



David Gierse

Berlin, der 31.01.2025

Die Zukunft der Daten: Wie moderne Architekturen Unternehmen verändern 🌐

Daten sind der Treibstoff der digitalen Wirtschaft. Doch wie wir mit Daten umgehen, hat sich in den letzten Jahren drastisch verändert. Unternehmen stehen vor der Herausforderung, immer größere und vielfältigere Datenmengen nicht nur zu speichern, sondern auch nutzbar zu machen. Moderne Datenarchitekturen wie **Data Lake**, **Data Warehouse**, **Data Lakehouse** und **Data Mesh** haben das Potenzial, den Umgang mit Daten zu revolutionieren.

In diesem Blog werfen wir einen Blick auf die Eigenschaften, Herausforderungen und Best Practices dieser Architekturen und wie sie Unternehmen voranbringen können.

Data Warehouse: Der Klassiker mit Struktur

Das Data Warehouse ist seit Jahrzehnten die Grundlage vieler BI-Lösungen. Es bietet strukturierte Daten, standardisierte Abfragen und ist ideal für historische Analysen.

Vorteile:

- Geeignet für strukturierte Daten.
- Verlässliche Berichte und Dashboards.
- Optimiert für SQL-Abfragen und BI-Tools.

Einschränkungen:

- Wenig flexibel für unstrukturierte oder semi-strukturierte Daten.
- Hohe Kosten bei Skalierung.
- Batch-Verarbeitung ist ineffizient für Echtzeitanalysen.

Beispiel: Ein Einzelhändler nutzt ein Data Warehouse, um historische Verkaufsdaten für monatliche Berichte zu analysieren. Die Integration neuer Datenquellen dauert jedoch oft Wochen.

Data Lake: Flexibilität für alle Daten

Der Data Lake wurde als Antwort auf die Limitierungen des Data Warehouse entwickelt. Er speichert große Mengen strukturierter, semi-strukturierter und unstrukturierter Daten im Rohformat.

Vorteile:

- Hoch skalierbar und flexibel.
- Unterstützt maschinelles Lernen und Big-Data-Analysen.
- Ideal für unstrukturierte Daten wie Bilder oder Texte.

Herausforderungen:

- Gefahr des „Data Swamp“: Ohne Governance werden Daten unorganisiert und schwer auffindbar.
- Mangelnde Unterstützung für standardisierte Berichte und SQL-Abfragen.

Best Practice: Implementierung eines Metadatenkatalogs wie Apache Atlas, um Datenquellen effizient auffindbar und verständlich zu machen. Über setzt solche Lösungen erfolgreich ein, um den Zugang zu den richtigen Daten zu erleichtern.

Data Lakehouse: Das Beste aus beiden Welten

Das Data Lakehouse kombiniert die Stärken von Data Lakes und Data Warehouses. Es bietet die Flexibilität eines Data Lakes und die Verlässlichkeit eines Data Warehouses.

Vorteile:

- Unterstützt sowohl SQL-Analysen als auch maschinelles Lernen.
- Bietet ACID-Transaktionen für Datenintegrität.
- Ermöglicht die Nutzung von Daten im Rohformat und verarbeiteten Formaten.

Herausforderungen:

- Neues Konzept mit begrenzter Marktreife.
- Erfordert spezialisierte Tools wie Delta Lake für ACID-Transaktionen und Apache Spark für parallele Datenverarbeitung.

Beispiel: Ein FinTech-Unternehmen nutzt ein Data Lakehouse, um Kundendaten für Echtzeitanalysen und maschinelles Lernen zu integrieren. Dies ermöglicht personalisierte Angebote in Sekundenschnelle.

Data Mesh: Datenorganisation für die Zukunft

Data Mesh ist weniger eine technische Lösung, sondern ein organisatorisches Paradigma. Es fördert die dezentrale Verantwortung für Daten in Domänen (z. B. Marketing, Vertrieb) und setzt auf föderierte Governance.

Vorteile:

- Fördert Eigenverantwortung und höhere Datenqualität.
- Ermöglicht schnellere Entscheidungen durch domänenspezifische Datenprodukte.
- Skaliert besser in großen Unternehmen mit unterschiedlichen Datenanforderungen.

Herausforderungen:

- Erfordert kulturellen Wandel und klare Verantwortlichkeiten.
- Bedarf einer starken Governance, um inkonsistente Datenprodukte zu vermeiden.

Best Practice: Zalando hat den Wandel zu Data Mesh erfolgreich gemeistert, indem es klare Ownership-Modelle etablierte, Teams datenkompetent machte und eine Self-Service-Infrastruktur bereitstellte.

Warum moderne Architekturen entscheidend sind

Moderne Datenarchitekturen bieten mehr als nur technologische Vorteile. Sie verändern die Art und Weise, wie Unternehmen Entscheidungen treffen und Innovationen vorantreiben.

1. **Wettbewerbsvorteile:** Schnelle, datenbasierte Entscheidungen sind heute unverzichtbar.
2. **Skalierbarkeit:** Moderne Architekturen passen sich wachsenden Datenmengen und Anforderungen an.
3. **Innovationsförderung:** Sie ermöglichen Fortschritte in KI und maschinellem Lernen.

Abfazit: Welcher Ansatz passt zu dir?

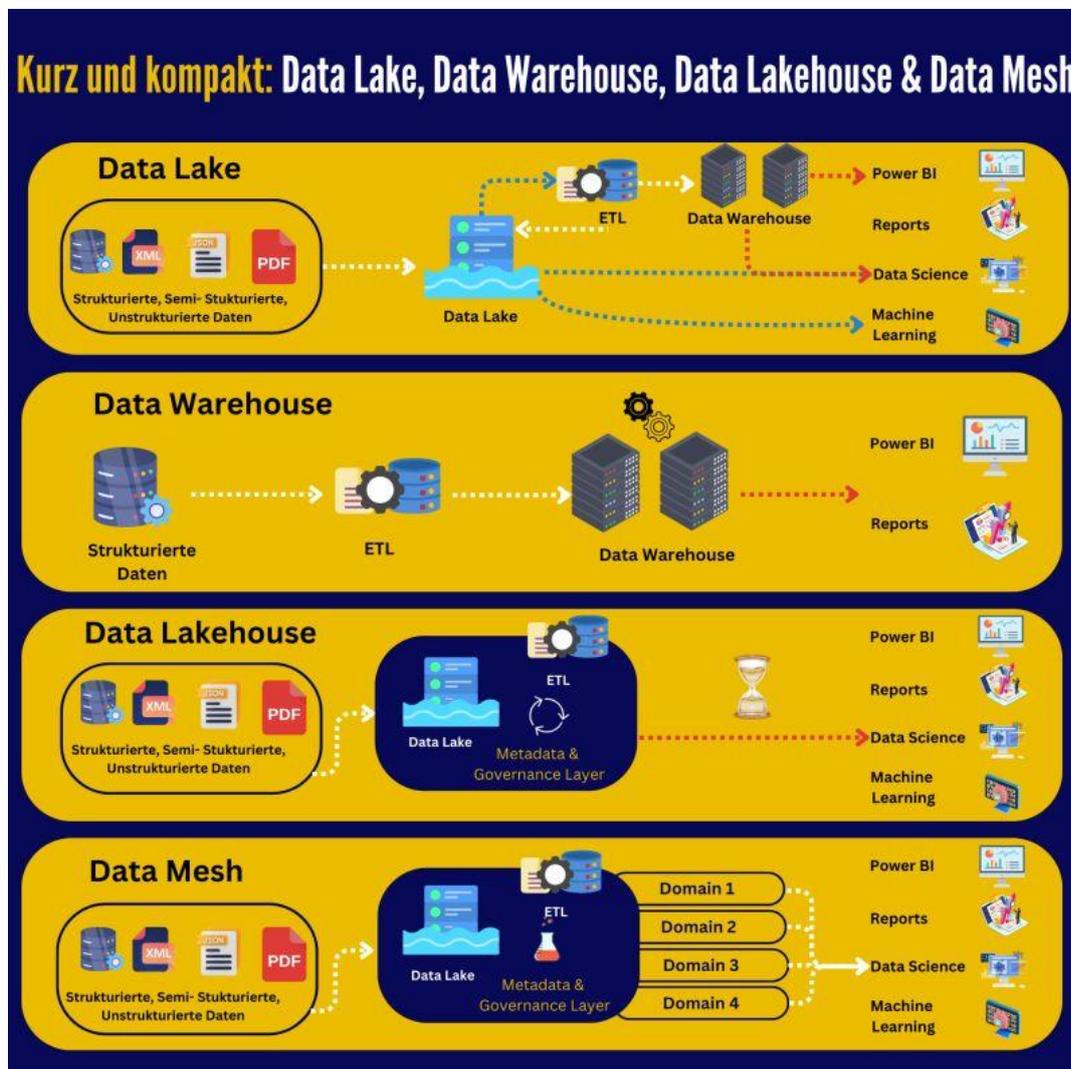
Die Wahl der richtigen Architektur hängt von den individuellen Anforderungen ab:

- **Data Warehouse** für strukturierte Berichte und Dashboards.
- **Data Lake** für unstrukturierte Daten und maschinelles Lernen.
- **Data Lakehouse** als Allrounder für Datenanalyse und KI.

- **Data Mesh** für Organisationen mit komplexen Datenanforderungen und dezentralen Teams.

Welche Architektur setzt ihr aktuell ein? Welche Erfahrungen habt ihr gemacht? Lasst uns in den Kommentaren diskutieren und voneinander lernen! 🚀

https://www.linkedin.com/posts/david-gierse-49a34718b_%F0%9D%97%97%F0%9D%97%B6%F0%9D%97%B2-%F0%9D%97%AD%F0%9D%98%82%F0%9D%97%B8%F0%9D%98%82%F0%9D%97%BB%F0%9D%97%B3%F0%9D%98%81-%F0%9D%97%B1%F0%9D%97%B2%F0%9D%97%BF-%F0%9D%97%97%F0%9D%97%AE%F0%9D%98%81%F0%9D%97%B2%F0%9D%97%BB-activity-7257405964562182144-a_gT?utm_source=share&utm_medium=member_desktop



<https://www.linkedin.com/in/david-gierse-49a34718b/>