Technology
Arts Sciences
TH Köln

Technische Hochschule Köln
Fakultät für Informatik und Ingenieurwissenschaften

BACHELORARBEIT

Kostenüberwachung und -optimierung für Cloud-Dienste am Beispiel von Amazon Web Services

Vorgelegt an der TH Köln Campus Gummersbach
im Studiengang Wirtshaftsinformatik

ausgearbeitet von:
Carlo Menjivar 11117929

Erstprüfer: Prof. Dr. Roman Majewski
Zweitprüfer: Thomas Raser

Gummersbach, im Feb<Monat der Abgabe>

Abstract

Zusammenfassung

Diese Arbeit beschäftigt sich damit, wie mehr Kontrolle über die Kosten von Cloud-Diensten erhalten wird, indem sie überwacht werden. In Kombination damit werden Maßnahmen und Werkzeuge untersucht, die zu erheblichen Kosteneinsparungen in der Cloud führen.
Angefangen bei der Wahl des richtigen Zahlungsmodells, über das automatische Herunterfahren ungenutzter Instanzen zu bestimmten Zeiten bis hin zur Implementierung von Autoscaling .

Die Arbeit ist auf der Grundlage von Empfehlungen von Amazon Web Services selbst, Erfahrungen von Experten in dem Fachgebiet und aktuelle Fachliteratur geschrieben.

Diese Arbeit ist für Nutzer von Cloud-Diensten relevant, die den Wechsel von klassischen Modellen bekannt als On-Premise zu On-Demand in der Cloud basierten Modelle planen und die unvorhersehbaren Kosten fürchten, die sich ihrer Kontrolle entziehen können. Es ist besonders interessant für Teams, die Cloud-Dienste in aktuellen Projekten verwalten und ihre Kosten optimieren wollen. Wenn die Kosten für Cloud-Dienste wie alle anderen Kosten betrachtet werden, ist es nur konsequent, über ihre Kontrolle und Optimierung nachzudenken.

Abstract

Platz für das englische Abstract...

Inhaltsverzeichnis

Abbildungsverzeichnis

Abkürzungsverzeichnis

aws Amazon Web Services

API Application Programming Interface

CI/CD Continuous Integration / Continuous Deployment

TCO Total Cost of Ownership

1 Einleitung

1.1 Motivation

Amazon Web Services, kurz AWS, wurde unter anderem für diese Arbeit ausgewählt wegen seiner frühen Präsenz (2006) als Cloudanbieter und seines großen Angebotes an Dienstleistungen, welche für zahlreiche Anwendungsfälle geeignet sind.
Eine Recherche von Gartner positioniert AWS als Marktführer in der Magic Quadrant für Cloud-Infrastruktur und Plattform-Services 2021. [25] Kostenoptimierung für Cloud-Dienste ist ein wichtiger Punkt, da man ohne Optimierungsmaßnahmen mit höheren Kosten rechnen muss als bei On-Premise Systemen.

> "Indeed, if you run the cloud the same way you run your on-premise data center, you are almost certain to incur higher expenses. It is necessary to use the following key cloud cost optimization techniques in order to successfully save money on the cloud." [1]

1.2 Problemstellung

Die Verwendung von Cloud-Diensten bringt viele Vorteile mit sich. Zum Beispiel kurzfristige Erhöhung oder Verringerung der Speicher- und Rechenkapazität, sowie Zugriff auf unterschiedliche Speicherarten, die genau an individuelle Anwendungsfälle angepasst sind. All diese Lösungen sind in wenigen Minuten einsatzfertig.

In einer Umfrage haben circa 50% der Unternehmen die Verwaltung der Kosten für den Betrieb von Cloud-Workloads als großes Hindernis genannt. Mehr als die Hälfte der Befrachter haben geäußert, Schwierigkeiten zu haben, alle Kosten für Cloud-Workloads zu rechtfertigen.

> „In its Stratecast Predictions 2018, Frost & Sullivan noted that 53% of IT leaders surveyed cited "managing costs to run cloud workloads" as a huge obstacle, and over 50% have difficulty justifying the expenses of some public cloud workloads." [27]

Diese Bachelorarbeit beschäftigt sich mit ebendieser Problematik, um herauszufinden, wie Unternehmen mit den passenden Werkzeugen die Kosten ihrer Cloud-Dienste überwachen und im Blick behalten können.
Zum Beispiel können frühzeitige Benachrichtigungen alarmieren, wenn Ressourcen mehr Kosten verursachen als geplant.

[1] [2], Seite 152

Außerdem sollte untersucht werden, wie sie mit der richtigen Auswahl an Diensten ihre Kosten optimieren können.

1.3 Fragestellung

In dieser Arbeit wird versucht, die folgenden Fragen beantworten.

- Wie können Kosten bei Cloud-Diensten überwacht werden und wie lassen sie sich optimieren? Am Beispiel von S3 Speichereinheiten und EC2-Instanzen.
- Welche Maßnahmen sind nötig, um unerwartet hohe Kosten bei Cloud-Diensten zu vermeiden.
- Was kann automatisiert werden, um Kosten zu vermeiden, die Nutzer von Cloud-Diensten verursachen.

Meine Hypothese ist, dass Kosten von Cloud-Diensten unter Kontrolle gehalten und reduziert werden können, wenn Überwachungs- und Optimierungswerkzeuge eingesetzt werden.

1.4 Zielsetzung

Daraus ergeben sich für die Arbeit die folgenden Ziele:

- Als Erstes wird gezeigt, wie mithilfe von bestehenden Werkzeugen die Kosten von Cloud-Diensten überwacht werden können.
- Als Nächstes wird anhand von Empfehlungen von Cloud-Experten identifiziert, welche Optimierungsmöglichkeiten bestehen.

1.5 Einschränkungen

Der Schwerpunkt dieser Arbeit liegt auf EC2-Instanzen, da diese in der Regel den größten Anteil an der Rechnung ausmacht. An zweiter Stelle stehen S3-Speichereinheiten, weil sie einen erheblichen Teil der Kosten darstellen.
[9, 10]
Diese Arbeit legt den Fokus auf die Optimierung der oben genannten Dienste. Als Überwachungswerkzeuge für die Kosten werden die AWS CloudWatch, der AWS Cost-Explorer und der AWS Trusted Advisor untersucht.

1.6 Struktur der Arbeit

Diese Bachelorarbeit ist in folgende Kapitel unterteilt:

Kapitel 2 befasst sich mit dem Begriff Cloud-Economy und erläutert das Nutzen der Cloud im wirtschaftlichen Sinne. Diese dienen als Grundlage für diese Arbeit.

Kapitel 3 zeigt die verschiedenen Zahlungsmodelle für Amazon Web Services.
Es werden Kriterien vorgestellt, die helfen, sich für das richtige Zahlungsmodell bei verschiedenen Szenarien zu entscheiden.

In Kapitel 4 werden die Werkzeuge eingeführt, die zur Überwachung der Kosten von Cloud-Diensten eingesetzt werden können.

Kapitel 5 befasst sich mit Optimierungsmaßnahmen insbesondere für EC2-Instanzen und S3 Speichereinheiten.

2 Grundlagen

In diesem Grundlagenkapitel werden Erfolgschancen für Unternehmen aufgelistet, die Cloud-Dienste in ihre Geschäftsprozesse integrieren. Es wird ebenfalls erklärt warum die Kostenoptimierung und -überwachung relevant für Unternehmen sind.

Folgende Ergebnisse können erreicht werden durch die Einführung von Überwachungs- und Optimierungsmaßnahmen:

- Die Möglichkeit, die Kosten verschiedener Projekte, die über dieselbe Infrastruktur laufen, zu trennen. Auf diese Weise kann zwischen Projekte, die mehr und Projekte, die weniger Ressourcen verbrauchen unterschieden werden.
- Eine beachtliche Erhöhung der finanziellen Rentabilität im Unternehmen.
- Eine geringere Ungewissheit bei der Umsetzung von cloudbasierten Systemen.
- Mehr Kontrolle über die Gesamtkosten des Betriebs (TCO) [2].

2.1 Ökonomie des Cloud Computing

[Date last review: 05.11 Sarah]

Cloud Economics auf Englisch, untersucht die Kosten und die Vorteile von Cloud Computing und der dahinter stehenden wirtschaftlichen Grundsätze. Basierend auf dem On-Demand Prinzip, besitzt die Flexibilität, die Rechenkapazität je nach Bedarf anzupassen. Es entfällt die Notwendigkeit, hohe Investitionen in Hardware zu tätigen, wie bei On-Premise-Systemen. Durch den Verzicht auf Hardware entfallen die Kosten für Reparatur, Wartung und eventuell damit verbundene Lizenzen.
Der Cloud-Anbieter übernimmt viele Verwaltungsaufgaben. Das führt zu einer Abnahme der nötigen Fachkraft. [24]
Die Nutzung von Cloud-Diensten is in unabhängiger Weise möglich. Mit anderen Worten in Selbstbedienung und mit der Freiheit, Ressourcen ohne Einschränkungen zu nutzen. Das bedeutet jedoch gleichzeitig, dass der Nutzer Verantwortung für die anfallenden Kosten übernimmt[3].

[Grafik der Kosten On-Premise/Demand?]

[2] TCO: Total Cost if Ownership
[3] Nutzer von Cloud-Diensten

2.1.1 Skalierbarkeit

Um die Leistung der Ressourcen aufrecht zu halten und bei Abnahme der Nachfrage diese zu reduzieren, ist es möglich zum Beispiel die Rechenkapazität hoch und runter zu skalieren.

Mit Auto Scaling wird sichergestellt, dass die Rechenkapazität in Zeiträumen von hoher Nachfrage automatisch hochskaliert.[AUCH RUNTER?] Auf diese Weise kann Zeit mit der Verwaltung von IT - Ressourcen gespart werden, um sich auf die wesentlichen Geschäftsaktivitäten zu konzentrieren[4].

Dies war der Fall bei Walgreens 2020 in den Vereinigte Staaten. Sie haben unter anderem 750 virtuelle Maschinen und SAP HANA auf Azure Instanzen migriert.

> „By getting out of the business of managing datacenters, WBA[Walgreens Boots Alliance] can spend less time worrying about managing IT resources and more time focusing on what it's really good at—delivering great healthcare and retail experiences to its customers. Azure also gives WBA an opportunity to better utilize the capabilities of its SAP implementation. "One of the key reasons for moving to Azure was so that we could take advantage of the scalability that SAP HANA is capable of," explains Regalado. "Instead of using extremely big SAP HANA Large Instances, we can start using smaller VMs[virtuelle Maschinen] and then scale out.„
>
> [21]

2.1.2 Flexibilität und Agilität

In den Amazon Web Services gibt es im Allgemeinen eine Auswahl zwischen folgenden Optionen:

- Verschiedene Betriebssysteme, ohne oder mit Lizenzierung.
- Die meistverbreiteten Programmiersprachen, unter anderem Java, C++, Go, JavaScript und Python.[4]
- Hosting für statische Webseiten und Webanwendungen. [5]
- Populäre relationale und nicht relationale Datenbanken. [11]
- Vielfältige Hardware-Konfigurationen.

[4] [20], Seite 29

Durch die Vielzahl der verfügbaren Ressourcen ist es möglich, Prototypen und Experimente in kurzer Zeit durchzuführen. [5].
Softwareprojekte können schnell auf den Markt gebracht werden.
Sollte ein Projekt kurzfristig stillgelegt werden, könnten alle damit verbundenen Kosten ausfallen.[WEIL...]

2.1.3 Selbstbedienung

Mit geringem Aufwand ist es möglich, Cloud-Dienste eigenständig einzurichten. Dies hat den Vorteil, dass keine weiteren Personen wie externe Spezialisten benötigt werden. Andererseits besteht die Gefahr, dass hohe ungewollte Kosten entstehen, wenn jemand versehentlich oder in unverantwortlicher Weise Dienstleistungen in Anspruch nimmt.

TODO: LOOK FOR A USE CASE WHEREE THIS HAPPEND BRINGT DIESE UNTERKAP. ETWAS ZUR ARBEIT BEI?

2.1.4 Keine Vorabkosten

Amazon Web Services bietet ein Pay-as-you-go-Modell für viele ihre Ressourcen. Wenn nur für die monatlich verbrauchten Ressourcen bezahlt wird, verringert sich oder sogar fällt die Anfangsinvestition in die IT-Infrastruktur weg. Es ist zu bedenken, dass weitere Investitionen wie technische Schulung für das Personal erforderlich werden.[WAS KOSTET EINE IT-INFRA./SERVER+Rack usw]

2.2 Server und Speicher ?

SOLLTEN BEIDE SUBSECTIONS IN EINER SUBSUBSECTION ERKLÄRT WERDEN?

2.2.1 Amazon Elastic Computing Instances EC2

WARUM WURDE DIESE AUSGEWÄHLT? WEIL DIESE CIRCA 80% der Rechnungen ausmachen.

2.2.2 Amazon Simple Storage Service / S3-Speichereinheiten ?

S3 ist der Speicherdienst für Objekte. Ein Objekt ist in AWS die Grundeinheit in den Dateien in der S3-Speichereinheiten gespeichert werden.
Neben die Objekte werden Metadaten wie das Datum der Objekterstellung, Datum der letzten Aktualisierung gespeichert. Laut der Rangliste viele Informatikwebseiten und der

[5] [24], Seite 7

AWS Solutions Architect Daniel Peña Silva[29] ist Amazon S3 ist einer der an den häufigsten genutzten AWS-Services. BESSER: S3 WIRD IN BÜCHER WIE t.ly/IJc1 GENANNT? THIS IS A SPANISH CITAT!

Darüber hinaus werden seit 2020 weltweit mehr Daten in Serverfarmen als auf lokalen Geräten gespeichert. Dies bietet Vorteile in Bezug auf die Geschwindigkeit der Arbeitsabläufe, birgt aber auch Risiken wie Datendiebstahl.

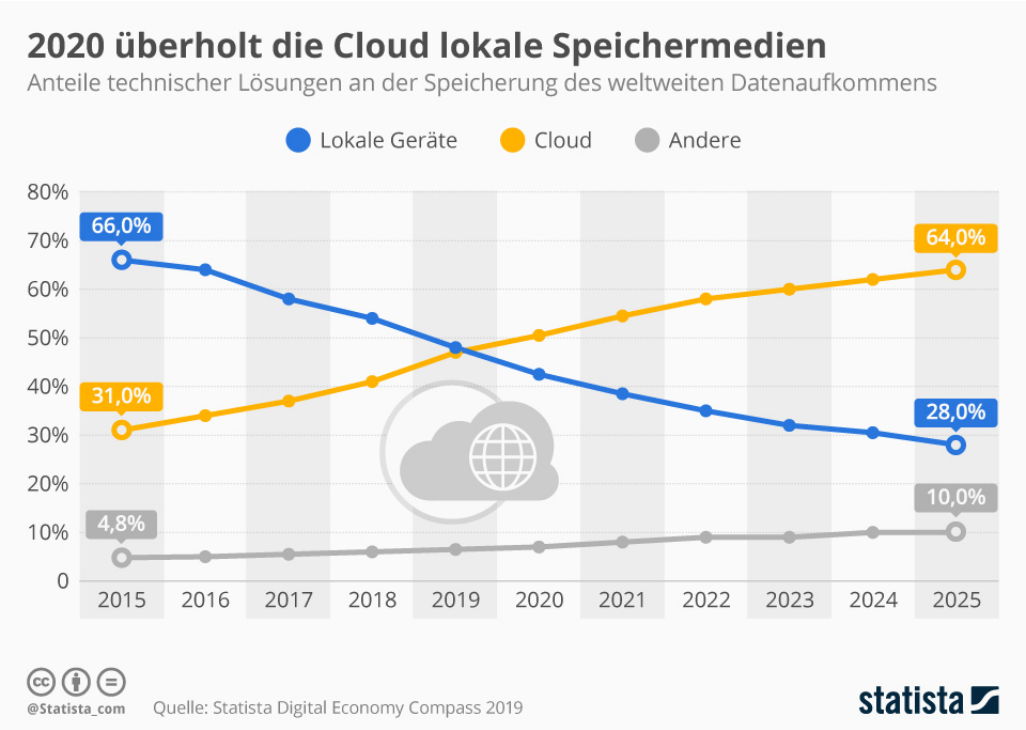

Abbildung 1
2020 überholt die Cloud lokale Speichermedien [29]

[ABSCHLUSS DES KAPITELS]

3 Zahlungsmodelle

Die Nutzung von EC2-Instanzen ist mit einem Zahlungsmodell verbunden. Die Wahl des Zahlungsmodells ist von entscheidender Bedeutung, um den besten Preis für EC2-Instanzen zu erzielen. Die von Amazon Web Services angebotenen Zahlungsmodelle werden im Folgenden dargestellt.

Das On-Demand-Modell beinhaltet keine langfristigen Verpflichtungen, sie ist daher die teuerste Alternative, die auf Stundenbasis berechnet wird. Die Modelle Saving Plans und Reserved Instances erfordern den Abschluss von Verträgen über 1 oder 3 Jahre, um günstige Preise zu erhalten. EC2-Spot-Instanzen sind das billigste Modell , haben aber den Nachteil, dass ihre Verfügbarkeit nicht immer garantiert ist.

Jedes Zahlungsmodell hat seine Vor- und Nachteile und eignet sich für unterschiedliche Anwendungsfälle. Gute Ergebnisse können auch durch die Kombination mehrerer Zahlungsmodelle erzielt werden[SAG DER CLOUD-EXPERT/FIRMA]. Dies wird in Unterkapitel 3.4 behandelt.[WIRD ES?]

In dieser Arbeit wird nicht darauf eingegangen, wie die richtige Server-Instanz ausgewählt werden sollte, da die Auswahl von individuellen Anforderungen abhängt, die von Fall zu Fall unterschiedlich sind. Im Allgemeinen, um die beste Leistung zu erzielen, wird empfohlen, Instanzen so nahe wie möglich an den anderen Ressourcen, mit denen sie kommunizieren werden, und an der Mehrzahl der Endnutzer, die die Dienste in Anspruch nehmen werden, zu platzieren.

3.1 On-Demand / Nutzungsabhängige Zahlung

Bei diesem Zahlungsmodell besteht keine Notwendigkeit, ein festes Anfangsbudget festzulegen. Die Kosten richten sich nach dem Verbrauch auf der Grundlage der Nutzungsstunden.
Dieses Modell eignet sich für Projekte, bei denen nicht viel vorhersehbar ist und die Möglichkeit besteht, dass das Projekt in kurzer Zeit abgeschlossen sein wird, so dass es keinen Sinn macht, eine langfristige Verpflichtung einzugehen.
Hier einige Beispiele von EC2-Instanzen im On-Demand Zahlungsmodell.

WARUM IST DIESE ABB.?

Der Preis für dieses Modell variiert je nach Instanztyp, Region und der übertragenen Datenmenge. Die aktuellen Preise für die verschiedenen Regionen sind auf der Amazon-Website in der Sektion EC2 - On-Demand-Preise[6] zu finden. In der Abbildung 2 werden

[6]https://aws.amazon.com/de/ec2/pricing/on-demand/

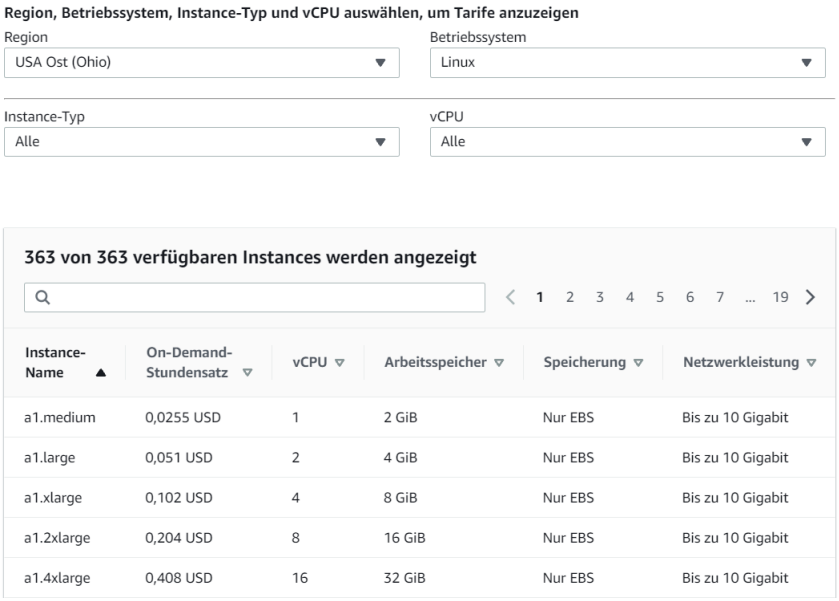

Abbildung 2
On-Demand Preise für Amazon EC2 [3]

die für die Region Ohio verfügbaren Linux-Instanzen gezeigt. Es ist zu beachten, dass Instanzen mit denselben Eigenschaften, aber in verschiedenen Regionen, unterschiedliche Preise haben können.

3.2 Reservierte Instanzen und Saving Plans

Die Zahlungsmodelle Reservierte Instanzen und Saving Plans sind sich sehr ähnlich. Beide kommen mit einer gleichbleibenden Nutzungsverpflichtung, die in €/Stunden gemessen wird.
Um die reduzierten Preise zu bekommen, müssen Verträge 1 oder 3 Jahre abgeschlossen werden.
Nachfolgend werden die prozentualen Einsparungen gemäß des jeweiligen Modells gezeigt.

Einsparungen nach Modell			
Compute Savings Plans	EC2-Instance Savings Plans	Convertible Reserved Instances	Standard Reserved Instances
bis zu 66%	bis zu 72%	bis zu 54%)	bis zu 72%

[8, 12]

Die ersten beiden Optionen in der obigen Tabelle, die Saving Plans, unterscheiden sich dadurch, dass die Compute Savings Plans die Flexibilität bieten, EC2-Instanzen nach Familie[7], Größe, Availability Zone (AZ), Betriebssystem oder Mandant zu wechseln.

[7] [20], Seite 95

> „Bei Compute Savings Plans können Sie beispielsweise jederzeit von C4- auf M5-Instances wechseln, eine Workload von EU (Irland) nach EU (London) verlagern oder eine Workload von EC2 auf Fargate oder Lambda verschieben. Dabei zahlen Sie automatisch weiterhin den Savings Plans-Preis." [12]

Bei den EC2-Instance Saving Plans hingegen muss eine Instance-Familie in einer bestimmten Region ausgewählt werden. Dies reduziert automatisch die Kosten für die ausgewählte Instanz-Familie in der jeweiligen Region, unabhängig von Availability Zone, Größe, Betriebssystem oder Mandant.
Die Festlegung eines festen Stundensatzes über einen langen Zeitraum bietet die Möglichkeit, künftige Kosten zu planen[ZITAT/BASIERT AUF...]
AUCH FÜR RIs?

3.2.1 Vorauszahlung

Zusätzlich gibt es bei Saving Plans und reservierten Instanzen die Option im Voraus zu zahlen. Im Gegenzug wird ein niedrigerer Preis angeboten.

Amazon bietet drei verschiedene Optionen an. Diese sind teilweise, keine oder vollständige Vorauszahlung.

Bei teilweiser Vorauszahlung ist eine Anzahlung von etwa 50% zu leisten.

(To-Do: wie viel kann in den verschiedenen Szenarien eingespart werden).

3.3 Spot Instanzen

EC2 Spot-Instances bieten die Möglichkeit aus von anderen Nutzern ungenutzter EC2-Instances zu profitieren. Mit einem Preisvorteil von bis zu 90 % gegenüber normalen On-Demand-Instanzen sind Spot-Instanzen ideal für fehlertolerante Anwendungen wie auf Containern ausgeführte Workloads, CI/CD, Bigdata-Anwendungen und ähnliches.

3.3.1 Unterbrechbarkeit

Es ist zu beachten, dass Spot-Instanzen jederzeit unterbrochen werden können.

Einer der Gründe ist die Preisüberschreitung[RICHTIG?] der Instanz. Wenn Spot-Instanzen angefordert werden, wird einen Maximalpreis festgelegt. Sollte der Preis der Instanz höher als das eingegebener Maximalpreis, wird die Instanz für die aktuelle Einstellung nicht mehr verfügbar sein.

Ein anderes Szenario ist, wenn der Instanzanbieter die Instanz erneut anfordert. Falls eine Spot-Instanz unterbrochen wird, benachrichtigt Amazon EC2 2 Minuten im Voraus. Dieses Ereignis ist verfügbar auf CloudWatch und es kann einen Alarm dafür eingestellt

werden.
Daher sollten kritische Anwendungen nicht nur mit EC2 Spot-Instanzen laufen.

3.4 Wann welches Zahlungsmodell?

Die folgende Tabelle fasst die Eigenschaften der Zahlungsmodelle für Instanzen zusammen und listet typische Applikationen je nach Zahlungsmodell auf. Abb. AKTUELL?

Vergleich der Zahlungsmodelle		
Eigenschaften		
Nutzungsabhängige Zahlung: On-Demand	**Optionen mit Verpflichtung: Reserved Instances and Saving Plans**	**Überschüssige Kapazität: Spot-Instances**
Erster Test oder erste Entwicklung	Verträge über 1 bis 3 Jahre	Unterbrechbare Instanzen
Keine langfristigen Verpflichtungen	Preisverpflichtung	Die billigste und riskanteste Option
Keine Vorabzahlungen		
Geeignete und übliche Anwendungen		
Allgemeine Anwendungen	Applikationen mit stabiler Arbeitbelastung	Bigdata-Applikationen
Experimente und Tests		Containern ausgeführte Workloads
Nicht unterbrechbare Applikationen		Fehlertolerante Applikationen
Applikationen mit unvorhersehbaren Arbeitsbelastungen		Batch-Workloads

Abbildung 3
Vergleich der Zahlungsmodelle
Quelle: Eigene Darstellung.

Summary/Fazit

In diesem Kapitel wurden die verschiedenen Zahlungsmodelle für EC2-Instanzen untersucht. Es wurden Hinweise für die Auswahl des richtigen Zahlungsmodells in verschiedenen Szenarien gegeben. Dies wurde erklärt, um die Preisvorteile[RICHTIG?] von den Zahlungsmodellen zu nutzen. Im Kapitel Kostenüberwachung wird X(CloudWatch?) vorgestellt, mit dem überprüft werden kann, ob das ausgewählte Zahlungsmodell tatsächlich das Richtige für den betreffenden Anwendungsfall ist. Für das On-Demand-Zahlungsmodell gibt es keine Kostenreduzierung, aber es gibt Maßnahmen, um die Nutzung von Instanzen zu reduzieren. Auf diese Maßnahmen wird im Kapitel über Optimierungsmaßnahmen näher eingegangen.

4 Kostenüberwachung

(CHECK Version/Carlo 21.11)
In diesem Kapitel werden Werkzeuge vorgestellt, mit denen Budgets mit Alarmen erstellt werden, diese informieren, wenn ein bestimmter Prozentsatz des festgelegten Budgets überschritten wurde. Die Erstellung von Budgets trägt zu einer besseren Planung-/Prognose- und Kostenkontrolle. ZITAT? Die Einstellung von Alarmen für relevante Ereignisse wie im Fall einer Budgetüberschreitung oder dem Start einer Instanz[IST GUT WEIL/TRÄG ZU...BEI].

Darüber hinaus ist es mit Werkzeugen wie CloudWatch möglich, die Abschaltung bestimmter Dienste zu automatisieren, wenn eine Budgetschwelle überschritten wurde. Diese Maßnahmen werden in dem Kapitel über die Optimierung behandelt.

Durch die Verwendung von Tags ist es möglich, die Ressourcen nach Kriterien wie Region, Umgebung, Projekt, Art der Ressource usw. zu visualisieren, dies ermöglicht, Kosten auf den von der Organisation festgelegten Ebenen zu verfolgen. [DIAGRAMM: BUDGET PRO ABTEILUNG]
Es könnte zum Beispiel ein Szenario entstehen, in dem eine Abteilung innerhalb der Organisation mehr Kosten verursacht als andere. In erster Linie ist dies nur bemerkenswert anhand eines Anstiegs der von Amazon generierten Rechnung, aber um den Grund für diesen Anstieg genauer zu verstehen, muss ihre Ursache untersucht werden. Werkzeuge wie Cost-Explorer machen diese Art von Analyse möglich.

TRUSTED ADVISOR(Einleitung)??

4.1 AWS CloudWatch

Amazon CloudWatch ermöglicht die Überwachung der Leistung von Resources, auch bei Ressourcen, die über verschiedene Regionen verteilt sind. CloudWatch sammelt operative Daten für die Verlaufsanalysen und die Entscheidungsfindung in Bezug auf Optimierung und Fehlerbehebung.

CloudWatch beschränkt sich nicht nur darauf, Daten aus der AWS-Umgebung zu empfangen. Externe CloudWatch fähige? Metriken sind ebenfalls zugelassen und können für eine einheitliche Analyse aggregiert werden.

Eine der Metriken, die mit Amazon CloudWatch überwacht werden kann, ist die CPU-Auslastung von EC2-Instanzen. Basierend auf einem Prozentsatz der CPU-Auslastung können Alarme?Benachrichtigungen?[WELCHES WORT?] und Aktionen konfiguriert werden
Eine dieser Aktionen ist die automatische Einrichtung neuer Instanzen zur Deckung des Kapazitätsbedarfs [8]. Diese Art von Aktionen werden im Kapitel 5 Optimierungsmaßnah-

[8] [20], Seite 185

men tiefer behandelt.

Im Folgenden werden die grundlegenden Bereiche und Begriffe von CloudWatch erläutert und wie sie zur Überwachung von Informationen über AWS-Ressourcen verwendet werden.

Metriken
Eine Metrik stellt eine Reihe von Daten über die Leistung einer Ressource in zeitlicher Reihenfolge dar. Standardmäßig werden viele kostenlose Metriken an CloudWatch übermittelt. Zum Beispiel kann der Durchschnitt von einer bestimmten API pro Stunde untersucht werden. Für eine detailliertere Überwachung ist es möglich, benutzerdefinierte Metriken zu konfigurieren, die eine Auflösung von bis zu 1 Sekunde zulassen. Ein praktisches Beispiel für benutzerdefinierte Metriken ist die Messung der Ladezeit einer Website. [EIN BEISPIEL MIT BEZUG AUF K:OPTIMIERUNG?oder WIESO IST DIE AUFLÖSUNG RELEVANT?]

Ereignisse / Events
Bei CloudWatch ist ein Ereignis eine Änderung bei einer Ressource in der AWS-Umgebung. AWS-Ressourcen können Ereignisse erzeugen, wenn sich ihr Status ändert. Beispielsweise, ein Ereignis wird erzeugt, wenn Amazon EC2 Auto Scaling, Instanzen gestartet oder beendet wird[9] oder wenn eine bestimmte Menge an Speicherplatz in einem Bucket erreicht wurde.

Regel
Eine Regel ordnet eintreffende Ereignisse zu und leitet diese zur Verarbeitung an Ziele weiter. Eine einzelne Regel kann an mehrere Ziele weiterleiten, die alle parallel verarbeitet werden[10].

Target / Ziele
Ziele sind Ressourcen, die aufgerufen werden, wenn eine Regel ausgelöst wird. EC2 instances, AWS Lambda functions und Amazon SNS topics sind unter anderem mögliche Ziele. Die Ziele einer Regel müssen sich in derselben Region wie die Regel befinden [11].

Alarme / Benachrichtigungen?
Benachrichtigt zu werden ist es wichtig, um relevante Ereignisse nicht zu verpassen und rechtzeitig Maßnahmen zu ergreifen. Mit CloudWatch können Alarme eingerichtet wer-

[9] [14], Seite 1
[10] [14], Seite 2
[11] [14], Seite 2

den, die durch Metriken wie die CPU-Auslastung und auch Gebühren[ANDERES WORT] auf AWS-Rechnungen ausgelöst werden. Benachrichtigungen können durch Amazon SNS oder zu einer E-Mail-Adresse geschickt werden.[NUR?]

Visualisierung von Metriken/ Dashboards
Mit Cloud-Watch Dashboards können [BESSERE FORMULIERUNG] relevante Metriken grafisch dargestellt werden. Durch die Dashboards[PASSENDES WORT?] können auch Alarmen erstellt werden. Für die Einrichtung der Alarme ist kein technisches Wissen benötigt[12]. Die in den Dashboards enthaltenen Informationen sind nicht nur für ihre Autoren von Relevanz. Weitere Personen innerhalb oder außerhalb der Organisation können Zugriff auf Dashboards mit nutzliche Informationen bekommen, um Prozesse zu beschleunigen[VORTEILE DES OPTIMIERUNGSPROZESS VERWEISEN] oder Probleme schneller zu beheben[USE CASE].

Dies ermöglicht einen schnelleren Kommunikationsfluss in Echtzeit[IM(Schwachstellenanalyse) IuM Technik/Check Book KCrmar]. Die Zugriffsverwaltung für geteilte Dashboards wird über AWS Identity and Access Management abgewickelt[13].

4.1.1 Fakturierungsalarme mit CloudWatch

AWS CloudWatch empfängt Abrechnungsmetriken von alle Ressourcen. Auf der Grundlage dieser Metriken ist es daher möglich, Regeln zu erstellen, die bei Überschreitung des geplanten Budgets einen Alarm auslösen. [KANN MAN NUR Total Estimated Charge verwenden ODER KÖNNTE MAN NACH PROJEKTE TRENNEN?] [WIE KANN MAN EIN BEZUG AUF PROJEKTMANAGEMENT/Kostenkontrolle MACHEN?]

4.1.2 Alarm bei Hoch- und Runterfahren von EC2-Instanzen

[AutoScaling WURDE NOCH NICHT THEMATISIERT]Obwohl Auto-Scaling dafür sorgt, die Rechenkapazität dynamisch anzupassen, ist es von größter Wichtigkeit, über Änderungen in der Infrastruktur informiert zu sein, ohne die Dashboards manuell überprüfen zu müssen. WEIL ....

4.2 AWS Cost-Explorer

Mit Cost-Explorer können Kosten der letzten 12 Monate und eine Schätzung der Kosten des laufenden Monats visualisiert werden. Darüber hinaus wird eine Kostenprognose für

[12] [15], Seite 28
[13] [15], Seite 18 und 39

die nächsten Monate erstellt. Die Prognose basieren auf die Kosten der vergangenen Monaten. Die Nutzung des Cost-Explorers ist kostenlos, nur API-Aufrufe sind kostenpflichtig [14].

Es gibt drei Arten von Berichten, die der Cost-Explorer bereitstellt:

- Bericht über die Nutzung und die in den letzten 12 Monaten entstandenen Kosten.
- Kostenprognose der kommenden drei Monaten.
- Empfehlungen zu Reserved Instances.

Amazon analysiert die bisherige Nutzung der Instanzen und gibt Empfehlungen zur Kostensenkung durch den Wechsel von EC2-Instanzen zu reservierten Instanzen. Diese ignorieren Kapazität, die bereits von anderen reservierten Instanzen abgedeckt wurden.

WAS KANN MAN MIT JEDER EINZELNE VON DIESEN INFOS? PM/Controlling etc. **Bugdetplanung**

Die Prognose der kommenden Kosten, dienen zu einer guter operativen Bugdetplanung...

4.3 AWS Trusted Advisor

AWS Trusted Advisor ist ein Werkzeug, das entwickelt wurde, um Kosten zu senken, um Systemverfügbarkeit und -leistung zu verbessern und um Sicherheit zu erhöhen. Es analysiert die Nutzung des AWS-Kontos und gibt Best-Practice-Empfehlungen. Es werden die Kategorien Leistungsgrenzen und Kostenoptimierung insbesondere betrachtet, da diese am relevanten für die vorliegende Arbeit sind.

Es ist zu berücksichtigen, dass nur limitierte Sicherheitsprüfungen (6 Prüfungen November 2021) für Konten in den Plänen Developer und Basic Support kostenlos sind. Prüfungen für die Kategorie Leistungsgrenzen sind kostenlos.

Detaillierte Informationen und Empfehlungen von der Kategorien Kostenoptimierung, Performance und Fehlertoleranz sind nur zugänglich, wenn ein Business- oder Enterprise-Konto vorliegt[15].

ES MUSS GEPRÜFT WERDEN, OB ES SINNVOLL IST, FÜR DIE OBEN GENANNTEN SUPPORT-PLÄNEN ZU ZAHLEN.

Die Abbildung 4 zeigt die 5 Kategorien von Trusted Advisor mit jeweils 3 Arten von Indikatoren. Die Indikatoren zeigen an, welche Prüfungen durchgeführt wurden.

Grün bedeutet, dass keine Fehler oder zu prüfenden Empfehlungen vorhanden sind. Warnungen werden durch orangefarbene Indikatoren und Fehler durch rote Indikatoren angezeigt.

Diese Empfehlungen sind eine Zusammenfassung auf hohem Niveau. Sie sind ein Startpunkt für die Untersuchung von Ressourcen mit Hilfe anderer Werkzeuge wie CloudWatch oder Cost-Explorer[ODER?].

[14] https://aws.amazon.com/de/aws-cost-management/pricing/
[15] https://aws.amazon.com/de/premiumsupport/technology/trusted-advisor/

Abbildung 4
AWS Trusted Advisor Kategorien

Kostenoptimierung

Die Empfehlungen zur Kostenoptimierung konzentrieren sich auf Möglichkeiten zur Kostensenkung, indem ungenutzte Ressourcen hervorgehoben werden. Sollten EC2-Instanzen mit geringer Auslastung gefunden werden, wird es diese bei Trusted Advisor signalisiert. Denn diese Instanzen verbrauchen Ressourcen und können terminiert oder pausiert werden. [HIER FEHLT...] Auch nicht zugewiesene Elastic IP-Adressen erzeugen Kosten. Diese können gegebenenfalls von Trusted Advisor gefunden werden.

Leistungsgrenzen

In dieser Kategorie werden Empfehlungen zur Vermeidung von Grenzwertüberschreitungen hervorgehoben. Es wird zum Beispiel nach einer Nutzung gesucht, die mehr als 80 % des Leistungsgrenzwerten für wichtige Dienste beträgt. Einige Beispiele sind Amazon EC2, Auto Scaling, Elastic Block Store, Simple Email Service und AWS CloudFormation.

Sich dieser Grenzen bewusst zu sein, gibt die Möglichkeit, rechtzeitig zu handeln und es trägt zu Kostenüberwachung bei. [SAGEN OB DIESE GRENZE DIE GLEICHE WIE BEI CloudWatch SIND] [ZEIGE BEISPIELE FueR Empfehlungen]
Bei der Erwägung von Trusted-Advisor ist zu überlegen, ob es kosteneffizient ist, für Pläne zu zahlen, die den Zugang zu allen Empfehlungen des Trusted Advisors ermöglichen. Das übergeordnete Ziel dieser Arbeit ist es, die Entstehung der Kosten auf eine praktikable Weise zu verstehen (Kostenüberwachung). Dies, um Optimierungsmaßnahmen zu ermöglichen.
Es wäre nicht sinnvoll, Kosten für Plänen wie Geschäfts- oder Enterprise Support zu übernehmen, wenn diese die möglichen Einsparungen übersteigen. Die Vorteile von Geschäfts- oder Enterprise Support-Plänen beschränken sich nicht auf Kosteneinsparungen und Kostenbegrenzung, sondern tragen auch zur Sicherheit und Leistung bei. Jedes Unternehmen muss selbst entscheiden, ob es diese Informationen benötigt. WIE TRIFFT MAN EINE

Überwachungswerkzeuge gemäß ihrer Verwendung			
	Cloud-Watch	**Cost-Explorer**	**Trusted-Advisor**
Visualisierung der CPU utilization	x		
Analyse von Kosten nach Tags, Monat...		x	
Benachrichtigung/Alarmen von Events	x		
Empfehlungen bezüglich RIs		x	x?
Um Ressourcen nach Tag zu		x	
Prognose für kommende Kosten			

Abbildung 5
Überwachungswerkzeuge gemäß ihrer Verwendung

ENTSCHEIDUNG IM UNTERNEHMEN? MODELLE?

4.4 Überwachungswerkzeuge gemäß ihrer Verwendung?

Handlungsempfehlungen

Überlegung 1:

Es kann in Erwägung gezogen werden, für einen begrenzten Zeitraum von 3 Monaten einen Support-Plan zu bezahlen, um aus den gegebenen Empfehlungen zu lernen. Oder Business-Plan alle 6 Monate für 1 Monat zu aktivieren.

Überlegung 2:

Ein Berater für eine Prüfung und Optimierung der Ressourcen kann in Deutschland zwischen x und b Dollar kosten. Dies ist nur eine Alternative zu den Plänen des Trusted-Advisor. Ein Berater, der alle 5 Kategorien abdeckt, könnte 555-999 kosten. Das sagt der Berater Juanito von der Firma XXX. [WIE KANN ICH DIESE ÜBERLEGUNGEN RICHTIG DARSTELLEN?]

Summary/Fazit

In diesem Kapitel wurde gezeigt, wie es mit CloudWatch möglich ist, Alarme auf Basis von Ereignissen einzurichten, die mit Amazon SNS [IWO ERKLÄREN WAS DAS IST;Für Kommunikation innerhalb von AWS] oder externen E-Mail-Adressen kommunizieren. Im nächsten Kapitel wird CloudWatch erneut behandelt. Diesmal nicht als Überwachungswerkzeug, sondern als Optimierungswerkzeug zur Erstellung von automatisierte Aktionen/Reaktionen. Dazu war es notwendig, die Rolle der von CloudWatch gesammelten Metriken zu verstehen, die die Grundlage für die Verwaltung von Aktionen wie Auto-Scaling-Gruppen bilden. Aus dem Blickwinkel des Kostenmanagements wurde gezeigt, wie man mit dem Cost-Explorer die Kosten der letzten 12 Monate analysieren, eine Einschätzung der Kosten im aktuellen Monat und eine Prognose für die nächsten Monate erhalten kann. Es ist möglich, die Kosten nach Tags und anderen Filtern zu trennen. Diese Informationen dient unter anderem zur Erstellung einer operativen Budgetplanung mit

genaueren Daten. Darüber hinaus wurde Trusted Advisor vorgestellt, die konkrete Optimierungsempfehlungen gibt und warnet über Leistungsgrenzen. Dies kann mit erheblichen Kosten verbunden sein und ist daher nicht für alle Arten von Unternehmen unmittelbar attraktiv. [WAS KOMMT IN NÄCHTEN KAP.?]

5 Optimierungsmaßnahmen

Die mit den Überwachungswerkzeuge gesammelte Informationen, bilden die Grundlage für die Optimierungsmaßnahmen.[HIER KONKRETER WERDEN] In diesem Kapitel werden die mit Hilfe der Werkzeuge gewonnenen Informationen genutzt[INOF X FUER WERKZEUG 1...], um über die am besten geeigneten Optimierungsmaßnahmen zu entscheiden.[VERSTÄNDLICH?]

5.1 EC2 Automatische Skalierung

Auto Scaling ist es hilfreich, um die richtige Anzahl von EC2 Instanzen zur Verfügung zu haben, um die Anwendungslast dynamisch abzudecken.

Die Abbildung 6 zeigt das wechselnde Verhalten einer Beispielanwendung, die vor allem unter der Woche Ressourcen verbraucht. Am Wochenende sinkt die Nachfrage nach Rechnerkapazität auf weniger als 25 % und lässt den Rest der Kapazität ungenutzt.

Die gelben Säulen stellen die tägliche genutzte Rechenkapazität dar. Die graue Zone entspricht ungenutzte Rechenkapazität und beträgt etwa ein Drittel der wöchentlichen Rechnerkapazität.

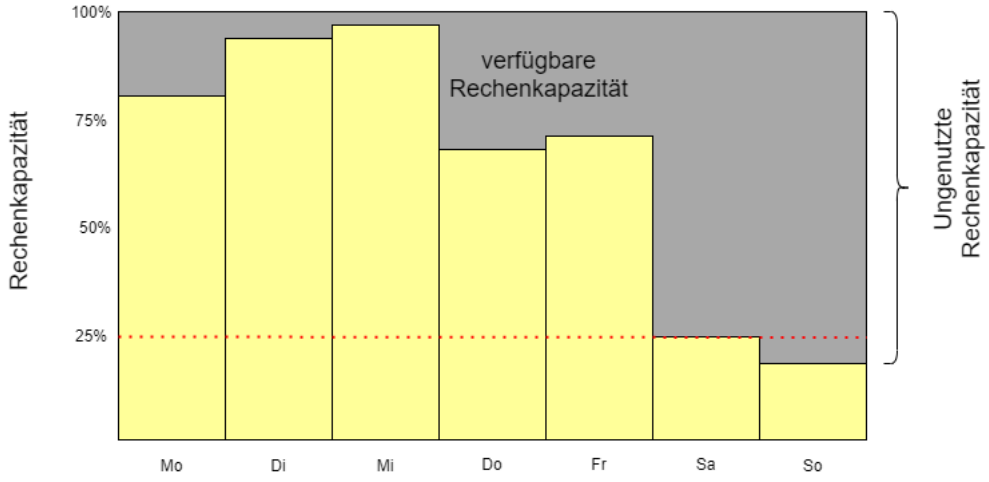

Abbildung 6
Ungenutzte Rechenkapazität ohne automatische Skalierung.
Quelle: Eigene Darstellung.

5.1.1 Zeitgesteuerte Skalierung

Nicht produktive Umgebungen??

In einem On-Premise-System macht es möglicherweise keinen Unterschied bei den Kosten, wenn die Instanzen aktiv bleiben. Im Gegensatz dazu ist es bei On-Demand-Zahlungsmodelle sinnvoll Zeiträume zu definieren, in denen Instanzen abgeschaltet werden können, um den Verbrauch an Ressourcen zu reduzieren.

Bei Systemen, die nur tagsüber und unter der Woche in Betrieb sein müssen, kann dies eine Einsparung von zu 67% bedeuten. Wenn zum Beispiel Test- und Beta-Umgebungen von Montag bis Freitag von 7 bis 20 Uhr laufen würden.

Zeitgesteuerte Skalierung von EC2-Instanzen		
	7:00-20:00 Uhr Montag-Freitag	24/7
Stunden inaktiv täglich	11	0
Stunden aktiv täglich	13	24
Tagen in der Woche	5	7
Stunden in der Woche	55	168
Stunden monatlich	239	730
Einsparung/Differenz %	**67.26%**	

Stundensatz	€0.1536	
Anzahl Instanzen	2	
On-Demand Kosten pro Monat*	€73.42	€224.26

Abbildung 7
Berechnung für ein nicht-produktive Umgebung mit Zeitgesteuerte Skalierung.
Quelle: Eigene Darstellung.

Der Stundensatz wurde am 23.11.2021 mit dem AWS Pricing Calculator[18] ermitellt für Linux Instanzen in Frankfurt mit 4vCPUs, 16 GB Arbeitsspeicher und Instanz-Familie t4g.xlarge in On-Demand-Zahlungsmodell.

Produktive Umgebungen??
Wenn der Zeitpunkt einer hohen Nachfrage bekannt ist, kann eine Erhöhung der Rechnerkapazität geplant werden, um Überlastungen zu vermeiden.

Beispiele für solche Zeiträume sind Weihnachten, Cyber-Monday und Black Friday. [STATISTIK?https://de.statista.com/statistik/daten/studie/1076963/umfrage/ausgaben-an-black-friday-und-cyber-monday-in-deutschland/]

5.1.2 Dynamische automatische Skalierung / Dynamisches Auto Scaling

Es kann jedoch zu schnelle und kontinuierliche Änderungen im Verhalten von Applikationen geben, häufig innerhalb von wenige Minuten. Bei solche Szenarien ist sinnvoller, Metriken zur automatischen Anpassung der Skalierung der Rechenkapazität festzulegen.

Beispiele für eine veränderte Nutzung von Applikationen finden sich bei Tinder und OkCupid, zwei der größten Dating-Applikationen in den vereinigten staaten. Die Abbildung 8 zeigt die Nutzungsspitzen bei den genannten Applikationen. Dieses wechselnde Verhalten wirkt sich unmittelbar auf die zu verschiedenen Tageszeiten benötigte Rechenkapazität aus und macht eine dynamische Skalierung der Rechenkapazität erforderlich, wenn das Ziel darin besteht, die Verschwendung von Ressourcen zu vermeiden oder zu verringern.

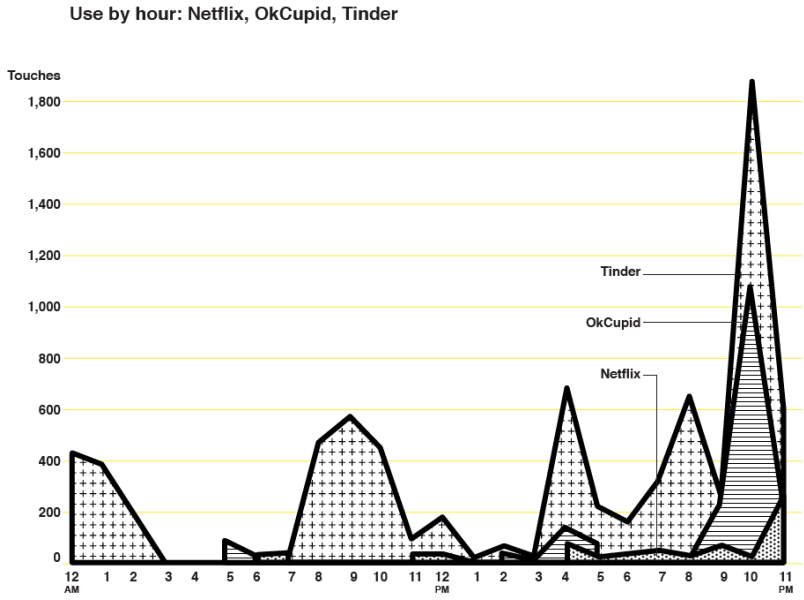

Abbildung 8
Nutzung von Tinder, OkCupid und Netflix pro Stunde. [28]
Mit Touches sind die Anzahl der Klicks, Swipes oder einfachen Interaktionen mit der Applikation gemeint.

Weitere Beispiele für solche Zeitpunkte sind das Feierabend, Mittagspause und beim Abendessen, wenn es um Applikationen für Unterhaltung und Online-Shopping geht[ZITAT/BELEG? Die für die automatische Skalierung erforderlichen Metriken wurden bereits im Kapitel Überwachungswerkzeuge erwähnt. Eine der Metriken, die von Optimierungsexperten/Cloudexperten [WELCHES?]benutzt wird, ist die gesamte CPU-Auslastung[BELEG]. Um die CPU-Auslastung als Metrik zu verwenden, werden mindestens zwei Schwellenwerte definiert. Eine für die Erhöhung von Rechenkapazität, Scale-Out genannt und eine für das Verringern von Rechenkapazität bezeichnet als Scale-In.

5.1.3 Manual Scaling CHECK/23.11

Für die Konfiguration einer Auto-Scaling-Gruppe werden die minimale, maximale und gewünschte Anzahl von Instanzen definiert. Wenn aufgrund von Bedingungen, die in der Konfiguration einer Auto-Scaling-Gruppe nicht berücksichtigt wurden mehr Rechenkapazität benötigt wird, ist es möglich, die Rechenkapazität manuell zu steuern. Dies geschieht, ohne dass die aktiven Instanzen unterbrochen werden.

5.1.4 Voraussagende Skalierung / Predective Scaling ??

Voraussagende Skalierung oder Predictive Scaling auf Englisch, nutzt maschinelles Lernen, um den Kapazitätsbedarf auf der Grundlage historischer Daten von CloudWatch vorherzusagen. Mit Hilfe der Predictive Scaling kann es die Kapazität vor der erwarteten Auslastung bereitstellen, im Gegensatz zur dynamischen Skalierung, die reaktiv ist. Für Instanzen, die viel Zeit für die Initialisierung benötigen kann die Zeit zwischen dem Beginn des Nachfrageanstiegs und der Initialisierung der Instanz vermieden oder verkürzt werden. [DIAGRAMM] Anders als Zeitgesteuerte Skalierung ist es nicht notwendig, die Verhaltensmuster der Anwendungen zu analysieren. [SOLLTE DAS ZITIERT WERDEN?]

5.2 S3 Optimierung

5.2.1 Richtige Speicherklassen wählen

Um die Speicherkosten zu optimieren, ist es daher notwendig, die richtige Speicherklassen für die jeweilige Applikation wählen. Um die richtige Wahl zu treffen, müssen die Anforderungen der Applikation verstanden werden. Klinische Patientendaten und eine Instagram-Story unterscheiden sich in der Zugriffshäufigkeit auf diese Daten und in der Länge der Aufbewahrungszeit[PASSENDES WORT?].

Amazon bietet verschiedene Speicherklassen an, die sich im Preis und in der Häufigkeit des Zugriffs auf die Objekte unterscheiden. Objekte sind in Behältern enthalten, die Buckets genannt werden. [ABB Speicherklasse ->Eigenschaften der Anwendung?] Wenn bekannt ist, dass die Daten über einen längeren Zeitraum gespeichert werden müssen. Falls die Anforderungen der Applikation dies vorschreiben oder für den Fall dass, per Gesetz auf die Informationen in der Zukunft zugegriffen werden muss.
Zusätzlich, wenn auf die Daten nicht häufig zugegriffen wird, sind Glacier und Glacier Deep Archive passende Speicherklassen. Die Entscheidung ist jedoch nicht immer so einfach und die Umstände können sich schnell ändern. Hinzu kommt, dass nicht alle Daten in einer Applikation immer die gleichen Zugriffsmuster haben. Für solche Fälle ist es möglich, Regeln zu definieren, die Dateien zwischen verschiedenen Speicherklassen abhängig von ihrem Alter übertragen.

5.2.2 Lebenszyklus-Konfiguration/Lifecycle Policies

Eine S3-Lebenszykluskonfiguration beschreibt in einer XML-Datei Regeln und Aktionen für die Manipulation von Objekten.

Aktionen wie das Verschieben von Objekten verursachen Kosten. Einige von ihnen werden in Abbildung 9 für die Berechnung der Speicherkosten verwendet.
Um konkretere Regeln zu zu definieren, ist es möglich Tags zu verwenden und somit eine Unterscheidung zwischen Objekten mit verschiedenen Tags zu treffen. Es ist zum Beispiel möglich, alle Objekte mit dem Tag: Dev nach 45 Tagen nach Standard Infrequent Access und nach 120 Tagen nach S3 Glacier zu verschieben.

```
  <LifecycleConfiguration>
  <Rule>
    <ID>example-id</ID>

 <Filter>
      <Tag>
         <Key>key</Key>
         <Value>Dev</Value>
      </Tag>
</Filter>

    <Status>Enabled</Status>
    <Transition>
      <Days>45</Days>
      <StorageClass>STANDARD_IA</StorageClass>
    </Transition>
    <Transition>
      <Days>120</Days>
      <StorageClass>GLACIER</StorageClass>
    </Transition>
    <Expiration>
      <Days>365</Days>
    </Expiration>
  </Rule>
</LifecycleConfiguration>
```

Angepasster Code auf Basis der Beispiele auf Seite 701 in Amazon Simple Storage Service - User Guide, [19]

Zur Veranschaulichung (der gezeigten Informationen[OHNE ODER DAMIT?]) wird

davon ausgegangen, dass ein Sicherheitsunternehmen, das Sicherheitsvideos speichern muss, im Durchschnitt 120 TB an Videos speichern muss. Viele von ihnen werden mindestens 5 Jahre lang aufbewahrt, falls sie vor Gericht als Beweismittel dienen. Ungefähr 50% der Videos werden mindestens einmal im Monat überprüft und müssen laut Gesetz sofort zugänglich sein. Die Software des Unternehmens speichert die Videos in S3-Buckets und hat eine durchschnittliche Größe von 3,4 GB.
Im Folgenden werden die Speicherkosten für ein Szenario berechnet, bei dem nur S3 Standard verwendet wird. Als nächstes wird die Kombination von S3 Standard Infrequent Access, S3 Glacier und S3 Standard für ein Szenario betrachtet, in dem die Dateien je nach Alter verschoben werden. Im letzten Szenario müssen die Kosten für den Übergang[RICHTIGES WORT?] zwischen Speicherklassen berücksichtigt werden.
Zur Vereinfachung der Berechnung wird angenommen, dass 20% der Dateien in S3 Standard Infrequent Access und 30% in S3 Glacier gespeichert werden.

Bei der Berechnung wurden die Kosten für das Verschieben von Dateien zwischen Speicherklassen berücksichtigt. Anhand der Berechnungen lässt sich erkennen,

Anhand der Berechnungen Sie sehen, dass ein Einsparungspotenzial von rund 1.000 Dollar pro Monat besteht, indem die notwendigen Regeln aufgestellt werden, um einen Teil der Dateien in anderen Speicherklassen zu verschieben, welche niedrigere Preise bieten.

5.2.3 Intelligent-Tiering

Intelligent-Tiering verschiebt Dateien auf der Grundlage von Zugriffsmustern. Diese Speicherklasse ist ideal für Daten mit wechselnden oder unbekannten Zugriffsmustern. Wie die Senior Product Manager für S3 Ruhi Dang erklärt, einige Unternehmen haben weder die Zeit noch die finanziellen Möglichkeiten, eine Person einzustellen, die ihre Daten sortiert und in die richtige Speicherklasse einordnet. Intelligent Auto Tiering ist eine attraktive Lösung für Unternehmen, die jährlich weniger als $100,000 für Speicher ausgeben [16].

Abbildung 10 zeigt, wie die Dateien in Abhängigkeit davon, ob auf sie zugegriffen wurde oder nicht, verschieben werden. Wird ein Datei zu einem späteren Zeitpunkt aus der Ebene der seltenen Zugriffe aufgerufen, wird es automatisch in eine Speicherklasse der häufigen Zugriffe zurückversetzt.

[16] [17], Minute: 21:12

Durchschnittliche Dateigröße	3.4	GB
Anzahl der Dateien	36,141	Überwachungsvideos
Gesamtspeicher	122880	GB
	120	TB

Ausschließlich S3-Standard verwenden		
	S3 Standard (erste 51200GB)	**S3 Standard (Nächste 450 TB /Monat)**
Speicherplatz in GB	51200	71680
Preis pro GB	$0.0245	$0.0235
Speicherverteilung	42%	58%
Anzahl der Dateien	15059	21082
Übertragungsgebühr (pro 1.000 Aufrufe)	-	-
Kosten für Verschiebung	0	0
Speicherkosten	$1,254.40	$1,684.48
Gesamtkosten		$2,938.88

Lebenszyklus-Konfiguration für die Verwendung von verschiedenen Arten von Speichern			
S3 Standard (erste 51200GB)	**S3 Standard (erste 51200GB)**	**S3 Standard Infrequent Access**	**S3 Glacier**
51200	10240	24576	36864
$0.0245	$0.0235	$0.0136	$0.0045
42%	8%	20%	30%
15059	3012	7228	10842
-	-	$0.0100	$0.0360
0	0	$0.72	$3.90
$1,254.40	$240.64	$334.23	$165.89
			$1,999.79

Abbildung 9
Kostenvergleich durch Nutzung von unterschiedlichen Speicherklassen.
Quelle: Eigene Darstellung. S3 Stundensätze: [10]
Der Punkt wurde als Dezimaltrennzeichen und das Komma als Tausendertrennzeichen verwendet.

6 Zusammenfassung und Ausblick

(To-Do:)
Kapitelweise Kurzdarstellung der Inhalte (inklusive Referenzierung auf die Kapitelnummerierung) => Nach dem Motto: *Was wurde wo beschrieben?*
Kurzdarstellung *Problem – Lösungsweg – Ergebnisse*
Rückkopplung auf die Einleitung: Wurde die Zielstellung der Arbeit und die Fragestellung zufriedenstellend beantwortet?
Kritische Bewertung (sofern nicht bereits im Hauptteil geschehen)

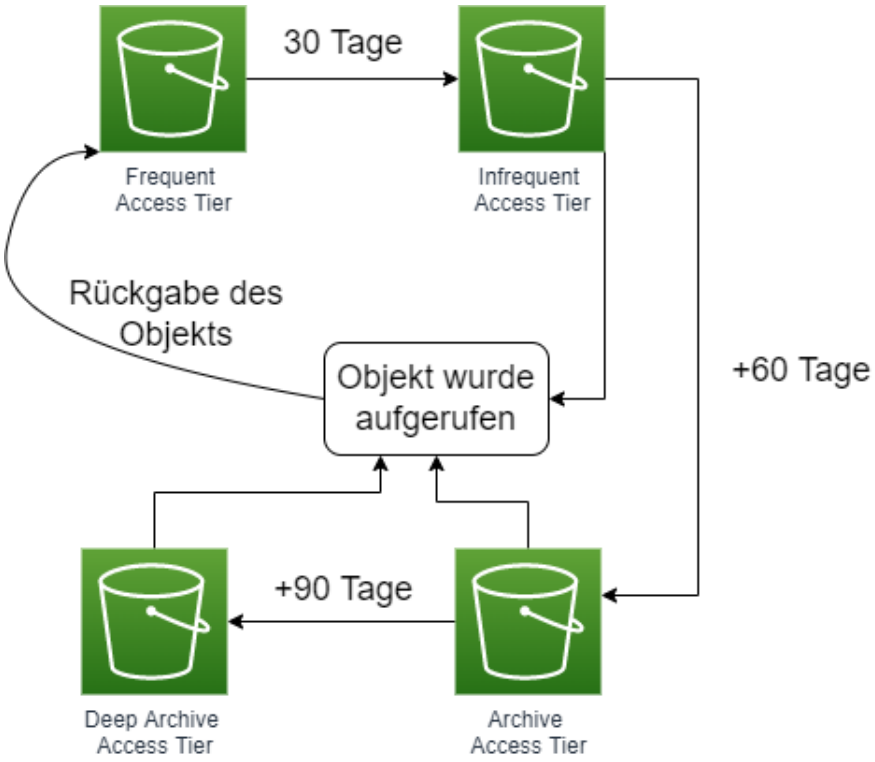

Abbildung 10
Funktionsweise von Intelligent-Tiering
Quelle: Eigene Darstellung.

Offene Probleme
Richtung der zukünftigen/möglichen Arbeiten
Erläuterung, warum welche Aspekte in der Arbeit nicht erläutert

6.1 Umweltbezogene Aspekte

Esta tesis habla sobre monitoreo y optimizacion de recursos de manera financiera. Pero esas dos areas enfocadas a la economia, tiene impacto en el medio ambiente por las emisiones generadas por las granjas de servidores. Estadisticas dicen que en Europa/Alemania se generan x toneladas de CO2 provenientes de centros de computo. Por tanto al monitoreat y reducir cosotos, se estan evitando despilfarros y al final tambien emisiones de CO2.

6.2 Test von den Werkzeugen und Maßnahmen

Da es in dieser Arbeit zeitlich nicht gelungen ist, die Überwachungswerkzeuge und Optimierungsmaßnahmen umzusetzen, bleibt es noch sie in einer echten Umgebung zu testen. Es wäre möglich zu verifizieren, ob die hier genannten Maßnahmen zur vergleichbaren

Einsparungen führen, wie die vom Cloud-Anbieter Amazon genannten.

Amazon bietet ein kostenloses Kontingent an, die jedoch für diese Tests nicht genug war.

6.3 Bewusstsein in der gesamten Organisation

Zusätzlich zu den bisher genannten Maßnahmen ist es wichtig, dass Verbraucher von Cloud-Diensten Bewusstsein für die Entstehung von Kosten entwikclen.[ODER sensibilisier werden?] Von dem Entwickler bis zum der IT-Manager, jeder sollte wissen, dass es so einfach ist, Cloud-Dienste mit ein paar Klicks zu beauftragen. Diese können in kurzer Zeit ungewünschte Kosten verursachen oder sogar über Jahre hinweg wirtschaftliche Schäden verursachen.

6.4 Die richtige Personen(Owneship verbreiten)

Die technischen Maßnahmen zur Überwachung und Kostenreduzierung wurden dargelegt, aber jemand muss diese Analysen, Anpassungen und Entscheidungen durchführen. Deshalb ist es wichtig, bestimmte Personen zu berücksichtigen, die die Verantwortung für das Geschehen in den Cloud-Systemen übernehmen. Idealerweise Menschen, die sich für das Thema interessieren und über die notwendigen Kenntnisse verfügen, um die gesetzten Ziele zu erreichen.

6.5 5G is comming

Mit 5G ist pronostiziert, dass mehr Daten[WIE VIELE AN WELCHEM JAHR?] automatisch von Maschinen produziert werden.

6.6 Langfristige Einsparungen sollten größer als Investitionen für Optimierung sein

Kostenoptimierung UND -Überwachung SOLLEN DIE Einsparungen NICHT ÜBERSCHREITEN . TRUSTED ADVISOR NICHT FÜR JEDE FIRMA.

Glossar

Cloud-Computing:

...

Cloud-Dienste:

...

On-Demand:

...

On-Premise:

...

Region:
Die Region ist ein völlig unabhängiges und eigenständiges geografisches Gebiet. Jede Region hat mehrere, physisch getrennte und isolierte Standorte, die als Availability Zones bekannt sind. Beispiele für Regionen sind London, Dublin, Sydney, usw [20], Seite 42.

Availability Zone:
Eine Verfügbarkeitszone ist einfach ein Datenzentrum oder eine Sammlung von Datenzentren. Jede Verfügbarkeitszone in einer Region verfügt über eine separate Stromversorgung, Netzwerk und Konnektivität, um die Gefahr eines gleichzeitigen Ausfalls in beiden Zonen zu verringern [17].

Instance family:
Instanzfamilien sind eine Sammlung von EC2-Instanzen, die nach dem Verhältnis von Speicher, Netzwerkleistung, CPU-Größe und Speicherwerten zueinander gruppiert sind. Zum Beispiel bietet die m4-Familie von EC2 eine ausbalancierte Kombination von Rechen-, Speicher- und Netzwerkressourcen. [18].

Instagram-Story

Tag

Buckets

[17] [20], Seite 42
[18] [20], Seite 95

Instagram-Story

Quellenverzeichnis

6.7 Literatur

[1] Stickel-Wolf, Christine; Wolf, Joachim (2011): Wissenschaftliches Lernen und Lerntechniken. Erfolgreich studieren—gewusst wie!. Wiesbaden: Gabler.

[2] Anders Lisdorf (2021): Cloud Computing Basics: a Non.-Technical Introduction. Apress.

6.8 Internetquellen

[1] Accenture Dienstleistungen GmbH. (Veröffentlicht am 13.11.2020, abgerufen am 12.04.2021). Hohe Erwartungen an die Cloud: Hürden meistern, Mehrwert maximieren
https://www.accenture.com/de-de/insights/technology/maximize-cloud-value

[2] AWS Introduction to EC2 Auto Scaling
https://www.aws.training/Details/Video?id=16387 (Abgerufen am 23.09.2021)

[3] AWS On-Demand Instances
https://aws.amazon.com/de/ec2/pricing/on-demand/ (Abgerufen am 20.10.2021)

[4] AWS-Entwicklerzentrum
https://aws.amazon.com/de/developer/ (Abgerufen am 21.10.2021)

[5] AWS Entwicklung kostenloser Websites und Webanwendungen
https://aws.amazon.com/de/free/webapps/ (Abgerufen am 21.10.2021)

[6] AWS Instance Scheduler(Abgerufen am 04.2021)
https://aws.amazon.com/de/solutions/implementations/instance-scheduler/

[7] AWS S3 Intelligent-Tiering Adds Archive Access Tiers
https://aws.amazon.com/de/blogs/aws/s3-intelligent-tiering-adds-archive-acce

#:~:text=What%20is%20S3%20Intelligent%2DTiering (Veröffentlicht am 09.11.2020)

[8] AWS Reserved Instances Pricing
https://aws.amazon.com/de/ec2/pricing/reserved-instances/ (Abgerufen am 22.10.2021)

[9] AWS für Amazon EC2 Spot Instances
https://aws.amazon.com/de/ec2/spot/pricing/ (Abgerufen am 25.10.2021)

[10] AWS S3 Pricing
https://aws.amazon.com/de/s3/pricing/ (Abgerufen am 25.10.2021)

[11] AWS Databases
https://aws.amazon.com/de/products/databases/learn/ (Abgerufen am 28.10.2021)

[12] AWS Saving Plans Pricing
https://aws.amazon.com/de/savingsplans/compute-pricing/ (Abgerufen am 02.11.2021)

[13] AWS Cloud Watch Features
https://aws.amazon.com/de/cloudwatch/features/ (Abgerufen am 03.11.2021)

[14] AWS Cloud Watch Events: User Guide
https://docs.aws.amazon.com/AmazonCloudWatch/latest/events/cwe-ug.pdf#WhatIsCloudWatchEvents (Abgerufen am 04.11.2021)

[15] AWS Cloud Watch : User Guide
https://docs.aws.amazon.com/AmazonCloudWatch/latest/monitoring/acw-ug.pdf#CloudWatch_Automatic_Dashboards_Focus_Service (Abgerufen am 04.11.2021)

[16] AWS Cloud Watch F.A.Q.
https://aws.amazon.com/de/cloudwatch/faqs/ (Abgerufen am 07.11.2021)

[17] AWS re:Invent 2019: Guidelines and design patterns for optimizing cost in Amazon S3
https://youtu.be/UPzsRk2lFWE?t=1279 (Abgerufen am 18.11.2021)

[18] AWS Pricing Calculator
https://calculator.aws/#/createCalculator/EC2
(Abgerufen am 23.11.2021)

[19] Amazon Simple Storage Service - User Guide
https://docs.aws.amazon.com/AmazonS3/latest/userguide/s3-userguide.pdf#lifecycle-transition-general-considerations
(Abgerufen am 24.11.2021)

[20] AWS Certified Solutions Architect - Associate (SAA-C02)
https://books.google.de/books?id=Dp__DwAAQBAJ&lpg=PA29&ots=T5WqfT25mA&dq=Increase%20efficiencies%3A%20Use%20automation%20to%20reduce%20or%20eliminate%20IT%20management%20activities%20that%20waste%20time%20and%20resources.&pg=PA29#v=onepage&q&f=false
(Abgerufen am 02.11.2021)

[21] Microsoft Customer Story-Walgreens Boots Alliance delivers superior customer service with SAP solutions on Azure
https://customers.microsoft.com/en-us/story/792289-walgreens-boots-alliance-retailers-azure-sap-migration
(Veröffentlicht am 10. Juni 2020)

[22] Bertelsmeier, Birgit (o. J.): Tipps zum Schreiben einer Abschlussarbeit. Fachhochschule Köln-Campus Gummersbach, Institut für Informatik.
http://lwibs01.gm.fh-koeln.de/blogs/bertelsmeier/files/2008/05/abschlussarbeitsbetreuung.pdf (29.10.2013).

[23] Halfmann, Marion; Rühmann, Hans (2008): Merkblatt zur Anfertigung von Projekt-, Bachelor-, Master- und Diplomarbeiten der Fakultät 10. Fachhochschule Köln-Campus Gummersbach.
http://www.f10.fh-koeln.de/imperia/md/content/pdfs/studium/tipps/anleitungda270108.pdf (29.10.2013).

[24] IDC Business Value ofAWS 2015
http://d0.awsstatic.com/analyst-reports/IDC_Business_Value_of_AWS_May_2015.pdf (Abgerufen am 22.10.2021)

[25] Raj Bala, Bob Gill, Dennis Smith, Kevin Ji, David Wright.
Magic Quadrant für Cloud-Infrastruktur und Plattform-Services
https://www.gartner.com/technology/media-products/reprints/AWS/1-271W1OSP-DEU.html (Abgerufen am 23.09.2021 / Veröffentlicht am 27. Juli 2021)

[26] LinkedIn: Listado de todos los Servicios de AWS
https://www.linkedin.com/pulse/listado-de-todos-los-servicios-amazon-web-ser
C3%B1a-silva/?originalSubdomain=es (Abgerufen am 18.11.2021)

[27] Stern, Adam, The Truth About Cloud Pricing
https://www.forbes.com/sites/forbestechcouncil/2018/11/16/the-truth-about-cloud-pricing/?sh=1f37bba42f33
(Veröffentlicht am 16.11.2018)

[28] Putting a Finger on Our Phone Obsession
https://blog.dscout.com/mobile-touches?_ga=2.18241977.1010253397.1637068725-1707869761.1637068725 (Abgerufen am 16.11.2021)

[29] Statista: 2020 überholt die Cloud lokale Speichermedien
https://de.statista.com/infografik/18231/cloud-vs-lokaler-speicher/
(Abgerufen am 18.11.2021)

A Anhang

(To-Do:)

A.1 Anhang X

Erklärung über die selbständige Abfassung der Arbeit

Ich versichere, die von mir vorgelegte Arbeit selbständig verfasst zu haben. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Arbeiten anderer entnommen sind, habe ich als entnommen kenntlich gemacht.
Sämtliche Quellen und Hilfsmittel, die ich für die Arbeit benutzt habe, sind angegeben. Die Arbeit hat mit gleichem Inhalt bzw. in wesentlichen Teilen noch keiner anderen Prüfungsbehörde vorgelegen.

_____

(Ort, Datum, Unterschrift)

Hinweise zur obigen *Erklärung*

- Bitte verwenden Sie nur die Erklärung, die Ihnen Ihr **Prüfungsservice** vorgibt. Ansonsten könnte es passieren, dass Ihre Abschlussarbeit nicht angenommen wird. Fragen Sie im Zweifelsfalle bei Ihrem Prüfungsservice nach.
- Sie müssen **alle abzugebende Exemplare** Ihrer Abschlussarbeit unterzeichnen. Sonst wird die Abschlussarbeit nicht akzeptiert.
- Ein **Verstoß** gegen die unterzeichnete *Erklärung* kann u. a. die Aberkennung Ihres akademischen Titels zur Folge haben.