

„Data Fabric“ - Ein neues Konzept oder Altbekanntes in neuer Verpackung?

Impulsvortrag zur Beleuchtung eines aktuellen Buzzwords
Round Table Data Campus 16.02.2022

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

Wer sind wir



Die cat out

- Unternehmensberatung für Daten- und Informationssicherheit – mit Schwerpunkt und Blick auf den Schutz und Monitoring von unerwünschtem Verlassen von Daten und Informationen aus dem Unternehmen
- In Kombination aus technischen Lösungen, organisatorischen Maßnahmen und Regulatorik
- Basierend auf Partnerschaften mit Spezialisten, Dienstleistern und Lösungsanbietern

Wer wir sind - Steckbrief

Marina Döhling

Diplom-Betriebswirt



Background: Datenbanken, Daten-Kommunikations-Systeme, Daten-Integration, Entwicklung

Tätigkeiten: Projektmanagement, Active Sourcing, Partner Management,
Business Development, Datenschutz und Daten/IT-Sicherheit, Regulatorik,

Langjährige Erfahrung im Umfeld Mainframe bis Cloud

Branchen: Handel (eCommerce), Logistik, Banken, Telco, Industrie

Jens Borchers

Diplom-Mathematiker, Physik, Computer Science



Background: Datenbanken, große OLTP-Systeme, Reengineering, Migration, Entwicklung

Schwerpunkte: Application Management, Near-/Off-Shoring, Software-Qualitätssicherung und Test-
Management, Datenschutz und Daten/IT-Sicherheit, Regulatorik (MaRisk, BAIT), Data Governance

Langjährige Projekterfahrung im Umfeld von Mainframe bis Cloud

Branchen: primär Banken, Finanzdienstleister, aber auch Handel, Industrie, Öffentliche Hand

Data Fabric – der Begriff



„Fabric“

- „Gefüge“, „Struktur“, „Gewebe“
- Nicht: „Fabrik“!
- Und nicht zu verwechseln mit „Data Fabrication“
 - Absichtliche Falschdarstellung von Forschungsergebnissen

Data Fabric - Definition nach Gartner



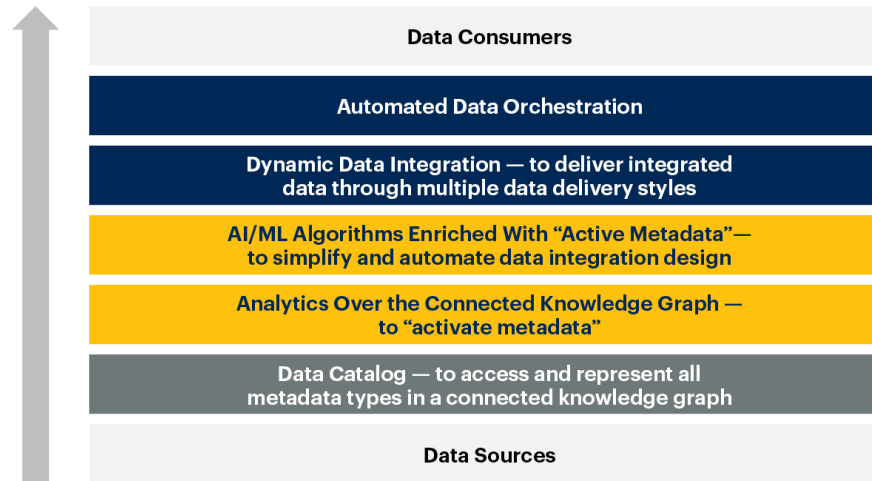
„Gartner defines data fabric as

- a **design concept** that serves as an integrated layer (fabric) of data and connecting processes.
- A data fabric utilizes continuous analytics over existing, discoverable and inferenced metadata assets to support the design, deployment and utilization of integrated and reusable data across all environments, including hybrid and multi-cloud platforms.“
- Quelle: <https://www.gartner.com/smarterwithgartner/data-fabric-architecture-is-key-to-modernizing-data-management-and-integration>

Data Fabric - Definition nach Gartner

Key Pillars of a Comprehensive Data Fabric

■ Data Integration Layer ■ Knowledge Graph and Active Metadata Analysis ■ Data Catalog/ Metadata Layer



gartner.com

Source: Gartner
© 2021 Gartner, Inc. All rights reserved. CTMKT_1274755

Gartner

Data Fabric – Definition nach Forrester



- „Forrester defines enterprise data fabric as:
- Orchestrating disparate data sources intelligently and securely in a self-service manner, leveraging data platforms such as data lakes, data warehouses, NoSQL, translytical, and others to deliver a unified, trusted, and comprehensive view of customer and business data across the enterprise to support applications and insights.“
- Forrester Enterprise Data Fabric Enables DataOps Advanced Level: Data Practices For Insights-Driven Businesses, August 2, 2021

„Data Fabric“ – die (üblichen?) Versprechungen



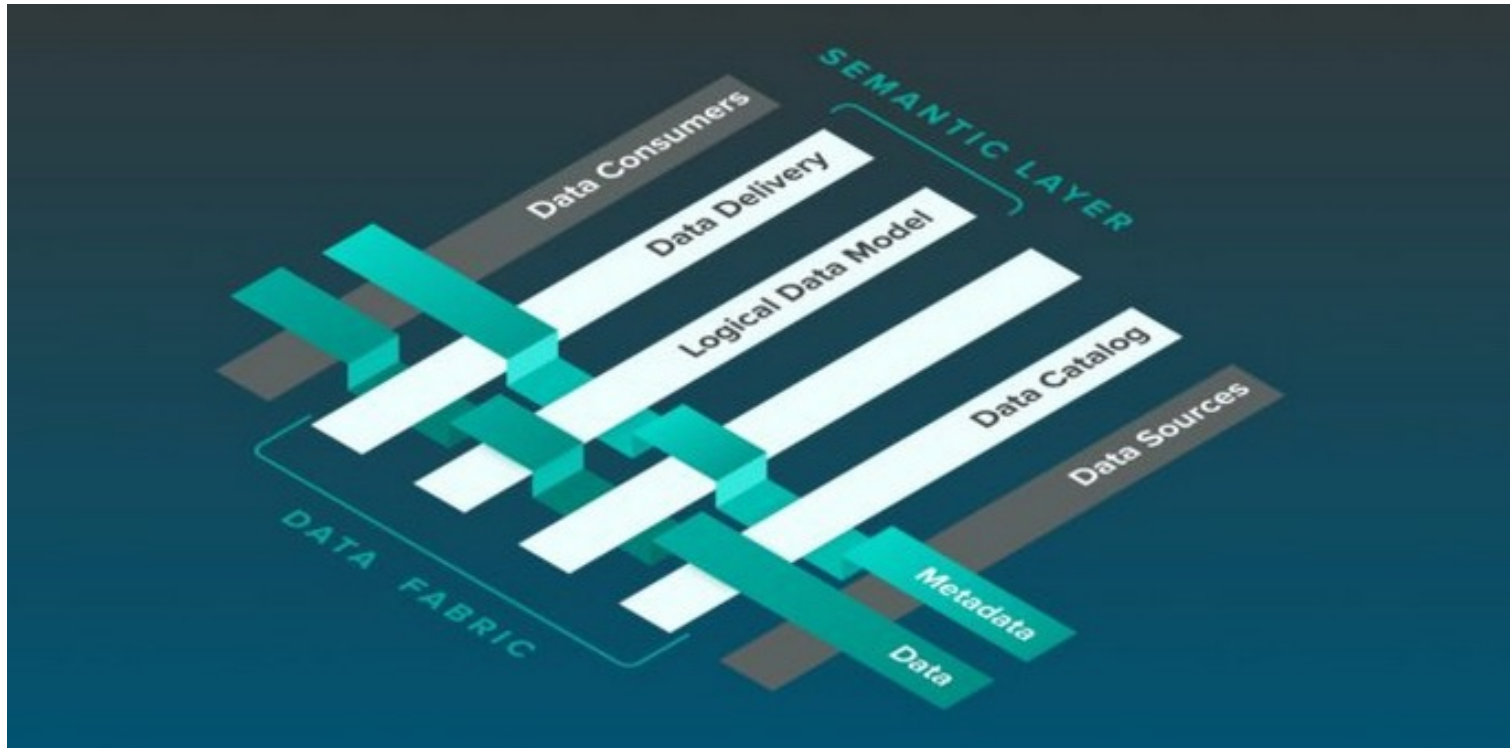
- Das Ende aller Daten-Silos
- Daten-Integration über alle Systeme hinweg – Clouds inklusive
- Kein Problem mit großen Datenmengen mehr
- Minimierung des Risikos von Verstößen gegen innerbetriebliche Vorgaben oder gesetzliche Vorschriften
- Schnell und einfach aufzubauen
- „Alles“ automatisch, weniger manueller Aufwand

Data Fabric - Produktmarkt nach Forrester



The Forrester Wave™ -
Enterprise Data Fabric Q2 2020

Data Fabric – ein Gewebe



Was soll Data Fabric bringen?



- Beschreibt eine Architektur, die die unternehmensweite Datenverwaltung und -bereitstellung umfasst
- Das Data-Fabric-Konzept ist entkoppelt von physischen Implementierungen und Infrastrukturen wie Clouds, eigenen Rechenzentren, IoT, lokalen Maschinen etc.
- Unabhängig vom Standort des Nutzers und der Daten liefert die Data Fabric eine konsistente und kohärente Sicht auf alle Daten, die der Nutzer benötigt, unter Berücksichtigung von Governance und Compliance

Was soll Data Fabric bringen?



- Dynamische Orchestrierung unterschiedlicher Datenquellen auf intelligente und sichere Weise im „Self-Service“-Verfahren
- Nutzung verschiedener Datenplattformen zur Bereitstellung integrierter und vertrauenswürdiger Daten zur Unterstützung verschiedener Anwendungen, Analysen und Anwendungsfälle
- Last but not Least: Aktives Metadatenmanagement auf Basis u.a. semantischer Wissensgraphen und AI / ML- Ansätzen

Data Fabric – auch als Integrationsarchitektur

Business Domains

Marketing

HR

Finance

Sales

Production

Services

Data Fabric

OLTP Databases /
Legacy Systems

Analytics/
BI/DWH

Data Lakes

Cloud Data
Stores

Repositoris, Docs
Applications u.a.

Data Fabric – ist (leider) nicht Hersteller-unabhängig



- Laut Gartner kann kein Hersteller alles vollständig
- Bekannte Player am Markt SAP, Denodo, Oracle, IBM, Talend, K2View
- Auch traditionelle Datenhaltungssysteme lassen sich integrieren
 - IBM z.B. verweist auf
 - IBM z-Mainframe in der Data Fabric
 - Anbindung über mehrere Integrationspunkte inkl. VSAM / ISAM

Data Fabric – Technologie-Komponenten

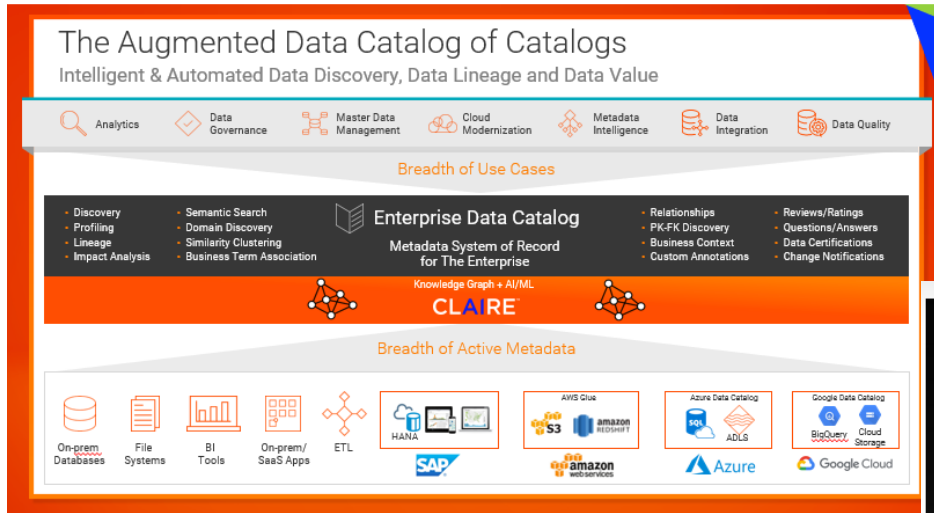


- **Data Virtualization** - Abstraktionsschicht für verteilte Ressourcen, einheitliche Ansicht, ob Cloud oder On-Premise, Option auf vordefinierte Datenansichten für grundlegende End-Nutzer - oder Experten-Anforderungen
- **Data Cataloging**– Verwendung von Metadaten beinhaltet das Profiling von Daten, auch Einsatz spezieller (ML)-Techniken zur Datenentdeckung und Datenklassifizierung
- **Knowledge Graph** – stellt selbstständig Beziehungen zwischen Entitäten her, die in verschiedene Datenmodellen und Daten entdeckt, mit Attributen versehen und in thematischen Kontext gebracht werden

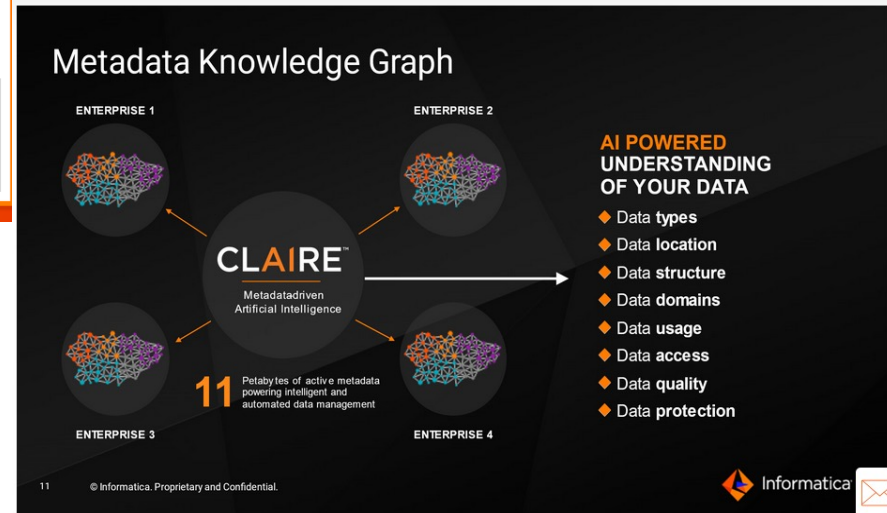
Data Fabric und Meta Daten

- Metadaten – Metainformationen / strukturierte Daten, die Informationen über die Merkmale anderer Daten enthalten.
- Metadaten Catalogue das Herzstück der Data Fabric
- Grundlage für die Beschreibung, Inventarisierung und das Verständnis von Daten für die diversen Anwendungsfälle
- Metadaten Catalogue dient z.B. dazu, Ort, Typ und Format der Daten ebenso aufzuzeigen wie die Beziehung der Datenobjekte zueinander.

Data Fabric – Meta-Daten und Knowledge Graph



<https://www.informatica.com/hk/resources/articles/data-fabric-the-transformative-next-step-in-data-management.html>



Data Fabric - Player



- Chief Data Officer
- Data Scientist
- Data Engineer
- Data Analyst
- Business Analyst
- Data Steward
- Stakeholder
- GRC Governance, Risk und Compliance

Wer definiert die Richtigkeit der Daten bei
Zusammenführung der Sicht über alle Daten?

Wer führt?

Gemeinsames Verständnis im Unternehmen?

Data Fabric Engineer – Beispiel einer Personal Anforderung

Die Position umfasst die Fähigkeit, komplexe technische Probleme anzugehen, einschließlich Solution Design & Build, Performance, Scalability und Deployment sowie Erfahrungen

- in der Arbeit mit und dem Aufbau von Big-Data-Lösungen, mit fundierten Kenntnissen über Data-Warehouse-Architekturen, Infrastrukturkomponenten, Data-Governance-Frameworks, Datenintegration, Migration und Business Analytics.
- mit relationalen / nicht relationalen Datenspeichern (Hadoop, Object Stores, Hadoop, Db2, Netezza, MongoDB, DataStax, EDB), Data Lake- und Data Fabric-Konzepten
- mit Cloud-Plattformen wie z.B. IBM Cloud, Amazon Web Services, Azure und Google Cloud
- Tiefes Verständnis von End-to-End-Lösungen und technischem Design
- mit ETL/ELT und Reporting-/Analyse-Visualisierungstools
- mit Skriptsprachen wie Shell, Python Expertise in relationalem SQL Expertise im Aufbau automatisierter Datenpipelines zum Bereinigen, Vorbereiten und Optimieren von Daten für die Aufnahme und Nutzung (Sicherheit) Erfahrung in der Arbeit mit verschiedenen Datenstrukturen und Algorithmen.

Data Fabric als neue „Single Source of Truth“?

- Zentrale Sicht auf alle Daten im Unternehmen
- Alle Business Domains im Unternehmen haben Zugriff auf die gleiche Datenqualität und Datenbasis
- Versorgung mit den richtigen Daten zum richtigen Zeitpunkt
- Problem von Daten-Dubletten lösbarer
- Abteilungsübergreifende Zusammenarbeit erforderlich – welche Daten und werden benötigt u.a.
- Compliance Richtlinien sind sorgfältig zu prüfen

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, DataWarehouse u.a. obsolet?
- Orchestrierung der Daten – DBMS, strukturierte/unstrukturierte u.a. per Data Fabric?

Data Fabric – weitere offene Punkte



- **Historie** – Realtime Verdichtung – was ist mit Historie, Speicherung von Transaktionsdaten?
- **Verteilte Daten – Standort der Daten**
echte Ad-hoc-Abfragen oder für die Verarbeitung von Analyse-/ML-Modellen, die Daten aus verstreuten Quellen (z. B. Sensoren am Rande des Unternehmens) benötigen, können die Daten nicht zwischengespeichert oder voraggregiert werden; sie werden bei Bedarf abgerufen - unabhängig von ihrem Standort. Dies kann zu erheblichen Latenzzeiten führen
- **Ergänzung und nicht als Ersatz** für Datenmanagement-Tools, -Praktiken und -Konzepte verstanden werden
- **Sicherheit** – Verschlüsselung, Benutzer und Rollen, User Login, Compliance

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