

Mehr Überblick, geringere Komplexität

Daten sind der Stoff, aus dem die Zukunft ist. Sie bestmöglich zu nutzen, verspricht das Data-Lake-Konzept. Allerdings lassen sich die Anforderungen der Fachabteilungen mit diesem zentralistischen Modell kaum erfüllen. Ein deutscher ITK-Provider baute deshalb ein Data-Mesh auf, das zentrale Governance mit dezentralem Datenmanagement verbindet. VON MARK MICHEL

SECHS DATA WAREHOUSES, realisiert mit drei unterschiedlichen Technologien – so stellte sich die Dateninfrastruktur bei dem deutschen ITK-Provider im Jahr 2020 dar. Um die Verfügbarkeit und die Nutzung der Daten zu erhöhen – und damit auch ihre Wirtschaftlichkeit – stellt das Unternehmen, auf eine Data-Lake-Infrastruktur um. Im Zuge der Einführung neuer Systeme wird der zentrale Data Lake verwendet, um den Aufwand für Schnittstellen zu minimieren: Relevante Daten werden nur einmalig in den Data Lake geschrieben und können dort administriert und weiterverarbeitet werden. Der erste Ansatz sah vor, den zentralen Datenpool auch zur Neugestaltung einer analytischen Landschaft zu nutzen. Das Ziel: Eine Unternehmenskultur, die Daten demokratisiert, um das darin enthaltene Potential effektiver zu nutzen.

Herausforderung Data Lake House

Die Neugestaltung der analytischen Infrastruktur erfolgte demnach weitgehend mit Open-Source Software in einer „Data Lake House“ Architektur. Hierbei werden die Rohdaten in einen Data Lake gespeichert, der zentral administriert und verwaltet wird. Anschließend können Daten zur Weiterverarbeitung in eines oder mehrere der fachspezifischen Data Warehouses geladen werden.

Das Data Lake House funktioniert reibungslos im Zusammenspiel mit dem Omnichannel-Contact-Center Genesys des Customer Care Bereichs. Dort lassen sich die Daten einfach durch Setzen eines Häkchens via Postgres SQL im- und exportieren. Andere Fachbereiche standen jedoch vor der Herausforderung, ihre bestehenden Systeme in die neue Landschaft zu überführen oder zu integrieren. Finance beispielsweise nutzt ein MS-SQL-Server-basiertes Data Warehouse. Mit erheblichem Aufwand erstellt und von hoch spezialisierten Fachkräften gepflegt, versorgt es alle anderen Domänen mit einer Vielzahl von Daten. Der Umzug auf eine andere Technologie hätte kurzfristig nur hohen Aufwand verursacht, ohne entsprechenden Mehrwert.

Fachbereiche in der Verantwortung

Um auch ohne technische Integration eine unternehmensweite Datennutzung zu erreichen, wurde der innovative Data-Mesh-Ansatz gewählt. Data Mesh ist ein Organisationsmodell für eine dezentrale Datenarchitektur, das Zhamak Dehghani 2019 erstmals vorstellte. Die wichtigsten Bestandteile des Modells sind

- Domain-Ownership,
- Data as a Product,
- Self-Service-Plattform und
- Federated Governance.

Vereinfachend zusammengefasst bedeutet Domain-Ownership: die interdisziplinär besetzten Datenmanagement-Teams der Fachbereiche (Domänen) übernehmen die Verantwortung dafür, dass ihr Fachbereich über alle benötigten Daten jederzeit aktuell in der benötigten Qualität verfügt. Nach dem Prinzip Data as a Product behan-

Der AUTOR Mark Michel

ist Senior Business Consultant Data & Intelligence bei NTT DATA DACH

den sie die Daten dabei immer so, dass sie auch anderen Domänen als Produkte zur Verfügung stehen. Dazu reichern sie ihre Datenprodukte um Metadaten an, die den Inhalt, die Quelle und die Verwendungsmöglichkeiten der Daten so beschreiben, dass sie auch für andere Nutzer verwertbar sind. Auf der Self-Service-Plattform können diese in einem Data Katalog mithilfe einer Suchmaschine gezielt recherchieren. Das Prinzip der Federated Governance sorgt dafür, dass die Kriterien zur Definition der Datenprodukte und zu ihrer domänenübergreifenden Nutzung in Abstimmung der Domänen untereinander einheitlich geregelt werden.

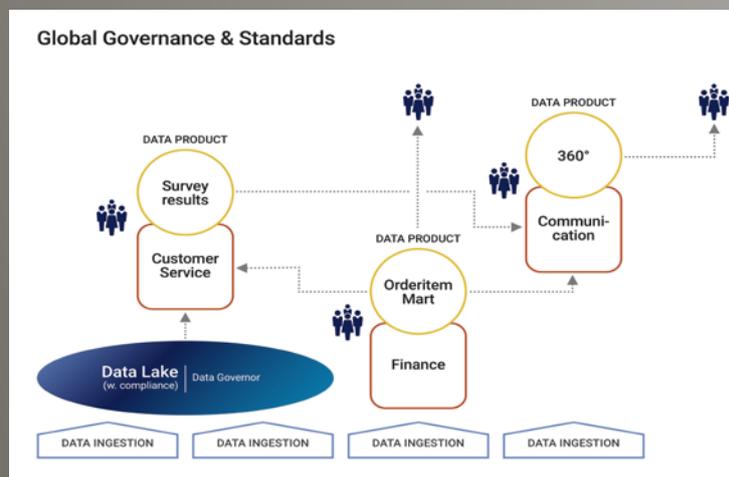
Fokus auf Datenprodukte

Kernbestandteile der Data-Mesh-Infrastruktur des ITK-Providers sind die Datenprodukte (Data Products). Hierzu ein Beispiel: Der Finanzbereich stellt den anderen Domänen das offizielle Vertriebsberichtswesen zur Verfügung. Das Datenprodukt „order item mart“ liefert detaillierte Informationen zu Produkt, Kanal, Organisation und relevanten Kennzahlen. Verschiedene Zuordnungen werden dabei exklusiv vergeben und dann unternehmensweit einheitlich verwendet. Das Datenprodukt wird in verschiedenen analytischen Anwendungen weiterverwendet, aber auch für Self-Service-Auswertungen per Plug-In und als Dashboard bereitgestellt.

Die Notwendigkeit eines solchen strukturierten Datenaustauschs wird anhand der Customer Journey deutlich: Verschiedene Stakeholder mit unterschiedlichen Analysebedarfen und Lösungen verfügen über Daten, die in Nachbarbereichen genutzt werden. Der bereits vorhandene Datenaustausch, teilweise über Systemgrenzen hinweg, muss nicht über einen zentralen Data Lake laufen, wenn per Data Mesh Verantwortlichkeiten, Struktur und Inhalte der für den Austausch verwendeten Datenprodukte geklärt werden.

Data Governor sichert Qualität

Erfolgsentscheidend für den Geschäftsprozess ist – nicht nur bei der Customer Journey – die Einhaltung der generellen Vorgaben für die Ablage von Rohdaten im Data Lake und Datenprodukte allgemein. Denn im laufenden Prozess führen unvollständige oder falsche Daten zu Fehlern und schlimmstenfalls zu unzufriedenen Kunden. Die Prüfung der Datenprodukte vor ihrer Ablage im Data Lake und der Aufnahme in den Data Catalog erfolgt im so genannten Data Governor. Dieses von der Unternehmens-IT selbst entwickelte Tool übernimmt Plausibilitäts- und Qualitätsprüfung sowie einfache Konvertierungsaufgaben. Beispielsweise werden alle UTC-Zeitstempel automatisch auf die deutsche Zeit umgerechnet, um Verwechslungen auszuschließen. Der Data Governor ist eine Komponente des Data Lakes. Hier haben die zentral agierenden Data Engineers die Verantwortung, den Domain Teams die Werkzeuge zu liefern, um ihre Datenprodukte zu managen. •



Zentrale Data Governance und dezentrale Verantwortung für einzelne Datenprodukte bilden die Ordnungsprinzipien im Data Mesh.