



# DATA FABRIC

**Digitalisierung zum Self Service 2.0**

**E-Book**

#### Herausgeber

SIGS DATACOM GmbH  
Lindlaustraße 2c  
53842 Troisdorf

[info@sigs-datacom.de](mailto:info@sigs-datacom.de)  
[www.sigs-datacom.de](http://www.sigs-datacom.de)

Copyright © 2021 SIGS DATACOM GmbH  
Lindlaustr. 2c  
53842 Troisdorf

Die vorliegende Publikation ist urheberrechtlich geschützt. Alle Rechte vorbehalten. Die Verwendung der Texte und Abbildungen, auch auszugsweise, ist ohne die schriftliche Zustimmung des Herausgebers urheberrechtswidrig und daher strafbar. Dies gilt insbesondere für die Vervielfältigung, Übersetzung oder die Verwendung in elektronischen Systemen.

Es wird darauf hingewiesen, dass die in der Broschüre verwendeten Soft- und Hardware-Bezeichnungen sowie Markennamen und Produktbezeichnungen der jeweiligen Firmen im Allgemeinen warenzeichen-, marken- oder patentrechtlichem Schutz unterliegen. Alle Angaben und Programme in dieser Broschüre wurden mit größter Sorgfalt kontrolliert. Weder Autor noch Herausgeber können jedoch für Schäden haftbar gemacht werden, die im Zusammenhang mit der Verwendung dieser Broschüre stehen.

Wo nicht anders angegeben, wurde auf die im Text verlinkten Quellen zurückgegriffen.

TDWI E-Book in Kooperation mit



**OPITZ CONSULTING**

<b>1</b>	<b>Einleitung</b>	<b>5</b>
<b>2</b>	<b>Reflektion der Digitalisierung – die Data Fabric im Kontext der Digitalisierung und des Self Service</b>	<b>6</b>
<b>3</b>	<b>Herausforderungen im Kontext der Datenbereitstellung</b>	<b>7</b>
3.1	Big-Data-Analytics	8
3.2	Self-Service-Analytics	9
3.2.1	Von der einfachen Ergebnisnutzung zur individuellen Ergebnisgestaltung	9
3.2.2	Wie weit soll ein Self Service gehen?	10
<b>4</b>	<b>Data Warehouse und Data Lakehouse – Basis der Data Fabric</b>	<b>11</b>
4.1	Zeitgenössische Architektur	11
4.2	Data Warehouse	12
4.3	Data Lake	13
4.4	Data Lakehouse	13
4.5	Konsequenzen für eine Data Fabric	15
4.5.1	Eine Data Fabric muss die Möglichkeit haben, alle Formen von Metadaten zu sammeln und zu analysieren	16
4.5.2	Eine Data Fabric muss die Möglichkeit haben, passive Metadaten zu analysieren und in aktive Metadaten zu konvertieren	16
4.5.3	Data Fabric muss die Möglichkeit haben, ein Wissensdiagramm zu erstellen, welches das Data-Fabric-Design operationalisieren kann	17
4.5.4	Data Fabric muss es Mitarbeitern im Fachbereich ermöglichen, Datenmodelle mit Semantik anzureichern	18
4.5.5	Empfehlungen für Datenanalyseverantwortliche	18
<b>5</b>	<b>TDWI Research: Sechs Fähigkeiten einer logischen Data Fabric</b>	<b>19</b>
5.1	Integration von Daten in Multicloud-Umgebungen	19
5.2	Automation manueller Aufgaben mit ergänzender Analysefähigkeit	20
5.3	Steigerung der Analytics-Performance durch eine schnelle Datenbereitstellung	20
5.4	Unterstützung der Datenentdeckung und Datenauswertung sowie von Data-Science-Aktivitäten	21
5.5	Analyse ruhender und dynamischer Daten	22
5.6	Katalogisierung aller Daten zur Discovery, Lineage und Verknüpfung	23
<b>6</b>	<b>Fazit</b>	<b>24</b>
	<b>Literatur</b>	<b>26</b>
	<b>Über unseren Sponsor</b>	<b>27</b>

## Vorwort

Die Verfügbarkeit und Wirtschaftlichkeit der Nutzung von Commodity-Hardware und -Software zur Entwicklung flexibler und skalierbarer Hochleistungsumgebungen hat in den letzten Jahren zugenommen. Das hat die Methoden und Prozesse für Reporting, Business Intelligence und Analytics verändert. Da Unternehmen sich für Cloud-Computing-Plattformen entscheiden und ihre Daten und Anwendungen in eine hybride Cloud-Umgebung migrieren, können sie mehrere Plattformen nutzen, um neue und wachsende Datentypen für Analysen zu unterstützen. Die Datenvirtualisierung, eine Kernkomponente einer Data Fabric, spielt eine wichtige Rolle bei der Unterstützung des Zugriffs auf Daten sowie deren Verwaltung und Analyse – und zwar über unterschiedliche Plattformen für ein traditionelles Berichtswesen und Business Intelligence. Dies gilt auch für moderne Anwendungsfälle wie maschinelles Lernen und künstliche Intelligenz, integrierte Analysen zugunsten einer automatisierten Entscheidungsfindung oder eine Analyse der Kombination ruhender Daten mit Echtzeit-Streaming-Daten. Das vorliegende eBook stellt das Konzept der Data Fabric im Spiegel seiner Anforderungen vor und geht darauf ein, wie Data-Science-Initiativen durch Datenermittlung und Analyse unterstützt werden können, um den Wert des Datenbewusstseins und der Verwendung einer Data Fabric als Unternehmensdaten-katalog herauszustellen.

# 1 Einleitung

Die Verfügbarkeit und Wirtschaftlichkeit der Nutzung von Commodity-Hardware und -Software zur Entwicklung flexibler und skalierbarer Hochleistungs-umgebungen hat in den letzten Jahren zugenommen.

Das hat die Methoden und Prozesse für Reporting, Business Intelligence und Analytics verändert. Da Unternehmen sich zunehmend für Cloud-Computing-Plattformen entscheiden und ihre Daten und Anwendungen in eine hybride Cloud-Umgebung migrieren, können sie mehrere Plattformen parallel nutzen, um neue und wachsende Datentypen für Analysen zu unterstützen.

Ohne bewährte Verfahren und Organisation für ihre hybride Datenumgebung besteht allerdings das Risiko erhöhter Komplexität, wenn Datenverbraucher nach den Datenressourcen suchen, die sie für ihre Arbeit benötigen, und darauf zugreifen. Eine Möglichkeit, Daten in dieser verteilten Umgebung zu verwalten und für Analysen zu verwenden, besteht in der Anwendung eines modernen Ansatzes im Sinne einer logischen Datenstruktur, um unterschiedliche Daten zu verfeinern und auf intelligente Weise zusammenzuführen.

Die Datenvirtualisierung, die eine Kernkomponente der logischen Datenstruktur ist, kann eine wichtige Rolle bei der Unterstützung des Zugriffs auf Daten über unterschiedliche Plattformen für herkömmliche Berichte und Business Intelligence (BI) hinweg sowie deren Verwaltung und die Analyse spielen. Das gilt auch für aktuelle Aufgaben wie Maschinelles Lernen und Artificial Intelligence, integrierte Analysen für automatisierte Entscheidungsfindung und Analysen, die ruhende Daten mit Echtzeit-Streaming-Datenquellen kombinieren.

Im Rahmen dieses eBooks werden sechs relevante Funktionen der logischen Datenstruktur beschrieben,

die sich mit modernen Datenmanagement- und Analysebemühungen beschäftigen und mehrere Plattformen, neue Datentypen und -quellen sowie erweiterte Analysen umfassen. Zu diesen Funktionen gehört die Aktivierung von Datenanalysten in der neuen hybriden Multicloud-Datenlandschaft. Dazu kommen umfassende Techniken zur nahtlosen Integration von Daten aus Multicloud-Plattformen. Diese zeigen, wie erweiterte Analysefähigkeit das Datenbewusstsein und die nahtlose Zugänglichkeit bei gleichzeitiger Aufrechterhaltung einer hohen betrieblichen Leistung erleichtert und wie die Leistungssteigerung durch Optimierungen die Verzögerungen im Zusammenhang mit der Datenlatenz reduziert. Darüber hinaus wird erläutert, wie Data-Science-Initiativen durch Datenermittlung und Analytics unterstützt werden können, die ruhende und bewegte Daten kombinieren. Schließlich zeigen die diskutierten Kriterien den Wert des Datenbewusstseins und der Verwendung einer Data Fabric als Unternehmensdatenkatalog auf:

- Datenbereitstellung wird zum Schlüssel, um eine Entkopplung zwischen Datenbereitstellung und Datennutzung zu schaffen.
- Identifikation von Daten zu identifizierten Themen der Digitalisierung.
- Verständnis der Data Fabric als Voraussetzung für die Funktionsfähigkeit des digitalisierten Unternehmens, bei dem eine Dezentralisierung der Datenverwendung mit einhergeht.

Das vorliegende eBook wird im zweiten Kapitel zunächst das Konzept der Data Fabric einführen. Kapitel 3 geht auf das Anwendungsszenario des Self Service ein, das als initiales Szenario für eine Data Fabric gilt. Im vierten Kapitel werden Konsequenzen aus den fachlichen und technischen Anforderungen an eine Data Fabric diskutiert und in Kapitel 5 erfolgt die Präsentation der Ergebnisse von TDWI Research USA zum Thema.