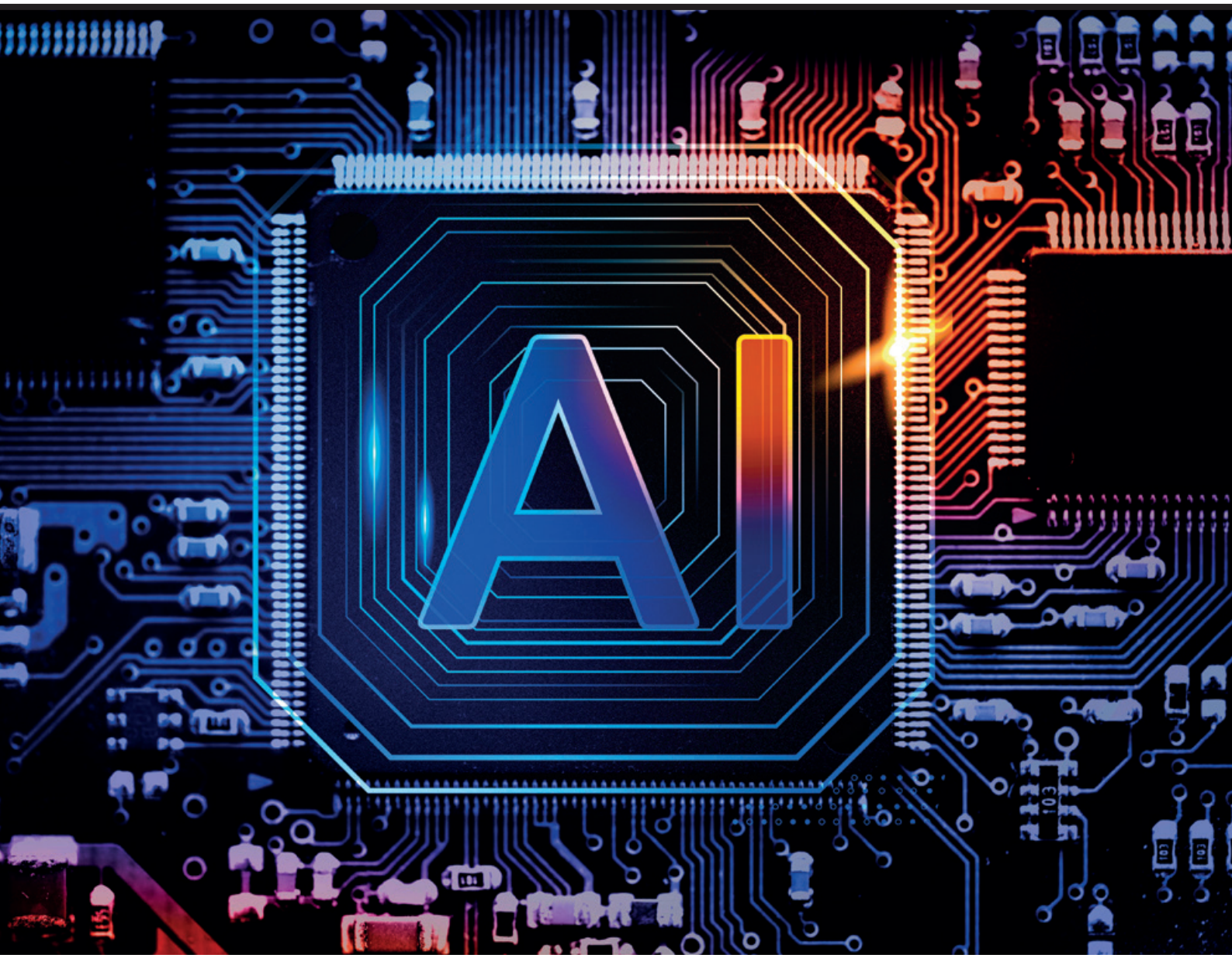
Nr. 4 | Juni 2024 | ISSN 2366-7915 | www.doag.org

Red Stack Magazin

DOAG SOUG swiss oracle user group AOUG AUSTRIAN ORACLE USER GROUP



KI & DIGITALISIERUNG

Aus der Praxis

Wie die KI das Planen für uns übernahm

Im Interview

Rainer Brehm, CEO Factory Automation, Siemens Digital Industries

KI

Warum wir unser Gehirn nun noch dringender brauchen

DIE DOAG ANWENDERKONFERENZ

ANWENDERKONFERENZ.DOAG.ORG

Save the Date

19. BIS 22. NOVEMBER

IN NÜRNBERG

2024

DOAG

Konferenz + Ausstellung

Oliver Szymanski
DOAG Vorstand KI, iJUG, JUG Nürnberg

Ausgabe Nr. 4/2024 auf Abruf!

Liebe Mitglieder, liebe Leserinnen und Leser,

Künstliche Intelligenz (KI) ist mit ungeheurer Wucht in unsere Welt gerückt. Doch statt revolutionär zu sein, wird bereits seit Jahrhunderten daran geforscht, falls man die Spekulationen Roger Bacons (1214–1292) dazu zählen möchte.

Seit circa 1950 werden Modelle erarbeitet, um aus Daten zu lernen und Muster zu erkennen. KI wurde gerade durch Computerspiele populär, da die Gegner scheinbar KI-gesteuert waren. Obwohl dies meist einfache Skripte ohne KI waren, wurde die Vorstellungskraft beflügelt.

Steve Grand ist britischer Wissenschaftler, bekannt für seine innovativen Ansätze in KI und Simulationen. Mit dem Spiel „Creatures" veröffentlichte er 1996 einen Meilenstein, welches die üblichen Grenzen durchbrach und virtuelle Wesen mit Lern- und Anpassungsprozessen simuliert. Darin „leben" kleine Kreaturen (Norns) mit simulierter DNA und neuronalen Netzwerken. Sie lernen aus Erfahrungen und entwickeln sich. Spieler füttern/erziehen sie, doch meist sehen sie ihnen beim Erkunden und Interagieren mit anderen Wesen zu.

„Creatures" zeigte erstmals den Einsatz von neuronalen Netzwerken und genetische Algorithmen in einem Spiel. Es simulierte komplexe Verhaltensweisen und Lernprozesse und beeindruckte damit die Spieleindustrie und wissenschaftliche Gemeinschaft. Steve Grand wurde 2000 dafür zum Officer of the Order of the British Empire ernannt. NESTA würdigte ihn 2002 mit folgenden Worten: „Steve has been called Britain's most intelligent man, as well as one of the 18 scientists most likely to revolutionize our lives during the 21st century."

Grand erschuf auch den Roboter Lucy, der wie ein menschliches Kind lernen kann, und schrieb darüber sein Buch „Growing up with Lucy" (2004).

Heute ist die Anwendung von KI über Computerspiele hinausgewachsen. KI wird in der Wirtschaft eingesetzt, um Prozesse zu optimieren, Vorhersagen zu treffen und Kundenbedürfnisse besser zu verstehen. In der Gesellschaft beeinflusst KI Bereiche wie Gesundheitswesen, Bildung und Kommunikation.

Dieses Red Stack Magazin mit dem Titelthema KI gibt spannende Einblicke in die aktuelle Nutzung von KI. Insbesondere die Interviews mit Siemens und Datev zeigen, wie KI heute in der Praxis angewendet wird und welches Potenzial sie für die Zukunft birgt.

Viel Spaß beim Lesen!

Oliver Szymanski

DOAG
WEBSESSION

Die DOAG WebSessions* bieten Ihnen in regelmäßigen Abständen spannende Online-Vorträge und -Diskussionen zu einer Vielzahl von Themenbereichen aus den jeweiligen DOAG Communities.

Freuen Sie sich auf WebSessions rund um die Themen Datenbank, Data Analytics und NetSuite oder beteiligen Sie sich bei den DOAG DevTalks an interessanten Gesprächsrunden zu aktuellen Development-Themen!

www.doag.org/go/websessions

*Die Buchung der WebSessions erfolgt ganz einfach über unseren Shop. Mitglieder erhalten im Buchungsprozess automatisch 100 % Rabatt.

Einleitung

3 Editorial

6 Timeline

7 „KI wird uns das Arbeitsleben erleichtern, Aufgaben verändern und neue schaffen, ohne den Menschen zu ersetzen."
Interview mit Rainer Brehm

KI

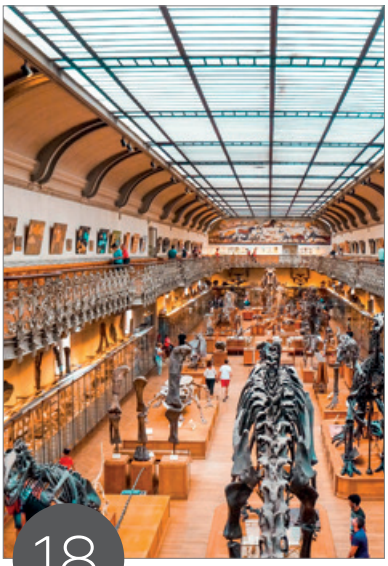
10 KI – Warum wir unser Gehirn nun noch dringender brauchen
Martin Talmeier

18 KI-Wertschöpfung in Unternehmen: Evolution statt Revolution
Dr. Konstantin Hopf

22 „Daher ist es essenziell, das Zustandekommen der Ergebnisse nachvollziehen zu können und diese noch einmal ‚von Hand' zu prüfen."
Interview mit Dr. Robert Helbig

26 Wie die KI das Planen für uns übernahm
Dr. Andreas Spies

KI – Warum wir unser Gehirn nun noch dringender brauchen



KI-Wertschöpfung in Unternehmen: Evolution statt Revolution

Interview mit Dr. Robert Helbig

Business Solutions

34 KI in den Oracle Applications – ein Zwischenbericht
Katharina Schraft

40 KI als Teamsport mit KI.NRW
Interview mit Dr. Christian Temath

Datenbank

43 Oracle Database 23c – bigger, better, faster, stronger, Teil 2: Die Revolution
Matthias Schulz

Development

50 Typische Fehler des Optimizers
Lothar Flatz

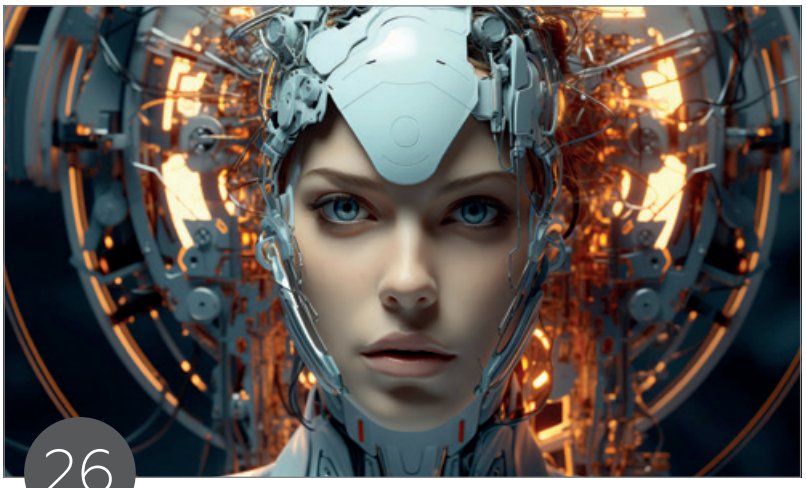
DevOps

54 DWH-Ops – the new normal
Dr.-Ing. Harald Philipp Gerhards

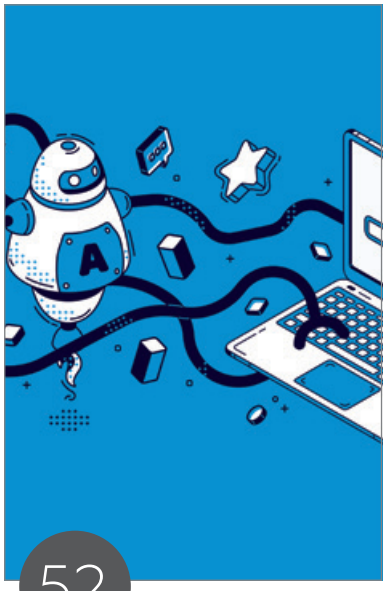
PL/SQL

61 PL/SQL-Laufzeitoptimierung durch Parallelisierung
Jan Gorkow

Interview mit Dr. Christian Temath



Wie die KI das Planen für uns übernahm



KI in den Oracle Applications – ein Zwischenbericht

PL/SQL-Laufzeitoptimierung durch Parallelisierung

News

53 Oracle Datenbanken Monthly News

60 Berliner Experimentseminare

Intern

68 APEX connect 2024 – eine Eins mit Sternchen für die Jubiläumsausgabe

72 Best of DOAG Online

73 Neue Mitglieder + Termine

74 Impressum + Inserenten

TIMELINE

22. bis 24. Februar 2024

Die APEX connect 2024 begrüßt mit neuem Konzept, bestehend aus dem Bootcamp & Solutions Day sowie zwei klassischen Konferenztagen, 341 Teilnehmende. Sie dürfen sich auf rund 77 Sessions freuen, die zu einem Viertel in englischer Sprache angeboten werden. Neben den zwei Keynotes aus Finnland und den USA stehen sechs DevTalks und zahlreiche 1-zu-1-Sessions mit Experten aus dem Oracle-APEX-Entwicklerteam auf der Agenda.

25. bis 27. April 2024

Das DOAG Leitungskräfteforum und die Delegiertenversammlung finden in Berlin statt. Es wird eine Satzungsänderung zur Neustrukturierung der Gruppen in der DOAG beschlossen. Diese sieht vor, dass die bestehende Struktur der Regionalgruppen für weitere Gruppen geöffnet werden soll, um damit die Zusammenarbeit in der DOAG zu fördern und zu erhöhen.

7. und 8. Mai 2024

Im Berliner Expertenseminar mit Marco Patzwahl geht es zwei Tage lang um das Thema „Oracle Reorg + Wartung“.

15. und 16. Mai 2024

Die DOAG 2024 Datenbank mit Exaday findet mit den Konferenz-Themen rund um Datenbanken und Engineered Systems vom 15. bis 16. Mai 2024 erneut im Düsseldorfer Airporthotel Van der Valk statt. Shasank Chavan, Vice President der Data, In-Memory and AI Technologies Group bei Oracle hält die Keynote.

16. Mai 2024

Im DevTalk mit Maik Becker geht es diesmal um eine APEX connect Review.

23. Mai 2024

Niels de Bruijn und Timo Herwix widmen sich im DOAG DevTalk dem Thema „APEX & ESG“.

4. Juni 2024

Das Regionaltreffen München/Südbayern findet in München statt. Zwei Themen stehen auf der Agenda: „APEX-Item-Plugin-Entwicklung für Formulare und Interactive Grids“ mit Referent Michael Schmid und „Mehrsprachige Anwendungen mit APEX – ganz einfach?“ mit Referentin Dr. Gudrun Pabst.

6. Juni 2024

Im DOAG DevTalk mit Oliver Lemm dreht sich alles rund um das Thema „APEX Office print“.

12. Juni 2024

Das Regionaltreffen Freiburg findet in Freiburg statt. Mit Referent Markus Schmidt wird die Oracle Cloud Infrastructure (OCI) vorgestellt.

„KI wird uns das Arbeitsleben erleichtern, Aufgaben verändern und neue schaffen, ohne den Menschen zu ersetzen.“

Martin Meyer, Redaktionsleiter des Red Stack Magazin, sprach mit Rainer Brehm, CEO Factory Automation, Siemens Digital Industries, über den Einsatz und die Entwicklung von KI-Technologie bei Siemens, Ethik und Datenschutz, den Einfluss auf Mitarbeiterstrukturen, die KI-Technologiestrategie bei Siemens sowie zu erwartende Trends.

Bitte stellen Sie sich unseren Lesern vor. Wer sind Sie und was sind Ihre Aufgaben bei der Siemens AG?

Ich bin seit Januar 2020 CEO der Siemens Business Unit Factory Automation. Gemeinsam mit rund 14.000 Mitarbeitenden kümmere ich mich um das Automatisierungsgeschäft von Siemens. Unter anderem bringen wir die zunehmende Integration von IT in die Automatisierung voran. Dabei stehen aktuell Technologien wie Cloud und Edge Computing oder KI im Vordergrund. Mit ihrer Hilfe können Industriekunden immer flexibler, effizienter und nachhaltiger produzieren sowie dem zunehmenden Fachkräftemangel begegnen.

Was war ihr persönlicher Wow-Effekt bezüglich KI und warum?

Mich faszinieren besonders die ersten Roboter mit flexiblen Greifern, die ohne Training den Greifpunkt bestimmen, komplexe Objekte erfassen und diese optimal platzieren. Mit dieser neuen Funktionalität sind Fertigungen viel anpassungsfähiger. Auch in der alltäglichen Arbeit gibt es viele Situationen, in denen mich KI beeindruckt. Neulich drehten wir auf Deutsch eine Grußbotschaft, die sich an alle unsere Mitarbeitenden richtete. Das Video wäre nur von einem Bruchteil der Kollegen und Kolleginnen verstanden worden. Deshalb hatte ich die Idee, den Ton in ein generatives KI-Tool einzuspeisen. Mithilfe des Tools konnten wir meine Originalstimme und selbst meine Lippenbewegungen in fünf weitere Sprache übersetzen. Im Video spreche ich beispielsweise fließend portugiesisch, obwohl mein Portugiesisch – trotz zwei Jahren in Brasilien – nicht mal annähernd so gut ist wie im KI-generierten Video.

Wie sieht Ihre Vision für den Einsatz von KI bei Siemens aus und wie trägt sie zur Erfüllung der Mission bei?

KI ist ein echter technologischer Durchbruch und wird wie das Internet vieles auf den Kopf stellen. Ich sehe den Einsatz von KI als einen Riesenhebel für unsere Kunden, für Siemens und als große Chance für den Industriestandort. Denn wie in vielen anderen Bereichen auch fehlen in der Industrie die hochqualifizierten Fachkräfte. Mithilfe von KI können Kolleginnen und Kollegen mit weniger Erfahrung oder unter Zeitdruck Fertigungen umbauen und Roboter einrichten. Sie können mit einer Maschine sprechen, als wäre sie ein Mensch. Dafür müssen wir nicht einmal eine spezielle Programmiersprache lernen. Indem wir KI-Anwendungen in die Produktionsprozesse integrieren, erhöhen wir die Produktivität. Wir können so mit weniger Ressourcen mehr erreichen und sichern damit unsere Wettbewerbsfähigkeit als Unternehmen.

Bei Siemens arbeiten wir bereits seit den 1970er Jahren an KI. Aber durch generative KI-Systeme bieten sich uns noch mal ganz andere Möglichkeiten. Generative KI ist der Anfang einer neuen Wachstumskurve – ein Supercharger für die digitale Transformation und Bewältigung des Klimawandels. Diese Technologie richtig einzusetzen, das ist die große Chance für unsere und kommende Generationen. Und es ist eine Riesenchance für Siemens.

Welche konkreten Anwendungsfälle sehen Sie für KI in Bezug auf ihre Arbeit?

Generative KI wird immer mehr Teil unseres Lebens und des Lebens unserer Kinder werden. Berufe werden sich ändern und Tätigkeiten wandeln. Wir müssen uns eine Art „KI-First"-Denken angewöhnen. Bei allem, was wir tun, müssen wir uns fragen: Kann generative KI das für mich erledigen? Im Privaten wäre das dann vielleicht die Reiseplanung oder eine Finanzberatung. Innerhalb unserer Organisation müssen wir uns auch entsprechend anpassen. Mein Ziel ist es, so schnell wie möglich dahin zu kommen, dass wir den maximalen Nutzen für die Industrie aus generativer KI herausziehen können. Dazu haben wir gerade ein internes „GenAI Bootcamp" veranstaltet und anschließend Anwendungsfälle für uns definiert. Mich hat unglaublich beeindruckt, was da herauskam. Natürlich muss bei unserer Arbeit immer im Mittelpunkt stehen: Welches Kundenproblem können wir lösen? Es gibt große Unterschiede, für welche Problemstellungen unsere Kunden KI einsetzen wollen. Der Fokus bleibt dabei auf den Menschen, die in Fabriken arbeiten, und wie generative KI sie bestmöglich unterstützen kann. In diesem Bereich sind wir als Siemens stark – denn hier liegt unser Domain-Know-how.

Zusammen mit Microsoft haben wir daher letztes Jahr den Siemens Industrial Copilot entwickelt und mit Schaeffler zwei konkrete Anwendungsfälle vorgestellt: schnelle Erstellung von Code für speicherprogrammierbare Steuerungen im Engineering und Fehlerbehebung im Bereich Operations. Die Entwicklungen gehen aber rasant weiter: Auf der Hannover Messe Ende April hat Siemens als erstes Industrieunternehmen ein Produkt für generative KI im Engineering gezeigt. Der Industrial Copilot ist an das Engineering-Framework Totally Integrated Automation (TIA) Portal von Siemens über die offene API-Schnittstelle TIA Portal Openness angebunden und wird Kunden über die Businessplattform Siemens Xcelerator zur Verfügung stehen.

In der Qualitätskontrolle kann KI visuelle Prüfungen mit Kameras durchführen, um Fehler oder Mängel an Produkten zu erkennen. Dafür kooperiert Siemens mit Basler und MVTec – zwei Branchenführern in der maschinellen und industriellen Bildverarbeitung. Diese Partnerschaft ermöglicht unseren Kunden, maschinelle Bildverarbeitungsanwendungen direkt in ihre bestehende Automatisierungstechnik zu integrieren und verbessert die Qualitätskontrolle.

Wie stellen Sie sicher, dass der Einsatz von KI ethisch und transparent ist, insbesondere im Hinblick auf Datenschutz? Und welche Rolle spielt die Nachvollziehbarkeit von KI-basierten Entscheidungen?

KI ist da und wird nicht verschwinden. Es bieten sich enorme Chancen, aber auch Risiken. Wir müssen eine KI schaffen, die verantwortungsvoll, verständlich und vertrauenswürdig ist. Das können wir erreichen, indem wir die Anwender im Umgang mit KI trainieren, sodass sie verstehen, wie die KI arbeitet, und dann die Ergebnisse ethisch beurteilen können. Dafür müssen wir als Siemens in die Ausbildung unserer Mitarbeitenden investieren, ihnen die entsprechenden Angebote zur Verfügung stellen und sie dazu ermutigen, sich mit der Technik zu befassen.

Das Thema Datenschutz können wir uns am Beispiel des Siemens Industrial Copilot anschauen: Wir verwenden für den Industrial Copilot eine private Instanz und die Daten werden nicht zum Training von Modellen verwendet. Des Weiteren be-

sitzt das TIA Portal die Möglichkeit, besonders sensiblen Code zu markieren, die sogenannte IP Protection. Dieser Code wird nie für den Industrial Copilot zugänglich. Datasharing-Ansätze diskutieren wir immer in enger Abstimmung mit unseren Partnern und Kunden.

Der Industrial Copilot funktioniert nur für und mit Menschen. Menschen werden in Zukunft KI-generierte Ergebnisse häufiger überprüfen und anpassen müssen als dies in der Vergangenheit der Fall war. Man spricht vom „human in the loop". Außerdem wenden wir die neusten Automatismen aus der Forschung an. Und wo immer möglich, stellen wir die Referenz zu bestehenden Daten her.

Welche Auswirkungen wird der Einsatz von KI in Ihrem Unternehmen auf Mitarbeiterstrukturen und -Positionen haben?

KI ist mitten unter uns. KI-Technologie wird bleiben und sich durch alle Branchen ziehen. KI wird uns das Arbeitsleben erleichtern, Aufgaben verändern und neue schaffen, ohne den Menschen zu ersetzen: Sie entlastet bei Routineaufgaben und schafft so Raum für kreativere und komplexere Aufgaben. Umgekehrt heißt das aber auch: Wir müssen die Menschen vorbereiten, sie in Bezug auf digitale Kompetenzen schulen und Ängste auffangen. Nicht alle sind auf demselben Level. Da sind einerseits Kollegen und Kolleginnen, die für die neue Technologie brennen, sie sofort verinnerlichen, und andererseits diejenigen, die an Bekanntem und Bewährtem festhalten. Die einzelnen Gruppen müssen wir als Unternehmen entsprechend unterstützen, weiterbilden und motivieren. Wichtig ist es sicherzustellen, dass wir alle Mitarbeitenden und ihre unterschiedlichen Bedürfnisse auf diesem Weg mitnehmen. Nicht nur die von KI-Begeisterten, sondern auch die der Skeptiker.

Was sind die wichtigsten Ziele, die Sie mit Ihrer (KI)-Technologiestrategie erreichen wollen und warum?

Der Klimawandel, eine alternde Bevölkerung und Fachkräftemangel in zahlreichen Branchen – die Welt steht vor riesigen Herausforderungen. Um diese zu bewältigen, brauchen wir Technologien. Technologien, die uns und unseren Kunden dabei helfen, effizienter, flexibler und nachhaltiger zu werden. Technologien, mit denen uns diese Transformation gelingt. Technologien wie KI. Siemens hat es sich zur Aufgabe gemacht, KI in der industriellen Welt anzuwenden, nutzbar zu machen und richtig einzusetzen. Bei uns arbeiten daran mehr als 1.500 KI-Expertinnen und -Experten. Bei Patenten für maschinelles Lernen und KI sind wir auf Augenhöhe mit den größten Tech-Unternehmen weltweit. Entscheidend ist jetzt aber, dass wir die guten Ideen in Produkte umsetzen, in industrie-taugliche Lösungen, in echten Mehrwert für unsere Kunden. Darin besteht die große Chance – für uns alle.

Welche Trends sehen Sie und welche Entwicklungen erwarten Sie in Zukunft bei KI und Digitalisierung? Und wie halten Sie sich darüber auf dem Laufenden?

Das Potenzial von KI ist riesig. Wahrscheinlich sehen wir noch nicht mal die Spitze des Eisbergs, welche Möglichkeiten sich uns bieten werden. Und die Entwicklungssprünge bei dieser Schlüsseltechnologie sind enorm. Bei Siemens arbeiten wir seit längerem daran, wie und wo wir KI – insbesondere generative KI-Systeme – überall anwenden können. Um das volle Potenzial nutzen zu können, muss generative KI „industrial grade" sein, sprich für den industriellen Einsatz in der Produktentwicklung und der Produktion tauglich: Die Systeme müssen robust, sicher und zuverlässig arbeiten. Bis dahin müssen wir noch einige Hürden überwinden. Viele KI-Anwendungen sind für das breite Produktionsumfeld noch nicht ausgereift genug, denn sie benötigen noch zu viel „human in the loop". Unscharfe Ergebnisse oder Halluzinationen beschäftigen unsere Experten-Teams. Oder die Frage, wie wir KI-Anwendungen für den industriellen Einsatz skalieren.

KI wird unsere Arbeitsplätze verändern und hilft uns dabei, zu wachsen und uns weiterzuentwickeln. Im abgelaufenen Geschäftsjahr haben fast 40.000 unserer Mitarbeitenden an KI-Trainings teilgenommen. Dieses Thema betrifft uns alle, egal ob in der Fertigung oder im Büro. Ich persönlich habe den Vorteil, dass ich mich bereits durch meine tägliche Arbeit sehr intensiv mit unseren Innovationen und den neuesten technischen Errungenschaften beschäftigen darf.

RAINER BREHM

ist CEO der Siemens Business Unit Factory Automation. Gemeinsam mit rund 14.000 Mitarbeitenden treibt er das Automatisierungsgeschäft von Siemens voran. Er ist Thought Leader im Bereich der Digitalisierung und Automatisierung, wo wegen der großen globalen Veränderungen eine schnellere Transformation in Richtung von Software-Defined Automation und Künstlicher Intelligenz erforderlich ist. Mit ihrer Hilfe können Industriekunden immer flexibler produzieren und somit effizienter und nachhaltiger werden sowie auf den zunehmenden Fachkräftemangel reagieren. Für eine firmenübergreifende Zusammenarbeit dabei setzt Brehm sich auch als Vorsitzender des Fachverbands Automation im Zentralverband der Elektrotechnik und Elektronikindustrie (ZVEI) ein.

KI – Warum wir unser Gehirn nun noch dringender brauchen

Martin Talmeier, Hasso-Plattner-Institut

In einer Ära, in der KI Mensch und Maschine verbindet, ist es entscheidend, die Stärken und Grenzen beider zu erkennen, um Synergien zu schaffen und Schwächen zu vermeiden. Warum wir trotz dieser scheinbaren Erleichterung unser Gehirn nun dringender denn je brauchen, wird im Folgenden aufgezeigt.

Der Faktor Mensch

KI hat das Potenzial, unsere Welt auf unvorstellbare Weise zu verändern. Um dieses Potenzial voll auszuschöpfen, MÜSSEN wir Menschen unser Verständnis und unsere Fähigkeiten weiterentwickeln. Die Art und Weise, wie wir KI nutzen, wird bestimmen, in welche Richtung diese (noch) unvorstellbaren Veränderungen ihren Lauf nehmen werden, die durch KI möglich sind *(siehe Abbildung 1)*. Wird sie uns das Leben dauerhaft erleichtern und verbessern oder vielleicht doch zu immensen Problemen führen?

Wir müssen KI bewusst und reflektiert nutzen. Die Auseinandersetzung mit KI sollte nicht von unbegründeten Ängsten oder übertriebenen Hoffnungen geleitet sein, sondern von einer fundierten Perspektive auf das, was KI wirklich leisten kann und wo ihre Grenzen liegen.

Der Mensch: Alte Strategien in einer neuen Welt – Kompliziertes beherrschbar machen

Geformt durch Jahrtausende der Evolution, beeinflussen Überlebensstrategien aus Vorzeiten der menschlichen Evolution noch immer, wie wir die Welt um uns herum wahrnehmen und wie wir mit ihr interagieren. Trotz des technologischen Fortschritts sind viele unserer Reaktionen und Entscheidungen von diesen urzeitlichen Impulsen geleitet *(siehe Abbildung 2)*.

Wenn wir uns heute in einer unbekannten Umgebung befinden und ein unerwartetes Geräusch hören, ist die unmittelbare Reaktion oft Vorsicht und wir empfinden Angst. Obwohl die realen Bedrohungen in unserer modernen Welt meist anders aussehen , bleibt diese instinktive Reaktion bestehen. Diese Urinstinkte, mehr oder weniger umfangreich angefüttert mit Wissen und vor allem Lebenserfahrung, leiten uns auch heute noch bei den täglichen privaten und beruflichen Entscheidungen.

Die moderne Welt, mit all ihren technologischen Wundern und komplexen Herausforderungen, verlangt, dass wir uns über diese instinktiven Reaktionen erheben. Unsere Realität ist von massiven Veränderungsprozessen und Unsicherheiten geprägt, in der nun die neue Unbekannte „künstliche Intelligenz“ erscheint – eine Realität, die unsere Vorfahren sich kaum hätten vorstellen können. Und doch sind wir mit den uns verfügbaren Fähigkeiten in der Lage, mit solchen Situationen intellektuell umzugehen.

Allerdings grätscht hier wieder die Überlebensstrategie unserer Vorfahren hinein, die als oberstes Gebot den sparsamen Gebrauch der eigenen Ressourcen gesetzt hat. Grund ist die damals ständige Sorge um die nächste verfügbare Nahrungsquelle. Daher greifen wir auch heute im Umgang mit komplexen Situationen vor allem auf einfache Strategien zurück, die uns wenig Aufwand kosten. Das oberflächliche Betrachten von Situationen vermeidet ebenfalls, dass wir durch Probleme beunruhigt werden.

Strategie Nummer eins ist dabei: ignorieren. Überlegen Sie, wie oft Sie Entscheidungen aufschieben oder versuchen, einfach wegzusehen. Damals sehr energieeffizient, im heutigen, durchorganisierten Leben dagegen eine Taktik, die riskant ist .

Strategie Nummer zwei bei komplexen Themen ist das Vereinfachen. Wir reduzieren dabei die Situation auf den Teil, den wir verstehen. Aber genau das, was wir nicht sofort verstehen, ist der Teil, der uns in Schwierigkeiten bringen kann. Daher ist dies keine durchdachte Lösung und so auch nicht zu empfehlen.

Strategie Nummer drei ist das „Trial-and-Error“-Prinzip. Man probiert eine potenzielle Herangehensweise aus und schaut dann, wie und ob es funktioniert.

Strategie Nummer vier ist die Denkweise „In God we trust – all others must bring data“ („Wir vertrauen in Gott – alle anderen müssen Daten herbeischaffen“).

Die ersten beiden Strategien sind wie beschrieben nicht empfehlenswert. Aber wie sieht es mit drei und vier aus?

Strategie 3 des Ausprobierens klingt nach einem Albtraum für Unternehmen. In einer Zeit der Unsicherheit, der sich laufend verändernden Umstände und Situationen, scheint es schwer, eine scheinbar unsichere Strategie umzusetzen. Aber in einer solchen Zeit, in der eine langfristige Planung sich inzwischen oft schon während der Planungszeit selbst überlebt, ist das aber tatsächlich die Strategie, die am erfolgversprechendsten ist. Allerdings nur, wenn das schrittweise Ausprobieren methodisch korrekt genutzt wird, was be-

Abbildung 1: The dawn of artificial intelligence (Quelle: Martin Talmeier mit DALL-E)

deutet, dass jeder Schritt sofort hinterfragt und das Zwischenergebnis auf die richtige Richtung überprüft werden muss. Das erfolgt durch stetige Prüf- und Lernschleifen eines „Build – Measure – Learn", dem methodischen Grundsatz aller modernen, auf Agilität ausgerichteten Innovationsmethoden wie Design Thinking oder Lean Startup.

Strategie 4 „Absicherung durch Daten" klingt modern, durchdacht und sicher. Hier stellt sich aber die Frage, wie diese Daten entstanden sind und interpretiert wurden. Zwar arbeiten die damit beauftragten Personen sicher professionell, aber inwieweit die Datenlage objektiv ist, bleibt undurchsichtig. Denn bei diesen Personen können sich alte Instinkte einschleichen oder gar menschliche Bias einfließen,

Damit aber nicht genug. Die menschliche Denk- und Handlungsweise schafft uns weitere urzeitlich entstandene Lebensvereinfachungen, die „damals" richtig waren, uns heute aber in große Schwierigkeiten bringen können.

Vier Wissenszustände sind es, in denen wir uns täglich bewegen:

- Erster Wissenszustand: Dinge, von denen wir wissen, dass wir sie wissen. Das ist ein angenehmer und sicherer Zustand.
- Zweiter Wissenszustand: Dinge, von denen wir wissen, dass wir sie nicht wissen. Das ist ein beherrschbarer Zustand, da wir unsere eigene Unwissenheit zu einer Fragstellung kennen und dadurch richtig damit umgehen.
- Dritter Wissenszustand: Dinge, von denen wir bis dahin nicht wissen, dass wir sie nicht wissen. Ein unangenehmer Zustand in dem Augenblick, in dem uns das bewusst wird. Bis dahin sind wir zumindest nicht beunruhigt, aber eben auch nicht achtsam.
- Der vierte Wissenszustand ist das eigentliche Problem für uns: Dinge, von denen wir glauben, dass sie richtig sind, die aber in Wirklichkeit falsch sind. Und dieser Wissenszustand, in dem wir mit voller Überzeugung entscheiden, ist nicht so selten, wie man vielleicht vermutet.

Warum werden diese menschlichen Schwächen im Zusammenhang mit KI aufgeführt?

Im Umgang mit Maschinen vertrauen wir wegen der oben aufgeführten Urinstinkte schnell auf eine vermeintliche Unfehlbarkeit und bleiben dann bei unserer Überzeugung, die aus einem guten Gefühl heraus entstanden ist. Tatsächlich wären wir auch nur in den wenigsten Fällen in der Lage, einzuschätzen, ob eine Maschine, ein Bauwerk oder etwas anderes tatsächlich sicher und zuverlässig konstruiert ist.

Je mächtiger und lebensbestimmender die Technologien werden, mit denen wir unser Leben teilen, gestalten und durch

sie gestalten lassen, desto wichtiger wird es, sich über die eigenen Strategien im Klaren zu sein. So kann im passenden Augenblick ganz bewusst das eigene Gehirn eingesetzt werden und wir können die tägliche Leichtfertigkeit verlassen, das eigene Vorgehen, die eigene Überzeugung nochmals zu überdenken und vielleicht auch eine dritte Meinung einzuholen.

Bei Letzterem muss ich aber noch ein menschliches Entscheidungsphänomen präsentieren: Das von allen als richtig akzeptierte Allgemeinwissen.

Gebe ich Ihnen den Satz „Ende gut, alles …“ und frage dabei, wie der Satz endet, werden Sie, wie die überwältigende Mehrheit der anderen Befragten, mit Überzeugung „gut“ antworten. Die Richtigkeit wird Ihnen dabei ohne das geringste Zögern von allen Befragten bestätigt werden. Das funktioniert auch, wenn ich zu dem unvollständigen Satz keinerlei weitere Information oder Aufgabe hinzufüge. Wir „denken mit“ und „wissen“, welche Aufgabe hier gestellt ist und handeln überzeugt. Und alle sind sich einig und zufrieden.

Führe ich dieses Experiment vor einer Gruppe, beispielsweise bei einem Vortrag durch und schert einer der Teilnehmenden mit einer anderen Antwort aus, zum Beispiel mit „Ende gut, alles auf Anfang“, führt das zu allgemeiner Erheiterung und Überraschung.

Schnell wird dann in der Gruppe klar, wie sehr wir uns auch in der Gemeinschaft auf die anscheinend einzig richtige Antwort einigen, obwohl die gemeinsam akzeptierte Antwort bei ein wenig Nachdenken nicht so absolutistisch richtig ist, wie wir gemeinsam, ohne nachzudenken, es bis dahin empfunden hatten.

Die Maschine: Neue Strategien in einer alten Welt – Kompliziertes berechenbar machen

Die Funktionsweise von KI zu verstehen, bedeutet, einen Blick in das Herz einer Technologie zu werfen, die unsere spontane menschliche Entscheidungsfähigkeit noch mehr in Frage stellt und

Abbildung 2: Urzeit-Höhle (Quelle: Martin Talmeier mit DALL-E)

Abbildung 3: Roboter mit Lupe (Quelle: Martin Talmeier mit DALL-E)

gleichzeitig erweitert. Dabei ist es wichtig, ebenso zu verstehen, wie ähnlich KI im Vergleich zu uns ist, und wo auch bei der KI (noch) Grenzen bestehen, die wir kennen sollten, um ChatGPT und Co. richtig einzusetzen.

Die KI navigiert durch die digitale Welt mit einer Präzision und Effizienz, die menschliches Denken in vielen Bereichen übertrifft. Die Grundlage dafür bildet ein komplexes Netz aus Algorithmen und Datenmodellen. Insbesondere die sogenannten Large Language Models (LLMs) sind in der Lage, Sprache zu imitieren, Sätze zu generieren und zu verstehen, indem sie auf gigantische Datenmengen beispielsweise aus dem Internet zurückgreifen, die dort das menschliche Wissen widerspiegeln.

Die Funktionsweise von KI kann vereinfacht mit unserer eigenen Denkweise beschrieben werden.

Wenn der Mensch eine Entscheidung trifft, durchläuft er im Bruchteil einer Sekunde einen Entscheidungsbaum mit – je nach verfügbarer Zeit bis zur erforderlichen Entscheidung – mehr oder weniger Entscheidungsebenen. Ist eine sofortige Handlung erforderlich, da es um Leben und Tod zu gehen scheint, springen die Reflexe ein. Hier wird nicht nachgedacht, ob der gerade gehörte Knall eine Bedrohung oder nur eine Lärmbelästigung ist, sondern sofort reagiert und weggeduckt.

Besteht etwas mehr Zeit, dann wird der erste Reflex hinterfragt, auf Erfahrungen aus vergleichbaren Situationen zurückgegriffen, es werden Konsequenzen durchdacht und dann nach bestem Wissen und Gewissen die gefühlt beste Entscheidung getroffen.

Der Denkprozess wird beendet, wenn entweder ein zufriedenstellendes Ergebnis vorliegt oder der zwingende Handlungsaugenblick gekommen ist – oder die Lust verloren gegangen ist, noch weiter nachzudenken. Letzteres passiert durch den Urinstinkt, generell nicht mehr Energie zu verschwenden als unbedingt notwendig.

Eine KI ist vergleichbar aufgebaut. Auch hier durchläuft die potenzielle Lösung eine Entscheidungsebene nach der anderen. In jeder Ebene (Layer) wird anhand von Algorithmen und darin enthaltenen mathematischen Formeln die Wahrscheinlichkeit der Sinnhaftigkeit/Richtigkeit der jeweiligen Option errech-

net und nach dem Durchlauf der Layer als Antwort ausgegeben.

Da hier auch der Zufall eine Rolle spielt, ist nicht vorhersagbar, wie genau in den einzelnen Layern die Entscheidungen entstehen. Damit bleibt einem Entwickler bei der Beurteilung, ob ein Algorithmus funktioniert, nur der Blick auf das Ergebnis. Ist das Ergebnis das, was der Entwickler als richtig erwartet, dann scheint der Algorithmus zu funktionieren.

Wir haben unter unseren KI-Demonstratoren eine Maschine, die Bilder betrachten und zu diesen Bildern eingegebene Fragen beantworten soll.

Hier wird zum Beispiel das Bild einer Baseball-Spielszene gezeigt. Zu sehen ist der knieende Fänger mit seinem Handschuh und seinem Helm auf dem Home Plate mit dem hinter ihm stehenden Schläger, mit seinem Baseball-Bat in der Hand. Die Szene ist genau in dem Moment festgehalten, in dem der Ball gerade geschlagen wurde und schon wieder etwas vom Schläger entfernt davonfliegt. Durch die hohe Schlaggeschwin-

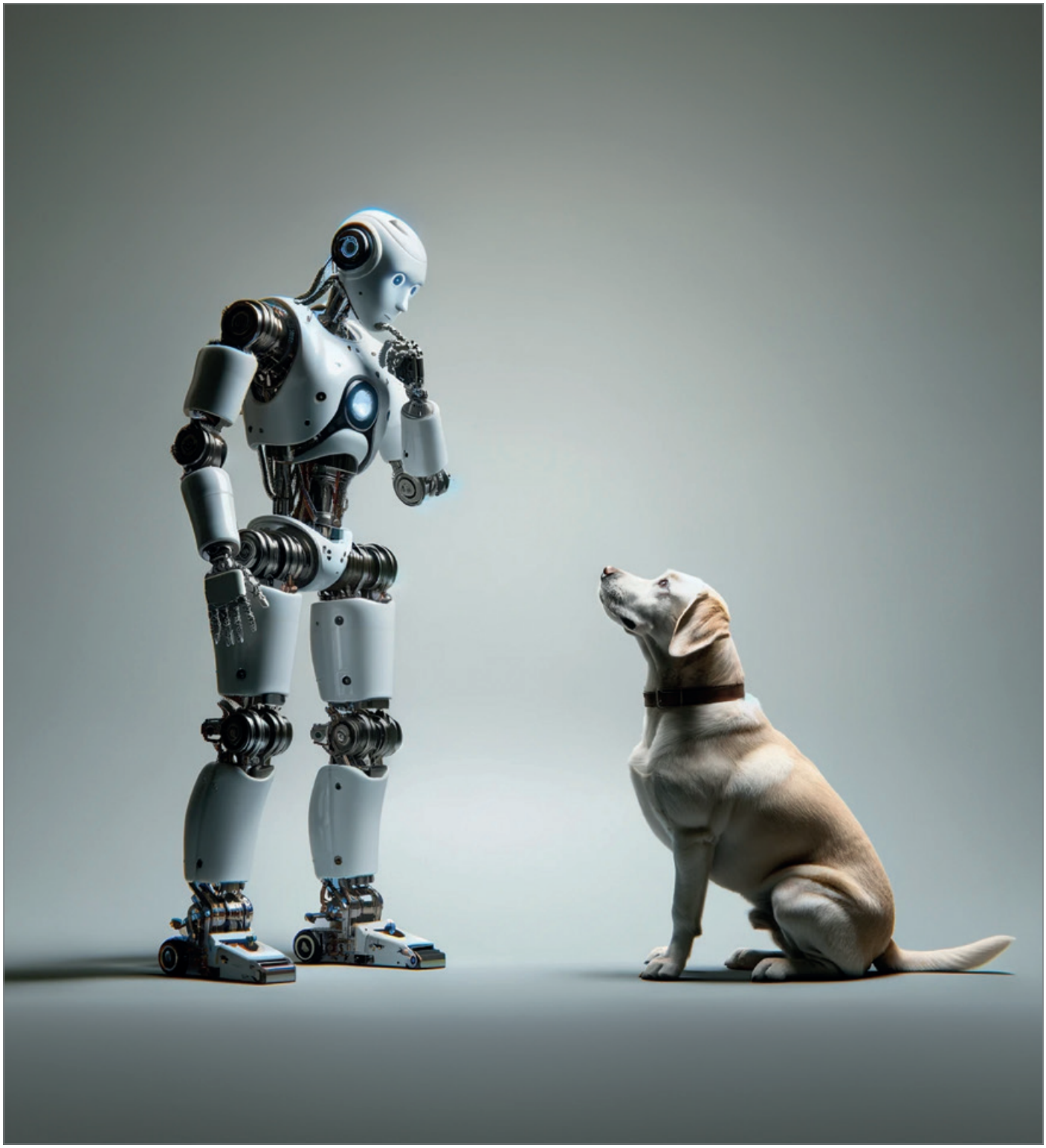


Abbildung 4: Roboter mit Hund (Quelle: Martin Talmeier mit DALL-E)

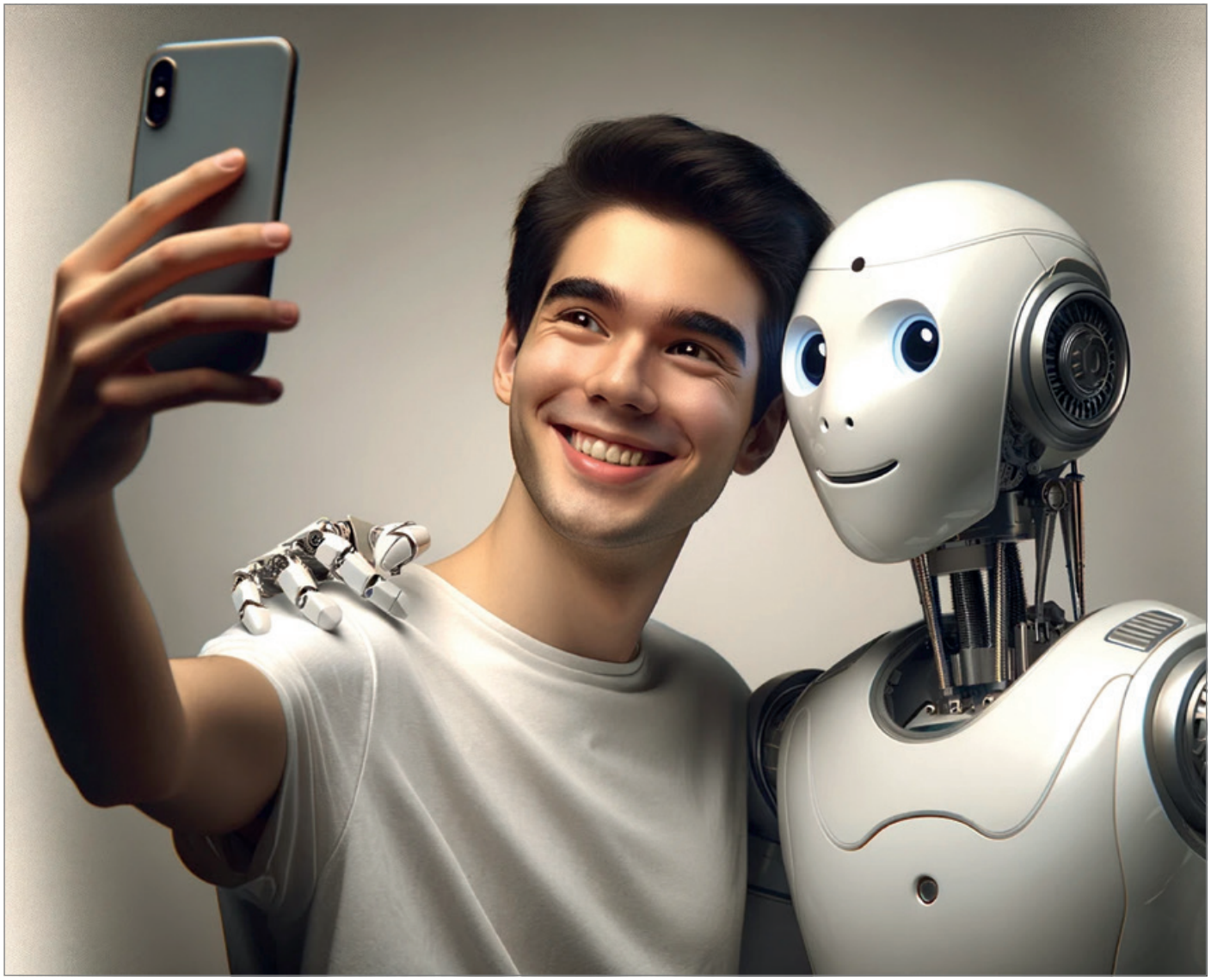


Abbildung 5: KI-Roboter im Einklang mit Mensch (Quelle: Martin Talmeier mit DALL-E)

digkeit ist beim Ball etwas Bewegungsunschärfe vorhanden.

Wird diese Maschine nun gefragt „Wo ist der Ball? ", kann man bei dem Demonstrator auch einsehen, welche Wahrscheinlichkeiten für die verschiedenen gefundenen Antwortvarianten die Maschine kurz vor Ausgabe der Antwort errechnet hat. Die Antwort mit der mit Abstand größten Wahrscheinlichkeit ist dabei „Der Ball ist in der Luft".

Das ist zum einen richtig und zum anderen beeindruckend, führt man sich die Zusammensetzung des Bildes vor Augen. Damit scheint der Algorithmus gut zu funktionieren.

Dieser Demonstrator bietet nun aber auch noch die Möglichkeit, sich anzeigen zu lassen, welche Stellen im Bild die Maschine für die Entscheidungsfindung herangezogen hat. Im Gegensatz zu uns Menschen kann eine Maschine ein Bild nicht in seiner Gesamtheit beurteilen, sondern „sieht" jeden einzelnen Pixel. Damit muss eine Bilderkennung das Bild nach Mustern durchsuchen und die Entscheidung treffen, welcher Bereich dann schlussendlich zur Erkennung ausgewählt wird. *(siehe Abbildung 3)*.

In diesem oben genannten Beispiel zeigt der Blick auf die Bereiche, die die Maschine für ihre Analyse herangezogen hat, dass sie sich nicht auf den Ball konzentriert hat, sondern auf den weißen Helm des Schlägers. Der Ball wurde gar nicht erkannt. Das Ergebnis klingt daher nur richtig, ist aber in Wirklichkeit falsch. Der Algorithmus ist somit entgegen der Vermutung alles andere als perfekt.

Komplexe KIs wie ChatGPT lassen solche Einblicke gar nicht mehr zu. Hier helfen nur intensive, lange Tests durch Anwender, die den Entwicklern zurückspielen, welche Fehler oder Ungereimtheiten sie entdeckt haben.

Womit wir bei den Besonderheiten von Sprachmodellen wären, deren frappierend menschenähnliche Form der Kommunikation und Interaktion uns Menschen in die nächste Schwierigkeit führt, die wir aber auch mithilfe unseres eingesetzten Gehirns meistern können.

Verhältnis Mensch – Maschine – „it's complicated"

ChatGPT, Gemini, Claude 2, PI, Llama, … all diese Modelle, um nur die bekanntesten zu nennen, kommunizieren durch die unfassbar große Datenmenge, die zum Erlernen der zwischenmenschlichen Kommunikation aus dem Internet heran-

gezogen wurde, genauso, wie ein Mensch mit uns kommunizieren würde. Da im Internet alle Varianten der dabei möglichen Tonalitäten, von freundlich, romantisch über formell, wissenschaftlich, umgangssprachlich, kindlich, jugendlich bis hin zu feindselig und hasserfüllt als Lernbeispiele unserer sprachlichen Vielfalt vorhanden sind und das in allen Sprachen, kann die Maschine gesteuert durch Algorithmen stets die höchste Wahrscheinlichkeit für das nächste passende Wort errechnen. Daraus ergeben sich fast perfekt formulierte Sätze.

Dies führt zu einer angeborenen emotionalen Reaktion beim menschlichen Nutzer. Wir bauen Sympathie auf. Es gibt keinen Grund, bei ChatGPT „bitte“ und „danke“ zu schreiben, die Mehrheit von uns tut es aber.

Schon das 1966 entwickelte Computerprogramm ELIZA des Informatikers Joseph Weizenbaum simulierte in einer der Variationen einen Psychologen. Im „Gespräch“ mit Patienten griff das Programm auf vorgegebene Antworten zurück, die sich an Stichworten der Patienten-Eingaben orientierten. Damit war es zwar keine künstliche Intelligenz, die Wirkung war aber überraschend. Die Patienten bevorzugten das Therapiegespräch mit der Maschine im Vergleich zu dem mit einem menschlichen Therapeuten und vertrauten der Maschine ihre intimsten Geheimnisse an.

ELIZA gilt als Prototyp für moderne ChatBots.

Und 60 Jahre später steht jedem von uns, der Zugang zu einem Computer mit Internetverbindung hat, ein viel leistungsfähigerer Gesprächspartner zur Verfügung.

Wie bei ELIZA schätzen wir die Fähigkeiten der Maschinen völlig falsch ein und gehen damit auch völlig falsch mit ihr um, beziehungsweise lassen uns durch die überzeugend vorgetragene Maschinenantwort verführen, die gesamte Denkarbeit der Maschine zu überlassen.

KI-Systeme verstehen keine Kausalität. Ihre „Entscheidungen“ basieren nicht auf einem Verständnis der Welt oder einem Bewusstsein ihrer selbst, sondern auf statistischen Wahrscheinlichkeiten und Korrelationen, also welches Wort wahrscheinlich auf ein anderes folgt.

Wenn wir „Hund“ lesen, dann haben wir Menschen sofort verschiedene Hunderassen vor Augen, empfinden je nach eigener Erfahrung Zuneigung oder Furcht. Wir wissen, wie unterschiedlich sich das Fell einer Langhaarrasse von dem einer Kurzhaarrasse anfühlt. Wir wissen, wie ein Hund riecht, was er aus uns macht, wenn er uns schwanzwedelnd ansieht oder zähnefletschend knurrt. Wir können uns sogar vorstellen, wie es schmecken und sich anfühlen könnte, wenn wir in das Fell beißen, was wahrscheinlich die wenigstens von uns je gemacht haben. Und dennoch können wir es uns leibhaft vorstellen *(siehe Abbildung 4)*.

Ein auf Sprachmodellen basierendes System versteht eigentlich Text, kann aber dank fortschrittlicher Algorithmen aus den Beziehungen in den Trainingsdaten menschlich wirkende, einfühlsame Aussagen kreieren. Es nutzt dabei eine reiche Sprache, um unsere komplexe Welt farbenfroh darzustellen, was uns manchmal fälschlicherweise glauben lässt, dass die Maschine genau versteht, worüber wir sprechen und was wir fühlen. Aber die Maschine fühlt und versteht nicht. Die Antworten der Maschinen sind das Ergebnis seelenloser statistischer Wahrscheinlichkeiten und Mustererkennung, nicht eines tiefgreifenden Verständnisses oder einer eigenständigen Reflexion.

Fazit

Die Herausforderung besteht darin, ein ausgewogenes Verhältnis zwischen der Anerkennung der beeindruckenden Leistungsfähigkeit von KI und einem realistischen Verständnis ihrer Begrenzungen zu finden. Wir müssen lernen, KI als das zu sehen, was sie ist: ein außergewöhnlich leistungsfähiges Werkzeug. Es ist in der Lage, bestimmte Aufgaben mit einer Effizienz und Präzision auszuführen, die menschliche Fähigkeiten übersteigt, das aber gleichzeitig von menschlicher Aufsicht, ethischen Überlegungen und einem tiefen Verständnis der zugrundeliegenden Technologie abhängig ist *(siehe Abbildung 5)*.

Über den Autor

Martin Talmeier ist beim Hasso-Plattner-Institut Lead-Coach und KI-Trainer im Projekt Mittelstand-Digital Zentrum Berlin. Er konzipiert und leitet dort Workshops und Trainings speziell für mittelständische Unternehmen. Unternehmen werden so dazu befähigt, ihre Position in einer sich schnell digitalisierenden Welt realistisch einzuschätzen, die Herausforderungen und Möglichkeiten der Digitalisierung zu beleuchten und die Potenziale durch Künstliche Intelligenz mutig auszuloten und sich damit auch in sehr unsicheren und agilen Zeiten nachhaltig zukunftssicher auszurichten.

Martin Talmeier
martin.talmeier@hpi.de

KI-Wertschöpfung in Unternehmen: Evolution statt Revolution

Dr. Konstantin Hopf, Otto-Friedrich-Universität Bamberg

Zur Entwicklung und der Umsetzung ihrer KI-Strategie können Unternehmen auf ein Prozessmodell zurückgreifen, das die Wertschöpfung von maschinellen Lernverfahren in Unternehmen erklärt. Das Modell zeigt Entwicklungspfade für KI-Initiativen auf und erlaubt einen agilen Blick auf die neue Technologie.

Die Fortschritte im Bereich der künstlichen Intelligenz (KI) verändern die Gesellschaft und die Arbeitswelt in einem atemberaubenden Tempo. Große Sprach- und Bildmodelle halten Einzug in den Alltag wissensintensiver Arbeitsbereiche wie Journalismus, Softwareentwicklung und viele mehr. Die KI-Werkzeuge können dort lästige Aufgaben abnehmen oder Inspiration bringen. Dass KI-Systeme selbst komplexe strategische Entscheidungen erfolgreich treffen können, zeigen uns die diversen Beispiele aus Online-Spielen bei denen KI-Agenten gegen ihre menschlichen Mitspieler gewinnen. Beeindruckend ist beispielsweise der Erfolg von Google's DeepMind im Spiel „Diplomacy" *[1]*. Hier demonstrierte das KI-System, dass es mit anderen (menschlichen) Spielern effektiv kommunizieren, Allianzen schmieden und diese aufgebauten Beziehungen dann für die eigene Agenda ausnutzen kann.

Viele Unternehmen befinden sich derzeit auf der Suche nach einer geeigneten strategischen Antwort auf diese Entwicklungen und es herrscht Unklarheit darüber, wie man greifbaren Geschäftswert aus KI-Anwendungen generiert oder KI effektiv in den Unternehmenskontext einbettet. Firmen engagieren sich in einer Vielzahl von KI-Projekten und experimentieren. Doch immer wieder wird berichtet, wie eine große Anzahl von KI-Projekten aufgrund von technischen und organisationalen Hürden scheitern.

Um dem KI-Engagement eine Orientierungshilfe zu geben, können Entscheiderinnen und Entscheider ein Prozessmodell verwenden, das die Wirtschaftsinformatikforschung kürzlich entwickelt hat. Das Modell, welches in *Abbildung 1* dargestellt ist, beschreibt drei grundlegende Mechanismen der betrieblichen Wertschöpfung mittels maschineller Lernverfahren (ML), eine der Kerntechnologien der KI. Es identifiziert Bedingungen die nötig sind, um einen entsprechenden Wertbeitrag im Unternehmen zu leisten.

Die Studie „*Shifting ML Value Creation Mechanisms: A process model of ML value creation*" erschien in der englischsprachigen Fachzeitschrift *The Journal of Strategic Information Systems* (Jahrgang 31, Ausgabe 3 im September 2022). Für diese Studie haben die Wirtschaftsinformatik-Professorin Arisa Shollo (Copenhagen Business School), Dr. Konstantin Hopf (Universität Bamberg), Dr. Tiemo Thiess (IT University of Copenhagen) sowie Prof. Oliver Müller (Universität Paderborn) im Zeitraum 2018 - 2022 zusammengearbeitet. Sie befragten Data Scientists und Manager zu ihrer Arbeit in 56 Projekten, in denen Anwendungen von maschinellen Lernverfahren in 29 Unternehmen entwickelt wurden.

Die Studie wurde als **bester Artikel des Jahrgangs** der Zeitschrift sowie als Bestes Papier der Wirtschaftsinformatik 2023 der internationalen Fachorganisation Association for Information Systems ausgezeichnet. Der Artikel kann kostenfrei (open access) unter *https://doi.org/10.1016/j.jsis.2022.101734* abgerufen werden *[2]*.

Die Mechanismen der Wertschöpfung sind entweder (i) die Schaffung von Wissen, (ii) die Unterstützung von Menschen bei der Entscheidungsfindung (mit Hilfe von ML-basierten Assistenzsystemen), wobei dem Einzelnen in Entscheidungssituationen eine Bandbreite von hohem bis niedrigem Ermessensspielraum bleibt, und (iii) autonome Agenten, die neue Angebote (also neue Produkte oder Dienstleistungen) oder effizientere Prozesse hervorbringen können. Jeder Mechanismus erfordert bestimmte Bedingungen, um erfolgreich eingesetzt zu werden, bringt aber auch einen einzigartigen Wertbeitrag im Unternehmen (zum Beispiel (i) Wissen, (ii) bessere Entscheidungen, (iii) neue Produkte oder effizientere Prozesse). Die systematische Gegenüberstellung der drei Arten der ML-Nutzung und der jeweiligen notwendigen Bedingungen hilft Unternehmen, ihre KI-Initiativen gezielter zu planen und zu überwachen.

Drei Mechanismen zur Schaffung von Unternehmenswert durch ML im Einzelnen

Der *erste Mechanismus* ist die Nutzung von ML zur *Wissensgenerierung.* Hier machen sich Unternehmen die Fähigkeit von ML-Algorithmen zunutze, Muster in Daten zu erkennen. Aus diesen Mustern können neue Erkenntnisse über das Geschäftsfeld, das Kundenverhalten und vieles mehr abgeleitet werden. Für diese Art der Wertschöpfung benötigen die Unternehmen datenwissenschaftliche Fähigkeiten und ein gutes Verständnis ihres Geschäfts.

Die *zweite Art* der ML-Nutzung sind *Assistenzsysteme*, die Menschen bei der Ausführung von Aufgaben unterstützen oder anleiten. Sie reichen von kleinen Erweiterungen bestehender Softwareprogramme bis hin zu spezialisierten Entscheidungsunterstützungssystemen und ermöglichen spezifische Erkenntnisse oder Handlungsempfehlungen. Beispielsweise die Schätzung einer Wahrscheinlichkeit, ob ein Kunde den Anbieter wechselt, oder welches Ersatzteil in einer Maschine ausgetauscht werden soll. Diese Einblicke können eher informativ sein oder sogar konkrete Handlungsanweisungen beinhalten, die Mitarbeiter später nur ausführen. So kann der Algorithmus gezielt den Arbeitsablauf beeinflussen

und effizienter gestalten. Da diese Art der ML-Wertschöpfung eine Integration in die Softwarelandschaft erfordert, brauchen Unternehmen auch eine gute Dateninfrastruktur, Kenntnisse über die Gestaltung von Benutzeroberflächen und Fähigkeiten, um ML-Modelle von der Entwicklung in den Betrieb zu überführen.

Die *dritte Art* der Wertschöpfung ist die der *autonomen Agenten*, bei der ML direkt in neue Produkte oder Dienstleistungen integriert wird, also neue Arten von Angeboten für Endkunden geschaffen werden, oder bei der ML-basierte Software nicht nur Teile von Geschäftsprozessen anleitet, sondern auch direkt ausführt. Um solche Einsatzszenarien zu realisieren, sind Unternehmen auch auf die Integration von ML-Anwendungen in betriebliche Prozesse und IT-Systeme, ein stabiles Geschäftsumfeld und wenige rechtliche und ethische Probleme mit ihren ML-Anwendungen angewiesen. Für den Einsatz von ML zur Automatisierung müssen die meisten Anforderungen erfüllt sein. Unter anderem muss der rechtliche und ethische Rahmen solche Anwendungen erlauben (vgl. „AI Act" der Europäischen Union). Die Vielzahl der notwenigen Bedingungen, die gegeben sein müssen, um autonome Agenten in der Praxis zu realisieren, ist eine mögliche Erklärung für das Scheitern von ML-Projekten.

Die drei Wertschöpfungsmechanismen unterscheiden sich nicht nur anhand ihrer Wertbeiträge, sondern auch anhand der verwendeten Technologien, wie *Abbildung 2* zeigt. Wo im Mechanismus der Wissensgenerierung (i) Menschen – vornehmlich Data Scientists – direkt mit dem ML-Modell in der Entwicklungsumgebung interagieren und Erkenntnisse ableiten, wird in Assistenzsystemen (ii) meist nur eine Vorhersage von Endnutzern verwendet. Wenn Entscheidungen durch ein System übernommen werden (Assistenzsysteme, die menschlichen Nutzern nur wenig Ermessensspielraum lassen oder autonome Agenten), wird neben der Vorhersage auch noch eine Entscheidungsfunktion benötigt, die spezifiziert wie die Vorhersage in eine Entscheidung(sempfehlung) übersetzt wird.

Agilität von ML-Anwendungen: Projekte wechseln die Art der Wertschöpfung über die Zeit

Durch eine eingehende Analyse der 56 ML-Projekte fand die Studie heraus, dass Unternehmen den Wertschöpfungsmechanismus von ML-Projekten im Laufe der Zeit wechseln. Dies wird notwendig, wenn Bedingungen nicht mehr erfüllt sind oder, wenn ML-Anwendungen weiterentwickelt werden und dabei zusätzliche Bedingungen erfüllt werden. Diese „Shifting"-Praxis lässt sich anhand einer ML-Anwendung bei einem Online-Einzelhändler verdeutlichen, der seine vollautomatische Anzeigenschaltung in Online-Medien zu Beginn der Covid-19-Pandemie aufgrund der schlechten Werbeleistung einstellen musste. Nachdem das Unternehmen im Anschluss menschliche Entscheidungen über Anzeigenplatzierung getroffen hatte, begann es, ML zu nutzen, um Wissen über die neue Situation zu sammeln. Mit dem gewonnenen Wissen wurde dann ein Dashboard zur Unterstützung der routinemäßigen menschlichen Entscheidungsfindung erstellt und schließlich trainierten die Entwickler neue Vorhersagemodelle, die auf die Pandemiesituation zugeschnitten waren, und kehrten schließlich zum alten Modus der ML-basierten Prozessautomatisierung zurück. Das Wissen, dass sich ML-Wertschöpfungsmechanismen über die Zeit ändern können, erlaubt Unternehmen eine Entwicklungsperspektive für ihre KI-Initiativen einzunehmen, anstelle einzelne Projekte zu früh als ‚gescheitert' anzusehen, deren angedachter Wert nicht erfüllt werden konnte.

Schlussendlich weist die Studie auf die zentrale Rolle von Data Scientists in der Wertschöpfung durch ML hin. Die Auto-

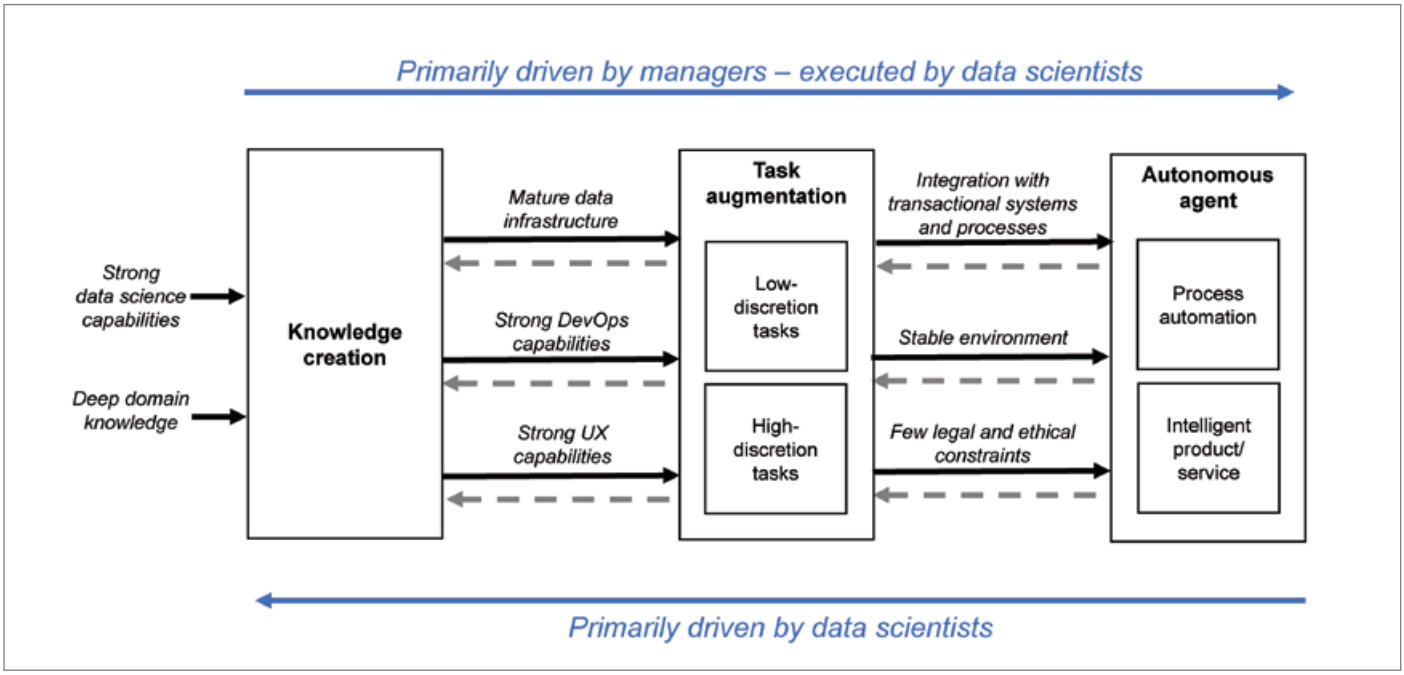


Abbildung 1: ML-Wertschöpfungsmechanismen und ihre notwendigen, aber nicht hinreichenden Bedingungen: Ein Prozessmodell der ML-Wertschöpfung in Organisationen. (Quelle: creative commons)

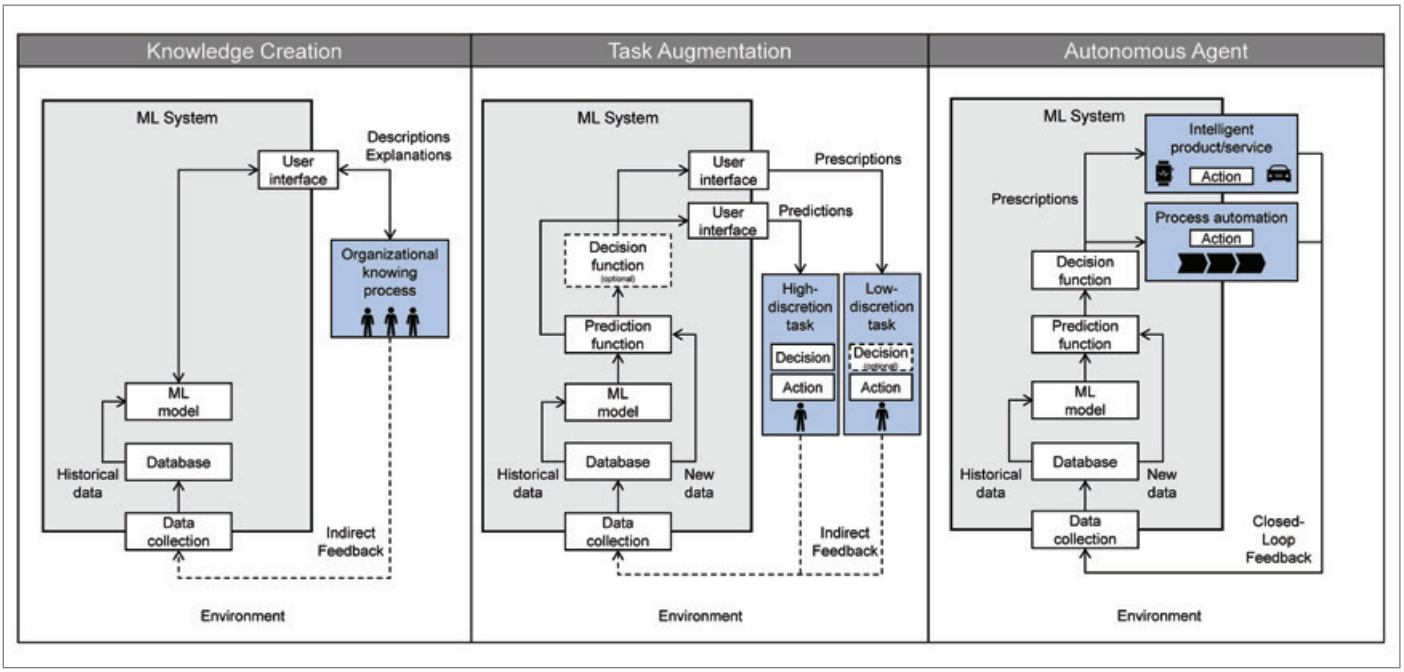


Abbildung 2: Gemeinsamkeiten und Unterschiede zwischen den ML-Wertschöpfungsmechanismen (Quelle: creative commons)

ren verdeutlichen, dass Data Scientists von Anfang bis Ende in die ML-Initiativen involviert sind und regelmäßig Manager über Fähigkeiten und Limitationen der Technologie aufklären müssen. Schlussendlich sind es die Data Scientists, die im Detail die Ressourcen und Fähigkeiten kennen, um Risiken für ML-Projekte abzuschätzen und Entwicklungspfade für ML-Anwendungen aufzudecken.

Quellen

[1] Google DeepMind, *https://deepmind.google/discover/blog/ai-for-the-board-game-diplomacy/*

[2] Shollo, A., Hopf, K., Thiess, T., Müller, O. (2022). "Shifting ML Value Creation Mechanisms: A process model of ML value creation", The Journal of Strategic Information Systems (Jahrgang 31, Ausgabe 3 im September 2022)

Über den Autor

Dr. Konstantin Hopf leitet die Data-Analytics-Forschungsgruppe am Lehrstuhl für Wirtschaftsinformatik, insbesondere Energieeffiziente Systeme, der Universität Bamberg. In industrienahen Forschungsprojekten entwickelt er betriebliche Anwendungen maschineller Lernverfahren. Er erforscht zudem Konzepte für das strategische Management von KI-Initiativen und Data-Science-Teams. Seine Forschungsergebnisse erscheinen in führenden Zeitschriften und Tagungsbänden der Wirtschaftsinformatik (zum Beispiel International Conference on Information Systems, California Management Review), aber auch in der Fach- und Tagespresse.

Dr. Konstantin Hopf
konstantin.hopf@uni-bamberg.de

„Daher ist es essenziell, das Zustandekommen der Ergebnisse nachvollziehen zu können und diese noch einmal „von Hand" zu prüfen."

Martin Meyer, Redaktionsleiter des Red Stack Magazin, sprach mit Dr. Robert Helbig, der den Bereich Inkubator und Technologische Innovation leitet und die zentrale Verantwortung für das Thema Generative AI bei der Datev eG übernimmt, über konkrete Anwendungsfälle bei der Datev, Datenschutz, den Einfluss auf Mitarbeiterstrukturen, die Nachvollziehbarkeit von KI-basierten Entscheidungen, die KI-Technologiestrategie bei der Datev sowie zu erwartende Trends.

Bitte stellen Sie sich unseren Lesern vor. Wer sind Sie und was sind Ihre Aufgaben bei der DATEV?

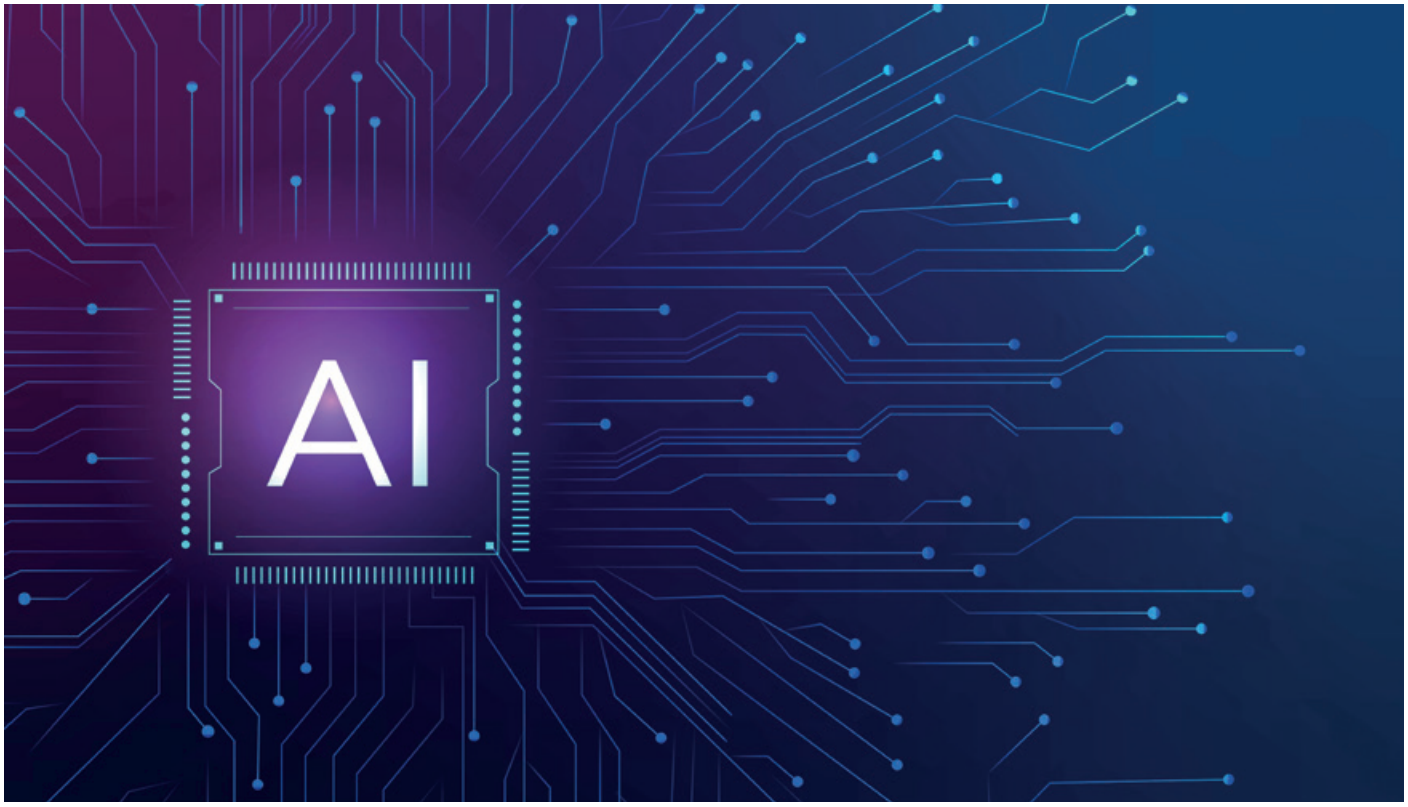
Mein Name ist Robert Helbig. Ich leite bei der DATEV eG den Bereich Inkubator und Technologische Innovation. Die Aufgabe des DATEV Inkubators ist es, innovative Geschäftsideen mit neuen Technologien zu verbinden, daraus erste Prototypen und Minimum Viable Products (MVPs) zu entwickeln und diese dann bei positiver Entscheidung weiter bis zur Produktreife zu treiben. In meinen Verantwortungsbereich fällt ebenso die zentrale Verantwortung für das Thema Generative AI. Wir verfolgen hierbei das Ziel, generative künstliche Intelligenz sachlich und chancenorientiert zu betrachten und sinnvoll anwendbar zu machen – sowohl für unsere Mitglieder als auch für unser Unternehmen selbst.

Was war ihr persönlicher Wow-Effekt bezüglich KI und warum?

Den kann ich sehr konkret benennen – Stichwort „AI Augmentation". Einer unserer Mitarbeitenden nimmt seine Gedanken für neue Ideen und Konzepte auf Ton auf, lässt sie durch Speech-to-Text AI transkribieren, schickt sie an ChatGPT, lässt sich die inhaltlichen Hauptpunkte zusammenfassen und fordert ChatGPT auf, Gegenargumente für die jeweiligen Punkte vorzuschlagen. Diese Möglichkeit, seine eigenen Gedanken derart anzureichern und herauszufordern hat mich tief beeindruckt.

Wie sieht Ihre Vision für den Einsatz von KI in der DATEV aus und wie trägt sie zur Erfüllung der Mission bei?

Unsere Mission lautet: Zukunft gestalten. Gemeinsam. Als Ausblick für den Einsatz von KI bei DATEV gibt es ein eindeutiges Zitat von unserem CEO Professor Dr. Robert Mayr: „Meine Erwartungshaltung ist an dieser Stelle aber auch ganz klar: Wir alle müssen uns mit dem Thema KI auseinandersetzen. KI muss zu unserem selbstverständlichen Begleiter im Arbeitsalltag werden. Jede und jeder von uns ist gefordert, sich diesbezüglich weiterzubilden und zu lernen." Dafür bauen wir zum Beispiel unser Weiterbildungsangebot zum Thema KI gezielt aus. So aufgestellt wollen wir dann auf strategischer Ebene unser Angebots-Portfolio sowie un-



ser Ökosystem weiterentwickeln, die Effizienz von Prozessen steigern und unseren Mitgliedern neue, zukunftsfähige Geschäftsmodelle ermöglichen.

Welche konkreten Anwendungsfälle sehen Sie für KI in Bezug auf ihre Arbeit?

Als Beispiele für konkrete Anwendungsfälle sehen wir die Nutzbarmachung von GPT-Technologie etwa im Rahmen eines so genannten DATEV-GPT, die KI-gestützte Verknüpfung interner Wissensquellen, um effektive und effiziente Recherche zu ermöglichen, den Einsatz von KI in der Softwareentwicklung im Rahmen eines AI-assisted Development-Ansatzes, die Unterstützung in den Bereichen Text-, Bild- und Videoproduktion sowie die Automatisierung von zeitaufwändigen Prozessen unter anderem in Organisation und Verwaltung.

Wie stellen Sie sicher, dass der Einsatz von KI ethisch und transparent ist, insbesondere im Hinblick auf Datenschutz?

Für unser Unternehmen und das damit verbundene Geschäftsmodell spielt Datenschutz seit jeher eine überragende Rolle. Daher beziehen wir auch zu allen KI-relevanten Themen die Kolleginnen und Kollegen der Abteilungen Recht, Datenschutz, IT-Security, Information Security und Privacy ein. Damit stellen wir sicher, dass alle rechtlichen Vorgaben und Richtlinien beim Einsatz von künstlicher Intelligenz gewahrt werden. Um möglichst viele Perspektiven bei Entscheidungen berücksichtigen zu können, holen wir uns darüber hinaus Unterstützung von unserer Personalabteilung sowie von Spezialistinnen und Spezialisten im Bereich Datenethik.

Welche Auswirkungen wird der Einsatz von KI in Ihrem Unternehmen auf Mitarbeiterstrukturen und -Positionen haben?

Der Aufbau der nötigen Skills spielt eine enorm wichtige Rolle, nachdem der Umgang mit künstlicher Intelligenz künftig für alle Mitarbeiterinnen und Mitarbeiter zum Handwerkszeug gehören wird wie heute vielleicht die Nutzung des Internets. Dementsprechend stimmen wir darauf unser Recruiting ab und investieren in den Aufbau und die Bereitstellung bedarfsgerechter Weiterbildungsmaßnahmen. Daneben blicken wir auch auf unsere Organisation und nehmen Veränderungen vor, sofern es die Rahmenbedingungen erfordern.

Welche Rolle spielt die Nachvollziehbarkeit von KI-basierten Entscheidungen und wie soll diese sichergestellt werden?

In unserem Kundensegment des steuerberatenden Berufsstandes (Steuerberater, Rechtsanwälte, Wirtschaftsprüfer) ist fachliche und rechtliche Richtigkeit absolut oberstes Gebot. Denn: Falsche Einschätzungen und Empfehlungen können schnell zu weitreichenden finanziellen und/oder rechtlichen Konsequenzen führen. Man denke an das Haftungsrisiko bei Steuerberaterinnen und Steuerberatern. Letztlich kann das dann auch die Reputation beeinträchtigen und damit mittel- bis langfristig negative Folgen haben. Deswegen spielt dieses Thema für uns eine absolut wichtige Rolle.

Gleichzeitig müssen wir gemeinsam mit unseren Kundinnen, Kunden und Mitgliedern lernen, dass KI immer wahrscheinlichkeitsbasiert arbeitet und damit Ergebnisse liefert, die nie zu 100 Prozent richtig sein müssen. Daher ist es essen-

ziell, das Zustandekommen der Ergebnisse nachvollziehen zu können und diese noch einmal „von Hand“ zu prüfen.

Ich sehe drei Aspekte, wie man dieser Thematik begegnen kann:

1. Kommunikation und Aufklärung: Man darf sich nicht blind auf KI-generierte Ergebnisse verlassen.
2. Die Anwendungen in sinnvollen kleinen Schritten arbeiten lassen, die den Anwenderinnen und Anwender helfen, das Zustandekommen der Ergebnisse nachvollziehen, kontrollieren und das Ergebnis bei Bedarf ändern zu können. Ich denke etwa an die Angabe von Quellen und Verlinkungen auf Volltexte.
3. Die technischen Möglichkeiten ausschöpfen, um bestmögliche Ergebnisse zu erzielen. Konkrete Beispiele dafür wären die Nutzung semantischer Ebenen oder Modelle auf Basis von Retrieval Augmented Generation (RAG).

Wie halten Sie sich über die neuesten technischen (KI)-Trends und Entwicklungen auf dem Laufenden?

Ich halte ich mich über die neuesten technischen (KI)-Trends und Entwicklungen auf dem Laufenden, indem ich verschiedene Methoden des Wissenstransfers nutze. Dazu gehört sowohl der unternehmensinterne als auch der unternehmensübergreifende Austausch. In unserem Unternehmen haben wir regelmäßige Meetings und Diskussionsrunden, in denen wir über aktuelle Entwicklungen sprechen und voneinander lernen können. Zudem nehmen wir auch an branchenspezifischen Messen und Konferenzen teil, um die neuesten Innovationen aus erster Hand zu erfahren und mit Experten auf diesem Gebiet in Kontakt zu treten.

Darüber hinaus nutze ich das Internet als wichtige Quelle für Informationen. Ich recherchiere regelmäßig auf spezialisierten Websites, Blogs und Foren, um auf dem neuesten Stand zu bleiben. Zudem lese ich Fachpublikationen, um detaillierte Einblicke in aktuelle KI-Trends und Forschungsergebnisse zu erhalten. Auch diverse Newsletter von renommierten Unternehmen, Fachleuten, Forschungseinrichtungen und Branchenverbänden helfen mir, über die neuesten Entwicklungen informiert zu bleiben.

Was sind die wichtigsten Ziele, die Sie mit Ihrer (KI)-Technologiestrategie erreichen wollen und warum?

Dazu haben wir bei DATEV im Jahr 2021 das Zielbild KI entwickelt. Das Zielbild KI beleuchtet das Thema künstliche Intelligenz aus fachlicher Sicht, zeigt den Kundennutzen auf, gibt Orientierung für die Teams und zeigt unter anderem die Potenziale des DATEV-Lösungsportfolios zur Erfüllung von Kundenbedürfnissen, die mit KI gelöst werden können. Die Ziele sind Wissen verfügbar zu machen, Eingabeeffizienz und Ausgabeintelligenz zu steigern, Prozesse zu automatisieren, Smarte Analysen zu ermöglichen, Beratungsqualität zu erhöhen und Compliance zu unterstützen. Da die Rohstoffe für KI-Anwendungen immer Daten sind, entstehen auf diese Weise Chancen für neue, datengetriebene Geschäftsmodelle.

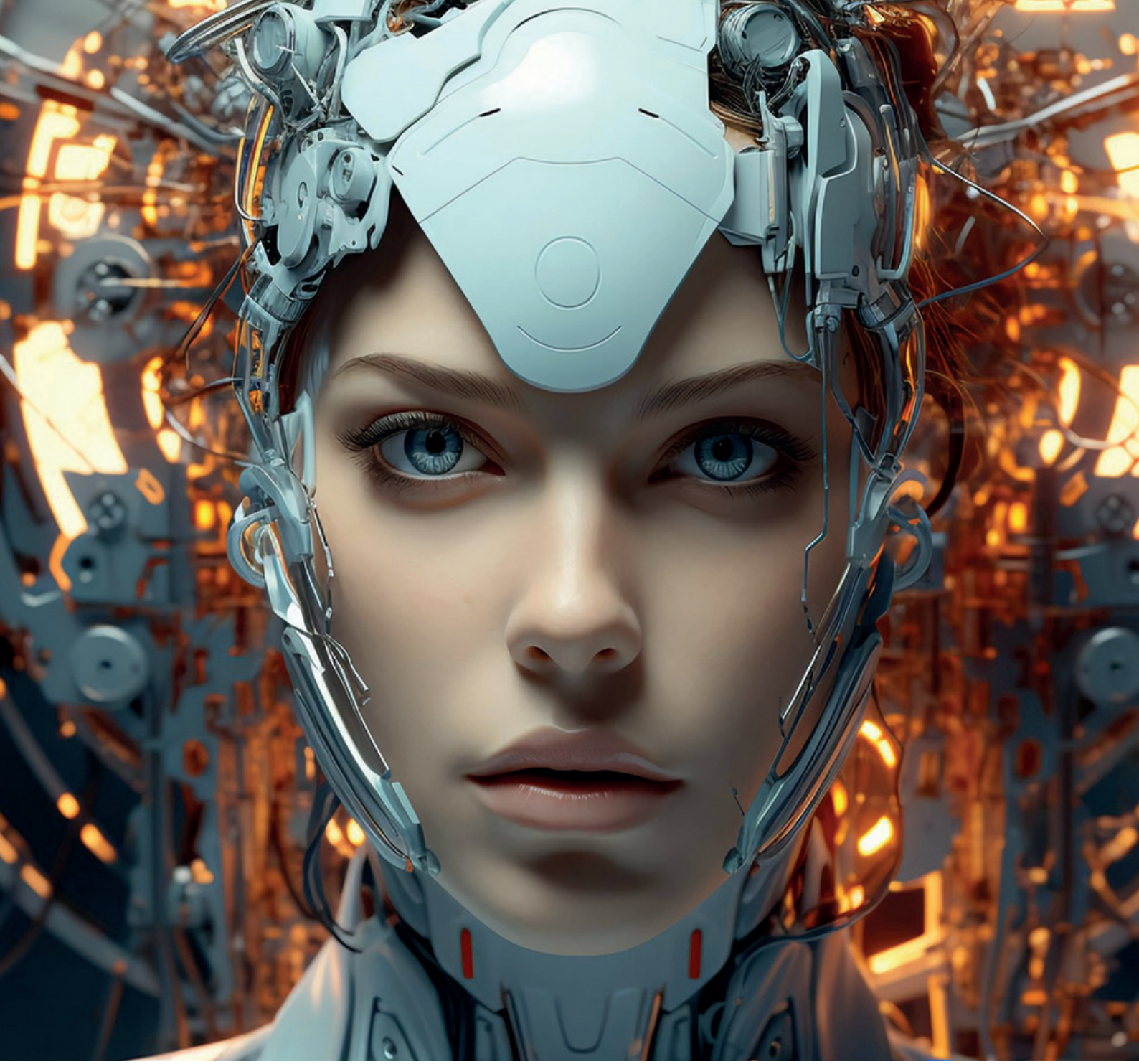
Welche Trends sehen Sie und welche Entwicklungen erwarten Sie in Zukunft bei KI und Digitalisierung?

Die Entwicklung ist derzeit dermaßen rasend schnell, dass ich mit einer Einschätzung zukünftiger Trends nur daneben liegen kann. Was ich derzeit wahrnehme, ist die zunehmende Awareness für die Themen Datenschutz, Datensicherheit und Datenethik. Daraus werden sicher sinnvolle und wichtige Impulse entstehen, wie wir als Unternehmen aber auch als Gesellschaft mit KI umgehen werden.

Zudem beobachten wir aktuell einen Trend zu Small Language Models (SLM), also kleineren Sprachmodellen, die explizit für spezielle Sachverhalte/Aufgaben trainiert werden, um Performance und Kostenentwicklung zu optimieren.

DR. ROBERT HELBIG

Dr. Robert Helbig ist Diplom-Volkswirt & Diplom-Kaufmann und hat zum Thema Steuerkomplexität promoviert. Er startete seine DATEV-Karriere im Jahr 2014 im Außendienst, wo er unter anderem für den Hochschulmarkt im Rahmen der DATEV-Bildungspartnerschaft verantwortlich war sowie die Leitung für mehrere Plattformprojekte innehatte. Im Jahr 2019 wechselte er dann in die Rolle eines Digital Transformation Officers.
Im selben Jahr übernahm er die Leitung und den Aufbau externer Entwicklungsteams bei DATEV. Zudem war er an der Konzeption des DATEV FlexHub beteiligt, einer Organisationseinheit, die im Sinne eines Marktplatzes Einheiten oder andere Ressourcen flexibel und zeitnah dort allokiert, wo Entwicklungspriorität & Skillnachfrage besonders hoch sind.
In seiner jüngsten Rolle übernimmt er die Leitung und den Aufbau des DATEV Inkubators & technologische Innovation als Chief Product Owner. Seit 2023 hat er zudem die zentrale Verantwortung für Generative AI bei DATEV.



Wie die KI das Planen für uns übernahm

Dr. Andreas Spies, Ordix

Die Ordix AG ist ein mittelständischer IT-Dienstleister mit einem reichhaltigen Seminarangebot. Die Herausforderung, eingehende Kundenanfragen bezüglich Seminarangeboten effizient zu verarbeiten, ist in Unternehmen mit wachsendem Seminarangebot und steigenden Anfragen evident. Wir zeigen im Folgenden die relevanten Komponenten und skizzieren unseren Ansatz, wie wir mit Hilfe künstlicher Intelligenz unser Ressourcenmanagement unterstützen.

Status Quo – Freiheit für Seminaranfragen? Man sieht den Sprecher vor lauter Referenten nicht!

Die steigende Anzahl von Seminaren und das Wachstum des Unternehmens machen es zunehmend aufwendig, den Überblick über angefragte Seminare, verfügbare Referenten und passende Termine zu behalten. Schnell sieht man den Wald vor lauter Bäumen nicht mehr. Die Komplexität wird durch die Form der Anfragen zudem erhöht. Sie liegen oft als Freitext vor, was die Identifikation des am besten geeigneten Seminars sowie verfügbarer Referenten erschwert. Das heißt, neben der reinen Prüfung der Verfügbarkeit möglicher Sprecher müssen Anfragen interpretiert und angefragte Inhalte und Titel mit dem vorliegenden Schulungsangebot abgeglichen werden.

Wir lichten den Seminarwald – So läuft's hier!

Die Business-Analyse legt aus wirtschaftlichen Gründen einen semi-automatisierten Prozess nahe. Er besteht aus den Komponenten Kunde, Front-End (User Interface) und KI-Backend und kann wie in *Abbildung 1* dargestellt werden.

Elektronisch eingehende Kundenanfragen werden manuell in eine semistrukturierte Form überführt und dabei in die wesentlichen Bestandteile, zum Beispiel angefragter Seminartitel, Inhalte und Zeiträume, gegliedert. Im KI-Backend wird automatisiert das bestgeeignete Seminar ermittelt und die zeitliche Verfügbarkeit potenzieller Referenten geprüft. Die Liste möglicher Seminare inklusive verfügbarer Referenten wird im User Interface anschließend zur menschlichen Verifikation angezeigt.

Was die Seminarwelt im Innersten zusammenhält – Der Matching-Algorithmus

Der Matching-Algorithmus versucht das am besten zur Kundenanfrage passende Seminar zu identifizieren. Neben der Kundenanfrage werden die Beschreibungen der Seminarinhalte aus dem Content-Management-System berücksichtigt. Der Algorithmus ist dabei in 3 Hierarchieebenen unterteilt. Zunächst wird in der Kundenanfrage über reguläre Ausdrücke nach einer Buchungs-ID gesucht. Führt dies zu einem Treffer wird das zugehörige Seminar gewählt. Ist die Suche erfolglos, folgt die zweite Hierarchiestufe. Basierend auf den angegebenen Titeln wird über die semantische Ähnlichkeit nach den bestgeeigneten Seminartiteln in der Seminardatenbank gesucht (*siehe* „Semantischer Abgleich" in *Abbildung 2*). Über einen Schwellenwert wird kontrolliert, wann ein semantisch ähnliches Seminar als potenzieller Kandidat betrachtet wird. Ist auch dieser Schritt erfolglos, werden in der letzten Ebene die angefragten Inhalte berücksichtigt („Aggregierter semantischer Abgleich" *siehe Abbildung 2*). Analog zum vorherigen Schritt, werden semantisch ähnliche Inhalte aus den Kundenanfragen und der Seminardatenbank ermittelt.

In den Tiefen der Suchmaschine – Was ist eine semantische Suche und wie funktioniert sie?

Das menschliche Kommunikationsmittel ist natürliche Sprache, ob geschrieben oder gesprochen. Dabei können gleiche Sachverhalte auf verschiedene Art und Weise ausgedrückt werden. Betrachten wir folgende Beispiele:

- Satz 1: „Künstliche Intelligenz wird zukünftig noch eine viel zentralere Rolle spielen."
- Satz 2: „Selbstlernende Algorithmen gewinnen immer mehr an Relevanz."
- Satz 3: „Agilität ist in einer schnelllebigen Gesellschaft von zentraler Bedeutung."

Wir erkennen leicht, dass die ersten beiden Sätze semantisch ähnlicher sind als der dritte Satz. Computer verstehen aber nur Zahlen. Die Aufgabe lautet also: Finde eine Möglichkeit, Sprache numerisch darzustellen, das heißt, wir brauchen ein Übersetzungstool von Sprache in Mathematik. Dabei muss berücksichtigt werden, dass Worte oder Textbausteine in unterschiedlichen Zusammenhängen unterschiedliche Bedeutungen haben.

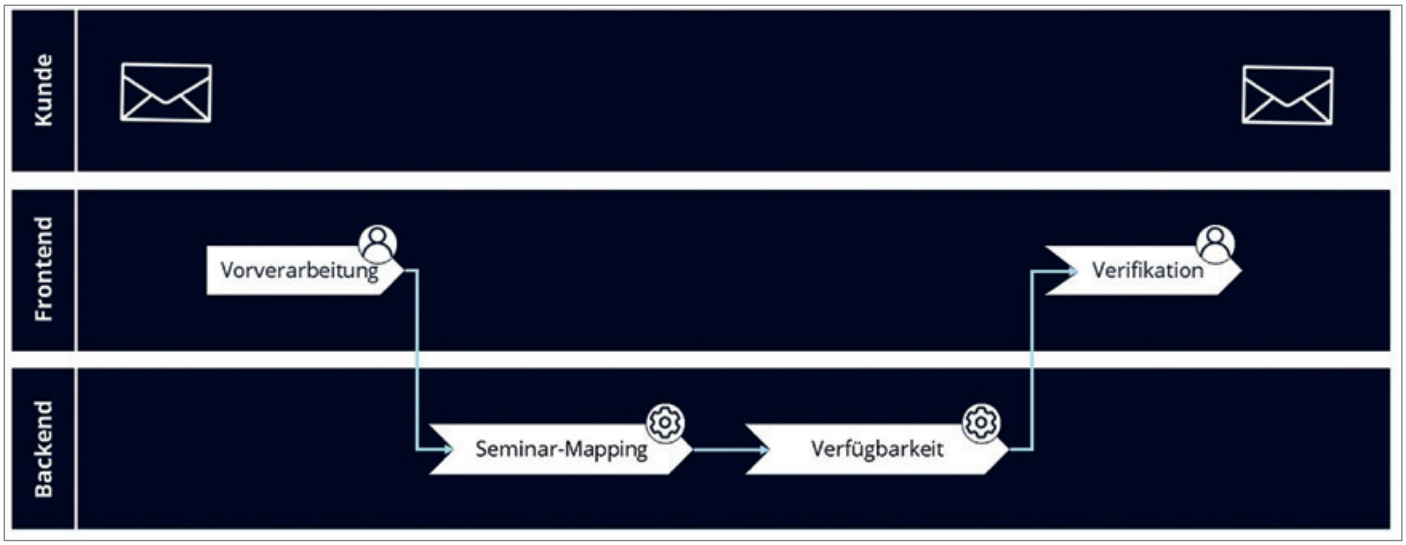


Abbildung 1: Schematischer Prozessablauf (Quelle: Dr. Andreas Spies)

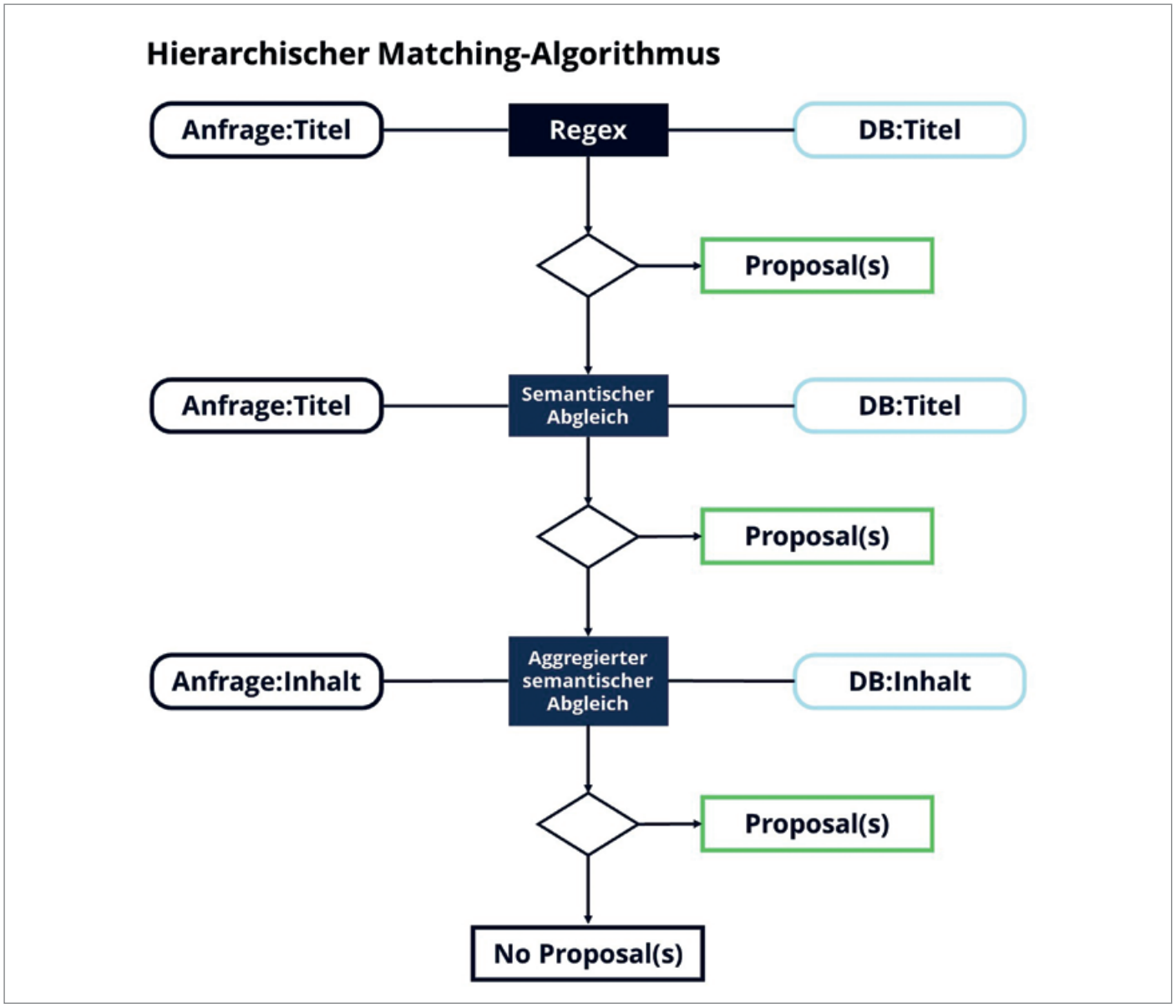


Abbildung 2: Hierarchischer Matching-Algorithmus (Quelle: Dr. Andreas Spies)

Die Übersetzung von Text in eine numerische Darstellung wird mit Hilfe von tiefen neuronalen Netzen und fortgeschrittenen Modellarchitekturen, sogenannte Transformer-Modellen *[1]*, realisiert. Im vorliegenden Fall wurde ein sogenannter *SentenceTransformer* verwendet *[2]*. Der Übersetzungsvorgang wird als Text-Embedding bezeichnet.

Nach der Übersetzung werden ähnliche Formulierungen durch ähnliche mathematische Repräsentationen ausgedrückt. Für die Darstellung eigenen sich hochdimensionale Vektoren. Im vorliegenden Fall handelt es sich um Vektoren der Dimension 1024. Die Ähnlichkeit lässt sich dann über den Winkel zwischen zwei Vektoren messen. Mathematisch gesprochen:

Vektor 1=Embedding(Ausdruck 1)
Vektor 2=Embedding(Ausdruck 2)
Semantische Ähnlichkeit=cos *(Vektor* 1,*Vektor* 2)

Zwei nahezu identische Vektoren, also semantisch ähnliche Textpassagen, schließen einen sehr kleinen Winkel ein. Stark unterschiedliche Vektoren hingegen umspannen einen großen Winkel.

In *Abbildung 3 „Semantische Ähnlichkeit"* sind die numerischen Vektoren der oben genannten Sätze graphisch dargestellt. In der Regel werden die Embedding-Vektoren in einen Raum (Embedding Space) eingebettet, der mehrere hundert Dimensionen umfasst. Aus Visualisierungsgründen haben wir uns hier auf zwei Dimensionen beschränkt. Die berechneten Ähnlichkeiten zwischen den einzelnen Sätzen sind in *Tabelle 1* zu finden.

Die Sätze 1 und 2 sind sich einander semantisch ähnlicher als die Satzpaare 1 und 3, sowie 2 und 3. Analog zur menschlichen Interpretation, „bewertet" auch der Computer die Ähnlichkeit der jeweiligen Sätze.

Im Kontext unsrer Seminarsuche werden sowohl die Kundenanfragen bestehend aus Titel und angefragten Inhalten als auch die in der Seminardatenbank hinterlegten Seminarinformationen in Vektoren übersetzt. Über den eingeschlossenen Winkel wird die semantische Ähnlichkeit zwischen Anfrage und Datenbankinhalt ermittelt. Dabei han-

Satz 1	Satz 2	Ähnlichkeit
Künstliche Intelligenz wird zukünftig noch eine viel zentralere Rolle spielen.	Selbstlernende Algorithmen gewinnen immer mehr an Relevanz.	0.63
Künstliche Intelligenz wird zukünftig noch eine viel zentralere Rolle spielen.	Agilität ist einer schnelllebigen Gesellschaft von zentraler Bedeutung.	0.33
Selbstlernende Algorithmen gewinnen immer mehr an Relevanz.	Agilität ist einer schnelllebigen Gesellschaft von zentraler Bedeutung.	0.26

Tabelle 1: Semantische Ähnlichkeit

delt es sich um eine 1:n-Beziehung, das heißt, für jede Anfrage müssen die Ähnlichkeiten zu allen n in der Datenbank vorliegenden Seminaren ermittelt werden. Das Anfrage-Datenbank-Paar mit der höchsten Ähnlichkeit wird als bestgeeignetes Seminar angenommen, sofern die Ähnlichkeit einen benutzerdefinierten Schwellwert übersteigt.

Speed is all you need – Laufzeitoptimierung

Die Berechnung von Text-Embeddings kann ressourcen- und zeitintensiv sein, abhängig von der Länge des angefragten Titels und der angefragten Inhalte. Im Schnitt enthält die Inhaltsangabe unserer Seminare etwa zehn Bullet-Points

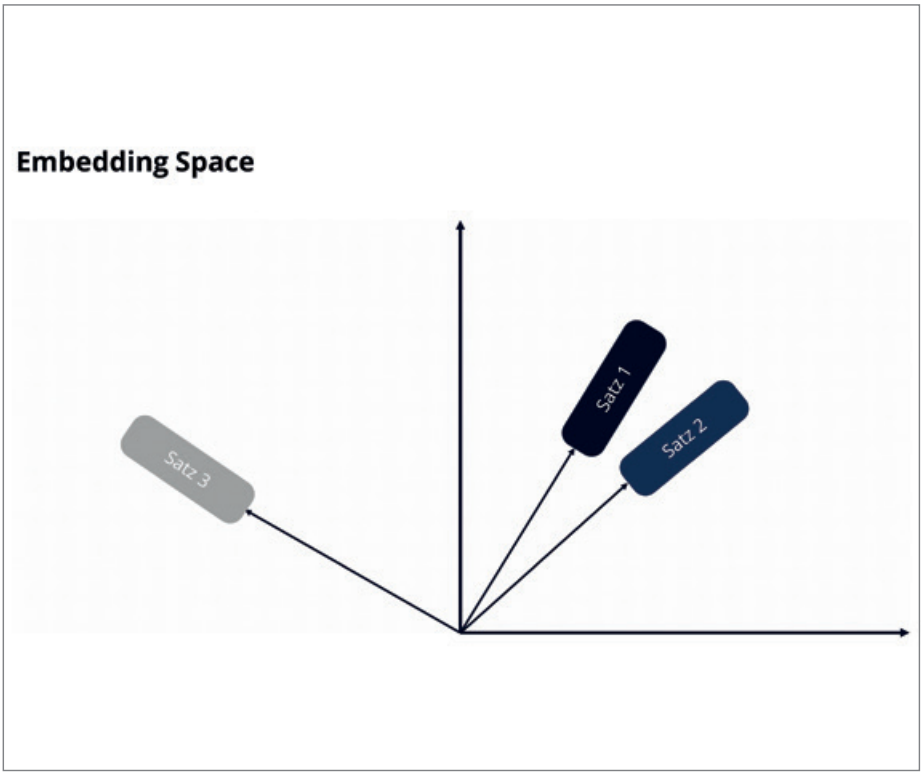


Abbildung 3: Semantische Ähnlichkeit (Quelle: Dr. Andreas Spies)

und die Berechnung eines Embeddings benötigt auf einer „herkömmlichen" CPU circa 0,5 Sekunden.

Im vorherigen Abschnitt haben wir gesehen, dass für die semantische Suche Text-Embeddings berechnet werden und anschließend die Ähnlichkeit zwischen den beiden Embeddings bestimmt wird. Pro Anfrage müssen also aktuell knapp 1500 Embeddings berechnet werden. Somit müssten unsere Kolleg:innen im Ressourcenmanagement etwas mehr als zehn Minuten auf die Antwort der KI-Anwendung warten.

Um die Laufzeit zu optimieren, wurde ein Vektor-Store entwickelt, der die eingebetteten Texte und die zugehörigen Metadaten als parquet-File speichert. Die Gültigkeit des Vektor-Stores wird in regelmäßigen Abständen geprüft. Im Falle eines veränderten Seminarangebots wird dann der Vektor-Store für neue oder geänderte Seminare neu berechnet, das heißt, neue Seminare werden hinzugefügt, veraltete Seminare entfernt. Ebenso werden die vorberechneten Vektoren aktualisiert, sollten sich Änderungen an den Inhalten der Seminare ergeben. Die Änderungen werden inkrementell durchgeführt, so dass kein regelmäßiger Full-Load notwendig ist.

Zur Laufzeit kann der Suchalgorithmus dann auf die vorberechneten Seminar-Embeddings zugreifen, so dass lediglich die im Frontend eingegebenen Seminaranfragen eingebettet werden müssen. Darauf basierend wird dann pro Anfrage die semantische Ähnlichkeit zum Seminarangebot bestimmt. Dadurch benötigen wir für die Ermittlung geeigneter Seminare nur noch 1 - 2 Sekunden.

Jemand zu Hause? Wir suchen verfügbare Referenten.

Nach dem Seminar-Matching sucht das System nach geeigneten Referenten. Neben der zeitlichen Verfügbarkeit der Sprecher wird bei der Suche auch die Dauer des ermittelten Seminars berücksichtigt. Der in Frage kommende Zeitraum, Z, ergibt sich also aus der Differenz des angefragten Zeitraums, AZ, und den Abwesenheitsperioden, AP *(siehe Abbildung 4)*. Für die in Frage kommenden Zeiträume, Z, dürfen nur zusammenhängende Zeiträu-

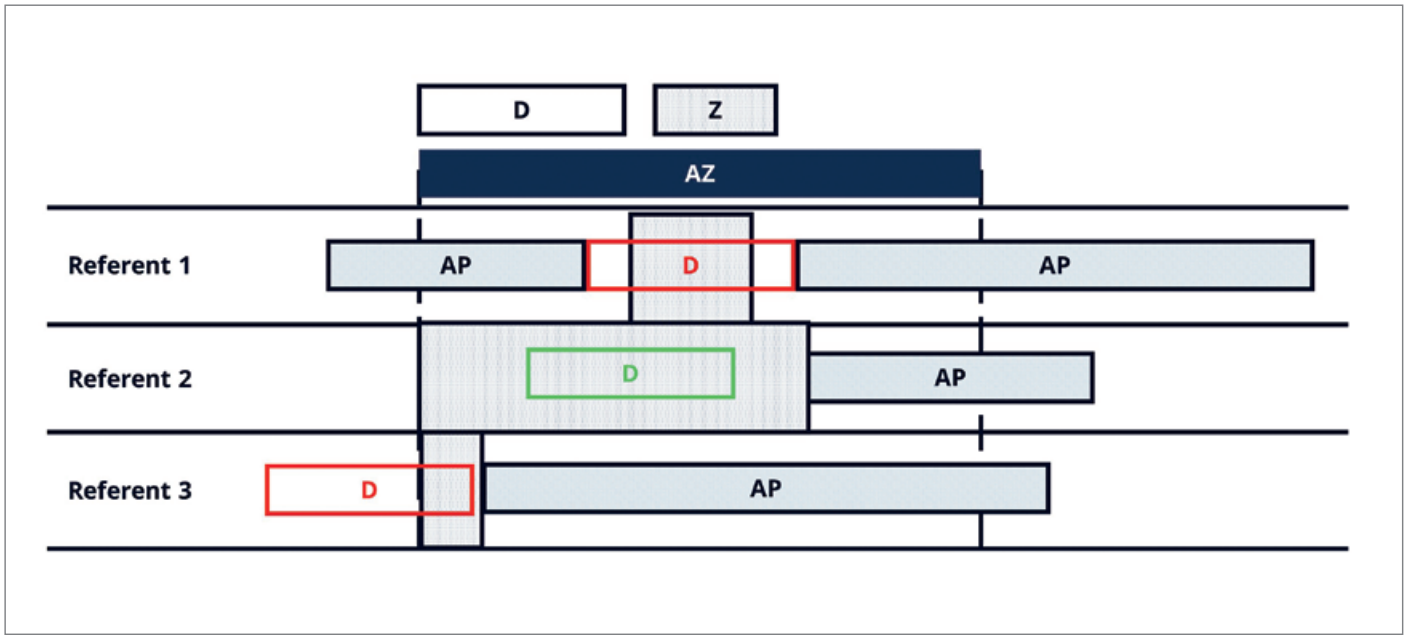


Abbildung 4:Verfügbarkeitsprüfung der Referenten (Quelle: Dr. Andreas Spies)

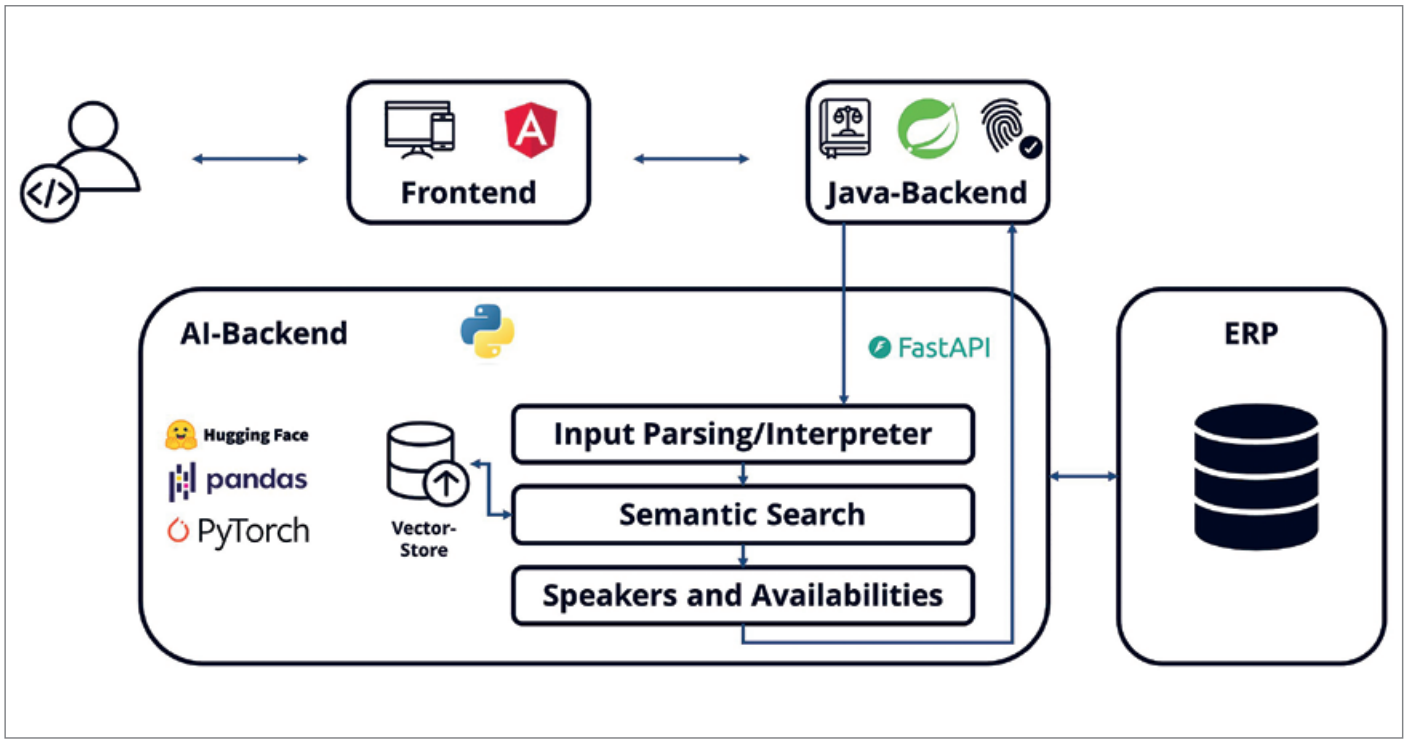


Abbildung 5: Gesamtarchitektur der KI-Lösung (Quelle: Dr. Andreas Spies)

me berücksichtigt werden. Ist die Dauer, D, des Seminars kleiner als der verfügbare Zeitraum, Z, wird ein Referent als „verfügbar" ermittelt.

Neben der reinen Verfügbarkeit werden pseudonymisierte Abwesenheitsgründe mit angereichert. So kann bei der manuellen Verifikation *(siehe Abbildung 1)* der mögliche Einsatz eines Referenten nach Rücksprache geändert werden.

Eine Tür macht noch kein Haus – Die Gesamtarchitektur der KI-Lösung

Die Gesamtarchitektur integriert ein Angular-Frontend, ein Java-basiertes Backend, das KI-Backend und das ERP-System der Ordix AG. Die Kommunikation der Web-Komponenten erfolgt über REST, wobei das KI-Backend über Datenbankabfragen mit OBDC mit dem ERP-System interagiert. Das Java-Backend übernimmt Aufgaben, wie die Authentifizierung und Autorisierung der im Frontend angemeldeten Benutzer und leitet die Anfragen zwischen dem Frontend und dem KI-Backend weiter. Auf diese Weise wird eine vollständige Kapselung des KI-Backends vom Endbenutzer erreicht *(siehe Abbildung 5)*.

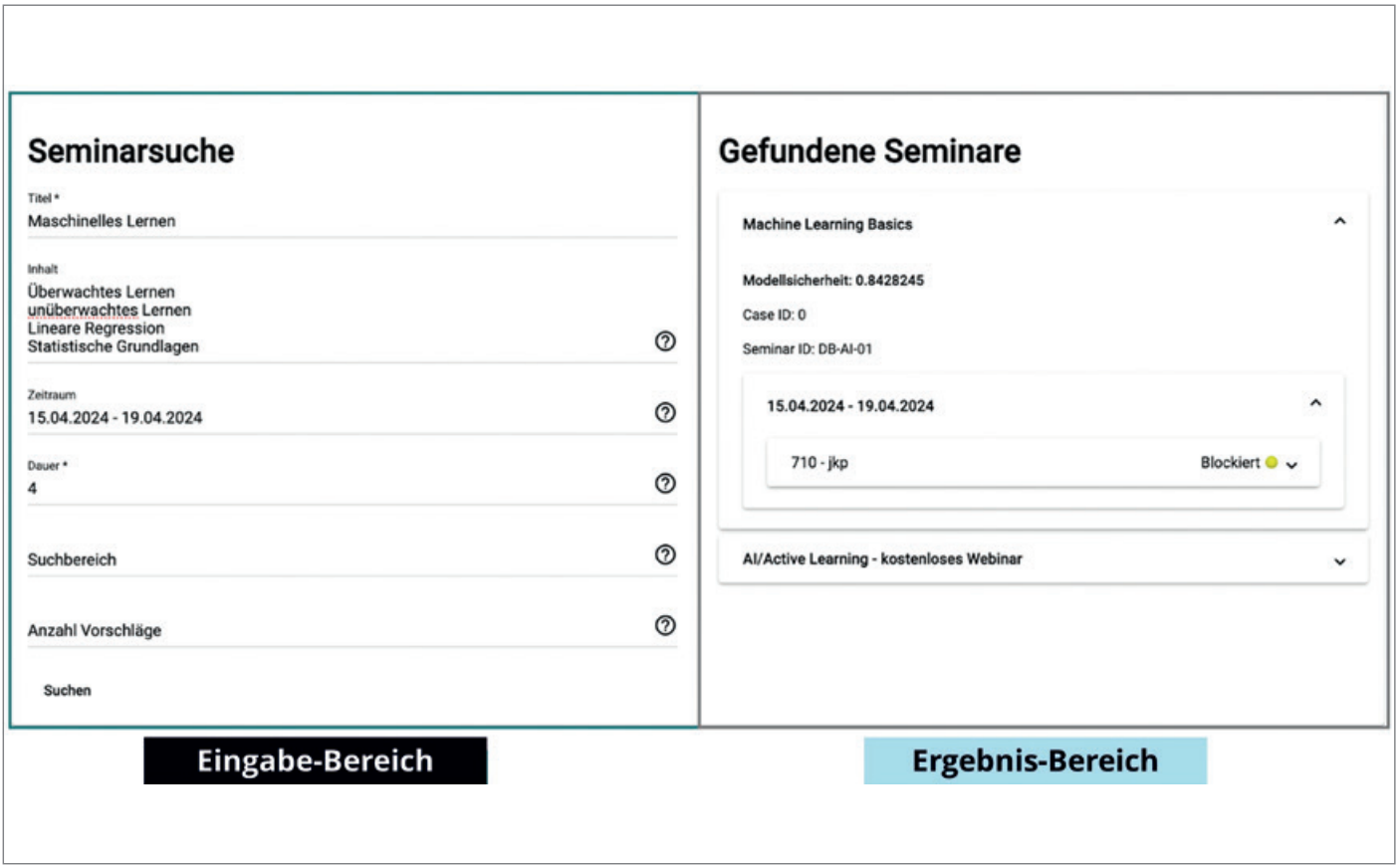


Abbildung 6: Web-Frontend der KI-Anwendung (Quelle: Dr. Andreas Spies)

Ein weiterer Baustein innerhalb des KI-Backends ist der bereits angesprochene Vektor-Store. Dieser wird vollständig vom KI-Backend verwaltet, mit den Seminardaten befüllt (und bei Bedarf aktualisiert) und ausgelesen.

Beispielhafte Suchanfrage:

Sehr geehrtes Seminar-Team,
wir suchen in der Woche vom 15.04.2024 - 19.04.2024 ein Seminar zum Thema maschinelles Lernen. Dabei sollen folgende Inhalte enthalten sein.

- *Überwachtes Lernen*
- *Unüberwachtes Lernen*
- *Lineare Regression*
- *Statistische Grundlagen*

Das Seminar sollte nicht länger als 4 Tage dauern.
Vielen Dank und viele Grüße,
Max Mustermann

Das Web-Frontend:

Die im Eingabe-Bereich angegebenen Informationen *(siehe Abbildung 6)* werden dann als JSON-Objekt an das KI-Backend über eine REST-API geschickt (*siehe Abbildung 7* „Beispiel JSON-Anfrage“).

All Good Things Come to an End – Fazit

Durch den Einsatz der entwickelten KI-Komponente konnte der Prozess der Referenten- und Seminarsuche vereinfacht und die Komplexität der Referentensuche und der damit verbundene zeitliche Aufwand für die Bearbeitung der Kundenanfragen erheblich reduziert werden. Die Kombination aus Ansätzen, basierend auf regulären Ausdrücken (Buchungs-ID) und der Methode der semantischen Suche (hierarchischer Ansatz, *siehe* oben), verbindet gleichzeitig die Möglichkeit, eindeutige Anfragen direkt über die Buchungs-ID zu verarbeiten, mit der Möglichkeit semantisch ähnliche Seminare zu ermitteln, ohne den exakten Wortlaut oder die exakten Inhalte zu nennen. Somit eignet sich dieses Vorgehen ideal für die Verarbeitung von Suchanfragen basierend auf natürlicher Sprache. Die Speicherung der vorberechneten Vektoren bietet eine weitere Zeitersparnis und verbessert die User Experience unserer Kolleg:innen im Ressourcenmanagement.

Die vorliegende Implementierung zeigt die nahtlose Integration von Methoden der künstlichen Intelligenz (semantische Seminarsuche), klassischen und „regelbasierten“ Algorithmen für die zeitliche Verfügbarkeit der Referenten und der manuellen Vorverarbeitung durch das Planungsteam und spiegelt damit den häufig interdisziplinären Einsatz von KI-Modulen wider.

Die Einsatzgebiete der verwendeten Methodik sind zudem erweiterbar auf eine Vielzahl an Anwendungsfällen, die eine semantische Zuordnung zweier oder mehrerer Komponenten enthalten.

So kann zum Beispiel im Bewerbungsprozess eine Vorauswahl geeigneter Kandidaten basierend auf dem semantischen Abgleich zwischen eingereichten Lebensläufen und/oder Berufserfahrungen und dem geforderten Stellen-/Skill-Profil erfolgen. Ein weiterer Anwendungsfall aus dem

```
{
    "title": "Maschinelles Lernen",
    "content":
    [
        "Statistische Grundlagen",
        "Überwachtes Lernen",
        "unüberwachtes Lernen",
        "Lineare Regression",
    ],
    "date":
    [
        "15.04.2024 - 19.04.2024",
    ],
    "duration": 4,
    "location": [],
}
```

Abbildung 7: Beispiel JSON-Anfrage (Quelle: Dr. Andreas Spies)

Bereich Human Ressources ist die Mitarbeiter:innen-Entwicklung bezogen auf stellen- oder rollenspezifische Qualifikationsanforderungen. Mit Hilfe der semantischen Zuordnung können gezielte Entwicklungspläne identifiziert und entworfen werden.

Sie möchten mit uns Ihren Anwendungsfall diskutieren? Wir stehen gerne zur Verfügung, um Ihre Anforderungen mit Ihnen zu besprechen.

Quellen

[1] Attention Is All You Need; Vaswani et. al.; 2017; Advances in Neural Information Processing Systems 30 (NIPS 2017); S. 5998 - 6008

[2] Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks; Nils Reimers, Iryna Gurevych; 2019; arXiv:1908.10084

Über den Autor

Dr. Andreas Spies ist Principal Consultant bei der Ordix AG, ausgebildeter Astrophysiker und leidenschaftlicher Data Scientist. Als Teamleiter im Bereich Data Science hat er knapp 10 Jahre Projektleitungs- und Hands-On-Erfahrung in der Bearbeitung von Use Cases der Finanz- und Versicherungsbranche aus den Bereichen Data Science, Computer Vision und Natural Language Processing. Hinzu kommt eine breite Erfahrung im Bereich der Bearbeitung „klassischer“ Data Science Use Cases mit strukturierten/tabellarischen Daten.

Dr. Andreas Spies
asp@ordix.de

10 JAHRE JAVALAND

JavaLand

www.javaland.eu

JAVALAND 2024 VERPASST?

ALLE ON-DEMAND-ANGEBOTE IM TICKETSHOP

JETZT ON-DEMAND-TICKET BUCHEN UND
VORTRAGSAUFZEICHNUNGEN ANSCHAUEN!

Präsentiert von: iJUG Verbund · Heise Medien · DOAG

Veranstalter: JavaLand

KI in den Oracle Applications – ein Zwischenbericht

Katharina Schraft, PROMATIS software GmbH

KI ist da, und sie beeinflusst alle Bereiche unseres Lebens. Gerade im privaten Bereich nutzen viele Menschen verfügbare Tools wie ChatGPT – oft findet der Einsatz auch wenig reguliert und nicht transparent im Geschäftsumfeld statt. Aber was wäre, wenn hilfreiche KI-Funktionalitäten bereits in Applikationen integriert wären? Hier ein Zwischenbericht.

Oracle baut auf verschiedene Arten der KI, um Lösungen für Applikationen bereitzustellen, die prägnantesten Beispiele gibt es im Bereich generative KI *(siehe Abbildung 1)*. Aktuell gibt es Tätigkeiten, für deren Lösung im schlimmsten Fall freie KI-Lösungen wie ChatGPT genutzt werden, die außerhalb des Systems liegen. Oracle möchte in Zukunft eine integrierte, sichere Lösung bereitstellen, welche die notwendigen Kontextdaten bereits zur Verfügung hat.

Bereits bei der Oracle CloudWorld hat Steve Miranda, Executive Vice President Applications Development bei Oracle, angekündigt, neue generative KI-Funktionen anzubieten – für jeden und ohne extra Kosten. Wir haben damals berichtet und frei zitiert: „Das Einzige, was wir wissen, ist, dass die Geschwindigkeit des Wandels zunimmt!“ *[2]*. Die heutige Welt und ihr Wandel stellt uns alle vor neue Herausforderungen, Wandel bietet aber auch immer viele neue Chancen. Sich vor den Neuerungen in der Informationstechnologie zu verschließen, ist keine Option und es müssen sichere Wege gefunden werden, den Fortschritt sinnvoll im eigenen System zu integrieren. Laut NewVantage Partners, die auch von Oracle zitiert wurden, hatten 2022 bereits 31% der Firmen KI im Einsatz, während 91% in KI-Initiativen investiert haben *[3]*. Davon möchten ERP-Anbieter wie Oracle gerne profitieren und so wird der Bereich KI öfter thematisiert: zum Beispiel über Artikel auf dem Oracle AI & Data Science Blog oder auch im Rahmen von regelmäßigen Webinare und Pressemeldungen.

Aber bemerkt man etwas dieser Bemühungen an den eigenen Applikationen?

Oracle und Cohere

Oracle ist eine Partnerschaft mit Cohere Technologies eingegangen *[3]*, um generative KI in die Oracle-Anwendungen

zu integrieren. Cohere Technologies ist ein Unternehmen, das sich auf die Entwicklung von KI-Technologien im Bereich generative KI spezialisiert hat. Die Besonderheit von Cohere im Bereich der generativen KI liegt in der Entwicklung fortschrittlicher Modelle für die natürliche Sprachverarbeitung (Natural Language Processing, NLP), die auf dem neuesten Stand der Forschung sind. Coheres Stärke liegt in der Fähigkeit, Modelle zu entwickeln, die nicht nur präzise und effizient sind, sondern auch menschenähnliche und kontextsensitive Antworten liefern können.

Dies wird durch die Nutzung von fortschrittlichen Techniken des maschinellen Lernens und des Deep Learning ermöglicht, und stellt Modelle bereit, die eine breite Palette von natürlichsprachlichen Aufgaben bewältigen können. Dazu gehören Aufgaben wie Übersetzung, Zusammenfassung, Frage-Antwort-Systeme und Chatbots.

Die Zusammenarbeit mit Cohere ermöglicht es den Oracle-Anwendungen, fortschrittliche Technologien zu verwenden, mit denen die Anwender generative KI als Unterstützung für ihre täglichen Aufgaben einfach und integriert nutzen können. Dies verbessert die Benutzerfreundlichkeit und ermöglicht eine effizientere Nutzung der Oracle-Anwendungen.

Fusion Applications

Fusion Applications, das Cloud-Flaggschiff der Oracle ERP-Systeme, soll in vielen Bereichen mit generativer KI angereichert werden. Dies wurde bereits bei der Oracle CloudWorld 2023 mit ersten Anwendungsbeispielen angekündigt *[1]*. Diverse KI-Services sollen hier in den verschiedenen Applikationen Mehrwert schaffen, insbesondere in Bezug auf digitale Assistenten, Dokumentenverständnis, Erkennung von Anomalien wie auch Forecasting *(siehe Abbildung 2)*.

In den **HCM**-Applikationen wurde insbesondere die Verfassung von Texten wie Stellenanzeigen (Assisted Authoring), die Auswertung von Ergebnissen und Zahlen (Summarization) oder auch das Erstellen von Fragebögen für Performancegespräche (Suggestions) vorgestellt. *Abbildung 3* zeigt ein Beispiel für diesen Einsatzbereich: Es findet eine Mitarbeiterreview statt, zum Beispiel persönlich mit dem Vorgesetzten oder Bereichsleiter. Mit Hilfe des Fragebogens, welcher im Vorfeld mit generativer KI erstellt werden kann, wird anhand konkreter Leistungen des Mitarbeiters eine Bewertung vorgenommen. Als Basis dienen beispielweise Aktivitäten, die während des Tagesgeschäfts im System erfasst wurden, zum Beisoiel die Erreichung von vertriebszielen oder fakturierten Tagen bei Beratern. Anhand der Auswertung gibt es Vorschläge für die weitere Entwicklung. Die KI kann über den Button „Use AI Assist“ angefragt werden. Diese explizite Aktion des Klickens sorgt dafür, dass es für den Mitarbeiter transparent sichtbar ist, welche Informationen von einer KI beisgesteuert werden. Diese müssen nach dem Einfügen durch den Bediener (zum Beispiel den bewertenden Vorgesetzten) übernommen werden. Am Ende wird dann der gemeinsam ausgefüllte Fragebogen gespeichert.

Es stellt sich bei vielen die Frage, warum das nötig sei. Für den Mitarbeiter und im Idealfall ebenso für die Führungskraft liegen die Ergebnisse des vergangenen Jahres vor, sodass eine Bewertung auch ohne die Nutzung von KI problemlos möglich ist. Die Gründe, warum es sinnvoll ist, die KI miteinzubeziehen, sind aber beispielsweise die Objektivität und Vollständigkeit der Inhalte. Wenn kein Zweifel an den vorliegenden Daten besteht, sparen alle Zeit in der Vorbereitung der Gespräche und durch das Einbringen von Fakten kann es einfacher gefördert werden, dass eine objektive Bewertung ohne Voreingenommenheit durchgeführt wird.

Im **Customer Experience Bereich** werden besonders der Support durch Chatbots und erstellte Wissensdatenbanken (Assisted Authoring) sowie die

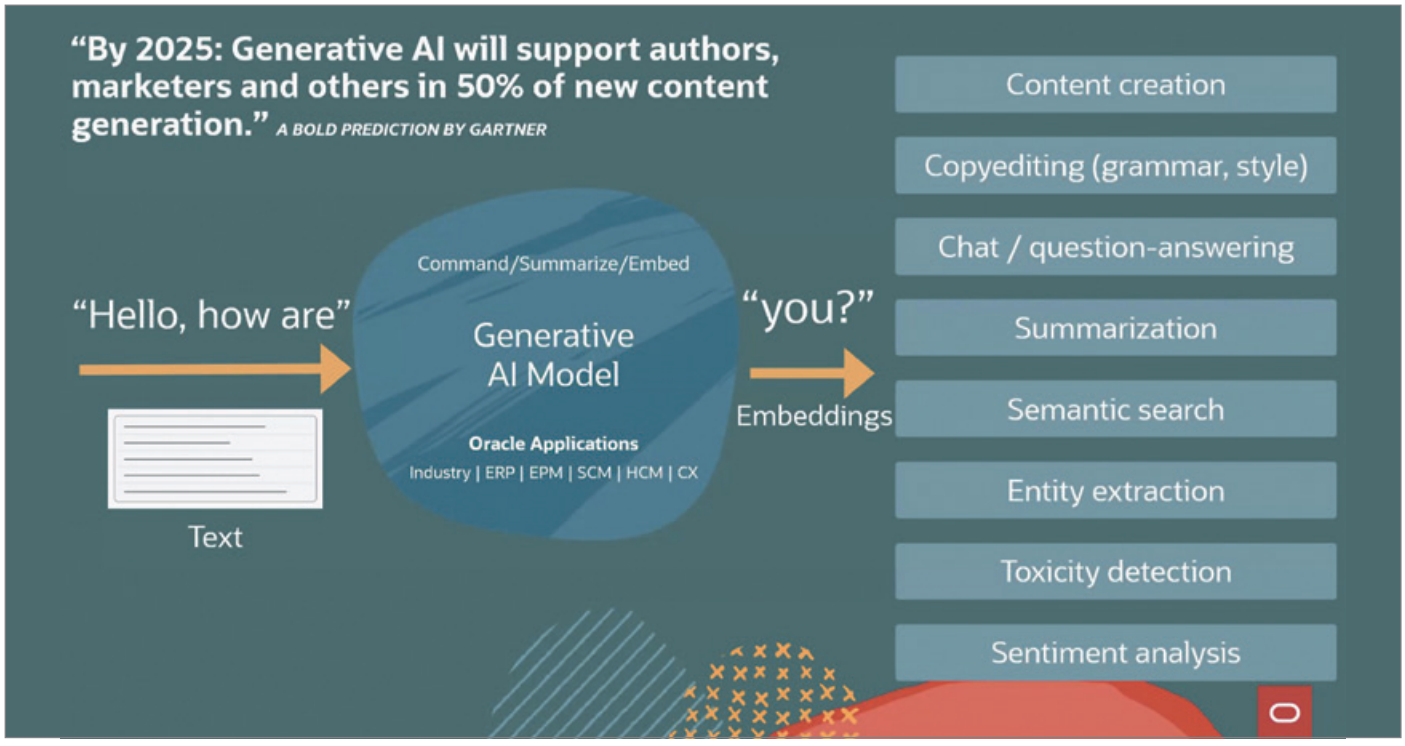


Abbildung 1: Generative KI Modelle [1]

Abbildung 2: Generative KI in Cloud-Applikationen [1]

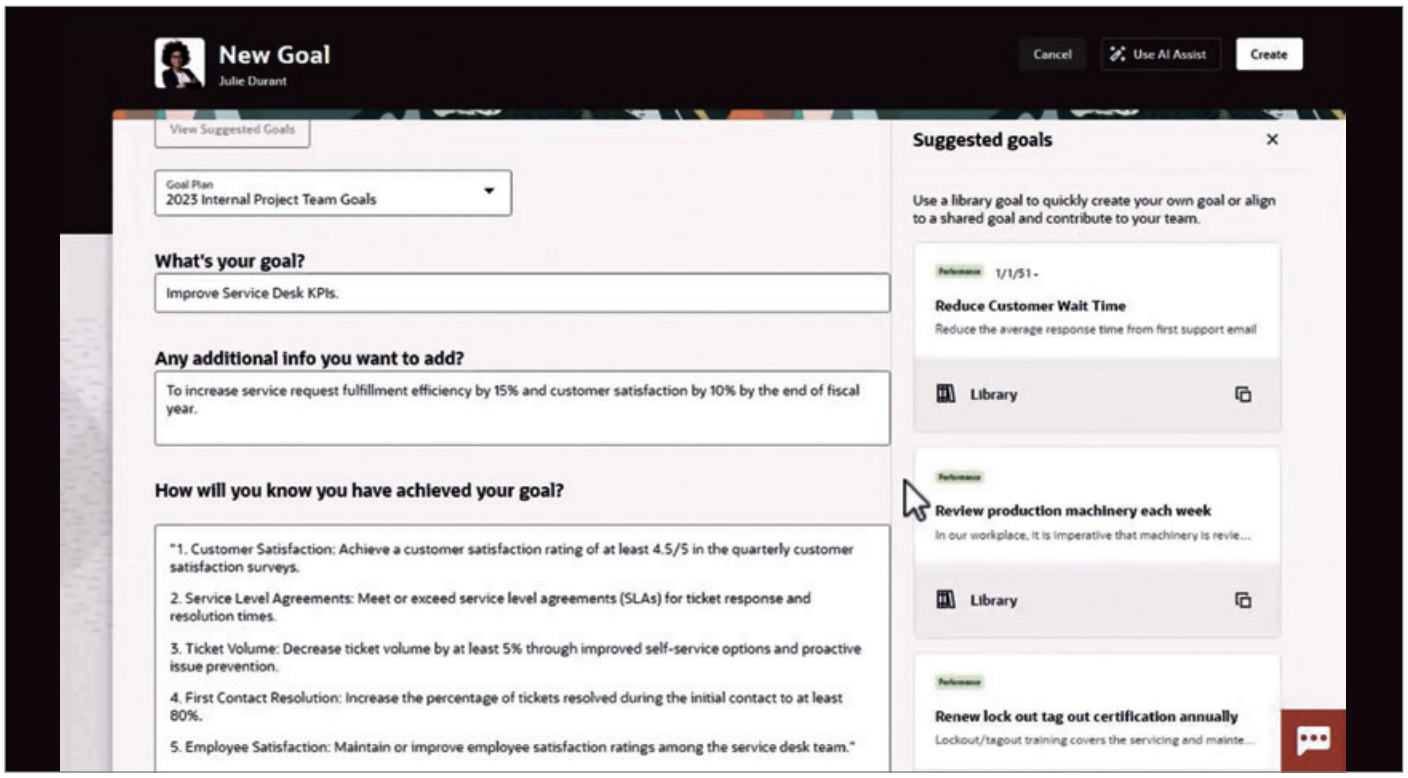


Abbildung 3: Performancegespräch Mitarbeiter [1]

Auswertungen von Kundenaktivitäten (Summarization) als positive Hilfe hervorgehoben. Kundenzufriedenheit und die effektive Nutzung von Supportressourcen ist bei vielen Firmen ein entscheidender Baustein des Erfolges. Zum Beispiel im Produktgeschäft entwickeln sich die Produkte, welche bei den Kunden im Einsatz sind, häufig weiter, manchmal gibt es auch verschiedene Versionen pro Kunde. Ein effektives Supportmanagement ist hier eine große Herausforderung, da eine Dokumentation immer aktuell gehalten werden muss. Hier kann generative KI helfen. Durch eine Chatbotlösung, welche auf existierende Informationen zugreift, kann ein Supportmitarbeiter effektiv entlastet werden, da er die benötigten Informationen vom Bot suchen lässt und so nicht selbst heraussuchen muss. Zudem ist es der generativen KI auch möglich, bei der professionellen Kommunikation zu unterstützen, indem sie mittels „Rephrase" am Ende die Antworten professionell formatieren kann *(siehe Abbildung 4)*.

Im **ERP**-Bereich liegt der Fokus auf narrativem Reporting und der Deutung von Ergebnissen, um Geschäftsentscheidungen (wie zum Beispiel die Auswahl des besten Lieferanten) zu treffen. Zudem kann im Finanzbereich KI genutzt werden, um das Abstimmen von Daten zu automatisieren, um somit

schneller und zuverlässiger an tragfähige Geschäftsergebnisse zu kommen. *Abbildung 5* zeigt ein Beispiel aus dem SCM-Umfeld. Hier werden die vorliegenden Daten im System genutzt, um Einblick in einen vergangenen Zeitraum darzustellen. Zudem können auch, basierend auf diesen Daten, Entscheidungen oder Tendenzen für die Zukunft empfohlen werden. Was den Anwender hier unterstützt, sind die Insights und Oracle AI Assist, welches Texte vorschlagen und nach Anwenderwünschen erweitern kann. Die Erstellung von Grafiken zur Veranschaulichung der Ergebnisse kann direkt vorgenommen werden.

Im **EPM** (Enterprise Performance Management) -Bereich werden bereits verschiedene Arten der KI eingesetzt. So bietet zum Beispiel die Planning & Budgeting Cloud die Funktionalität „IPM Insights". Diese ermöglicht schnellere Datenanalyse, um schneller Maßnahmen ergreifen zu können. Oracle sagt selbst: „Mit eingebetteter KI und maschinellem Lernen (ML) können Sie Ihre Pläne, Prognosen und Abweichungen kontinuierlich überwachen. So werden Sie über Anomalien, Vorurteile sowie verborgene Zusammenhänge informiert. Sie erhalten direkt relevante Einblicke, sodass Sie rechtzeitig die richtigen Maßnahmen ergreifen können." *[6]* Maschinelles Lernen macht den Unterschied. Hier liegt der Fokus, im Gegenteil zur generativen KI, nicht auf dem Erzeugen von Inhalten, sondern auf dem Auswerten von großen Datenmengen, um daraus Schlüsse ziehen zu können. Durch die Integration von verschiedenen Datenquellen in der EPM-Plattform kann eine breite Informationsbasis aufgebaut werden. Die klare Analyse und Auswertung ermöglicht den Blick auf das Wesentliche und vermeidet blinde Flecken. Auch die Nutzung von eingebettetem ML in eigenen Lösungen wird unterstützt. Laut Oracle ist die Nutzung dieser Technologie sehr einfach: „Jeder – von Klickern bis zu Programmierern – kann eingebettetes ML verwenden, um benutzerdefinierte, geschäftsspezifische Modelle für eine bessere Entscheidungsfindung zu erstellen. Geschäftsbenutzer benötigen keine speziellen technischen oder Programmierkenntnisse für die Verwendung von ML." *[7]*

In Kombination mit generativer KI ermöglicht die Nutzung von narrativem Reporting, diese großen Datenmengen schnell und effektiv in aussagekräftige Auswertungen und Berichte zu transformieren.

Viele dieser Funktionalitäten sind bereits nutzbar *[8]* und werden nach und nach für die einzelnen EPM-Komponenten, auch außerhalb der Enterprise Lizenz, zur Verfügung gestellt. So zum Beispiel in Oracle NetSuite Planning and Budgeting, das auf dem Planning and Budgeting Cloud Service basiert und für die Nutzung mit dem Oracle NetSuite ERP optimiert ist.

Oracle NetSuite hat bei der SuiteWorld im Jahr 2023 ebenfalls das Thema KI vertieft und hat nun für das Release 2024.1 Funktionalitäten im Bereich Textverbesserung und Rechnungserfassung (NetSuite Bill Capture) angekündigt *[5]*. Auch Oracle Analytics IPM (siehe Abschnitt über Enterprise Performance Management) wird für NetSuite Planning und Budgeting zur Verfügung gestellt.

Neu erscheint NetSuite Text Enhance. Diese Funktionalität hilft mittels generativer KI, Texte in NetSuite anzureichern. Dies soll die Arbeit der Mitarbeiter beschleunigen. Angekündigte Bereiche sind die Produktbeschreibungen für Artikel, Tätigkeitsbeschreibungen und das Verfassen von Stellenausschreibungen, das Entwerfen von Mitarbeiterzielen und das Erstellen von Kundendienstkommunikation – also Funktionalitäten, die fast den angekündigten Funktionalitäten der Oracle Cloud Applications entsprechen.

Im Fall von NetSuite Text Enhance gibt es einen AI Assist Button, der Texte entsprechend der eigenen Wünsche aufarbeitet, ein Beispiel ist in *Abbildung 6* zu sehen. Hier handelt es sich um ein Beschreibungsfeld eines Artikels, welches im Folgeprozess beispielsweise in einem Webshop angezeigt werden kann. Dieses Beispiel ist aus einem offiziellen Video von NetSuite entnommen *[10]*.

Wenn man sich diese Funktionalität in einer Release Preview 2024.1 in Deutschland anschauen möchte, wird die Freude allerdings schnell getrübt, denn diese Funktionalität, wie auch NetSuite Bill Cap-

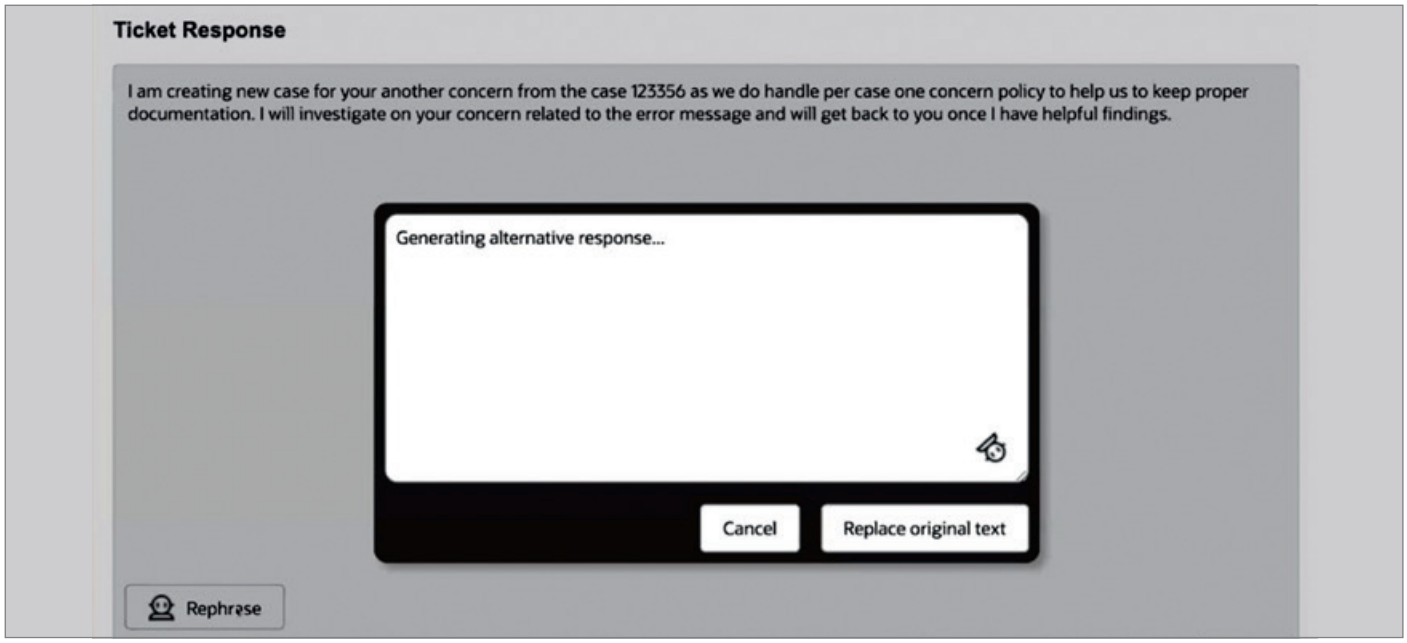


Abbildung 4: Verbesserung der Ticketbearbeitung [1]

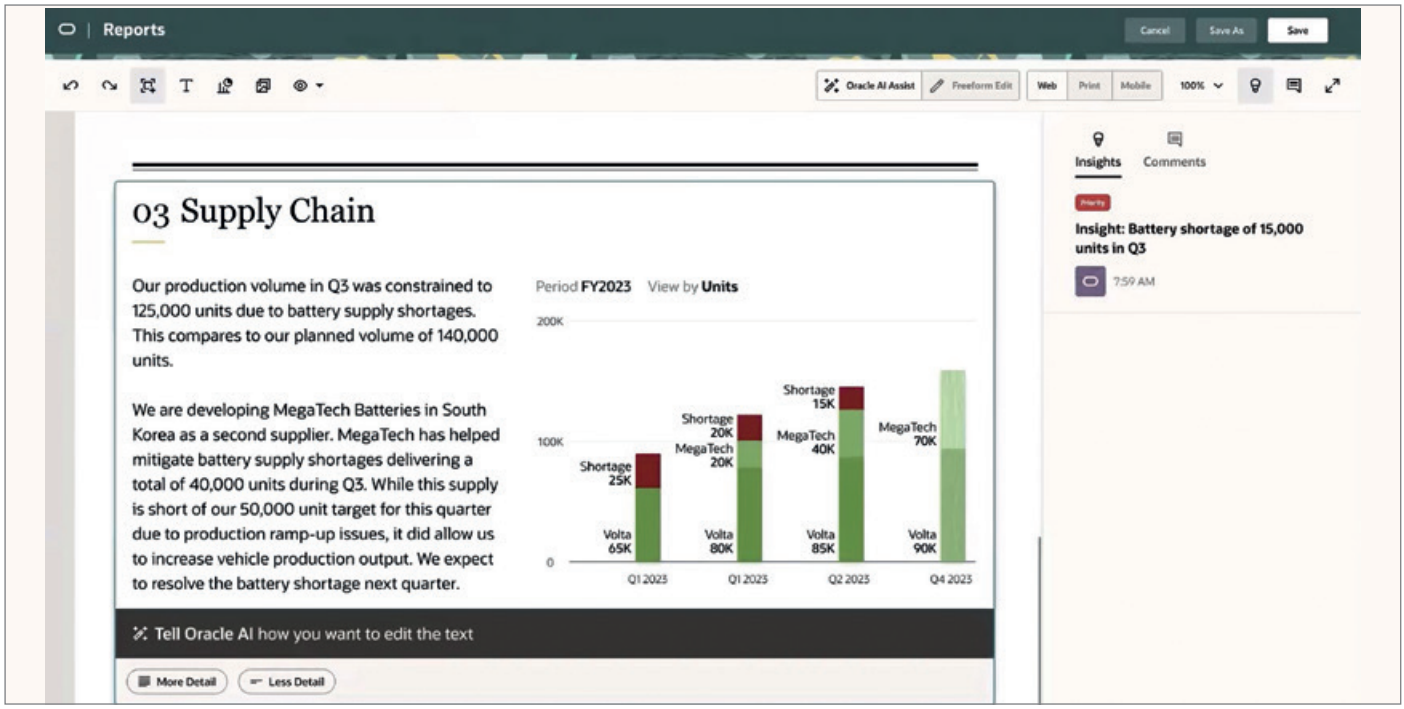


Abbildung 5: Narratives Reporting [1]

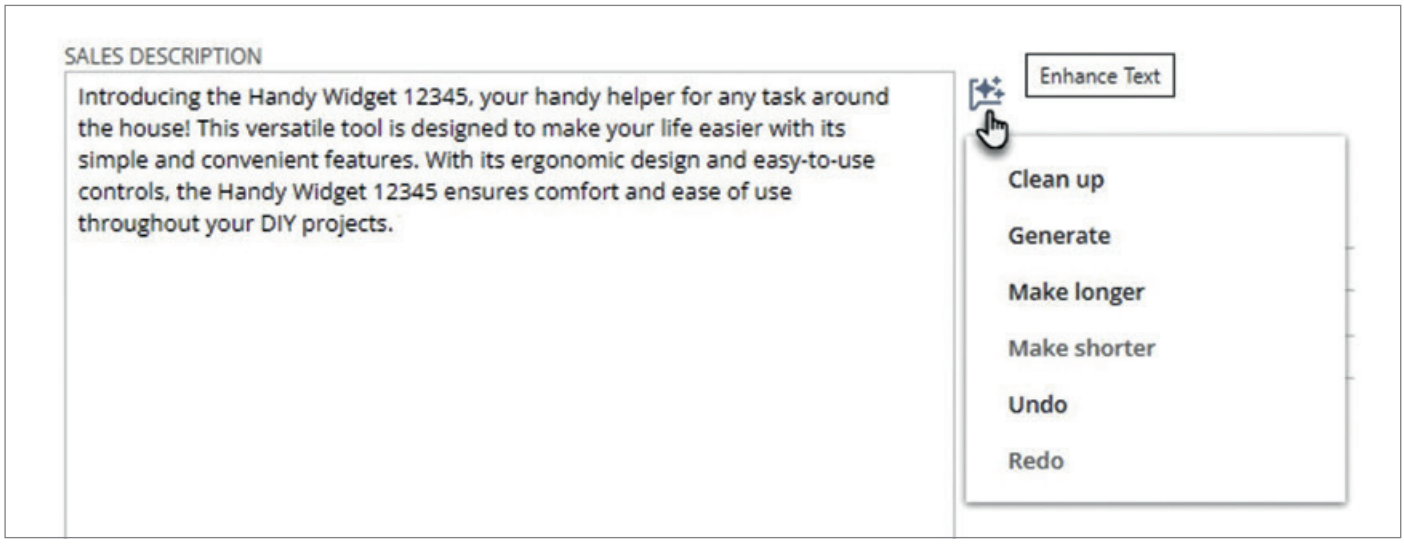


Abbildung 6: Beispiel einer Erstellung eines Beschreibungsfeldes mit Hilfe von NetSuite Text Enhance [9]

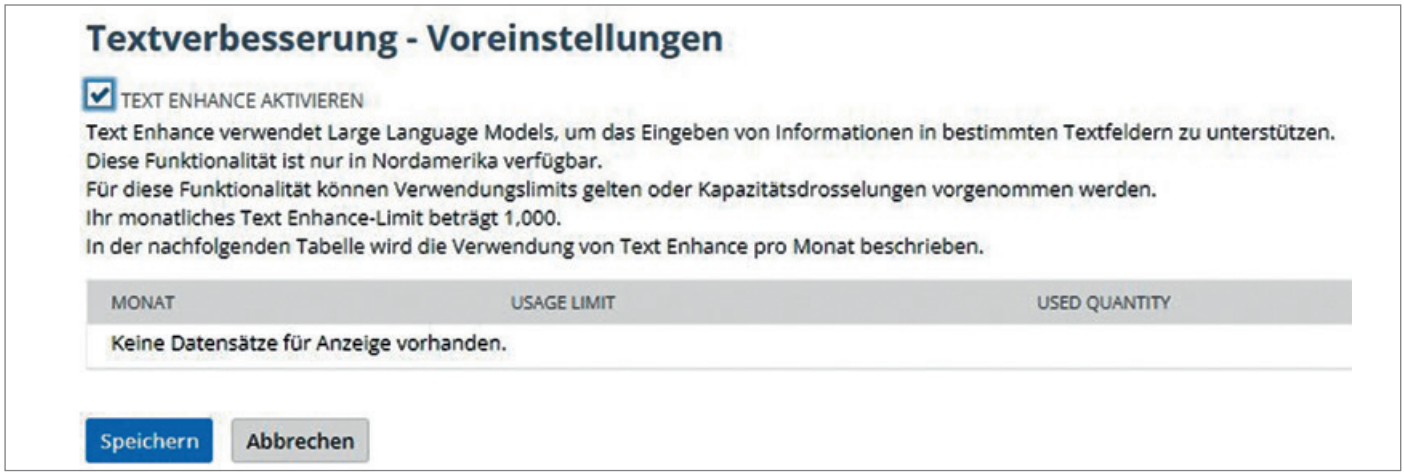


Abbildung 7: NetSuite Text Enhance in Deutschland (selbst erstellt)

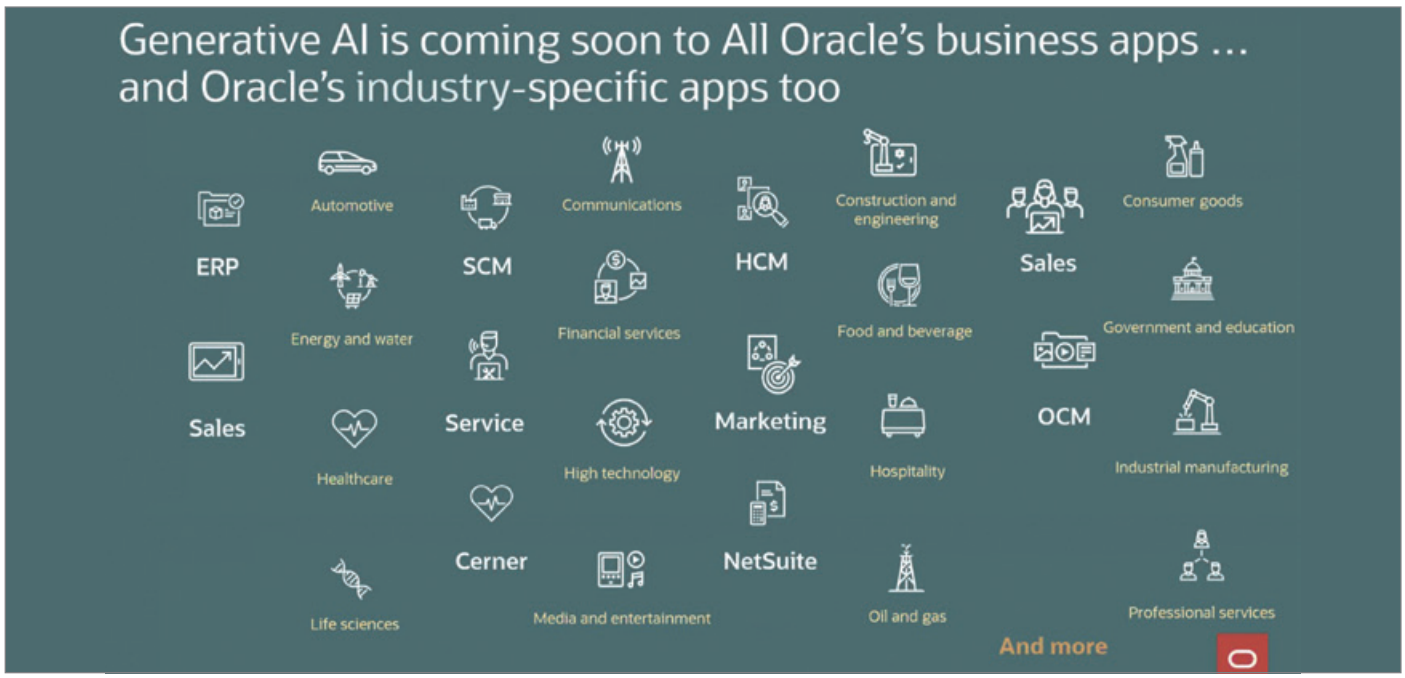


Abbildung 8: Oracles Zukunft in Bezug auf generative KI [1]

ture, steht momentan nur in Nordamerika zur Verfügung wie in *Abbildung 7*, als Ausschnitt aus dem Systemsetup ersichtlich.

Fazit – wie sieht sich Oracle?

Oracle nimmt die Umsetzung der generativen KI in Applikationen ernst und stellt sich, gemeinsam mit Cohere und aufbauend auf der Oracle Cloud Infrastructure, stark auf *[1]*. Der Fokus liegt vor allem darauf, die Funktionalitäten kostenfrei oder mit geringen Kosten zur Verfügung zu stellen, aber auch ein KI-Toolkit für eigene Entwicklungen zur Verfügung zu stellen. Die LLM (Large Language Models) sollen aktuell und leicht anpassbar sein, dabei aber zugleich performant und sicher. Dieser Fortschritt soll in allen Cloud Applikationen und auch industriespezifischen Lösungen spürbar sein *(siehe Abbildung 8)*.

Quellen

[1] OPN Webinar „Bringing Generative AI to the Enterprise“, 27.02.2024, John Menhinick, Michael Nasrany, Oracle. Aufzeichnung über Oracle University via *https://learn.oracle.com/ols/course/bringing-generative-ai-to-the-enterprise/89350/134284*, letzter Zugriff 04.04.2024

[2] Beitrag „Die Geschwindigkeit des Wandels nimmt zu“, 22.09.2023, Katharina Schraft und Marcos López, *https://www.doag.org/de/home/news/die-geschwindigkeit-des-wandels-nimmt-zu/*, letzter Zugriff 14.03.2024 um 08:52.

[3] Pressemitteilung „Oracle to Deliver Powerful and Secure Generative AI Services for Business“, *https://www.oracle.com/de/news/announcement/oracle-to-deliver-powerful-and-secure-generative-ai-service-for-business-2023-06-13/*, letzter Zugriff am 14.03.2024 um 10:50.

[4] Bericht „Data & AI Leadership Executive Survey 2022“, *https://www.wavestone.com/en/insight/data-ai-leadership-executive-survey-2022/*, letzter Zugriff am 14.03.2024 um 08:59.

[5] Bericht „AI-driven applications are building the future of business“, *https://diginomica.com/ai-driven-applications-are-building-future-business*, letzter Zugriff 05.03.2024 um 11:00.

[6] Pressemitteilung „NetSuite 2024.1 ermöglicht die Einführung von KI mit Textverbesserung, Rechnungserfassung und mehr“, *https://www.netsuite.com/portal/assets/pdf/newrelease/embedded-ai-de-de.pdf*, , letzter Zugriff am 05.03.2024 um 11:00.

[7] Webseite „Intelligentes Leitungsmanagement“, *https://www.oracle.com/de/performance-management/platform/#rc30p5*, letzter Zugriff am 05.03.2024 um 13:15.

[8] Artikel „Analysegesteuerte Kultur mit maschinellem Lernen schaffen“, *https://www.oracle.com/de/business-analytics/analytics-platform/#rc30p5*, letzter Zugriff am05.03.2024 um 13:15.

[9] Blogbeitrag „How every analytics user can benefit from machine learning“, *https://blogs.oracle.com/analytics/post/how-every-analytics-user-can-benefit-from-machine-learning*, letzter Zugriff am 14.03.2024 10:30.

[10] Artikel „Text Enhance“, *https://docs.oracle.com/en/cloud/saas/netsuite/ns-online-help/article_2171112518.html#Text-Enhance*, letzter Zugriff am 14.03.2024 um 11:30.

[11] Video „NetSuite Text Enhance Demo: Item Description“, *https://youtu.be/d2jX-R29aB-w?si=Yugum5DBtUOcLcEe*, letzter Zugriff am 14.03.2024 um 10:43.

Über die Autorin

Katharina Schraft ist seit 2010 bei der PROMATIS software GmbH in Ettlingen bei Karlsruhe tätig und hat ihren Schwerpunkt in der ERP-Beratung auf den Bereich Finance gelegt. Seit 2018 ist sie Vice President der BU NetSuite und berät Kunden unter anderem in strategischen Buchhaltungs- und Compliance-Fragestellungen. Seit 2023 leitet Katharina die Business Solutions Community der DOAG.

Katharina Schraft
katharina.schraft@promatis.de

KI als Teamsport mit KI.NRW

Im Interview von Armin Wildenberg, Leiter der Data Analytics Community der DOAG, mit Dr. Christian Temath, Geschäftsführer von KI.NRW, werden die Aufgaben und Projekte von KI.NRW vorgestellt. Des Weiteren erläutert Dr. Temath die Herangehensweise bei der Unterstützung von Unternehmen und gibt eine aktuelle Einschätzung zu den Herausforderungen beim Einsatz von KI.

Herr Dr. Temath, Sie sind Geschäftsführer von KI.NRW. Wann wurde die Kompetenzplattform ins Leben gerufen und wo liegen ihre Hauptaufgabengebiete?

Wir sind sehr froh, dass die Kompetenzplattform KI.NRW bereits Ende 2018 ins Leben gerufen wurde. Die Gründung beruhte auf einem gemeinsamen Beschluss des Wissenschafts- und Wirtschaftsministeriums von NRW. Wir hatten dadurch die Chance, schon lange vor dem Hype um ChatGPT und den rasanten Entwicklungen, die sich in den letzten zwölf bis 18 Monaten ergeben haben, eine Basis aufzubauen und Angebote im Bereich der sinnvollen, erfolgreichen Anwendung von Künstlicher Intelligenz zu entwickeln.

Die Aufgabengebiete von KI.NRW sind im Wesentlichen, dass wir den Transfer von KI-Anwendungen aus der Forschung in die praktische Anwendung begleiten. Insbesondere kleine und mittelständische Unternehmen in Nordrhein-Westfalen unterstützen wir bei ihrer KI-Reise. Diese Begleitung kann durch Informationsveranstaltungen erfolgen oder aber durch ganz konkrete Angebote, wie Workshopformate, dem KI.Schnellstart oder persönliche Beratungen in einer KI.Sprechstunde.

„Wir begleiten kleinere und mittlere Unternehmen auf ihrer KI-Reise“

Gleichzeitig wollen wir das KI-Ökosystem in Nordrhein-Westfalen zusammenbringen, vernetzen und stärken, um dadurch eine Sichtbarkeit über die Grenze von NRW hinaus sicherzustellen und am Ende des Tages auch in der Champions League für Angewandte Künstliche Intelligenz mitzuspielen.

Champions League ist eine schöne Bezeichnung für NRW. Sind Sie mit Ihrer Kompetenzplattform ausschließlich in NRW tätig oder haben Sie auch Tätigkeitsfelder in anderen Bundesländern?

Die grundsätzlichen Angebote für KMU richten wir primär an Unternehmen aus Nordrhein-Westfalen, um im hiesigen Wirtschaftsraum die KI-Reise der hier ansässigen Unternehmen zu unterstützen und zu begleiten. Wir haben ebenso einen starken Fokus darauf, das KI-Ökosystem in NRW, ob Forschungsprojekte, KI-Anwender*innen oder KI-Anbieter*innen, über unsere digitale KI.NRW-Landkarte sichtbar zu machen. Gleichzeitig sind wir mit bundesweit aktiven Initiativen sehr stark vernetzt und im Austausch auf europäischer Ebene, um auch über Deutschland hinaus die Sichtbarkeit von KI-Themen aus Nordrhein-Westfalen zu erhöhen.

Eines unserer Flagship-Projekte, „ZERTIFIZIERTE KI", gehört zu denjenigen Themen, bei denen wir sehr froh sind, dass wir es in Nordrhein-Westfalen führend voranbringen. Damit zeigen wir: NRW kann KI. Gleichzeitig hat dieses Leuchtturmprojekt eine wichtige Wirkung auf andere Bundesländer und strahlt zudem über die Grenzen von Deutschland hinaus.

Sie haben das Projekt „ZERTIFIZIERTE KI" angesprochen: Handelt es sich dabei um ein Vorhaben zur Regulierung oder zur Qualitätssicherung von KI-Anwendungen?

Das Projekt „ZERTIFIZIERTE KI" hat eine lange Historie. Es startete weit vor der Frage, wie eine konkrete Regulierung auf europäischer Ebene aussehen könnte. Es beschäftigte sich zum Beispiel mit der Fragestellung, wie wir eine vertrauenswürdige KI für uns definieren wollen und welche Kriterien sie umfasst. Im Rahmen dieses Projektes wurden dann entsprechende Kriterienkataloge erstellt. Und vor knapp drei Jahren ist ein Prüfkatalog entstanden, der heute konkret von Unternehmen eingesetzt wird, um KI-Anwendungen anhand von verschiedenen Dimensionen der Vertrauenswürdigkeit zu prüfen und zu testen.

„Datenschutz und Datensicherheit sind Teil von vertrauenswürdigen KI-Anwendungen."

Über den Begriff vertrauenswürdig kommen wir schnell zum Datenschutz. In einer aktuellen Umfrage nannten über 80 % der Unternehmen die Themen Datenschutz gefolgt von Fachkräftemangel und Zeit als Hauptgründe für die schleppende Digitalisierung. Gehört Datenschutz zu den Aspekten Ihrer KI-Zertifizierung?

In der Tat, die Themen Datenschutz und Datensicherheit sind auch ein Teil von vertrauenswürdigen KI-Anwendungen. Sie gehören zu den sechs Prüfungsdimensionen.

Die von Ihnen benannten Themen Datenschutz, Fachkräftemangel und Zeit stellen für Unternehmen in Bezug auf ihre KI-Reise eine große Herausforderung dar, insbesondere wenn in den Bereichen Datenschutz und Regulierung Unsicherheiten bestehen. Da taucht dann schnell die Frage auf, wie bewege ich mich richtig und sicher? Oder: Soll ich in das Risiko gehen oder soll ich es lieber lassen, weil ich als Unternehmer*in Angst davor habe, irgendwelche entsprechenden Vorgaben zu verletzen. In diesem Kontext stellen wir fest, dass in Unternehmen oftmals das nötige Wissen fehlt und die Notwendigkeit besteht, grundlegende Themen sauber erklärt und aufbereitet zu bekommen. Das gehört zu den Themen, die uns bei KI.NRW stark bewegen, wie auch die Aufklärung, dass viele KI-Anwendungen mit nichtpersonenbezogenen Daten sehr gut funktionieren können.

Kommen wir zum Faktor Zeit, den sie zurecht ansprechen: Unternehmen fragen häufig, wie viel Zeit sie für eine KI-Lösung investieren müssen? In der aktuellen Situation haben die Unternehmen durchaus mehr als eine Herausforderung auf dem Tisch: Es geht um Energiepreise, um Lieferketten und um Innovationsdruck. Hinzu kommt das Thema KI, das auch vorangetrieben werden soll. Genau hier wollen wir unterstützen und lassen die Unternehmen wissen: So aufwendig muss es gar nicht sein. Ihr könnt auch mit kleinen, aber immer wieder weiteren Schritten eure KI-Reise erfolgreich starten und voranbringen.

Sie möchten kleinen und mittleren Unternehmen helfen, erfolgreich mit KI zu starten und ihre KI-Anwendungen voranzubringen. KI.NRW hat zwei Flagship-Projekte in der Medizin. Gibt es daneben noch andere, die Sie gerade betreiben oder wo Sie gerade Unterstützung anbieten?

Sie haben gerade das Thema Medizin genannt. Da haben wir sicherlich mit dem „Innovative Secure Medical Campus" des Universitätsklinikums Bonn und dem „SmartHospital.NRW" der Universitätsmedizin Essen zwei Flagship-Projekte, die ganz explizit Künstliche Intelligenz im medizinischen Bereich in die Anwendung bringen. Wir haben gleichzeitig aber auch KI-Flagships aus dem Bereich des Handwerks und des Handels. Bei Letzterem soll für kleinere und mittlere Handelsunternehmen der Einsatz von und der Zugang zu KI erleichtert werden. Von daher sind wir in der Flagship-Initiative gut aufgestellt und versuchen damit auch, die Bandbreite der Wirtschaftsstruktur in Nordrhein-Westfalen abzubilden. Denn wir sehen, dass in der Breite die Unternehmen, die KI-Innovationen für sich nutzbar machen wollen, in allen Branchen angesiedelt sind.

Sie sehen KI-Innovationen in allen Branchen. Wie steht es um die öffentlichen Verwaltungen und Behörden?

Das Thema KI in der öffentlichen Verwaltung ist eines, das uns ebenfalls beschäftigt und das wir auch regelmäßig mit verschiedenen Akteuren im Land besprechen. Meines Erachtens gibt es im Bereich der öffentlichen Verwaltung enorme Potenziale, zum einen im Hinblick auf die künftige Verrentungswelle, um Serviceangebote langfristig weiter anbieten zu können. Auf der anderen Seite kann mithilfe von KI die Zugänglichkeit von Diensten für viel breitere Bevölkerungsschichten verbessert werden, etwa wenn es um Sprachbarrieren oder um das richtige Ausfüllen von komplizierten Formularen geht. Insbesondere durch generative KI ergeben sich sehr viele Möglichkeiten, das Zusammenspiel zwischen Bürger*innen und der Verwaltung nochmal wesentlich zu verbessern und eine deutlich höhere Servicequalität zu erreichen.

Jetzt ist das Thema Fachkräftemangel im öffentlichen Sektor mindestens genauso präsent wie bei den Unter-

nehmen Sie haben auch das Thema Verrentung schon angesprochen, aber wie sehen Sie überhaupt die Umsetzungsmöglichkeiten von KI-gestützten Anwendungen beziehungsweise einer stärkeren Digitalisierung?

Sie sprechen da einen sehr wichtigen Punkt an. Beim Thema Digitalisierung beschäftigt mich die Frage, wie denken Unternehmen, wie denken öffentliche Verwaltungen über zur Verfügung stehende Daten nach, an welchen Stellen sind welche Daten hilfreich, um bessere Entscheidungen zu treffen und besseren Service anzubieten? Das ist ein Thema, das bei dem Hype um die aktuellen KI-Entwicklungen sehr wichtig bleibt. Auf einer soliden, guten Datenbasis kann man sehr erfolgreich KI-Anwendungen entwickeln. Auch bin ich der Ansicht, dass die Daten in einer vernünftigen Art und Weise vorliegen müssen und dass sie natürlich auch sicher gespeichert werden. An dieser Stelle spielen die wichtigen Themen Cybersicherheit und Digitalisierung mit der Komponente KI zusammen, um überhaupt diese Potenziale nutzbar zu machen.

„Wie können wir KI als Teamsport verstehen?“

Sicherheit ist ein wichtiger Punkt! Sie haben gerade gesagt, wir müssen darüber nachdenken, wie wir unsere Daten verarbeiten wollen. Ein digitaler Prozess sieht nicht genauso aus wie der manuelle Prozess vorher. Brauchen wir dafür andere Menschen oder, besser formuliert, in den Unternehmen eine andere, neue Rolle für das Thema Prozesse und Digitalisierung, die wir bisher so in der Definition nicht haben?

Meine erste Frage wäre eher, wie können wir KI als Teamsport verstehen? Teamsport setzt bei den Menschen an, die an beziehungsweise in einem bestimmten Prozess arbeiten. Wenn wir wollen, dass sich dieser Prozess durch Digitalisierung unterstützen und effizienter gestalten lässt, ist die wichtigste Frage: Wie schaffen wir es, diese Mitarbeitenden mitzunehmen, die mit diesen Prozessen arbeiten müssen? Denn im Zweifelsfall wird der Prozess durch Digitalisierung oder KI-Unterstützung ein anderer sein. Damit der Einsatz von Künstlicher Intelligenz also erfolgreich wird, ist es wichtig, dass die Mitarbeitenden vollständig mitgenommen werden. Neben der frühzeitigen Einbindung des Teams ist aus meiner Sicht ebenfalls das Thema der Aus- und Weiterbildung essenziell, um mitzuteilen, wie ein KI-Tool unterstützen kann. Aber auch um klarzumachen, was das Tool nicht kann und an welchen Stellen der oder die verantwortliche Mitarbeitende im Arbeitsprozess die Qualität sichern muss.

Die Menschen mitnehmen! Ein schöner Schlussakkord für unser Interview!

Ja, wir sind der festen Überzeugung: Wir wollen den Menschen immer in den Mittelpunkt einer vertrauenswürdigen KI stellen. Um das zu tun, sind insbesondere die Menschen in den Unternehmen, die mit den KI-Tools arbeiten, einzubinden und entsprechend aus- beziehungsweise weiterzubilden. So lässt sich Künstliche Intelligenz in Unternehmen, Behörden und in der Verwaltung erfolgreich zum Einsatz bringen.

Vielen Dank für Ihre Zeit, Herr Dr. Temath.

Über KI.NRW

Die Kompetenzplattform KI.NRW baut Nordrhein-Westfalen zu einem bundesweit führenden Standort für angewandte Künstliche Intelligenz (KI) aus und etabliert das Land in internationalen Netzwerken. Als zentrale Landes-Dachorganisation für Künstliche Intelligenz vereint KI.NRW den Dreiklang aus Spitzenforschung, Innovation und Unternehmertum. Ziel ist es, den Transfer von KI aus der Spitzenforschung in die Wirtschaft zu beschleunigen, eine Leitregion für berufliche Qualifizierung in KI aufzubauen und Impulse im gesellschaftlichen Dialog zu setzen. Dabei stellt KI.NRW den Menschen in den Mittelpunkt einer vertrauenswürdigen KI. Die Kompetenzplattform wird durch die Landesministerien MWIKE und MKW gefördert und geleitet von einem der europaweit führenden Forschungsinstitute auf den Gebieten der angewandten Künstlichen Intelligenz und des Maschinellen Lernens, dem Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS in Sankt Augustin.

DR. CHRISTIAN TEMATH

Dr. Christian Temath ist seit September 2020 Geschäftsführer der Kompetenzplattform KI.NRW und arbeitet am Fraunhofer IAIS in Sankt Augustin mit seinem Team daran, die Marke »KI made in NRW« zu etablieren und die technologische Souveränität des Landes NRW zu stärken. Als promovierter Wirtschaftsinformatiker verfügt er über langjährige Erfahrung in der Managementberatung im Bereich Technologie sowie in der praktischen Anwendung von KI-Technologien bei einem internationalen e-Commerce-Unternehmen. Sein Team aus KI-Manager*innen bündelt Wissen und Erfahrung aus den Bereichen Künstliche Intelligenz, Digitalisierung, Netzwerkarbeit und Beratung mit dem Ziel, Unternehmen, mittelständischen Firmen und Start-ups den Einstieg in die KI zu erleichtern.

Copyright © 2024 Schulz IT Services GmbH

Oracle Database 23c – bigger, better, faster, stronger Teil 2: Die Revolution

Matthias Schulz, Schulz IT Services

Die Oracle Datenbank 23c ist mit über 300 neuen Features eines der umfangreichsten und spannendsten Datenbank-Releases seit langem. Begleiten Sie mich in dieser Artikelserie auf eine Reise zu den großen Neuheiten und vielen kleinen, aber feinen Verbesserungen.

Teil 2 dieser Artikelserie beschäftigt sich mit dem neuen **JSON Datatype** und einer Revolution, die alles verändern wird:

„JSON Relational Duality Views".

1. Die JSON-Revolution

Die wichtigste Neuerung der Oracle Database 23c wird die Welt der Datenbanken fundamental verändern:

> *Oracle's JSON Relational Duality,*
> *a truly revolutionary solution,*
> *is perhaps one of the most*
> *important innovations*
> *in information science in 20 years.*

Carl Olofson, Research VP, Data Management Software, IDC.

Aber was macht JSON Relational Duality Views so revolutionär?

Der Name liefert bereits einige Hinweise: Es geht um die Dualität von JSON **und** relationalen Daten.

Betrachten wir also zunächst diese beiden Technologien sowie ihre Vor- und Nachteile.

1.1 Dokumentenorientierte Datenbanken (JSON)

Dokumentenorientierte Datenbanken, zum Beispiel die MongoDB, finden immer größere Verbreitung. Sie fordern die klassischen relationalen Datenbanken heraus und nehmen ihnen sogar Marktanteile weg.

Aber was macht sie so populär, vor allem unter Anwendungsentwicklern?

Entwickler haben eine genaue Vorstellung davon, wie die Daten für ihr Programm strukturiert und verarbeitet werden sollten und möchten diese Sicht auch beim Lesen und Speichern der Daten vorfinden. Und genau dies bieten dokumentenorientierte Datenbanken.

Ein Programm zur Anzeige und Pflege von Filmen und Schauspielern benötigt nur ein einfaches JSON-Dokument, um voll funktionsfähig zu sein:

- Filme:
  - Out of Sight
    - Schauspieler:
- George Clooney
- Jennifer Lopez
  - Ocean's Eleven
    - Schauspieler:
- George Clooney
- Brad Pitt

Warum also sollte man nicht auch dieses Dokument genauso in einer Datenbank speichern?

Die Probleme dieser Art der Datenspeicherung treten erst später zu Tage, wenn ein zweites (Teil-)Programm hinzukommt, dass die Hierarchie umdreht und Schauspieler und deren Filme bearbeiten soll:

- Schauspieler:
  - George Clooney
    - Filme:
- Out of Sight
- Ocean's Eleven
  - Brad Pitt
    - Filme:
- Out of Sight
- Ocean's Eleven

Die Daten für Schauspieler und Filme sind nun redundant in zwei Dokumentarten gespeichert und müssen bei jeder Änderung überall gefunden und angepasst werden *(siehe Abbildung 1)*.

Und meist kommen bei erfolgreichen Anwendungen, im Laufe der Zeit, immer neue Daten und Sichten hinzu:

- Regisseure, Filme und Schauspieler
- Filmstudios, Filme, Regisseure, Schauspieler
- Länder, Filmstudios, Filme, Schauspieler, Regisseure
- ...

Mit jeder neuen Sicht steigt der Aufwand die redundant, in verschiedenen Dokumenten gespeicherten Daten, zu finden und aktuell zu halten *(siehe Abbildung 2)*.

Wir leiden nun unter genau den Problemen, die zur Entwicklung und zum Erfolg der relationalen Datenbanken geführt haben.

1.2 Relationale Datenbanken

Relationale Datenbanken speichern Daten normalisiert in Tabellen, um eine anwendungsunabhängige, redundanzfreie, flexible und dauerhafte Speicherung der Informationen zu ermöglichen *(siehe Abbildung 3)*.

Den Vorteilen bei der Speicherung und Aufbewahrung der Daten steht jedoch der Aufwand bei der Nutzung der Daten durch Programme gegenüber. Die Tabellen der Datenbank müssen erst mit den Strukturen eines Programmes verbunden werden.

Hierzu kann, um einige der Optionen zu nennen, ein Objekt-relationaler Mapper verwendet, SQL manuell erstellt und gemapped, oder eine JSON-Zwischenschicht programmiert werden.

All diese Optionen bedeuten Entwicklungsaufwand, mögliche Fehlerquellen und bergen nicht zuletzt auch potenzielle Mapping-Probleme.

1.3 Das Beste aus beiden Welten

Beide Modelle haben ihre individuellen Vor- und Nachteile. Das relationale Modell eignet sich besser für die Speicherung der Daten, während das Dokumentenmodell Programmstrukturen besser unterstützt.

Ideal wäre es, die Vorteile beider Modelle nutzen zu können, ohne deren Nachteile und ohne sich um die Konversion zwischen Tabellen und JSON-Dokumenten kümmern zu müssen.

Mit der Version 23c der Oracle Database wird dieser Wunsch Realität!

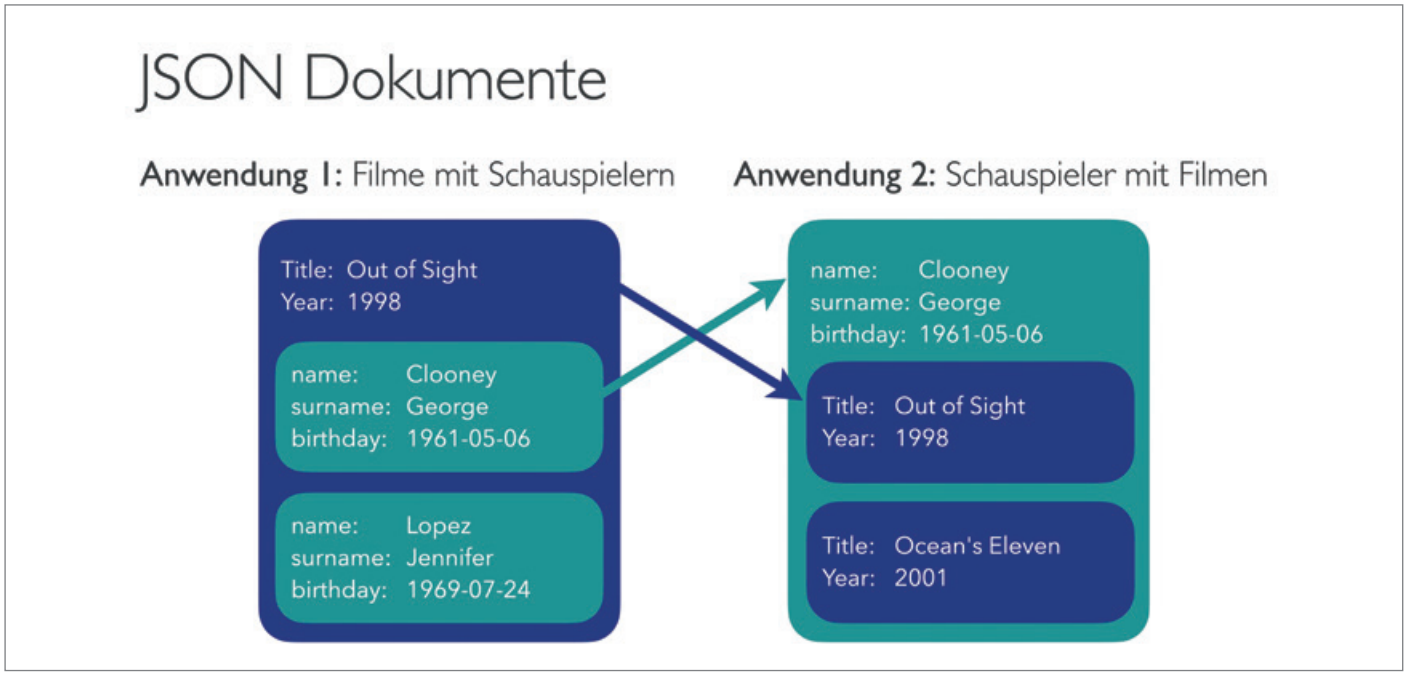


Abbildung 1: JSON-Dokumente (Quelle: Matthias Schulz)

```
CREATE OR REPLACE JSON RELATIONAL DUALITY VIEW V_MOVIES_ACTORS AS
MOVIE @insert @update @delete
  {movieId   : MOVIE_ID
   title     : TITLE
   year      : year
  actors     : ACTOR @insert @update @delete
                [
                  {personID : PERSON_ID
                   movieId  : MOVIE_ID,
                              PERSON @update /* hier nur @update zulässig!!! */
                                {actorId  : PERSON_ID
                                 name     : NAME
                                 surname  : SURNAME
                                 birthday : BIRTHDAY
                                }
                  }
                ]
  };
```

Listing 1: JSON Relational Duality View: Filme mit Schauspielern

1.4 JSON Relational Duality Views

Die JSON Relational Duality Views schließen die Lücke zwischen den relationalen Datenbanken und der Dokumentensicht der Anwendung und das mit minimalem Aufwand und ohne Nachteile!

Die Daten werden normalisiert in relationalen Tabellen gespeichert und für jedes gewünschte JSON-Dokument wird einfach eine View definiert – fertig!

Jedes Programm erhält hierdurch exakt die JSON-Dokumente, die es benötigt und in der jeweils optimalen Form – und das ohne redundante Datenhaltung, manuell erstellte SQL-Befehle oder Objekt-relationale Mapper.

Die Anwendung liest die Daten als JSON-Dokument, ändert dieses nach Belieben und speichert es wieder. Die Daten des Dokuments stammen aus Tabellen, die Änderungen am Dokument landen wieder in diesen Tabellen, und all das komplett automatisch. Selbst Optimistic Locking wird automatisch und ohne Extraspalten bereitgestellt.

1.5 View-Erstellung

Die Views, die diese Magie ermöglichen, lassen sich in einer einfachen Syntax erstellen, die nur die Namen der gewünschten Tabellen und Spalten benötigt.

Um ein JSON-Dokument für „Filme“ zu erstellen,

```
{movieId : MOVIE_ID
 title   : TITLE
 year    : year
};
```

benötigen wir nur diese einfache View:

```
CREATE OR REPLACE
JSON RELATIONAL DUALITY VIEW V_
MOVIE
AS
MOVIE @insert @update @delete
  {movieId   : MOVIE_ID
   title     : TITLE
   year      : year
  };
```

In *Listing 1* sehen Sie die komplette View für „Filme mit Schauspielern“.

In *Listing 2* die View für "„Schauspieler mit Filmen“.

1.6 View-Berechtigungen

In der View-Definition kann für jede verwendete Tabelle einzeln festgelegt werden, welche Datenänderungen erlaubt sein sollen. Um diese Berechtigungen zu erteilen, können die folgenden Annotationen hinter dem Tabellennamen hinzugefügt werden:

- **@insert** – Datensätze hinzufügen
- **@update** – Daten ändern
- **@delete** – Datensätze löschen

1.7 Lesen

Zum Lesen verwenden wir die oben erstellte View V_MOVIES:

```
SELECT * FROM V_MOVIE t;
```

```
{"movieId":10,"title":"Out
of Sight","year":"1998-01-
01T00:00:00","_metadata":{"etag":
"A985DBEC80B51EE99BCB0066C3AA5428
","asof":"0000000000DAFBE6"}}
```

Um eine formatierte Ausgabe des JSON-Dokuments zu erhalten, kann die Funktion **json_serialize** mit der Option **pretty** verwendet werden und mit der **Dot-Notation** (**t.data."title"**) kann direkt auf Keys des Dokuments zugegriffen werden, zum Beispiel, um nur einen Film anzuzeigen:

```
SELECT json_serialize(data
pretty)
   FROM V_MOVIE t
  WHERE t.data."title" = 'Out
of Sight';
```

```
{
  "_metadata" :
  {
    "etag" : "A985DBEC-
80B51EE99BCB0066C3AA5428",
    "asof" :
"0000000000DAFC05"
  },
  "movieId" : 10,
  "title" : "Out of Sight",
  "year" : "1998-01-
01T00:00:00"
}
```

1.8 Optimistic Locking

Die Datenbank fügt unserem JSON-Dokument den Key **„_metadata“** hinzu, welcher die für das Optimistic Locking benötigten Informationen enthält.

Das Attribut **„_etag“** enthält hierfür einen Hash, der aus dem Inhalt des Datensatzes generiert wird, um zu überprüfen,

```
CREATE OR REPLACE JSON RELATIONAL DUALITY VIEW V_ACTORS_MOVIES AS
PERSON @insert @update @delete
  {actorId  : PERSON_ID
   name     : NAME
   surname  : SURNAME
   birthday : BIRTHDAY
   movies   : ACTOR   @update
              [
                {personIdFk : PERSON_ID
                 movieIdFk  : MOVIE_ID,
                              MOVIE @unnest
                                {movieId : MOVIE_ID
                                 title   : TITLE
                                 year    : YEAR
                                }
                }
              ]
  };
```

Listing 2: JSON Relational Duality View: Schauspieler mit Filmen

```
-- bisher: ------------------------

CREATE TABLE T_FRUITS (
  ID         NUMBER GENERATED ALWAYS AS IDENTITY PRIMARY KEY,
  JSON_DATA  CLOB CHECK(JSON_DATA IS JSON)
);

-- ab Oracle 23c: -----------------

CREATE TABLE T_FRUITS (
  ID         NUMBER GENERATED ALWAYS AS IDENTITY PRIMARY KEY,
  JSON_DATA  JSON
);
```

Listing 3: JSON Datatype

```
-- Daten-Eingabe: -----------------
DECLARE
    l_varchar2  VARCHAR2(32767);
    l_clob      CLOB;
    l_blob      BLOB;
    l_json      JSON;
BEGIN
    l_varchar2 := '{"fruit":"apple","quantity":10}';
    l_clob     := '{"fruit":"orange","quantity":20}';
    l_blob     := utl_raw.cast_to_raw('{"fruit":"banana","quantity":30}');
    l_json     := json('{"fruit":"cherry","quantity":40}');

    INSERT INTO T_FRUITS (json_data) VALUES (json(l_varchar2));
    INSERT INTO T_FRUITS (json_data) VALUES (json(l_clob));
    INSERT INTO T_FRUITS (json_data) VALUES (json(l_blob));
    INSERT INTO T_FRUITS (json_data) VALUES (l_json);
    COMMIT;
END;
/

-- Daten-Ausgabe: -----------------
SELECT * FROM T_FRUITS;

ID JSON_DATA
1  {"fruit":"apple","quantity":10}
2  {"fruit":"orange","quantity":20}
3  {"fruit":"banana","quantity":30}
4  {"fruit":"cherry","quantity":40}
```

Listing 4: Verwendung des JSON Datatype

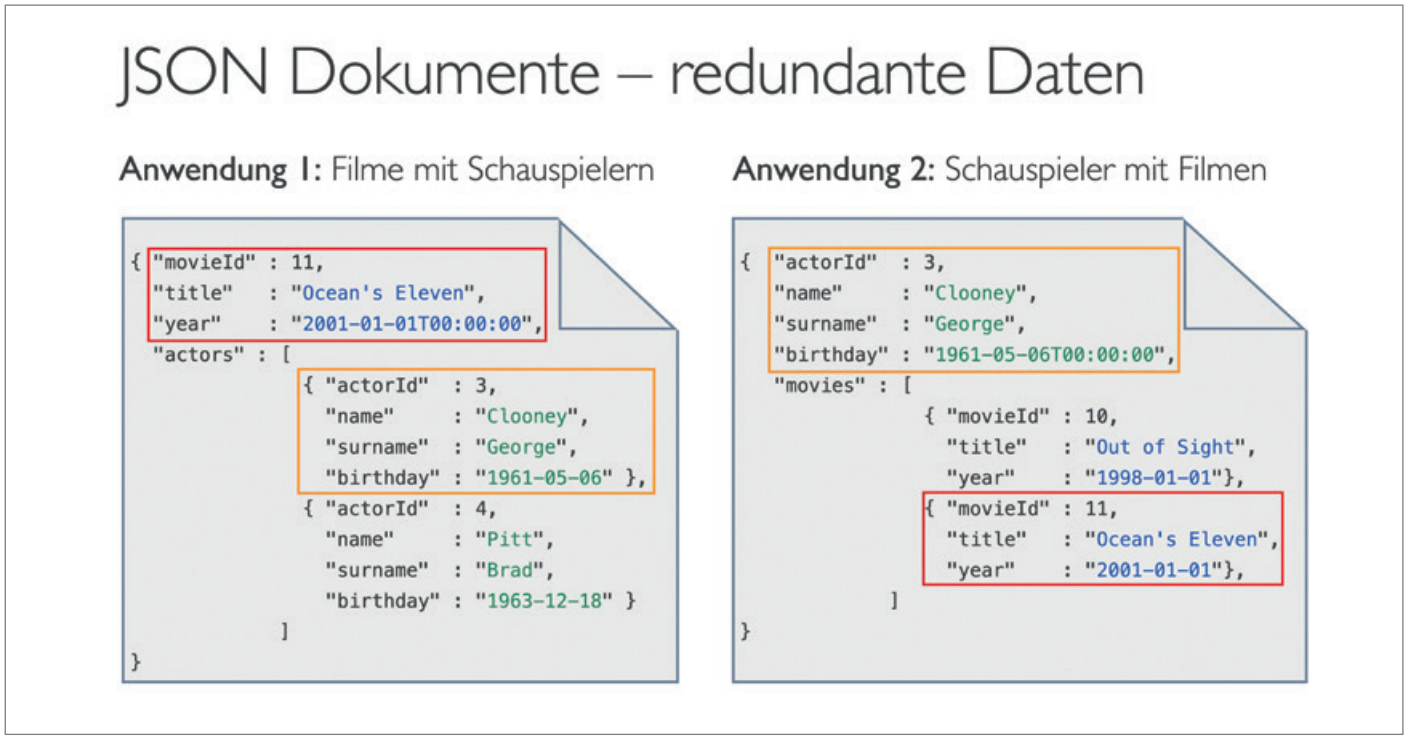


Abbildung 2: JSON-Dokumente – redundante Daten (Quelle: Matthias Schulz)

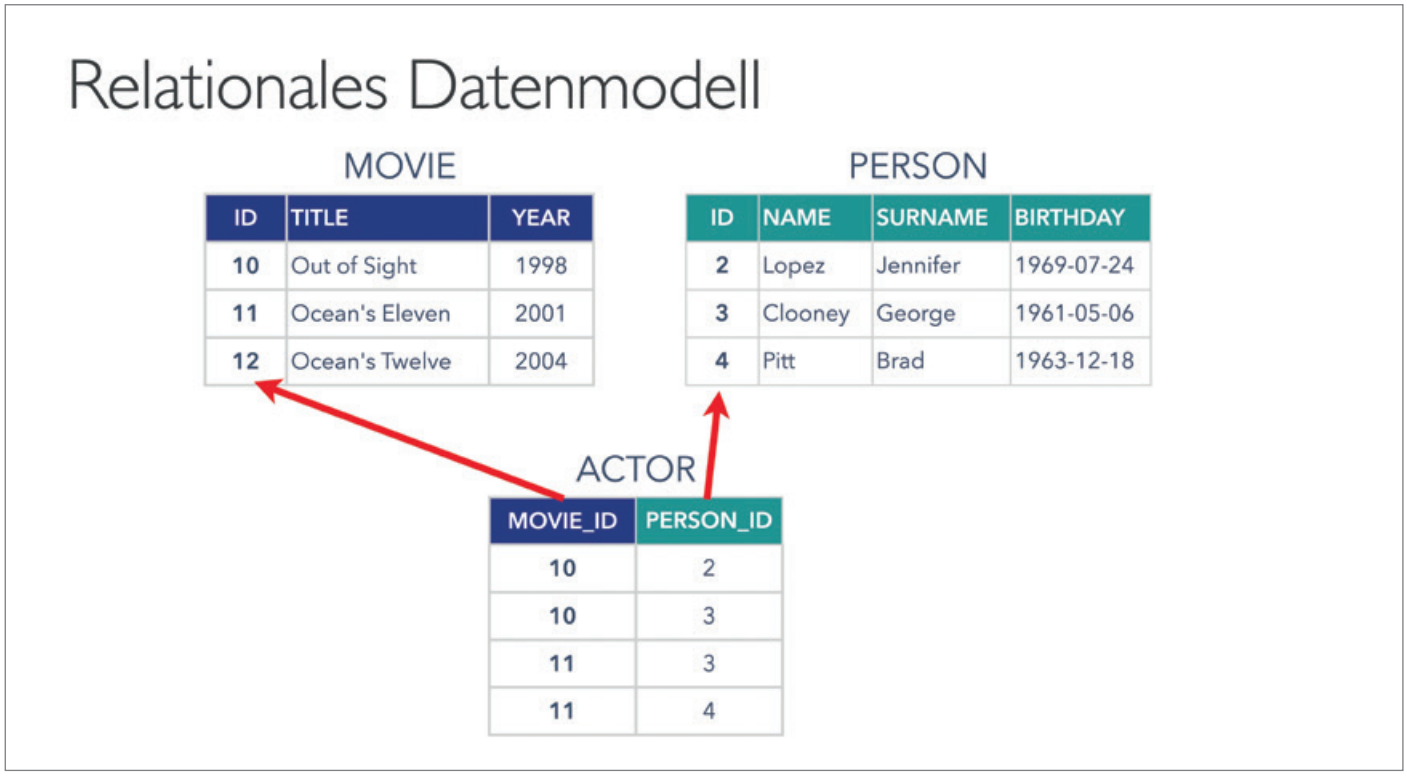


Abbildung 3: Relationales Datenmodell (Quelle: Matthias Schulz)

```
SELECT JSON{*} FROM PERSON;

{"PERSON_ID":2,"NAME":"Lopez","SURNAME":"Jennifer","BIRTHDAY":"1969-07-24T00:00:00"}
{"PERSON_ID":3,"NAME":"Clooney","SURNAME":"George","BIRTHDAY":"1961-05-06T00:00:00"}
{"PERSON_ID":4,"NAME":"Pitt","SURNAME":"Brad","BIRTHDAY":"1963-12-18T00:00:00"}
```

Listing 5: JSON{}*

ob sich der Inhalt des Datensatzes seit der Erstellung des JSON-Dokuments geändert hat.

1.9 Insert, Update, Delete

Zunächst fügt man den Datensatz als JSON-Dokument hinzu:

```
INSERT INTO V_MOVIE (data)
VALUES ('{
  "movieId" : 12,
  "title" : "Ocean''s 12",
  "year" : "2004-01-01"
}');
```

So kann der Datensatz über ein JSON-Dokument aktualisiert werden:

```
UPDATE V_MOVIE t
   SET t.data = ('{
  "movieId" : 12,
  "title" : "Ocean''s Twelve",
  "year" : "2004-01-01"
}')
 WHERE t.data."movieId" = 12;
```

Und schließlich kann man den Datensatz wie folgt löschen:

```
DELETE FROM V_MOVIE t
 WHERE t.data."movieId" = 12;
```

1.10 JSON Columns und Flex Column

Ein Duality View kann neben herkömmlichen Spalten (NUMBER, VARCHAR2, DATE, und Weitere) auch eine beliebige Anzahl von JSON-Spalten enthalten. Zudem kann pro Tabelle eine dieser JSON-Spalten als „Flex Column“ definiert werden.

Eine Flex Column ermöglicht es dem Nutzer des JSON-Dokuments flexibel JSON-Felder hinzuzufügen und neu zu definieren. Dies verleiht einer Duality View und den zugrunde liegenden Tabellen eine gewisse Schemaflexibilität.

2. JSON Datatype

Bisher konnten JSON-Dokumente in der Oracle-Datenbank nur als Text gespeichert werden (VARCHAR2, CLOB, BLOB) gespeichert werden. Mit einem Check-Constraint („IS JSON“) konnte man immerhin sicherstellen, dass die Datenbank weiß, dass es sich um ein JSON-Dokument handelt und, dass nur gültige JSON-Dokumente gespeichert werden können.

Mit Oracle 23c kommt nun der neue binäre Datentyp **„JSON“** hinzu, der den direkten Zugriff auf die Elemente des Dokuments ermöglicht, ohne dieses jedes Mal parsen zu müssen *(siehe Listing 3)*.

Bestehende textbasierte JSON-Daten können mit dem Befehle "„json()“" in den neuen Datentyp umgewandelt werden *(siehe Listing 4)*.

Zusammenfassend lassen sich folgende Vorteile feststellen:

- Binäre Speicherung
- Schnellzugriff auf einzelne Werte
- Schnelle Wert-Updates
- Erweiterte Datentypen (Date, Timestamp, Double, ...)
- Multi-Value-Index

3. JSON{}

Hier ein Beispiel einer Ausgabe der kompletten Zeile als JSON-Dokument *(siehe Listing 5)*.

Ausgabe ausgewählter Spalten pro Zeile als JSON-Dokument *(siehe Listing 6)*.

```
SELECT JSON{NAME, SURNAME} FROM PERSON;

{"NAME": "Lopez",   "SURNAME": "Jennifer"}
{"NAME": "Clooney", "SURNAME": "George"}
{"NAME": "Pitt",    "SURNAME": "Brad"}
```

Listing 6: JSON{} – ausgewählte Spalten

Oracle 23c free

Durch die Verfügbarkeit der kostenlosen Version "„Oracle Database 23c free“" *(https://www.oracle.com/database/free/)*, können alle in dieser Artikelserie vorgestellten Features sofort ausprobiert werden.

Das jeweils neueste Docker Image der Oracle 23c free kann direkt mit dem folgendem Docker-Befehl geladen werden:

docker pull container-registry.oracle.com/database/free:latest

Über den Autor

Matthias Schulz ist seit 2002 Geschäftsführer der Schulz IT Services GmbH und als Consultant, Trainer und Konferenzsprecher im Datenbankumfeld tätig. An der Dualen Hochschule Baden-Württemberg ist er seit 2016 Dozent für Informatik (Datenbanken).

Tätigkeitsschwerpunkte sind Data Vault 2.0, High-End-Systeme (Exadata, Exasol), Oracle Cloud und die Oracle Datenbank (Entwicklung, Tuning, Architektur, ETL, Testing und Java in der Datenbank).

Matthias Schulz
schulz@schulz-it-services.de

CloudLand
WWW.CLOUDLAND.ORG

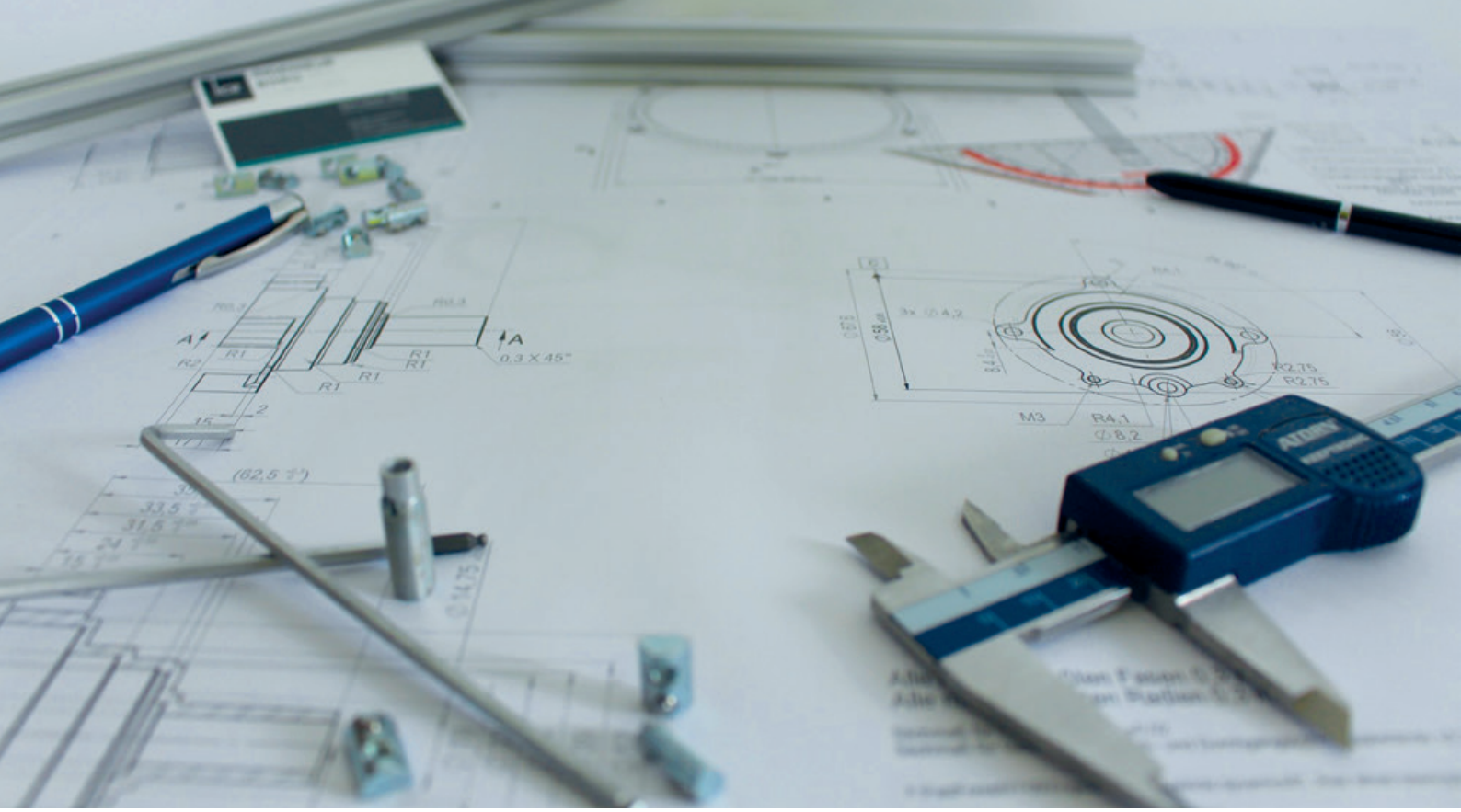
DAS CLOUD NATIVE FESTIVAL

ON DEMAND

CLOUDLAND VERPASST?

JETZT ON-DEMAND-TICKET BUCHEN UND VORTRAGSAUFZEICHNUNGEN ANSCHAUEN!

ALLE ANGEBOTE IM ON-DEMAND-TICKET-SHOP



Typische Fehler des Optimizers

Lothar Flatz

Manche meiner Kunden sind geradezu beleidigt, wenn der Oracle Optimizer einen Fehler macht. Es scheint, ihren Glauben an die Technik zu stören.In seinem Aufsatz *Is Query Optimization a "Solved" Problem? [1]* bezieht Guy Lohman einen gegenteiligen Standpunkt. Er fragt: *„Why would the optimizer ever pick a decent plan?"* Es ist in der Tat ein hoher Anspruch, zu jeder beliebigen SQL-Abfrage in jedem Datenmodell innerhalb von Sekundenbruchteilen einen optimalen Plan zu erzeugen. Das sollte man sich immer vor Augen halten, wenn man den folgenden Text liest. Tatsächlich sind die grundlegenden Konstruktionsbausteine des Optimizers leicht verständlich. Man sollte aber nicht in den Fehler verfallen, das wäre schon alles, was es zu wissen gäbe. Lieber sollten wir uns gemeinsam darüber freuen, dass das Wesentliche des Optimizers einfach und verständlich ist.

Kardinalitätsschätzungen

Der Optimizer hat die Aufgabe, zu jedem beliebigen SQL-Statement einen optimalen (daher Optimizer) Execution-Plan zu erstellen. Der Execution-Plan definiert, wie auf die abzufragenden Tabellen zugegriffen werden soll und wie das Resultat dieser Zugriffe verknüpft werden soll.

Die Art des Zugriffes und der Verknüpfung wird wesentlich durch die erwarteten Datenmengen bestimmt. Wird eine Ergebnismenge im Bereich von 5% oder weniger der Tabellengröße erwartet, so ist ein geeigneter Index (sofern er existiert) das Mittel der Wahl. Wird eine weitaus größere Ergebnismenge erwartet, so wäre ein *full table scan* wesentlich effizienter. Für das gezielte Verknüpfen von kleinen Ergebnismengen ist der *Nested Loop Join* das optimale Mittel, für große Ergebnismengen wäre ein *Hash Join* besser. Diese Grundsätze gelten nicht uneingeschränkt, haben sich aber in der Praxis als Anhaltspunkte durchaus bewährt. Tatsächlich berechnet der Optimizer für jeden Zugriff fiktive Kosten, jedoch sind diese Kosten stark von den Ergebnisschätzungen abhängig.

Als Mensch ist man meist gut damit beraten, die intransparenten Kosten, welche zum Beispiel von meist fiktiven CPU-Geschwindigkeiten abhängig sind, zu ignorieren und sich an die klaren und einfach zu verstehenden Kardinalitätsschätzungen zu halten.

Um es mit den Worten des Mathematikers und Performanceexperten Wolfgang Breitling zu sagen: *„The costs are a bypro-*

duct, derived from the cardinality estimates. If those estimates are right, the CBO does an excellent job of finding the optimal plan. Conversely, whenever the plan is wrong it can be tracked down to incorrect cardinality estimates." [3]

Herausforderungen beim Erzeugen eines Execution-Planes

Um den Execution-Plan zu erzeugen, hat der Optimizer nur Sekundenbruchteile Zeit. Auch wenn die Aufgabe komplex ist, muss sie doch schnell erledigt werden können. Manche Fakten sind auch erst zur Laufzeit bekannt, dennoch muss der Plan vor der Durchführung erzeugt werden.

Das zwingt den Optimizer zu Abkürzungen und Vereinfachungen, die manchmal zu fehlerhaften Plänen führen. Einige Wissenslücken des Optimizers müssen auch durch vereinfachende Annahmen geschlossen werden. Mit der gegenwärtigen Technologie kann man wenig dagegen machen. Der Optimizer hat nur beschränkt Zeit zur Verfügung und die Schwächen sind daher unvermeidlich. Ich denke, dass die Künstliche Intelligenz uns hier weiterbringen wird, aber noch ist es nicht so weit.

Schwächen des Optimizers

Es ist schon länger bekannt, dass es einige charakteristische Schwächen des kostenbasierten Optimizers gibt, was an der Aufgabenstellung und der Architektur der Optimizer liegt. Interessant ist dabei, dass die Probleme anscheinend datenbankübergreifend existieren.

Dr. Guy Lohmann, der selbst am IBM DB2 Optimizer mitgearbeitet hat, nennt 2014 die folgenden grundsätzlichen Probleme des Optimizers:

- Suchvariablen von schwankender Qualität
- Selektivität von Join-Variablen
- Korrelation von Spalten
- Redundante Prädikate

Ich möchte noch einen weiteren Punkt hinzufügen:

- Berechnete Ausdrücke (ohne Statistik)

Wolfgang Breitling [2], der schon 2003 zu diesem Thema sehr fundiert geschrieben hat, stimmt im Wesentlichen Dr. Lohmann zu, obwohl er über den Oracle Optimizer schreibt. Als Mathematiker hat Breitling das Thema etwas abstrakter dargestellt als Dr. Lohmann. Im Endresultat meinen beide oft das Gleiche, ich habe mich hier aber mehr an die etwas konkretere Darstellung von Dr. Lohmann gehalten.

Detaillierte Darstellung der Schwachpunkte

Unterschiedlich gute Suchkriterien in Bindevariablen

Im Folgenden ist es sehr wichtig, sich vor Augen zu halten, dass der Optimizer die Pläne nicht zur Laufzeit der Abfragen, sondern vorher erstellt. In diesem Sinne kennt der Optimizer die Inhalte der Suchvariablen auch nicht, sondern muss einen Kompromissplan entwerfen, der auf alle möglichen Suchbedingungen passt. Es ist im Optimalfall schon möglich, Kategorien von Plänen zu bilden. Zum Beispiel Pläne für Suchbedingungen, die in wenigen gefundenen Datensätzen resultieren, oder solchen, die viele Sätze als Resultat haben. Der Optimizer kann natürlich mehrere Pläne erzeugen, die je nach Suchwert gewählt werden können. Man nennt solche Pläne „bind sensitiv". In der Praxis ist das jedoch sehr schwierig und funktioniert nicht immer. Wenn dann der erzeugte Plan mit einem konkreten Suchwert durchgeführt wird, kann es immer wieder dazu kommen, dass der Plan nicht auf den Suchwert passt. Das ist insbesondere dann der Fall, wenn abhängig von den Suchwerten die Mengen der Resultate stark schwanken. Dies hängt wieder vom konkreten Geschäft des Kunden ab.

Nehmen wir beispielsweise die Schweizer Postbank, die einen großen Teil des Zahlungsverkehrs in der Schweiz abwickelt. Wenn man die Anzahl der Zahlungen eines Kunden schätzen will, kommt es entscheidend darauf an, wer der Kunde ist.

Der Kunde könnte eine Privatperson, aber auch eine große Handelskette wie Migros oder Coop sein. Je nachdem müssen die Pläne ganz unterschiedlich sein.

In der Oracle-Datenbank hilft es natürlich, wenn auf der Spalte Kundennummer in der Tabelle Zahlung ein Histogramm existiert. Das ist auch die in diesem Fall angebrachte Gegenmaßnahme.

So hat die Datenbank zumindest eine Chance, zwischen kleineren und größeren Kunden zu unterscheiden. Ich war als Berater bei der Schweizer Postbank tätig und weiß, dass auch das nicht immer wie gewünscht funktioniert.

Selektivität von Join-Variablen

Noch schlimmer wird es im obigen Beispiel, wenn wir nicht anhand der Kundennummer, sondern anhand des Kundennamens suchen.

Dazu muss die Kundennummer erst zu Laufzeit aus der Kundentabelle ermittelt werden. Zum Zeitpunkt der Planerstellung kennt man höchstens den Kundennamen, die Kundennummer jedoch nicht. Das Histogramm existiert aber für die Kundennummer und nicht für den Kundennamen. Das heißt, es existiert noch nicht einmal die theoretische Möglichkeit zwischen kleinen und großen Kunden bezüglich der Zahlungen zu unterscheiden.

Der Optimizer kann hier nur mit Durchschnittswerten arbeiten, die im Extremfall falsch sein werden.

Die einzige Möglichkeit ist hier, dass SQL in zwei Schritten auszuführen. In einem ersten Statement ermittelt man die Kundennummer, um dann im nächsten Schritt damit auf die Zahlungen zuzugreifen

Korrelation von Spalten

Grundsätzlich wird die Unabhängigkeit von Suchkriterien unterstellt. Das heißt, wenn man zwei Suchkriterien hat, sind nach dieser Annahme beide voll und gleich wirksam. In der Realität ist das aber nicht immer richtig.

Sehen wir uns dazu ein Beispiel an: Unter den Neuzulassungen von Autos in Österreich im Januar 2019 betrug der Anteil von VW 17%. Der Anteil des Golf betrug 4%.

Wie hoch ist der Anteil von Hersteller = VW und Marke = Golf?

Die Datenbank würde wie folgt rechnen: 4% von 17% sind 0.68 %. Wir Menschen wissen es besser, es sind natürlich 4%, weil alle Golfs immer von VW stammen. Die Suchkriterien sind also in diesem Beispiel voneinander abhängig.

```
-------------------------------------------------------------------------------
| Id  | Operation          | Name | Rows  | Bytes | Cost (%CPU)| Time     |
-------------------------------------------------------------------------------
|   0 | SELECT STATEMENT   |      |       |       |     3 (100)|          |
|   1 |  SORT AGGREGATE    |      |     1 |     3 |            |          |
|*  2 |   TABLE ACCESS FULL| EMP  |     7 |    21 |     3   (0)| 00:00:01 |
-------------------------------------------------------------------------------

Predicate Information (identified by operation id):
---------------------------------------------------

   2 - filter("DEPTNO">20)
```

Tabelle 1 : Ausführungsplan zu Abfrage 1

```
-------------------------------------------------------------------------------
| Id  | Operation          | Name | Rows  | Bytes | Cost (%CPU)| Time     |
-------------------------------------------------------------------------------
|   0 | SELECT STATEMENT   |      |       |       |     3 (100)|          |
|   1 |  SORT AGGREGATE    |      |     1 |     3 |            |          |
|*  2 |   TABLE ACCESS FULL| EMP  |     1 |     3 |     3   (0)| 00:00:01 |
-------------------------------------------------------------------------------

Predicate Information (identified by operation id):
---------------------------------------------------

   2 - filter("DEPTNO"+1>21)
```

Tabelle 2 : Ausführungsplan zu Abfrage 2

Solche Abhängigkeiten sind nicht die Regel, kommen aber doch recht häufig vor. So hatte ich einmal einen Fall bei einem Mobilfunkanbieter, bei dem die Anzahl der Anrufe pro Kunde und Mobilfunknummer in einem Monat geschätzt werden sollten. In den allermeisten Fällen hat ein Kunde genau eine Mobilfunknummer und daher genügt das eine oder das andere für die Suche, jedoch ist das leider nicht immer der Fall. Es kommt vor, dass abgemeldete Nummern an neue Kunden vergeben werden oder Firmenkunden viele Mobilfunknummer zu einer Kundennummer haben.

Da der Optimizer die Mobilfunknummer und die Kundennummer als unabhängige Suchkriterien betrachtete, hat er die Menge der Anrufe dramatisch zu niedrig geschätzt – mit dem Resultat eines massiv schlechten Ausführungsplanes.

Glücklicherweise gibt es seit Version 11 die Möglichkeit, Spaltenkorrelationsstatistiken zu erzeugen, seit 12.1 können sie auch automatisch erzeugt werden *[5]*, *[4]*.

Redundante Prädikate

Dies tritt auf, wenn in einer Abfrage die gleichbedeutende Suchbedingung zwei Mal auftritt. Die einfachen Fälle kann der Optimizer auch von selbst erkennen. Damit ein Problem auftritt, müssen die zwei Bedingungen sich schon etwas unterscheiden.

Im Kontext des stark verbreiteten SCOTT-Schema wäre es zum Beispiel die Abfrage:

```
select * from emp where mgr is
null or JOB='PRESIDENT';
```

Ein solcher Fall ist in der Praxis relativ selten, ich habe ihn nur einmal erlebt. Im Prinzip kann man das als einen Spezialfall der Korrelation von Spalten *(siehe oben)* betrachten.

Ich habe den Verdacht, dass Guy Lohmann den Punkt nur deshalb getrennt aufführt, um von einem eigenen Tuning-Erfolg zu berichten. Wer will, kann das Beispiel gerne bei ihm nachlesen *[1]*.

Berechnete Ausdrücke (ohne Statistik)

Eine grundsätzliche Schwäche, wenn auch vielleicht kein Fehler, des Optimizers ist es, dass er grundsätzlich keine berechneten Ausdrücke schätzen kann.

Hier ein kleines Beispiel aus dem SCOTT-Schema:

Abfrage 1

```
select count(*) from emp where
deptno > 20;
```

(Ausführungsplan zu Abfrage 1 *siehe Tabelle 1)*

Abfrage 2

```
select count(*) from emp where
deptno+1 > 20+1;
```

(Ausführungsplan zu Abfrage 2 *siehe Tabelle 2)*

Gesucht wird beides Mal nach der Anzahl der Mitarbeiter mit einer Abteilungsnummer größer als 20. Die richtige Antwort ist 6.

Während bei der ersten Abfrage eine einigermaßen vernünftige Schätzung resultiert, schätzt der Optimizer bei Abfrage 2 einen einzigen Datensatz (*siehe* Spalte Rows).

Zu der berechneten Spalte deptno +1 gibt es keine Statistiken, deshalb wird einfach ein default-Wert verwendet, meistens ist es 1% der Tabellengröße.

(Anmerkung: Das Resultat der Schätzung des Optimizers wird immer auf mindestens einen Datensatz aufgerundet.)

Der Entwickler trägt die Verantwortung dafür, seine Abfragebedingungen so zu schreiben, dass der Optimizer die Statistiken verwenden kann.

Allerdings gibt es auch bei berechneten Spalten die Möglichkeit, spezielle Statistiken zu erzeugen *[6]*.

Zusammenfassung

Ich habe in diesem Text, ohne Anspruch auf Vollständigkeit, die wichtigsten Stolpersteine des Oracle Optimizers gelistet.

Es empfiehlt sich, die Grenzen des Optimizers zu kennen, wenn man gute Execution-Pläne für seine Abfragen möchte. Der Optimizer kann die Komplexität der Realität nicht abbilden und daher wird es immer Schwächen in den Plänen geben. Wie Dr. Lohmann bin ich davon nicht enttäuscht, sondern eher erfreut, dass immer noch so oft gute Pläne erzeugt werden.

Quellen

[1] Guy Lohman, „Is Query Optimization a "Solved" Problem?", ACM Sigmod Blog, *https://wp.sigmod.org/?p=1075*

[2] Wolfgang Breitling, Fallacies of the Cost Based Optimizer, *http://www.centrexcc.com/*

[3] Wolfgang Breitling, Fallacies of the Cost Based Optimizer, Überblicksseite, *http://www.centrexcc.com/papers.html*

[4] Tim Hall, Automatic Column Group Detection in Oracle Database 12c Release 1 (12.1), *https://oracle-base.com/articles/12c/automatic-column-group-detection-extended-statistics-12cr1*

[5] Tim Hall, Multi-Column (Column Group) Statistics, *https://oracle-base.com/articles/11g/statistics-collection-enhancements-11gr1#multi_column_statistics*

[6] Tim Hall, Expression Statistics, *https://oracle-base.com/articles/11g/statistics-collection-enhancements-11gr1#expression_statistics*

Über den Autor

Ich habe im Jahr 1989 auf Version 5 meine erste Oracle-Datenbank-Schulung bekommen. Ich war 15 Jahre Mitarbeiter der Oracle Corp., darunter zwei Jahre als Mitglied der Real World Performance Group als einer von fünf Europäern.

Als Berater war ich in Europa und Asien tätig. Meine Kunden waren unter anderem das IOC, die FIFA, die WTO, die UNO, CERN und diverse Großbanken. Ich bin Oracle ACE und Mitglied in der Oaktable.

Lothar Flatz
l.flatz@bluewin.ch

Oracle Datenbanken Monthly News

Auf dem deutschsprachigen Oracle-Blog ist die Mai-Ausgabe der News-Serie erschienen.

DOAG Online

Es ist wieder so weit: die neue Ausgabe ist online! Das sechsköpfige Redaktionsteam von Oracle Deutschland hat wieder Neuigkeiten rund um die Oracle-Datenbank für On-Premises und Cloud-Installation zusammengestellt.

Alles wird wieder in einem Video präsentiert.

In der aktuellen Ausgabe wird wieder ein zusätzliches Quick Link Posting (in Englisch) zur Verfügung gestellt, um einen schnellen Zugriff auf die zugehörigen Links zu gewährleisten.

https://www.doag.org/de/home/news/oracle-datenbanken-monthly-news-34/

DWH-Ops – the new normal

Dr.-Ing. Harald Philipp Gerhards, infologistix

DevOps-Vorgehensweisen und cloudnative Architekturansätze lassen sich auch auf das Data Warehouse (DWH) anwenden. Diese Kombination ermöglicht die Automatisierung von Prozessen, Self-Service-Ansätze und Test Driven Development. Höhere Geschwindigkeit und mehr Flexibilität sind das Ergebnis. Durch die zusätzliche Integration von automatisch arbeitenden Test-Daten-Generatoren kann die Prozesseffizienz noch weiter gesteigert werden.

Seit 2018 haben wir bei infologistix den Weg unserer Kunden weg von einzelnen IT-Systemen und klassischen BI-Architekturen hin zu vereinheitlichten Plattformen begleitet. Da aber über 80 % unserer Kunden nicht oder nicht ausschließlich in der public Cloud unterwegs sind, war es nicht verwunderlich, dass schon ein Jahr später der Wunsch nach der automatischen Provisionierung und Verwaltung von Oracle- Datenbanken aufkam. Unsere Antwort darauf ist das Konzept, dass ich hier vorstellen möchte, welches wir über die Jahre immer weiter entwickelt haben.

Die Multitenant-Architektur als neuer Standard

Die Grundlage dafür legte die Einführung der Multitenant-Architektur der Oracle- Datenbanken bei unseren Kunden. Der Umstieg von klassischen Oracle-Datenbanken zur Multitenant-Architektur stellt eine signifikante Veränderung in der Verwaltung und dem Betrieb von Datenbankressourcen dar. Mit Oracle 12c wurde die Multitenant-Architektur eingeführt *[1]*. In Version 21c wurde die klassische Single-Datenbank-Architektur offiziell aus dem Support genommen, wodurch die Multitenant-Architektur zum neuen Standard avancierte *[2]*. In dieser Architektur werden mehrere Pluggable Databases (PDBs) innerhalb einer Container Database (CDB) verwaltet, was eine effizientere Ressourcennutzung und eine vereinfachte Administration ermöglicht.

ORDS macht PDB-Operationen automatisierbar

Die Automatisierung von Oracle PDBs kann durch den Einsatz von Oracle REST Data Services (ORDS) erreicht werden,

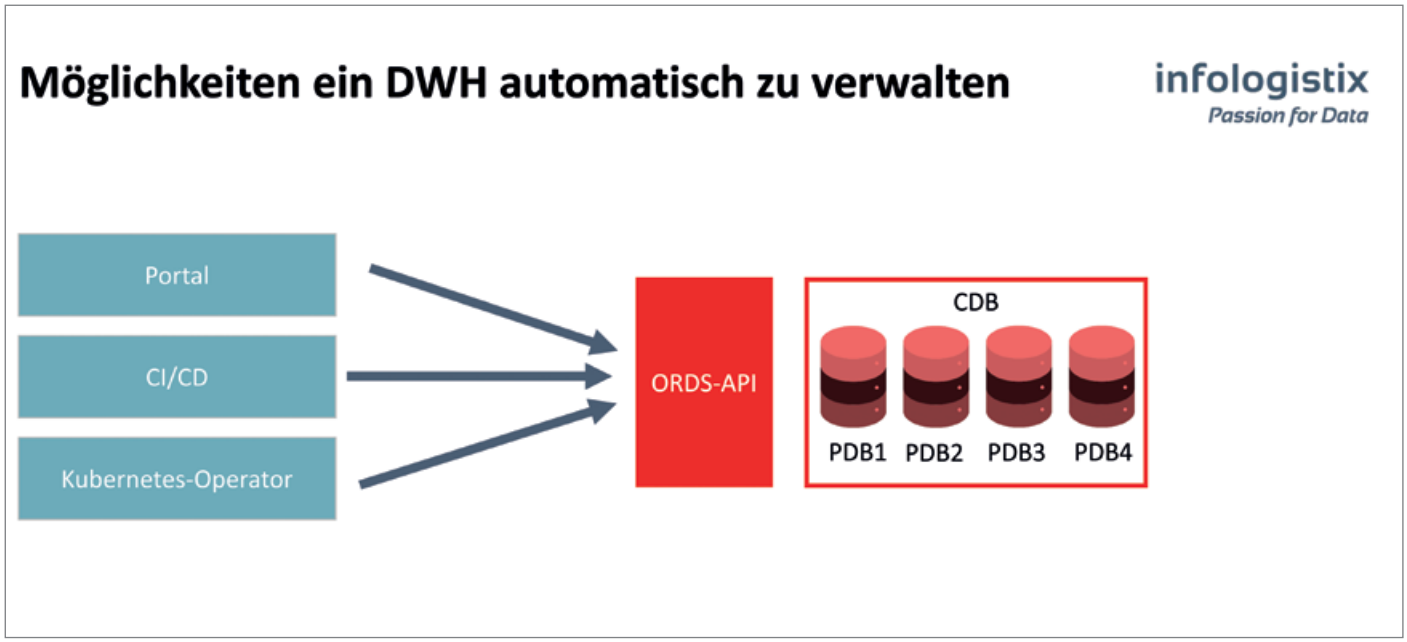


Abbildung 1: Möglichkeiten ein DWH über ORDS automatisch zu verwalten (Quelle: infologistix GmbH)

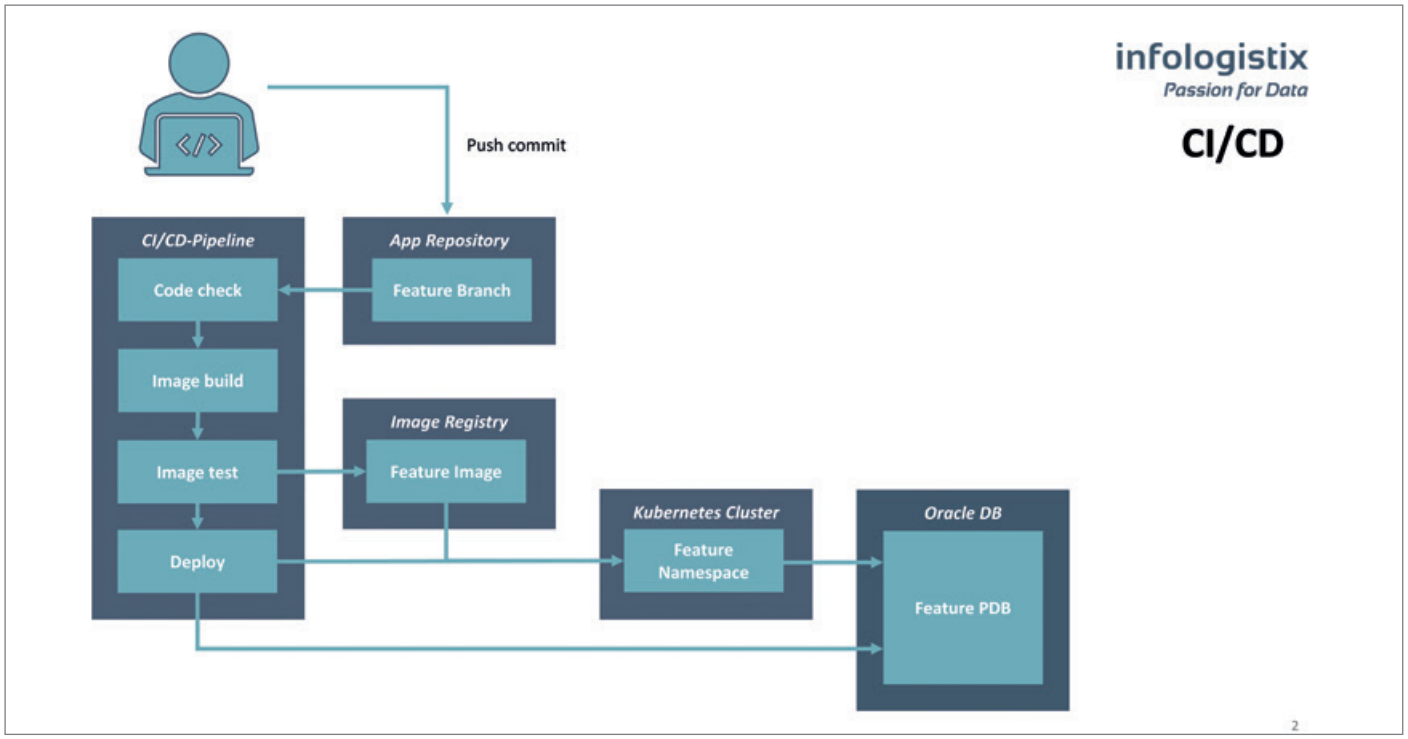


Abbildung 2: Deployment einer Änderung einer Kubernetes-Workload und einer zugehörigen über ORDS automatisierten PDB mit einer CI/CD-Pipeline (Quelle: infologistix GmbH)

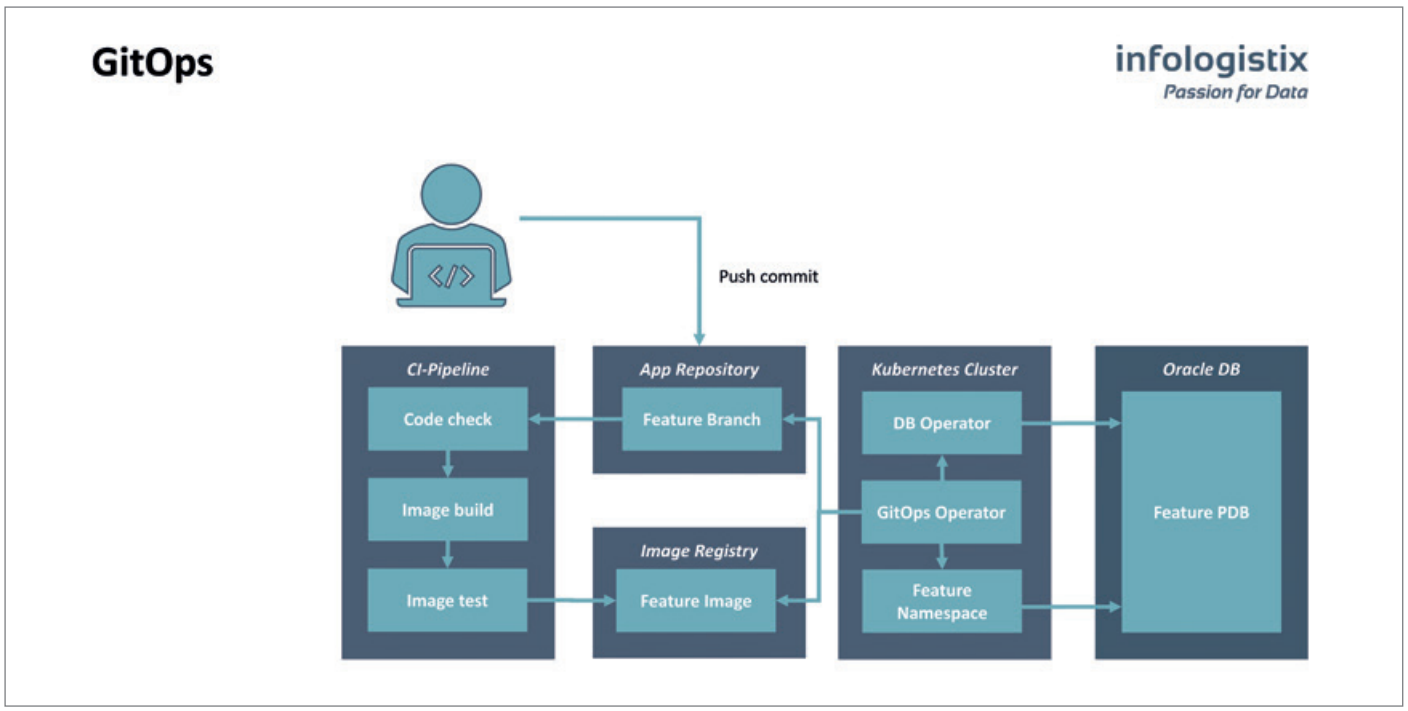


Abbildung 3: Deployment einer Änderung einer Kubernetes-Workload und einer zugehörigen über ORDS automatisierten PDB durch einen GitOps- und einen Datenbank-Operator (Quelle: infologistix GmbH)

welche ab der Oracle-Datenbankversion 19.1 die Steuerung des PDB-Lifecycles unterstützt *[3]*. ORDS ermöglicht es, eine Vielzahl von Datenbankoperationen über eine REST-API durchzuführen, wie beispielsweise das Anlegen, Klonen und Löschen von PDBs sowie das Überwachen von Performance und Sessions *(siehe Tabelle 1)*.

Abbildung 1 zeigt die verschiedenen Möglichkeiten die ORDS-API aufzurufen. Neben dem manuellen Aufruf über ein grafisches User Interface (Portal) sind auch automatische Aufrufe der API möglich. Beispielsweise kann der Aufruf über eine CI/CD-Pipeline imperativ *(siehe Abbildung 2)* oder über einen Kubernetes-Operator deklarativ *(siehe Abbildung 3)* erfolgen.

Da die gesuchte Lösung sich in die GitOps-Automatisierung der Entwick-

API Call	Effekt
create_service(p_admpwd IN VARCHAR2, p_ziel IN VARCHAR2) RETURN VARCHAR2;	Erzeugt eine neue PDB mit Namen und Passwort für Administrator.
clone_service(p_quelle IN VARCHAR2, p_ziel IN VARCHAR2) RETURN VARCHAR2;	Erzeugt eine neue PDB auf Basis einer 1:1-Kopie einer PDB.
clone_snapservice(p_quelle IN VARCHAR2, p_ziel IN VARCHAR2) RETURN VARCHAR2;	Erzeugt eine neue PDB auf Basis einer existierenden PDB. Speichert Änderungen als Differenz in p_ziel und liest Originaldaten aus.
drop_service(p_service IN VARCHAR2) RETURN VARCHAR2;	Die angegebene PDB wird restlos gelöscht und aus der Whitelist ausgetragen.
move_service(p_quelle IN VARCHAR2, p_ziel IN VARCHAR2) RETURN VARCHAR2;	Benennt eine existierende PDB um.
exist_service_as_string(p_service IN VARCHAR2) RETURN VARCHAR2;	Prüft, ob es eine PDB mit angegebenem Namen schon gibt.
get_service_list_all RETURN VARCHAR2;	Liefert die kommagetrennte Liste der Namen aller PDBs.
get_host_port RETURN VARCHAR2;	Liefert den Host und Port des provisionierten Services zurück.
is_job_alive RETURN VARCHAR2;	Healthcheck für die API. Liefert das aktuelle Datum zurück.
is_service_up(p_service IN VARCHAR2) RETURN VARCHAR2;	Prüft den Status einer PDB (mounted, read only, read write).
set_state(p_service IN VARCHAR2, p_state IN VARCHAR2) RETURN VARCHAR2;	Setzt den Zustand eines Service (read write, read only).
get_awr(p_service IN VARCHAR2, p_begin_point IN NUMBER, p_end_point IN NUMBER,) RETURN CLOB;	Liefert einen Automatic Workload-Report (AWR) der PDB zu dem angefragten Zeitraum in Stunden zurück.

Tabelle 1: Beispiele für DB-Operationen über ORDS

lungsteams einfügen sowie deklarativ und cloud-nativ sein sollte, wurde entschieden, einen Kubernetes-Operator zu entwickeln. Dieser Datenbank-Operator (wie auch seit 2022 der Oracle Database Operator (OraOperator)) erweitert die Kubernetes-API mit benutzerdefinierten Ressourcen und Controllern, um das Management von Oracle-Datenbanken kubernetes-nativ durch deklarative YAML-Files zu ermöglichen. Ein Beispiel eines YAML-Files für das Klonen einer PDB gibt *Abbildung 4*.

Dabei enthält dieses Custom Ressource-YAML folgende Informationen:

- Name: Name der Ressource
- Maintanier: Referenz Ansprechpartner
- Description: Anlass für die Instanz
- sourceDB: Name der Klonvorlage
- thinProvisioned: Vollständige Kopie oder differenziell zur Vorlage verwaltet
- lastRollbackTimestamp: Letztes Rollback
- serviceName: Name des DB-Services
- deadline: Datum automatische Löschung
- userStorageInit: Initial allokierte Bytes für User-Table
- userStorageMax: Max. allokierte Bytes für User-Table
- tempStorage: Storage für Temp-Tables
- noOfSnapshots: Zahl gespeicherter Snapshots

DevOps fürs DWH = DWH-Ops

Durch diese Automatisierung konnte die Integration der Datenbankerstellung und -verwaltung in die DevOps-Automatisierung der Entwicklerteams erreicht und sogar die Vorzüge einer GitOps-Automatisierung für die Verwaltung der Datenbanken ermöglicht werden:

- Durch die Versionierung sind die Änderungen immer nachvollziehbar.
- Die externe Speicherung im Git macht die Konfigurationen jederzeit wiederherstellbar.

Es ist somit möglich sowohl das Datenmanagement (ETL und Applikationen) als auch die Datenhaltung durch DevOps-Automatisierung zu verwalten: Kurz DWH-Ops

Die automatische Bereitstellung kurzlebiger Datenbanken ermöglicht auch eigene Datenbanken für Feature Branches, was automatisierte Tests vereinfacht. Allerdings sind wir mit unserer Entwicklung an diesem Punkt nicht stehen geblieben.

DevSecOps und Secure by Design

Zuerst gibt es Aspekte der Sicherheit, die durch die nahtlose technische und organisatorische Integration erreicht wurden:

- Vollständige Transportverschlüsselung durch Service Mesh
- Durchgehende Authentifizierung in CI, Plattform und Datenbank
- Ausfallsicherheit der Automatisierung durch Kubernetes
- Integration von Tests in die Workflow-Automatisierung

Da diese Sicherheitsaspekte auch in der Konzeption (Secure by Design) und Entwicklung (DevSecOps) berücksichtigt werden, sind auch Anforderungen der demnächst geltenden NIS-2-Verordnung erfüllt.

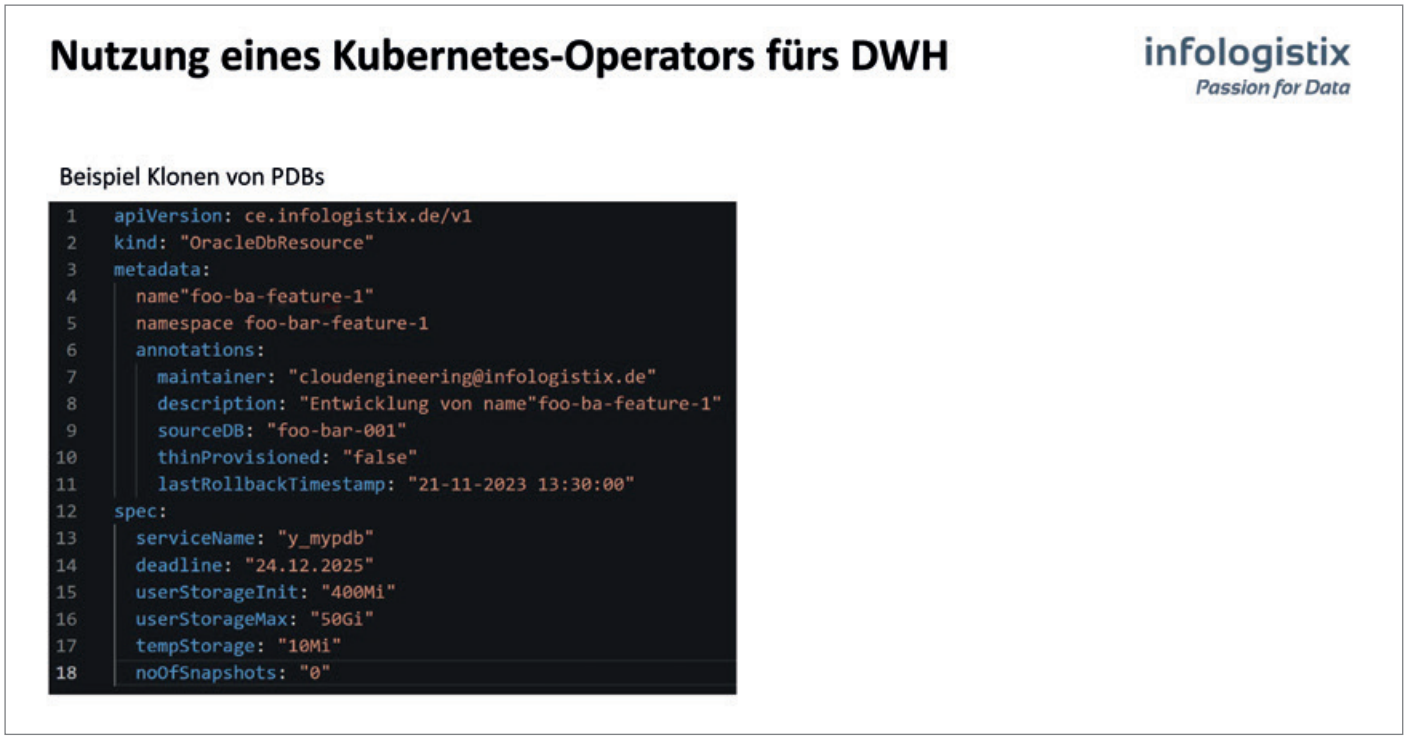


Abbildung 4: Custom Ressource zum Klonen einer PDB (Quelle: infologistix GmbH)

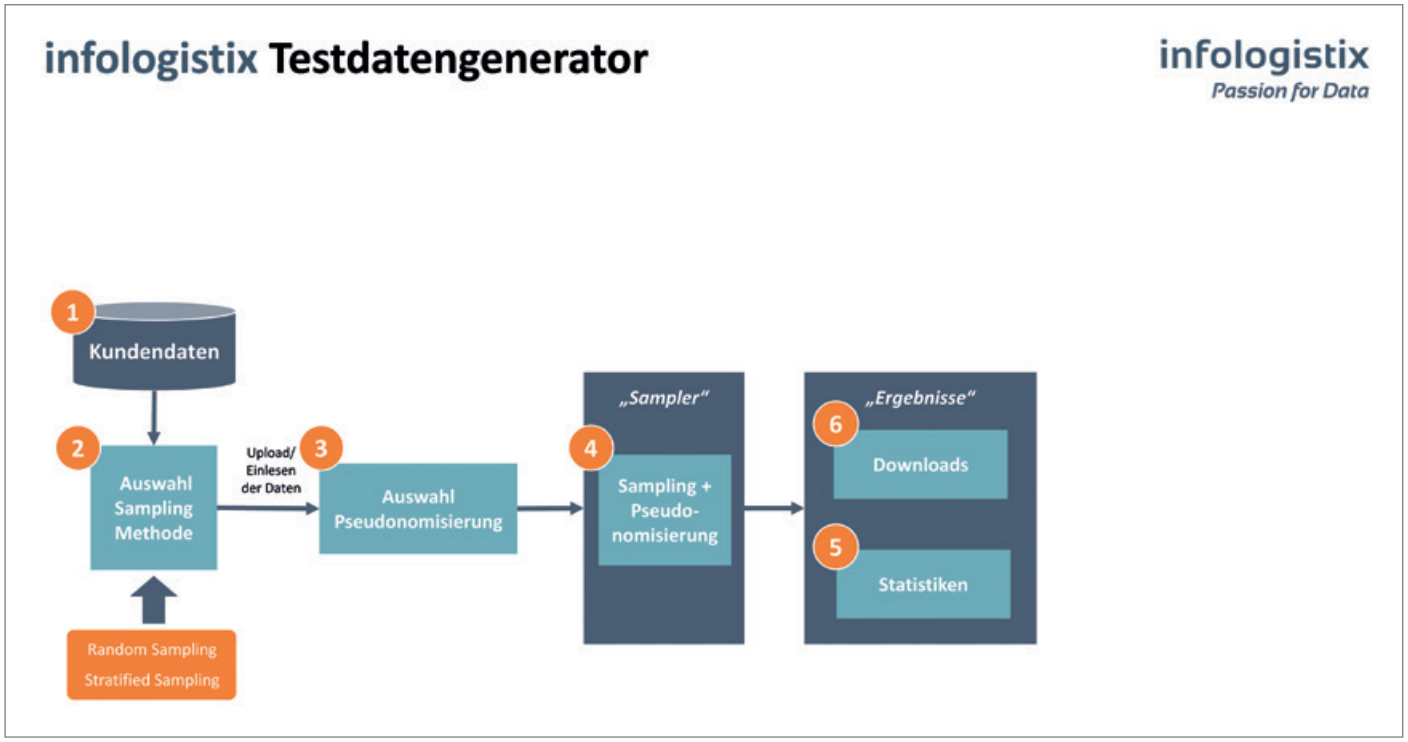


Abbildung 5: infologistix Testdatengenerator (Quelle: infologistix GmbH)

Test driven development

Durch die Integration der Datenbank-Administration in die CI ergeben sich zudem auch Möglichkeiten, die Datenbank und entsprechende Testdatensätze dynamisch in automatisierte Tests zu integrieren. Gerade für test driven development (TDD) bei dem die Tests sich aus den Kundenanforderungen und nicht aus der Programmierung der Anwendung ergeben, erweitert das die Möglichkeiten. Allerdings setzt das auch voraus, dass entsprechende produktionsreferente Testdaten zur Verfügung stehen.

infologistix Testdatengenerator

Dafür wurde der infologistix Testdatengenerator *(siehe Abbildung 5)* entwickelt, der aus einem großen Originaldatenset ein repräsentatives Sample erstellt und dieses anschließend pseudonymisiert.

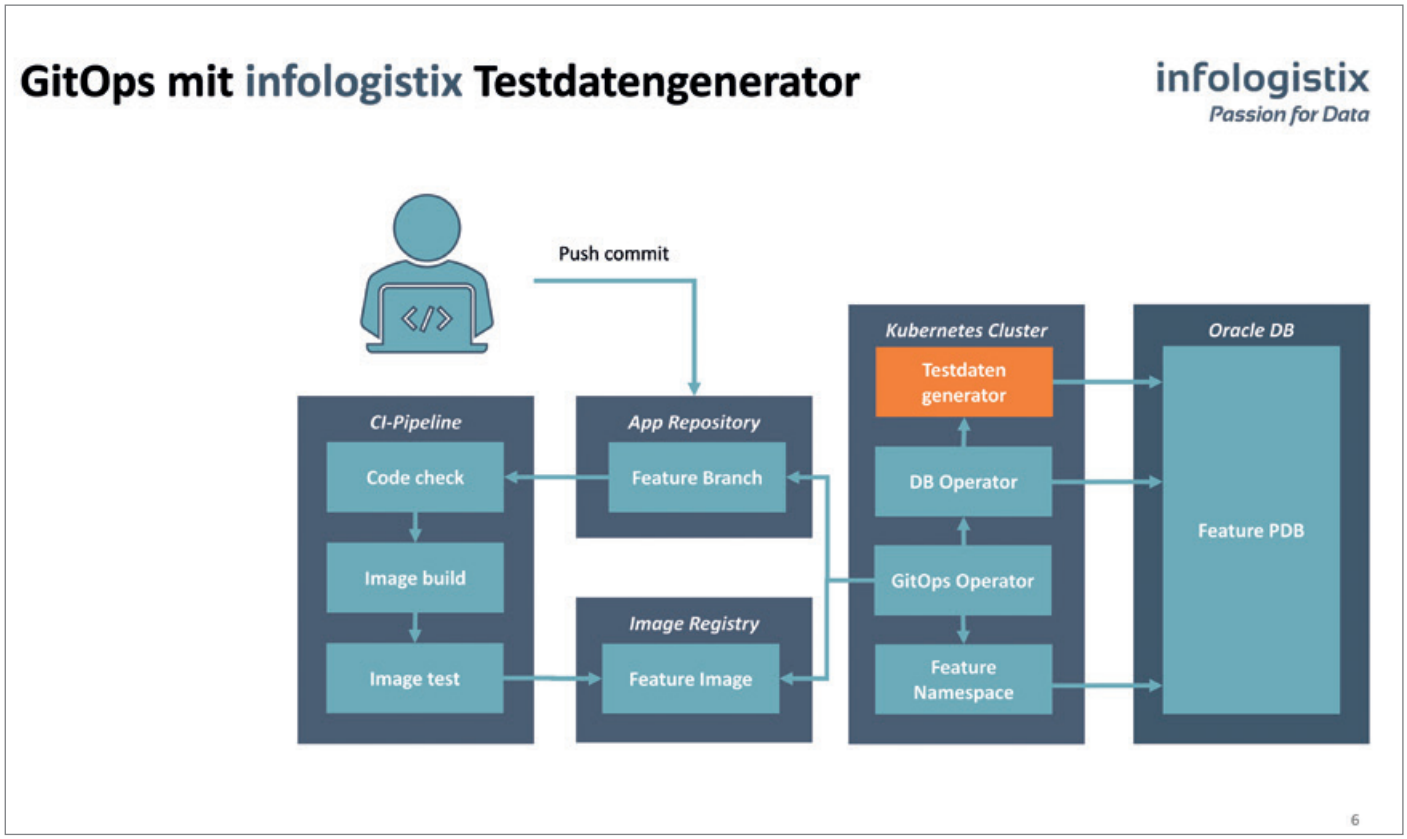


Abbildung 6: Integration des infologistix Testdatengenerator in den GitOps-Ansatz (Quelle: infologistix GmbH)

Nr.	Prozess	Beschreibung
1	Kundendaten	Originaldaten werden ausgelesen.
2	Auswahl Sampling Methode	Auswahl nach welcher Methode die Daten gesampelt werden sollen. Aktuell stehen „Random Sampling“ und „Stratified Sampling“ zur Verfügung.
3	Auswahl Pseudonymisierung	Während des Einlesens der Daten werden die Datentypen (Zahlen, String, ...) analysiert und entsprechende Pseudonymisierungen vorgeschlagen.
4	Sampling + Pseudonymisierung	Hier wird zunächst ein Sample aus den Originaldaten erstellt, welches anschließend pseudonymisiert wird.
5	Statistiken	Vergleich vom Originaldatenset und gesampelten Datenset sowohl tabellarisch als auch visuell
6	Downloads	Bereitstellung des gesampelten Datensets, des pseudonymisierten gesampleten Datensets und der ausgewählten Konfiguration für die Pseudonymisierung als Download.

Tabelle 2: Erläuterungen zum infologistix Testdatengenerator

Dabei kann aus zwei unterschiedlichen Samplingmethoden ausgewählt werden, welche vorab mithilfe statistischer Verfahrensweisen evaluiert wurden. Außerdem stehen verschiedene Pseudonymisierungen zur Verfügung. Zum Schluss wird ein Download der Testdaten und ein kurzer Bericht mit einer Gegenüberstellung der Original- und Testdaten zur Verfügung gestellt *(siehe Tabelle 2)*.

Der Testdatengenerator fügt sich auch entsprechend einfach in unser GitOps-Konzept ein *(siehe Abbildung 6)*. Dadurch ist eine automatische Testdatenerzeugung und Bereitstellung in einer kurzlebigen Testdatenbank beispielsweise für einen Feature Branch ohne manuelle Aufwände möglich.

Fazit und Ausblick

Durch unseren Ansatz konnten viele Operationen des DB-Admins automatisiert werden, was die Prozesse sehr beschleunigt. Die Verwendung von GitOps für die Datenbank-Operationen macht alle Änderungen versioniert, nachvollziehbar und jederzeit wiederherstellbar. Der Einsatz kurzlebiger Datenbanken für Feature Branches und automatisierte Tests ist ohne Mehraufwand möglich. Die IT-Sicherheit konnte durch mTLS, RBAC und getrennte Datenbanken erhöht und NIS-2 erfüllt werden. Durch die Integration des Testdatengenerators können passende Testdaten instantan erzeugt werden, um Fehler durch immer gleiche Testdaten zu vermeiden. Nächste Schritte unserer Entwicklung befassen sich aktuell mit der vollautomatisierten Bereitstellung von Testumgebungen und der Möglichkeit auch andere Datenbank-Typen durch diese Automatisierung zu ermöglichen.

Quellen

[1] *https://dbaclass.com/article/oracle-12c-multitenant-architecture/*
[2] *https://oracle-base.com/articles/21c/upgrade-non-cdb-to-21c-autoupgrade*
[3] *https://blogs.oracle.com/coretec/post/pdb-lifecycle-management-with-ords*

Über den Autor

Harald Gerhards ist Bereichsleiter Cloud Engineering bei der infologistix GmbH und Lehrbeauftragter für Cloud Native Computing an der Technischen Hochschule Würzburg-Schweinfurt (THWS). Er ist zertifizierter Projekt- und Qualitätsmanager (TÜV) und berät als Senior Systemarchitekt Kunden mit den Schwerpunkten Applikationsplattformen und Erweiterungen bestehender DWH- und BI-Architekturen.

Dr.-Ing Harald Philipp Gerhards
harald.gerhards@infologistix.de

BERLINER EXPERTENSEMINARE

26.06.- 27.06.2024

Entdecke die Welt der Künstlichen Intelligenz: Von Grundlagen zu Generativen Modellen

Berliner Expertenseminar mit Philip Weinmann, AI-Xperience GmbH

In diesem Seminar werden wir Ihnen einen umfassenden Einblick in die Funktionsweise von KI`s geben und die vielfältigen Möglichkeiten aufzeigen, wie Daten verarbeitet werden können. Wir beginnen mit einer Reise durch die verschiedenen Schichten, Gewichte und Aufgaben von KI-Modellen und führen Sie durch den Prozess des Modelltrainings, der Etikettierung, der Validierung und der Bewertung von Modellen mithilfe von Begriffen wie Epochen, Verlustfunktionen, Lernraten und Rückpropagierung. Wir werden uns auch mit wichtigen Konzepten wie Merkmalen, Genauigkeit, Präzision und Rückruf beschäftigen und den Übergang von einfachen Machine-Learning-Methoden zu neuronalen Netzwerken wie ReLu und Sigmoid diskutieren. Darüber hinaus werden wir Ihnen einen Einblick in einige der grundlegenden Frameworks für KI wie PyTorch, TensorFlow, Keras und Theano geben. Schließlich werden wir uns in die Welt der Generativen KI wagen und Ihnen sogar einen Einblick in die Bildgenerierung mit fortschrittlichen Techniken wie Stable Diffusion geben. Abgerundet wird das Seminar mit einer kurzen Erklärung zu Neuralen Sprach- und Sprachmodellen sowie deren Anwendungen. Dieses Seminar bietet Ihnen eine solide Grundlage, um die Möglichkeiten der Künstlichen Intelligenz voll auszuschöpfen und in die Welt der Generativen Modelle einzutauchen.

PL/SQL-Laufzeitoptimierung durch Parallelisierung

Jan Gorkow, SD&C Solutions Development & Consulting

Mit Hilfe von PL/SQL lässt sich datennahe Business Logik sauber strukturiert sowie hoch performant in der Oracle Datenbank abbilden. Einige Ansätze zur performanceoptimierten Entwicklung finden sich in dem Beitrag „PL/SQL-Performancesteigerung durch optimale Datentypen und native Kompilierung“, erschienen in der Februar-Ausgabe 02/2024 dieses Magazins. Doch was tun, wenn trotz aller Optimierungen die geforderten Laufzeiten komplexer, sequenziell abgebildeter Logiken nicht erreicht werden können und dadurch Anwender beispielsweise unerträglich lange vor einem APEX-Web-Frontend auf Ergebnisse warten müssen? Dann kommt Parallelisierung ins Spiel! Dieser Artikel beleuchtet drei Ansätze, wie komplexe Programmlogik innerhalb der Datenbank über mehrere Prozesse parallelisiert werden kann und geht anschießend auf entsprechende Architekturmodelle ein.

Die heutzutage verfügbare und eingesetzte Hardware unterstützt Parallelisierung auf vielen Ebenen. Im hier relevanten Bereich der zur Verfügung stehenden Rechenleistung je Zeit wird dies insbesondere durch eine Vielzahl von Cores je CPU, mehreren CPUs je Server sowie der Zusammenschaltung von mehreren Servern zu einem Cluster erreicht.

Die Oracle-Datenbank skaliert über alle diese Ebenen ideal. DBMS ist seit jeher mit einer auf Parallelisierung ausgelegten Prozess-Architektur ausgestattet, da sich die Skalierungsmöglichkeiten durch den Einsatz von RAC & Co. inzwischen weit über die Grenzen von Rechnern, ja sogar Rechenzentren, erstrecken. So sind parallele Nutzerzahlen von mehreren tausend, die alle gleichzeitig auf ein und dieselbe Oracle-Datenbank zugreifen, inzwischen alles andere als unüblich. Auch die Verteilung der Workloads einzelner Benutzersessions auf mehrere Prozesse/CPU-Threads ist durch Parallel Query/Parallel DML gängige Praxis und tief im Oracle DBMS verankert.

Im Vergleich zur teilweise bereits vollautomatisch beziehungsweise deklarativ oder konfigurativ erfolgenden Parallelisierung von SQL ist die parallele Abarbeitung von PL/SQL-Programmen nicht ganz so trivial. Ein PL/SQL-Programm, welches in einer Nutzersession gestartet wird, wird zunächst nur von einem einzigen Thread abgearbeitet. Nachfolgend werden daher Varianten aufgezeigt, mit denen die gewünschte Verteilung auf mehrere Threads beziehungsweise Prozesse erreicht werden kann.

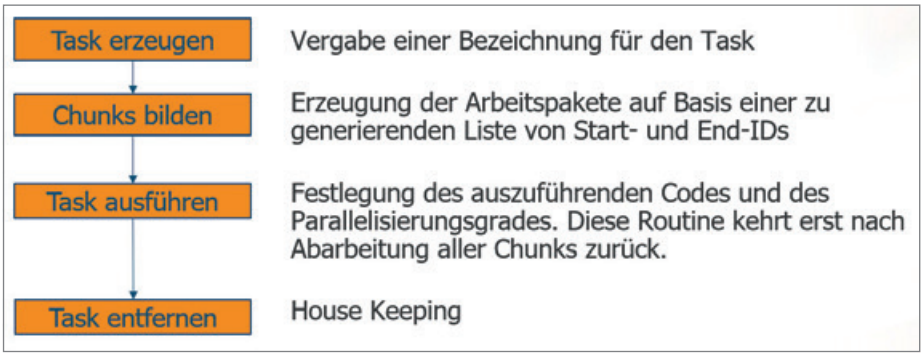


Abbildung 1: Ablauf DBMS_PARALLEL_EXECUTE (Quelle: Jan Gorkow)

Variante 1 – Parallel Pipelined Table Functions

Als Einstieg zur Erklärung dieser Variante soll zunächst auf die verschiedenen verfügbaren Funktionstypen eingegangen werden:

- Eine normale skalare Funktion nimmt 0 bis n Parameter entgegen und gibt einen errechneten und/oder aus einer Tabelle geladenen skalaren Wert zurück.
- Eine Table Function erzeugt eine tabellarische Datenmenge und gibt diese mit einem Mal zurück. Dieser Funktionstyp kann direkt in PL/SQL oder in der FROM-Klausel von SQL-Anweisungen aufgerufen werden.
- Eine Pipelined Table Function gibt eine tabellarische Datenmenge zurück, allerdings satzweise und kann daher nur in FROM-Klauseln von SQL-Anweisungen verwendet werden.
- Darauf aufbauend handelt es sich bei Parallel Pipelined Table Functions um eine parallelisierbare Variante, welche nun näher erläutert werden soll.

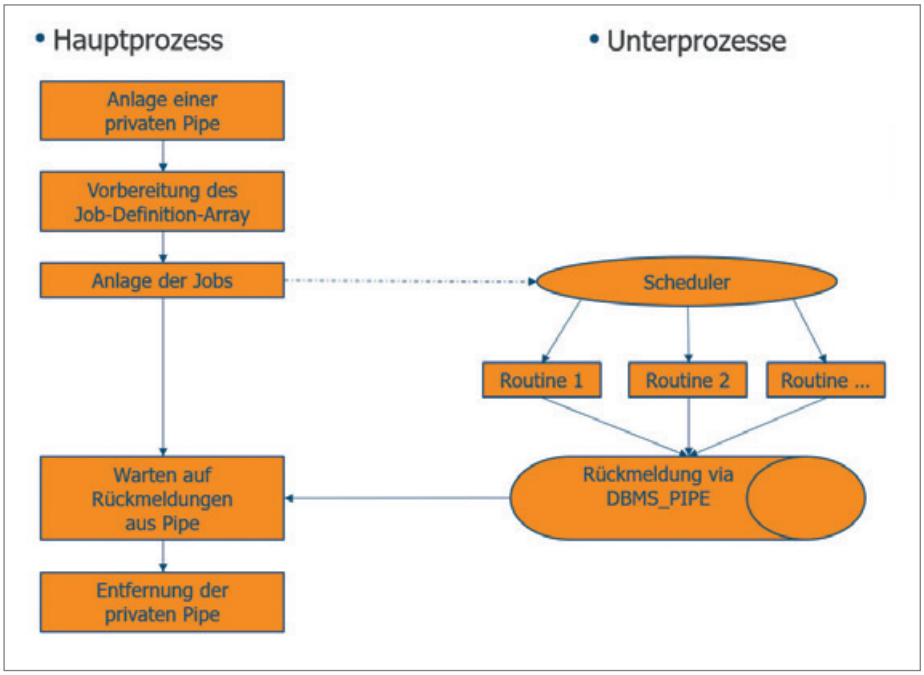


Abbildung 2: Ablauf bei direkter Scheduler-Nutzung – Multitasking-Architektur (Quelle: Jan Gorkow)

Parallel Pipelined Table Functions zeichnen sich dadurch aus, dass sie mindestens einen Ref-Cursor-Parameter annehmen müssen, über welchen der Funktion eine tabellarische Datenmenge übergeben wird. Die eigentliche Parallelisierung wird über die PARALLEL_ENABLE-Klausel auf Basis des Ref-Cursor-Parameters erreicht, indem angegeben wird, über welche Kriterien die eingehende Datenmenge partitioniert werden soll, um die daraus resultierenden Teilmengen an jeweils eine eigene Funktionsinstanz zu übergeben. Die Partitionierung kann entweder wahlfrei (ANY), per HASH oder per RANGE über die angegebene Spalte erfolgen. *Listing 1* zeigt die Syntax zur Erzeugung sowie den groben schematischen Aufbau einer Parallel Pipelined Table Function.

Innerhalb der Funktion werden Daten per FETCH aus dem übergebenen Ref Cursor abgerufen und per implementierter Logik auf dieser Basis Ergebnissätze erzeugt, welche per PIPE-ROW-Anweisung zurückgegeben werden. Oracle kümmert sich um die

Zusammenführung der Ergebnissätze aller Funktionsinstanzen und gibt diese an das aufrufende Statement zurück. Die Parallel Pipelined Table Functions wiederum können nur innerhalb von FROM-Klauseln in SQL-Anweisungen referenziert werden. Der Grad der Parallelisierung kann bei der Übergabe des Ref Cursors innerhalb der ihm zugrunde liegenden SQL-Anweisung beeinflusst werden, zum Beispiel mittels des PARALLEL Hints oder durch die auf Tabellenebene gesetzten PARALLEL- Eigenschaften.

Diese Variante ist immer dann sinnvoll, wenn es um die Beschleunigung komplexer Transformationslogiken geht, die auf große Datenmengen angewendet werden sollen. Hierbei muss es jedoch möglich sein, die Gesamtdatenmenge in einzelne, unabhängig voneinander verarbeitbare Teilmengen zu partitionieren. Ist diese Voraussetzung gegeben, so kann mit geringem Implementierungsaufwand eine sehr gute Skalierung erreicht werden. Durch die Verzahnung von SQL und PL/SQL kommt es bei der Ausführung naturgemäß zu einer Reihe von Kontext-Switches. Für die Parallelisierung von reinem PL/SQL-Code ist diese Variante daher eher nicht geeignet.

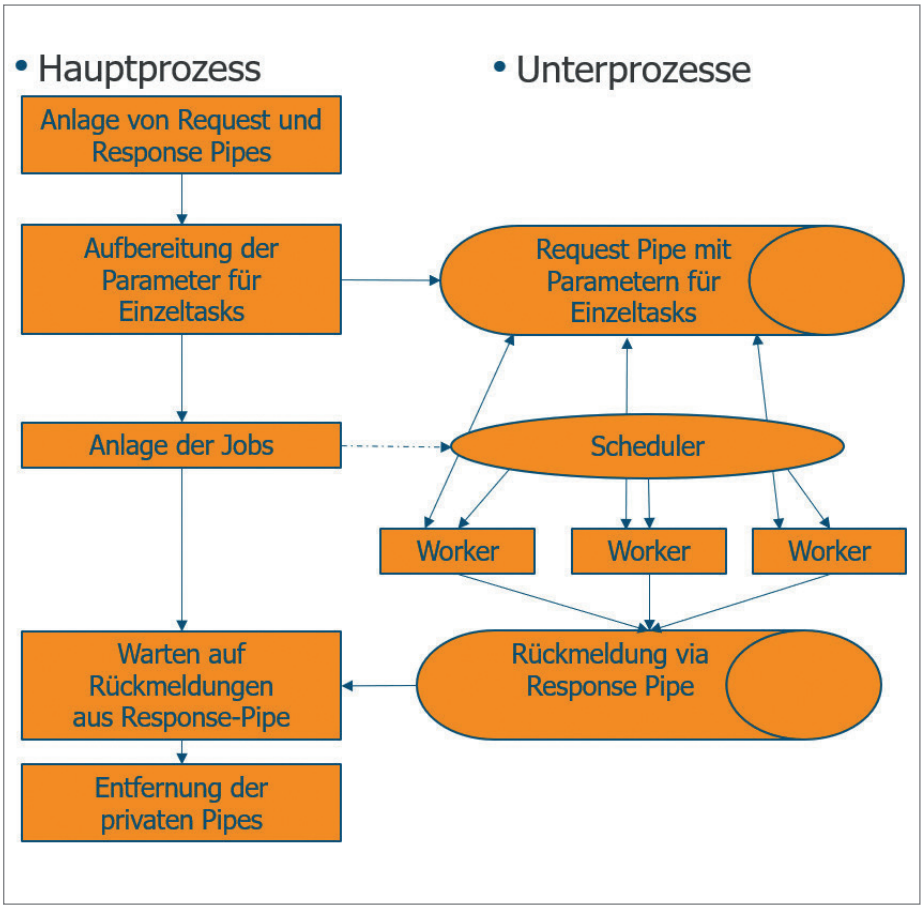


Abbildung 3: Architektur – Worker-Modell (Quelle: Jan Gorkow)

Variante 2 – DBMS_PARALLEL_EXECUTE

Mit der Oracle Version 11g wurde das Built-in Package DBMS_PARALLEL_EXECUTE eingeführt. Hintergrund war primär die Parallelisierung von Updates auf Massendaten im Data-Warehouse-Umfeld, aber es ist auch für die Parallelisierung von PL/SQL einsetzbar. Über Aufrufe der API des Packages werden die in *Abbildung 1* dargestellten Schritte durchlaufen.

Listing 2 zeigt ein Anwendungsbeispiel. Der Prozedur prc_execute_parallel wird über den Parameter instances die Anzahl der parallel zu startenden Verarbeitungsinstanzen einer Routine prc_workload übergeben. Nach Anlage des Tasks PARALLEL_EXECUTE (create_task) werden per SQL-Anweisung die Chunks definiert (create_chunks_by_sql). Die Ergebnisstruktur der übergebenen SQL-Anweisung ist hierbei von Oracle vorgegeben, das heißt, es müssen immer die Spalten START_ID und END_ID geliefert werden. In diesem Bei-

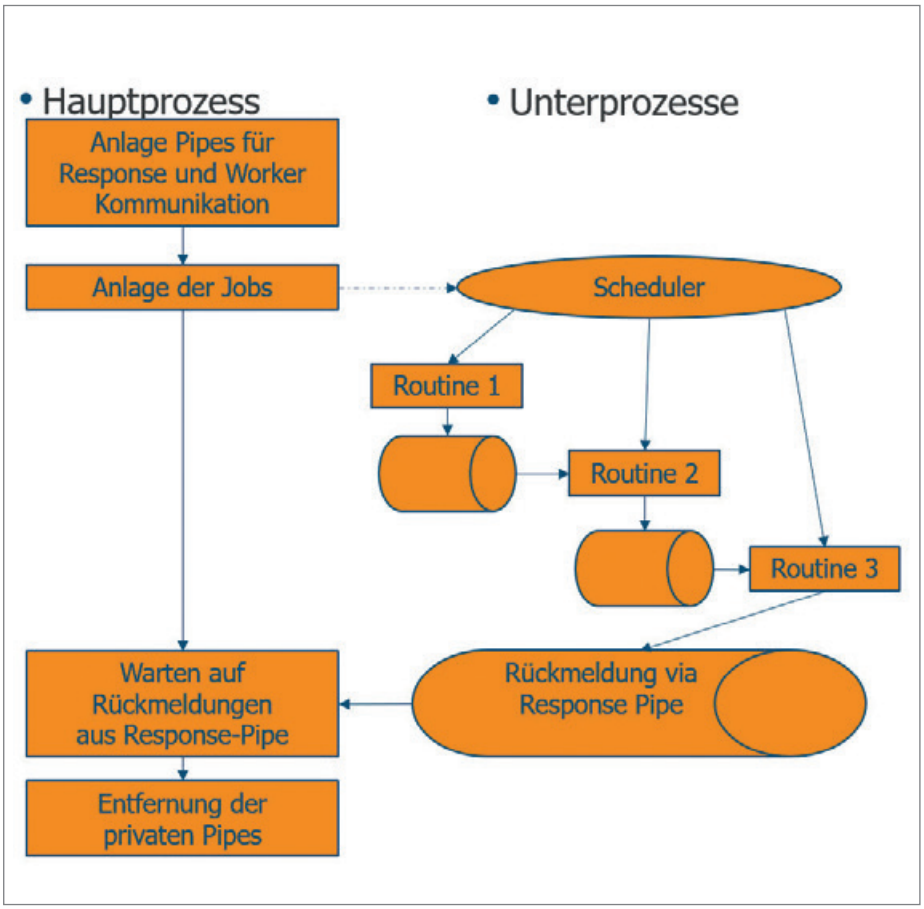


Abbildung 4: Architektur – Pipelining-Modell (Quelle: Jan Gorkow)

```
CREATE FUNCTION pp_tab_function(p_ref_cursor ref-cursor-type)
  RETURN return_tab_type PIPELINED
  PARALLEL_ENABLE (PARTITION p_ref_cursor BY [{HASH | RANGE} (spalten) | ANY]) IS
  v_return_record return_record_type;
  …
BEGIN
  LOOP
    FETCH p_ref_cursor INTO …
    EXIT WHEN p_ref_cursor%NOTFOUND;

    --Business Logik auf Basis der gefetchten Daten
    …
    PIPE ROW(v_return_record);
  END LOOP;
END pp_tab_function;
/
```

Listing 1: Template zur Erstellung einer Parallel Pipelined Table Function

spiel erhalten beide Spalten den gleichen Wert und es werden entsprechend dem übergebenen Parameter die Anzahl an Datensätzen erzeugt. Anschließend kann der Task ausgeführt werden (run_task). Neben dem eigentlichen Workload wird an dieser Stelle auch erst der Grad der Parallelisierung festgelegt. Im Hintergrund wird die Anzahl der geforderten Verarbeitungsprozesse gestartet. Jeder einzelne Arbeitsprozess wird dann kontinuierlich so lange mit Arbeit versorgt, wie zu verarbeitende Chunks vorhanden sind. Erst nach Abarbeitung aller Chunks kommt der run_task-Aufruf zurück und der Hauptprozess läuft weiter. Zum Abschluss muss nur noch etwas House Keeping betrieben und der Task wieder entfernt werden (drop_task).

Die Vorteile dieser Variante ergeben sich aus all dem, was dem Entwickler von dem Package DBMS_PARALLEL_EXECUTE abgenommen wird:

- Prozesserzeugung unter interner Verwendung des Scheduler
- Prozesssynchronisation nach Abarbeitung aller Chunks
- Einfach zu konfigurierender Parallelisierungsgrad
- Logging und Monitoring

All das hat natürlich seinen Preis. Vor dem Hintergrund des ursprünglich angedachten Anwendungsfalls im DWH-Umfeld, muss ein Task bezüglich der Abarbeitung seiner Chunks wiederanlauffähig sein und unter anderem nach einem Neustart der Datenbank weiterlaufen. Um dies sicherstellen zu können, müssen eine ganze Reihe von Metadaten gepflegt werden, was wiederum eine Vielzahl impliziter Transaktionen erfordert. Jede Einzelne davon ist mit Festplattenzugriffen verbunden und damit im Hinblick auf den Anwendungsfall der Parallelisierung von reinem, hoch performanten PL/SQL-Code stark Overhead-behaftet und verlängert die Gesamtlaufzeit. So kann zwischen dem Abschluss des letzten Chunks und der Rückkehr der run_task-Routine im Hauptprozess schon mal eine ganze Sekunde vergehen, was im Hinblick auf die Beschleunigung rechenintensiver Logiken im Frontend-Kontext schon eine kleine Ewigkeit und eigentlich auch schon zu viel an Overhead ist.

Variante 3 – Direkte Nutzung des Schedulers

Die dritte und letzte im Rahmen dieses Artikels betrachtete Variante basiert auf der direkten Nutzung des Schedulers sowie dem Einsatz von privaten Pipes zur erforderlichen Interprozess-Kommunikation. *Abbildung 2* stellt den Ablauf diesbezüglich schematisch dar.

Zunächst wird eine private Pipe angelegt, welche dazu dient, den Hauptprozess über den Status, insbesondere den Abschluss der Unterprozesse zu informieren.

Anschließend wird über den Scheduler der Start der Unter- beziehungsweise eigentlichen Arbeitsprozesse vorbereitet. Hierfür können mehrere Jobs in Form eines Job Definition Arrays zusammengefasst und mittels eines Aufrufes an den Scheduler zu Ausführung übergeben werden.

Bei der Definition von Scheduler Jobs muss der Job Style mit angegeben werden. Hier stehen folgende Ausprägungen zur Auswahl:

- REGULAR: Regulärer Job mit vollem Set an Metadaten im Data Dictionary.
- LIGHTWEIGHT: Job mit reduzierten Metadaten. Für die aufgerufene Programm-Einheit muss zuvor ein Scheduler-Programm definiert worden sein.
- IN_MEMORY_RUNTIME: Weiter reduzierte Festplattenzugriffe durch verringerte Metadaten und den Einsatz von In Memory-Cache-Mechanismen.
- IN_MEMORY_FULL: Eingeführt mit Oracle 12c. Kompletter Verzicht auf Metadaten und dadurch komplett transaktionslos.

IN_MEMORY_FULL-Jobs können ausschließlich sofort ausgeführt und nicht per Schedule geplant werden. Die maximale Performance durch den vollständigen Verzicht auf persistierte Metadaten macht sie ideal für die Parallelisierung von PL/SQL-Programmen im Frontend-Kontext.

Der letzte Schritt im PL/SQL-Code jedes einzelnen Arbeitsprozesses ist die Rückinfo an den Hauptprozess. Sobald diese sämtliche Rückmeldungen erhalten haben, läuft das Hauptprogramm weiter und räumt die dann nicht mehr benötig-

```
CREATE OR PROCEDURE prc_workload
    (instance IN SIMPLE_INTEGER)IS
-----------------------------------------------------------------
--Routine zur Simulation eines Workloads von 2 Sekunden. Ueber
--den Parameter instance koennen bei Bedarf die verschiedenen,
--parallel laufenden Verarbeitungs-Instanzen unterschieden
--werden.
-----------------------------------------------------------------
BEGIN
   DBMS_SESSION.sleep (seconds => 2);
END prc_workload;
/

CREATE OR REPLACE PROCEDURE prc_execute_parallel
    (instances IN SIMPLE_INTEGER) IS
-----------------------------------------------------------
--Parallele Abarbeitung der uebergebenen Anzahl an
--Arbeitspaketen durch die parallele Ausfuehrung von
--Verarbeitungsinstanzen unter Nutzung des Packages
--DBMS_PARALLEL_EXECUTE.
-----------------------------------------------------------
   const_task_name   CONSTANT VARCHAR2 (30 CHAR)
     := 'PARALLEL_EXECUTE';
   v_result                   INTEGER;
BEGIN
   DBMS_PARALLEL_EXECUTE.create_task (
      task_name => const_task_name);
   DBMS_PARALLEL_EXECUTE.create_chunks_by_sql (
      task_name   => const_task_name,
      sql_stmt    =>
            'WITH instances (instance)'
         || '        AS (SELECT 1 AS instance FROM DUAL'
         || '            UNION ALL'
         || '            SELECT instances.instance + 1'
         || '                     AS instance'
         || '              FROM instances'
         || '             WHERE instances.instance < '
         || TO_CHAR (prc_execute_parallel_execute.instances)
         || ') '
         || 'SELECT instances.instance AS start_id,'
         || '       instances.instance AS end_id'
         || '  FROM instances',
      by_rowid    => FALSE);
   DBMS_PARALLEL_EXECUTE.run_task (
      task_name        => const_task_name,
      sql_stmt         =>
            'DECLARE'
         || '   v_start_id PLS_INTEGER := :start_id;'
         || '   v_end_id   PLS_INTEGER := :end_id;'
         || 'BEGIN'
         || '   prc_workload ('
         || '     instance => v_start_id);'
         || 'END;',
      language_flag    => DBMS_SQL.native,
      parallel_level   => prc_execute_parallel_execute.instances);
   DBMS_PARALLEL_EXECUTE.drop_task (
     task_name => const_task_name);
END prc_execute_parallel_execute;
/
```

Listing 2: Beispiel für die PL/SQL-Parallelisierung per DBMS_PARALLEL_EXECUTE

te Pipe wieder weg. *Listing 3* zeigt sowohl die Parallelisierungslogik als auch die angepasste Workload-Routine.

Der Vorteil dieser Variante liegt in dem kaum vorhandenen Overhead der Parallelisierung. Aufgrund des umgehenden Starts der Jobs bei Nutzung des Job-Styles IN_MEMORY_FULL sowie der extrem schnellen Interprozess-Kommunikation per Pipe wird tatsächlich fast die komplette Laufzeit des Hauptprogramms über von mindestens einem Arbeitsprozess produktiv gearbeitet. Die Kunst liegt hier in der sinnvollen Schneidung der Arbeitspakete im Hinblick auf ähnlich lange Laufzeiten, damit der Effekt der Parallelisierung sich optimal auswirken kann.

Die Nachteile des höheren Entwicklungsaufwandes sowie des fehlenden integrierten Loggings und Monitorings sollen hier nicht unerwähnt bleiben.

Architekturmodelle

Werfen wir noch einen Blick auf eine Auswahl verschiedener Architekturmodelle, die im Hinblick auf die Planung und Implementierung der Schneidung und Parallelisierung von PL/SQL-Code eine Rolle spielen können.

Die mit dem Ablauf in *Abbildung 2* dargestellte Architektur entspricht einem Multitasking-Ansatz. Hierbei werden wenige, voneinander unabhängige und gegebenenfalls vollkommen verschiedenartige Einzeltasks parallelisiert, wobei alle Einzeltasks idealerweise eine ähnliche Laufzeit haben sollten.

Hat man es bezüglich des Workloads hingegen mit einer Vielzahl gleichartiger Tasks zu tun, so kommt eher die in *Abbildung 3* dargestellte Worker-Architektur in Frage, da es nicht sinnvoll beziehungsweise sogar kontraproduktiv ist, für jeden der vielen Einzeltasks einen eigenen Job über den Scheduler zu starten. Vielmehr wird nur eine definierte und im Hinblick auf die Parallelisierung sinnvolle Anzahl von Jobs gestartet, die sich aus einem gemeinsamen Arbeitsvorrat in Form einer Pipe so lange die Parameter für ihre Aufgaben holen und abarbeiten, bis der Vorrat leer ist. Sobald ein Worker nichts mehr zu tun hat, gibt er eine entsprechende Rückmeldung per Pipe an den Hauptprozess und sobald dieser alle

```
CREATE OR REPLACE PROCEDURE prc_workload_callback
   (instance IN SIMPLE_INTEGER) IS
-----------------------------------------------------------------
--Routine zur Simulation eines Workloads von 2 Sekunden mit
--anschliessendem Callback
-----------------------------------------------------------------
   v_result   INTEGER;
BEGIN
   DBMS_SESSION.sleep (seconds => 2);
   DBMS_PIPE.pack_message (item =>
      prc_workload_callback.instance);
   v_result := DBMS_PIPE.send_message (
      pipename => const_pipename);
END prc_workload_callback;

PROCEDURE prc_execute_parallel_scheduler
   (instances IN SIMPLE_INTEGER) IS
-----------------------------------------------------------------
--Parallele Abarbeitung von x Arbeitspaketen unter Nutzung des
--DBMS_SCHEDULER-Packages
-----------------------------------------------------------------
   v_result                 INTEGER;
   v_job_definition_array   job_definition_array;
BEGIN
   v_result :=
      DBMS_PIPE.create_pipe (
         pipename      => 'PIPE_CALLBACK',
         maxpipesize   => 1048576,
         private       => TRUE);

   v_job_definition_array := NEW job_definition_array ();
   v_job_definition_array.EXTEND (
      prc_execute_parallel_scheduler.instances);

   FOR i IN 1 .. prc_execute_parallel_scheduler.instances
   LOOP
      v_job_definition_array (i) :=
         NEW job_definition (
                job_name       => 'JOB_EXECUTE_PARALLEL_'
                                  || TO_CHAR (i),
                job_style      => 'IN_MEMORY_FULL',
                program_name   => 'SP_WORKLOAD_CALLBACK',
                arguments      =>
                   NEW jobarg_array (
                          NEW jobarg (arg_position   => 1,
                                      arg_value      =>
                                          TO_CHAR (i))),
                enabled        => TRUE);
   END LOOP;

   DBMS_SCHEDULER.create_jobs (
      jobdef_array       => v_job_definition_array,
      commit_semantics   => 'TRANSACTIONAL');

   FOR v_instance IN 1 .. prc_execute_parallel_scheduler.instances
   LOOP
      v_result :=
         DBMS_PIPE.receive_message (
             pipename   => const_pipename,
             timeout    => const_timeout);
   END LOOP;

   v_result := DBMS_PIPE.remove_pipe (
                  pipename => const_pipename);
END prc_execute_parallel_scheduler;
/
```

Listing 3: Parallelisierung per Scheduler und Pipe

Rückmeldungen hat, läuft das Hauptprogramm weiter. Ein entscheidender Vorteil der Worker-Architektur ist, dass auch dann eine optimale Parallelisierung erreicht werden kann, wenn die Laufzeit der Einzeltasks stark variiert.

Eine weitere mögliche Architektur ist in *Abbildung 4* mit dem sogenannten Pipelining- Modell dargestellt. Hierfür muss sich die Gesamtverarbeitung in einzelne, aufeinander aufbauende Schritte aufteilen lassen. Ferner müssen die Folgeprozesse mit Teil-Ergebnismengen beziehungsweise Einzelwerten ihres jeweiligen Vorgängerprozesses arbeiten können. Der Vorteil am Pipelining-Modell liegt darin, dass dem Hauptprozess sehr schnell die ersten Ergebnisse vorliegen, mit welchen dieser dann wiederum bereits arbeiten kann.

Fazit

Sicherlich ist nicht immer eine Parallelisierung der in PL/SQL abgebildeten komplexen Business Logik möglich, nötig oder sinnvoll. Meine Empfehlung ist, bei bestehenden Performance- beziehungsweise Laufzeitproblemen zunächst sämtliche auf sequenziellen PL/SQL-Code anwendbaren Optimierungsmaßnahmen auszuschöpfen, zum Beispiel native Kompilierung, Einsatz von SIMPLE_ - Datentypen und/oder Caching. Erst wenn hiermit die geforderte Laufzeit nicht erreicht werden kann, sollten die Möglichkeiten der Parallelisierung ins Auge gefasst werden.

Was die dann eingesetzte Parallelisierungstechnik angeht, so fällt die klare Empfehlung des Autors auf Variante 3, insbesondere aufgrund des sehr geringen Overheads und der hohen Flexibilität, die sich aus der erforderlichen Individualentwicklung ergibt.

Wichtig ist allerdings zunächst die Auswahl der optimalen Parallelisierungsarchitektur, da dies wiederum starken Einfluss auf die Schneidung des vorhandenen Codes beziehungsweise die Implementierung der Workload-Routinen hat. Ferner muss der optimale Parallelisierungsgrad sinnvoll bestimmt oder ermittelt werden. Hierbei gilt es zu beachten, dass sich die Umgebungsparameter und Rahmenbedingungen zwischen den Systemen sowie auch

zeitabhängig stark unterscheiden können. Nur weil eine Parallelisierung über 10 Prozesse auf dem Entwicklungssystem mitten in der Nacht für einen Entwickler optimale Ergebnisse liefert, muss das bei der gleichen Anwendung in der Produktion bei 100 gleichzeitig aktiven Benutzern am Montagvormittag nicht auch so sein. Grundsätzlich sollte daher der Parallelisierungsgrad sehr fein steuer- beziehungsweise konfigurierbar sein.

Über den Autor

Jan Gorkow ist gelernter Fachinformatiker in der Anwendungsentwicklung und hat sich bereits sehr früh in seiner Ausbildung auf Oracle-Datenbankentwicklung, seinerzeit noch unter Oracle 8i, spezialisiert. Seit 2011 ist er als Senior Managing Consultant für die SD&C Solutions Development & Consulting GmbH tätig, welche seit Ende 2021 zur Dataciders Gruppe gehört. Mit seinem Team von Oracle-Datenbank- und APEX-Spezialisten konzipiert und realisiert er für Kunden hoch performante DB-Anwendungen, um das in den Daten steckende Potential zu erschließen und optimale Business-Entscheidungen zu ermöglichen, getreu der Dataciders-Vision: Niemand muss jemals wieder schlechte Entscheidungen treffen.

Jan Gorkow
jan.gorkow@sd-c.de

Zukunftssicherheit: Ihr Partner für Digitale Intelligenz

Nachhaltige Technologien für Unternehmen von morgen

Sie haben die Daten, wir die Vernetzung zum Erfolg – individuell für Ihre Bedürfnisse. Mit innovativen Cloud-Lösungen und Oracle-Applikationen, die eine praxisnahe Umsetzung garantieren. Sicher. Einfach. Überall.

Jetzt informieren: **www.promatis.de**

ORACLE | Partner

PROMATIS

APEX connect 2024 – eine Eins mit Sternchen für die Jubiläumsausgabe

Text & Fotos: Marcos López, DOAG Redaktion

Nach 2018 hatte man sich für die 10. Ausgabe der Oracle-Application-Express-Fachkonferenz „made in Germany" erneut für Düsseldorf entschieden. Ein Volltreffer!

Die APEX connect 2024 begrüßte mit neuem Konzept, bestehend aus dem Bootcamp & Solutions Day für absolute Beginner und mit Lösungen auf Basis von APEX sowie zwei klassischen Konferenztagen, über 341 Teilnehmende. Das Programm, das zu einem Viertel in englischer Sprache angeboten wurde, bestand aus 77 Sessions à 45 Minuten. Neben den Keynotes aus Finnland und den USA wurden sechs DevTalks und zahlreiche 1-zu-1-Sessions mit Experten aus dem Oracle-APEX-Entwicklerteam realisiert. 27 Oracle ACEs hatten zudem ihre Anwesenheit auf dieser Konferenz zugesagt.

Mit dem mittlerweile legendären APEX-Power-Ruf, der einer aus Fußballstadien bekannten La-Ola-Choreographie nachempfunden ist, schloss die perfekte Eröffnung der Jubiläumsausgabe in dem aus allen Nähten platzenden Raum des Van der Valk Airporthotel. Den 24 Sessions des ersten Tages folgte die mittlerweile bestens etablierte „Happy Hour sponsored by Oracle“, sowie das vorzügliche Live Cooking im Atrium-Foyer, das den bereits zum Mittagessen gewonnenen Eindruck bestätigte: Frisch zubereiteter Fisch und gebratenes Fleisch, Salate und eine große Auswahl an raffinierten Beilagen sowie vegetarische und vegane Optionen ließen nichts zu wünschen übrig und sorgten auch in den kommenden Tagen für Begeisterung.

Volles Programm, vorzüglicher Abend

Tag 2 der APEX-Konferenz präsentierte das aus den Jahren zuvor etablierte „klassische Konferenzprogramm“, bestehend aus PL/SQL, JavaScript und natürlich APEX, in Form von insgesamt 37 unterschiedlichen Sessions. Der Tag war für einige schon früh losgegangen, denn bereits um 7 Uhr trafen sich zahlreiche Frühaufsteher der Konferenz, um bei frostigen 0° joggend eine 5-Kilometer-Runde um den Düsseldorfer Silbersee zu drehen und sich so eine Extra-Portion Sauerstoff zu gönnen. Der offiziellen Begrüßung um 9:15 Uhr folgte dann die Keynote des US-Amerikaners Michael Hichwa, Erfinder von APEX – APEX ist das Akronym für Application Express. Hichwa widmete sich in seinem Vortrag der mittlerweile zwanzigjährigen Geschichte von Oracle Application Express, verbunden mit Frage, was noch alles möglich ist beziehungsweise wohin die Reise noch gehen könnte.

Besondere Beachtung bekam die NGO Fellas For Europe e.V., die mit einem in 2022 zu Beginn des russischen Angriffskriegs attackierten Krankenwagen aus der Ukraine vor dem Haupteingang des Van der Valk Station machte. Hier wurde die Geschichte des mit Einschüssen übersäten Fahrzeugs erzählt und auf die aktuelle Situation in der Ukraine hingewiesen.

Zum Abend brach die APEX-Community schließlich mit drei großen Reisebussen zu Roncalli's Apollo Varieté am malerischen Ufer des Rheins auf. Hier wartete das neue Programm namens „Traumschiff“ auf seine gutgelaunten Gäste, die nach den ersten Getränken samt Vorspeise von der Crew auf eine musikalisch-maritime Reise mit hochklassiger Akrobatik und Pop durch zahlreiche Länder Europas mitgenommen wurden.

6k+
Developers
41k+
Applications

FELLAS FOR
EUROPE E.V.
NAFO
OFAN
KD

Von Anfang bis Ende eine tolle Energie

Erstes Highlight des dritten Tages war die Keynote von Heli Helskyaho. Die Finnin hatte ihren Sohn Matias mitgebracht und zusammen nahmen sie das Publikum mit auf eine Reise über Generative AI, Large Language Models (LLM) und Vektoren. In der letzten offiziellen Vormittags-Kaffeepause der Konferenz herrschte so ein reges Treiben, dass man das Gefühl bekommen konnte, sie hätte gerade erst begonnen; an sämtlichen Ständen der Aussteller und Sponsoren herrschte Hochbetrieb. Gegen 15 Uhr erklang schließlich ein letztes Mal das Signal des Van der Valk, um den pünktlichen Beginn der letzten Vorträge der APEX connect 2024 anzukündigen und damit das Finale einzuläuten. Ein Blick durch das Foyer bestätigte den über den gesamten Verlauf der Konferenz gewonnen Eindruck: rundherum zufriedene Gesichter.

Stimmen zur APEX connect 2024

„Ich habe die Konferenz als sehr angenehm wahrgenommen – die interessanten Stände, die netten Leute, die vielen Gespräche, die man führen konnte. Das persönliche Highlight war für mich diese Situation vor der Tür des Apollo Varietés, wo wir alle Teilnehmer noch einmal begrüßen und angucken konnten und man die freudigen Gesichter gesehen hat. Das war sehr schön.“
Kai Donato, APEX connect Stream JavaScript

„Es war eine ganz tolle Konferenz mit einer super Stimmung von der ersten Minute an – das hat mir richtig gut gefallen. Ich denke, dass auch dieser neue Tag, der Bootcamp & Solutions Day, sehr gut funktioniert hat – wir sollten das auf jeden Fall weiterführen. Eine Sache möchte ich noch mal ganz besonders herausstellen: Wir hatten diesmal ganz tolle Studenten hier, die wirklich einen guten Eindruck hinterlassen haben, die engagiert dabei waren und auch richtig Interesse gezeigt haben für das Thema, das wir auch so lieben.“
Maik Becker, APEX connect NextGen

„Für mich war ein Highlight, dass wir diesmal auch Kollegen dabeihatten, die APEX noch gar nicht kannten – und die waren richtig geflasht davon. Aus den vielen Gesprächen nehme ich wiederum mit, wie offen die Leute hier sind und was man so alles auf den Weg bringen kann. Lobenswert ist auch die besondere Unterstützung seitens Oracle, dass Leute wie Kris Rice und Michael Hichwa extra aus Amerika rübergekommen sind, um auch zu hören, wie die Community hier so lebt in Deutschland und wie sie mit APEX umgeht. Sie können viel Feedback mit zurücknehmen. Es hat einfach alles gepasst. Ich freu mich schon auf nächstes Jahr.“
Niels de Bruijn, Konferenzleiter APEX connect

Anmerkung der Redaktion

Sämtliche Sessions der APEX connect 2024 wurden aufgezeichnet und können mit einem regulär erworbenen Ticket nachträglich über die On-demand-Agenda aufgerufen werden. Dies betrifft ebenso die Präsentationsunterlagen. Mit einem On-demand-Ticket der APEX connect 2024 ist dies auch nachträglich möglich. Es ist ab Kauf ein Jahr gültig.

by DOAG APEX connect

BEST OF DOAG ONLINE

Eine Auswahl der besten DOAG News April/Mai 2024

JavaLand 2024: Jubiläumsfeier am Nürburgring

Vom 9. bis 11. April feierte die JavaLand ihren zehnten Geburtstag am Nürburgring in der Eifel. Wir erzählen euch, was alles los war.

DOAG Datenbank Kolumne: Schnelle Lookups mit MemOptimized Rowstore

Wussten Sie schon, dass es einen speziellen Memory-Bereich in der Oracle-Datenbank für schnelle In-Memory Lookups gibt?

Observability und Datenbanken – Oracle-DBAs sind im Vorteil

Steven Lemme von Oracle gibt eine ausführliche Einführung in das Thema Observability für Datenbankadministratoren.

Aus Oracle 23c wird Oracle 23ai

Oracle hat – kurz vor der "DOAG Datenbank mit Exaday"-Konferenz – die neue Datenbankversion 23ai veröffentlicht.

DOAG.tv mit Prof. Dr. Stefanie Scholz und Andreas Buckenhofer über Chancen und Risiken des Citizen Data Scientist

Der DOAG Themenverantwortliche für Data Governance und Data Quality spricht in DOAG@Talk mit der Professorin für Data Science in Social Economy.

Highway to CloudLand – volle Fahrt voraus für die DOAG Roadshow

Die erste Woche der H2C-Tour lief blendend: Fünf Stationen wurden in Bayern mit dem DOAG-Tourbus für Talks, Pizza und Kaltgetränke angefahren. Die zweite Woche führt nun auch nach Hessen.

Wir begrüßen unsere neuen Mitglieder

Natürliche Mitglieder:

- Eva-Maria Ursula Schaller
- Sascha Schünemann
- Werner Fischer
- Florian Schafhauser
- Richard Allen

Korporative Mitglieder:

- rola Security Solutions GmbH, Repräsentant: Peter Weiß
- vS-Management Services GmbH, Repräsentantin: Sükran Aydin

Termine

Juni 06

18. - 21.06.2024
CloudLand 2024 – Das Cloud Native Festival
Community-Veranstaltung rund um die Themen Cloud- und Container-Technologien, Continuous Delivery, Microservices und DevOps
Brühl (Phantasialand)

20.06.2024
DOAG DevTalk: Datenmodellierung – Data Vault in operativen Systemen
DevTalk mit Matthias Schulz und Christian Schwitalla
Online

25.06.2024
Regionaltreffen Osnabrück/Bielefeld/Münster
Datenbank Konsolidierung – der Weg in die Cloudifizierung (von Exadata zu ExaCC)

26. - 27.06.2024
Expertenseminar: Entdecke die Welt der Künstlichen Intelligenz – Von Grundlagen zu Generativen Modellen
Berliner Expertenseminar mit Philip Weinmann
Berlin

Juli 07

15.07.2024
Regionaltreffen München/Südbayern
1. Was die Welt im Innersten zusammenhält – Oracle Transaktionen
2. Das Alert Log
München

September 09

10. - 11.09.2024
Expertenseminar: DSGVO 2024 – Aktuelles, Altbewährtes, Achtung!!!
Berliner Expertenseminar mit Sandra Leist
Berlin

Impressum

Red Stack Magazin wird gemeinsam herausgegeben von DOAG 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 ist die Community-Publikation für angewandte Informations- und Kommunikationstechnologie (ITK) im Raum Deutschland, Österreich und Schweiz. Es setzt bewusst auf Technologieoffenheit mit Blick auf Anwendung und IT-Innovationen.

Es bildet die Interessensschwerpunkte der Anwenderinnen und Anwender ab – von Cybersicherheit bis Datenschutz, von Datenbank und Development über Data Analytics bis Digitalisierung, von Cloud und Infrastruktur über Künstliche Intelligenz bis Open Source und Soft Skills – vermittelt praktisches Wissen und fördert den Know-how-Transfer und die Netzwerkbildung zwischen den Leserinnen und Lesern.

Die Inhalte des Red Stack Magazin werden von ausschließlich ehrenamtlichen Autorinnen und Autoren eingereicht und von der Redaktion aufbereitet.

Red Stack Magazin 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. DOAG e.V. hält 100 Prozent der Stammeinlage der DOAG Dienstleistungen GmbH. DOAG e.V. wird gesetzlich durch den Vorstand vertreten; Vorsitzender: Björn Bröhl

Redaktion:
Sitz: DOAG Dienstleistungen GmbH
(Anschrift s.o.)
ViSdP: Fried Saacke
Redaktionsleitung Red Stack Magazin:
Martin Meyer
Kontakt: redaktion@doag.org

Autorinnen und Autoren dieser Ausgabe
(in alphabetischer Reihenfolge):
Rainer Brehm, Lothar Flatz,
Dr. Harald Gerhards, Jan Gorkow,
Dr. Robert Helbig, Konstantin Hopf,
Marcos López, Martin Meyer,
Dr. Andreas Spies, Katharina Schraft,
Matthias Schulz, Oliver Szymanski,
Martin Talmeier, Dr. Christian Temath,
Armin Wildenberg

Titel, Gestaltung und Satz:
Diana Tkach
DOAG Dienstleistungen GmbH
(Anschrift s.o.)

Fotonachweis:
Titel: © rawpixel.com | *www.freepik.com*
S. 10: © KELLEPICS | *www.pixabay.com*
S. 18: © adventurous_blondine | *www.pixabay.com*
S. 24: © rawpixel.com | *www.freepik.com*
S. 26 © Mohammad_usman | *www.pixabay.com*
S. 34: @ upklyak | *www.freepik.com*
S. 50: © CR-3D | *www.pixabay.com*
S. 54: © Karabo_Spain | *www.pixabay.com*
S. 61: © aitoff | *www.pixabay.com*
S. 73: © freepik | www.freepik.com

Anzeigen:
sponsoring@doag.org

Mediadaten und Preise:
www.doag.org/go/mediadaten

Druck:
WIRmachenDRUCK GmbH,
www.wir-machen-druck.de

Inserentenverzeichnis

DOAG e.V. *www.doag.org* **U 2, U 3, U 4**

DOAG e.V *www.doag.org* **S. 3, S. 33, S. 49**

Promatis Gruppe *www.promatis.de* **S. 67**

European NetSuite User Days

November
2024
18 - 19

at NCC Ost
Nuremberg

netsuite.doag.org

#NetSuiteUserDays

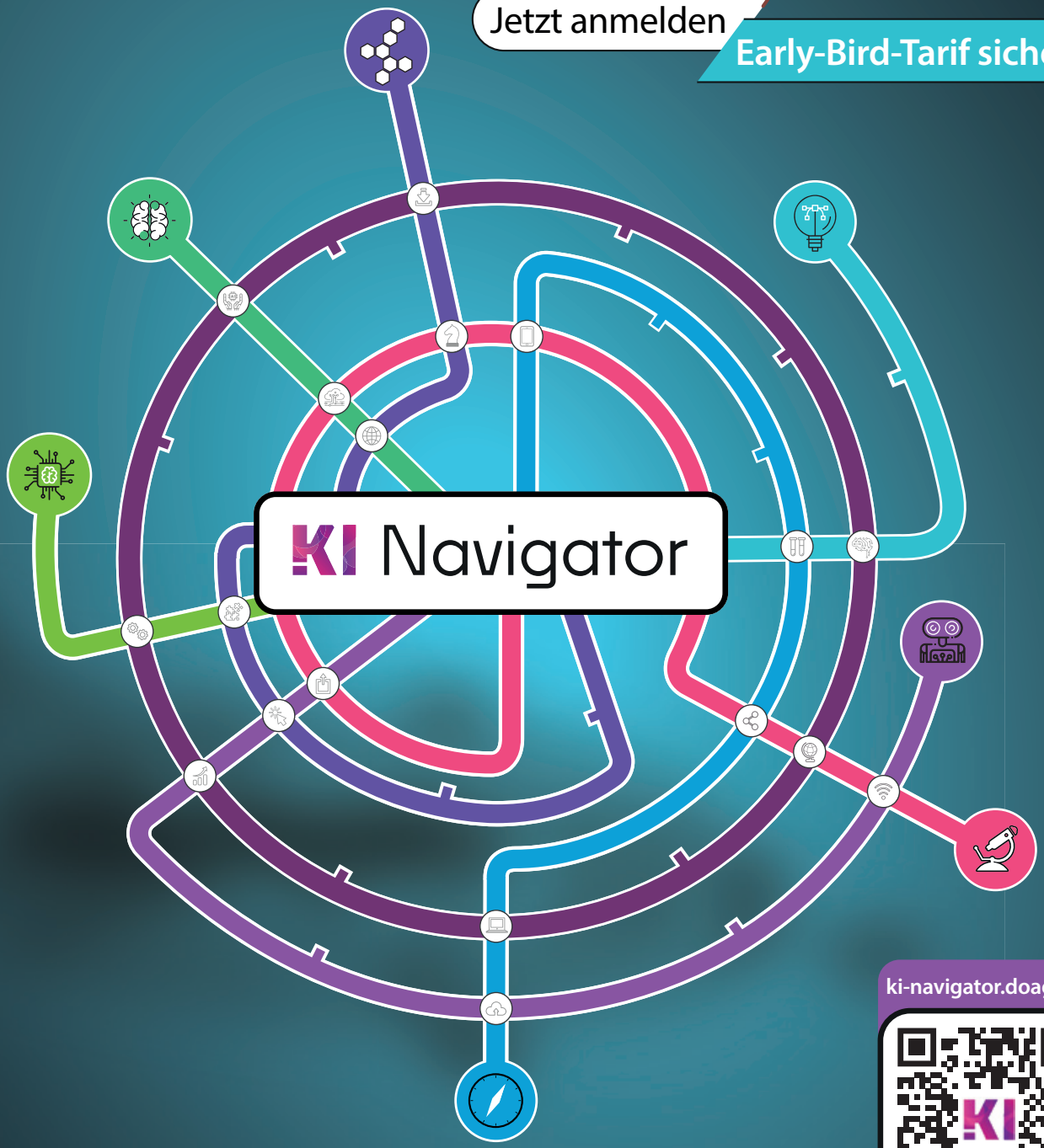
KI Navigator 2024

Konferenz zur Praxis der KI in IT, Wirtschaft und Gesellschaft

20. + 21. November in Nürnberg

Jetzt anmelden

Early-Bird-Tarif sichern



ki-navigator.doag.org

Veranstalter: DOAG Heise Medien dpunkt