

Die 5 größten Datenherausforderungen und wie **Sie sie bewältigen**

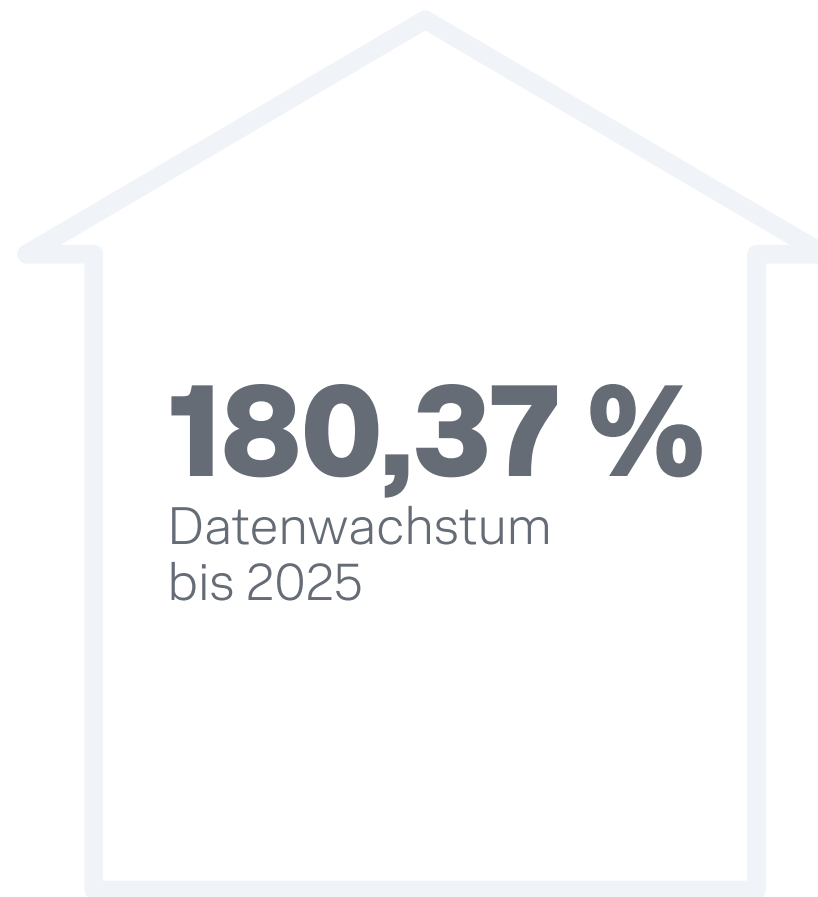


splunk>

Wir sind mitten im **Datenzeitalter**, es gibt kein Zurück mehr. IDC Research schätzt, dass die Gesamtmenge der Daten, die weltweit erstellt, erfasst, kopiert und genutzt werden, bis 2025 auf **mehr als 180 Zettabyte** ansteigen wird. 2020 waren es noch 64,2 Zettabyte.

Wenn Sie sich jetzt fragen, wie Ihr Unternehmen mit den wechselnden Trends und den Skalierungsproblemen von immer mehr Cloud-nativen Anwendungen, ausufernder Cloud-Infrastruktur und den vielen On-premises-Datenbanken am Edge zurecht kommen soll – und das im Bewusstsein, dass all dies auch weiterhin so schnell wachsen wird, dass Sie unmöglich Schritt halten können –, dann sind Sie auf das falsche Problem fixiert. Kämpfen Sie nicht gegen den Wandel an, sondern lassen Sie ihn für Sie arbeiten.

Fakt ist, dass die enormen Datenmengen und das rasende Tempo bleiben werden. Das bedeutet: Gewinnen werden diejenigen Unternehmen, die jetzt das Potenzial ihrer Daten nutzen und damit geschäftlichen Mehrwert erzielen können.



Dazu ist jedoch mehr nötig als neue Cloud-Speicher, ein Haufen Sonderlösungen für einzelne Analysebereiche und mehr als die Bereitstellung zusätzlicher Mittel. Wer eine tragfähige Datenstrategie für diese neue Welt will, braucht eine umfassende Plattform, mit der das Unternehmen intelligent und einfach bestimmen kann, welche Daten es bei Bedarf durchsuchen will, und zwar ohne Rücksicht auf Quelle und Format. Was zunächst nach komplexen Architekturen und Workflows klingt, muss keineswegs kostspielig und kompliziert sein – nicht wenn die Datenmethodik und die Datenperspektive stimmen.

Data Tiering, die strategische Datenspeicherung unter den Gesichtspunkten von Performance, Verfügbarkeit, Kapazität und Kosten, ist ein zentraler Aspekt jeder modernen Datenstrategie. Unternehmen müssen dazu ihre Datenmengen nach Priorität in Schichten sortieren, also je nachdem, wie häufig die Daten gebraucht werden, wie rasch sie zugänglich sein sollen und wie wichtig sie für den Geschäftsbetrieb sind – das alles mit Blick auf eine kosteneffiziente Speicherung.

In diesem Leitfaden erfahren Sie, worin die fünf größten Herausforderungen von Big Data liegen und warum Data Tiering der beste Weg in die Zukunft ist.



D1

Die Formatvielfalt erschwert den geschäftlichen Nutzen

Die Big-Data-Welt ist geprägt von den Komplexitäten unterschiedlicher Datenformate. In dieser Vielfalt steckt ein Kommunikationsproblem: Wenn Daten nicht mit anderen Daten sprechen oder zusammen funktionieren können, bleiben sie insofern wertlos, als sich daraus nur schwerlich geschäftliche Erkenntnisse ziehen lassen.

Zum einen haben wir das ganze Spektrum der unstrukturierten, halbstrukturierten und strukturierten Daten mit ihren jeweils eigenen Standardformaten. Im Gegensatz zu strukturierten Daten, die einem definierten Schema folgen und von daher nur bestimmte Nutzungsformen zulassen, bieten unstrukturierte Daten flexiblere Suchmöglichkeiten; ihre Verwaltung erfordert jedoch Fachkenntnisse, und man braucht spezielle Tools, die mit solchen Daten live umgehen können. Weil ca. **80 % aller Geschäftsdaten unstrukturiert** sind, müssen Unternehmen diese Aufgabe irgendwie bewältigen.

Zum anderen gibt es noch die Frage nach der Wohlgeformtheit von Daten. Die Herausforderung für Unternehmen besteht darin, jedes einzelne Format mit sämtlichen Kombinationen aus den einzelnen Anwendungen abzugleichen, die unter Umständen eigene, nicht standardkonforme Log-Formate verwenden.

Von den Unterschieden bei der Erfassung, Analyse und Speicherung von Big Data einmal abgesehen, geht es letztlich darum, aus den Daten geschäftlichen Nutzen zu ziehen und auf der Grundlage von Daten sichere Entscheidungen zu treffen, die das Unternehmen voranbringen.

Von den Unterschieden bei der Erfassung, Analyse und Speicherung einmal abgesehen, geht es letztlich darum, aus den Daten geschäftlichen Nutzen zu ziehen und auf dieser Grundlage sichere Entscheidungen zu treffen, die das Unternehmen voranbringen.

Wie also ziehen Sie aus Big Data den maximalen Mehrwert? Indem Sie zuerst ein Framework zur Datenintegration etablieren, damit Sie wissen, welche Daten im Unternehmen vorhanden sind, wo sie liegen, wie alt sie sind und wie sie zur Lösung von Problemen oder für Use Cases zur Erschließung neuer Chancen beitragen können. Das ist eine Grundvoraussetzung jeder modernen Datenstrategie. Ein solcher Rahmen wird Ihnen auch die Investitionsentscheidungen in puncto Datenverarbeitung erleichtern, weil Sie dann die Option von Plattformen und Lösungen haben, mit denen Sie selbst aus disparaten Datenformaten geschäftlichen Nutzen ziehen.

Zum Start Ihres Frameworks brauchen Sie eine Plattform, die sämtliche Schemata aufnehmen und formal normalisieren kann und damit weitere Untersuchungen beschleunigt. Außerdem soll die Plattform erweiterbar sein, sie muss Daten aus unterschiedlichsten Quellen verarbeiten können, mit den jeweils neuesten Funktionen integrierbar sein und dabei Richtlinienkonformität gewährleisten.

Die Menge unstrukturierter Daten wird jährlich um

55–65 %

zunehmen

02

Datensilos verhindern verteilten Datenzugriff

Ein gewachsenes Unternehmen, das Server vor Ort, Multi-Clouds oder hybride Infrastrukturen verwendet, hat meist diverse Datensilos, die so unterschiedlich sein können wie die Daten selbst. Datensilos entstehen im Laufe des Unternehmenswachstums wie von selbst, sie sind eine nahezu unvermeidliche Begleiterscheinung. Treiber dieser Entwicklung sind Faktoren wie alternde (Legacy-)Infrastrukturen, Fusionen und Übernahmen, konkurrierende Technologien und Tools oder die ganz normale Fehlkommunikation zwischen Teams und Abteilungen.

Die Folge davon ist laut [Lagebericht Dateninnovation](#) (State of Data Innovation Report), dass nur 40 % der Unternehmen ihre Daten umfassend aggregieren, während 29 % zu Protokoll geben, dass bei ihnen einige oder alle Geschäftsbereiche ihre Daten getrennt oder unter Verschluss halten.

Manche Tools und Silos eignen sich gut für stark strukturierte Daten, auf die in regelmäßigen Mustern zugegriffen wird, andere sind besser für unstrukturierte Daten und speziellere Zugriffsmuster geeignet, auch angesichts dessen, dass sich die Datenformate laufend entwickeln und ändern. Doch die Vielzahl der Tools und die Menge der oft isolierten Datenspeicher führen dazu, dass die Daten in wesentlichen Aspekten unsichtbar bleiben, etwa was Entstehung, Umwandlungen oder den Gesamtzustand betrifft. Ein einheitlicher Datenzugriff wird dadurch extrem schwierig.

Laut Lagebericht Dateninnovation aggregieren nur 40 % der Unternehmen ihre Daten umfassend, während 29 % sagen, dass bei ihnen einige oder alle Geschäftsbereiche ihre Daten getrennt oder unter Verschluss halten.

Verschärft wird das Problem noch in Umgebungen mit zahlreichen Einzelösungen, die eigene Silos erfordern, welche abermals einer gemeinsamen Datennutzung durch die Teams im Wege stehen. Unternehmen haben daher nach wie vor damit zu kämpfen, ihre jeweiligen regionalen Geschäftsanforderungen über Datensilos hinweg zu erfüllen – eine Aufgabe, die in Anbetracht von Sicherheits- und Compliance-Vorschriften nicht leichter wird.

Unternehmen müssen also Wege finden, diese Mauern zu durchbrechen, damit jedes einzelne Team die Datenvorteile nutzen kann, ohne doppelte und mehrfache ETL-Prozeduren (Extract, Transform, Load) und ohne Kontextwechsel zwischen Tools, Speicherorten und Quellen.

Daten sind wertvoll, wenn sie zugänglich sind und man sie nutzen kann, um Erkenntnisse zu gewinnen und Maßnahmen zu ergreifen. Also machen Sie sich vorerst keine Sorgen um Silos – lassen Sie Ihre Daten getrost dort, wo sie sind. Konzentrieren Sie Ihre Anstrengungen stattdessen darauf, aus sämtlichen Silos die Daten zu sammeln. Das geht am besten mit einer Konsolidierungsplattform, die ohne Rücksicht auf Silomauern funktioniert.

Nur 29 %
der Unternehmen haben
ein Klassifizierungssystem,
das die meisten/alle Daten
bestimmt

03

Geringe Datenqualität mindert den Wert

Von Haus aus sind Daten normalerweise unsauber und verrauscht. Sie sind redundant und roh, unvollständig und fehlerhaft, der Kontext fehlt oder ist unklar, sie enthalten ungültige Werte oder unzulässige PII-Daten, mit denen einzelne Personen identifiziert werden können – all dies schmälert die Qualität der Analysen, wenn man sich nicht darum kümmert. Insofern ist kaum erstaunlich, dass einer [Trifacta-Studie](#) zufolge 60 % der IT-Fachleute mindestens die Hälfte ihrer Arbeitszeit damit verbringen, Daten zu bereinigen, aufzubereiten oder anderweitig deren Qualität zu sichern.

Bei Daten hoher Qualität stehen die Aufbereitung und die Formgebung durch Umwandlungen, Aggregationen und Anreicherungen im Vordergrund. Dann können Sie diese Daten besser befragen, aussagekräftigen Analysen unterziehen und daraus Erkenntnisse gewinnen. Daten, die unnötig aufgebläht oder gar nicht lesbar sind, können Sie nicht gebrauchen. Die Bereinigung und Formatierung der Daten, unabhängig von der Quelle oder dem Speicherort, ist ein entscheidender Schritt im Vorfeld, mit dem Sie die Kosten senken und dafür sorgen, dass die Daten für alle User effizient nutzbar werden.

Insgesamt benötigen Unternehmen eine zentrale Möglichkeit, Daten zu und von den jeweiligen Datenspeichern zu leiten, und zwar so, dass die Daten konsistent bleiben, Fehlzuordnungen vermieden werden und die Gemeinkosten des Transports minimal bleiben.

Die Datenaufbereitung kostet Unternehmen weltweit mehr als

\$ 450 Mrd.

04

Die Datenrelevanz hängt von Alter und Nutzen ab

Daten sind nicht alle gleich. Zwar gelten alle Daten zunächst einmal als erfolgskritisch und werden dementsprechend dem teuersten Speicher und der stärksten Analytik zugewiesen. Aber Ihr Unternehmen wirft Geld aus dem Fenster und vergibt wertvolle Erkenntnisse, wenn es auf Dauer alle Daten so behandelt.

Wenn Sie Ihre Speicherpraxis überdenken und auf Relevanz ausrichten, bieten sich zwei Kriterien an: das Alter und der maximale Nutzen der Daten während ihrer Lebensdauer. Denn anders als ein guter Wein werden Daten mit der Zeit nicht besser. Vielmehr verlieren sie schon kurz nach der Entstehung an Wert. Außerdem sinkt der Wert nicht bei allen Daten in gleichem Maße.

Je schneller Sie Informationen erfassen, analysieren und nutzen können, desto größer ist der geschäftliche Wert. [Forrester](#) schätzt jedoch, dass 60 bis 73 % aller Daten, die in der Wirtschaft entstehen, analysetechnisch ungenutzt bleiben. Der rasche Wertverfall bedeutet aber nicht, dass Sie noch mehr Daten sammeln müssen. Sondern dass Sie Ihre Daten nach Nutzwert gewichten sollten, je nachdem, ob die Daten bei Untersuchungen oder Analysen eine Hauptrolle spielen oder eher eine Nebenrolle bei Compliance, Forensik etc. übernehmen.

Ein Unternehmen, das Alter und Nutzen seiner Daten berücksichtigt, findet ein gut ausgewogenes Verhältnis von Kosten, Leistung und Funktionsumfang, mit dem sich alle Use Cases ohne Kostenrisiko bedienen lassen.

Forrester schätzt, dass 60 bis 73 % aller Daten, die in der Wirtschaft entstehen, analysetechnisch ungenutzt bleiben.

05

Starre Kostenstrukturen belasten die Budgets

Zu den größten Posten unter den Ausgaben für Big Data zählen normalerweise die Ausgaben für die entsprechenden analytischen Datenbanken. Diese Datenbanken skalieren in der Regel mit den steigenden Kosten für Datenverarbeitung und -speicherung. Wenn Unternehmen dann noch versuchen, mit Einzellösungen die Lücken in der Datenbankfunktionalität zu flicken, dann blähen sich die Ausgaben weiter auf, während der Produktivität die Luft ausgeht.

Gartner geht davon aus, dass die Ausgaben für Software 2022 um 9,6 % auf 806,8 Milliarden US-Dollar steigen und der Markt für IT-Services um 6,2 % auf 1,3 Billionen US-Dollar wächst, wobei in den Unternehmen der größte Teil des Budgets für Security und Analytics verwendet wird. Das Datenvolumen dürfte unterdessen um ganze 180 % zulegen. Bei vielen Preismodellen von Punkt-lösungen bleibt dieses Ungleichgewicht jedoch unberücksichtigt.

Die Anbieter von Cloud-Services haben mit elastischen Tarifen reagiert, mit austauschbaren Einheiten über mehrere Dienste hinweg. Viele Tools haben aber weiterhin eine Preisstruktur, bei der Kunden nur schwer oder gar nicht von Skaleneffekten profitieren können.

Eine Datenplattform, die ihren Kunden die Vorteile elastischer Tarife bietet, ermöglicht es ihnen auch, Datenstrategien zu entwickeln, die nach Rechenleistung und Arbeitslast flexibel sind.

Ein Ungleichgewicht:

180 % Datenwachstum,
9,6 % mehr Budget

Das Datenzeitalter ist Wirklichkeit geworden, und es bestätigt, dass Daten in der Tat das wertvollste und strategisch wichtigste Gut für Unternehmen sind. Es lehrt uns auch, dass Daten wie ein lebender Organismus eine Entwicklung durchmachen und dabei unterschiedliche Aufmerksamkeit erfordern.

Sie können nun entweder Zeit und Ressourcen darauf verwenden, Ihre Daten zu bereinigen, umzuwandeln und nutzbar zu machen. Oder Sie investieren in eine Datenplattform, die Daten bereits bei der Aufnahme so formen kann, dass Transaktionen, Trends und Muster automatisch zu finden sind. Eine kohärente Datenarchitektur ist das Resultat einer Datenstrategie, die vielschichtig und am Nutzwert der Daten ausgerichtet ist und auf jeder durchsuchbaren Ebene von leistungsstarken Analysen gestützt wird.

Wir bei Splunk sind davon überzeugt, dass Data Tiering der beste Weg ist, die Komplexität von Big Data so in den Griff zu bekommen, dass Sie aus Ihren Daten maximalen geschäftlichen Mehrwert ziehen.

Erfahren Sie mehr über den Stand der Splunk-Forschung zu Data Tiering und lernen Sie ein Framework kennen, bei dem Sie mit eigenen Augen sehen, was Ihre Daten Ihnen sagen wollen.

[Mehr erfahren](#)

Splunk, Splunk> und Turn Data Into Doing sind Marken und eingetragene Marken von Splunk Inc. in den Vereinigten Staaten und anderen Ländern. Alle anderen Markennamen, Produktnamen oder Marken gehören den entsprechenden Inhabern.
© 2022 Splunk Inc. Alle Rechte vorbehalten.

22-25319-Splunk-Top 5 Challenges in Data and What You Need to Do About Them-LS-106_GER



splunk>
turn data into doing®