

# Neue Ansätze im Datenmanagement

**Know-how** Um grosse, verteilte Datenmengen zu beherrschen, sind in den letzten Jahren Daten-Architekturkonzepte wie Data Mesh neu erdacht worden. Was spricht für den Trend und Ansatz des Data Mesh – und was dagegen?

Von Stephan Meier

**G**rosse Unternehmen sind mit ständig wachsenden Datenmengen und einer stark steigenden Anzahl Datenanalysen konfrontiert. Auch sind einzelne Unternehmensbereiche teilweise so gross geworden, dass sie heute wie eine Firma in der Firma geführt werden. Die Anforderungen an die Datenaufbereitung übersteigen so oftmals die Kapazitäten von zentralen Data-Warehouse-Systemen (DWH) und Teams.

In grossen, global vertretenen Unternehmen ist es nicht ungewöhnlich, dass mehrere DWHs auf unterschiedlichen technischen Plattformen und mit unterschiedlichen Standards betrieben werden. Die Komplexität des Datenaustauschs ist hoch, der Koordinationsaufwand ist es ebenfalls. Diese Herausforderungen haben in den letzten Jahren zur Entwicklung von neuen Architekturkonzepten geführt, welche solche Probleme adressieren und dabei zentrale, IT-betriebene Data Warehouses ablösen oder ergänzen sollen.

Im Folgenden werden zwei der neuen Ansätze erläutert: Data Fabric und – vertiefter betrachtet – Data Mesh. Dabei wird unter anderem der Frage nachgegangen, ob ein Data-Mesh-Ansatz auch für mittelständische Unternehmen Sinn macht.

## Herausforderung Datenintegration

Die Datenintegration, also das Zusammenführen von Daten aus verschiedenen, innerhalb und ausserhalb des Unternehmens verteilten Datenquellen, war und ist seit jeher eine Herausforderung. Der Bedarf nach mehr Eigenständigkeit und mehr Daten in den Fachabteilungen ist dabei ein Spannungsfeld für Unterneh-

men, deren Data Warehouses zentral von der IT betrieben werden.

In diesem Sinne klassische DWH-Organisationen können mit den Bedürfnissen der einzelnen Fachbereiche kaum schritthalten. Sie sind im Normalfall konstant überlastet und können neue Anforderungen oft nicht in der gewünschten Zeit umsetzen. Ausserdem dürfen, zum Beispiel aus rechtlichen Gründen, nicht immer alle Daten und Auswertungen eines Fachbereichs oder eines Tochterunternehmens in ein zentrales DWH überführt werden. Deshalb wurden in der Vergangenheit auch schon Business-Intelligence (BI)-Kompetenzen in die Fachabteilungen ausgelagert, oder diese haben selbst die Beschaffung von BI-Tools in die Hand genommen.

Hersteller werben seit Jahrzehnten für ihre Lösungen mit der Unabhängigkeit der Fachanwender und der Befreiung aus den Zwängen der Abhängigkeit von den IT-Abteilungen. Dabei entstanden oft neue Probleme wie die des Tool-Wildwuchses, unkoordinierte lokale DWH-Initiativen oder die Datenhaltung in den BI-Frontend-Anwendungen.

## Data Fabric

Data Fabric ist eine Architektur, die in erster Linie versucht, das Problem des Zugriffs auf in Unternehmen verteilte Datenquellen auf technischem Weg zu lösen. Bei einer Data-Fabric-Architektur werden die Daten nicht in ein zentrales DWH transportiert – was bei einigen Datenquellen allein schon wegen der schieren Datenmenge unmöglich ist. Stattdessen werden Zugriffsmechanismen zur Verfügung gestellt, die es erlauben, Daten übergreifend aus verschiedenen Quellen zum Ab-

fragezeitpunkt zu verbinden. Diese Form der Integration wird im Fachjargon auch als virtuelle Datenintegration oder Datenvirtualisierung bezeichnet. Zusammen mit einem starken Metadaten-Management (Welche Dateninhalte befinden sich in welcher Quelle?) soll die zunehmende Anzahl von Quellen, Datensilos und Zugriffsmethoden den Benutzern als einheitliche Datenbasis präsentiert werden.

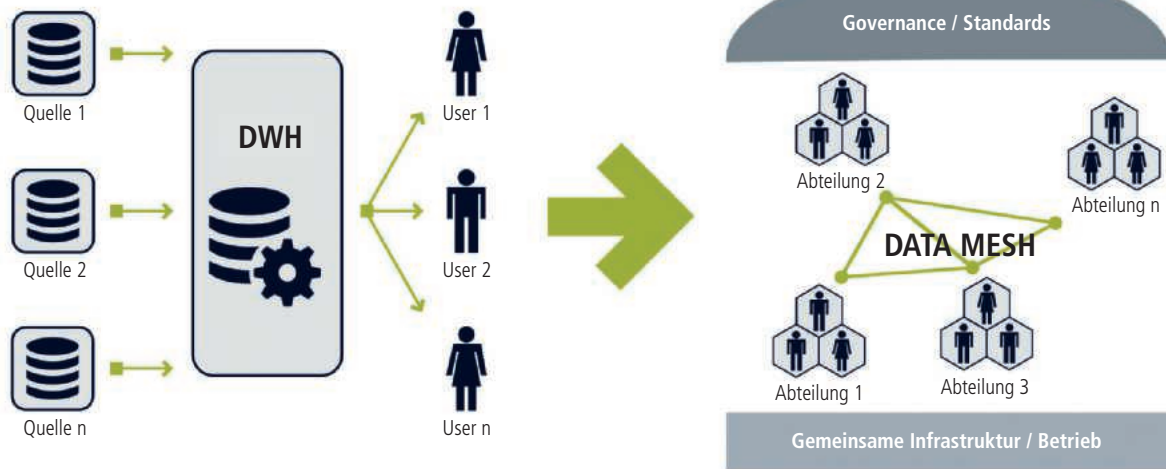
Dieser sehr technische Ansatz wird hauptsächlich von Tool-Anbietern getrieben. Das Versprechen, dank intelligenter Technologie keine aufwendige Datenaufarbeitung nach dem ETL-Prinzip (Extract, Transform, Load) betreiben zu müssen, geht aber mit diversen Problemen einher. So ist die Skalierbarkeit stark von den Tools abhängig und mit zunehmender Komplexität beschränkt.

Mit einer steigenden Anzahl Datenquellen und immer grösseren Datenmengen nehmen auch die Integrationsprobleme zu. Historisierung, Zugriffssteuerung und Transaktionsintegrität lassen sich oft nur über zusätzliche Mechanismen implementieren. Auch dürften sich Quellsystem-Owner (und Anwender) wenig über unvorhersehbare Lasten und damit verbundene Leistungsver schlechterungen auf operativen Systemen freuen. Besonders analytische Abfragen führen in auf operative Prozesse optimierten Systemen zu Performance-Problemen. Schleicht sich dann bei einer Query noch ein Fehler ein, kann ein Quellsystem auch mal lahmgelegt werden.

## Data Mesh

Das Konzept des Data Mesh wurde 2018 erstmals von Zhamak Dehghani, Director of Emerging Technologies des US-Unter-

## DATA WAREHOUSE VS. DATA MESH



Die Grafik zeigt die konzeptionellen Unterschiede zwischen einer klassischen Datawarehouse-Architektur und dem Data-Mesh-Konzept.

Quelle: IT-Logix

nehmens Thoughtworks, beschrieben und hat seither in der Datenindustrie grosse Aufmerksamkeit erlangt. Der Grundgedanke ist es, nicht nur die BI-Kompetenz, sondern auch die Datenaufbereitung und den BI-Applikationsbetrieb zu dezentralisieren und Service-orientiert zu betreiben. Anstelle eines von der IT zentral betriebenen DWHs, sollen die Daten von Entwicklern in den Fachabteilungen als Teil der dort angebotenen Produkte und Dienstleistungen angesehen werden. Die Datenlieferungen und Datenservices, deren Qualität und die Dokumentation wechseln damit in den Verantwortungsbereich der jeweiligen Fachbereiche.

Data Mesh basiert auf den folgenden vier Prinzipien:

- ▶ Fachbereichsbezogene (Domänen-) Daten-Ownership (und -Architektur)
- ▶ Daten als Produkt
- ▶ Selbstbedienungsinfrastruktur als Plattform
- ▶ Zentral verwaltete computergestützte Governance

Wie erwähnt, wurden Data-Mesh-Ansätze zuerst vorwiegend in sehr grossen Unternehmen angewendet. Bekannte Beispiele sind etwa BMW, Zalando oder die ABN Amro Bank. Ihnen allen gemeinsam ist der Umgang mit sehr grossen, global verteilten Datenmengen (Big Data). Weiter sind diese Unternehmen häufig bereits sehr datengetrieben und datenorganisiert. Ihre Geschäftsmodelle sind geprägt von sich kurzfristig

ändernden und neu zu entwickelnden Geschäftsfeldern. Zudem ist eine agile Applikationsentwicklung mit kombiniertem Betrieb dieser Applikationen (Stichwort Devops) oft bereits stark verbreitet.

Zumindest einzelne Aspekte von Data Mesh können aber auch für mittelständische Unternehmen sinnvoll anzuwenden sein. Selbst aus Sicht eines zentralen DWH-Betreibers und auch im Kontext der Data Awareness eines Unternehmens ist es wünschenswert, wenn Fachbereiche mehr Verantwortung für ihre Daten und Datenanalyse übernehmen. Das Fach soll also seine eigenen Produkte und Services verwalten («Eat your own Dogfood»).

Ein Grund, warum Data-Mesh-Aspekte neben der Verlagerung der Datenverantwortung (Ownership) auch für mittelständische Unternehmen relevant sein können, sind Daten, die aus unterschiedlichen Gründen (regulatorischen, technischen) nicht in ein zentrales DWH überführt werden können, von denen aber Teilmengen für andere Firmenbereiche durchaus sehr nützlich sein können.

### Technisch möglich, in der Praxis nicht einfach

Es gibt verschiedene Formen, wie ein Data-Mesh-Ansatz in einem Unternehmen umgesetzt werden kann. Zuerst sei der ursprüngliche Anwendungsfall erwähnt, welcher vor allem für Grosskonzerne relevant ist: Die komplette Dezentralisierung der Datenaufbereitung und -haltung. Hier haben die Fachbereiche

(oder Tochterunternehmen) eigene lokale Standards, Prozesse und Infrastruktur. Der Data-Mesh-Ansatz impliziert in so einem Umfeld aber paradoxerweise den Aufbau einer starken, zentralen Governance. Diese regelt vor allem Aspekte des Austauschs und der Dokumentation von Daten zwischen den dezentralen Daten-systemen.

Ein Data Mesh bedingt die Delegation von Kompetenzen, was in Unternehmen nicht immer einfach umsetzbar ist. Die Ownership muss neu auf Seite der Fachabteilungen gelebt werden, was dort einen nicht unerheblichen Zusatzaufwand bedeutet. Denn es sind einige Aufgaben zu erfüllen: Datenaufbereitung, Dokumentation, Sicherstellung der Datenqualität, Security und nicht zuletzt die Erstellung von Datenprodukten, die von anderen Anwendern genutzt werden können. Dafür werden in den Fachabteilungen Mitarbeitende benötigt, die Daten aufbereiten, analysieren und ihr Domänenwissen einbringen können. Werden dabei noch unterschiedliche, technische Plattformen eingesetzt, benötigt jede weitere Plattform Spezialistenwissen und erhöht die Gesamtkomplexität der Infrastruktur. Entsprechende Fachkräfte sind heute schon sehr gesucht.

Eine für mittelständische Unternehmen angepasste Umsetzung des Data-Mesh-Gedankens umfasst daher eine allen Fachbereichen gemeinsame technische Basis. Diese erlaubt es, zusätzlich auch Entwicklungsstandards und eventuell auch

den technischen Betrieb zentral zu kontrollieren, während die (Weiter-)Entwicklung der Datenprodukte durch die Fachbereiche erfolgt. Der Vorteil einer solchen Lösung ist, dass das technische Know-how auf eine Plattform beschränkt werden kann und Expertenwissen zwischen den Fachbereichen austauschbar wird.

### Vereinfachung durch Automatisierungs-Tools

Eine weitere Möglichkeit, die Datenaufbereitung in den Fachbereichen zu vereinfachen, ist der Einsatz von DWH-Automatisierungs-Tools. Durch die Generierung von Code aus Metadaten (die technische Dokumentation) können Lade Strecken und Transformationen in einer sehr standardisierten Form erstellt werden, Entwickler werden produktiver und es ist auch möglich, dass Datenexperten ohne grosse Programmiererfahrung beim Aufbau einer Fachbereichsdatenumgebung (Domäne) mitarbeiten. Ein erwünschter Nebeneffekt eines solchen Vorgehens ist, dass im Rahmen der Entwicklung gleichzeitig die Dokumentation erstellt wird, die Einarbeitungszeit neuer Mitarbeiter verkürzt wird und Umsetzungsstandards, die in der Governance definiert sind, erzwungen werden. Diese Standards, Zugriffskontrolle, Datenschutz, Datenkataloge werden zentral verwaltet und vorgegeben.

Eine der grössten Hürden für ein Unternehmen dürfte es sein, Datenprodukte und Service Level Agreements (SLAs) zu entwickeln, deren Anforderungen sich

potenziell schnell anpassen lassen, auch durch Anfragen anderer Fachbereiche. Genaue Spezifikationen funktionieren hier für gewöhnlich nicht, denn eine Auswertung oder ein Dashboard ist selten nach der ersten Iteration fertig. Diese Anpassungsaufwände werden von den einzelnen Fachbereichen getragen – etwas, was früher einfach der IT oder dem DWH-Team übergeben wurde. Für den BI-Anwender sollte sich ein Data Mesh im Vergleich zu einem zentralen DWH gleich anfühlen. Lediglich seine Ansprechpartner wechseln: Daten werden nicht mehr von der IT, sondern von den Fachabteilungen geliefert.

### Kein Produkt, das man kaufen kann

Data Mesh ist vor allem ein Organisations- und Datenkulturthema. Es ist kein Produkt, das man kaufen kann. Wenn die Ownership für die Daten in die Fachbereiche verlagert wird, steigert dies das Bewusstsein für die eigenen Daten und den Aufwand für deren Pflege, der sonst von den zentralen Stellen für die Bereiche geleistet wird. Damit wird die Datenkompetenz insgesamt gesteigert und das Thema in der Firma über ein reines Bedürfnis nach Analysen hinaus tiefer verankert. Ein datengetriebenes Unternehmen und datengetriebene Entscheidungen bleiben nicht länger reine Schlagwörter.

Durch DWH-Automation als Basis der Entwicklung wird auch der Dokumentationsstand deutlich verbessert. Die dafür erfassten Metadaten dienen nicht nur als

Basis für Ladeprozesse, sondern können beispielsweise auch im Rahmen von Datenschutzgesetzprojekten verwendet werden. Schliesslich fördert ein Data-Mesh-Ansatz die Wiederintegration von unabhängig betriebenen Analyseplattformen. Aber längst nicht jedes Unternehmen entwickelt oder betreibt mehrere Datawarehouses oder hat Quantitäts- oder Speed-Probleme, die sich nicht durch Skalierung oder Automatisierung der bestehenden DWH-Architektur lösen liessen. Hier gilt es im konkreten Fall, zu prüfen, welche Kompetenzen gezielt in den Fachbereichen des Unternehmens aufgebaut und dorthin delegiert werden sollen und welche Aufgaben weiterhin bei einer zentralen IT- oder BI-Organisation verbleiben. ■

### DER AUTOR

Stephan Meier ist Senior Consultant Data & Analytics beim auf Business Intelligence, Data Warehousing, Data Science und Big Data spezialisierten IT-Beratungshaus IT-Logix. Als ausgewiesener Data- und BI-Fachmann in Consulting- wie auch Managementrollen unterstützt er Firmen bei der Erarbeitung von Datenstrategien, beim Aufbau von Data-Teams sowie bei der Etablierung von datengetriebenen Prozessen, Strukturen und Architekturen.



# Rezepte für die Digitalisierung

IT-Strategien & Praxis in Schweizer Unternehmen

Interviews | Fallstudien | Know-how

Jeden Monat in **Swiss IT Magazine**

Kostenloses Probe-Abonnement unter [www.itmagazine.ch/abo](http://www.itmagazine.ch/abo)