



**Hewlett Packard
Enterprise**

Verbesserte Nutzung des KI-Potenzials durch mehr Ordnung in Ihrer Datenumgebung

Für Ihre KI-Strategie muss zunächst eine starke Datenstrategie her. Hier sind fünf Möglichkeiten, um sicherzustellen, dass Ihre Daten für die Verarbeitung durch KI vorbereitet sind.



Unternehmen überschlagen sich regelrecht, um in KI zu investieren. Laut IDC werden sich die weltweiten Ausgaben für KI-Anwendungen, -Infrastrukturen und -Services in den nächsten vier Jahren voraussichtlich verdoppeln und bis 2028 einen Wert von 632 Milliarden US-Dollar erreichen.¹ Leider implementieren viele dieser Unternehmen KI-Lösungen vorschnell, bevor sie die notwendigen Grundlagen dafür geschaffen haben. Ein entscheidender erster Schritt wird dabei oft übersehen, nämlich sicherzustellen, dass die zur Ausführung von KI-Workloads erforderlichen Daten den Anforderungen entsprechen.

„Ist die Datenqualität schlecht, bekommt man ein schlechtes Ergebnis – ganz egal, wie gut das verwendete Modell für maschinelles Lernen ist“, sagt Chad Smykay, AI CTO für vertikale Branchen und angesehener Technologie bei Hewlett Packard Enterprise. Laut Smykay ist das Geheimnis hinter einer guten Datenqualität überhaupt keines: Ihre Daten müssen aufbereitet, bereinigt und gut geordnet werden, damit Data Scientists und Entwickler im gesamten Unternehmen sie einsetzen können. Doch die Produktion großer Volumen an bereinigten Daten, um die Leistungsanforderungen der heutigen datenintensiven KI-Workloads zu erfüllen, ist zeit- und arbeitsaufwendig. Außerdem stellt sie eine enorme Herausforderung in der gegenwärtigen Situation dar, in der es an Datenexpert:innen mangelt und Geschwindigkeit von entscheidender Bedeutung ist.

Wie können diese Herausforderungen überwunden und die Datenqualität verbessert werden? Smykay nennt fünf konkrete Schritte, mit denen Sie Ihre KI-Bereitstellung durch bessere Datenprozesse beschleunigen können.

¹ „Worldwide Spending on Artificial Intelligence Forecast to Reach \$632 Billion in 2028, According to a New IDC Spending Guide“, IDC, August 2024

1. Legen Sie intelligente Richtlinien zur Daten-Governance fest.

Jedes Unternehmen benötigt eine solide Governance-Grundlage für Daten, die gewährleistet, dass Daten bei ihrer Erfassung ordnungsgemäß definiert und geordnet werden, bevor sie einem Algorithmus zugeführt oder zum Trainieren eines Large Language Model (LLM) herangezogen werden. Diese Richtlinien verlangen, dass Unternehmen sich einige wichtige Fragen stellen, etwa um welchen Datentyp es sich handelt, woher sie stammen, ob sie auf den jeweiligen geschäftlichen Anwendungsfall anwendbar sind, wie alt sie sind und wo sie gespeichert werden sollen.

Es ist nicht nur eine Frage einer guten Betriebspraxis. Mit einer effektiven Daten-Governance stellen Sie sicher, dass die Daten hinter den KI-Modellen Ihres Unternehmens nicht nur ordnungsgemäß definiert werden, sondern auch sicher und vertrauenswürdig sind. Gleichzeitig bereiten Sie Ihr Unternehmen damit auf künftige Datenschutzgesetze und sonstige rechtliche Bestimmungen vor, die höchstwahrscheinlich in den nächsten fünf Jahren aufkommen werden. Diese Regeln, wie die Datenschutz-Grundverordnung der Europäischen Union (DSGVO), können Unternehmen dazu anhalten, nachzuweisen, woher die Daten stammen und auf welche Art sie zum Trainieren von Algorithmen verwendet wurden.

Zwar versprechen neue Softwareprodukte, die Daten-Governance zu automatisieren, doch Smykay zufolge sollte die Daten-Governance zumindest am Anfang ein von Menschen gesteuerter, an Technologie ausgerichteter und von einem Chief Technology Officer, Chief Data Officer oder Chief Data und AI Officer geleiteter Prozess sein. Der Grund? Software kann höchstwahrscheinlich nicht unterscheiden, welche Daten tatsächlich nützlich sein werden, um die geschäftlichen Anforderungen und Ziele eines Unternehmens rund um KI zu erreichen.

„Um aus KI Dateneinblicke zu gewinnen, die Ihnen dabei helfen sollen, datenorientierte Entscheidungen zu treffen, ist es entscheidend, dass die Daten, die Sie klassifizieren, bereinigen und speichern, für die KI-Anwendungsfälle Ihres Unternehmens relevant sind“, führt er aus.

„Personen, die verstehen, wie das Unternehmen Nutzen aus der KI ziehen wird, müssen infolgedessen auch diejenigen sein, die die Regeln bestimmen.“

2. Vereinfachen Sie Ihr Datamanagement mit einem globalen Namespace.

Bei der Optimierung Ihrer Daten für erfolgreiche KI-Bereitstellungen ist es wichtig, sich um eine Vereinfachung des Managements Ihres globalen Datenbestands zu kümmern. Um die angestrebte Einfachheit zu erreichen, empfiehlt Smykay, dass Unternehmen eine Data Fabric mit einem globalen Namespace für das Management sämtlicher Unternehmensdaten aufbauen. Stellen Sie sich dies als einzelnen Snapshot vor, der alle Daten-Assets aus der verteilten hybriden Computing-Umgebung umfasst und mit einem einheitlichen und konsistenten Benennungssystem für die verschiedenen Datensätze, die KI-Entwickler eventuell nutzen möchten, klar organisiert ist.

„Ein globaler Namespace ist eine Lösung, die Entwicklern die einfachste Möglichkeit bietet, Daten zu finden, die an mehreren Standorten gespeichert sind – sei es in mehreren Public Clouds, in verschiedenen Rechenzentren oder am Edge“, sagt Smykay. „Sie können die Daten, die sie benötigen, an jedem Speicherort schnell einsehen und abrufen.“

Ein weiteres Problem, das durch den Einsatz eines globalen Namespace gelöst werden kann, ist die sogenannte „Datenverschwendung“. „Auf einen Großteil der Unternehmensdaten wird nie zugegriffen“, sagt Smykay. „Mit einem globalen Namespace können Sie diese Daten abrufen und dann insbesondere für KI-Workloads sinnvoll nutzen.“



3. Schaffen Sie eine flexible, offene und portierbare Datenplattform.

In der komplexen Unternehmensumgebung von heute gibt es eine Vielfalt an Daten, die sich darauf verteilen, woher sie stammen, wo sie gespeichert werden und sogar in welcher Form sie vorliegen. Strukturiert oder unstrukturiert, intern generiert oder aus dem Internet bezogen – die Daten von heute werden in unzähligen Formaten gespeichert und müssen über mehrere Anwendungsprogrammierschnittstellen (APIs) abgerufen werden, die unter Umständen nicht miteinander kommunizieren.

„Weit verstreute und unzusammenhängende Daten können Probleme verursachen, daher ist eine Konsolidierung die beste Lösung“, so Smykay. Die Bereitstellung all dieser verschiedenen und verteilten Daten-Assets eines Unternehmens auf einer einheitlichen Data Fabric ist ein kritischer Schritt, um Entwicklern hochwertige Daten zur Verfügung stellen zu können. Eine einzige Datenzugriffsplattform mit offener API, die mehrere Datentypen unterstützen kann, liefert Entwicklern die entscheidende Zutat für ihren Erfolg: Flexibilität. „Im Bereich von Analysen und KI benötigen Entwickler eine Datenplattform, die mehrere APIs nutzen kann. Denn dadurch erhalten sie die flexibelste Möglichkeit, um die benötigten Daten abzurufen und aufzubereiten, egal, in welchem Format sie vorliegen“, sagt Smykay.

4. Erhöhen Sie die Skalierbarkeit Ihres Datenbestands.

Daten sind der Kraftstoff, der das Training und die Ausgaben von KI-Modellen antreibt. Wer die leistungsbezogenen Herausforderungen datenintensiver Workloads bewältigen will, etwa die riesigen Datenvolumen, die zum Trainieren der LLMs der generativen KI benötigt werden, benötigt eine enorme Speicherkapazität. „Um die Erfassung, Aufbereitung und Bereitstellung von Daten für KI zu beschleunigen, benötigen Unternehmen eine skalierbare Datamanagement-Lösung, welche die gesamte Hybrid Cloud umfasst“, so Smykay.

Für die Exabytes an Daten, die für die KI-Modelle von heute benötigt werden, ist Hybrid eine der kostengünstigsten und am besten skalierbaren Datamanagement-Strategien, erläutert er. „Traditionelle,



große Hyperscaler-Lösungen in der Public Cloud sind insbesondere für diese Arten von KI-Workloads einfach zu teuer geworden“, führt Smykay aus. „Die Hybrid Cloud kann Skalierbarkeitsanforderungen erfüllen, indem sie die Assets überall dort verteilt, wo Daten generiert oder eingespeist werden, von einer Private Cloud vor Ort bis zum Edge.“

Die Handhabung dieser Daten entwickelt sich zu einer enormen Herausforderung, da das Volumen der generierten Daten in einem wahnwitzigen Tempo zunimmt. Dem IDC zufolge wird erwartet, dass sich die Menge der jährlich erzeugten Daten bis zum Jahr 2026 gegenüber dem Stand von 2022 mehr als verdoppeln wird, wobei Unternehmensdaten doppelt so schnell zunehmen wie Verbraucherdaten.²

Smykay ergänzt, dass die Datenebenenstrategie der Hybrid Cloud darüber hinaus flexibel und zukunftsweisend ist und sich daher bestens eignet, um innovative KI-Modelle zu unterstützen, die im weiteren Verlauf entstehen werden.

² „Global DataSphere, Data Marketplaces, and Data as a Service“, IDC, August 2023

5. Beschleunigen Sie Datenpipelines, um KI-Modelle schneller zu versorgen.

KI-Workloads erfordern nicht nur viele, sondern auch genaue Daten, damit sie maximale Vorteile liefern können. „Die Nützlichkeit der meisten Daten ist jedoch zeitlich begrenzt“, sagt Smykay. Daher müssen KI-Modelle diese schnell verarbeiten, solange sie noch aktuell sind.

Obwohl Eile angezeigt ist, geben 85 % der Fortune 500-Unternehmen an, dass sie ihre Daten nicht effektiv nutzen können.³ Streaming-Lösungen, die gut in eine einheitliche Datenplattform integriert sind, können diesen Prozess allerdings beschleunigen. „Sie haben Ihre KI-Modelle erstellt und trainiert. Jetzt lautet die Herausforderung, sie weiterhin zu pflegen“, sagt Smykay. „Die gute Nachricht ist, dass Sie den Datenfluss in Echtzeit beibehalten können, wenn Sie die Aufnahme der Daten am Backend mit einem Streaming-Format abändern.“

Smykay merkt an, dass eine schnellere Datenaufnahme äußerst wertvoll sein kann, um die Datenqualität zu gewährleisten, weil Entwickler damit eine kürzere Vorlaufzeit für die Datenbereinigung und -aufbereitung erhalten. Es gibt zwar immer wieder neue automatisierte Tools, die die Aufbereitung von Daten für KI-Modelle unterstützen, doch Entwickler und Data Scientists wissen, dass die Qualitätskontrolle von Daten zu wichtig ist, um sie einem Algorithmus zu überlassen. Mit den richtigen Datenlösungen können Unternehmen sich bestens dafür rüsten, hochwertige Daten für eine schnelle, flexible und sichere KI-Bereitstellung zu optimieren.



Das Datenwachstum wird nicht langsamer. Tatsächlich wird es im Zuge der kontinuierlichen Weiterentwicklung von KI und Geschäftsanalysen sogar noch an Geschwindigkeit zulegen. Es ist daher angebracht, sich auf diese Situation vorzubereiten. Bei der Planung und Gestaltung der Datengrundlage Ihres Unternehmens ist eine Data Fabric von entscheidender Bedeutung. Sie ermöglicht Ihnen die Aufnahme von Daten aus Ihrer gesamten Hybrid Cloud, unterstützt zahlreiche Datenformate, lässt sich gemeinsam mit Ihrem Unternehmen skalieren und hilft ihm bei der Entwicklung einer robusten Daten-Governance-Strategie, mit der Sicherheit und Vertrauenswürdigkeit gewährleistet werden.

³ „Drowning in Data for Want of Information: Is Data Minimization Really Possible?“, IDC, September 2024

Weitere Informationen unter

HPE.com/AI/insights