DER SUPERVISED DATA LAKE

DIE GRUNDLAGE FÜR IHRE BIG DATA-INITIATIVEN

Ein Data Lake dient als Grundlage für Big Data-Initiativen sowie als Ergänzung oder Ersatz für ein bestehendes Data Warehouse. Eine Organisation profitiert an vielen Stellen durch die gewonnene Flexibilität. Doch ohne Konzept und strukturiertes Vorgehen wird ein Data Lake schnell zum Data Swamp, der nur noch Kosten verursacht und keinerlei Mehrwert liefert.

VERKÜRZEN SIE MIT FLEXIBLER DATENHALTUNG DIE TIME-TO-ANALYSIS

Ein Data Lake resultiert in einer für Analysezwecke optimierten zentralen Datenplattform, die Daten aus mehreren, in der Regel heterogenen, Quellen bereitstellt. Die Zusammenführung und Verdichtung ist dabei keine notwendige Bedingung. Konzeptions- Integrationsaufwände werden dadurch drastisch reduziert. Einem Wunsch nach Bereitstellung neuer Daten kann somit sehr schnell entgegengekommen werden. Big Data-Technologien ermöglichen dabei eine sehr günstige Datenhaltung und -verarbeitung. Neben einer sehr schnellen und agilen Bereitstellung neuer Daten – bspw. für die prototypische Entwicklung digitaler Lösungen – kann eine gewachsene Systemlandschaft von der Ergänzung eines Data Lake auf mehreren Ebenen profitieren.

IHR INDIVIDUELLER SUPERVISED DATA LAKE

Die Erschaffung eines unternehmensweiten Supervised Data Lake ist ein hochgradig kundenindividuelles und komplexes Projekt – nicht nur auf technologischer Ebene. Der Supervised-Ansatz folgt dem „Data Governance First“ Gedanken unter der Prämisse größtmöglicher Flexibilität. Entwickelt wurde dieser mit den Erkenntnissen jahrelanger Erfahrungen im Datenmanagement. Ein flexibles Konzept, wie der Data Lake, bietet nicht nur die Möglichkeit vorhandene Analyselandschaften zu modernisieren oder zu ersetzen. Getreu dem Motto „Think Big, Start Small, Act Now“ erarbeiten wir mit Ihnen gemeinsam ein für Sie passendes Konzept und begleiten Sie bei der Einführung sowie der Umsetzung erster Use Cases des Data Lake.

Die fünf Projektphasen

1. VISION & ZIELE
2. KONZEPTION DATA LAKE
3. IMPLEMENTIERUNG BASIS-UMGEBUNG
4. DATA LAKE RAMP-UP
5. LAUFEND: KORREKTUR & EVOLUTION

DER SUPERVISED DATA LAKE

DIE GRUNDLAGE FÜR IHRE BIG DATA-INITIATIVEN

DIE INITIALISIERUNG EINES SUPERVISED DATA LAKE

Erarbeiten der Vision, Ziele und Rahmenbedingungen

- Sensibilisierung für Digitalisierung und Big Data
- Treiber / Motivation
- Vorgehen
- Technologie / Konzepte
- Use Cases
- Prozesse / Organisation
- Standortbestimmung
- Basic Governance
- Bestimmung der langfristigen Nutzergruppen
- Vendorstrategie

Konzeption des Data Lake

- Erarbeitung eines kundenspezifischen Konzepts, abhängig vom Ziel des Data Lake und den Basic Governance-Anforderungen
- Sizing der Basis-Umgebung für die Ramp-Up-Phase und entwickeln der Scale-Out-Strategie

Implementierung der Basis-Umgebung

- Anpassen der Infrastruktur (Firewall, Proxies, ...)
- Anlegen des Metadatenkonzeptes
- Anlegen des Bereinigungskonzeptes

Data Lake Ramp-Up

- Ersten Case heraussuchen und mit einer ausgewählten Gruppe starten – Start Small
- Erkenntnisse erfordern ggf. eine Anpassung des Konzeptes und der Parametrisierung der Basis-Umgebung
- Anpassung der Data Governance-Regelungen

Laufende Korrektur und Evolution

- Sowohl neue Technologien als auch Erkenntnisse aus dem laufenden Betrieb können zu einer Korrektur am Konzept und den Komponenten der Basis-Umgebung führen
- Das Konzept sollte laufend auf Optimierungspotenziale und Übereinstimmung mit Compliance-Anforderungen geprüft werden
- Erweiterung des Clusters

DER SUPERVISED ANSATZ BRINGT DEN UNTERSCHIED

Mit dem Supervised-Ansatz werden innerhalb der Konzeption u. a. die folgenden Themen behandelt:

- Kostenoptimierung durch die Auswahl optimaler Speicherformate, Entwicklung eines Speicherformat-Guides, optimale Replikationseinstellungen
- Self-Service durch Bereitstellung von Data Upload-Möglichkeiten, umfangreiches Sandboxing, geringe Einschränkung im Softwareportfolio
- Katalogisierung durch zentrales Register aller Datenquellen im Data Lake, fachliches Auszeichnen von Daten
- Regulatorik u. a. durch Sicherstellung regulatorischer Anforderungen, Berücksichtigung von Mindest- und Maximalhaltedauern, Anonymisierung / Pseudonymisierung
- Sicherheit u. a. durch Sicherstellung von Zugriffsbeschränkungen, Erstellung eines Berechtigungskonzeptes
- Automatische Reinigung u. a. durch periodisches Löschen nicht mehr benötigter Daten, Säubern von Daten, welche den Mindestanforderungen nicht genügen
- Anforderungen an die Datenaufnahme u. a. Data Ownership, Metadatenanreicherung, Ablageortkonzept, Nomenklatur

GERT JAN FEICK
Manager

INFOMOTION GMBH
Westhafenplatz 1
60327 Frankfurt

T +49 69 56608-3000

gert.jan.feick@infomotion.de

www.infomotion.de

Die INFOMOTION GmbH ist das führende Beratungsunternehmen für Business Intelligence, Big Data und Digital Solutions im deutschsprachigen Raum. Das umfangreiche Portfolio reicht von der strategischen Beratung über die Konzeption, Implementierung und den Betrieb nachhaltiger BI-Lösungen bis hin zur Schulung Ihrer Mitarbeiter.

INFOMOTION
we love data