

# Performance-Nutzen-Analyse: Automatisierung eines Data Warehouse

## Untersuchung am Beispiel des Datavault Builder auf verschiedenen Zielplattformen in der Azure Cloud

Kevin Ullrich

Technische Hochschule  
Mittelhessen

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

Arthur Arendt

OPITZ CONSULTING  
Deutschland GmbH

Am Weidenring 56  
61352 Bad Homburg v. d. H.  
E-Mail:  
[arthur.arendt@opitz-consulting.com](mailto:arthur.arendt@opitz-consulting.com)

### Kategorie

Masterarbeit

### Schlüsselwörter

Cloud Computing, Data Warehouse, Automatisierung,  
Performance, Nutzen, TPC-H Benchmark

### Zusammenfassung

In der vorliegenden Arbeit wird eine Performance-Nutzen-Analyse vorgenommen, untersucht am Beispiel des Datavault Builder zur automatischen Data Warehouse-Erstellung auf verschiedenen Zielplattformen (konkret: Azure SQL-Datenbank, Exasol und Snowflake) in der Azure Cloud. Die Aufgabenstellung im Zusammenhang mit der vorliegenden Arbeit besteht im Vergleich verschiedener Zielplattformen in Kombination mit dem Datavault Builder und darin, eine Aussage über die Vorteilhaftigkeit einer bestimmten Zielplattform für den Datavault Builder im Hinblick auf die Performance und Nutzung geben zu können.

Zudem soll die Kenntnis der Arbeit es ermöglichen, die Thematik der Data Warehouse-Automatisierung tiefer zu durchdringen, um die jeweils bestmögliche Nutzung in Kundenprojekten zu garantieren. Dazu wird ein Benchmark in unterschiedlich großen Skalierungsausprägungen verwendet. Qualitative Aspekte werden mit Hilfe einer Nutzwertanalyse bewertet.

Als Ergebnis wird deutlich, dass sich die Datenbank Exasol für alle drei TPC-H-Datensatzgrößen trotz des schlechteren Nutzwertes als vorteilhafter herausgestellt hat: zum einem deshalb, weil sie in jeder Skalierungsstufe die geringste Gesamtlaufzeit gegenüber der Azure SQL-Datenbank und Snowflake erzielt hat;

zum anderen aus dem Grund, weil sie nur einen Bruchteil der Kosten der anderen beiden Zieldatenbanken verursacht.

Die Azure SQL-Datenbank ist aufgrund ihrer langen Laufzeiten und den damit verbundenen Kosten weniger für den Einsatz mit dem Datavault Builder zu empfehlen.

Snowflake hingegen ist zwar nicht viel langsamer als Exasol, verursacht aber erheblich höhere Kosten. Der Vorteil, den Snowflake jedoch gegenüber Exasol hat, ist darin zu sehen, dass Snowflake als Software-as-a-Service-Lösung bereitgestellt wird, bei der Snowflake alle Aspekte der Verwaltung abdeckt.