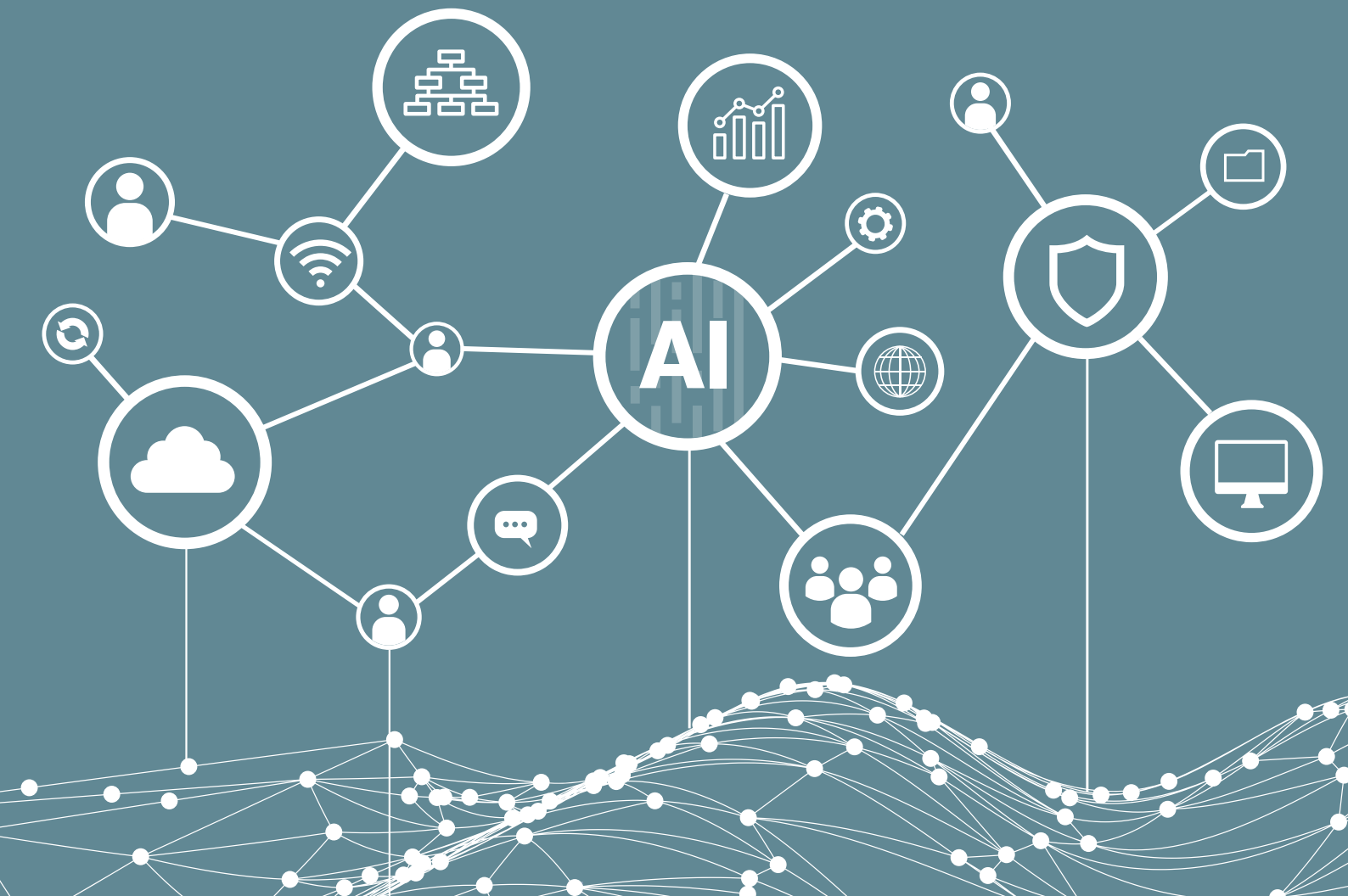


Aufbau eines leistungsstarken Daten- und KI-Unternehmens





Vorwort

„Aufbau eines leistungsstarken Daten- und KI-Unternehmens“ ist ein MIT Technology Review Insights-Bericht, der von Databricks gesponsert wurde. Für diesen Bericht führte MIT Technology Review Insights eine weltweite Umfrage unter 351 Chief Data Officers, Chief Analytics Officers, Chief Information Officers und weiteren Führungskräften aus dem Technologiebereich durch. Die Befragten sind zu jeweils gleichen Anteilen in Nordamerika, Europa und dem asiatisch-pazifischen Raum ansässig. In der Stichprobe sind 14 Branchen vertreten, und alle Befragten arbeiten in Unternehmen mit einem Jahresumsatz von mindestens 1 Mrd. US-Dollar. Die Untersuchung umfasste auch eine Reihe von Interviews mit Führungskräften, die in ihrem jeweiligen Unternehmen für Datenmanagement, Analysen und die damit verbundene Infrastruktur zuständig sind. Denis McCauley ist Verfasser, Francesca Fanshawe Lektorin und Nicola Crepaldi Erstellerin dieses Berichts. Die Untersuchungen erfolgten redaktionell unabhängig, und zum Ausdruck gebrachte Ansichten spiegeln die Meinung von MIT Technology Review Insights wider.

Wir möchten den folgenden Personen dafür danken, dass sie uns ihre Zeit und ihre Erkenntnisse zur Verfügung gestellt haben:

Patrick Baginski, Senior Director Data Science, McDonald's (USA)

Bob Darin, Chief Data Officer, CVS Health, und Chief Analytics Officer, CVS Pharmacy (USA)

Naveen Jayaraman, VP Data, CRM & Analytics, L'Oréal (USA)

Michel Lutz, Group Chief Data Officer, Total (Frankreich)

Mainak Mazumdar, Chief Data and Research Officer, Nielsen (USA)

Andy McQuarrie, Chief Technology Officer, Hivery (Australien)

Sol Rashidi, Chief Analytics Officer, The Estée Lauder Companies (USA)

Ashwin Sinha, Chief Data and Analytics Officer, Macquarie Bank (Australien)

Don Vu, Chief Data Officer, Northwestern Mutual (USA)



01 Executive Summary	4
02 Wachstum und Komplexität.....	6
Die Perspektive von Databricks:	
Der Aufstieg des Lakehouse-Effekts	7
03 Anpassen und Umsetzen von Strategien	9
Leistungsträger in Sachen Daten	11
Nielsen: Datentransformation	
für ein datenabhängiges Unternehmen.....	13
04 Skalieren von Analysen und Machine Learning	14
Paradigmenwechsel bei CVS Health.....	15
Hindernisse auf dem Weg zur Skalierung	16
Schutz für den ROI.....	17
Technologie, Demokratie und Kultur	18
05 Zukunftsvisionen	19
Wunschliste eines CDO für eine neue Architektur	19
06 Fazit.....	21

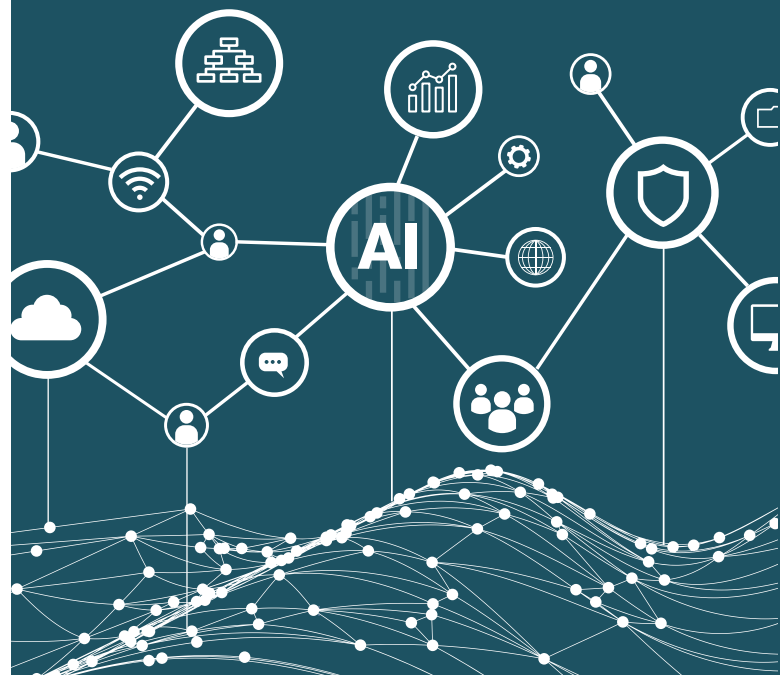


01 Executive Summary

COOs und Vorstände haben erkannt, dass die Fähigkeit ihres Unternehmens, aus Daten verwertbare Erkenntnisse zu gewinnen – und das möglichst noch in Echtzeit – von höchster strategischer Bedeutung ist. Die im vergangenen Krisenjahr erfolgte heillose Flucht der Verbraucher in die Online-Welt hat definitiv alle Zweifel diesbezüglich ausgeräumt. Damit Unternehmen datengestützt werden können, setzen sie zunehmend fortschrittliche Technologien auf Cloud-Basis ein, wie z. B. Analysetools mit Funktionen für Machine Learning (ML). Die Ergebnisse dieser Tools sind jedoch nur von begrenztem Wert, wenn nicht genügend hochwertige und leicht zugängliche Daten verfügbar sind.

In diesem Zusammenhang gehört ein effektives Datenmanagement zu den Grundvoraussetzungen eines datengestützten Unternehmens. Die Verwaltung von Daten in einem Unternehmen ist jedoch äußerst komplex. Mit der Einführung neuer Datentechnologien wächst die Belastung durch veraltete Systeme und Datensilos, sofern sich diese nicht integrieren oder anderweitig in den Griff bekommen lassen. Die Fragmentierung der Architektur bereitet vielen Chief Data Officers (CDOs) Kopfzerbrechen. Das liegt nicht nur an den Silos, sondern auch an der Vielzahl von On-Premises- und Cloud-basierten Tools, die in vielen Unternehmen eingesetzt werden. Zusammen mit einer geringen Datenqualität führen diese Probleme dazu, dass die Datenplattformen von Unternehmen – und die von diesen unterstützten Machine-Learning- und Analysemodelle – nicht die Geschwindigkeit und Skalierbarkeit bieten, die zum Erzielen der gewünschten Geschäftsergebnisse unverzichtbar sind.

Um zu verstehen, wie sich das Datenmanagement und die Technologien, auf die es sich stützt, inmitten solcher



Herausforderungen entwickeln, befragte MIT Technology Review Insights 351 CDOs, Chief Analytics Officers (CAOs – wir bezeichnen diese und die CDOs an verschiedenen Stellen des Berichts auch als „Data Leaders“) sowie Chief Information Officers (CIOs), Chief Technology Officers (CTOs) und weitere Führungskräfte im Technologiebereich. Außerdem haben wir ausführliche Interviews mit mehreren für Technologie zuständigen Führungskräften geführt. Im Folgenden sind die wichtigsten Ergebnisse dieser Untersuchung aufgeführt:

- **Nur 13 % der Unternehmen sind bei der Umsetzung ihrer Datenstrategie erfolgreich.** Diese ausgewählte Gruppe der „Leistungsträger“ liefert unternehmensweit nachweisliche Geschäftsergebnisse. Sie sind erfolgreich, weil sie Ihr Augenmerk den Grundlagen einer soliden Datenverwaltung und Architektur gilt. Diese ermöglichen es ihnen, Daten zu „demokratisieren“ und so von Machine Learning zu profitieren. Die Grundlagen sorgen für weniger Datenduplizierung, einen einfachen Zugang zu relevanten Daten, die Fähigkeit zur schnellen Verarbeitung großer Datenmengen und eine höhere Datenqualität. Darüber hinaus sind Leistungsträger fortgeschrittene Cloud-Anwender: 74 % betreiben mindestens die Hälfte ihrer Datenservices oder ihrer Infrastruktur in einer Cloud-Umgebung.



Die Prioritäten im Datenbereich, die für Unternehmen in den kommenden zwei Jahren am wichtigsten sein werden, lassen sich in drei Bereiche einteilen, die alle durch die zunehmende Verbreitung von Cloud-Plattformen unterstützt werden: **Verbesserung des Datenmanagements, Optimierung von Datenanalysen und ML sowie Ausweitung der Nutzung aller Arten von Unternehmensdaten**, darunter auch Streaming- und unstrukturierte Daten.


- **Technologiegestützte Zusammenarbeit schafft eine funktionierende Datenkultur.** Die für die Studie befragten CDOs messen der Demokratisierung von Analyse- und ML-Fähigkeiten große Bedeutung bei. Diese mithilfe fortschrittlicher Datentechnologien komplett auszureizen, hilft Endanwendern, fundiertere geschäftliche Entscheidungen zu treffen – und genau das ist das Markenzeichen einer starken Datenkultur. Möglich wird diese aber erst mit einer modernen Datenarchitektur. Ein CDO brachte es auf den Punkt, als er meinte, erfolgreiches Datenmanagement sei dann erreicht, wenn die richtigen Benutzer Zugriff auf die erforderlichen Daten haben, um schnell Erkenntnisse zu gewinnen, die einen Mehrwert für das Unternehmen darstellen.
- **Die Auswirkungen von Machine Learning auf das Unternehmen werden durch Schwierigkeiten beim Verwalten des End-to-End-Lifecycles relativiert.** Das Skalieren von ML-Anwendungsfällen ist für viele Unternehmen äußerst komplex. Die größte Herausforderung, so meinten 55 % der Befragten, sei das Fehlen einer zentralen Anlaufstelle zum Speichern und Ermitteln von ML-Modellen. Dieser Mangel sowie häufige Fehler bei der Übergabe von der Data Science an die Produktion und ein Mangel an qualifizierten ML-Ressourcen – beides wurde von 39 % der Befragten genannt – deuten auf erhebliche Schwierigkeiten bei der Umsetzung der Zusammenarbeit zwischen ML-, Daten- und Geschäftsanwenderteams hin.
- **Unternehmen wünschen sich Cloud-native Plattformen, die Datenmanagement, Analysen und Machine Learning unterstützen.** Die Prioritäten im Datenbereich, die für Unternehmen in den kommenden zwei Jahren am wichtigsten sein werden, lassen sich in drei Bereiche einteilen, die alle durch die zunehmende Verbreitung von Cloud-Plattformen unterstützt werden: Verbesserung des Datenmanagements, Optimierung von Datenanalysen und ML sowie Ausweitung der Nutzung aller Arten von Unternehmensdaten, darunter auch Streaming- und unstrukturierte Daten. Für „Nachzügler“ – d. h. Unternehmen, die Schwierigkeiten bei der Umsetzung ihrer Datenstrategie haben – stellen Verbesserungen beim Datenmanagement alle sonstigen Prioritäten in den Schatten (so zumindest äußerten sich 59 % der Befragten aus dieser Gruppe). Die Leistungsträger dagegen konzentrieren sich mehrheitlich (53 %) auf die Weiterentwicklung ihrer ML-Anwendungsfälle.
- **Offene Standards sind die wichtigste Voraussetzung für künftige Datenarchitekturstrategien.** Wenn Befragte eine neue Datenarchitektur für ihr Unternehmen aufbauen könnten, wäre der wichtigste Vorteil gegenüber der bestehenden Architektur eine stärkere Berücksichtigung von Open-Source-Standards und offenen Datenformaten. Wenn es darum geht, Innovation zu beschleunigen und für eine Auswahl der wirklich besten Tools von Drittanbietern zu sorgen, erkennen Data Leaders mittlerweile den Wert von Open-Source-Standards. Es überrascht nicht, dass mehr Sicherheit und eine bessere Verwaltung ganz oben auf der Liste der Anforderungen der Befragten stehen.

02 Wachstum und Komplexität

Die Geschwindigkeit, mit der sich das Datenmanagement in den Unternehmen verändert, ist gleichermaßen atemberaubend wie frustrierend. Einst von Führungskräften als Nebenprodukt des operativen Geschäfts betrachtet, gelten Daten heute als wichtigster Faktor für den Geschäftswert. Die generierten Datenmengen wachsen weiterhin rasant. Sie umfassen strukturierte, teilstrukturierte und unstrukturierte Datentypen, die Unternehmen nun speichern und analysieren können müssen.

Während Unternehmen noch vor nicht allzu langer Zeit gerade einmal auf eine Handvoll Technologieriesen zurückgreifen konnten, wenn es darum ging, ihren Bedarf an Dateninfrastruktur und -tools zu decken, haben sie heute die Qual der Wahl zwischen Hunderten von Anbietern in einem gigantischen Datenökosystem. Diese Akteure entwickeln laufend – und mittlerweile auch unter Zuhilfenahme von Machine Learning – neue Analysetools, die Daten in nie dagewesener Geschwindigkeit, Detailtreue und Raffinesse analysieren. Immer größer werdende Clouds bieten Unternehmen einen riesigen Speicherplatz und eine enorme Leistung zur Verarbeitung ihrer Daten – und das auf immer kosteneffizientere Weise. Nicht zuletzt haben sich auf verschiedenen Ebenen neue Rollen und Strukturen herausgebildet, wie z. B. der Aufstieg von Chief Data Officers (CDOs) und Chief Analytics Officers (CAOs), denn sie sollen die Datenkapazitäten des Unternehmens so kanalisieren, dass sie einen neuen, auf die strategischen Ziele abgestimmten geschäftlichen Mehrwert generieren.

„Früher war es für mich schwierig und kostspielig, Daten über viele Elemente unseres Kundenerlebnisses zu erhalten“, sagt Bob Darin, Chief Data Officer bei CVS Health (und Chief Analytics Officer bei CVS Pharmacy).



Die Cloud, die einst als optionale Technologieumgebung galt, ist heute Grundlage für die Modernisierung des Datenmanagements: **63 % der Befragten** nutzen Cloud-Dienste oder -Infrastrukturen umfassend in ihrer Datenarchitektur.

„Jetzt erhalte ich Erkenntnisse über unsere Kunden, unsere Lieferkette und die Arbeitsweise der Menschen, die mir vorher nicht zur Verfügung standen. Wir haben alle Tools, um diese Daten großflächig zu analysieren, und die Kosten für diese Tools sinken sogar noch. So können wir in großem Umfang Erkenntnisse gewinnen und so integrieren, dass sie Teil unserer Patienten- und Kunden-Workflows sind und wir unseren Kunden ein stärker personalisiertes und relevanteres Erlebnis bieten können.“

Die Cloud, die einst als optionale Technologieumgebung galt, ist heute Grundlage der Modernisierung des Datenmanagements, denn sie bietet immer mehr Speicherplatz und Rechenkapazität zu immer niedrigeren Kosten. Von den Unternehmen, die an unserer Untersuchung teilnahmen, nutzen 63 % in ihrer Datenarchitektur umfassend Cloud-Services oder -Infrastrukturen. Von diesen betreiben etwas mehr als ein Drittel (34 %) sogar mehrere Clouds.

Und doch tritt beim Datenmanagement immer wieder Frustration auf. Unternehmen, die versuchen, ein Upgrade ihrer Datenplattformen durchzuführen, werden häufig mit

Die Perspektive von Databricks: Der Aufstieg des Lakehouse-Effekts

Jedes Unternehmen spürt den Drang, ein Datenunternehmen zu werden, und setzt zunehmend auf künstliche Intelligenz, um deren enormes geschäftliches Potenzial auszuschöpfen. Doch gelingt es, wie dieser Bericht zeigt, nur 13 % der Unternehmen, ihre Datenstrategie erfolgreich umzusetzen. Die im Bereich Daten und Analysen führenden Unternehmen leiten ihren Erfolg zum großen Teil daraus her, dass sie die Grundlagen der Datenverwaltung beherrschen. Warum also tun sich dann so viele andere schwer damit?

Die erste Herausforderung ist die Datenarchitektur. Die Studie legt nahe, dass Unternehmen vier verschiedene Stacks konstruieren müssen, um alle erforderlichen Daten-Workloads zu bewältigen: Business Analytics, Data Engineering, Streaming und ML. Die vier Stacks erfordern jeweils sehr

unterschiedliche Technologien und kommen sich manchmal auch gegenseitig ins Gehege.

Das Technologieökosystem für Data Warehouses und Data Lakes verkompliziert die Architektur zusätzlich. Das Ganze ist ausgesprochen teuer und in der Verwaltung ressourcenaufwändig. Diese Komplexität hat Folgen für die Datenteams. Daten- und Unternehmenssilos können die Kommunikation ungewollt verlangsamen, Innovation behindern und zur Entstehung unterschiedlicher Ziele der verschiedenen Teams führen. Folge sind Datenduplizierung, das Fehlen eines konsistenten Sicherheits- und Governance-Modells, geschlossene Systeme und eine abnehmende Produktivität der Datenteams.

Gleichzeitig bleibt auch Machine Learning ein

Fortsetzung auf der nächsten Seite

den alten On-Premises-Silos konfrontiert, die sich einer einfachen Integration widersetzen, hohe Kosten aufwerfen oder aufgrund von Datenduplizierung und schlechter Qualität Probleme verursachen. Hierdurch entsteht eine Dateninfrastruktur, die sich durch eine hohe Komplexität auszeichnet. Die Cloud kann trotz ihrer bahnbrechenden Auswirkungen ebenfalls zur Komplexität beitragen, da Unternehmen ihre Daten weiterhin bei verschiedenen Anbietern speichern, um sich nicht an einen einzigen Anbieter zu binden, regionalen Anforderungen gerecht zu werden oder für Best-of-Breed-Lösungen ein optimales Umfeld zu schaffen. Zudem haben Datenarchitekturen sich in relativ kurzer Zeit weiterentwickelt, sodass Unternehmen gleichzeitig On-Premises-Datenbanken, Data Warehouses, Data Lakes und weitere neuartige Datenarchitekturen zusammen mit verschiedenen Cloud-basierten Tools für Konfiguration, Governance und sonstige Funktionen nutzen können.

„Architekturen sind richtig kompliziert geworden – das aber vor allem, weil wir dazu neigen, sie zu sehr zu verkomplizieren“, sagt Sol Rashidi, Chief Analytics Officer

bei The Estée Lauder Companies. „Wir machen das, weil wir das Wichtigste aus den Augen verlieren. Allzu oft implementieren wir wieder und wieder die neuesten und besten Technologien und Plattformen, weil wir glauben, dass sie unsere Probleme lösen werden. Wenn aber das Unternehmen noch gar nicht bereit ist, die Tools zu nutzen, und ihm die Reife fehlt, um aufschlussreiche Erkenntnisse zu extrahieren, und Prozesse und Logik sich nicht im Einklang befinden, dann ist es kein Wunder, dass das Architekturchaos immer größer wird.“

Unternehmen, die nicht in der Lage sind, die Komplexität zu bewältigen, leiden in der Regel im Nachgang unter verpassten Chancen (z. B. wenn ML-Anwendungsfälle keinen Gewinn abwerfen), höheren Kosten (weil etwa mehrere sich überschneidende Systeme verwaltet und unterstützt werden müssen) und Schwierigkeiten bei der Erfüllung der immer strikter werdenden gesetzlichen Anforderungen an Daten und sind so dem Wettbewerb letztendlich schutzlos ausgeliefert.

Trotzdem überwiegen, wie unsere Untersuchung zeigt, Enthusiasmus und Optimismus bei den für Daten und Technologie Verantwortlichen, wenn es um ihre aktuelle und zukünftige Fähigkeit geht, Daten für ihr Unternehmen wirksam zu verwalten.

Partner- perspektive

schwer zu erreichendes Ziel. Mit dem Aufkommen der Lakehouse-Architektur sind Unternehmen nicht mehr an die Beschränkungen und Komplexitäten von Legacy-Architekturen gebunden. Durch die Kombination der Leistung, Zuverlässigkeit und Governance von Data Warehouses mit der Skalierbarkeit, den günstigen Kosten und der Workload-Flexibilität von Data Lakes stellt die Lakehouse-Architektur einen flexiblen und leistungsstarken Entwurf für verschiedene Datenanwendungen dar. Dies schließt Echtzeit-Streaming, Batch-Verwaltung, SQL-Analysen, Data Science und ML ein.

Wir von Databricks erwecken die Lakehouse-Architektur mit unserer Lakehouse-Plattform zum Leben. Der Schlüssel zu dieser Innovation ist Delta Lake. Delta Lake ist das Herzstück der Plattform und erstellt kuratierte Data Lakes, die die für Data Warehouses typische Zuverlässigkeit, Leistung und Governance direkt in den bestehenden Data Lake einbringen. Unternehmen erhalten einen besseren Überblick über die unternehmensweite Datenverwaltung.

Die Lakehouse-Plattform von Databricks glänzt insbesondere in dreierlei Hinsicht:

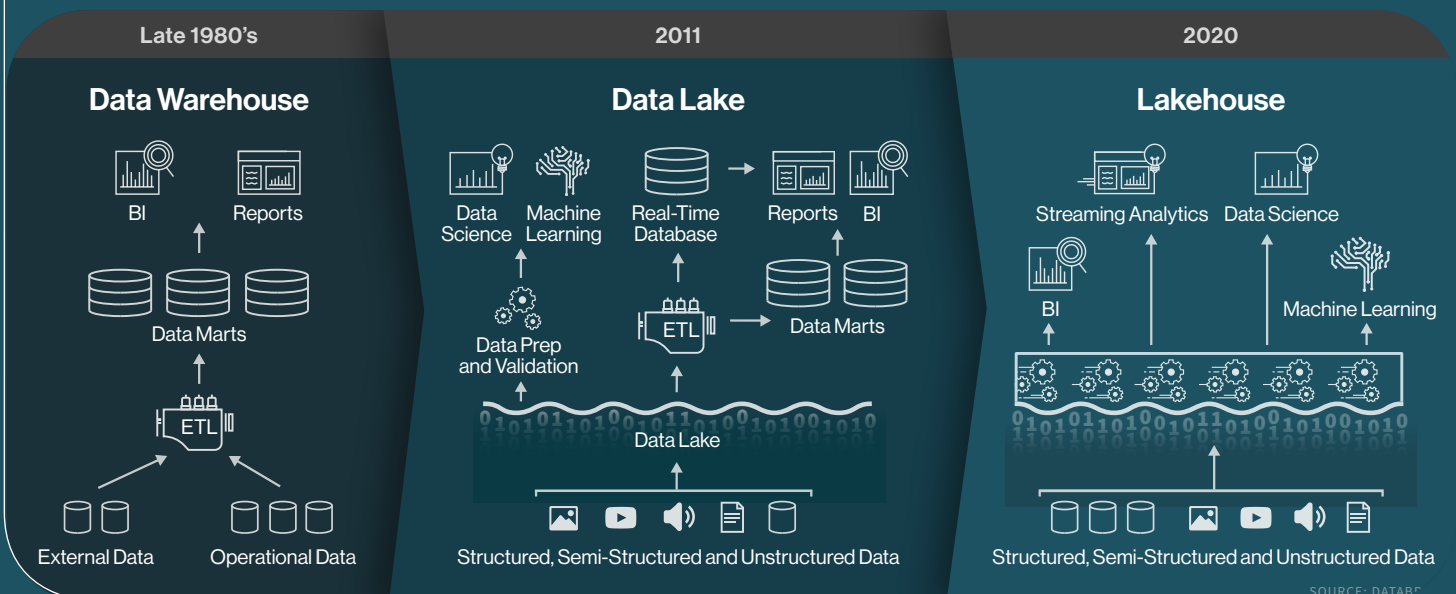
Sie ist einfach: Daten müssen nur einmalig vorhanden sein, um alle Workloads auf einer gemeinsamen Plattform zu unterstützen.

Sie ist offen: Da die Plattform auf Open Source und offenen Standards basiert, lässt sie sich problemlos mit bestehenden Tools integrieren. Proprietäre Formate werden vermieden.

Sie ist teamfähig: Data Engineers, Analysten und Data Scientists können gemeinsam effizienter arbeiten.

Allein die Kosteneinsparungen, die Effizienz und die Produktivitätssteigerungen, die mit der Databricks Lakehouse-Plattform einhergehen, wirken sich auf das Endergebnis von Unternehmen in allen Branchen und Regionen aus. Ganz ohne überkomplexe Architektur bietet Databricks eine gemeinsame Cloud-basierte Datenbasis für alle Daten und Arbeitslasten bei sämtlichen wichtigen Cloud-Anbietern. Da dank vorausschauender Wartung tägliche Planungsarbeiten und die damit einhergehenden Komplexitäten wegfallen, können Vorreiter in den Bereichen Daten und Analyse eine datengesteuerte Kultur fördern, in deren Zentrum die Wertschöpfung steht.

Ob Videostream-Analyse oder Customer-Lifetime-Value, ob Gesundheitsschutz oder die Suche nach Leben auf dem Mars: Daten sind immer Teil der Lösung. Daten zu verstehen, ist der Schlüssel, der alle Türen öffnet.



03 Anpassen und Umsetzen von Strategien

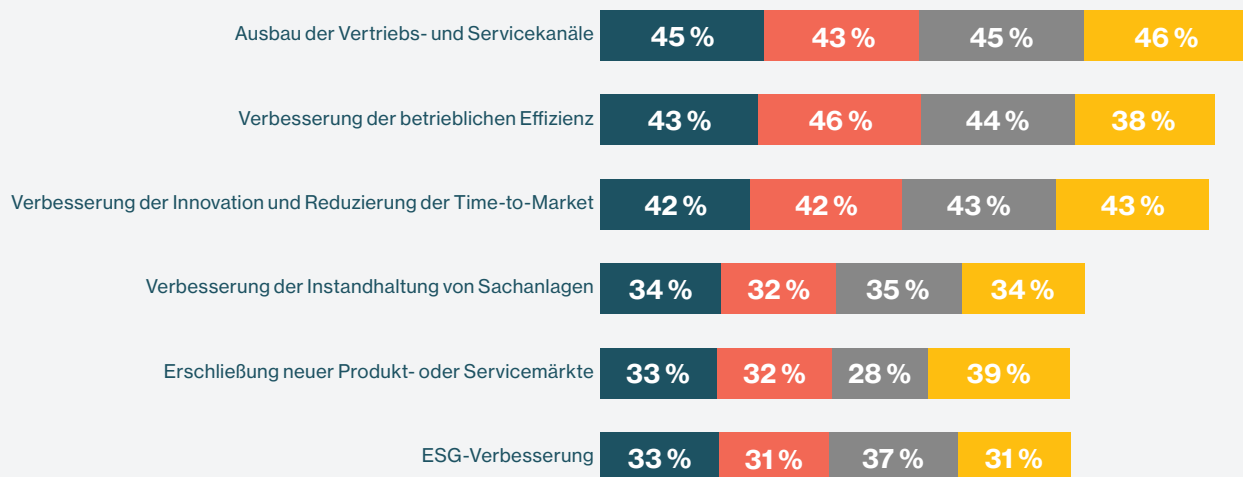


Inmitten eines weltweiten Wirtschaftsabschwungs, wie es ihn seit fast einem Jahrhundert nicht mehr gegeben hat, könnte man eigentlich erwarten, dass die Unternehmen ihre Ambitionen zügeln und vor allem auf den Gewinn achten. Nichtsdestoweniger scheinen jedoch viele der von uns befragten Unternehmen wachstumsorientiert zu sein. Auf die Frage nach den wichtigsten geschäftlichen Zielen, die sie sich für ihre Unternehmensdatenstrategie in den nächsten zwei Jahren gesetzt haben, betonen 45 % der Befragten eine Priorisierung der Umsatzsteigerung durch Ausbau von Vertriebs- und Servicekanälen. Dieser Wert übersteigt die Anzahl derjenigen, die auf Effizienzsteigerungen abzielen (43 %). Knapp dahinter folgen mit 42 % die Innovationsoptimierung und die Reduzierung der Time-to-Market neuer oder verbesserter Produkte.

Ein Blick auf die wichtigsten Dateninitiativen der befragten Unternehmen für die kommenden zwei Jahre deutet auf eine weitgehend konsequente Ausrichtung auf eine wachstumsorientierte Geschäftsstrategie hin. Ferner zeigt sich, dass die Unternehmen die Dringlichkeit der zur Unterstützung dieser Geschäftsziele erforderlichen Verbesserungen beim Datenmanagement erkannt haben. Die umfassende Einführung Cloud-nativer Plattformen wird diese und weitere Initiativen untermauern.

Die mit 48 % meistgenannte Priorität ist ein besseres Datenmanagement durch Optimierung der Datenqualität und -verarbeitung. (Dabei belaufen sich die Werte in der Mineralölbranche auf 74 % und in der

Abbildung 1: Wichtigste Geschäftsziele der Unternehmen für die Datenstrategie in den nächsten zwei Jahren (meistgenannte Antworten, % der Befragten)



Eine Verbesserung des Datenmanagements durch Optimierung der Datenqualität ist entscheidend dafür, dass wachstumsorientierte (also beispielsweise ML-gestützte) Bemühungen zügig zum Erfolg führen.

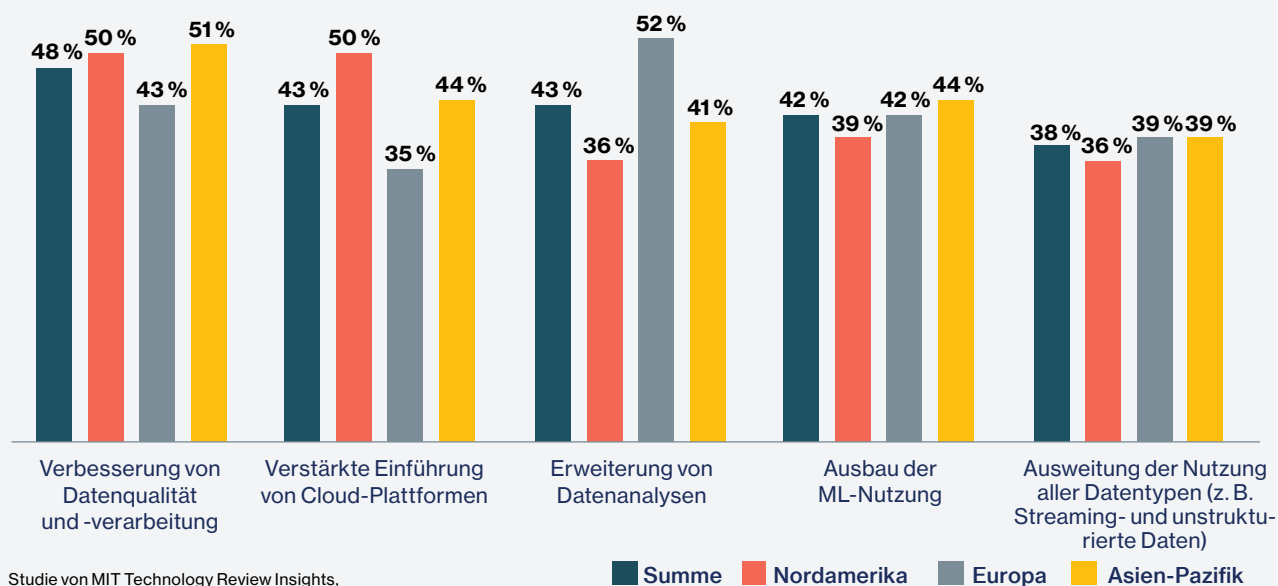
Konsumgüterindustrie auf 67 %.) Solche Anstrengungen sind entscheidend dafür, dass wachstumsorientierte (also beispielsweise ML-gestützte) Bemühungen zügig zum Erfolg führen. Für Hivery, ein in Australien ansässiges Handelsunternehmen, dessen Produkte auf künstlicher Intelligenz (KI) beruhen, ist die Qualität der Kundendaten noch wichtiger als die Möglichkeit der Erfassung großer Datenmengen. Andy McQuarrie, Chief Technology Officer bei Hivery, erklärt in diesem Zusammenhang, dass umso weniger Probleme bei der Erfassung auftreten, je sauberer die Kundendaten sind.

Die weiteren Hauptprioritäten der befragten Unternehmen im Datenbereich – die verstärkte Nutzung von Cloud-Plattformen (von 43 % genannt), die Verbesserung der Datenanalyse (43 %) und die Ausweitung der Anwendung von ML (42 %) – werden, sofern sie umgesetzt werden, den Datenteams mehr Potenzial bei Kapazität, Leistung und Skalierung bieten, um beispielsweise schnell neue Vertriebs- und Servicemöglichkeiten zu erschließen und die Entwicklung neuer Datenprodukte zu unterstützen. Sie unterstützen natürlich auch voll und ganz das Ziel einer Optimierung der betrieblichen Effizienz. Eine weitere, von 38 % genannte Priorität ist die Ausweitung der Nutzung

von Streaming-Daten, unstrukturierten Daten und weiteren Datentypen.

Dass die Datenstrategie eng mit den allgemeinen Geschäftszielen abgestimmt sein sollte, gilt heute eigentlich als selbstverständlich, aber die Wichtigkeit einer solchen Handlungsweise war nicht immer klar. Laut Don Vu, Chief Data Officer des US-amerikanischen Finanzdienstleisters Northwestern Mutual, ist die Abstimmung von Daten- und Geschäftsstrategien in vielen Unternehmen sehr viel enger geworden, denn die CDOs haben ihren Einfluss geltend gemacht, und die Unternehmensstrukturen wurden optimiert, um die Zuständigkeiten in Datenfragen besser zusammenzuführen. In seinem Unternehmen, so Vu, „war zwar bekannt, dass eine Angleichung grundsätzlich wichtig ist, aber so richtig kristallisierte sich dies erst heraus, als unsere Teams sich intensiver mit der Frage befassten, wie wir die verschiedenen Initiativen zur Geschäftsstrategie tatsächlich umsetzen wollten. Die Verknüpfung von Ideen wie dem Vertrauen in notwendige vertrauenswürdige Datenquellen oder einer Demokratisierung der Datennutzung mit der Unternehmensstrategie wurde dadurch viel deutlicher.“

Abbildung 2: Wichtigste unternehmensweite Datenstrategieinitiativen in den nächsten zwei Jahren
(meistgenannte Antworten, % der Befragten)



Leistungsträger in Sachen Daten

Es gibt nicht viele große Unternehmen, die beim Datenmanagement brillieren. Dies spiegelt sich in der Umfrage wider, in der nur 13 % der Befragten die Leistung ihres Unternehmens in Bezug auf die Umsetzung der Datenstrategie hoch einschätzten, d. h. es auf einer Skala von 1 bis 10 mit einem hohen Wert (also 9 oder 10) einstufen. Diese „Leistungsträger“ in Sachen Daten erzielen nach Aussage ihrer Führungskräfte

spartenübergreifend messbare Auswirkungen auf das Geschäft. Ihnen gegenüber steht eine ähnlich große Gruppe von Nachzüglern (12 % der Stichprobe), deren Leistungsfähigkeit auf der Datenskala mit 6 oder schlechter bewertet wurde.

Zwischen diesen beiden Gruppen zeigen sich bei bestimmten Eigenschaften und Absichten erhebliche

Abbildung 3a: Ausmaß, in dem Unternehmen ihre Datenstrategie erfolgreich umsetzen
(Selbsteinschätzung auf einer Skala von 1–10, wobei 10=erfolgreich)

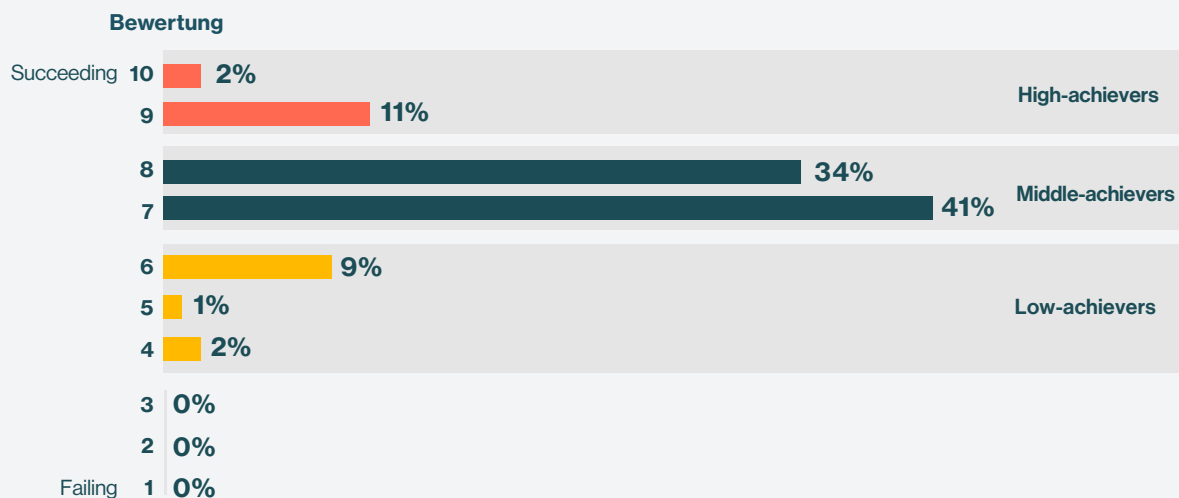


Abbildung 3b: Leistungsträger: Befragte, die ihr Unternehmen auf den Wert 9 oder 10 für die Umsetzung der Datenstrategie mit spartenübergreifend messbaren Auswirkungen auf das Geschäft einstufen
(insgesamt, Regionen und ausgewählte Branchen)

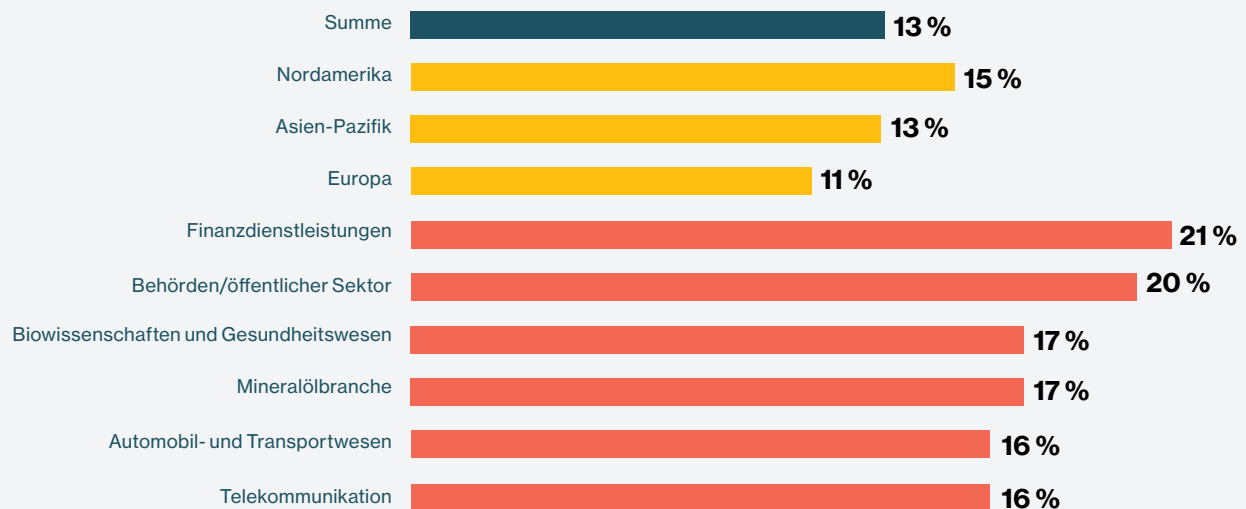
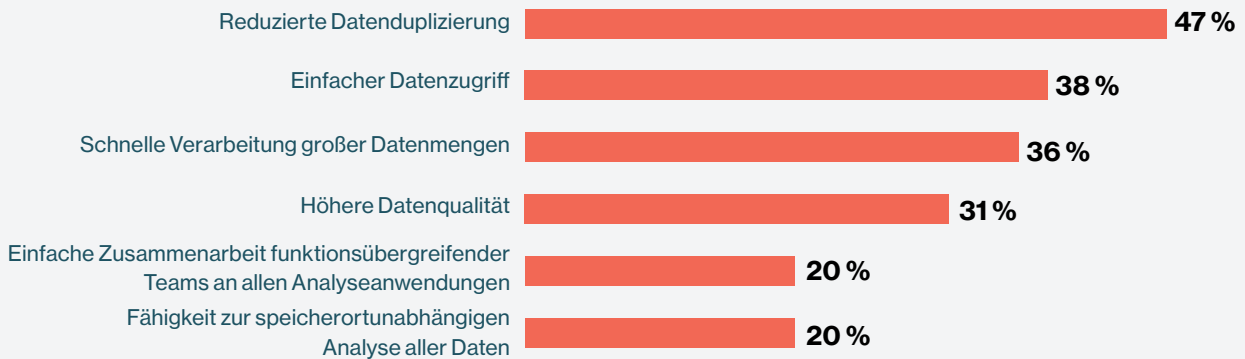


Abbildung 4: Wichtigste Erfolgsfaktoren für „Leistungsträger“ bei der Umsetzung ihrer Datenstrategieinitiativen (meistgenannte Antworten; % der Befragten)



Studie von MIT Technology Review Insights, 2021

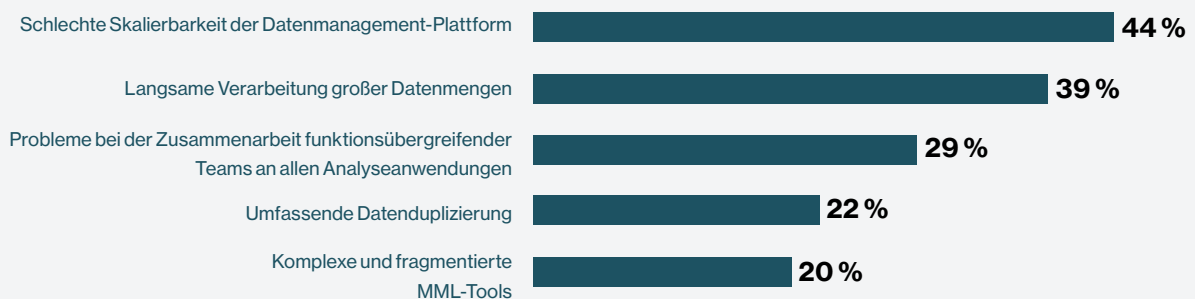
Diskrepanzen. So spielt beispielsweise die Cloud in der Datenarchitektur von Leistungsträgern eine größere Rolle: 74 % dieser Gruppe betreiben mindestens die Hälfte ihrer Datenservices oder ihrer Infrastruktur in einer Cloud-Umgebung, bei den Nachzüglern liegt der Wert dagegen nur bei 60 %. Bei den Datenprioritäten wollen die meisten leistungsschwächeren Unternehmen (59 %) in den nächsten zwei Jahren den Schwerpunkt auf die Verbesserung des Datenmanagements (Datenqualität und -verarbeitung) legen, während die am häufigsten genannte Initiative der Leistungsträger die Ausweitung der ML-Nutzung ist (53 %).

Die Leistungsträger betrachten die Grundlagen eines soliden Datenmanagements dabei nicht als selbstverständlich, sondern führen ihren Erfolg darauf zurück, dass sie diesen Grundlagen eine hohe Aufmerksamkeit geschenkt haben. Dies umfasst etwa die Reduzierung duplizierter Daten, einen vereinfachten Datenzugriff für Endanwender im Unternehmen – ein charakteristisches Merkmal der sogenannten

„Demokratisierung“ von Daten – und die Verarbeitung großer Datenmengen mit hoher Geschwindigkeit. Die Datenduplizierung findet in großen Unternehmen auf mehreren Ebenen statt, z. B. in Data Warehouses, auf operativen Systemen, in Berichten, Dashboards und Desktop-Tools. Sie habe erhebliche Auswirkungen auf Kosten, Risikomanagement und Zuverlässigkeit, so Ashwin Sinha, Chief Data & Analytics Officer bei der Macquarie Bank. Die Duplizierung von Daten wirkt sich zudem auf die Fähigkeit aus, Machine Learning unternehmensweit zu skalieren und effektiv zu nutzen.

Auf die Frage, was den Fortschritt in ihrem Unternehmen ausbremst, nannten die meisten Nachzügler die begrenzte Skalierbarkeit ihrer Datenmanagementplattform. Weitere häufig genannte Beeinträchtigungen sind die langsame Verarbeitung großer Datenmengen und Schwierigkeiten bei der Einrichtung der Kollaboration. Wie wir noch sehen werden, sind Skalierung, Geschwindigkeit und Kollaboration Herausforderungen für U

Abbildung 5: Zentrale Herausforderungen für „Nachzügler“ bei der Umsetzung ihrer Datenstrategieinitiativen (meistgenannte Antworten; % der Befragten)



Studie von MIT Technology Review Insights, 2021

Nielsen: Datentransformation für ein datenabhängiges Unternehmen



Für das berühmte Marktforschungsunternehmen Nielsen ist die Bedeutung eines soliden Datenmanagements kaum zu überschätzen. Es handelt sich hierbei um eines der wenigen hundertjährigen Unternehmen, bei denen Daten vom ersten Tag an zentraler Bestandteil des Geschäftsmodells waren. Die Nielsen-Panels geben Konsumgüterunternehmen Aufschluss darüber, welche Produkte Kunden kaufen und wie sich Verhaltensweisen wandeln. Außerdem beraten die Panels diese Unternehmen auch zu der Frage, wann und wo sie TV-Werbung schalten sollten. In seinem zweiten Jahr als Chief Data Officer und seinem fünften Jahr als Chief Research Officer leitet Mainak Mazumdar die Transformation des Datenmanagements und der Infrastruktur des Unternehmens. „Noch vor wenigen Jahren“, erinnert er sich, „haben wir uns an der Datenfragmentierung aufgegeben: jede Menge Daten in Silos und dazu undokumentiertes Wissen, das für den Zugriff darauf unentbehrlich war. Außerdem fehlte es an Metadaten und auch an Governance. Gleichzeitig wuchsen die Volumina tagtäglich um mehrere Petabytes an.“

Für heute zeichnet Mazumdar ein anderes Bild: „Mittlerweile sind wir in der Lage, schnell von 20 bis 30 Spezialisten auf einer Plattform auf über 300 zu skalieren. Wir arbeiten auf einer Cloud-Plattform mit einem Data Lake, in dem die Daten kuratiert, beschriftet, definiert, mit Metastores versehen und konsolidiert werden. Wir haben unsere eigene Analyse-Engine entwickelt. Mehr noch: Vieles von dem, was früher die Software Engineers machten, wird

heute von meinem Team erledigt und direkt in die Produktion überführt.“ Laut Mazumdar haben die Änderungen die Zykluszeit für sein Team um 50 % reduziert. „Unserer Modelle heute arbeiten mit etwa 50-facher Geschwindigkeit. Was früher 20 Minuten dauerte, können wir heute in einer Minute oder noch schneller erledigen. Gleichzeitig erfassen und verarbeiten wir enorme Datenmengen, die für Data Scientists leicht zugänglich sind. Das ist eine riesige Veränderung.“

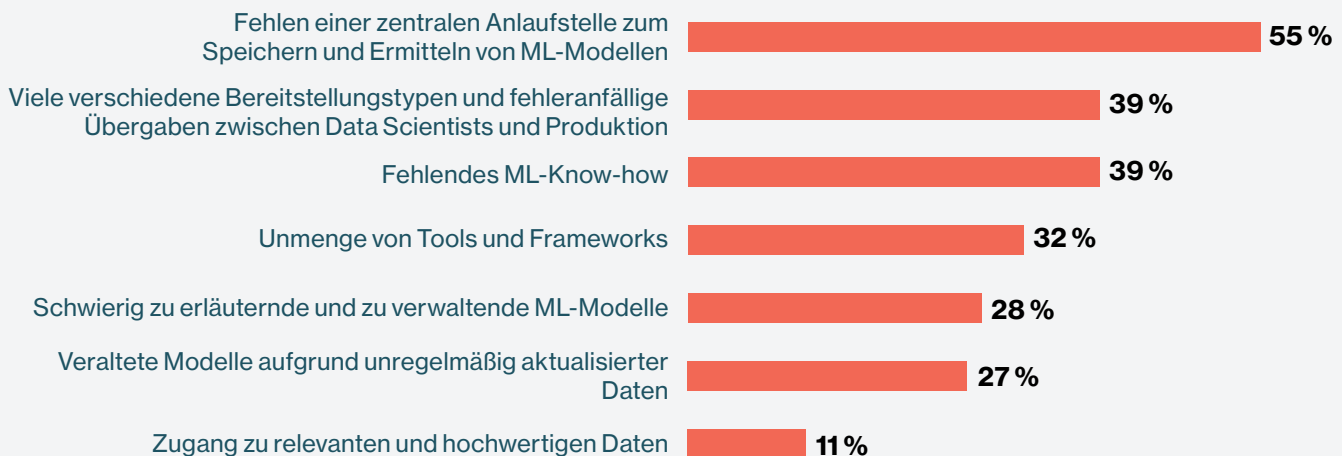
Ein Beispiel dafür, wie Nielsen mithilfe dieser neuen Fähigkeiten Wachstum ermöglicht hat, ist die Einführung eines neuen Bewertungsprodukts in den rund 200 lokalen Märkten des Unternehmens in den USA. Basierend auf der Verarbeitung großer Datenmengen von TV-Set-Top-Boxen ermöglicht das „High-Recognition-Deep-Learning-Modell“ es Nielsen nicht nur, vorherzusagen, was sich ein Zuschauer zu einem bestimmten Zeitpunkt wahrscheinlich gerade ansieht, sondern auch, wer in einem Haushalt gerade zuschaut – und das war bislang nicht möglich gewesen. „Wir hätten dieses Produkt ohne Anpassungen unseres Datenmanagements nicht einführen können, und wir hätten auch nicht solche Datenmengen erfassen können“, sagt Mazumdar. „Durch die Aufnahme immer größerer Datenmengen werden die Modelle – und damit auch das Produkt – immer besser.“

Skalieren von Analysen und Machine Learning

Führungskräfte wissen, dass die Fähigkeit ihres Unternehmens, mit der Nachfrage Schritt zu halten und sie zu antizipieren, den Wettbewerbsdruck zu bewältigen, effektiv Innovation zu betreiben und effizient zu arbeiten, mit der Beherrschung von Analysen und ML steht und fällt. Unternehmen in praktisch jeder Branche befassen sich gegenwärtig mit der Entwicklung von Analyse- und ML-Anwendungsfällen, die einen erheblichen geschäftlichen Einfluss haben werden. Die meisten großen Unternehmen betrachten es inzwischen nicht mehr als rein optional, ein umfassendes Anwendungsportfolio zu entwickeln und auszubauen. CDOs und ihre Teams werden zunehmend danach beurteilt, welchen Beitrag sie zur Umsetzung solcher Anwendungen leisten.

Allerdings tun sich viele Unternehmen insbesondere damit schwer, die für eine spürbare Wirkung nötige Größenordnung zu erzielen. Nach Meinung von Sol Rashidi von Estée Lauder ist ein Grund dafür zu viel Ehrgeiz: „Allzu oft wollen Unternehmen beim Machine Learning die Krabbel- und Gehphasen überspringen und direkt loslaufen, ohne sich mit den Grundlagen vertraut gemacht zu haben.“ Für andere CDOs – z. B. Don Vu von Northwestern Mutual – liegen die größten Herausforderungen in der Auswahl der passenden Anwendungsfälle für die Überführung in die Produktion. Ohne den Input geschäftlicher Anwender, so meint er, steige die Wahrscheinlichkeit einer Entscheidung für Anwendungen, die sich einem Geschäftsziel gerade nicht eindeutig zuordnen lassen.

Abbildung 6: Wesentliche Schwierigkeiten, auf die Unternehmen bei der Skalierung von ML-Anwendungen stoßen
(meistgenannte Antworten; % der Befragten)



Paradigmenwechsel bei CVS Health



Apotheken spielen seit jeher eine wichtige Rolle in der Gesellschaft, jedoch noch nie so sehr wie im vergangenen Jahr. Apothekenketten wie CVS, das umsatzstärkste Apothekenunternehmen der USA, stellen mithilfe von immer fortschrittlicheren Datenfunktionen sicher, dass ihre Kunden über verschriebene Arzneien und die Inanspruchnahme weiterer Gesundheitsleistungen auf dem Laufenden bleiben, damit die aktuelle körperliche Verfassung nicht zu ernsteren gesundheitlichen Folgen führt. Bob Darin ist als Chief Data Officer bei CVS Health sowie als Chief Analytics Officer des Apothekengeschäfts des Unternehmens federführend bei vielen derartigen Initiativen.

Schon seit langem nutzt das Unternehmen Datensysteme, um dafür zu sorgen, dass sich Kunden über ihre Medikamente auf dem Laufenden halten. Dazu gehörten beispielsweise freundliche Erinnerungen via Telefon oder SMS, die gezielte Ansprache in der Apothekenfiliale oder Empfehlungen an Patienten, mit ihrem Arzt über bestimmte Nachsorgemaßnahmen oder die Medikamentierung zu sprechen. In den letzten Jahren wurden diese Initiativen in die Data Science eingebettet, um sie stärker zu personalisieren, so Darin. „Wir wissen jetzt, welche Gedächtnisstützen, Programme, Kommunikationsformen und Botschaften am effektivsten sind, um Patienten dabei zu unterstützen, an die Einnahme ihrer Medikamente zu denken, ihre Abholtermine zu

planen und Möglichkeiten zur Kosteneinsparung zu verstehen.“ All das wird jetzt durch Analysemodelle gesteuert, meint er – früher sei hingegen ein Universalverfahren für alle Punkte zum Einsatz gekommen.

„Für uns war das ein Paradigmenwechsel“, sagt Darin. „Das betrifft nicht nur die Art und Weise, wie wir Analysen nutzen, sondern auch, wie wir unsere Datenplattform und -architektur verwalten.“ Seinen Angaben zufolge schloss dies auch Investitionen in Data Warehousing, operatives Reporting und Analysen ein. Seine Teams seien jedoch auf dem Weg, „mehr Geschäftsfunktionen in eine hybride, mehrdimensionale Datenumgebung zu verlagern – eine Umgebung, die nicht nur Anwendungen für Data Warehousing und Descriptive Analytics unterstützt, sondern auch den Einsatz fortschrittlicherer Verfahren für maschinelles Lernen, die Entwicklung von Algorithmen und Optimierungen im großen Maßstab.“

Eine solche Umgebung müsse nach den Worten von Darin „sowohl fortschrittlichere Data-Science-Anwendungen als auch im Tagesgeschäft eingesetzte Geschäftsanalysen unterstützen, wie z. B. Beschreibungen und Ad-hoc-Abfragen – eben die Art von Informationen, die für geschäftliche Entscheidungen tagtäglich benötigt werden. Wir brauchen Architekturen und Datenplattformen, die dies beides unterstützen.“

Drei Viertel der Befragten aus Energiewirtschaft und von Versorgungsunternehmen wie auch mehr als zwei Drittel der Befragten bei Behörden und in Konsumgüterunternehmen geben an, dass das Fehlen einer zentralen Möglichkeit für Speicherung und Ermittlung von ML-Modellen ein zentrales Hindernis für die Skalierung darstellt.

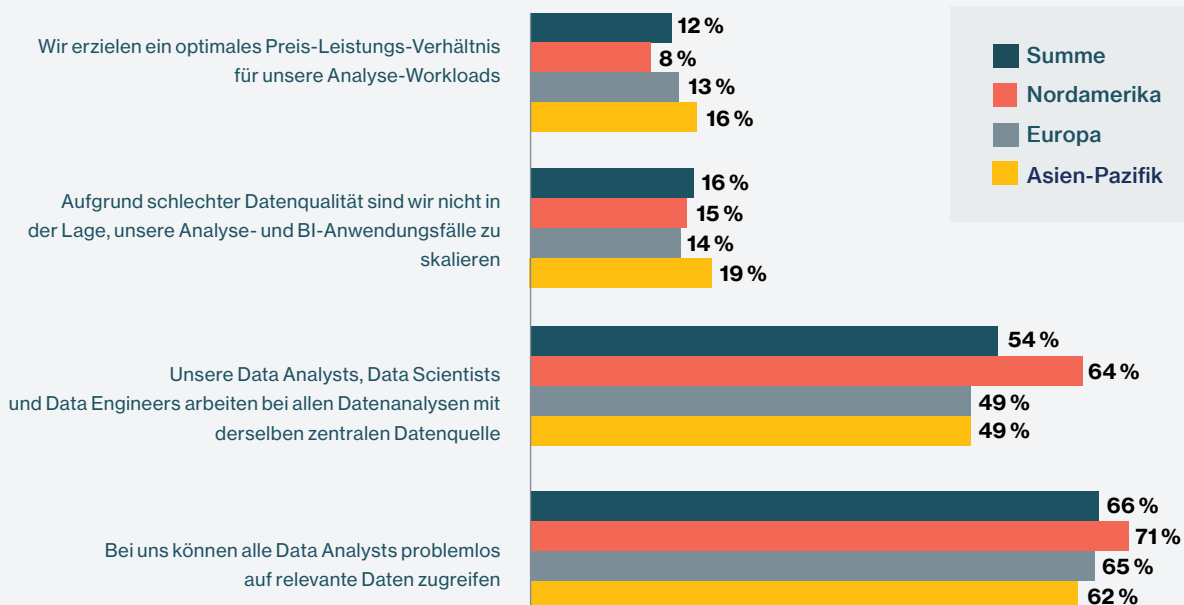
Ein solider Dialog zwischen Entwicklern von ML-Modellen und geschäftlichen Anwendern beugt dem sogenannten „Shiny-Algorithm-Effekt“ vor, bei dem das Data-Science-Team ML-Anwendungsfälle priorisiert, die zwar für Informatiker hochgradig spannend, für Geschäftsanwender jedoch nur von geringem Nutzen sind. „Wenn Ihre ML-Anwendung nicht eng an einen geschäftlichen Anwendungsfall angeglichen ist“, so sagt Andy McQuarrie von Hivery, „dann frage ich mich schon, was das soll.“

Hindernisse auf dem Weg zur Skalierung

Unternehmen tun sich mit der komplexen Verwaltung des gesamten ML-Lebenszyklus eindeutig schwer. Dies lässt sich auch der Liste der Schwierigkeiten entnehmen, auf die die Unternehmen der Befragten bei der Skalierung von Anwendungsfällen stoßen (Abbildung 6). Die größte Herausforderung sei nach Ansicht von 55 % der Befragten

das Fehlen einer zentralen Anlaufstelle zum Speichern und Ermitteln von ML-Modellen. Drei Viertel der Befragten aus Energiewirtschaft und von Versorgungsunternehmen wie auch mehr als zwei Drittel der Befragten bei Behörden und in Konsumgüterunternehmen geben an, dass dies ein beträchtliche Hürde auf dem Weg zur erfolgreichen Skalierung sei. Eine unzureichende Zusammenarbeit zwischen Data Science und Produktion, die sich in Mehrfachimplementierungen und fehleranfälligen Übergaben niederschlägt, ist laut 39 % ein zentrales Hindernis für die Skalierung. Bei den Befragten aus den Bereichen Biowissenschaften und Gesundheitswesen liegt diese Zahl sogar bei 50 %. Zudem haben viele Unternehmen mit einer Unmenge von Tools und Frameworks sowie einer unregelmäßigen Aktualisierung der Modelldaten zu kämpfen (32 % bzw. 27 %). Das Problem der Mehrfachbereitstellung und der fehleranfälligen Übergabe zwischen Data Science und

Abbildung 7: Befragte, die den Aussagen zu Analysen und Business Intelligence (BI) in ihrem Unternehmen tendenziell oder voll und ganz zustimmen (% der Befragten)





„Mehrfachbereitstellung und eine fehleranfällige Übergabe zwischen Data Science und Produktion stellen ein erhebliches Problem dar. Es gibt oft eine Diskrepanz zwischen dem Output der Data Scientists und den Ergebnissen, die wir nach der Operationalisierung erhalten.“

– Naveen Jayaraman, VP Data, CRM & Analytics, L'Oréal

Produktion ist eines, das die von uns befragten CDOs offenbar gut kennen. „Das ist eine echte Herausforderung“, sagt Naveen Jayaraman, VP Data, CRM & Analytics von L'Oréal, einem Anbieter von Körperpflegeprodukten. „Es gibt oft eine Diskrepanz zwischen dem Output der Data Scientists und den Ergebnissen, die wir nach der Operationalisierung erhalten.“

Laut Bob Darin ist es CVS Health gelungen, die Übergabeproblematik in den Griff zu bekommen: „In der Vergangenheit erstellten unsere Data Scientists Modelle und gewannen daraus aussagekräftige Informationen. Diese gaben sie zur Implementierung an ein anderes Team weiter. Diese ‚Wasserfallmethodik‘ hat aber nicht richtig funktioniert, und die Modelle waren nicht gut skalierbar.“ Seinen Worten nach hat die Integration von Data Science und Produktion in einzelnen Teams, die mit der IT-Abteilung zusammenarbeiten, bei CVS Health dazu beigetragen, diese Schwierigkeiten zu überwinden. „Das ist viel besser, als wenn ein Team Erkenntnisse entwickelt und ein anderes die Skalierung vornimmt.“

Die Kompetenzlücke ist und bleibt ein systemisches Problem für Unternehmen, die erfolgreiche ML-Praktiken aufbauen wollen. Das Fehlen von handfestem Know-how im Bereich Machine Learning wird von 39 % der Befragten als zentrales Hindernis für die Skalierung von Anwendungen genannt und erweist sich insbesondere für die an der Umfrage teilnehmenden Fertigungsunternehmen als echter Hemmschuh (64 %). Die Leistungsträger sind jedoch anscheinend in der Lage, solche Lücken zu überwinden, denn nur 27 % von ihnen

nannten dieses Problem. Nachzügler hingegen haben damit schwer zu kämpfen: 59 % gaben an, dass dies für sie eine wesentliche Herausforderung sei.

Schutz für den ROI

Schwierigkeiten beim Aufbau von ML- und Analyseanwendungen in der geplanten Größenordnung wirken sich auf den Return on Investment (ROI) aus. So meinten beispielsweise nur 12 % der Befragten, dass sie ein optimales Preis-Leistungs-Verhältnis für ihre Analyse-Workloads erzielen würden. Die bereits erwähnten Probleme mit der Datenduplizierung tragen wahrscheinlich ebenfalls zur Leistungsbeeinträchtigung bei, ebenso wie die Kosten für Beauftragung ohnehin knapper Ressourcen, sich mit der anspruchsvollen Bereinigung von Daten minderer Qualität zu befassen. Um ein unternehmensweites Ausmaß zu erreichen, sind häufig Investitionen in die Cloud-Infrastruktur und das Datenmanagement erforderlich, sagt Jayaraman. „Das ist schwieriges Terrain“, erläutert er. „Solange eine Anwendung keine großen Gewinne im Vertrieb abwirft oder massive Effizienzsteigerungen zur Folge hat, wird der ROI-Aspekt immer wieder diskutiert werden.“

Das Erfüllen oder Übertreffen von Leistungserwartungen kann dagegen mehr als nur einen ROI-Nutzen bieten, meint Patrick Baginski, Senior Director Data Science beim Fastfoodriesen McDonald's. „Bei ML ist die Bedeutung der Time-to-Value immens. Je schneller Sie den Nutzen von ML und Data Science nachweisen, desto schneller können Sie die Nutzer überzeugen und die Zuversicht der Geschäftsleitung betreffs des potenziellen Mehrwerts für das Unternehmen stärken.“ Dies wiederum trägt wesentlich zur Entwicklung der Datenkultur bei.

Technologie, Demokratie und Kultur

Trotz des Wandels, den das Datenmanagement bei Systemen, Führungskräften und Wahrnehmung des Geschäftswerts durchlaufen hat, besteht in den meisten Unternehmen nach wie vor eine immense Kluft zwischen Datenteams und Endanwendern, d. h. den Mitarbeitern im Front- oder Back-Office, die tagtäglich datengestützte Informationen benötigen, um Entscheidungen zu treffen. Viele CDOs versuchen, diese Kluft zu überbrücken, indem sie Data Scientists direkt in die Geschäftsbereiche einbinden, damit sie dort regelmäßig mit Anwendern interagieren. Eine weitere Möglichkeit, die Lücke zu schließen, besteht darin, die Analysen den Nutzern direkt zur Verfügung zu stellen, sodass sie bei Bedarf selbst Erkenntnisse daraus ziehen können. Das allerdings, so Patrick Baginski von McDonald's, mache es auch erforderlich, dass die Daten näher an den „Rand“ verlagert werden – also dorthin, wo der Anwender sich befindet.

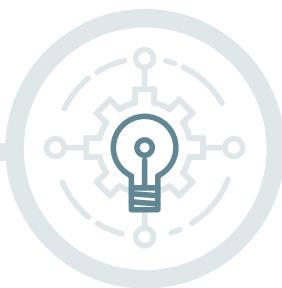
Dies ist ein Ziel, das sich alle von uns befragten CDOs und CAOs auf die Fahnen geschrieben haben. „Angesichts der Weiterentwicklung der Tools und der zunehmenden Vertrautheit der Menschen mit erweiterten Analysen und Data Science müssen wir Nutzern die Möglichkeit bieten, Analysen selbst durchzuführen, statt nur von anderen erstellte Analysen zu verwenden“, so Bob Darin.

So sieht Datendemokratisierung in der Praxis aus. Für einige CDOs stellt sie zudem ein Mittel dar, um nicht nur Datenkompetenz aufzubauen, sondern auch die Entwicklung einer Datenkultur im Unternehmen zu fördern. Dateninfrastruktur und Daten-Tools tragen unmittelbar dazu bei. „Unser Ziel als zentrales Datenteam ist es, Tools zu entwickeln, die einen Großteil der Verantwortung und Entscheidungsbefugnis an andere abgeben“, sagt

Don Vu. „Wir versuchen daher, die Leistungsfähigkeit unserer Datenplattform an den Rand zu verlagern. Wir wollen unter den Anwendungsteams und Anwendern in Bezug auf die Funktionen eine Anziehung hervorheben. Wir wollen, dass die Leute diese Plattform gerne nutzen, weil sie Aspekte wie Datenqualität, Datenschutz und Zugriff gewährleistet.“

Nach Angaben von Ashwin Sinha von der Macquarie Bank hat der erleichterte Zugriff von Endanwendern auf Daten und Analysen einen direkten Beitrag zur Verbesserung der Datenkultur im Unternehmen geleistet. „Dies gilt nicht nur für Analysten“, sagt er. „Führungskräfte im Unternehmen, die analytisch denken, nutzen Datenvisualisierungstools mit einfachen Oberflächen, um Daten zu analysieren und Erkenntnisse zu gewinnen.“ Visualisierungstools sind nach Sinhas Meinung wichtig, aber Gleiches gilt auch für die Nutzung der Cloud-Infrastruktur durch das Team zur Steigerung von Zuverlässigkeit und Leistung. „Unsere Daten-Workloads – ob regulativ, ML-bezogen oder analytisch – wurden in den letzten drei Jahren samt und sonders auf eine Cloud-Datenplattform verlagert. Die Datenplattform nutzt eine Vielzahl von Cloud-Services, Open-Source-Paketen und Datenmanagement-Tools, um sicherzustellen, dass die Daten im gesamten Unternehmen standardisiert, integriert und in zweckmäßiger und geregelter Weise bereitgestellt werden – einschließlich vollständiger Datenherkunft und Datenrückverfolgbarkeit.“

Die robuste Data Governance, der unkomplizierte Zugriff und die Vereinfachung tragen nach Ansicht von Sinha insgesamt dazu bei, das Vertrauen der Nutzer in Daten und Analysen zu stärken. Ohne Vertrauen auf eine zuverlässige Datenplattform kann sich keine Datenkultur entwickeln.



„Man muss wissen, welche Daten am wichtigsten sind, Prioritäten setzen, die Disziplin entwickeln, diese zu schützen und zu verwalten, und die Daten schließlich demokratisieren, damit Ihre Datenexperten und Endanwender die Erkenntnisse gewinnen, die sie für ihre Innovationen benötigen.“

– Sol Rashidi, Chief Analytics Officer, The Estée Lauder Companies

05 Zukunftsvisionen

Die Hälfte der befragten Führungskräfte (und sogar etwa zwei Drittel derjenigen in Technologie- und Fertigungsunternehmen) geben an, dass sie derzeit eine neue Datenplattform evaluieren oder implementieren, um ihre aktuellen Datenprobleme zu lösen. Weitere 9 % würden dies tun, stoßen aber auf Hindernisse bei der Modifikation ihrer Architektur. Auch hier sind die Unterschiede zwischen Vorreitern und Nachzüglern extrem ausgeprägt. Die meisten Vorreiter (56 %) sind mit ihrer derzeitigen Architektur zufrieden, obwohl fast ein Drittel von ihnen gegenwärtig neue Plattformen prüft oder Anbieter evaluiert. Dagegen zeigen sich nur 24 % der Nachzügler mit ihrer bestehenden Architektur zufrieden, während 59 % sie besser früher als später ändern wollen. Die Wahrscheinlichkeit, dass Technologieunternehmen und Hersteller ein Upgrade ihrer Plattformen in Angriff nehmen, ist weitaus größer als in anderen Sektoren.

Wunschliste eines CDO für eine neue Architektur

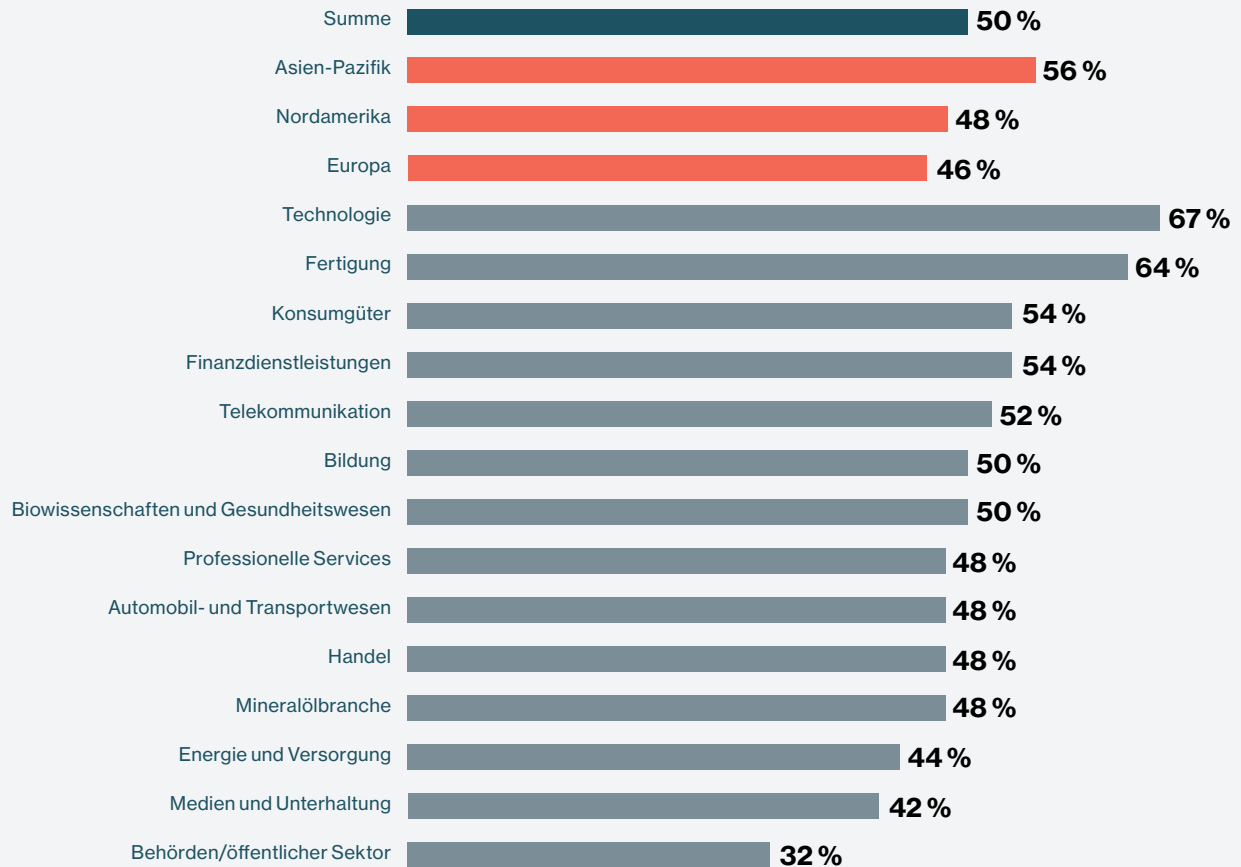
Nehmen wir einmal an, Data Leaders und Führungskräfte im Technologiebereich dürften selbst eine neue Datenarchitektur für ihr Unternehmen aufbauen. Wir haben nachgefragt, welches für sie die wichtigsten Vorteile gegenüber dem bestehenden System wären. Ganz oben auf der Wunschliste stehen bei 50 % der

Befragten (und bei mindestens 60 % derjenigen aus den Bereichen Technologie, Versorgungsunternehmen und Behörden) Open-Source-Standards und Open-Data-Formate. Für Andy McQuarrie von Hivery ist dies keine Überraschung: „Offene Standards ermöglichen es Ihnen, die Entwicklung zur Reife ganz einfach zu durchlaufen. Kurzfristig können Sie Dienste über einen Managed Service nutzen. Wenn der Reifegrad Ihres Unternehmens zunimmt und Sie feststellen, dass Sie eine Komponente hinzufügen müssen, haben Sie mit Open Source die Möglichkeit dazu, ohne dass dies die gesamte Architektur beeinträchtigen würde. Anders als früher ist es heute nicht mehr notwendig, gleich die gesamte Technik auszutauschen.“

Als nächste Punkte stehen keine neuen Funktionen auf der Wunschliste, sondern eher Funktionsverbesserungen in Bereichen, in denen Technologieführer ständig nach Optimierungspotenzial suchen: mehr Sicherheit, bessere Governance und ein besseres Preis-Leistungs-Verhältnis für Infrastruktur, Betrieb, Wartung und weitere Architekturelemente. Die Befragten bestehen auch darauf, dass jede neue Architektur alle Analyseanwendungen unabhängig davon unterstützt, ob sie auf ML, Data Science oder Business Intelligence basiert.

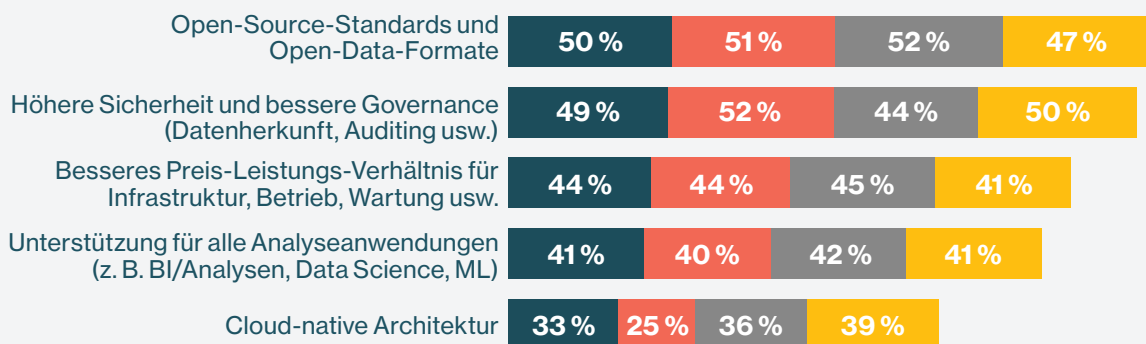
Open-Source-Standards und Open-Data-Formate stehen ganz oben auf der Wunschliste der Technologieführer, auf den Rängen folgen mehr Sicherheit und Governance sowie ein besseres Preis-Leistungs-Verhältnis für Architekturelemente.

Abbildung 8: Befragte, die derzeit eine neue Datenplattform evaluieren oder implementieren, um ihre unternehmensweite Datenstrategie oder Lösungen zur Bewältigung ihrer aktuellen Herausforderungen zu unterstützen
(% der Befragten)



Studie von MIT Technology Review Insights, 2021

Abbildung 9: Die wichtigsten Vorteile einer idealen neuen gegenüber der bestehenden Architektur
(meistgenannte Antworten; % der Befragten)



Studie von MIT Technology Review Insights, 2021

■ Summe ■ Nordamerika ■ Europa ■ Asien-Pazifik

„Die native Integration in Kundensysteme ist etwas, worauf wir wirklich fokussiert sind. Wie machen wir es unseren Kunden so einfach wie möglich, ihre Daten an uns zu übermitteln?“

– Andy McQuarrie, Chief Technology Officer, Hivery

Die von uns befragten CDOs hegen zusätzliche Hoffnungen und Pläne für die Weiterentwicklung ihres Datenmanagements und ihrer Infrastruktur. Michel Lutz, Group Chief Data Officer beim Energiekonzern Total, will die ML-basierte Automatisierung seines Metadatenmanagements ausbauen. Dieser Schritt ist für geowissenschaftliche Daten bereits erfolgt, doch plant er für das kommende Jahr eine Ausweitung auf alle Datenbereiche der Gruppe. Auch das Potenzial der Data-Mesh-Architektur hat Lutz auf dem Schirm. „Dies ist sicherlich einer der nächsten Schritte bei unserer Datenarchitektur, denn es ermöglicht eine bessere unternehmensweite Skalierbarkeit und eine stärkere Spezialisierung auf bestimmte Bereiche“, erläutert er.

Naveen Jayaraman von L'Oréal erkennt in ähnlicher Weise die Vorteile domänenorientierter Datenprodukte für Unternehmen wie das seine, die ihre Datenverwaltungsfunktionen umfassend verteilen möchten. Seiner Meinung nach haben Architekturansätze, die ein Daten-Mesh verwenden, für seine Zwecke Potenzial, sofern einige Elemente, wie z. B. die polyglotte Speicherung, zentralisiert werden können.

Mainak Mazumdar von Nielsen freut sich auf die Einführung von „KI für die KI“, also die automatische Auswahl von KI-Modellen. „Wir hoffen, eine weitere Intelligenzebene erstellen zu können, die die besten Entscheidungen über die zu verwendenden Modelle trifft.“

Längerfristig denkt Ashwin Sinha über Möglichkeiten einer „agilen Echtzeitarchitektur“ nach, die ausreichend flexibel für eine Anpassung an künftige Innovationen im Data Engineering ist. Seiner Meinung nach wäre dies eine Möglichkeit, eine erst vor kurzem entworfene Architektur zumindest mittelfristig zukunftssicher zu machen. „Wir würden gerne einen Architekturansatz verfolgen, der so flexibel ist, dass er nicht schon nach zwei oder drei Jahren veraltet ist.“

06 Fazit

Wie ein roter Faden zog sich die Erkenntnis durch unsere Untersuchungen und insbesondere unsere Gespräche mit CDOs, dass es eine direkte Linie von der Dateninfrastruktur über Data Science bis hin zur Haltung der Mitarbeiter auf allen Ebenen des Unternehmens gegenüber Daten gibt und all dies mit den geschäftlichen Auswirkungen verknüpft ist. Diese Linie ist die Datenkultur. Die vorliegende Studie bietet einige Erkenntnisse für CDOs, die ihre Ressourcen zur Förderung einer Datenkultur in ihrem Unternehmen einsetzen wollen.

- **Halten Sie es einfach und flexibel.** Viele CDOs wollen übermäßig komplexe Architekturen vereinfachen, aber genauso wichtig ist es, dass Analyseoberflächen für Endanwender benutzerfreundlich sind und Spaß machen. Die Architektur muss außerdem so flexibel bleiben, dass auch künftige geschäftliche Anforderungen erfüllt werden können, ohne dass zum Hinzufügen neuer Technologien umfassende Migrationen oder Umgestaltungen erforderlich wären. Je mehr Anwender in der Lage sind, mit ihren bevorzugten Tools ständig aktualisierte Daten abzurufen, mit ihnen zu spielen und ihnen ggf. aussagekräftige Erkenntnisse zu entlocken, desto größer wird das Vertrauen in diese Technologien.
- **Setzen Sie auf gute Governance.** Nichts wird das Mitarbeitererlebnis bei Analysen stärker beeinträchtigen

als das mehrfache Entdecken von Fehlern in den Daten. CDOs müssen nicht daran erinnert werden, wie wichtig eine solide Data Governance ist, aber vielleicht sind sich nicht alle im Klaren darüber, wie weit die Auswirkungen von Governance-Fehlern im Unternehmen reichen können. Aufbau und Pflege einer „einzigen Version der Wahrheit“ sind für alle von Bedeutung.

- **Erklären und die Werbetrommel rühren.** Einige CDOs beauftragen Teile ihrer Teams damit, Spartenabteilungen und andere Mitarbeiter über Data Science aufzuklären und sie in der Verwendung von Analysen und weiteren Tools zu schulen. Viele sind auch der Meinung, dass eine breit angelegte Werbekampagne für die positiven Auswirkungen der Nutzung von Daten und Analysen und den damit zu erzielenden Mehrwert den Wunsch der Mitarbeiter danach wecken kann. Außerdem trägt der stete Hinweis auf die durch eine starke Governance bedingte Richtigkeit von Daten ebenfalls zur Vertrauensbildung bei.
- **Knüpfen Sie Kontakte zu anderen CxOs.** Für CDOs und CAOs mag es selbstverständlich sein, die Beziehungen zu den anderen CxOs zu vertiefen, aber ein mangelnder Austausch über das Datenmanagement etwa mit CIOs und CTOs könnte Misstrauen schüren. Die Datenstrategie muss die bestehende Technologieinfrastruktur einbeziehen und sicherstellen, dass sie stets auf die Erfüllung der Unternehmensprioritäten ausgerichtet ist.



Über MIT Technology Review Insights

MIT Technology Review Insights ist die unabhängige Verlagsabteilung des *MIT Technology Review*, des weltweit ältesten Technologiemaßgebens, und wird somit von der global führenden Technologieinstitution unterstützt. Wir veranstalten Live-Events und leiten Untersuchungen zu aktuellen zentralen Herausforderungen in den Bereichen Technologie und Business. Insights führt qualitative und quantitative Untersuchungen und Analysen in den USA und anderen Ländern durch und publiziert eine Vielzahl von Inhalten, wie Artikel, Berichte, Infografiken, Videos und Podcasts. Dank des immer umfangreicher werdenden globalen Panels von MIT Technology Review ist Insights in der unvergleichlichen Lage, immer mehr Führungskräfte, Innovatoren und Vordenker auf der ganzen Welt im Rahmen von Umfragen und ausführlichen Interviews befragen zu können.

Über den Sponsor

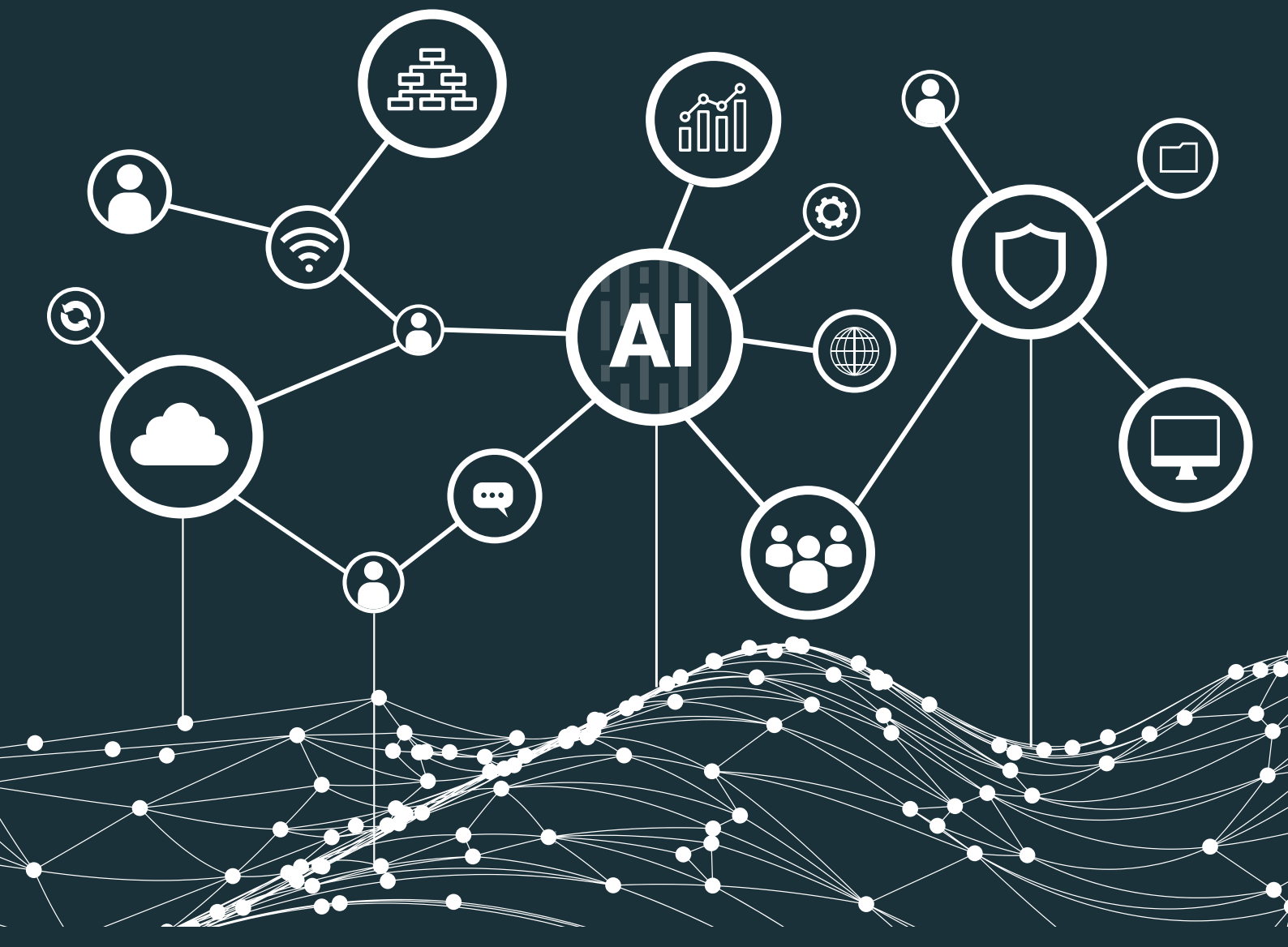
Databricks ist das Unternehmen für Daten + KI. Auf eine moderne Lakehouse-Architektur aufsetzend, kombiniert Databricks das Beste aus Data Warehouses und Data Lakes und stellt eine einfache, offene und kollaborative Plattform für sämtliche Daten-Workloads Verfügung. Mehr als 5.000 Unternehmen weltweit – darunter Shell, Comcast, CVS Health, HSBC, T-Mobile und Regeneron – nehmen die Leistungen von Databricks in Anspruch, um Data Engineering in riesigen Ausmaßen, aussagekräftige Data Science, den gesamten Lebenszyklus des Machine Learning und Business Analytics zu implementieren. Mit einer globalen Präsenz und Hunderten von Partnern, darunter Microsoft, Amazon, Google und Tableau, hat Databricks es sich zur Aufgabe gemacht, Datenteams in aller Welt bei der Lösung schwierigster Probleme zu unterstützen.



Abbildungen

Davooda: 5, 9, 13, 14; fad82:13; Friday Studio: 2, 3, 6, 15, 17, 22; Nadiinko: 18; ninamalin:18, papipo:15; Prostock Studio: Titel- und Rückseite, 2, 3, 4, 5, 13,19; Rashes Ashor: 19; Starline: Titel- und Rückseite, 2. Alle Abbildungen stammen von Shutterstock und wurden von Scott Shultz Design zusammengestellt.


Obwohl alle Anstrengungen unternommen wurden, um die Richtigkeit der vorliegenden Informationen zu überprüfen, übernimmt MIT Technology Review Insights keinerlei Verantwortung oder Haftung für den Fall, dass die Richtigkeit des vorliegenden Berichts oder der darin enthaltenen Informationen, Meinungen oder Schlussfolgerungen durch Dritte angenommen wird.



MIT Technology Review Insights

 www.technologyreview.com

 @techreview @mit_insights

 insights@technologyreview.com