

# DataOps für cloudbasierte Artefakte und Anwendungen

Laura Leester

Prof. Dr. Harald Ritz

Dr. Bernd Heßbrügge

Technische Hochschule Mittel-  
hessen

Technische Hochschule Mittel-  
hessen

INFOMOTION GmbH

Fachbereich MNI  
Wiesenstr. 14  
35390 Gießen  
E-Mail: [laura.leester@mni.thm.de](mailto:laura.leester@mni.thm.de)

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

BU Strategy & Digital Solutions  
Westhafenplatz 1  
60327 Frankfurt  
E-Mail: [bernd.hessbruegge@infomotion.de](mailto:bernd.hessbruegge@infomotion.de)

## Kategorie

Bachelorarbeit

## Schlüsselwörter

DataOps, Cloud Computing

## Zusammenfassung

Mit wachsendem Interesse befassen sich deutsche Unternehmen mit der Thematik Datenspeicherung und dem damit verbundenen Datenmanagement. Hier wurde das Potential erkannt, durch datengetriebene Prozesse und Entscheidungen eine bessere Position im Wettbewerb zu erlangen und so nach außen und innen Vorteile für das Unternehmen erwirken zu können. Die Problemstellung umfasst die (unbekannteren) Möglichkeiten, das Datenmanagement mit DataOps zu verbessern. Diese Thesis befasst sich speziell mit Unternehmen, die unter Einsatz von Cloud Computing Artefakte und Anwendungen nutzen oder produzieren. Die Thesis stellt auf Basis dieses Interesses im Verlauf die Untersuchung der Eignung der DataOps-Methode für cloudbasierte Artefakte und Anwendungen dar.

Um die Forschungsfrage zu beantworten werden zunächst Definitionen der verschiedenen Themengebiete aufgesetzt. Deren Eigenschaften werden extrahiert und in Bezug zueinander gesetzt, um Gemeinsamkeiten und Unterschiede der Ziele und Methoden der verschiedenen Themengebiete aufzuzeigen und eine Eignung und einen gemeinsamen Nutzungskontext abzuleiten. Die DataOps-Methode stellte sich als eine agile, datenbezogene Vorgehensweise für Rollen im Data Science-Bereich dar. Es fällt auf, dass diverse DataOps-Prinzipien darauf hinarbeiten, dass die Datenkompetenz eines Unternehmens erhöht wird, da DataOps Daten als wertschöpfenden Faktor betrachtet und diesen einen hohen Wert beimisst. Beiwerke zu DataOps sind die verwandten DevOps-, Agile- und Lean Manufacturing-Methoden. Davon übernimmt DataOps die Agilität mit kurzen Arbeitszyklen und dem Ziel, schnell Ergebnisse zu produzieren, um Kundenbedürfnisse abzudecken. Ähnlich wie DevOps auch Rollen an Mitarbeiter vergibt oder eine Struktur zwischen Teams aufstellt, so legt auch DataOps Rollen für Data Governance, Analysten und Ingenieure fest, um durch die übersichtliche Struktur die Kommunikation zwischen den Teams besser zu koordinieren. Anschließend erfolgt

die Betrachtung cloudbasierter Artefakte und Anwendungen. Artefakte treten in Unternehmen im Data-Science-Bereich in diversen Varianten auf und es ist wichtig, hier klar abzutrennen, dass von allen im Data-Science-Life-Cycle anfallenden Artefakten gesprochen wird, nicht nur jenen, die in der Modellierung und Entwicklung auftreten. Den Einsatz von Cloud Computing rechtfertigen dabei nicht alle Varianten der Artefakte, die als irrelevant betrachtet werden hier durch den Vergleich mit den Eigenschaften von Cloud-Services herausgestellt und für die weitere Betrachtung vernachlässigt. Cloudbasierte Anwendungen lassen sich dagegen relativ simpel auf bereits bekannte Cloud-Services zurückführen: das Software-as-a-Service-Angebot (SaaS) als direkt nutzbare Software sowie das Platform-as-a-Service-Angebot (PaaS) zur Softwareentwicklung in der Cloud. Die als cloud-relevant eingestuften Artefakte und Anwendungen werden mit den Eigenschaften und Anforderungen der DataOps-Methode verglichen. Hier zeigt sich, dass der Umgang mit cloudbasierten Artefakten durch den Einsatz von DataOps profitieren kann. Auch PaaS und SaaS wurden hierzu überprüft, wobei sich zeigt, dass SaaS losgelöst von der Einwirkung von DataOps ist, PaaS aber den Nutzern erlaubt, Einfluss auf Faktoren zu nehmen, die DataOps zu verbessern sucht. Der innovative Umgang mit Daten in auf PaaS bereitgestellte Testumgebungen oder beliebiges Einrichten von Schnittstellen werden nicht restringiert. Bei SaaS sind Anwender auf die Einstellungen und Konfigurationsmöglichkeiten angewiesen, die der Anbieter zur Verfügung stellt, ohne dass sie weiteres Customizing vornehmen können. Die Integration in vorhandene Systeme kann so deutlich komplexer ausfallen als bei PaaS-Systemen. So macht PaaS das Datenmanagement einfacher und individualisierbar. Die Data Governance kann besser eingehalten werden und unternehmensspezifische Ziele im Data&Analytics-Bereich werden mit PaaS leichter erreicht, als mit generischen SaaS-Angeboten. Auch zeigt sich, bei welchen cloudbasierten Artefakten und Anwendungen der Einsatz von DataOps einen Nutzen bringt und wie DataOps in diesem Fall eingesetzt werden kann. Es ist zu beachten, dass diese Thesis davon ausgeht, dass cloudbasierte Artefakte und Anwendungen eingesetzt werden, es erfolgt keine Evaluierung der generellen Nutzung von DataOps für klassische on-prem Data-Center.