

# Aktuelle Methoden für die Datenintegration: ETL/ELT, Datensynchronisierung, -virtualisierung und -streaming

Luca Münch

Technische Hochschule  
Mittelhessen

Fachbereich MND  
Wilhelm-Leuschner-Straße 13  
61169 Friedberg  
E-Mail:  
[luca.muench@mnd.thm.de](mailto:luca.muench@mnd.thm.de)

Prof. Dr. Harald Ritz

Technische Hochschule  
Mittelhessen

Fachbereich MNI  
Wiesenstraße 14  
35390 Gießen  
E-Mail:  
[harald.ritz@mni.thm.de](mailto:harald.ritz@mni.thm.de)

Prof. Dr. Peter Hohmann

Technische Hochschule  
Mittelhessen

Fachbereich MNI  
Wiesenstraße 14  
35390 Gießen  
E-Mail:  
[peter.hohmann@mni.thm.de](mailto:peter.hohmann@mni.thm.de)

## Kategorie

Bachelorarbeit

## Schlüsselwörter

Datenintegration, ETL, ELT, Datensynchronisierung, Datenvirtualisierung, Datenstreaming, KI, Internet der Dinge, Business Intelligence

## Zusammenfassung

Die Thesis setzt sich mit der zunehmenden Relevanz der Datenintegration im Kontext der Digitalisierung auseinander. Angesichts der exponentiell wachsenden Datenmengen, die aus vielfältigen Quellen stammen, wird die effiziente Verarbeitung und Speicherung von Daten zu einem entscheidenden Faktor für Privatpersonen sowie Unternehmen, unabhängig davon, ob die Daten in der Cloud oder lokal gespeichert sind. Ziel dieser Arbeit ist es, einen umfassenden Einblick in die verschiedenen Methoden der Datenintegration zu geben, insbesondere ETL/ELT, Datensynchronisierung, -virtualisierung und -streaming.

Im Allgemeinen bezeichnet Datenintegration die Standardisierung von Datendefinitionen und -strukturen in einer Sammlung von Datenquellen, indem ein gemeinsames konzeptionelles Schema verwendet wird. Die Datenintegration stellt sicher, dass die Bedeutung und Verwendung der Daten im Laufe der Zeit und bei verschiedenen Nutzern unverändert bleibt, und dass die Daten in verschiedenen Systemen oder Datenbanken konsistent oder logisch kompatibel sind.

Der Analyseprozess beginnt mit einer eingehenden Klärung des Konzepts der Datenintegration, gefolgt von einem umfassenden Überblick über die genannten Methoden. Dabei werden nicht nur ihre Vor- und Nachteile beleuchtet, sondern auch aktuelle Anwendungen detailliert betrachtet. Besonderes Augenmerk liegt auf einem gründlichen Vergleich dieser Methoden, einschließlich Bewertungen wie einem Vergleich zwischen Datenstreaming durch Stream- und

Batch-Verarbeitung. Das übergeordnete Ziel besteht darin, optimale Lösungen für eine Vielzahl von Anwendungsfällen zu identifizieren und deren Stärken sowie Schwächen herauszuarbeiten.

Die Ergebnisse dieser eingehenden Untersuchung werden anschließend bewertet und zusammengefasst, wobei die vielfältigen Ansätze der Datenintegration hervorgehoben werden. Traditionelle ETL und ELT Methoden erweisen sich dabei als bewährte Ansätze, wobei die Wahl zwischen ihnen von den spezifischen Anforderungen und Ressourcen abhängt. Datensynchronisierung, als kontinuierlicher Prozess der Datenaktualisierung, erfordert eine individuelle Anpassung an die Sicherheitsanforderungen, während Datenvirtualisierung als Integration ohne physische Datenverschiebung eine sorgfältige Abwägung von Kosten, Qualität und Performance erfordert. Datenstreaming, speziell in Echtzeit-Szenarien, bringt hingegen Herausforderungen in Bezug auf Datenverlust und -integrität mit sich und wird als Methode zur Identifikation und Verarbeitung wertvoller Datenelemente in Echtzeit betrachtet.

Der Ausblick auf zukünftige Entwicklungen unterstreicht die weiterhin zunehmende Bedeutung der Datenintegration angesichts wachsender Datenmengen und identifiziert mögliche Fortschritte, darunter die Verbesserung der Echtzeitdatenverarbeitung und die Integration von KI, um Prozesse automatisch zu optimieren. In einer zunehmend vernetzten Welt der Internet der Dinge könnten Datensynchronisierung und -streaming weiter an Bedeutung gewinnen, während Datenvirtualisierung dazu beitragen kann, mit den stetig komplexer werdenden Datenlandschaften besser umzugehen.

Abschließend wird betont, dass die Datenintegration ein dynamisches und sich entwickelndes Feld bleibt, das eng mit den Fortschritten in der Datenverarbeitung und dem BI-Umfeld verbunden ist. Organisationen, die in der Lage sind, flexibel auf diese Veränderungen zu reagieren und die passenden Methoden auszuwählen, werden

besser positioniert sein, um aus ihren Daten wertvolle Erkenntnisse zu gewinnen und wettbewerbsfähig zu bleiben.

### **Literatur**

Armbrust, Michael/ Ghodsi, Ali/ Xin, Reynold/ Zaharia, Matei: Lakehouse: A New Generation of Open Platforms that Unify Data Warehousing and Advanced Analytics, Creative Commons Attribution License, 2021

Goodhue, Dale/ Kirsch, Laurie/ Wybo, Michael: The Impact of Data Integration on the Costs and Benefits of Information Systems, Management Information Systems Research Center, 1992

Martin, James: Information Engineering, Savant Research Studies, Carnforth, 1986

Rossak, Ines (Hrsg.): Datenintegration – Integrationsansätze, Beispielszenarien, Problemlösungen, Talend Open Studio, München, 2013

Schön, Dietmar: Planung und Reporting im BI-gestützten Controlling – Grundlagen, Business Intelligence, Mobile BI, Big-Data-Analytics und KI, Wiesbaden, 2022

Spotfire (Hrsg.): What is data streaming?, o.O., o.J., URL: <https://www.spotfire.com/glossary/what-is-data-streaming> (Abruf 20.10.23)