

# Die Data Fabric- Architektur bietet 3 unmittelbare Vorteile

Integrieren und aktivieren Sie alle  
Daten Ihres Unternehmens für  
bessere Geschäftsergebnisse

## **Inhalt**

- 3** Die Herausforderungen von Daten heute
- 5** Geschäftsvorteile einer Data Fabric
- 6** Wie IBM eine Data Fabric liefert
- 7** Data Fabric in Aktion
- 7** Data Fabric ist der transformative nächste Schritt für Ihr Unternehmen

# Die Herausforderungen von Daten heute

Daten sind ein integraler Bestandteil der digitalen Transformation für Unternehmen.

Doch wenn Unternehmen versuchen, ihre Daten zu nutzen, stoßen sie auf Herausforderungen, die sich aus unterschiedlichen Datenquellen, -typen, -strukturen, -umgebungen und -plattformen ergeben. Dieses multidimensionale Datenproblem wird noch komplizierter, wenn Unternehmen hybride und Multi-Cloud-Architekturen nutzen. In vielen Unternehmen sind die operativen Daten bisher weitgehend isoliert und verborgen geblieben, was zu einer enormen Menge an ungenutzten Daten geführt hat.

Nehmen Sie das Beispiel eines nordamerikanischen Energieunternehmens, das sich als datengesteuertes Unternehmen neu erfinden wollte - ein Unternehmen, in dem Data Science-Fähigkeiten über mehrere Geschäftsbereiche hinweg leicht zugänglich waren. Sie erkannten bald, dass ihre digitale Transformation durch Silodaten, inkonsistente Tools und unterschiedliche Qualifikationsniveaus behindert wurde, was alles zu kritischen Lücken zwischen den Datenkompetenzen führte.

Das Problem, mit dem sie sich konfrontiert sahen, war nicht einzigartig für ihr Unternehmen; tatsächlich ist es eine häufige Folge von Datenlandschaften, die aus ihren Datenmanagement-Architekturen herausgewachsen sind.

Was den Energiekonzern letztlich wieder auf einen erfolgreichen Weg der digitalen Transformation brachte, war der Einsatz eines neuen Datenarchitekturkonzepts, der sogenannten Data Fabric.

Was genau ist eine Data Fabric, wie unterscheidet sie sich von vorherigen Architekturen, was kann sie für Unternehmen leisten und welche Rolle spielt IBM bei der Implementierung? Dieses Whitepaper beantwortet diese Fragen.

## Was ist eine Data Fabric?

In der Vergangenheit haben Unternehmen versucht, Probleme beim Datenzugriff entweder durch Punkt-zu-Punkt-Integration oder durch die Einführung von Daten-Hubs zu lösen. Beides ist nicht geeignet, wenn die Daten stark verteilt und siloartig angeordnet sind. Punkt-zu-Punkt-Integrationen verursachen exponentielle Kosten für jeden zusätzlichen Endpunkt, der angeschlossen werden muss, was bedeutet, dass dies ein nicht skalierbarer Ansatz ist. Daten-Hubs ermöglichen eine einfachere Integration von Anwendungen und Quellen, erhöhen aber die Kosten und die Komplexität, um die Qualität und das Vertrauen in die Daten innerhalb des Hubs zu erhalten.

Die Data Fabric ist eine aufstrebende Architektur, die darauf abzielt, die Datenherausforderungen zu bewältigen, die sich aus einer hybriden Datenlandschaft ergeben. Der Grundgedanke von ist es, ein Gleichgewicht zwischen Dezentralisierung und Globalisierung durch die Funktion als virtuelles Bindegewebe Bindegewebe zwischen den Datenendpunkten herzustellen (siehe Seite 4).

Durch Technologien wie die Automatisierung und Erweiterung der Integration, föderierte Governance sowie die Aktivierung von Metadaten ermöglicht eine Data Fabric-Architektur eine dynamische und intelligente Datenorchestrierung über eine verteilte Landschaft hinweg und schafft so ein Netzwerk von sofort verfügbaren Informationen, die ein Unternehmen antreiben.

Eine Data Fabric ist unabhängig von Einsatzplattformen, Datenprozessen, geografischen Standorten und Architekturansätzen. Es erleichtert die Nutzung von Daten als Unternehmensvermögen. Eine Data Fabric stellt sicher, dass Ihre verschiedenen Arten von Daten erfolgreich kombiniert, abgerufen und verwaltet werden können - effizient und effektiv.

## Fähigkeiten und Prinzipien einer Data Fabric

Das Herzstück der Data Fabric-Architektur ist eine Datenverwaltungsplattform, die die gesamte Bandbreite integrierter Datenverwaltungsfunktionen einschließlich Discovery, Governance, Zusammenstellung und Orchestrierung ermöglicht.

Eine Data Fabric entwickelt sich jedoch von traditionellen Datenmanagementkonzepten wie DataOps, die sich nur auf etablierte Praktiken konzentrieren, hin zu einem höheren Grad der Operationalisierung von Daten. Es basiert auf einer verteilten Architektur und fortschrittlicher Technologie, die in der Lage ist, die Anforderungen zu erfüllen, die sich aus der extremen Vielfalt und Verteilung von Datenbeständen ergeben.

Eine Data Fabric kann logisch in vier Fähigkeiten (oder Komponenten) unterteilt werden:

### Wissen, Erkenntnisse und Semantik

- Bietet einen Datenmarktplatz und ein Einkaufserlebnis
- Reichert entdeckte Datenbestände automatisch mit Wissen und Semantik an, damit Verbraucher die Daten finden und verstehen können

### Einheitliche Governance und Compliance

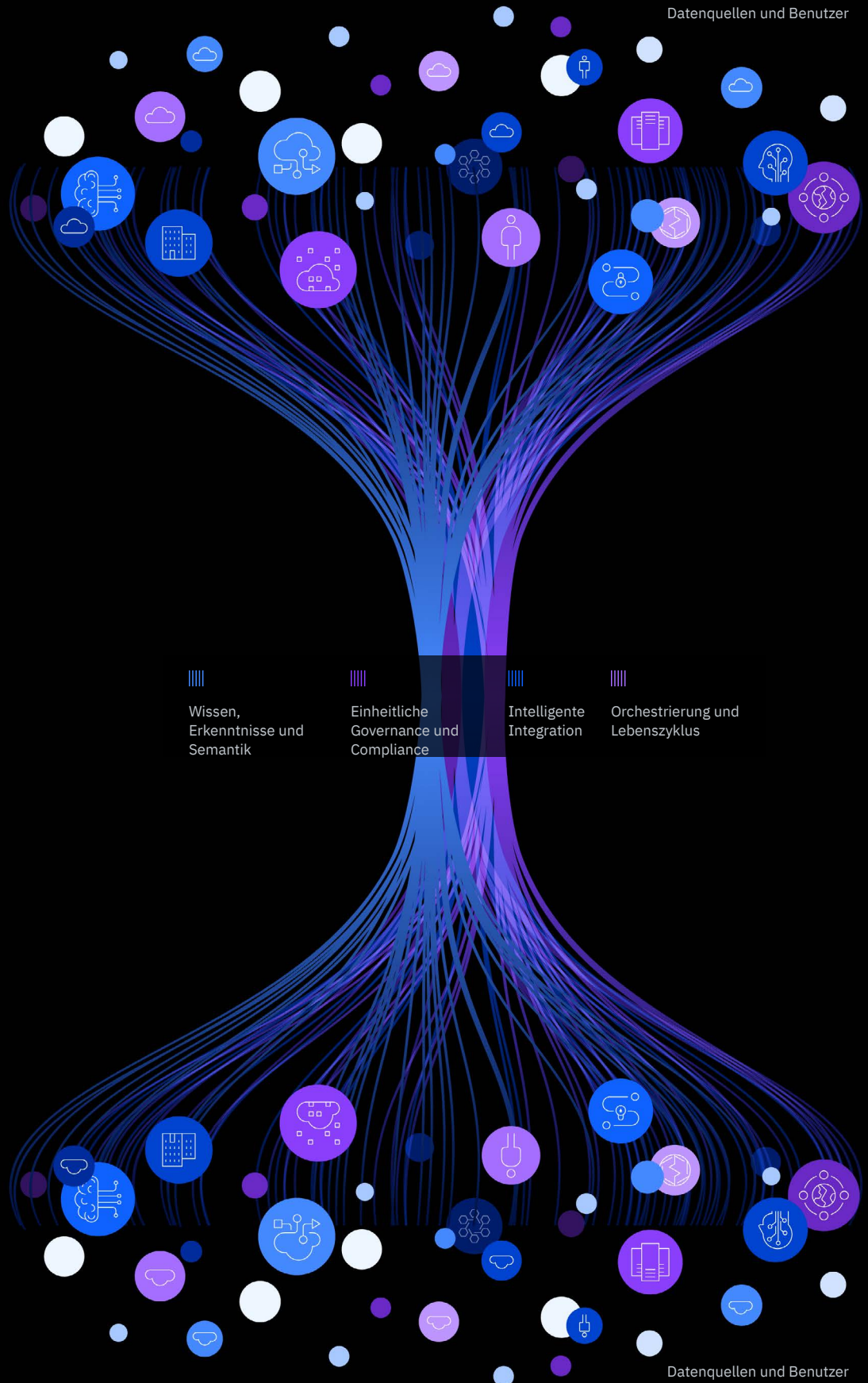
- Ermöglicht lokale Verwaltung und Steuerung von Metadaten, unterstützt aber eine globale, einheitliche Ansicht und die Durchsetzung von Richtlinien
- Automatische Anwendung von Richtlinien auf Datenbestände in Übereinstimmung mit globalen und lokalen Regeln
- Nutzt erweiterte Funktionen zur Automatisierung der Klassifizierung und Zusammenstellung von Datenbeständen
- Richtet automatisch abfragbare Zugriffswege für beliebige katalogisierte Assets ein, um die Aktivierung der Daten zu erhöhen

### Intelligente Integration

- Beschleunigt die Aufgaben des Ampere-Dateningenieurs durch automatisierten Fluss und Pipelineerstellung horizontal verteilter Datenquellen
- Ermöglicht Self-Service-Ingestion und Datenzugriff über beliebige Daten mit lokaler und globaler tiefgreifender Durchsetzung von Datenschutzrichtlinien
- Ermittelt automatisch die bestmögliche Ausführung durch optimierte Arbeitslastverteilung und Selbstabstimmung sowie Korrektur von Schemadriften

### Orchestrierung und Lebenszyklus

- Ermöglicht den Aufbau, das Testen, den Betrieb und die Überwachung von Datenpipelines
- Integration von KI-Funktionen in den Datenlebenszyklus zur Automatisierung von Aufgaben, Selbstoptimierung, Selbstheilung und Erkennung von Quelldatenänderungen, was automatisierte Updates erleichtert



# Geschäftsvorteile einer Data Fabric

Daten liefern nur dann einen Geschäftswert, wenn sie kontextualisiert sind und für jeden Benutzer oder jede Anwendung im Unternehmen zugänglich sind. Richtig implementiert, trägt eine Data Fabric dazu bei, dass diese Werte im gesamten Unternehmen auf möglichst effiziente und automatisierte Weise zur Verfügung stehen. Als solches hat das Gewebe drei wesentliche Vorteile:

1. Ermöglicht Self-Service-Datenkonsum und Zusammenarbeit
2. Automatisiert Governance, Schutz und Sicherheit; ermöglicht durch aktive Metadaten
3. Automatisiert Data Engineering-Aufgaben und erweitert die Datenintegration über hybride Cloud-Ressourcen hinweg

## Ermöglicht Self-Service-Datenkonsum und Zusammenarbeit

Durch die Integration von Daten aus verschiedenen Quellen und die Analyse eines größeren Teils der enormen Datenmenge, die täglich erzeugt wird, erhalten Unternehmen bessere Einblicke und können schneller auf veränderte Geschäftsanforderungen reagieren. Eine Data Fabric leitet Daten schnell denjenigen zu, die sie benötigen. Self-Service ermöglicht es dem gesamten Unternehmen, geeignete Daten schneller zu finden und mehr Zeit damit zu verbringen, diese Daten zu nutzen, um greifbare Erkenntnisse zu gewinnen.

Vorteile von Data Fabric für den Selbstbedienungsdatenverbrauch:

- Geschäftsanwender haben einen einzigen Zugangspunkt, um Daten im gesamten Unternehmen zu finden, zu verstehen, zu gestalten und zu konsumieren.
- Eine zentralisierte Daten-Governance und Lineage helfen den Anwendern zu verstehen, was die Daten bedeuten, woher sie kommen und wie sie mit anderen Assets in Beziehung stehen.
- Umfangreiche und anpassbare Metadatenverwaltung skaliert leicht und ist über APIs zugänglich.
- Der Self-Service-Zugriff auf vertrauenswürdige und kontrollierte Daten ermöglicht die Zusammenarbeit mit anderen Anwendern.

Eine Forrester Total Economic Impact Studie<sup>1</sup> ergab, dass diese Fähigkeiten bedeuten können:

USD  
**5,8M** bei Leistungen

ROI  
**459 %**

## Automatisiert Governance, Schutz und Sicherheit; ermöglicht durch aktive Metadaten

Eine verteilte aktive Governance-Ebene für alle Dateninitiativen reduziert Compliance- und regulatorische Risiken, indem sie Vertrauen und Transparenz schafft.

Sie ermöglicht die automatische Durchsetzung von Richtlinien für jeden Datenzugriff und bietet ein hohes Maß an Datenschutz und Compliance.

Durch den Einsatz von KI- und maschinellen Lerntechnologien können Data Fabric-Anwender ihren Automatisierungsgrad erhöhen, z. B. durch die automatische Extraktion von Data Governance-Regeln auf Basis von Sprache und Definitionen in regulatorischen Dokumenten. So können Unternehmen in wenigen Minuten branchenspezifische Governance-Regeln anwenden, um kostspielige Bußgelder zu vermeiden und die ethische Nutzung von Daten sicherzustellen, egal wo sie sich befinden.

Vorteile einer Data Fabric für kontrollierte Virtualisierung:

- Die Agilität, Sicherheit und Produktivität von Dateningenieuren, Datenwissenschaftlern und Business-Analysten wird erhöht.
- Mehrere globale Datenquellen erscheinen als eine Datenbank.
- Eine neue, branchenführende Erkennung von persönlich identifizierbaren Informationen (PII) und kritischen Datenelementen ist in massivem Umfang möglich.

Diese Fähigkeiten können bedeuten:

USD

**2,4M**

in Leistungen<sup>1</sup>

**430 %**

Leistungsverbesserung<sup>2</sup>

## Automatisieren von Data Engineering-Aufgaben und Erweitern der Datenintegration

Fortgeschrittenes Data Engineering bedeutet, dass praktisch jeder Datenzugriff oder -bereitstellungsprozess automatisiert ist und keinen langwierigen oder fehleranfälligen Codierungsprozess erfordert. Die Erweiterung der Integration nutzt Metadaten, um die Datenbereitstellung und den Zugriff zu optimieren.

Vorteile einer Data Fabric für Data Engineering und Integration:

- Automatisch optimierte Datenintegration hilft, die Datenbereitstellung zu beschleunigen.
- Automatischer Workload-Ausgleich und elastische Skalierung bedeuten, dass die Jobs für jede Umgebung und jedes Datenvolumen bereit sind.
- Ausfallsicherheit und CI/CD-Automatisierung sind integriert.
- Der automatisierte Prozess zur Erfassung von Änderungen in Echtzeit unterstützt die Bereitstellung von Qualitätsdaten für Geschäftsprozesse.
- Maschinelles Lernen kann benutzerdefinierte Prozesse zur Datenerkennung, -klassifizierung und -kuratierung automatisieren und erweitern, was zu einer schnelleren Time-to-Value führt.
- Kontinuierliche Analysen können automatisch und in Echtzeit durchgeführt werden, egal wo die Daten liegen.

Ergebnisse von einem führenden Einzelhändler<sup>3</sup> zeigen:

**60X**

Beschleunigung der Datenlieferzeit

**20X**

Schnellere Kundenaffinitätsanalysen

# Wie IBM eine Data Fabric liefert

IBM Cloud Pak® for Data macht dieses Konzept einer Data Fabric möglich. IBM Cloud Pak for Data ist eine Insight-Plattform, die das Sammeln, Organisieren und Analysieren von Daten vereinfacht und automatisiert. beschleunigt die Einführung von KI in Ihrem Unternehmen.

Mit seinen Fähigkeiten, Daten überall zu verbinden, Workloads überall auszuführen und KI in hybriden Cloud-Umgebungen in großem Umfang zu erstellen, bereitzustellen und zu verwalten, ist IBM Cloud Pak for Data der Enabler für eine digitale Transformation des Unternehmens.

Die Plattform bietet nahtlose Integration in einem hybriden Unternehmen für:

- IBM Cloud Pak for Data-Partner
- Externe Anwendungen und Datenquellen
- Erweiterte KI-basierte Funktionen für Datenmanagement und Governance

Diese Grundlage macht kuratierte Daten für die Verbraucher mit der optimalen Balance von Kosten, Leistung und Compliance verfügbar und verfügt über die Intelligenz zur Orchestrierung und Optimierung der Datenverarbeitung auf der Grundlage von Arbeitslasten, Datenlokalisierung und Datenrichtlinien.

IBM Cloud Pak for Data nutzt die folgenden Funktionen, um die geschäftsfähigen Daten zu liefern, die eine Datenstruktur erfordert. All diese Funktionen spielen eine Rolle bei der Unterstützung der Data Fabric-Architektur.

**1. Metadatenbasierter Wissenskern** Erleichtert das Auffinden von Datenquellen und Katalogen, reichert Datenbestände an und führt Analysen durch, um Erkenntnisse für eine stärkere Automatisierung mithilfe von KI zu gewinnen. Der Wissenskern wird genutzt, um semantische Suche in den Marktplat zu einzubringen.

2. **Self-Service-Datenmarkt** Der Datenkatalog der nächsten Generation, der Datenkonsumenten, wie z. B. Geschäftsanalysten, dabei unterstützt, Daten aus der gesamten Datenlandschaft des Unternehmens abzurufen.
3. **Intelligente Integration** Ermöglicht die Datennutzung durch Extraktion, Virtualisierung, Umwandlung und Streaming von Daten. Sie ist in den Wissenskern integriert, um die Datenintegration zu automatisieren, und verfügt über die Intelligenz, zu entscheiden, welcher Integrationsansatz auf der Grundlage von Arbeitslasten und Datenrichtlinien am besten geeignet ist. Sie kann auch für die Datenaufbereitung im Rahmen von Data Engineering Workloads oder zur Erstellung von Datenprodukten verwendet werden. Schließlich bietet sie die Möglichkeit, Aktualisierungen von Datenprodukten zu veröffentlichen.
4. **Governance** katalogisiert und kuratiert Metadaten, definiert Datenrichtlinien für den Datenschutz, kuratiert Daten, erfasst die Datenabfolge und führt andere Aufgaben in Bezug auf Sicherheit und Compliance aus. Diese Schicht versteht das Datenformat (z. B. strukturiert vs. unstrukturiert) und die Datenbedeutung (z. B. öffentlich vs. geschützt) und wendet die richtigen Richtlinien auf jedes Bit an Daten und jeden potenziellen Benutzer an. Anstatt Standards und Regeln manuell auf Daten anzuwenden, bedeutet diese integrierte Plattformfunktion, dass sie auf organisatorischer Ebene angewendet werden können und sich bei Bedarf in den verschiedenen Daten Ressourcen ausbreiten. Analytische Modelle in verschiedenen Tools können miteinander kommunizieren; die Durchsetzung von Datenrichtlinien auf der granularen Ebene kann weitgehend automatisiert werden.
5. **Einheitliche Entwicklung und Unternehmensaktivitäten** Ermöglicht einen einheitlichen Lebenszyklus zur Konfiguration und zum Betrieb aller Aspekte der Datenplattform in der Produktion.

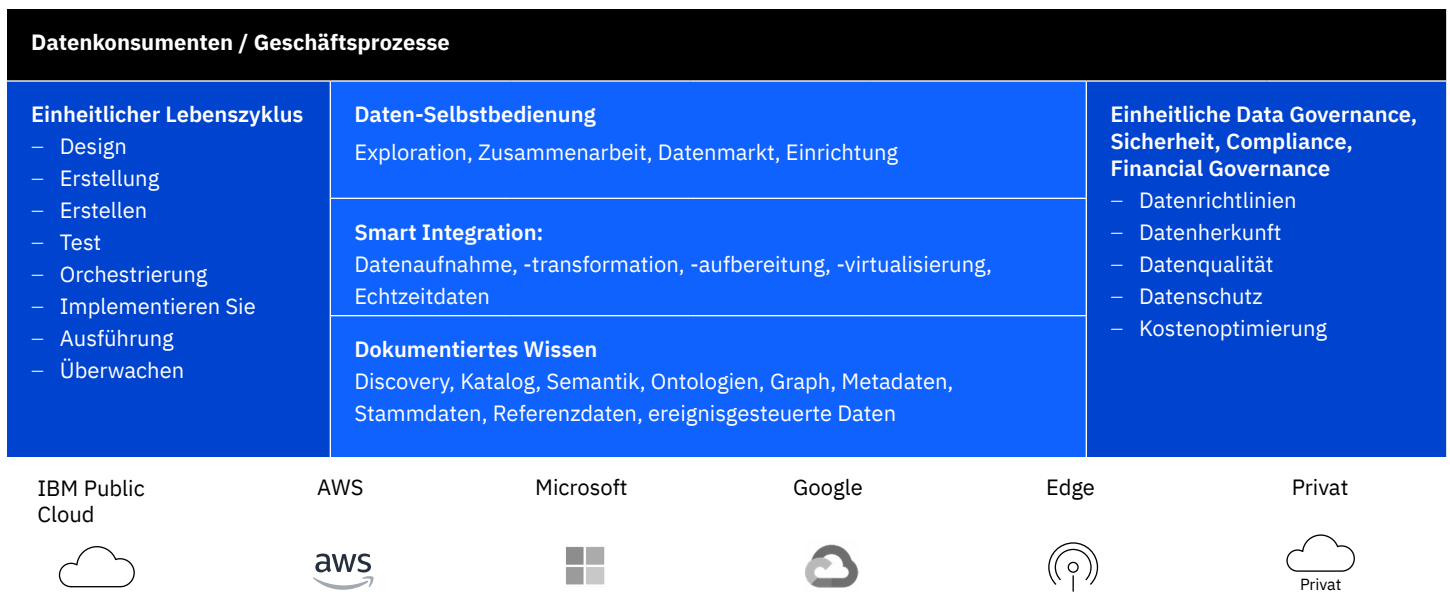


Abbildung 1. Die Funktionen der IBM Cloud Pak for Data-Plattform - die alle die KI-Entwicklung und die Data Fabric unterstützen



# Data Fabric in Aktion

Der beste Weg, den Wert einer Data Fabric zu verstehen, ist, die geschäftlichen Vorteile zu sehen, die tatsächliche Unternehmen durch Implementierung einer Data Fabric auf einer Insight-Plattform erzielt haben.

## Energieunternehmen

Im Abschnitt "Herausforderungen" zu Beginn dieses Whitepapers wurde der Fall eines nordamerikanischen Energieunternehmens vorgestellt, das seine digitale Transformation durchlief.

Dieser Kunde beauftragte IBM mit der Implementierung einer Data Fabric-Architektur auf Basis von IBM Cloud Pak for Data.

Mit einer flexiblen und integrierten Datenstruktur war das Unternehmen in der Lage, eine Reihe wichtiger Datenprojekte über verschiedene Abteilungen hinweg zu implementieren, darunter:

- eMobilität
- Gasbetrieb Dokumentenerkennung, einschließlich Handschriftextraktion
- Stromkundensegmentierung und Lastprognose
- Asset-Management
- COVID-19 Lasteinwirkungen
- Risikomodell für die Rückkehr an den Arbeitsplatz

### Mit einem Datengewebe ist diese Organisation in der Lage:

- Bieten Sie mehreren Geschäftseinheiten direkten Datenzugriff über eine Self-Service-Insight-Plattform.
- Nutzen Sie die zweckgebundene Rechenkapazität, um Modelle effizient auf Milliarden von Datenzeilen auszuführen.
- Entwickeln Sie gemeinsam Modelle und setzen Sie diese Modelle einfach ein, um im gesamten Unternehmen Erkenntnisse zu gewinnen.

## Verringerung der Bewegungen und Verbesserung der Übersichtlichkeit

Ein großer Industriedienstleister verbesserte die Data Governance und erleichterte die Einhaltung gesetzlicher Vorschriften.

Diese Organisation hatte Schwierigkeiten, große Datenmengen in ihren Cloud-basierten Data Lake zu verschieben. Für sie bestanden zwei Anforderungen, und zwar, jede Instanz des Datenzugriffs mit einer starken Governance-Ebene zu versehen und die Datenqualität zu bestimmen, bevor sie den Geschäftsanwendern Zugriff gewähren.

Durch die Implementierung einer Data Fabric, die auf IBM Cloud Pak for Data aufbaut, erzielte dieses Unternehmen enorme Verbesserungen in den Bereichen Data Governance, Data Compliance und Datentransformationsprozesse. Mit einer Data Fabric können sie Geschäftsanwendern einen einfachen und sicheren Zugriff auf Hunderte von Datenquellen in ihrem Cloud-basierten Data Lake und auf lokale SAP-Datenquellen bieten.

Sie nutzen auch Tools zur Beschleunigung der Regulierung der Branche, um Datenquellen auf mögliche PII-Daten zu scannen, die der GDPR- und CCPA-Regulierung unterliegen würden. Die automatische Datenkennzeichnung hilft bei der Identifizierung von Datenquellen, die für den Umgang mit PII verwaltet werden müssen, sowohl vor Ort als auch in der Cloud.

Mit besserem Zugriff und schnellerem PII-Screening können Geschäftsanwender relevante Daten für wichtige Erkenntnisse auswerten, ohne lange Wartezeiten für den Datenzugriff oder das Risiko, geschützte Daten preiszugeben.

### Mit einem Datengewebe ist diese Organisation in der Lage:

- Stellen Sie eine angemessene Data Governance sicher und nutzen Sie gleichzeitig Daten aus dem gesamten Unternehmen.
- Erhalten Sie vertrauenswürdige Daten und reduzieren Sie den Aufwand für die Datenaufbereitung.
- Unterstützen Sie bei der Einhaltung von Datenschutzbestimmungen wie GDPR und CCPA.

## Data Fabric ist der transformative nächste Schritt für Ihr Unternehmen

Mit einer Data Fabric, die auf der IBM Cloud Pak for Data-Technologie aufbaut, können Sie die Datenermittlung, Data Governance und Datennutzung in einer hybriden und Multi-Cloud-Datenlandschaft hyper-automatisieren. Verwenden Sie eine Data Fabric, um eine schnellere Time-to-Value für Geschäftsanwender, eine höhere Produktivität für Data Engineering und Operations sowie eine größere Governance und Compliance-Treue zu ermöglichen.

[Um loszulegen, testen Sie IBM Cloud Pak for Data kostenlos.](#)

[Vereinbaren Sie ein kostenloses persönliches Beratungsgespräch mit einem IBM Data and AI-Mitarbeiter.](#)



© Copyright IBM Corporation 2021

IBM Deutschland GmbH  
IBM-Allee 1  
71139 Ehningen  
ibm.com/de

IBM Österreich  
Obere Donaustraße 95  
1020 Wien  
ibm.com/at

IBM Schweiz  
Vulkanstrasse 106  
8010 Zürich  
ibm.com/ch

IBM, das IBM-Logo und IBM Cloud Pak sind Marken oder eingetragene Marken der International Business Machines Corporation, in den USA und/oder anderen Ländern. Weitere Produkt- und Servicennamen können Marken von IBM oder anderen Unternehmen sein. Eine aktuelle Liste der IBM Marken finden Sie unter [ibm.com/trademark](http://ibm.com/trademark).

Dieses Dokument ist zum Datum seiner Erstveröffentlichung aktuell und kann jederzeit von IBM geändert werden. Nicht alle IBM Angebote sind in jedem Land, in welchem IBM tätig ist, verfügbar.

Die erwähnten Leistungsdaten und Kundenbeispiele dienen lediglich der Veranschaulichung. Die tatsächlichen Performance-Ergebnisse können je nach spezifischen Konfigurationen und Betriebsbedingungen variieren. Die Verantwortung für den Betrieb von Produkten, Programmen und Services anderer Anbieter mit IBM Produkten und Programmen liegt beim Kunden. Vertragsbedingungen und Preise erhalten Sie bei den IBM Geschäftsstellen und/oder den IBM Business Partnern. Die Produktinformationen geben den derzeitigen Stand wieder. Für IBM Produkte gelten die Gewährleistungen, die in den Vereinbarungen vorgesehen sind, unter denen sie erworben werden.

Erklärung zu geeigneten Sicherheitsvorkehrungen: Zur Sicherheit von IT-Systemen gehört der Schutz von Systemen und Informationen in Form von Vorbeugung, Erkennung und Reaktion auf unbefugten Zugriff innerhalb des Unternehmens und von außen. Unbefugter Zugriff kann dazu führen, dass Informationen geändert, gelöscht oder veruntreut werden. Ebenso können Ihre Systeme beschädigt oder missbräuchlich verwendet werden, einschließlich zum Zweck von Angriffen. Kein IT-System oder Produkt kann umfassend als sicher betrachtet werden. Kein einzelnes Produkt, kein einzelner Service und keine einzelne Sicherheitsmaßnahme können eine unbefugte Verwendung oder einen unbefugten Zugriff mit vollständiger Wirksamkeit verhindern. IBM Systeme und Produkte werden als Teil eines dem Gesetz entsprechenden, umfassenden Sicherheitskonzepts entwickelt, sodass die Einbeziehung zusätzlicher Betriebsprozesse erforderlich ist. Ferner wird vorausgesetzt, dass andere Systeme, Produkte oder Services so effektiv wie möglich sind. **IBM ÜBERNIMMT KEINERLEI GEWÄHR DAFÜR, DASS SYSTEME, PRODUKTE ODER DIENSTLEISTUNGEN VOR BÖSWILLIGEM ODER RECHTSWIDRIGEM VERHALTEN EINER PARTEI GESCHÜTZT SIND ODER IHR UNTERNEHMEN DAVOR SCHÜTZEN.**

Der Kunde ist für die Einhaltung der geltenden Gesetze und Verordnungen selbst verantwortlich. IBM erteilt keine Rechtsberatung und gibt keine Garantie bzw. Gewährleistung bezüglich der Konformität von IBM Produkten oder Services mit den geltenden Gesetzen und gesetzlichen Bestimmungen.

- 1 Eine Forrester New Technology: Projected Total Economic Impact™ Studie im Auftrag von IBM, Februar 2020
- 2 Die Leistungsmessungen wurden in einer kontrollierten Testumgebung in den IBM Silicon Valley Labs unter Verwendung der IBM Datenvirtualisierung gegen verschiedene 100 TB große Datenquellen durchgeführt. Die Messungen im Mai 2019 und die Leistungsgewinne werden mit IBM Federation verglichen.
- 3 Basierend auf den Erfahrungen eines IBM Kunden.