



# DATA LAKE

**Der Data Lake als zentrales Element in  
Analytics-Architekturen**

**E-Book**

## Zusammenfassung

Das hier vorliegende eBook greift mit dem Thema des Data Lake in analytischen Architekturen eine aktuelle und intensive Diskussion in Unternehmen auf. Dieser Begriff wird mit einem engen Fokus auf die Bereitstellung von Rohdaten verbunden, was jedoch eher seiner technischen Natur entspricht. Jedoch stecken die Herausforderungen bei der Gestaltung eines Data Lake nicht nur auf der technischen Seite, sondern ist dies auch mit einem organisatorischen Rollenkonzept und einhergehenden Verantwortlichkeiten sowie insbesondere datenschutzrechtlichen Aspekten verbunden. Daher sind in der Konsequenz Make-or-Buy-Entscheidungen bezüglich der architektonischen Bausteine zu treffen, andererseits ein Change-Management zu betreiben, um den organisatorischen und rechtlichen Rahmen zu schaffen, der die Nutzung von Daten in den Themenkomplexen der zunehmenden Digitalisierung in Unternehmen wertstiftend ermöglicht. Daher trägt das vorliegende eBook eine managementorientierte Diskussion, Erfahrungswerte aus Data-Warehouse-Projekten im Spannungsfeld des Data-Lake-Ansatzes zusammen, um einen Beitrag zur Diskussion rund um die digitale Transformation zu leisten.

## Inhaltsverzeichnis

1 Das Entstehen des Data Lake .....	3
2 Das Data Warehouse und der Data Lake – integrierte Parallelwelten.....	4
2.1 Der Klassiker – das Data-Warehouse-Konzept.....	4
2.1.1 Definition und charakterisierende Merkmale .....	5
2.1.2 Struktur und Datenfluss .....	7
2.2 Das Umdenken – der Data Lake zur Rohdatenbereitstellung .....	8
2 Architektonische Aspekte.....	10
3 Datenflussorientierte Betrachtungen zum Data Lake .....	12
3.1 Metadata Driven Onboarding .....	12
3.2 Streamlined Data Refinery .....	15
4 Rechtliche und organisatorische Aspekte .....	16
4.1 Rechtliche Perspektive .....	16
4.2 Organisatorische Perspektive.....	22
4.2.1 Der Head of ACC.....	24
4.2.2 Das Project Lab .....	24
4.2.3 Die Data Science Community .....	24
4.2.4 Die Fachbereiche .....	24
4.2.5 Die IT.....	24
4.2.6 Das Analytics Governance Gremium .....	24
5 TDWI Research - Prioritäten für Data Lakes.....	25
5.1 Verständnis über den Data Lake .....	25
5.2 Technologische Vorteile .....	26
5.3 Plattform .....	26
5.4 Hybride Architektur .....	26
5.5 Hadoop ergänzen .....	26
5.6 Graphical User Interface (GUI) .....	26
5.7 Vorsicht vor Datendumping .....	27
5.8 Gestaltung des Data Lake .....	27
5.9 Konzentration auf Rohdaten .....	27
5.10 Keine generellem, sondern eine individuelle Steuerung des Data Lake .....	27
5.11 Spezialisten für Datenmanagement.....	28
5.12 Vernetzung der Stakeholder.....	28
6 Fazit – Data Oceans und Data Swamps .....	28
Literatur.....	30