

**Exposé für die Bachelor-Thesis im Studiengang Social Media Systems**

**Wintersemester 2020/21**

## Stand: 03.09.20

Betreut durch: AW Thema: Cloud Data Warehousing:

Evaluation von Amazon Redshift und Snowflake anhand eines Kriterienkatalogs und einer prototypischen Implementierung

i

**Motivation und Problemstellung:** Schnelle Innovationszyklen und neue Technologien verändern das Data Warehousing und schaffen neue Möglichkeiten. Das Cloud Computing als eine dieser neuen Technologien, öffnet Türen für das Data Warehousing, denn neben dem Einsparen von Ressourcen ermöglicht die Cloud auch das Auslagern von IT-Infrastrukturen. Zwei der führenden Anbieter im Bereich Cloud Data Warehousing sind Amazon Redshift und Snowflake. Die beiden Cloud Services treiben den Wandel im Bereich des Cloud DWH voran und ein Vergleich liefert wertvolle Erkenntnisse für die Zukunft und das Arbeiten mit DWHs in der Cloud.

# Forschungsfrage:

### Im Rahmen der Abschlussarbeit soll folgende Frage beantwortet werden:

Welche Ergebnisse lassen sich aus dem Vergleich der Cloud Data Warehouses ableiten und welche Rückschlüsse können daraus gezogen werden?

# Ziel der Arbeit:

Das Ziel der Arbeit ist es, einen Kriterienkatalog zum Vergleichen der beiden Cloud Data Warehouses Amazon Redshift und Snowflake zu erarbeiten. Durch das Implementieren von Beispieldaten in beiden Services soll zudem ein praxisbezogener Vergleich auf Basis der Kriterien durchgeführt werden, der in einem Ergebnis zusammengefasst wird und letztlich Rückschlüsse über die Nutzungsmöglichkeiten der beiden Cloud Data Warehouses erlaubt.

# Methodisches Vorgehen:

Auf Basis wissenschaftlicher Literatur werden Grundlagen des Data Warehousing, des Cloud Computing und der bei den Cloud Data Warehouses, Amazon Redhsift und Snowflake, aufbereitet und vermittelt.

Danach wird ein Kriterienkatalog mit Qualitätsmerkmalen auf Basis von Normen für Software (ISO 25010, ISO 9126) und POCs (Proof of Concept) für die Cloud erarbeitet, der durch die Implementierung mit Testdaten in beiden Cloud Data Warehouses und durch eine Bewertung anhand des Kriterienkatalogs einen faktenbasierten Vergleich ermöglicht. Zuletzt werden die Ergebnisse aus dem Vergleich dargestellt und bewertet.

ii

# Gliederung:

### Einleitung

* 1. Problemstellung und Motivation
  2. Zielsetzung
  3. Methodik und Vorgehensweise

*Relevanz der Arbeit erklären und Ziele aufführen*

### Einführung in die Cloud und das Data Warehousing

* 1. Einführung in das Cloud Computing
  2. Einführung in das Data Warehousing

*Grundlagen für das Verständnis von Cloud, neue Möglichkeiten, Data Warehousing auch speziell in Verbindung mit der Cloud*

### Cloud Data Warehousing

* 1. Einführung in das Cloud Data Warehousing
  2. Marktüberblick Cloud Data Warehousing
  3. Amazon Redshift: Grundlagen
  4. Snowflake: Grundlagen

*Überblick über aktuelle Marktsituation und genauere Betrachtung von Amazon Redshift und Snowflake*

iii

### Konzeption und Erarbeitung des Kriterienkatalogs

* 1. Einleitung
     1. ISO 25010 und ISO 9126
     2. Proof of Concept für die Cloud
  2. Auswahl der Kriterien für den Vergleich
  3. Aufbau eines Kriterienkatalogs für die Implemtierung

*Für was Kriterien? Was ist wichtig? Welche Kriterien sollten wie gewichtet werden und wie macht man die beiden DWHs vergleichbar?*

### Implementierung und Vergleich

* 1. Rahmenbedingungen und Testdaten für die Implementierung
  2. Erläuterung der DWH Umgebung
     1. Aufbau der Umgebung in Redshift
     2. Aufbau Umgebung in Snowflake
  3. Dokumentation der Implementierung in den beiden Cloud DWHs
     1. Implementierung der Umgebung in Redshift
     2. Implementierung der Umgebung in Snowflake

5.4 Vergleich der Implementierung in den beiden Cloud DWHs

*Erklärung und Aufbau der Demo-Umgebung und bewerten der Kriterien, um zu vergleichen und eine Beurteilung und Bewertung zu ermöglichen*

### Ergebnis und Zusammenfassung

* 1. Zusammenfassung der Erkenntnisse aus der Implementierung
  2. Auswertung der Ergebnisse auf Basis des Kriterienkatalogs

iv

*Ergebnisse Zusammenfassen und Auswerten, sodass eine Aussage auf Basis der Kriterien möglich ist und Rückschlüsse gezogen werden können*

### Fazit und Ausblick

*Ergebnisse prägnant darstellen, kritisch diskutieren und einen Abschluss setzen, Ausblick geben*

# Literatur:

Amazon (2020a): Amazon Redshift Database Developer Guide.<https://docs.aws.amazon.com/redshift/latest/dg/redshift-dg.pdf#proof-of-concept-playbook>, abgerufen am 22.08.2020

*Redshift Dokumentation*

Amazon Web Service (2020b): Overview of Amazon Web Services. [https://d0.awsstatic.com/whitepapers/aws-overview.pdf,](https://d0.awsstatic.com/whitepapers/aws-overview.pdf) abgerufen am 22.08.2020

*Übersicht über AWS und die verschiedenen Services*

Amazon (2020c): Whitepapers: latest aws-overview.<https://docs.aws.amazon.com/whitepapers/latest/aws-overview/aws-overview.pdf>, aufgerufen am 22.08.2020

*Ebenfalls neueste Übersicht über AWS Services*

[Bauer, Andreas](https://hds.hebis.de/thm/Search/Results?lookfor0%5b%5d=Bauer%2C%20Andreas&amp;type0%5b%5d=author&amp;lastposition) (2013): [Data-Warehouse-Systeme : Architektur, Entwicklung, Anwendung](https://hds.hebis.de/thm/Record/HEB323406939)

4. Aufl., überarbeitete und erweiterte Auflage. Heidelberg : dpunkt-Verlag.

*Data Warehousing komplett erklärt – Grundlagen und Anwendung*

[Baun, Christian](https://hds.hebis.de/thm/Search/Results?lookfor0%5b%5d=Baun%2C%20Christian&amp;type0%5b%5d=author&amp;lastposition): [Cloud Computing (2011) : Web-basierte dynamische IT-Services](https://hds.hebis.de/thm/Record/HEB234187778)

2. Aufl. Heidelberg: Springer.

Cloud Computing und damit verbundene Chancen; Risiken etc.

Bitkom (2020): Wie wichtig sind die folgenden Kriterien und Leistungen bei der Auswahl eines Cloud Providers für Ihr Unternehmen? In: Statista, das StatistikPortal. [(https://de.statista.com/statistik/daten/studie/545924/umfrage/kriterien-bei-der-auswahl-eines-cloud-](https://de.statista.com/statistik/daten/studie/545924/umfrage/kriterien-bei-der-auswahl-eines-cloud-providers-in-deutschen-unternehmen/) [providers-in-deutschen-unternehmen/,](https://de.statista.com/statistik/daten/studie/545924/umfrage/kriterien-bei-der-auswahl-eines-cloud-providers-in-deutschen-unternehmen/) abgerufen am 23.08.2020

*Umfrage zur Auswahl von Cloud-Providern*

v

[Finger, Ralf](https://hds.hebis.de/thm/Search/Results?lookfor0%5b%5d=Finger%2C%20Ralf&amp;type0%5b%5d=author&amp;lastposition) (2018): [BI & Analytics in der Cloud: Architektur, Vorgehen und Praxis](https://hds.hebis.de/thm/Record/HEB446238570) Heidelberg : dpunkt.verlag.

*Wie funktioniert BI und Analytics in der Cloud? Grundlagen und Anwendung*

Gartner (2020a): Umsatz der führenden Anbieter mit Infrastructure-as-a-Service (IAAS) weltweit. In: Statista, das Statist Portal. [https://de.statista.com/statistik/daten/studie/754661/umfrage/umsatz-der-](https://de.statista.com/statistik/daten/studie/754661/umfrage/umsatz-der-fuehrenden-anbieter-mit-infrastructure-as-a-service-weltweit/) [fuehrenden-anbieter-mit-infrastructure-as-a-service-weltweit/,](https://de.statista.com/statistik/daten/studie/754661/umfrage/umsatz-der-fuehrenden-anbieter-mit-infrastructure-as-a-service-weltweit/) aufgerufen am 22.08.2020

*Marktvolumen und Anbieter-Vergleich (AWS, Azure etc.)*

Gartner (2020b): Umsatz mit Cloud Computing weltweit von 2010 bis 2019 und Prognose bis 2022 nach Segment. In: Statista, das Statistik Portal. [https://de.statista.com/statistik/daten/studie/284706/umfrage/prognose-zum-umsatz-mit-cloud-computing-](https://de.statista.com/statistik/daten/studie/284706/umfrage/prognose-zum-umsatz-mit-cloud-computing-weltweit-nach-segment/) [weltweit-nach-segment/,](https://de.statista.com/statistik/daten/studie/284706/umfrage/prognose-zum-umsatz-mit-cloud-computing-weltweit-nach-segment/) abgerufen am 22.08.2020

*Umsatzentwicklung des Cloud Computing*

Helmer, Sven (2012): Integrating Star and Snowflake Schemas in Data Warehouses. In: International Journal of Data Warehousing and Mining. [https://www.researchgate.net/publication/262239999\_Integrating\_Star\_and\_Snowflake\_Schemas\_in\_Data](https://www.researchgate.net/publication/262239999_Integrating_Star_and_Snowflake_Schemas_in_Data_Warehouses)

[\_Warehouses,](https://www.researchgate.net/publication/262239999_Integrating_Star_and_Snowflake_Schemas_in_Data_Warehouses) abgerufen am 22.08.2020, abgerufen am 22.08.2020

*Vergleich und Implementierung von Star und Snowflake Schema im DWH*

ISO 25000 (2019): ISO/ IEC [25010. https://iso25000.com/index.php/en/iso-25000-standards/iso-25010](https://iso25000.com/index.php/en/iso-25000-standards/iso-25010),

abgerufen am 05.09.2020

*Qualitätsmerkmale IS0 25010*

Johner Institut (2015): ISO 9126 und ISO 25010. [https://www.johner-institut.de/blog/iec-62304-](https://www.johner-institut.de/blog/iec-62304-medizinische-software/iso-9126-und-iso-25010/) [medizinische-software/iso-9126-und-iso-25010/,](https://www.johner-institut.de/blog/iec-62304-medizinische-software/iso-9126-und-iso-25010/) abgerufen am 22.08.2020

*Qualitätsmerkmale/ Kriterien für Software im Allgemeinen (Sicherheit, Performance, Usability etc.)*

[Köppen, Veit](https://hds.hebis.de/thm/Search/Results?lookfor0%5b%5d=K%C3%B6ppen%2C%20Veit&amp;type0%5b%5d=author&amp;lastposition) (2014): [Data Warehouse: Technologien](https://hds.hebis.de/thm/Record/HEB445404876). 2. Aufl. Heidelberg: mitp.

*Zeigt verschiedene Technologien für DWHs, außerdem werden Grundlagen des DWH vermittelt*

[Reinheimer, Stefan](https://hds.hebis.de/thm/Search/Results?lookfor0%5b%5d=Reinheimer%2C%20Stefan&amp;type0%5b%5d=author&amp;lastposition) (2018): [Cloud Computing: Die Infrastruktur der Digitalisierung](https://hds.hebis.de/thm/Record/HEB428860133) Wiesbaden: Springer Fachmedien.

*Einführung in das Gebiet des Cloud-Computings*

Gupta, A./ Agarwal, D/ Tan, D./ Kulesza, J./ Pathak, R./ Stefani, S./ Srinivasan, V. (2015): Amazon Redshift and the Case for Simpler Data Warehouses. In: Research Gate.

vi

[https://www.researchgate.net/publication/300581416\_Amazon\_Redshift\_and\_the\_Case\_for\_Simpler\_Dat](https://www.researchgate.net/publication/300581416_Amazon_Redshift_and_the_Case_for_Simpler_Data_Warehouses) [a\_Warehouses,](https://www.researchgate.net/publication/300581416_Amazon_Redshift_and_the_Case_for_Simpler_Data_Warehouses) abgerufen am 22.08.2020

*… über AWS und Data Warehousing*

Snowflake (2020): Snowflake Documentation. [https://docs.snowflake.com/en/,](https://docs.snowflake.com/en/) aufgerufen am 25.08.2020

*Dokumentation zu Snowflake, Funktionen, Anwenudng etc.*

Statista Technology Market (2020): Infrastructure-as-a-Service In: Statista, das Statistik Portal. [https://de.statista.com/outlook/16110/100/infrastructure-as-a-service/weltweit,](https://de.statista.com/outlook/16110/100/infrastructure-as-a-service/weltweit) abgerufen am 22.08.2020

*Statistiken zu IaaS, Cloud und Relevanz*

vii