

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 Wirtschaftsinformatik

ausgearbeitet von:

CARLO MENJIVAR 11117929

Erstprüfer: Prof. Dr. Roman Majewski

Zweitprüfer: Thomas Raser

Gummersbach, im Feb<Monat der Abgabe>

1 Abstract

Zusammenfassung

In dieser Arbeit wird es gezeigt, 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 man die Kosten für Cloud-Dienste wie alle anderen Kosten betrachtet, ist es nur konsequent, über ihre Kontrolle und Optimierung nachzudenken.

Abstract

Platz für das englische Abstract...

Inhaltsverzeichnis

1	Abstract	1
	Abbildungsverzeichnis	5
2	Einleitung	6
2.1	Motivation	6
2.2	Problemstellung	6
2.3	Darstellung der Ergebnisse und Einschränkungen	7
2.4	Fragestellung	7
2.5	Zielsetzung	7
2.6	Struktur der Arbeit	8
3	Grundlagen	9
3.1	Ökonomie des Cloud Computing	9
3.2	Skalierbarkeit	9
3.3	Flexibilität und Agilität	10
3.4	Selbstbedienung	10
3.5	Keine Vorabkosten	10
4	Zahlungsmodelle	11
4.1	On-Demand / Nutzungsabhängige Zahlung	11
4.2	Saving Plans	11
4.3	Versteigerung? / Spot Instanzen	11
4.4	Wann welches Zahlungsmodell?	12
4.5	Vorauszahlung	12
5	Kostenüberwachung	13
5.1	Werkzeuge	13
5.1.1	AWS CloudWatch	13
5.1.2	AWS Cost-Explorer	13
5.1.3	AWS Trusted advisor	13
6	Optimierungsmöglichkeiten	14
6.1	Optimierungsmaßnahmen	14
6.1.1	Auto Scaling Group	14
6.1.2	Automatisiere das Hoch- und Herunterfahren von Dev/Test Umgebungen	14
6.1.3	(auto) Tiering	15
6.1.4	Automatisierung mit Lambda Funktionen	15

6.1.5	Benachrigungen, wenn x% Kapazität unterschreiten wurde	15
7	Zusammenfassung und Ausblick	16
7.1	Bewusstsein in der gesamten Organisation	16
7.2	Die richtige Personen(Owneship)	16
8	Glossar	17
9	Quellenverzeichnis	18
9.1	Literatur	18
9.2	Internetquellen	18
A	Anhang	20
A.1	Anhang X	20
	Erklärung über die selbständige Abfassung der Arbeit	21

Abbildungsverzeichnis

2 Einleitung

2.1 Motivation

Amazon Web Services, kurz AWS, wurde unter anderen ausgewählt wegen ihrer frühen Präsenz (2006) als Cloudanbieter und ihr großes Angebot an Dienstleistungen, welche sich an sehr zahlreiche Anwendungsfälle anpassen.

Eine Recherche von Gartner positioniert AWS als Marktführer in der Magic Quadrant für Cloud-Infrastruktur und Plattform-Services 2021. [7]

Für viele Unternehmen ist eine große Herausforderung, die Kosten von Cloud-Diensten übersichtlich zu halten und Optimierungsmöglichkeit leicht zu erkennen. Zusätzlich besteht die Gefahr, unangenehme Überraschungen in einer Rechnung zu bekommen, weil keine Grenze für den Konsum von Cloud-Diensten festgelegt wurde.

Die Kostenoptimierung für Cloud-Dienste, ist so wichtig, dass wenn keine Optimierungsmaßnahmen ergriffen werden, wird es mit Sicherheit mehr bezahlt werden als bei On-Premise Systeme.

”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.” [2]

2.2 Problemstellung

Die Verwendung von Cloud-Diensten bringt mit sich viele Vorteile, wie zum Beispiel kurzfristige Erhöhung oder Verringerung der Speicher- und Rechenkapazität oder Zugriff auf unterschiedliche Speicherarten, die genau an individuellen Anwendungsfälle passen. Alle diese Lösungen sind in wenigen Minuten erreichbar.

Viele Unternehmen befürchten jedoch, dass der Wechsel von On-Premise zu On-Demand zu hohen Kosten führen könnte.

Diese Bachelorarbeit beschäftigt sich mit dieser Problematik, um herauszufinden, wie Unternehmen mit den passenden Werkzeugen, die Kosten ihrer Cloud-Dienste überwachen und im Blick halten können. Wie mit frühzeitigen Benachrichtigungen, alarmiert wird, wenn Systeme mehr Kosten verursachen als geplant.

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

2.3 Darstellung der Ergebnisse und Einschränkungen

Nach Angaben von Amazon Web Services ist es möglich bis zu n% der Kosten für S3 und bis zu 90% für EC2 zu sparen.¹ Diese Arbeit legt den Fokus auf die Optimierung der oben genannten Diensten. Als Überwachungswerkzeuge für die Kosten werden die AWS CloudWatch, der AWS Cost-Explorer und der AWS Trusted Advisor untersucht.

2.4 Fragestellung

„Bei der Umfrage von Stratecast Predictions 2018, Frost & Sullivan, nannten 53 % der IT-Führungskräfte die Verwaltung der Kosten für den Betrieb von Cloud-Workloads als großes Hindernis.“ [3]

Ausgehend von dieser Beobachtung untersucht die Arbeit die folgenden Fragen.

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

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

2.5 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.

¹AWS: Amazon Web Services

2.6 Struktur der Arbeit

Diese Bachelorarbeit ist in folgenden Kapiteln unterteilt:

In Kapitel 3 befasst sich mit dem Begriff Cloud-Economy und erläutert das Nutzen der Cloud im wirtschaftlichen Sinne. Diese dienen als Grundlage für diese Arbeit.

Kapitel 4 zeigt die verschiedenen Zahlungsmodelle für Amazon Web Services.

Es werden Kriterien vorgestellt, die helfen, sich für das richtige Zahlungsmodell für verschiedene Szenarien zu entscheiden.

In Kapitel 5 werden die Werkzeuge eingeführt, die zur Überwachung der Kosten von Cloud-Diensten eingesetzt werden können.

In Kapitel 6 auf Optimierungsmaßnahmen, (Benachrichtigungen von relevanten Ergebnissen) und Limitierung des Konsums von Cloud-Diensten, insbesondere auf EC2-Instanzen.

3 Grundlagen

Das nächste Grundlagenkapitel listet Erfolgschancen für Unternehmen auf, die Cloud-Dienste in ihre Geschäftsprozesse integrieren.

Es wird erklärt warum die Kostenoptimierung und -überwachung relevant für Unternehmen sind.

Die Optimierungsergebnisse können zu folgenden Ergebnissen führen.

- Die Möglichkeit, die individuellen Kosten verschiedener Projekte die über dieselbe Infrastruktur laufen, zu erkennen. Auf diese Weise ist es auch möglich, zwischen Kunden, die mehr Ressourcen verbrauchen, und solchen, die weniger verbrauchen, zu unterscheiden.
- Eine beachtliche Erhöhung der (finanziellen) Rentabilität im Unternehmen.
- Eine geringere Unsicherheit, bei der Umsetzung von cloudbasierten Systemen.
- ...

3.1 Ökonomie des Cloud Computing

[to check:]

Cloud Economics auf Englisch, basiert auf das On-Demand Prinzip basiert, gibt die Flexibilität, Rechnerkapazität je nach Bedarf entweder manuell oder automatisch anzupassen. Es entfallen große Investitionen in Hardware, wie bei On-Premise-Systemen. Durch den Verzicht auf Hardware entfallen die Kosten für Reparatur, Wartung und eventuell damit verbundene Lizenzen, und im Falle von Systemen, die dies benötigen, entfällt der Bedarf an zuständigem Personal.

Die Nutzung dieser Dienste ist innerhalb weniger Minuten und selbstständig möglich (Selbstbedienung).

Grafik der Kosten On-Premise/Demand?

3.2 Skalierbarkeit

Mit Auto Scaling ist es möglich sicherzustellen, dass die Anzahl der Amazon Server-Instanzen während Nachfragespitzen nahtlos hochskaliert werden, um die Leistung aufrechtzuerhalten, und in Nachfrageflauten automatisch reduziert wird, um die Kosten zu minimieren. Auf diese Weise kann es weniger Zeit mit der Verwaltung von IT-Ressourcen verbracht werden und sich mehr auf ihr Geschäft konzentrieren können.

Das war der Fall von Walgreens in der USA. Sie haben unter anderen 750 virtuelle Maschinen und SAP HANA auf Azure Instanzen migriert.

„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“,
 Dan Regalado: Vice President of Global Technology Transformation and Strategic Partnerships Walgreens Boots Alliance [10]

3.3 Flexibilität und Agilität

3.4 Selbstbedienung

3.5 Keine Vorabkosten

Amazon Web Services bietet ein Pay-as-you-go-Modell für Ressourcen in On-Demand Zahlungsmodell. Diese werden nach einer bestimmten Einheit wie GB/Stunde berechnet für Speichereinheiten.

Im Fall von Server-Instanzen wird es ein Stundensatz auf der Grundlage der Servereigenschaften festgelegt.

Hier einige Beispiele von EC2-Instanzen l.

Region, Betriebssystem, Instance-Typ und vCPU auswählen, um Tarife anzuzeigen

Region USA Ost (Ohio) ▼	Betriebssystem Linux ▼
Instance-Typ Alle ▼	vCPU Alle ▼

363 von 363 verfügbaren Instances werden angezeigt

Instance-Name ▲	On-Demand-Stundensatz ▼	vCPU ▼	Arbeitsspeicher ▼	Speicherung ▼	Netzwerkleistung ▼
a1.medium	0,0255 USD	1	2 GiB	Nur EBS	Bis zu 10 Gigabit
a1.large	0,051 USD	2	4 GiB	Nur EBS	Bis zu 10 Gigabit
a1.xlarge	0,102 USD	4	8 GiB	Nur EBS	Bis zu 10 Gigabit
a1.2xlarge	0,204 USD	8	16 GiB	Nur EBS	Bis zu 10 Gigabit
a1.4xlarge	0,408 USD	16	32 GiB	Nur EBS	Bis zu 10 Gigabit

Abbildung Unterabschnitt 3.5: On-Demand Preise für Amazon EC2 ²

²Vgl. u.a.[8]

4 Zahlungsmodelle

Die Wahl des Zahlungsmodells ist von entscheidender Bedeutung, um den besten Preis für die vertraglich vereinbarten Dienstleistungen zu erzielen.

Die drei von Amazon Web Services angebotenen Zahlungsmodelle werden im Folgenden dargestellt.

Die On-Demand-Option erfordert keine langfristigen Verpflichtungen, sie ist daher die teuerste Alternative, die auf Stundenbasis berechnet wird. Die Saving Plans erfordern den Abschluss von Verträgen über 1 oder 3 Jahre, um günstige Preise zu erhalten.

Spots schließlich sind die billigste Option, haben aber den Nachteil, dass ihre Verfügbarkeit nicht immer garantiert ist.

Jedes Zahlungsmodell hat ihre Vor- und Nachteile und eignet sich für unterschiedliche Anwendungsfälle. Gute Ergebnisse können auch durch die Kombination mehrerer Zahlungsmodelle erzielt werden.

(noch in Arbeit =>) Vor- und Nachteile noch tabellarisch aufzulisten!!

4.1 On-Demand / Nutzungsabhängige Zahlung

Bei diesem Zahlungsmodell besteht keine Notwendigkeit, Budgets 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.

4.2 Saving Plans

Mit Saving Plans, verpflichten wir uns zu einer konstanten Nutzungsmenge, die in \$/Stunden über 1 oder 3 Jahre gemessen wird, und im Gegenzug gewährt Amazon uns einen Rabatt für diese Nutzung. EC2 Spot-Instances können mit On-Demand, Reserved Instances und Saving-Plans kombiniert werden, um sowohl feste als auch dynamische Last abzudecken.

4.3 Versteigerung? / Spot Instanzen

EC2 Spot-Instances bieten die Möglichkeit aus 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.

4.4 Wann welches Zahlungsmodell?

(To-Do:) Was sind die Kriterien für die Auswahl eines oder mehrerer Zahlungsmodelle?

4.5 Vorauszahlung

Zusätzlich zu den oben beschriebenen Zahlungsmodellen motivieren uns einige Anbieter von Cloud-Diensten, im Voraus zu zahlen, im Austausch bieten sie bessere Preise. Dies ist bei Amazon der Fall, das derzeit (2021) drei verschiedene Optionen anbietet: keine, teilweise und vollständige Vorauszahlung.

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

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

5 Kostenüberwachung

5.1 Werkzeuge

(To-Do: Intro)

Welche Metriken und Informationen lassen sich mit den hier erwähnten Werkzeugen finden?

Welche Kosten sind mit der Nutzung dieser Werkzeuge verbunden?

Es sollte gezeigt werden, die Arten/Kategorien von Kunden, die kostenlosen Zugang zu diesen Werkzeugen haben?

5.1.1 AWS CloudWatch

Amazon CloudWatch ermöglicht die Überwachung der Leistung von Diensten. Es ist relevant für alle die Cloud-Dienste verbrauchen, wie DevOps-Manager, Ingenieure und Entwickler. Dieses Werkzeug sammelt operative Daten für die Analyse und Entscheidungsfindung.

Most popular parts of CloudWatch:

Dashboards:

Alarms: notificar o actuar

Metrics

Logs

Events: para on/off EC2 Lo que nos interesa es "Billing"

Para EC2 podemos ver la carga a CPU y luego escalar.

5.1.2 AWS Cost-Explorer

...

5.1.3 AWS Trusted advisor

Die 5 Kategorien von dem Trusted Advisor:

- Kostenoptimierung
- Leistung
- Sicherheit
- Fehlertoleranz
- Leistungsgrenzen.

6 Optimierungsmöglichkeiten

Teil der Information, die mit den Überwachungswerkzeuge gesammelt wurde, bilden die Grundlage für die Optimierungsmaßnahmen.

6.1 Optimierungsmaßnahmen

6.1.1 Auto Scaling Group

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

Folgende Abbildung zeigt das Verhältnis von einer Anwendung über eine Woche ohne Auto Scaling. Die graue Zone entspricht ungenutzte Kapazität einer EC2-Instanz. Dies bedeutet, es wird für ungenutzte Ressourcen bezahlt.

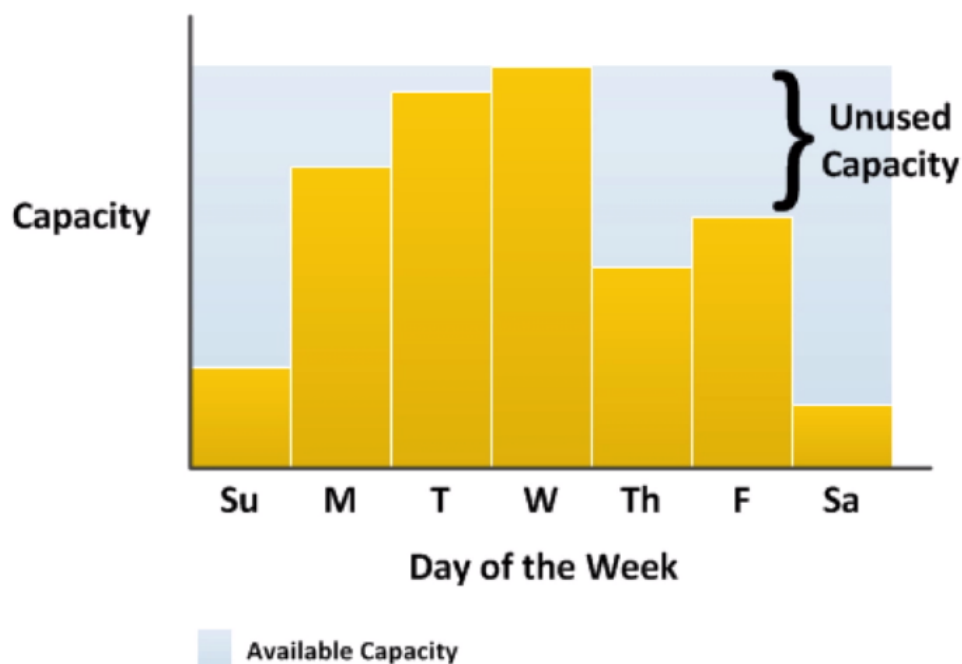


Abbildung Unterunterabschnitt 6.1.1: Ungenutzte Ressourcen ³

6.1.2 Automatisiere das Hoch- und Herunterfahren von Dev/Test Umgebungen

Grund: weil i.d.R., kein Entwickler 24/7 arbeitet. Wie?: mit Tagging, Lambda oder mit Auto Scaling Groups.

³Vgl. u.a.[8]

6.1.3 (auto) Tiering

Grund: nicht alle Dateien brauchen eine hohe Verfügbarkeit.

6.1.4 Automatisierung mit Lambda Funktionen

Grund: einmal programmiert, funktioniert es für immer.

(To-Do:) Möglichkeiten untersuchen, bewerten und die passende Auswählen.

6.1.5 Benachrigungen, wenn x% Kapazität unterschreiten wurde

Grund: um relevante Ereignisse nicht zu verpassen und rechtzeitig Maßnahmen zu ergreifen

warum ist das empfehlungswert?

7 Zusammenfassung und Ausblick

(To-Do:)

7.1 Bewusstsein in der gesamten Organisation

Zusätzlich zu den bisher genannten Maßnahmen ist es wichtig, dass Verantwortliche für die Kostenerzeugung Bewusstsein entwickeln. Von dem Entwickler bis zum IT-Manager, jeder sollte wissen, dass es so einfach ist, Cloud-Dienste mit ein paar Klicks zu beauftragen. Diese können in kurzer Zeit unglückliche/ungeplante Kosten verursachen.

7.2 Die richtige Personen(Owneship)

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.

8 Glossar

Cloud-Computing:

...

Cloud-Dienste:

...

On-Demand:

...

On-Premise:

...

Akronyme:

AWS

Amazon Web Services

Amazon EC2

Amazon Elastic Compute Cloud

DOM

Document Object Model

API

Application Programming Interface

AJAX

Asynchronous JavaScript And XML

CI/CD

Continuous Integration/ Continuous Deployment

9 Quellenverzeichnis

9.1 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. Seite 163.

9.2 Internetquellen

- [1] 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).
- [2] 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).
- [3] 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)
- [4] 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>
- [5] AWS Instance Scheduler (Abgerufen am 04.2021) <https://aws.amazon.com/de/solutions/implementations/instance-scheduler/>
- [6] S3 Intelligent-Tiering Adds Archive Access Tiers <https://aws.amazon.com/de/blogs/aws/s3-intelligent-tiering-adds-archive-access-tiers/#:~:text=What%20is%20S3%20Intelligent%2DTiering> (Abgerufen am 09. 11 2020)

-
- [7] Magic Quadrant für Cloud-Infrastruktur und Plattform-Services
<https://www.gartner.com/technology/media-products/reprints/AWS/1-271W10SP-DEU.html> (Abgerufen am 23.09.2021 / Veröffentlicht am 27. Juli 2021)
- [8] Introduction to EC2 Auto Scaling <https://www.aws.training/Details/Video?id=16387> (Abgerufen am 23.09.2021)
- [9] AWS On-Demand Instances <https://aws.amazon.com/de/ec2/pricing/on-demand/> (Abgerufen am 20.10.2021)
- [10] 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)

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.