

Vergleichende Bewertung von Data-Warehouse-Landschaften in der Cloud am Beispiel von Microsoft Azure Synapse Analytics und Snowflake

Thao-An Hua

Technische Hochschule Mittelhessen

Fachbereich MND
Wilhelm-Leuschner-Str. 13
61169 Friedberg
thao-an.hua@mnd.thm.de

Prof. Dr. Harald Ritz

Technische Hochschule Mittelhessen

Fachbereich MNI
Wiesenstraße 14
35390 Gießen
harald.ritz@mni.thm.de

Wadim Myza (B.Sc.)

INFOMOTION GmbH

BU Veltzé
Westhafenplatz 1
60327 Frankfurt am Main
wadim.myza@infomotion.de

Kategorie

Bachelorarbeit

Schlüsselwörter

Business Intelligence, Cloud-Computing, Softwareevaluation, Bewertungsmatrix, Microsoft Azure Synapse Analytics, Snowflake

Zusammenfassung

Die Welt befindet sich in einem zunehmend rapiden gesellschaftlichen Wandel. Vieles soll schneller, virtueller und digitaler werden. Infolge der digitalen Transformation ist das unverzügliche Treffen fundierter Entscheidungen von signifikanter Bedeutung, um kompetitiv zu bleiben. Durch die frühzeitige Reaktion auf Änderungen im Marktumfeld kann ein Wettbewerbsvorteil gegenüber der Konkurrenz erzielt werden. Zur Entscheidungsunterstützung greifen viele Unternehmen auf Data-Warehouse-Landschaften in der Cloud zurück, um große Datenmengen zu analysieren. Die zentralen Merkmale des Cloud-Computings, wie die Skalierbarkeit, tragen zu einer effizienten Entscheidungsfindung bei.

Die adäquate Auswahl einer Cloud-Datenplattform-Lösung stellt eine Herausforderung für Unternehmen dar. Angesichts der stetig voranschreitenden Digitalisierung und jene damit einhergehende unterschiedliche Reifegrade der am Markt verfügbaren Auswahlmöglichkeiten ergibt sich für viele Unternehmen die Frage, welche Cloud-Datenplattform und dessen Datenlösung die beste Unterstützung bietet. Infolgedessen ist das Untersuchungsziel der Bachelorarbeit die Konzeption einer Bewertungsmatrix zur fundierten Auswahl einer Cloud-Datenplattform-Lösung. Am Beispiel der Gegenüberstellung des PaaS-Produkts Azure Synapse Analytics von Microsoft (MS) und des SaaS-Produkts Snowflake ist zu eruieren, welche Lösung in Bezug auf die beleuchteten Anforderungen am besten geeignet ist.

Mit Microsoft steht Unternehmen einer der größten Cloud-Anbieter zur Verfügung. In Anbetracht der breiten

Palette an Produkten, wie beispielsweise Synapse Analytics, werden allgemein moderne Lösungen für diverse Anwendungsszenarien angeboten. Snowflake ist eine vom gleichnamigen Unternehmen entwickelte SaaS-Lösung. Der Komplettservice bietet eine vollständige Verwaltung der Infrastruktur in der Cloud an, sodass der Fokus auf die Analyse und Verwaltung der Daten gelenkt werden kann.

Zur Ermittlung relevanter Kriterien wurden verschiedene Quellen gesichtet. Die gewonnenen Erkenntnisse sind in die Konzeption der Bewertungsmatrix eingeflossen, die aus 17 Kriterien besteht. Diese sind in funktionale und qualitative Anforderungen gegliedert und in den sieben Kategorien Administration, Architektur, Interoperabilität, Kosten, qualitative Kriterien, Data Governance und Flexibilität unterteilt.

Des Weiteren wurde jeweils ein Datenmodell in MS Azure Synapse Analytics und Snowflake, im Kontext eines fiktiven Anwendungsszenarios umgesetzt. Die vierstufige Data-Warehouse-Grundarchitektur stellte hierbei die Basis. Als Datenquelle wurde ein Ausschnitt der Daten der aktuellen COVID-19-Pandemie verwendet. Bei der praktischen Umsetzung stellte sich dabei heraus, dass die Implantierung je nach gesammelten Erfahrungen ein mehr oder weniger kompliziertes Unterfangen darstellt. Auf Basis der jeweiligen Datenmodelle wurde jeweils ein Dashboard mittels Power BI Desktop entworfen und realisiert. Anhand dessen und durch die jeweilige Umsetzung konnten sowohl für MS Azure Synapse Analytics als auch für Snowflake praktische Beispiele zur Bewertung der einzelnen Kriterien herangezogen werden. In diesem Zusammenhang wurden beispielsweise die jeweiligen Kosten für den Entwicklungszeitraum berechnet und Performance-Tests durchgeführt.

Im Hinblick auf die beleuchteten Kriterien gelten sowohl MS Azure Synapse Analytics als auch Snowflake als moderne Analyseplattform. Bei der vergleichenden Bewertung hat sich zum Betrachtungszeitpunkt herausgestellt, dass Snowflake in Bezug auf die Anforderungen an eine

moderne Analyseplattform die sieben Kriterien Auto-Resume, Bereitstellung, Skalierbarkeit, Performance, Einsatz von Drittanbieter-Tools, Abrechnung und Wartbarkeit besser qualifiziert erfüllt als MS Azure Synapse Analytics. Trotzdem kann auch MS Azure Synapse als moderne Analyseplattform einen Mehrwert für das Unternehmen liefern, da die individuellen Anforderungen eines Unternehmens je nach spezifischen Anwendungsfall variieren.

Literatur

Hartmut, B.; Hartmut, H.: Vertrieb und Marketing in der digitalen Welt, Springer Gabler, Wiesbaden, 2018

Kraynak, J.; Baum, D.: Cloud Datawarehousing, 2. Auflage, John Wiley & Sons, New York, 2020

Schnider, D.; Jordan, C.; Welker, P.; Wehner, J.: Data Warehouse Blueprints, Carl Hanser Verlag GmbH & Co. KG, München, 2016

Norman, B.: Treiber einer „Data Warehouse as a Service“-Lösung, In: Finger, R. (Hrsg.): *BI & Analytics in der Cloud: Architektur, Vorgehen und Praxis*, 1. Auflage, dpunkt.verlag, Heidelberg, 2018