



## Sicher unterwegs in Data Lake und Cloud

DSGVO-konformes DWH-Offloading mit dem Azure Governed Data Lab von Microsoft

Das Inkrafttreten der Europäischen Datenschutzgrundverordnung (EU-DSGVO) stellt Unternehmen jetzt und in Zukunft vor enorme Herausforderungen. Die Umsetzung der Richtlinien stellt einen regelrechten Kraftakt dar, bei dem sowohl fachliche Themen als auch technische Innovationen häufig das Nachsehen haben. Wie kann eine schnelle Umsetzung der Datenschutzkonformität erreicht werden? Wie kann man einem möglichen Innovationsmangel entgegenwirken? Und wie wird gleichzeitig der Weg in die Cloud geebnet? Finden Sie jetzt heraus, wie Sie diese Herausforderungen mit dem Azure Governed Data Lab meistern können.

### Datenschutz vs. Innovation

Die Umsetzung der Regelungen der neuen Europäischen Datenschutz-Grundverordnung (EU-DSGVO) wurde bei vielen Unternehmen mit Hochdruck vorangetrieben. Der Fokus auf die Einhaltung der Rechtskonformität führt häufig dazu, dass die Umsetzung fachlicher Anforderungen in den Hintergrund gerät.

Auch innovative Themen wie die Modernisierung der bestehenden Informationslandschaft sowie die Erschließung weiterer relevanter Bereiche wie Big Data, Advanced Analytics, Automatisierung etc. standen gerade in der jüngeren Vergangenheit in der Priorisierung nicht an vorderster Stelle.

Auch nach Inkrafttreten der DSGVO und bereits implementierten Umsetzungen wird der Prozess der Anforderungsumsetzung tendenziell aufwendiger, da die entsprechenden Governance-Richtlinien eingehalten werden müssen. Die Gefahr einer längeren Time-to-Market wächst und droht, weitere Innovationen zu blockieren.

Der Spagat zwischen der Berücksichtigung regulatorischer Anforderungen und dem Vorantreiben innovativer Themen stellt eine große Herausforderung dar. Wir zeigen Ihnen hier einen Weg, der es ermöglicht, beides miteinander zu vereinen.

### DWH-Offloading und Data Lake in der Cloud

Ein Szenario hinsichtlich der DWH-Modernisierung stellt das DWH Offloading dar. Beim DWH Offloading werden die Daten aus dem bestehenden DWH in einen Data Lake transferiert. In dieser kostengünstigen Speicheralternative können Daten aus dem DWH problemlos gesichert und etablierte und qualitätsgesicherte ETL-Prozesse beibehalten werden.

Setzt man das Offloading zusätzlich noch innerhalb einer Cloud-Architektur um, entfallen etwaige Investitionskosten in eine geeignete On-Premises-Architektur. Bei dem Transfer in den Data Lake ist zu beachten, dass ausschließlich datenschutzkonforme Daten überführt werden.

Durch die Nutzung eines Data Lakes stoßen Sie das Tor zu weiteren Analysemöglichkeiten auf. So wird es möglich, weitere Datenquellen, auch unstrukturierte Daten, in den Datenbestand aufzunehmen, und diese mit qualitätsgesicherten Daten aus dem etablierten DWH zu verknüpfen, um neue Erkenntnisse aus den alten und neuen Daten zu ziehen.

## Das Azure Governed Data Lab

Insbesondere wenn es darum geht, ein Offloading-Szenario schnell umzusetzen und dabei datenschutzkonform zu bleiben, hilft eine vorgefertigte und standardisierte Lösung, die per One-Click-Provisionierung sowohl die Plattform als auch die notwendigen Anonymisierungs- bzw. Pseudonymisierungsroutinen bereitstellt. Diese Funktionalität bietet das Azure Governed Data Lab von Microsoft.

Hierzu wurde innerhalb der Microsoft Azure Cloud folgendes Standardszenario umgesetzt:

- Das bereits on-premises existierende Data Warehouse wird auf personenbezogene Daten hin analysiert. Die personenbezogenen Daten werden im nächsten Schritt on-premises anonymisiert bzw. pseudonymisiert.
- Der nächste Schritt ist die Provisionierung des Data-Lake-Setups. Hierzu wurde eine Referenzarchitektur umgesetzt (siehe Abbildung). Das Herzstück stellt der Azure Data Lake Store dar, in dem die Daten schlussendlich gespeichert werden. Um einen wohlorchestrierten Abzug der On-Premises-Daten durchzuführen, wird die Azure Data Factory eingesetzt, die eine Vielzahl von Verbindungen zu unterschiedlichsten lokalen Datenbanken herstellen kann. Diese Architektur wird in Form von Automatisierungsskripten umgesetzt und kann auf Abruf innerhalb weniger Minuten deploy werden.
- Für das Reporting auf den nun pseudonymisierten Daten in der Azure Cloud wird konsequenterweise ein Cloud-basiertes Reporting Tool eingesetzt. Im Fall des Azure Governed Data Lab wurde dies mit Power BI umgesetzt, das auch in Azure zur Provisionierung bereitsteht.

Durch die Einrichtung des Data Lake Storages besteht nun die Möglichkeit, weitere unterschiedliche Datenquellen per Schema-on-Read-Prinzip in die Landing Zone des Data Lake Stores zu laden, um diese Daten dann weiter zu verwenden. Die Lösung verbindet somit zwei verschiedene Modernisierungsvorhaben: den Aufbau einer Data-Lake-Architektur und den Schritt in die Cloud.

Überzeugen Sie sich von den Möglichkeiten, die Ihnen diese hochgradig standardisierte Lösung in Microsoft Azure für Ihre aktuellen und zukünftigen Analytics-Anforderungen bietet.

### Unser Angebot auf einen Blick

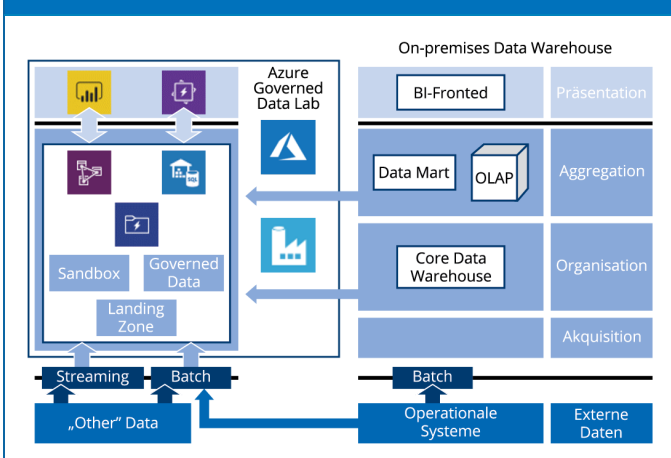
#### So helfen wir Ihnen

- Schneller Aufbau eines Data Lakes mit Microsoft Azure
- Unterstützung bei den ersten Schritten in Richtung Cloud Analytics
- Bereitstellung von Anonymisierungs- und Pseudonymisierungsroutinen
- Empfehlung der zu Ihnen passenden Azure Komponenten

#### Ihre Vorteile

- Datenschutzkonformität
- Risikominimierung durch Cloud-Ansatz
- Skalierbare Plattform
- Schnelle Ergebnisse

### Architekturskizze



#### Sprechen Sie uns an:



Jürgen Fischer  
Consulting Manager

Kontakt:  
+49 2261 6001-1122  
[juergen.fischer@opitz-consulting.com](mailto:juergen.fischer@opitz-consulting.com)

Unsere Leistungen im Bereich BI, Analytics und Big Data:  
[www.opitz-consulting.com/bi-analytics-big-data](http://www.opitz-consulting.com/bi-analytics-big-data)

