

„Data Fabric Revisited“ –

**Was hat sich in den letzten 2 Jahren
in diesem Feld getan?**



Vortrag zur erneuten Beleuchtung eines „Buzzwords“

Marina Doehling, Executive Consultant, cat out
Jens Borchers, Executive Consultant, cat out

cat out - IT-BERATUNG UND IMPULSGEBER

Lösungskompetenz für komplexe IT-Herausforderungen technisch, organisatorisch und rechtlich

- **Beratung:** Individuell, ganzheitlich oder in spezifischen Teilgebieten.
- **Analyse:** Bestandsaufnahme, Identifizierung Maßnahmen und Prioritäten
- **Projektleitung:** Effektive Steuerung von Projekten und Dienstleistern
- **Umsetzung:** Realisierung Projekte in enger Zusammenarbeit mit den Unternehmen.
- **Team-Building:** Aufbau und Ergänzung von Projektteams mit den passenden Talenten.
- **Netzwerk:** Zugang zu einem breiten Netzwerk von Experten und Spezialisten.

Kompetenzbereiche:

- **Branchen:** Handel, Logistik, Banking und Finance, Industrie, Telekommunikation
- **Schwerpunkte:** Digitalisierung, Transformation, Migration, Qualitätssicherung, User Management, Daten Management, Integration, Software Development
- **Technologien:** Beherrschung heterogener Systemlandschaften von Mainframe bis Cloud.



Wer wir sind - Steckbrief

Marina Döhling Diplom-Betriebswirt

Consultant mit umfassenden Kenntnissen in Datenbanken, Datenkommunikation, Integration und Softwareentwicklung. Profunde Fähigkeiten in Projektmanagement, Active Sourcing, Partnermanagement und Business Development, verstärkt durch Expertise in Datenschutz, IT-Sicherheit und Compliance. Erfahrungen Kenntnisse in IT-Infrastrukturen, Mainframes und Cloud-Technologien. Branchenübergreifendes Know-how in Handel, E-Commerce, Logistik, Bankwesen, Telekommunikation und Industrie.



Jens Borchers Diplom-Mathematiker, Physik, Computer Science

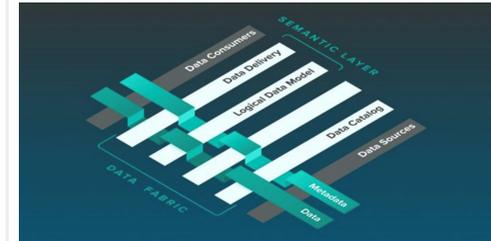
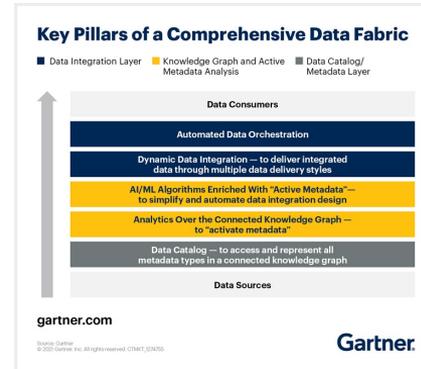
IT-Experte mit tiefgreifenden Kenntnissen in der Verwaltung und Optimierung von großen OLTP-Datenbanksystemen. Spezialisiert auf Reengineering, Migration und Softwareentwicklung über diverse Plattformen, von Mainframes bis hin zu Cloud-Technologien. Verfügt über spezialisierte Fähigkeiten im Application Management, Qualitätssicherung und Test-Management. Erfahren in Near-/Off-Shoring Projekten. Tiefgehende Kompetenzen im Datenschutz, Daten- und IT-Sicherheit, insbesondere mit Bezug auf MaRisk und BAIT, sowie Data Governance. Branchenkenntnisse konzentrieren sich auf den Banken- und Finanzsektor, inklusive Erfahrung im Handel, in der Industrie und im öffentlichen Sektor.



Data Fabric – Was haben wir damals gesehen und erwartet?

Unser Vortrag am 22.04.2022

- Data Fabric war damals ein neuer Ansatz für das Management und den Zugriff auf die (weit) verteilten Daten eines Unternehmens.
- Sollte – wie immer – alle Probleme des Datenmanagements „ganzheitlich“ lösen.
- Heute schauen wir, was sich in den letzten beiden Jahren – z.T. drastisch – verändert hat.



Ist Data Fabric die Lösung?



- Kontrolle über rasant wachsende, verteilte Datenmengen
- Kontrolle über die Daten im gesamten Unternehmen inkl. Peripherie
 - DBA nicht mehr der Master
 - Unternehmensweites Datenmodell „Illusion“?
- Überblick über Daten-Redundanz, Verteilte Daten, Dark Data
- Data Lakes, Data Warehouse u.a. obsolet?
- Orchestrierung der Daten – DBMS, strukturierte/unstrukturierte u.a. per Data Fabric?

Data Fabric – Grundlegende Definition



Data Fabric ist eine Architektur und Set von Datenmanagement-Diensten, die darauf abzielen, eine konsolidierte Perspektive auf Unternehmensdaten zu ermöglichen, ohne physische Duplikate von Ursprungsdaten zu erzeugen. Dabei wird eine semantische Ebene etabliert, welche die primären Datenquellen repräsentiert und die Durchführung von Suchvorgängen und Analysen erleichtert.

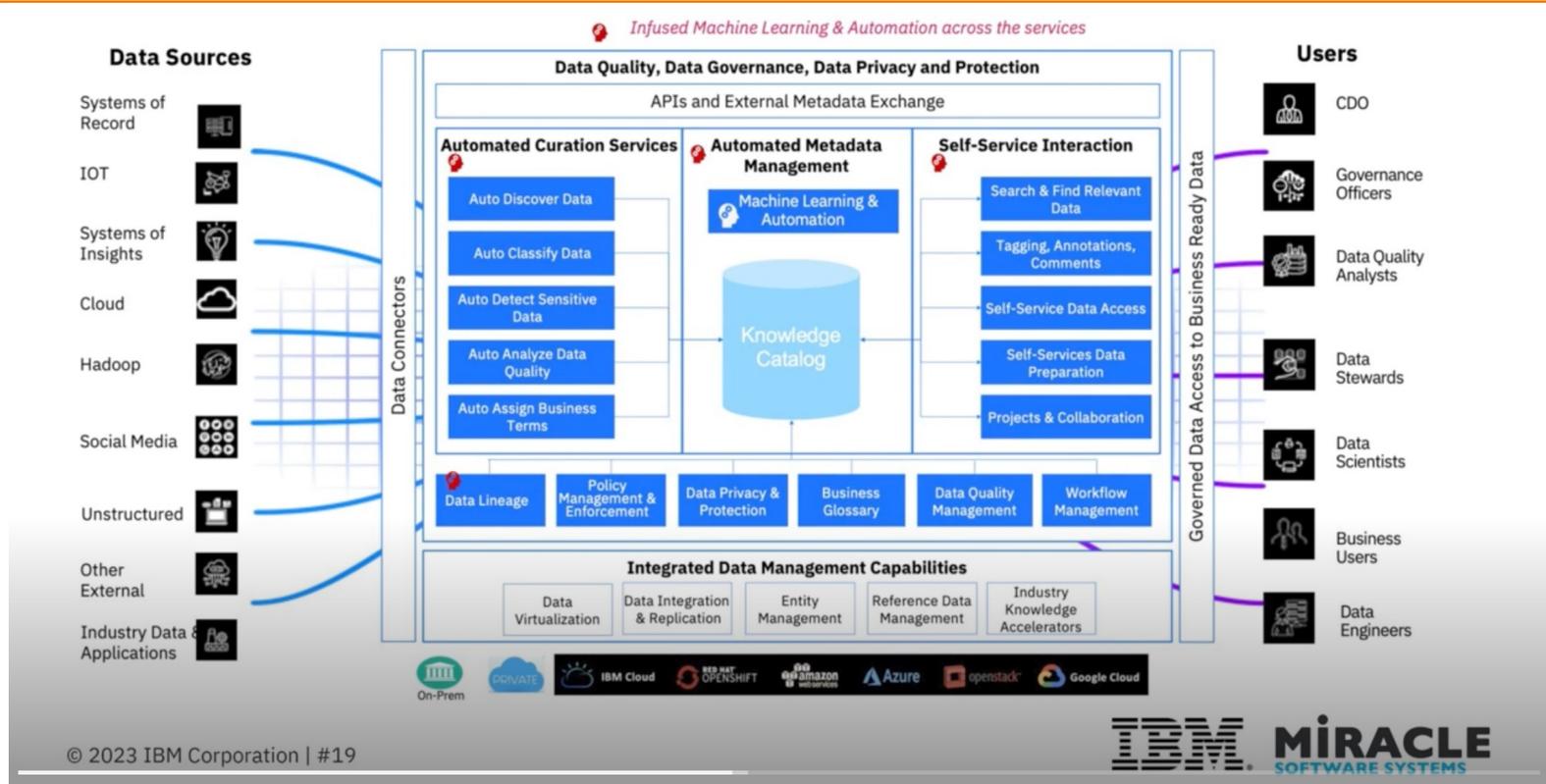
Ziel: eine reibungslose Datenintegration und Zugänglichkeit über diverse Plattformen und geografische Standorte hinweg, um eine vereinfachte Handhabung und Verwaltung von Daten über unterschiedliche Speichersysteme und Betriebsumgebungen zu ermöglichen.

Data Fabric – Kernkonzepte



- **Verteilte Datenarchitektur:** Zugriff auf allen Formen von Daten (roh, konsolidiert), unabhängig vom Speicherort, sei es in Clouds, On-Premise-Servern oder in Edge-Geräten.
- **Agilität und Flexibilität:** Durch APIs und Microservices schnelle Anpassung an neue Datenquellen und -formate => flexible Datenintegration.
- **Selbstservice-Reporting:** Endnutzer, insbesondere Nicht-IT-Experten, greifen auf Daten, ohne genaues Wissen über Speicherort und Formatierung
- **Automatisierte Datenmanagementprozesse:** KI und maschinelles Lernen zur Automatisierung Daten-Katalogisierung, -Qualitätsmanagement, Metadatenmanagement.
- **Sicherheit und Governance:** Datenintegrität und -konformität werden sichergestellt.

Data Fabric – Gesamtarchitektur gemäß IBM



Data Fabric – vs. DATA WAREHOUSE, DATA LAKE, DATA MESH

Eigenschaft	Data Warehouse	Data Lake	Data Mesh	Data Fabric
Hauptfokus	Strukturierte Datenspeicherung für Analysen	Rohdatenspeicherung in nativem Format	Dezentrale Datenverwaltung und -bereitstellung	Integrierte Datenverwaltung über verschiedene Umgebungen
Datentypen	Hauptsächlich strukturierte	Strukturierte, semi-strukturierte, unstrukturierte	Alle Typen, mit Fokus auf Produktorientierung	Alle Typen, über verschiedene Quellen, Formate und Plattformen
Schema	Schema-on-Write (bei Speicherung)	Schema-on-Read (bei Abfrage)	Schema-flexibel, orientiert sich an domain-driven Design	Schema-“frei“, flexibel, abhängig von Implementierungen
Optimiert für	Schnelle Abfragen und Berichterstattung	Flexible Datenspeicherung und -erkundung	Autonomie der Teams und Interoperabilität	Datenzugriff, -integration und -management in Unternehmens-Maßstab
Architektur	Zentralisiert	Zentralisiert	Dezentralisiert	Netzwerkverbunden, zentralisiert oder dezentralisiert
Einsatzgebiet	Business Intelligence, strukturierte Berichte	Data Science, Big Data-Analysen	komplexe Datenlandschaften und Autonomie-Anforderungen	agile und integrierte Datenlandschaft, Basis für eigene GenAI-Anwendungen

Data Fabric - Primary Point of Truth

„Primary Point of Truth“ (PPoT) in einer Data Fabric-Architektur

- Zentrales Element für Datenverlässlichkeit und -integrität
- Einheitliche Sicht auf die Datenbestände, auch wenn diese physisch in diversen Systemen oder an verschiedenen Orten gespeichert sind
- Voraussetzung ist ein Metadaten-Hub, die zentrale Sicht auf Metadaten, die für eine effektive Datenverwaltung und -nutzung entscheidend ist
- Stellt sich die Frage der „Historisierung“ (wie z.B. in Data Warehouses Standard)



Data Fabric – ENTWICKLUNGEN SEIT 2022

Datenmanagement-Funktionen:

- **Datenqualitätsmanagement:** Tools wie Talend Data Quality bieten Funktionen zur Überwachung und Verbesserung der Datenqualität, Bereinigung und Anreicherung von Daten.
- **Metadatenmanagement:** Lösungen wie Informatica Enterprise Data Catalog ermöglichen ein umfassendes Metadatenmanagement, um die Herkunft, Nutzung und Bedeutung von Daten im gesamten Unternehmen zu verstehen.

Datenzugriff: Unterstützung für Multi- und Hybrid-Cloud-Umgebungen

- **Datenintegration über Clouds hinweg:** Snowflake ermöglicht Datenintegration und -analyse über verschiedene Cloud-Plattformen und On-Premises-Umgebungen hinweg
- **Datenmobilität:** Google Anthos bietet Tools für die Datenmobilität, um Anwendungen und Daten zwischen verschiedenen Cloud- und On-Premises-Umgebungen zu verwalten.

Data Fabric – ENTWICKLUNGEN SEIT 2022

Automatisierung und Self-Service-Fähigkeiten:

- **Automatisierte Datenpipelines:** Alteryx bietet eine Drag-and-Drop-Umgebung, in der Benutzer Datenpipelines erstellen können, die Daten vorbereiten, mischen und analysieren, ohne Code schreiben zu müssen.
- **Self-Service-Datenzugriff:** Tableau ermöglicht es Endnutzern, ohne umfangreiche technische Kenntnisse auf Daten zuzugreifen und interaktive Dashboards zu erstellen.

Unterstützung durch Künstlicher Intelligenz (KI) und Maschinelles Lernen (ML):

- **Automatisierte Datenkatalogisierung:** Tools wie IBM Watson Knowledge Catalog nutzen KI, um Daten automatisch zu katalogisieren und zu klassifizieren
- **Anomalie-Erkennung:** Plattformen wie Splunk verwenden ML, um ungewöhnliche Muster in Daten zu erkennen zur Früherkennung von Sicherheitsbedrohungen

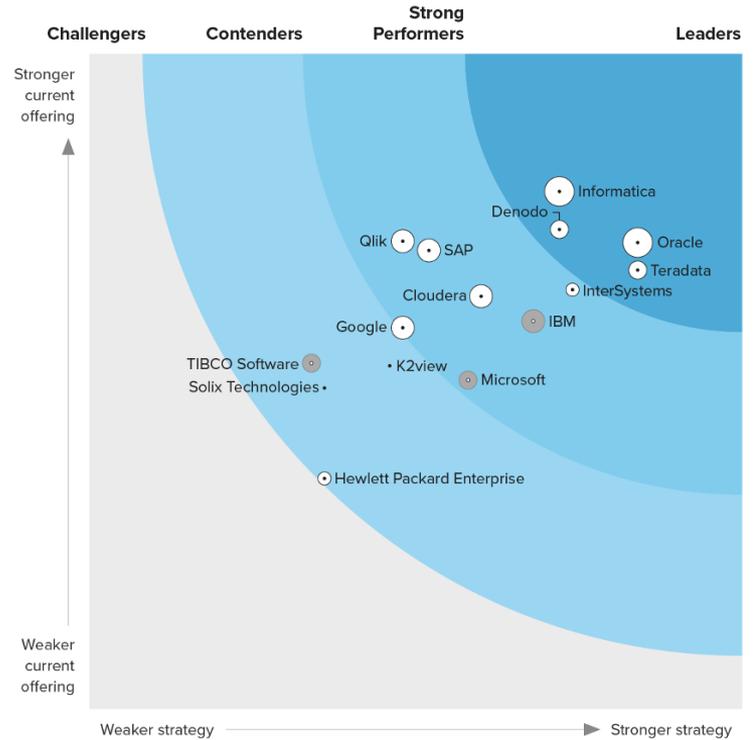
Data Fabric – ENTWICKLUNGEN SEIT 2022

- **Wenig grundlegende Veränderungen:** Seit 2022 gab es im Bereich Datenmanagement keine revolutionären Durchbrüche. Die etablierten Systeme und Praktiken bleiben weitgehend unverändert.
- **Fokus auf bestehende Technologien:** Unternehmen konzentrieren sich weiterhin auf die Optimierung und Effizienzsteigerung vorhandener Datenmanagement-Systeme.
- **Integration und Konsolidierung:** Ein anhaltender Trend ist die Integration verschiedener Datenquellen und -systeme, um einen umfassenderen und effizienteren Zugriff auf Daten zu ermöglichen.
- **Use Cases** im Finanzsektor – primär von Anbietern, generell verhaltend, noch „übersichtlich“

Data Fabric – FORRESTER WAVE™ 2022 VS. 2024



The Forrester Wave™ - Enterprise Data Fabric Q2 2022



Forrester Wave™: Enterprise Data Fabric Scorecard, Q1 2024

Data Fabric - AKTUELLE BEWERTUNGEN VON ANALYSTEN



■ „A Dream too Soon“

- PBT Group – „Data Fabric braucht noch 10 Jahre“
- <https://www.pbtgroup.nl/data-fabric-a-dream-too-soon/>

■ „Data Fabric Overview“

- Spotlight by Bloor Research – „Noch keine durchgängige Lösung vorhanden“
- <https://www.bloorresearch.com/technology/data-fabric/>

Data Fabric - AKTUELLER STAND UND EINSATZFÄHIGKEIT

- PBT Group: Julian Thomas - „Data fabric: A dream too soon?“, February 2023
 - Nicht viele Firmen haben heute ein ganzheitliches Informationsmanagement implementiert. Aber ohne ein solches wird auch Data Fabric keinen Effekt bringen.
 - Unternehmen haben kaum an der Oberfläche gekratzt, um die Grundlagen der Datenanalyse richtig zu beherrschen, ohne auch nur einen Blick auf die Möglichkeiten zu werfen, die KI und robotergestützte Prozessautomatisierung bieten können.
 - Die Realität ist, dass die meisten Unternehmen darauf warten, dass eine weltweit anerkannte Methodik verfügbar wird, bevor sie den Übergang zu Data Fabric in Betracht ziehen.
 - Gibt es eine dringende Notwendigkeit für Data Fabric? Vielleicht in 10 Jahren wird es einfacher „alles“ zu integrieren.

Data Fabric - AKTUELLER STAND UND EINSATZFÄHIGKEIT

- Bloor Research: Andy Hayler – Data Fabric Overview, March 2024 (last update)
 - Das Herz von Data Fabric bleibt der Data Catalog mit einem Fachbegriff-Glossar
 - Data Fabric ist nicht all-umfassend – andere Ansätze bleiben erhalten (z.B. MDM)
 - Wie wird Data Fabric redundante und ggf. inkonsistente Daten derselben Entität handhaben?
 - Letztlich hat heute keiner der Anbieter eine durchgängige „fertige“ Lösung
 - Use Cases meist von den Anbietern publiziert und eher „übersichtlich“

Data Fabric und KI – der „Game Changer“ ?

- Data Scientists und Data Engineers erstellen eine Vielzahl von Analysen, verbringen viel Zeit In Anwendungen und Systemen – Anfragestau
- Umsetzung Sprache in z.B. SQL-Query durch Einsatz von KI z.B. ChatGPT
- Fachabteilung kann allein Anfragen in natürlicher Sprache stellen, mit Hilfe der KI wird diese Anfrage formuliert, mit entsprechenden API die Verarbeitung angestoßen und die Daten zur Verfügung gestellt
- Einsatz derzeit für Unternehmen mit passender IT Abteilung, für KMU´s nur begrenzt erreichbar



Data Fabric – KI-ENTWICKLUNG UND DATEN-MANAGEMENT

- **Rasante Fortschritte in der KI:** Seit 2022 hat die Entwicklung im Bereich der künstlichen Intelligenz (KI) signifikante Fortschritte gemacht, insbesondere in den Bereichen maschinelles Lernen, natürliche Sprachverarbeitung (LLM) und generative Ansätze (GPT)
- **Automatisierung von internen Reporting-Prozessen:** KI-Technologien ermöglichen die Automatisierung komplexer Reporting-Prozesse, Verbesserung der Genauigkeit von Vorhersagemodellen und Trendanalysen
- **Selbstbedienungs-Reporting:** KI-getriebene Tools erleichtern das Self-Service-Reporting, indem sie Benutzern ermöglichen, ohne technisches Know-how auf komplexe Datenanalysen zuzugreifen (z.B. ChatGPT, aber auf unternehmenseigenen Daten).

Data Fabric – KI-ENTWICKLUNG UND DATEN-MANAGEMENT

- Während das Datenmanagement seit 2022 relativ konstant geblieben ist, haben die Fortschritte in KI-Anwendungen das Potenzial, die Art und Weise, wie Unternehmen mit ihren (zusammen mit extern verfügbaren) Daten umgehen und sie für das interne Reporting nutzen, grundlegend zu verändern.
- Der Einsatz von KI wird auch auf Data Fabric-Architekturen wesentlichen Einfluss haben:
 - KI als Unterstützung zum Aufbau der für Data Fabric unverzichtbaren **fachlichen** Metadaten
 - Mustererkennung, Anomalien, Abgleich Modelle vs. reale Daten
 - KI als Nutzer von Data Fabric-Architekturen
 - Verbindung von LLMs mit unternehmensinternen Daten

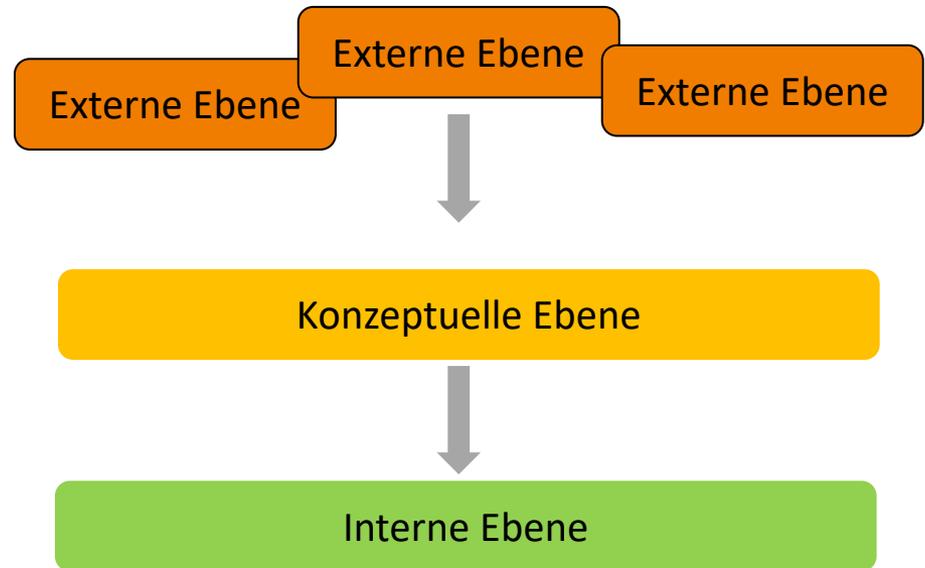
Data Fabric – ANSI/X3/SPARC-MODELL (50 JAHRE!)

1975 - American National Standards Institute, Standards Planning and Requirements Committee (SPARC)

Interaktion / Sicht der Benutzer und Systeme,
Bereitstellung verschiedener Ansichten
auf die Daten

einheitliche Sicht auf die gesamte Datenbank,
welche Daten gespeichert sind , in welchen
Beziehungen sie stehen, unabhängig vom
physischen Speicherort

physischer Speicher- und
Datenorganisationstechniken



KONTAKT



Dipl.-Betriebsw. Marina Döhling
Executive Consultant

M. +49 (0) 172 43 00 429
marina.doehling@cat-out.com



Dipl.-Math. Jens Borchers
Executive Consultant

M. +49 (0) 174 88 99 675
jens.borchers@cat-out.com

cat out gmbh

info@cat-out.com

www.cat-out.com