



# Fünf Mythen über den Data Lake

Sobald Unternehmen den Nutzen und die Relevanz ihrer Daten für neue geschäftliche Chancen und Umsatzmodelle erkennen, beschäftigt sie eine zentrale Frage: Welche der unendlich vielen Technologien ist am besten geeignet, um alle Daten zu verwalten und wertvolle Erkenntnisse zu gewinnen?

Beim Data Lake handelt es sich um eine Referenzarchitektur, die momentan aus gutem Grund im Fokus steht. Doch es gibt eine Reihe von Mythen über den Data Lake, die Unternehmen und IT-Manager wertvolle Zeit bei der Suche nach der richtigen Lösung kosten.

## 1.

### **Der Data Lake kann nur in einer einzelnen Cloud bereitgestellt werden → Falsch**

Der Data Lake ist nicht auf einen bestimmten Standort oder eine einzelne Cloud beschränkt. Er ist nicht auf die Vor-Ort-Bereitstellung beschränkt. Sie können einen Data Lake über mehrere Clouds und mit Hybridoptionen anlegen. Ein Data Lake ist eine technologieunabhängige Referenzarchitektur. Mit diesem Ansatz rücken Unternehmen Daten ins Zentrum ihrer betrieblichen Abläufe, um den Zugriff auf verschiedenste Datentypen in großem Umfang zu ermöglichen und Benutzern Self-Service-Analysen zu erlauben.

## 2.

### **Hadoop ist der einzige Data Lake → Falsch**

Auch wenn der Begriff Data Lake häufig mit Hadoop oder Hadoop-orientierter Objektspeicherung in Verbindung gebracht wird, kann ein Data Lake durchaus ohne Hadoop entwickelt und effektiv eingesetzt werden. Beispielsweise könnte ein effizienter Data Lake auf verschiedenen Managementsystemen für relationale Datenbanken basieren. Ein Data Lake kombiniert vielfältige Technologien in einem System of Insight, damit Data Scientists Geschäftsanforderungen mithilfe agiler Datenexploration umsetzen können.

### 3.

#### **Der Data Lake generiert einen Speicherauszug der Daten – ganz ohne Governance → Falsch**

Software und Hardware sind die Schlüsselkomponenten einer Data-Lake-Lösung, aber auch die Katalogisierung, Qualität, Governance und Verwaltung von Daten spielen eine wichtige Rolle.

Genauso wie Unmengen an Daten in manchen Data Warehouses in einem Schwarzen Loch verschwinden, kann auch ein Data Lake zu einem Datensumpf werden, wenn wirksame Governance-Richtlinien fehlen.

### 4.

#### **Der Datenzugriff ist die wichtigste Kennzahl für den Erfolg eines Data Lakes → Falsch**

Nur Daten an einem zentralen Ort zu sammeln, ist von einer echten Analyselösung sehr weit entfernt. Das Ziel besteht in der Durchführung von Datenanalysen, die aussagekräftige Geschäftseinblicke für neue Umsatzströme, Kundenbindungsmodelle oder Produkterweiterungen liefern.

Die verwendeten Daten müssen jedoch vertrauenswürdig, relevant und für alle Datennutzer verfügbar sein. Ein Data Lake setzt einen intelligenten Metadatenkatalog voraus, der mit Geschäftsbegriffen verknüpft ist. So werden auch kryptische Daten ausgewertet und in einem verständlicheren Kontext präsentiert. Darüber hinaus berücksichtigt der Data Lake die Quelle und Qualität strukturierter wie unstrukturierter Datenbestände sowie Governance-Fabric. Auf diese Weise verfügen Unternehmen über geschützte, standardisierte, effizient verwaltete und vertrauenswürdige Daten.

### 5.

#### **Der Data Lake ersetzt das Data Warehouse → Falsch**

Der Data Lake kann mehrere Enterprise Data Warehouses (EDW) sowie weitere Datenquellen wie Social Media oder IoT-Daten zusammenführen. Alle diese Informationen fließen in den Data Lake, der mit Governance kombiniert werden kann, um die vertrauenswürdige Erkennung von Daten unternehmensweit zu vereinfachen.

Der Data Lake erweitert also Ihre EDW-Umgebung. Mit dieser Lösung können Data Scientists und Analysten Daten leicht durchsuchen, um neue Perspektiven, Erkenntnisse und Chancen zu entdecken und Innovation und Wachstum voranzubringen.



# Die Vorteile eines Multi-Cloud-Data-Lake

Lassen Sie sich nicht von fehlerhaften Informationen verwirren; ein regulierter Data Lake kann im gesamten Unternehmen den Zugriff auf vielfältige strukturierte und unstrukturierte Daten gewährleisten und ortsunabhängig dazu beitragen, dass diese Daten zuverlässig und sicher bleiben.

Ein entsprechend den Geschäftsanforderungen optimierter Data Lake kann Analysen beschleunigen und Informationen exakter verarbeiten, denn:



Die Daten befinden sich auf einer sicheren und zuverlässigen Infrastrukturbasis.



Durch kontrollierte Dateneinspeisungen wird der Data Lake mit zuverlässigen Informationen gefüllt und anschließend werden die Datenbestände, ihre Metadaten und ihr Geschäftskontext dokumentiert. So ergibt sich ein Datenfluss in Echtzeit in den Data Lake hinein.



Die Qualität, der Ursprung und die Abstammung der Daten sind bekannt.



Die Daten werden in Geschäftssprache hinterlegt, sodass Data Scientists sofort mit der Arbeit beginnen können und sich nicht mit der Bedeutung kryptischer Ausdrücke befassen müssen.



Die Daten werden ordnungsgemäß klassifiziert, geschützt und reguliert.

# Die Wahrheit ist irgendwo da draußen

Das **IBM Datenmanagement-Portfolio** hat Kunden erfolgreich geholfen, typische Probleme und Mythen beim Thema Data Lake zu vermeiden und die kritischen Schritte zur erfolgreichen Implementierung eines regulierten Data Lakes zu bewältigen.

## Entdecken Sie Data-Lake-Lösungen von IBM

Fördern Sie intelligentere Entscheidungen durch die Nutzung einer größeren Bandbreite an Datentypen aus mehr Datenquellen.



## E-Book lesen

Erfahren Sie, wie regulierte Data Lakes Chancen eröffnen, aus Daten wichtige Geschäftseinblicke abzuleiten.

