Grundlagen und Anwendungsbeispiele für das NOSQL-Datenbankmanagementsystem DynamoDB

Seminararbeit

vorgelegt am 25. Januar 2017

Fakultät Wirtschaft

Studiengang Wirtschaftsinformatik

Kurs WWI2014I

von

|  |  |
| --- | --- |
| Betreuer in der Ausbildungsstätte: | DHBW Stuttgart: |
| Prof. Dr. Nikolai Preiß DHBW Stuttgart | [Titel, Vorname und Nachname des wissenschaftl. Betreuers/Prüfers] |

Kerstin Farke  
Dhruv Mahandru  
Henning Mohr  
Käthe Vrettos

**Inhaltsverzeichnis**

[Abkürzungsverzeichnis (bei Bedarf) IV](#_Toc469838817)

[Abbildungsverzeichnis (bei Bedarf) V](#_Toc469838818)

[Tabellenverzeichnis (bei Bedarf) VI](#_Toc469838819)

[1 Einleitung 1](#_Toc469838820)

[2 Theoretische Grundlagen 2](#_Toc469838821)

[2.1 Definitionen 2](#_Toc469838822)

[2.1.1 Datenbank 2](#_Toc469838823)

[2.1.2 Datenbankmanagementsystem 2](#_Toc469838824)

[2.1.3 Software-as-a-Service 2](#_Toc469838825)

[2.1.4 Database-as-a-Service 3](#_Toc469838826)

[2.2 Datenbank Prinzipien 3](#_Toc469838827)

[2.2.1 ACID 3](#_Toc469838828)

[2.2.2 BASE 4](#_Toc469838829)

[2.2.3 CAP-Theorem 4](#_Toc469838830)

[2.3 Entwicklungen in der Datenbankbranche 5](#_Toc469838831)

[2.4 Datenbankmodelle 6](#_Toc469838832)

[2.4.1 Relationale Datenbanken 6](#_Toc469838833)

[2.4.2 NoSQL Datenbanken 7](#_Toc469838834)

[2.4.3 Zusammenfassung 9](#_Toc469838835)

[3 DynamoDB 11](#_Toc469838836)

[3.1 Generelle Informationen über AWS 11](#_Toc469838837)

[3.1.1 Infrastructre as a Serivice (IaaS)/Kunden 11](#_Toc469838838)

[3.1.2 Gründe für die Verwendung von AWS/Entwicklungsgeschichte (evtl.) 12](#_Toc469838839)

[3.1.3 Ziele 12](#_Toc469838840)

[3.1.4 Vor- und Nachteile von AWS 12](#_Toc469838841)

[3.2 Was ist DynamoDB 12](#_Toc469838842)

[3.2.1 Database as a Service (DBaaS) Definition 12](#_Toc469838843)

[3.2.2 Geschichte der DynamoDB 12](#_Toc469838844)

[3.2.3 Verwendung der DynamoDB in der Amazon Infrastruktur 12](#_Toc469838845)

[3.2.4 Preismodell 12](#_Toc469838846)

[3.3 Art der Datenverwaltung/Datenmodell 12](#_Toc469838847)

[3.3.1 Tabelle, Elemente und Attribute Konzept 12](#_Toc469838848)

[3.3.2 Eigenschaften (Fully managed, Durable, Scalable, Fast, Simple Administration, Flexible, Fault Tolerance, Indexing, Secure, Cost Effective) 12](#_Toc469838849)

[3.3.3 Datenspeicherung über die SSD 12](#_Toc469838850)

[3.3.4 Key-Value Eigenschaft 12](#_Toc469838851)

[3.3.5 Datentypen 12](#_Toc469838852)

[3.4 Vergleich mit anderes NoSQL-Lösungen 12](#_Toc469838853)

[3.4.1 Microsoft Azure 12](#_Toc469838854)

[3.4.2 Redis 12](#_Toc469838855)

[3.4.3 Vorteile und Nachteile von DynamoDB 12](#_Toc469838856)

[4 Praxis 13](#_Toc469838857)

[4.1 Insgesamt zum Zurechtfinden 13](#_Toc469838858)

[4.2 Testdatenbank mit Testdaten 13](#_Toc469838859)

[4.3 Datenbanksprache 13](#_Toc469838860)

[4.4 GUI/API 13](#_Toc469838861)

[4.5 Ablageform der Daten 13](#_Toc469838862)

[5 Schluss 14](#_Toc469838863)

[5.1 Fazit 14](#_Toc469838864)

[5.2 Ausblick 14](#_Toc469838865)

[Anhang (bei Bedarf) 15](#_Toc469838866)

[Quellenverzeichnis 16](#_Toc469838867)

[Ehrenwörtliche Erklärung 17](#_Toc469838868)

[Literaturverzeichnis 18](#_Toc469838869)

# Abkürzungsverzeichnis (bei Bedarf)

AktG = Aktiengesetz

AWS = Amazon Web Services

BFH = Bundesfinanzhof

# Abbildungsverzeichnis (bei Bedarf)

[Abb. 1: Eine Abbildung 2](#_Toc444006772)

# Tabellenverzeichnis (bei Bedarf)

[Tabelle 1: Eine Beispieltabelle 2](#_Toc444006773)

1. Einleitung
2. Theoretische Grundlagen
   1. Definitionen
      1. Datenbank

Nach Definition von Schicker in seinem Buch „Datenbanken und SQL: Eine praxisorientierte Einführung mit Anwendungen in Oracle, SQL Server und MySQL“ ist eine Datenbank „eine Sammlung von Daten, die untereinander in einer logischen Beziehung stehen und von einem eigenen Datenverwaltungssystem (Database Management System, DMBS) verwaltet werden“.[[1]](#footnote-2) Demnach stellt eine Datenbank eine organisierte Datenkollektion dar, welche den Zugriff zu Inhalten einfacher gestaltet.[[2]](#footnote-3)

(aus meiner PA kopiert 🡪 muss noch bearbeitet werden!)

* + 1. Datenbankmanagementsystem

Nach Leimeister besteht „[j]edes Datenbanksystem [..] in seiner Grundstruktur aus einem Datenbankmanagementsystem und mehreren, untereinander verknüpften Daten, der eigentlichen Datenbank“[[3]](#footnote-4). Das DBMS beinhaltet verschiedene Funktionen, unter anderem den Auf-bau, die Verwaltung und die Manipulation der Daten einer Datenbank, sowie die Steuerung und Kontrolle aller peripheren Geräte, das Belegen von Speicherbereichen und die Regelung von Datenkollisionen. Letzteres geschieht in Zusammenarbeit mit dem Betriebssystem.[[4]](#footnote-5)

(aus meiner PA kopiert 🡪 muss noch bearbeitet werden!)

* + 1. Software-as-a-Service

Die Art und Weise wie Software vertrieben wird hat sich durch die Popularität des Internets verändert. So wird Software nicht mehr gekauft, sondern an verschiedene Nutzer vermietet und über das Internet zur Verfügung gestellt.[[5]](#footnote-6) Software-as-a-Service (SaaS) bezeichnet die Bereitstellung von existierenden Applikationen in einem ganzen oder partiellen Remote-Service.[[6]](#footnote-7) So können Nutzer in Form von beispielsweise web-basierten Applikationen auf die Applikationen des Bereitstellers zugreifen und den Zugang der Software Applikation, welcher mehreren Nutzer gleichzeitig bereitsteht, mittels „pay-per-use“-Basis zahlen.[[7]](#footnote-8)

* + 1. Database-as-a-Service

Database-as-a-Service (DbaaS) stellt die Lieferung einer Datenbanksoftware mit verbundenem physikalischem Datenbankspeicher als Service dar. Als ein Managed Service wird dem Kunden auf einer „pay-per-usage“-Basis ein auf Anfrage bestehender Zugang zu einer Datenbank gegeben.[[8]](#footnote-9) DbaaS stellt Transaktionen und Funktionen auf einem entfernten Datenbankserver zur Verfügung, welcher mit anderen Nutzern genutzt wird, aber logisch wie eine lokale Datenbank funktioniert. Dadurch ist DbaaS eine Spezialisierung von Software-as-a-Service (SaaS).[[9]](#footnote-10)

* 1. Datenbank Prinzipien

Datenbanken folgen unterschiedlichen Prinzipien. So basieren relationale Datenbanken einem anderen Prinzip als beispielsweise NoSQL Datenbanken. Im Folgenden werden die drei meistgenutzten Prinzipien erläutert und differenziert.

* + 1. ACID

Das ACID-Prinzip steht für Atomarität, Konsistenz, Isolation und Dauerhaftigkeit der durchzuführenden Transaktionen,[[10]](#footnote-11) und wird besonders in der SQL Umgebung genutzt.[[11]](#footnote-12)

Die Atomarität steht sicher, dass eine Transaktion ganz oder gar nicht durchgeführt wird. Kann ein Teil der Transaktion also nicht komplett ausgeführt werden, so schlägt es als Gesamtes fehl.[[12]](#footnote-13) Damit wird entweder die gesamte, oder keine, Transaktion ausgeführt.[[13]](#footnote-14) Die Konsistenz sichert, dass die Datenbank sowohl vor als auch nach der Transaktion einen soliden Status vorweisen kann,[[14]](#footnote-15) während die Isolation sicherstellt, dass mehrere Transaktionen, welche zur selben Zeit ausgeführt werden, nicht die jeweils andere Transaktion beeinflussen, und somit isoliert durchgeführt werden. Dementsprechend müssen gleichzeitig ausgeführte Transaktionen „serializable“ sein.[[15]](#footnote-16) Zur Sicherstellung des permanenten Status einer Transaktion wird das ACID-Prinzip der Dauerhaftigkeit genutzt. Sobald eine Transaktion abgeschlossen worden ist und Daten gespeichert wurden, sollten die geschriebenen Daten dauerhaft verfügbar sein, auch bei einem Crash oder Fehler des Systems.[[16]](#footnote-17)

* + 1. BASE

Das BASE-Paradigma steht für „Basically Available, Soft State, Eventual Consistency“ und stellt generell das Gegenstück zu den ACID-Proportionen dar.[[17]](#footnote-18) Die Konsistenz ist nach einer Abfrage nicht mehr in einem „solid state“, sondern in einem „soft state“ und soll nicht direkt nach Ende einer Transaktion erreicht werden, sondern nach Ende einer Operation.[[18]](#footnote-19) Der Fokus der BASE-Prinzipien besteht in der permanenten Verfügbarkeit,[[19]](#footnote-20) und nicht in der stetigen Konsistenz, welche nach dem ACID-Prinzip bestehen soll.

Mit Blick auf NoSQL Datenbanken sollte die Datenbank zwischen ACID und BASE bestehen.[[20]](#footnote-21)

* + 1. CAP-Theorem

Das CAP-Theorem wurde zum ersten Mal von Eric Brewer im Jahr 2000 vorgestellt.[[21]](#footnote-22) Dabei steht das Theorem aus drei verschiedenen Faktoren: Konsistenz (Consistency), Verfügbarkeit (Availability) und Ausfalltoleranz (Partition Tolerance).[[22]](#footnote-23)

Diese Prinzipien sagen Folgendes aus:[[23]](#footnote-24)

1. Die verfügbaren Daten sind auf allen Maschinen gleich und Updates sollen frequent durchgeführt werden
2. Die Daten müssen permanent verfügbar, und ständig zugänglich sein
3. Bei Fehlern der Maschine sollte die Datenbank weiterhin laufen, ohne die Arbeit stoppen zu müssen.



Abb. 1: CAP-Theorem[[24]](#footnote-25)

Das Theorem besagt, dass nur zwei von den drei Aspekten in einem verteilten System ermöglicht werden können. Demnach muss sich im Voraus also für zwei Aspekte entschieden werden.

* 1. Entwicklungen in der Datenbankbranche

Relationale Datenbanken wurden in den 70er Jahren von E.F. Codd[[25]](#footnote-26) mit dem Ziel entwickelt, strukturierte Daten in einer organisierten Art und Weise in Tabellen zu speichern,[[26]](#footnote-27) und die hierarchischen oder netzwerkartigen Datenbanken als Vorläufer der relationalen Datenbanken abzulösen.[[27]](#footnote-28)

Die hauptsächliche Verantwortlichkeit besteht bei relationalen Datenbanken zurzeit in der Datenverarbeitung, besonders in Verbindung mit der Speicherung von finanziellen Informationen und persönlichen Daten.[[28]](#footnote-29) Beschränkungen bestehen jedoch besonders in dem starken Datenwachstum, durch welchen Datenabfragen durch das große Volumen der Daten nicht mehr so effizient gestaltet werden kann. Außerdem zeichnen sich Schwierigkeiten in der Speicherung und dem Management von größeren Datenbanken ab.[[29]](#footnote-30) Viele Unternehmen setzen aus diesen Gründen nicht-relationale Datenbanken, auch unter dem Namen NoSQL Datenbanken, in ihren Betrieben ein.[[30]](#footnote-31)

NoSQL wurde zuerst in den 1998er Jahren von Carlo Strozzi genannt, wuchs aber erst um 2009 zu einem Konkurrenten der relationalen Datenbanken.[[31]](#footnote-32) Besonders mit dem Aufkommen von des Internets und der Vielzahl von webbasierten Anwendungen gewannen die nicht-relationalen Konzepte gegenüber den relationalen an Gewicht.[[32]](#footnote-33)

* 1. Datenbankmodelle
     1. Relationale Datenbanken

Relationale Datenbanken werden primär durch die ACID-Prinzipien repräsentiert. Daraus schließend sollten relationale Datenbanken besonders dann gewählt werden, wenn Konsistenz von hoher Relevanz ist.[[33]](#footnote-34)

[[34]](#footnote-35)Vorteile bestehen besonders in der langzeitigen Nutzung dieses Modells. Als Industriestandard können relationale Datenbanken schon viel Erfahrung abzeichnen. Die Dokumentation von Professionals, welche in ihrer Karriere die gleichen Paradigmen genutzt haben, stellt weiterhin nützliche Hilfestellung bei der Entwicklung von Datenbanken dar. Ein weiterer Vorteil von relationalen Datenbanken ist die gegebene Reife. Global Player wie International Business Maschine (IBM) und Microsoft, sowie die Open-Source Community bieten stetige Verbesserungen zu dem bestehenden Datenbankmodell und vereinfacht somit die Nutzung. Auch die hohen Levels an Support und die Möglichkeit komplexe Joins zu kreieren, damit Datenbank Administratoren tiefe und komplexe Abfragen implementieren können, stellen durchaus positive Aspekte der relationalen Datenbanken dar.

Negative Aspekte bestehen besonders in der Pflicht, Daten vor der Speicherung definieren zu müssen.[[35]](#footnote-36) Weiterhin unterstützen relationale Datenbanken nicht die hohe Skalierbarkeit, soweit nicht eine bessere Hardware zu diesem Zweck implementiert werden kann.[[36]](#footnote-37) Ein weiterer Defizit besteht in der Abfragesprache in Form von SQL, welche bei strukturierten Daten zwar effizient ist, bei unstrukturierten Daten jedoch höchst komplex wird.[[37]](#footnote-38)

* + 1. NoSQL Datenbanken

NoSQL Datenbanken existieren in dem Spektrum zwischen den ACID und BASE-Prinzipien[[38]](#footnote-39) und werden primär durch die fehlenden Relationen zwischen den verschiedenen Datensätzen charakterisiert.[[39]](#footnote-40) NoSQL Implementation baiseren auf Entitäten und unterstützen die Funktionen von relationalen Datenbankmanagementsystemen wie das Sortieren, Indexen, Projektieren und Abfragen von Daten. Zusätzlich werden Joins und ACID Garantien für eine höhere Transaktionsgeschwindigkeit eingetauscht.[[40]](#footnote-41) Popularität für diesen alternativen Ansatz wurde durch die einfache Skalierbarkeit, sowie einer verbesserten Sicherheit gewonnen.[[41]](#footnote-42)

Meier und Kaufmann erläutern in ihrem Buch „SQL- & NoSQL-Datenbanken“, dass folgende Bedingungen für eine NoSQL Datenbank vorliegen müssen[[42]](#footnote-43):

1. Das zugrundeliegende Datenbankmodell ist nicht relational
2. Das Datenbanksystem erfüllt die Anforderungen für umfangreiche Datenvolumen, flexible Datenstrukturen und Echtzeitverarbeitung, auch bekannt unter den drei V’s (Volume, Variety und Velocity)
3. Das Datenbankschema unterliegt keinem fixen Schema
4. Die Datenbankarchitektur unterstützt horizontale Skalierbarkeit und verteilte Webanwendungen
5. Das Datenbanksystem unterstützt Datenreplikation
6. Durch das CAP-Theorem ist Konsistenz nur verzögert gewährleistet, falls hohe Verfügbarkeit und Ausfalltoleranz gewährleistet angestrebt wird

Generell kann man bei NoSQL Datenbanken zwischen vier verschiedenen Arten differenzieren:

* Key-Value Store
* Document Store
* Column Store
* Graphendatenbank

Bei dem Key-Value Store werden alle gelagerten Daten durch ein „key-value pair“ je Datei repräsentiert. Damit ist jeder Schlüssel einzigartig und erlaubt es die Datensatzinformationen als Werte anzufordern. Diese Struktur ist auch unter dem Namen „Hash Table“ bekannt, in dem Datenabrufe mit Hilfe des Schlüssels durchgeführt werden.[[43]](#footnote-44) Zudem können in dem System des Key-Value Stores sowohl strukturierte als auch unstrukturierte Daten gespeichert werden.[[44]](#footnote-45)

Die Document Store Datenbanken sind dafür ausgelegt, die gespeicherten Daten, welche in Dokumenten verschiedener Formate, wie beispielsweise XML oder JSON, gespeichert werden, zu handhaben.[[45]](#footnote-46) Der Nutzer kann eine unbegrenzte Anzahl von Felder jeder Länge zu einem Dokument hinzufügen[[46]](#footnote-47) und ist daher komplexer als der Key-Value Store.[[47]](#footnote-48)

Der Column Store besitzt eine Struktur, welche ähnlich zu der der relationalen Datenbank ist. Hier werden alle Daten in Reihen und Spalten gespeichert. Spalten, welche ähnliche Daten besitzen und vermehrt zusammen abgerufen werden, können zudem gruppiert werden.[[48]](#footnote-49) Der Unterschied zu relationalen Datenbanken besteht in dem Sachverhalt der unstrukturierten Daten. Während Datensätze in relationalen stark strukturiert sein müssen, so müssen diese in dem Column Store nur eng miteinander verwandt sein.[[49]](#footnote-50)

Graphendatenbanken werden genutzt, wenn gespeicherte Daten mit verflochtenen Elementen wie Social Networking oder Road Maps, als Graph dargestellt werden sollen.[[50]](#footnote-51)

[[51]](#footnote-52)Vorteile der NoSQL Datenbanklösungen bestehen zum einem in einem einfachen Applikationsmanagement, welches die Notwendung von Applikationsänderungen oder Datenbankschema-Änderungen nimmt. Weiterhin bieten NoSQL Datenbanken die Möglichkeit, Clusters und Nodes transparent zu nutzen, ohne dass das Datenbankadministrationsmanagement oder eine manuelle Distribution von Informationen benötigt wird. Zudem sind NoSQL Datenbanken dafür ausgelegt, Daten automatisch zu administrieren und wiederherzustellen.

Wird die Literatur zu NoSQL Datenbanken betrachtet, so wird als Vorteil primär die Skalierbarkeit in den Vordergrund gezogen. Die horizontale Skalierbarkeit erlaubt es zum einen Anfragen durch niedrig-preisliche Server zufriedenzustellen, wodurch ein Unternehmen seine Kosten verringern kann.[[52]](#footnote-53) Außerdem können einzelne Dokumente oder „key-value pairs“ zwischen verschiedenen Servern verschoben werden, ohne dass die Datenbank als Ganzes an Integrität verliert. Dieser Sachverhalt vereinfacht die horizontale Skalierbarkeit und Datenredundanz.[[53]](#footnote-54)

Nachteilig zu sehen ist die unzuverlässige Natur der Datenverfügbarkeit.[[54]](#footnote-55) Die letztendliche Konsistenz bedeutet, dass zwei Personen mit verschiedenen Kopien des gleichen Dokuments eine aktualisierte und nicht-aktualisierte Version eines Dokuments sehen können, wenn kurz zuvor eine write-Operation durchgeführt worden ist.[[55]](#footnote-56) Eine weitere Schwäche der NoSQL Datenbanken besteht in der Datenredundanz und der Speicherung unstrukturierter Daten, welche eine finanzielle Belastung für das Unternehmen darstellen kann.[[56]](#footnote-57) Weiterhin ist die Technologie der NoSQL-Datenbanken noch sehr jung und dadurch teilweise fehlerhaft.[[57]](#footnote-58) Eines der größten Schwachstellen ist jedoch das CAP-Theorem, bei welchem zwischen zwei der drei Eigenschaften gewählt werden muss.[[58]](#footnote-59)

* + 1. Zusammenfassung

Mahamed, Altrafi und Ismail fassen in ihrem Zeitschriftenartikel die Unterschiede von relationalen Datenbanken und noSQL Datenbanken nach verschiedenen Kriterien zusammen. In Form einer Tabelle können diese wie folgt dargestellt werden:

|  |  |  |
| --- | --- | --- |
|  | **Relationale Datenbanken** | **NoSQL Datenbanken** |
| **Transaktionsausfallsicherheit** | Garantiert sehr hohe Transaktionsausfallsicherheit | Spanne von BASE und ACID |
| **Datenmodell** | Daten als Tupel repräsentiert und durch Relationen gruppiert  Die Relationen beinhaltet Tupel (Reihen), mit Attributen (Spalten) Attribut kann durch Domain (Menge von Werten) angesprochen werden | Nehmen Modelliertechniken wie Key-Value Stores, Graphen- oder Dokumenten Stores.  NoSQL Datenbanken kann aus einem oder mehreren Datenmodellen bestehen |
| **Skalierbarkeit** | vertikal | horizontal |
| **Cloud** | Nicht gut für den Cloud-Betrieb geeignet, da sie nicht wichtige Funktionen nicht unterstützen und die nur bis zu einem bestimmten Limit skaliert werden kann | Beste Lösung für Cloud Datenbanken, da alle Charakteristiken, welche NoSQL definieren, in der Cloud Umgebung gewünscht werden |
| **Big Data Handling** | Lösung ist Skalierbarkeit und Datenverteilung in Form von horizontaler oder vertikaler Skalierbarkeit – nur begrenzt möglich | Entworfen um Big Data zu handhaben – vorhandene Methoden erleichtern und verbessern die Performance vom Speichern und Auslesen der Daten |
| **Data Warehouse** | Über Zeit nimmt die Anzahl der gespeicherten Daten stark zu, wodurch ein Big Data Problem entsteht – führt zu Performance Abbau | Design fokussiert sich auf hohe Performance, Skalierbarkeit, Verfügbarkeit und das Speichern von Big Data – dadurch eine gute Lösung für Data Warehouses |
| **Komplexität** | Erhöht sich, da Nutzer Daten in konvertieren müssen, damit sie in die Tabellen passen | Haben die Fähigkeit unstrukturierte und semi-strukturierte Daten zu speichern |
| **Crash Recovery** | Crash Recovery durch den Recovery Manager, welcher dafür verantwortlich ist sicherzustellen, dass Transaktionen atomar und dauerhaft sind [durch log-files und ARIES algorithmus] | Benötigt Replikationen als Backup, um von einem Crash zu recovern |
| **Sicherheit** | Hat sehr sichere Mechanismen um Security Services bereitzustellen – trotzdem gibt es Sicherheitsrisiken wie SQL Injections, Cross Site Scripting, Root Kits etc. | Wurde eher für die Verbesserung von Performance und Big Data auf den Markt gebracht, nicht um Sicherheitslücken zu füllen |

Tabelle 1: Relationale vs. NoSQL Datenbanken[[59]](#footnote-60)

1. DynamoDB

Das zweite theoretisch ausgelegte Kapitel beschäftigt sich mit der Einführung in die *Amazon Web Services* und stellt im weiteren Verlauf die DynamoDB im Detail vor. Der Einstieg widmet sich den Hintergründen der *Amazon Web Services*, um ein grobes Verständnis für die Infrastruktur zu erzeugen. Anschließend sollen grundlegende Informationen zu DynamoDB hinterlegt werden. In Folge dessen gilt es, die Art der Datenverwaltung und das Datenmodell der NoSQL-Datenbank zu erläutern und zu untersuchen. Abschließend soll ein Vergleich mit alternativen NoSQL-Lösungen die Einschränkungen bzw. die besonderen Merkmale von DynamoDB ausarbeiten.

* 1. Generelle Informationen über AWS

Die DynamoDB ist ein Teil der *Amazon Web Services* und daher ist es relevant, die Merkmale und Besonderheiten der Infrastruktur zu erläutern, bevor eine Einführung in Datenbanklösung gleistet werden kann. Das folgende Unterkapitel ist auf die Einführung in AWS ausgerichtet. Im Verlauf dieses Unterkapitels werden die Ziele und Absichten der Dienstleistung aufgezeigt. Im Anschluss folgt eine Gegenüberstellung der Vorteile und Nachteile der Infrastruktur.

* + 1. Infrastructre as a Serivice (IaaS)/Kunden

Das Modell von *Infrastructure as a Service* stellt laut einem Artikel von Wolfgang Sommergut „die IT-Ressourcen, wie Rechenleistung, [Speicherplatz] oder Netzwerkkapazitäten zur Verfügung“[[60]](#footnote-61). Außerdem habe der Anwender die Chance, das Betriebssystem, sowie die Applikationen eigenständig zu steuern und die Infrastruktur nach eigenem Bedarf zusammenzustellen. Im Wesentlichen handelt es sich bei AWS also um ein *IaaS,* die die Möglichkeiten zur flexiblen Datenspeicherung und den Ausbau der verfügbaren Rechenleistung bereitstellt. Neben den genannten Ausprägungen, bietet AWS eine Vielfalt von Ressourcen an, die in Tabell 1 zusammengefast sind.

|  |  |
| --- | --- |
| **IT-Ressource** | **Beschreibung** |
| Online-Speicher | Es handelt sich hierbei um Cloud-Speicher. |
| Near-line Speicher | Speichermedien mit längeren Zugriffszeiten |
| Server | Server zum Bereitstellen von diversen Diensten |
| Amazon Simple Queue Services | Dienste zur Nachrichtenübermittelung |
| Datenbanksystem | NoSQL-Lösungen zur Speicherung verschiedener Datenformate |

Tabelle 2 Einige IT-Ressourcen von AWS[[61]](#footnote-62)

* + 1. Gründe für die Verwendung von AWS/Entwicklungsgeschichte (evtl.)

Die in 2006 gegründete Infrastruktur gehört zu den ersten seiner Art und ist mittlerweile Marktführer (siehe Abbildung 1) in diesem Bereich. Aus diesem Grund verlassen sich ebenfalls Plattformen, wie Netflix, AirBnB und Spotify auf die Dienstleistungen des Anbieters. Angesichts der soliden Etablierung im Markt, ist das Unternehmen in der Lage vielseitige Servicepakete und Ressourcen bereitzustellen. . Im Vergleich zu den Wettbewerbern, könne Amazon bis zu 10-mal mehr Kapazitäten bereitstellen[[62]](#footnote-63).

Abbildung 1 IaaS Martkanteile[[63]](#footnote-64)

Laut Linton sind die „Flexibilität“ und die „Kontrolle“ die größten Vorteile der *Amazon Web Services*. Dabei habe der Anwender die Möglichkeit Ressourcen beliebig zu reduzieren oder zu erweitern. Für den Fall, dass die Dienstleistungen nicht zufriedenstellend sind, habe der Anwender die Möglichkeit die Applikationen auf einer alternativen Plattform zu basieren. Zusätzlich bietet AWS mit seinen 15 geographischen Standorten einen hohen Grad an Verfügbarkeit und Sicherheit für die Kunden[[64]](#footnote-65). Das Preismodell von AWS richtet sich auf den Bedarf der Ressourcen. Hierbei können Kapazitäten beliebig zur Applikation hinzugefügt werden, wenn diese notwendig sind.

* + 1. AWS Anwendungsbeispiele

Die Infrastruktur erlaubt aufgrund der vielfältigen Dienstleistungspakete eine große Anzahl an Anwendungsbeispielen. Dieses Unterkapitel stellt einige dieser Anwendungsbeispiele vor und ermöglicht somit einen Einblick in die Möglichkeiten, die Kunden mit AWS haben.

**Hosten von Webseiten mit EC2**

Die Serverfarmen der Infrastruktur können für das Hosten von Webseiten verwendet werden.[[65]](#footnote-66) Dies ist besonders für Großhändler lukrativ, da diese mit umfangreiche Internetseiten bspw. auch mit einem Webshop auftreten. Die *Elastic Compute Cloud* Infrastruktur von Amazon fängt hierbei die Anfragen ab und kollaborieren mit einer Datenbank, die weitere Daten zu Produkten enthält**.[[66]](#footnote-67)** Zusätzlich erhält der Kunde professionelle Wartungsmöglichkeiten und technische Beratung, die die Zuverlässigkeit und Verfügbarkeit der Webseiten erhöhen.

**Ausführen von Unternehmens Applikationen**

Hierfür wird die Amazon Cloud eingerichtet und mit dem Unternehmensnetzwerk mit einem Virtual Private Network (VPN) verbunden.[[67]](#footnote-68) Im Weiteren werden Applikationsserver auf virtuelle Maschinen (VM) installiert, auf denen anschließend die Applikation ausgeführt werden kann. Die Applikationen laufen nun auf der Amazon Cloud und können unabhängig von geografischen Grenzen gepflegt und bedient werden.

**Datenspeicherung und Datenarchivierung mit S3**

Neben dem Hosten und Ausführen, bietet AWS die Möglichkeit der dauerhaften Datenspeicherung und Datenarchivierung. Das bekannteste Produkt hierbei ist der *Simple Storage Service*.[[68]](#footnote-69) Dieses Angebot visiert insbesondere die gesetzlichen Vorschriften zur sicheren Datenaufbewahrung für einen gewissen Zeitraum. Dabei werden Daten dezentral gespeichert, um Verluste zu verhindern.

* + 1. Vor- und Nachteile von AWS

**Vorteile**

Die geografischen Distributionen der Datenzentren ermöglichen den Kunden den Standort der gespeicherten Daten selbständig zu wählen, dadurch erhält der Kunde mehr Kontrolle über die Daten. Die Amazon Infrastruktur erweitert das Serviceportfolio kontinuierlich, dadurch kommen neue Servicepakete und Verbesserung zur Cloud stetig hinzu.[[69]](#footnote-70) Ein weiterer Vorteil des Cloud Anbieters ist die Flexibilität und Skalierbarkeit, die es dem Kunden erlaubt Ressourcen arbiträr aufzustellen.[[70]](#footnote-71) Aus diesem Grund zahlt der Kunde ausschließlich für die Ressourcen, die tatsächlich verwendet werden. Aus ökonomischer Perspektive, reduziert die AWS Cloud die Produkteinführungszeit[[71]](#footnote-72). Das bedeutet, wenn ein Anwendungsfall zusätzliche Ressourcen benötigt, können diese in wenigen Minuten vollständig benutz werden.

**Nachteile**

Es besteht ein hoher Grad an Abhängigkeit gegenüber einem Drittanbieter. Im In Bezug auf den Standortwahl der Kunden, ist es schwer nachvollziehen, ob sich die Daten in der Cloud tatsächlich am ausgewählten oder priorisierten Standort befinden. Für den Kunden liefert dies eine Intransparenz. Zusätzlich ist die große Auswahl von AWS für Kunden und Entwickler nicht immer überschaubar[[72]](#footnote-73). Das führt dazu, dass die Dienstleistungen nicht optimal genutzt werden.

Zum Thema Kosten ist AWS in einigen Beriechen teurer als die direkten Wettbewerber, wie Microsoft Azure[[73]](#footnote-74).

* 1. Was ist DynamoDB
     1. Database as a Service (DBaaS) Definition
     2. Geschichte der DynamoDB
     3. Verwendung der DynamoDB in der Amazon Infrastruktur
     4. Preismodell
  2. Art der Datenverwaltung/Datenmodell
     1. Tabelle, Elemente und Attribute Konzept
     2. Eigenschaften (Fully managed, Durable, Scalable, Fast, Simple Administration, Flexible, Fault Tolerance, Indexing, Secure, Cost Effective)
     3. Datenspeicherung über die SSD
     4. Key-Value Eigenschaft
     5. Datentypen
  3. Vergleich mit anderes NoSQL-Lösungen
     1. Microsoft Azure
     2. Redis
     3. Vorteile und Nachteile von DynamoDB

1. Praxis

Im Praxisteil werden folgende Themen behandelt!!!!!!!

* 1. Insgesamt zum Zurechtfinden
  2. Testdatenbank mit Testdaten
  3. Datenbanksprache
  4. GUI/API
  5. Ablageform der Daten

1. Schluss
   1. Fazit
   2. Ausblick

Anhang (bei Bedarf)

# Quellenverzeichnis

**Literaturverzeichnis**

**Preiß, N. (2007):** Entwurf und Verarbeitung relationaler Datenbanken, München/Wien: Oldenbourg

**Steger, J. (2006):** Kosten- und Leistungsrechnung, 4. Aufl., München/Wien: Oldenbourg

**Stoi, R. (2003):** Management und Controlling von Intangibles, in: Studium & Praxis, 4. Jg., Nr. 1, S. 34-46

**Verzeichnis der Internet- und Intranetquellen (bei Bedarf)**

**Gesprächsverzeichnis (bei Bedarf)**

# Ehrenwörtliche Erklärung

Wir versichern hiermit, dass wir unsere Seminararbeit mit dem Thema: [ Thema einfügen ] selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt haben. Wir versichern zudem, dass die eingereichte elektronische Fassung mit der gedruckten Fassung übereinstimmt.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Ort, Datum) (Kerstin Farke)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Ort, Datum) (Dhruv Mahandru)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Ort, Datum) (Henning Mohr)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Ort, Datum) (Käthe Vrettos)

Literaturverzeichnis

Abramova, Veronika; Bernardino, Jorge; Furtado, Pedro (2014): Experimental Evaluation of NoSQL Databases. In: *International Journal of Database Management Systems (IJDMS)* 6 (3), S. 1–16. Online verfügbar unter http://s3.amazonaws.com/academia.edu.documents/44185663/6314ijdms01.pdf?AWSAccessKeyId=AKIAJ56TQJRTWSMTNPEA&Expires=1480620894&Signature=Xuo9y8raSPFKC7sf2q3deopev5o%3D&response-content-disposition=inline%3B%20filename%3DEXPERIMENTAL\_EVALUATION\_OF\_NOSQL\_DATABAS.pdf.

Antonopoulos, N.; Gillam, L. (2010): Cloud Computing: Principles, Systems and Applications: Springer London. Online verfügbar unter https://books.google.de/books?id=SbSbdkqibwIC.

Edward, S. G.; Sabharwal, N. (2015): Practical MongoDB: Architecting, Developing, and Administering MongoDB: Apress. Online verfügbar unter https://books.google.de/books?id=7iI3CwAAQBAJ.

Ellenberg, J. (2014): Preispolitik im Software-as-a-Service Markt: Deskriptive Analyse und Bewertung: Diplomica Verlag. Online verfügbar unter https://books.google.de/books?id=7GWQAwAAQBAJ.

Hammes, Dayne; Medero, Hiram; Mitchell, Harrison (2014): Comparison of NoSQL and SQL Databases in the Cloud. In:. Information Systems Conference. Macon, GA, USA, March 21st - 22nd 2014. Proceedings of the Southern Association for Information Systems Conference. Macon, GA, USA, S. 1–8.

Jatan, Nishtha; Puri, Sahil; Ahuja, Mehak; Kathuria, Ishita; Gosain, Dishant (2012): A Survey and Comparison of Relational and Non-Relational Database. In: *International Journal of Engineering Research & Technology (IJERT)* 1 (6), S. 1–5. Online verfügbar unter http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.678.9352&rep=rep1&type=pdf.

Leimeister, Jan Marco (2015): Einführung in die Wirtschaftsinformatik. 12. Aufl. Berlin [u.a.]: Springer (Springer-Lehrbuch).

Link, Jörg (2013): Customer Relationship Management. Erfolgreiche Kundenbeziehungen durch integrierte Informationssysteme. Berlin, Heidelberg: Springer Berlin Heidelberg.

Mahamed, Mohamed A.; Altrafi, Obay G.; Ismail, Mohammed O. (2014): Relational vs. NoSQL Databases: A Survey. In: *International Journal of Computer and Information Technology* 3 (3), S. 598–601.

Meier, Andreas; Kaufmann, Michael (2016): SQL- & NoSQL-Datenbanken. 8., überarb. u. erw. Aufl. 2016. Berlin, Heidelberg: Springer Berlin Heidelberg (eXamen.press).

Narang, R. (2011): Database Management Systems. o. O.: Prentice-Hall of India. Online verfügbar unter https://books.google.com/books?id=B-cJhhsh4NAC.

Padhy, Rabi Prasad; Patra, Manas Ranjan; Satapathy, Suresh Chandra (2011): RDBMS to NoSQL. Reviewing Some Next-Generation Non-Relational Database's. In: *International Journal of Advanced Engineering Sciences and Technologies* 11 (1), S. 15–30. Online verfügbar unter http://liacs.leidenuniv.nl/~stefanovtp/courses/StudentenSeminarium/Papers/DB/3.IJAEST-Vol-No-11-Issue-No-1-RDBMS-to-NoSQL-Reviewing-Some-Next-Generation-Non-Relational-Database's-015-030.pdf.

Schicker, Edwin (2014): Datenbanken und SQL. Eine praxisorientierte Einführung mit Anwendungen in Oracle, SQL Server und MySQL. 4., überarb. Aufl. 2014. Wiesbaden: Springer Vieweg (SpringerLink : Bücher).

Sharma, Vatika; Dave, Meenu (2012): SQL and NoSQL Databases. In: *International Journal of Advanced Research in Computer Science and Software Engineering* 2 (8), S. 20–27. Online verfügbar unter http://s3.amazonaws.com/academia.edu.documents/33559632/V2I800154.pdf?AWSAccessKeyId=AKIAJ56TQJRTWSMTNPEA&Expires=1480620329&Signature=IgyAiXJM9JNlTsXPAjBJK3dQTUQ%3D&response-content-disposition=inline%3B%20filename%3DSQL\_and\_NoSQL\_Databases.pdf.

Wang, L.; Ranjan, R.; Chen, J.; Benatallah, B. (2011): Cloud Computing: Methodology, Systems, and Applications: CRC Press. Online verfügbar unter https://books.google.de/books?id=8UbNBQAAQBAJ.

1. Schicker 2014, S. 3 [↑](#footnote-ref-2)
2. Vgl. Narang 2011, S. 1 [↑](#footnote-ref-3)
3. Leimeister 2015, S. 86 [↑](#footnote-ref-4)
4. Vgl. Link 2013, S. 184 [↑](#footnote-ref-5)
5. Vgl. Ellenberg 2014, S. 2 [↑](#footnote-ref-6)
6. Vgl. Wang et al. 2011, S. 14 [↑](#footnote-ref-7)
7. Vgl. Wang et al. 2011, S. 14 [↑](#footnote-ref-8)
8. Vgl. Wang et al. 2011, S. 15 [↑](#footnote-ref-9)
9. Vgl. Antonopoulos und Gillam 2010, S. 53 [↑](#footnote-ref-10)
10. Vgl. Schicker 2014, S. 18 [↑](#footnote-ref-11)
11. Vgl. Sharma und Dave 2012, S. 21 [↑](#footnote-ref-12)
12. Vgl. Mahamed et al. 2014, S. 598 [↑](#footnote-ref-13)
13. Vgl. Schicker 2014, S. 18 [↑](#footnote-ref-14)
14. Vgl. Mahamed et al. 2014, S. 598 [↑](#footnote-ref-15)
15. Vgl. Mahamed et al. 2014, S. 598 [↑](#footnote-ref-16)
16. Vgl. Mahamed et al. 2014, S. 598 [↑](#footnote-ref-17)
17. Vgl. Sharma und Dave 2012, S. 21 [↑](#footnote-ref-18)
18. Vgl. Mahamed et al. 2014, S. 598f. [↑](#footnote-ref-19)
19. Vgl. Sharma, Dave 2012, S. 21, vgl. dazu auch Mahamed et. al. 2014, S.598f. [↑](#footnote-ref-20)
20. Vgl. Mahamed et al. 2014, S. 598–599, vgl. dazu auch Sharma, Dave 2012, S.21 [↑](#footnote-ref-21)
21. Vgl. Mahamed et al. 2014, S. 599 [↑](#footnote-ref-22)
22. Vgl. Mahamed et al. 2014, S. 599 [↑](#footnote-ref-23)
23. Vgl. auch im Folgenden Sharma und Dave 2012, S. 21–22 [↑](#footnote-ref-24)
24. Entnommen aus: Edward und Sabharwal 2015, S. 16 [↑](#footnote-ref-25)
25. Vgl. Jatan et al. 2012, S. 2 [↑](#footnote-ref-26)
26. Vgl. Abramova et al. 2014, S. 1 [↑](#footnote-ref-27)
27. Vgl. Meier und Kaufmann 2016, S. 18 [↑](#footnote-ref-28)
28. Vgl. Mahamed et al. 2014, S. 598 [↑](#footnote-ref-29)
29. Vgl. Abramova et al. 2014, S. 1 [↑](#footnote-ref-30)
30. Vgl. Mahamed et al. 2014, S. 598 [↑](#footnote-ref-31)
31. Vgl. Mahamed et al. 2014, S. 598 [↑](#footnote-ref-32)
32. Vgl. Meier und Kaufmann 2016, S. 18 [↑](#footnote-ref-33)
33. Vgl. Abramova et al. 2014, S. 1–2 [↑](#footnote-ref-34)
34. Vgl. Hammes et al. 2014, S. 1 [↑](#footnote-ref-35)
35. Vgl. Abramova et al. 2014, S. 1–2 [↑](#footnote-ref-36)
36. Vgl. Jatan et al. 2012, S. 3 [↑](#footnote-ref-37)
37. Vgl. Jatan et al. 2012, S. 3 [↑](#footnote-ref-38)
38. Vgl. Mahamed et al. 2014, S. 599 [↑](#footnote-ref-39)
39. Vgl. Abramova et al. 2014, S. 1–2 [↑](#footnote-ref-40)
40. Vgl. Hammes et al. 2014, S. 1 [↑](#footnote-ref-41)
41. Vgl. Hammes et al. 2014, S. 1 [↑](#footnote-ref-42)
42. Vgl. Meier und Kaufmann 2016, S. 20 [↑](#footnote-ref-43)
43. Vgl. Abramova et al. 2014, S. 3–4 [↑](#footnote-ref-44)
44. Vgl. Padhy et al. 2011, S. 16 [↑](#footnote-ref-45)
45. Vgl. Abramova et al. 2014, S. 3–4 [↑](#footnote-ref-46)
46. Vgl. Padhy et al. 2011, S. 16 [↑](#footnote-ref-47)
47. Vgl. Abramova et al. 2014, S. 3–4 [↑](#footnote-ref-48)
48. Vgl. Abramova et al. 2014, S. 3–4 [↑](#footnote-ref-49)
49. Vgl. Padhy et al. 2011, S. 16 [↑](#footnote-ref-50)
50. Vgl. Abramova et al. 2014, S. 3–4 [↑](#footnote-ref-51)
51. Vgl. Abramova et al. 2014, S. 3–4 [↑](#footnote-ref-52)
52. Vgl. Abramova et al. 2014, S. 3–4 [↑](#footnote-ref-53)
53. Vgl. Hammes et al. 2014, S. 3 [↑](#footnote-ref-54)
54. Vgl. Jatan et al. 2012, S. 4–5 [↑](#footnote-ref-55)
55. Vgl. Hammes et al. 2014, S. 3–4 [↑](#footnote-ref-56)
56. Vgl. Hammes et al. 2014, S. 3–4 [↑](#footnote-ref-57)
57. Vgl. Hammes et al. 2014, S. 3–4 [↑](#footnote-ref-58)
58. Vgl. Hammes et al. 2014, S. 3–4 [↑](#footnote-ref-59)
59. Mahamed et al. 2014, S. 600 [↑](#footnote-ref-60)
60. Artikel aus Computerwochen von Wolfagang Sommergut

    http://www.computerwoche.de/a/was-sie-ueber-die-cloud-wissen-muessen,2504589,2 [↑](#footnote-ref-61)
61. # Vgl. Linton 2011,S.8

    [↑](#footnote-ref-62)
62. Matthew Finnegan (2016): http://www.computerworlduk.com/it-vendors/microsoft-azure-vs-amazon-aws-public-cloud-comparison-which-cloud-is-best-for-enterprise-3624848/ [↑](#footnote-ref-63)
63. Statista 2015 [↑](#footnote-ref-64)
64. Vgl. Wittig 2016, S.5 [↑](#footnote-ref-65)
65. Vgl. Wittig 2016, S.6a [↑](#footnote-ref-66)
66. Vgl. Jeff 2010, S. 99 [↑](#footnote-ref-67)
67. Vgl. Wittig 2016, S.6b [↑](#footnote-ref-68)
68. Vgl. Jeff 2010, S. 59 [↑](#footnote-ref-69)
69. Vgl. Wittig 2016, S.10 [↑](#footnote-ref-70)
70. Vgl Narula et.al 2015, S.3 [↑](#footnote-ref-71)
71. Vgl. Wittig 2016, S.12 [↑](#footnote-ref-72)
72. Vgl. Bargavi und Scharma 2012, S.6 [↑](#footnote-ref-73)
73. Vgl. Madhuri und Sowjanya 2016, S.6 [↑](#footnote-ref-74)