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, 1x Feb 2021

Abstract

Zusammenfassung

In dieser Arbeit werden Werkzeuge untersucht, die einen klareren Überblick über die finanziellen Ressourcen schaffen. Mit den gesammelten Informationen dienen sie dazu, direkte Maßnahmen zu ergreifen. Darüber hinaus werden allgemeine Optimierungsmaßnahmen aufgezeigt, die bereits über die Jahre hinweg von anderen Nutzern getestet wurden und von Amazon Web Services (als Best Practices) empfohlen werden.

Die Grundlage dieser Recherche sind Empfehlungen von Cloud-Anbietern bezüglich Kostenüberwachung und -optimierung, Erfahrungen von Experten in dem Fachgebiet und aktuelle Fachliteratur.

Es ist besonders interessant für Teams, die Cloud-Dienste in aktuellen Projekten nutzen und ihre Kosten in der Cloud besser verstehen und optimieren wollen. Wenn die Kosten für Cloud-Dienste wie alle anderen Kosten betrachtet werden, ist es konsequent, über ihre Kontrolle und Optimierung nachzudenken. Ein häufiges Problem ist, dass Kosten entstehen, die sich der Kontrolle der Nutzer entziehen.Aus diesem Grund stehen Unternehmen die bereits On-Premise IT-Infrastruktur nutzen, einem Wechsel kritisch gegenüber, obwohl ihnen die Flexibilität von Cloud-Diensten bessere Wettbewerbsvorteile bieten würde.

Deshalb sind die in dieser Arbeit aufgezeigten Werkzeuge und Maßnahmen relevant für diejenigen, die von einem Wechsel von klassischen Modellen, bekannt als On-Premise, zu Cloud basierten Modellen profitieren möchten.

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

EC Elastic Compute

Einleitung

Motivation

Die zunehmende Digitalisierung von Geschäftsmodellen, die auch durch die Corona-Pandemie vorangetrieben wird[37], lässt Cloud-basierte Applikationen an Bedeutung gewinnen[36]. Als direkte Folge davon steigt die Nachfrage nach Server- und Speicherkapazität. Amazon Web Services, kurz AWS, wurden unter anderem als Fallbeispiel 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[1] für Cloud-Infrastruktur und Plattform-Services 2021[28].

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." [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 Befragten haben geäußert, dass sie Schwierigkeiten haben, alle Kosten für Cloud-Workloads zu rechtfertigen.

[1] Der Magic Quadrant ist eine zweidimensionale Matrix mit vier Quadranten. Jeder Quadrant steht für einen Unternehmenstypus im Markt. Im Uhrzeigersinn von links unten beginnend sind dies: Niche Players, Challengers, Leader und Visionaries.

[2] [2], Seite 152

> „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.“ [33]

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. Außerdem sollte untersucht werden, wie mit der richtigen Auswahl an Diensten Kosten optimiert werden. In dieser Arbeit wird versucht zu beantworten, wie Kosten bei Cloud-Diensten überwacht werden können. Auf Grundlage dieser Information werden Optimierungsmöglichkeiten gesucht. Es wird untersucht, welche Maßnahmen nötig sind, um unerwartet hohe Kosten bei Cloud-Diensten zu vermeiden. Darüber hinaus werden Empfehlungen von Cloud-Experten berücksichtigt, um Kosten von Cloud-Diensten zu minimieren. Diese Arbeit untersucht speziell S3-Speichereinheiten und EC2-Server-Instanzen.

Zielsetzung

Die vorliegende Arbeit betrachtet die von AWS angebotenen Überwachungswerkzeuge, um ein tiefergehendes Verständnis der Entstehung von Kosten durch die Nutzung von Cloud-Diensten zu gewährleisten. Mit den von AWS zur Verfügung gestellten Maßnahmen sollen die Nutzung und damit die Kosten von Cloud-Diensten reduziert werden.

Struktur der Arbeit

Diese Bachelorarbeit ist in folgende Kapitel unterteilt:

Kapitel 1 befasst sich mit dem Begriff Cloud-Economy und erläutert das Potenzial der Cloud-Diensten im wirtschaftlichen Sinne. Die Cloud-Dienste EC2-Instanzen und S3 Speichereinheiten werden ebenfalls kurz erklärt.

Kapitel 2 zeigt die verschiedenen Zahlungsmodelle für EC2-Instanzen. Es werden Kriterien vorgestellt, die helfen sollen, sich für das richtige Zahlungsmodell bei verschiedenen Szenarien zu entscheiden.

In Kapitel 3 werden die Werkzeuge eingeführt, die zur Überwachung der Kosten von Cloud-Diensten eingesetzt werden.

Kapitel 4 befasst sich mit Optimierungsmaßnahmen insbesondere für EC2-Instanzen und S3 Speichereinheiten.

1 Grundlagen

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

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

- Die Möglichkeit, die Kosten verschiedener Projekte, die über dieselbe Infrastruktur laufen, zu trennen. Auf diese Weise kann zwischen Projekten, die mehr, und Projekten, 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) [3].[ZITAT]

[AUFLISTUNG WIRD IN FLIEßTEXT NEU FORMULIERT]

1.1 Cloud Economics

Cloud Economics untersucht die Kosten und die Vorteile von Cloud Computing und die, der dahinterstehenden wirtschaftlichen Grundsätze. Das 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 verbundenen Lizenzen.
Der Cloud-Anbieter übernimmt viele Verwaltungsaufgaben. Das führt zu einer Abnahme der nötigen Fachkraft. [27]
Die Nutzung von Cloud-Diensten ist in unabhängiger Weise möglich; in Selbstbedienung und mit der Freiheit Ressourcen ohne Einschränkungen zu nutzen. Das bedeutet jedoch gleichzeitig, dass der Nutzer Verantwortung für die anfallenden Kosten übernimmt[4].

[Grafik der Kosten On-Premise/Demand?]

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

1.1.1 Skalierbarkeit

Um die Leistung der Ressourcen aufrecht zu halten und diese bei Abnahme der Nachfrage zu reduzieren, ist es möglich die Rechenkapazität hoch- und runterzuskalieren.

Auf diese Weise kann Zeit mit der Verwaltung von IT - Ressourcen gespart werden, welche dann genutzt werden kann, um sich auf die wesentlichen Geschäftsaktivitäten zu konzentrieren[5].

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.„
>
> [23]

1.1.2 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.[5]
- Hosting für statische Webseiten und Webanwendungen. [6]
- Populäre relationale und nicht relationale Datenbanken. [11]
- Vielfältige Hardware-Konfigurationen.

[5] [3], Seite 29

Durch die Vielzahl der verfügbaren Ressourcen ist es möglich, Prototypen und Experimente in kurzer Zeit durchzuführen. [6].
Softwareprojekte können schnell auf den Markt gebracht werden. Je nach ihrem Erfolg ist es möglich, sinnvolle Entscheidungen zu treffen. Wenn ein Projekt, aus welchen Gründen auch immer, kurzfristig eingestellt werden muss, könnten alle damit verbundenen Kosten ausfallen.
Denn im Gegensatz zu On-Premise-Infrastrukturen gibt es keine Bindung an kostspielige Hardware.

1.1.3 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. [TRÄGT DIESES ETWAS ZUR ARBEIT BEI?]

1.1.4 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 die Anfangsinvestition in die IT-Infrastruktur oder fällt ganz weg. Es ist zu bedenken, dass weitere Investitionen wie technische Schulungen für das Personal erforderlich werden. TÜV Rheinland bietet Kurse zur Ausbildung von Cloud Architekten an. Die Kurse dauern drei Tage und kosten 2.136,05 € pro Teilnehmer. Maßnahmen wie die genannten Kurse wirken einem der Hauptprobleme entgegen, mit denen Unternehmen bei der Migration in die Cloud konfrontiert sind. In der von Accenture im Jahr 2020 durchgeführten Umfrage gaben 38% der Befragten an, dass fehlende Kompetenzen im Unternehmen im Bezug auf die Cloud ein Hindernis für eine Cloud-Migration ist[7].

1.2 Amazon Cloud-Dienste

Einer der am häufigsten genutzten AWS-Dienste ist Amazon Elastic Computing Instances EC2, mit dem virtuelle Maschinen erstellt werden können[38]. Ein weiterer wichtiger Dienst ist Amazon Simple Storage Service (S3), der zum Speichern von Objekten verwendet wird. Deshalb konzentrieren sich in dieser Arbeit die Überwachungs- und Optimie-

[6] [27], Seite 7
[7] [2], Seite 11

rungsmaßnahmen hauptsächlich auf EC2-Instanzen und S3-Speichereinheiten. Wie Lynn Langit, eine erfahrene Cloud-Architektin, feststellt, können bis zu 80% der Rechnung aus Gebühren für EC2-Instanzen bestehen[8].

Objekte sind in AWS die Grundeinheit in welchen Dateien in den S3-Speichereinheiten gespeichert werden. Neben den Objekten werden Metadaten, wie das Datum der Objekterstellung und das Datum der letzten Aktualisierung gespeichert. Laut des AWS Solutions Architekten Daniel Peña Silva[35] ist Amazon S3 einer der am häufigsten genutzten AWS-Dienste.

Wie in Abbildung 1 zu sehen ist, werden darüber hinaus seit 2020 weltweit mehr Daten in Serverfarmen als auf lokalen Geräten gespeichert. Dies bietet Vorteile im Bezug auf die Geschwindigkeit der Arbeitsabläufe, birgt aber auch Risiken wie Datendiebstahl. Dies wird in dieser Arbeit nicht behandelt; da es den Rahmen der Recherche sprengen würde.

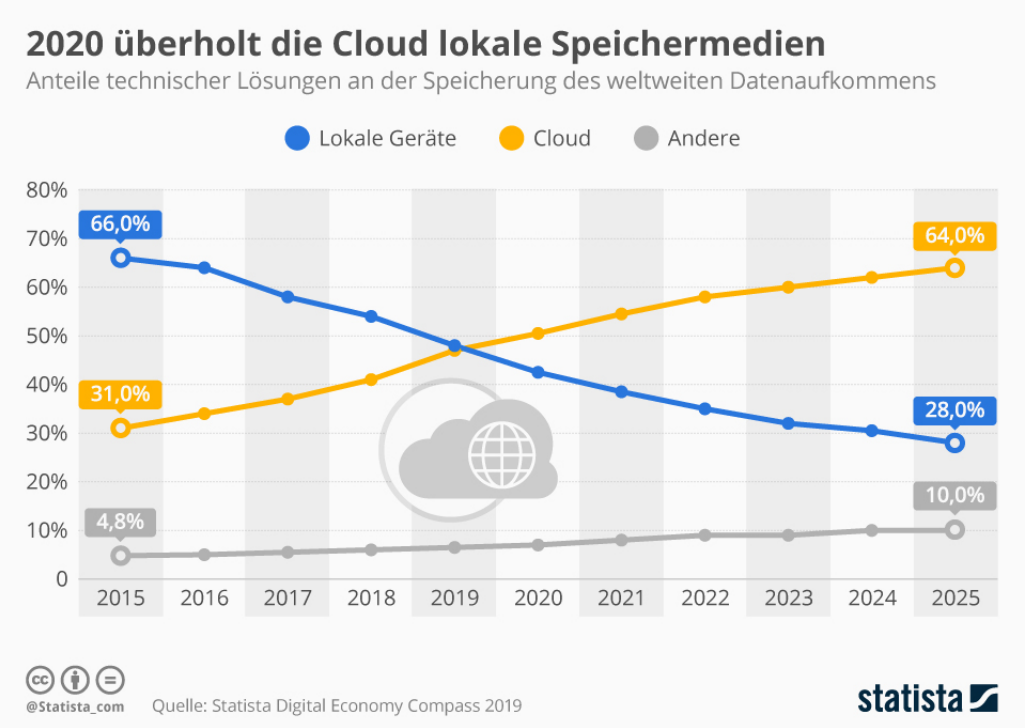

Abbildung 1
2020 überholt die Cloud lokale Speichermedien [35]

Dieses grundlegende Kapitel hat einige potenzielle Vorteile der Nutzung von Cloud-Diensten für Unternehmen aufgezeigt. Darüber hinaus geht der Trend in den letzten Jah-

[8] [30], Sektion 2 Control Costs by Service, Video AWS service type Minute 00:30

ren zur Nutzung von Cloud-basierten Diensten. Das nächste Kapitel befasst sich mit Zahlungsmodellen für EC2-Instanzen und den Überlegungen, die bei der Wahl dieser Modelle in verschiedenen Szenarien zu berücksichtigen sind.

2 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, es ist daher die teuerste Alternative, die auf Stundenbasis berechnet wird. Die Modelle Saving Plans und Reserved Instances erfordern den Abschluss von Verträgen über ein oder drei Jahre, um günstige Preise zu erhalten. EC2-Spot-Instanzen sind das kostengünstigste Modell, sie 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[ZU ERGÄNZEN]. Dies wird in Unterkapitel 2.4 behandelt.

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 wird empfohlen Instanzen so nahe wie möglich an den Ressourcen[9], mit denen sie kommunizieren werden, zu platzieren. Die beste Leistung wird außerdem angestrebt, indem sich diese Instanzen in räumlicher Nähe zur Mehrzahl der Endnutzer, befinden.

2.1 On-Demand-Instanzen

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, deren Entwicklung unvorhersehbar ist und die Möglichkeit besteht, dass das es in kurzer Zeit abgeschlossen sein wird, sodass es nicht Sinnvoll ist, eine langfristige Verpflichtung einzugehen.

Die Preise beim dem On-Demand Zahlungsmodell variiert je nach Instanz Typ, Region und der übertragenen Datenmenge. Die aktuellen Preise für die verschiedenen Regionen sind auf der Amazon-Website in der Sektion EC2 - On-Demand-Preise[10] zu finden. In der Abbildung 2 werden die für die Region Ohio verfügbaren Linux-Instanzen gezeigt.

[9] Mit Ressourcen sind Cloud-Dienste gemeint
[10] https://aws.amazon.com/de/ec2/pricing/on-demand/

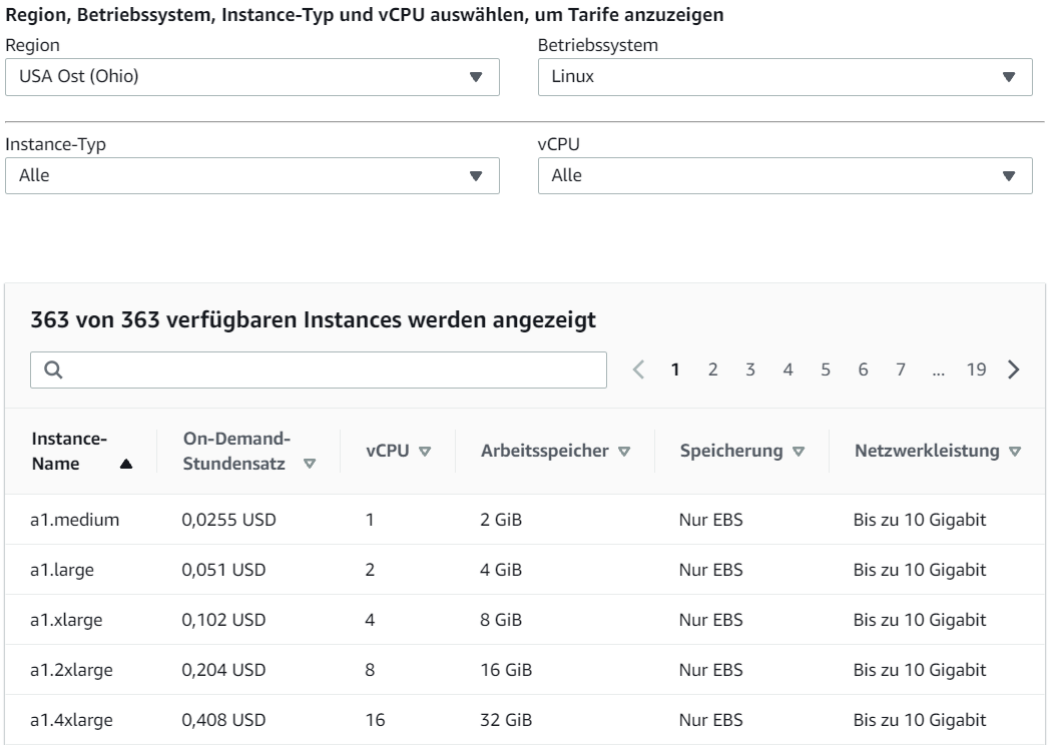

Abbildung 2
On-Demand Preise für Amazon EC2 [4]

Es ist zu beachten, dass Instanzen mit denselben Eigenschaften, aber in verschiedenen Regionen, unterschiedliche Preise haben können.

2.2 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 über ein oder drei Jahre abgeschlossen werden. Nachfolgend werden die prozentualen Einsparungen gemäß des jeweiligen Modells gezeigt.

Einsparungen nach Modell			
Compute Saving Plans	EC2-Instance Saving 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 Saving Plans die Flexibilität bieten, EC2-Instanzen nach

Familie[11], Größe, Availability Zone (AZ), Betriebssystem oder Mandant zu wechseln.

> „Bei Compute Saving 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 Saving 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.

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

Die Abbildung 3 zeigt den Vergleich zwischen den drei Optionen für Vorauszahlungen. Hier wird deutlich, dass es kaum einen Unterschied zwischen eine teilweise Vorauszahlung und keine Vorauszahlung zu machen gibt. Eine erhebliche Einsparung ergibt sich, wenn man für den gesamten Zeitraum der gebuchten Instanzen im Voraus bezahlt.
Die Berechnungen wurden mit dem AWS Pricing Calculator [18] für Instanzen der Familie t4g.xlarge, in der EU (Frankfurt) und für eine Laufzeit von 3 Jahren durchgeführt.

2.3 Spot Instanzen

EC2 Spot-Instanzen bieten die Möglichkeit aus den ungenutzten EC2-Instanzen anderer Nutzer 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.

Unterbrechbarkeit

Es ist zu beachten, dass Spot-Instanzen jederzeit unterbrochen werden können. Einer der Gründe ist die Preisüberschreitung der Instanz. Wenn Spot-Instanzen angefordert

[11] [3], Seite 95

Zahlungsmodell		EC2 Instance Saving Plans	
Anzahl der Instanzen	20		
Dauer	36	Monate	
Vorauszahlung	**keine**	**teilweise**	**vollständig**
Gesamtkosten pro Monat	$967.98	$519.62	$0.00
Vorabkosten gesamt	$0.00	$16,135.92	$30,327.12
Gesamtbetrag	$34,847.28	$34,842.24	$30,327.12
Prozentuale Einsparung	-	0.01%	12.96%
Monetäre Einsparung	-	$5.04	$4,515.12

Ohne Elastic Block Storage (EBS)

Abbildung 3
Mögliche Einsparungen durch Vorauszahlungen für EC2 Instanzen in Saving Plans Zahlungsmodell
Eigene Darstellung. Quelle: [18]

werden, wird einen Maximalpreis festgelegt. Ist der Preis der Spot-Instanz höher als der eingegebene Maximalpreis, ist die Spot-Instanz für die aktuelle Einstellung nicht mehr verfügbar. Ein anderes Szenario ist, wenn der Instanz Anbieter die Spot-Instanz erneut anfordert. Falls eine Spot-Instanz unterbrochen wird, benachrichtigt Amazon EC2 zwei Minuten im Voraus. Dieses Ereignis ist verfügbar auf CloudWatch, damit weitere Alarmen eingestellt werden. Diese und andere Funktionalitäten von CloudWatch werden in Kapitel 3 näher erläutert.
Da Spot-Instanzen anfällig für Unterbrechungen sind, ist es nicht empfehlenswert, für Produktionsumgebungen nur Spot-Instanzen zu verwenden.

2.4 Wahl des Zahlungsmodells

Die folgende Tabelle fasst die Eigenschaften der Zahlungsmodelle für EC2-Instanzen zusammen und listet typische Applikationen je nach Zahlungsmodell auf. [Abb. VOLLSTÄNDIG?]

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 von den Zahlungsmodellen zu nutzen. Beginnend mit dem On-Demand-Zahlungsmodell, gefolgt von Reserved Instanzen

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 4
Vergleich der Zahlungsmodelle nach Eingenschaft und Anwendungsfall
Eigene Darstellung. Quelle: [4, 8, 12, 20], [31] Seite 9

und Saving Plans. In dieser Reihenfolge sinkt der Preis und mit ihm steigt die Verpflichtung, sich langfristig zu binden. Schließlich mit Spot-Instanzen, die die niedrigsten Preise bieten, aber keine volle Verfügbarkeit sicherstellen.

Im nächsten Kapitel wird CloudWatch[UND...] 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 4 näher eingegangen.

3 Kostenüberwachung

In diesem Kapitel werden Werkzeuge vorgestellt, mit denen Budgets mit Alarmen erstellt werden, diese informieren die Nutzer, wenn ein bestimmter Prozentsatz des festgelegten Budgets überschritten wurde. Die Erstellung von Budgets trägt zu einer besseren Planung-/Prognose- und Kostenkontrolle. [WEIL, Cost Controlling] 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 nächsten Kapitel genauer behandelt. [TRUSTED ADVISOR fehlt noch in der Einleitung] [SOLLTE AB HIER EINE NEUE UNTERKAPITEL ANFANGEN? zb. TAGS ZUR TRENNUNG DER Ressourcen] 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.

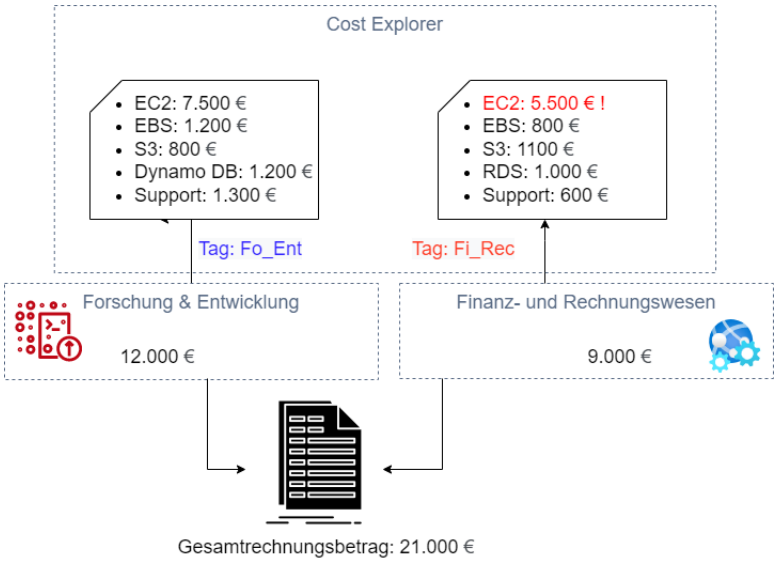

Abbildung 5
Trennung der Kosten durch Tags.
Die Angaben dienen nur als Beispiel und entsprechen keiner realen IT-Infrastruktur.

Es könnte zum Beispiel ein Szenario entstehen, in dem eine Abteilung innerhalb einer Organisation mehr Kosten verursacht als Andere. Dies ist nur durch den Anstieg 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.

3.1 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 Metriken, die mit CloudWatch kompatibel sind, 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 Benachrichtigungen und Aktionen konfiguriert werden. Eine dieser Aktionen ist die automatische Einrichtung neuer Instanzen zur Deckung des Kapazitätsbedarfs[12]. Diese Art von Aktionen werden im Kapitel 4 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 eine Sekunde zulassen. Ein praktisches Beispiel für benutzerdefinierte Metriken ist die Messung der Ladezeit einer Website. [ANDERES BEISPIEL?]

Ereignisse
Bei CloudWatch wird eine Änderung bei einer Ressource in der AWS-Umgebung Ereignis genannt. AWS-Ressourcen können Ereignisse erzeugen, wenn sich ihr Status ändert. Beispielsweise, wird ein Ereignis erzeugt, wenn Amazon EC2 Auto Scaling, Instanzen gestartet oder beendet werden[13] oder wenn eine bestimmte Menge an Speicherplatz in einem

[12] [3], Seite 185
[13] [14], Seite 1

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[14].

Ziele

Ziele oder Targets sind Ressourcen, die aufgerufen werden, wenn eine Regel ausgelöst wird. EC2 instances, AWS Lambda functions und Amazon SNS sind unter anderem mögliche Ziele. Die Ziele einer Regel müssen sich in derselben Region wie die Regel befinden [15].

Benachrichtigungen

Benachrichtigt zu werden ist wichtig, um relevante Ereignisse nicht zu verpassen und rechtzeitig Maßnahmen zu ergreifen. Mit CloudWatch können Alarme eingerichtet werden, die durch Metriken wie die CPU-Auslastung und auch Gebühren[PASSENDES WORT?] auf AWS-Rechnungen ausgelöst werden. Benachrichtigungen können durch Amazon SNS[16] oder zu einer E-Mail-Adresse geschickt werden.

Visualisierung von Metriken

Mit Cloud-Watch Dashboards können relevante Metriken grafisch dargestellt werden. Durch die Dashboards können auch Benachrichtigungen erstellt werden. Für die Einrichtung der Benachrichtigungen ist kein technisches Wissen nötig[17]. 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 nützlichen Informationen bekommen, um Prozesse zu beschleunigen[VORTEILE DES OPTIMIERUNGSPROZESS VERWEISEN] oder Probleme schneller zu beheben[USE CASE].

Dies ermöglicht einen schnelleren Kommunikationsfluss in Echtzeit[Fach Informationsmanagement: (Schwachstellenanalyse) IuM Technik/Check Book KCrmar]. Die Zugriffsverwaltung für geteilte Dashboards wird über AWS Identity and Access Management abgewickelt[18][UND?].

[14] [14], Seite 2
[15] [14], Seite 2
[16] Amazon SNS ist ein Dienst, der die Benachrichtigung an Personen und an Applikationen ermöglicht. https://aws.amazon.com/de/sns/
[17] [15], Seite 28
[18] [15], Seite 18 und 39

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. [BEZUG AUF PROJEKTMANAGEMENT/-Kostenkontrolle?]

Benachrichtigung bei Hoch- und Runterfahren von EC2-Instanzen

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 ....]

3.2 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 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 [19].

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.

Budgetplanung

Die Budgetplanung ist eine Methode der Kostenkontrolle, die beim Start eines neuen Projekts eingesetzt wird[26]. Der Bericht über die Nutzung und die in den letzten 12 Monaten entstandenen Kosten zusammen mit der Prognose der Kosten der kommenden drei Monaten tragen zu einen guten Budgetplanung bei. Durch die Möglichkeit, die in den letzten Monaten angefallenen Kosten nach Ressourcen, Projekt oder Abteilung zu trennen,

[19] https://aws.amazon.com/de/aws-cost-management/pricing/

ist es möglich, operative Budgetplanungen aus vergangenen Projekten mit Genauigkeit zu erstellen.

> Bei der operativen Planung wird von einem Zeithorizont von einem Jahr ausgegangen. Hier liegt der Fokus darauf, Ressourcen konkret zuzuweisen und detailreicher zu planen. Welche Mittel werden wofür verwendet und welche kurz- und mittelfristigen Ziele sollen durch diesen Mitteleinsatz erreicht werden.[25]

3.3 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[20].

Abbildung 6
AWS Trusted Advisor Kategorien[21]

Die Abbildung 6 zeigt die fünf 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

[20] https://aws.amazon.com/de/premiumsupport/technology/trusted-advisor/

oder Cost-Explorer.

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. Auch nicht zugewiesene Elastic IP-Adressen erzeugen Kosten. Diese können gegebenenfalls von Trusted Advisor gefunden werden. [BEISPIELE]

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. [GLEICHE GRENZEN WIE BEI CloudWatch?] 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. [ZEIGE BEISPIELE FueR Empfehlungen]

3.4 Überwachungswerkzeuge gemäß ihrer Verwendung

Die Abbildung 7 fasst die Überwachungswerkzeuge zusammen und listet die Verwendung der einzelnen Werkzeuge auf. [NOCH NICHT VOLLSTÄNDIG]

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 7
Überwachungswerkzeuge gemäß ihrer Verwendung
Eigene Darstellung[13, 21, 22].

Handlungsempfehlungen

[SIND SIE HIER RICHTIG PLAZIERT? SOLLTEN LIEBER IN FAZIT SEIN?;NOCH ZU VERVOLSTÄNDIGEN]

Handlungsempfehlung 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.

Handlungsempfehlung 2:

Ein Berater für eine Prüfung und Optimierung der Ressourcen kann in Deutschland zwischen x und N-EUR kosten. Dies ist eine Alternative zu den Plänen des Trusted-Advisor. Ein Berater, der alle 5 Kategorien abdeckt, könnte [BETRAG] kosten.

Fazit

In diesem Kapitel wurde gezeigt, dass es mit CloudWatch möglich ist, Alarme auf Basis von Ereignissen einzurichten, die mit Amazon SNS 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, dass mit Cost-Explorer eine Analyse von Kosten der letzten 12 Monate, eine Einschätzung der Kosten im aktuellen Monat und eine Prognose für die nächsten Monate möglich ist. Kosten können nach Tags und anderen Filtern getrennt werden. Diese Informationen dient unter anderem zur Erstellung einer operativen Budgetplanung mit genaueren Daten. Darüber hinaus wurde Trusted Advisor vorgestellt, die konkrete Optimierungsempfehlungen gibt und warnt über Leistungsgrenzen. Dies kann mit erheblichen Kosten verbunden sein und ist daher nicht für alle Arten von Unternehmen unmittelbar attraktiv. Obwohl sich nicht alle Unternehmen die Prüfungen von Trusted Advisor leisten können, sollten die kostenlosen Empfehlungen im Überwachungs- und Optimierungsplan berücksichtigt werden. [WAS KOMMT IN NÄCHTEN KAP.?]

4 Optimierungsmaßnahmen

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

4.1 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 8 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.

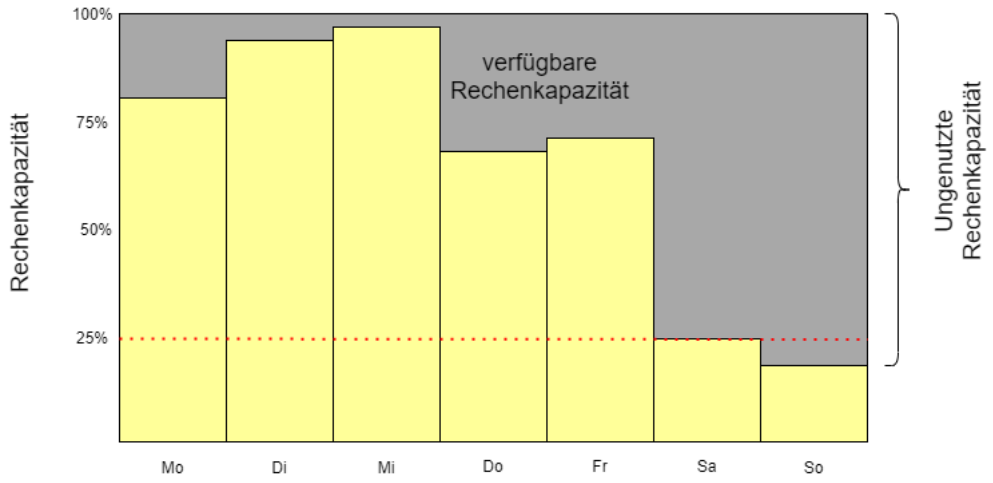

Abbildung 8
Ungenutzte Rechenkapazität ohne automatische Skalierung.
Quelle: Eigene Darstellung mit [fiktiven?] Angaben.

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

4.1.1 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 9
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 Cyber-Monday und Black Friday[39]. [WEITERE ERKLÄRUNG]

4.1.2 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 10 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. Die für die automatische Skalierung erforderlichen Metriken wurden

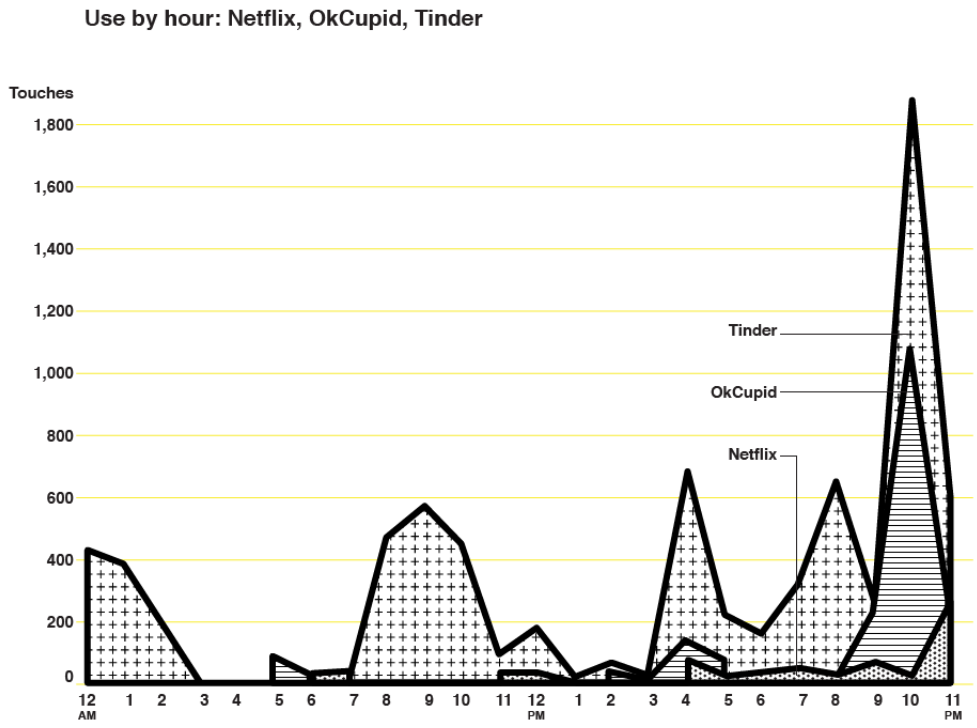

Abbildung 10
Nutzung von Tinder, OkCupid und Netflix pro Stunde. [34]
Mit Touches sind die Anzahl der Klicks, Swipes oder einfachen Interaktionen mit der Applikation gemeint.

bereits im Kapitel Überwachungswerkzeuge erwähnt. Eine der Metriken, die von Cloudexperten [Telefonat mit Deepak von der Einheit CIS[Cloud Capgemini]]benutzt wird, ist

die gesamte CPU-Auslastung. 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.

4.1.3 Manual Scaling

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.

4.1.4 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. [(EIGENES) DIAGRAMM] Anders als Zeitgesteuerte Skalierung ist es nicht notwendig, die Verhaltensmuster der Anwendungen zu analysieren.

4.2 S3 Optimierung

4.2.1 Die 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 Daten über einen längeren Zeitraum gespeichert werden müssen, weil die Anforderungen

der Applikation dies vorschreiben oder für den Fall dass, per Gesetz auf die Informationen in der Zukunft zugegriffen werden muss. [UMFORMULIEREN] 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.

4.2.2 Lebenszyklus-Konfiguration

Eine S3-Lebenszykluskonfiguration oder lifecycle policy 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 11 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, \cite{AMZ18}
```

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, 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.

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)**
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 (Nächste 450 TB)**	**S3 Standard Infrequent Access**	**S3 Glacier**
Speicherplatz in GB	51200	10240	24576	36864
Preis pro GB	$0.0245	$0.0235	$0.0136	$0.0045
Speicherverteilung	42%	8%	20%	30%
Anzahl der Dateien	15059	3012	7228	10842
Übertragungsgebühr (pro 1.000 Aufrufe)	-	-	$0.0100	$0.0360
Kosten für Verschiebung	0	0	$0.72	$3.90
Speicherkosten	$1,254.40	$240.64	$334.23	$165.89
Gesamtkosten				$1,999.79

Abbildung 11
Kostenvergleich durch Nutzung von unterschiedlichen Speicherklassen.
Quelle: Eigene Darstellung. S3 Stundensätze: [10]
Der Punkt wurde als Dezimaltrennzeichen und das Komma als Tausendertrennzeichen verwendet.

4.2.3 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 [21].

Abbildung 12 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

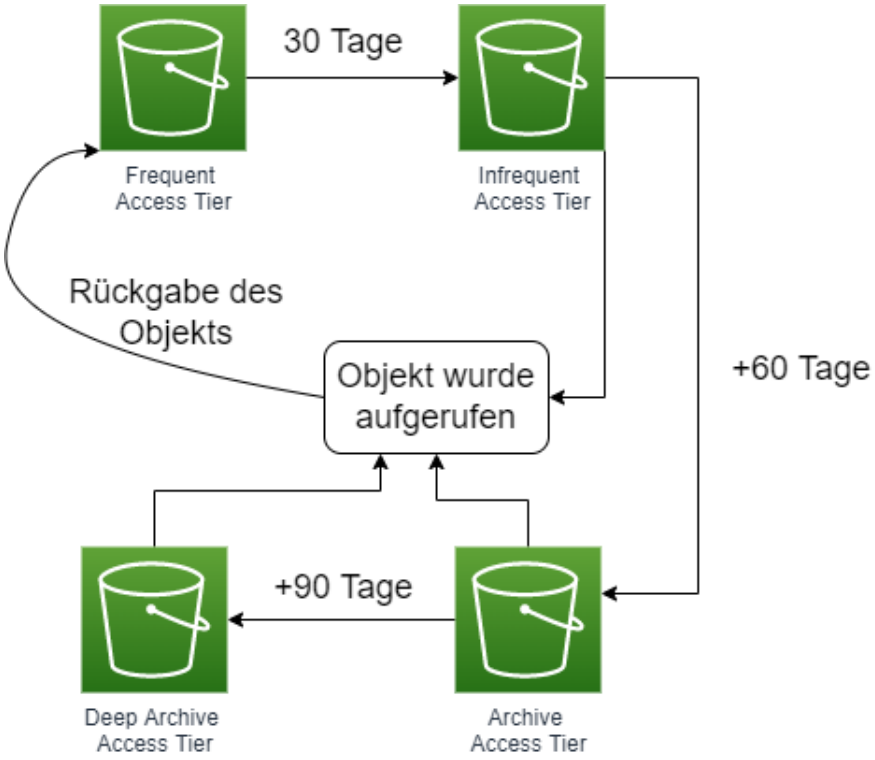

Abbildung 12
Funktionsweise von Intelligent-Tiering
Quelle: Eigene Darstellung

Ebene der seltenen Zugriffe aufgerufen, wird es automatisch in eine Speicherklasse der häufigen Zugriffe zurückversetzt.

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

Zusammenfassung und Ausblick

Umweltbezogene Aspekte

[WORK IN PROGRESS]
In dieser Arbeit geht es um die Überwachung und Optimierung von Cloud-Diensten unter finanziellen Aspekten.

Serverfarmen und ihre Speichereinheiten haben Auswirkungen auf die Umwelt, da sie eine große Menge an Strom benötigen. [t.ly/XYMJ]

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.

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[22]. Diese können in kurzer Zeit ungewünschte Kosten verursachen oder sogar über Jahre hinweg wirtschaftliche Schäden verursachen.

Die richtige Personen finden, 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

[22] [31], Seite 5

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.

5G/IoT generierte Daten

Mit 5G ist pronostiziert, dass mehr Daten[WIE VIELE / WANN?] automatisch und schnell von Maschinen produziert werden.

Rentabilität bei der Optimierungsmaßnahmen

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 [3], 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 [23].

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. [24].

Instagram-Story

[23] [3], Seite 42
[24] [3], Seite 95

Tag

Buckets

Quellenverzeichnis

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.
ISBN-13 (pbk): 978-1-4842-6920-6

[3] 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
ISBN: 9780137325160
(Abgerufen am 02.11.2021)

Internetquellen

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

[2] Accenture GmbH: Navigating the barriers to maximizing cloud value
https://www.accenture.com/de-de/insights/technology/maximize-cloud-value
(Veröffentlicht July-August 2020, abgerufen am 29.11.2021)

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

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

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

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

[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] Amazon EC2-Spot-Instances
https://aws.amazon.com/de/ec2/spot/?cards.sort-by=item.additionalFields.startDateTime&cards.sort-order=asc
(Abgerufen am 26.11.2021)

[21] AWS Trusted Advisor
https://aws.amazon.com/de/premiumsupport/technology/trusted-advisor/
(Abgerufen am 26.11.2021)

[22] AWS Cost Explorer
https://aws.amazon.com/de/aws-cost-management/aws-cost-explorer/
(Abgerufen am 26.11.2021)

[23] 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)

[24] 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 (Veröffentlicht am 29.10.2013).

[25] SevDesk: Definition von Budgetplanung https://sevdesk.de/lexikon/budgetplanung/#budgetplanung-definition (Abgerufen am 28.11.2021)

[26] Indeed:Cost Control Methods: Definitions and Examples https://www.indeed.com/career-advice/career-development/cost-control-methods (Abgerufen am 29.11.2021)

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

[28] 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)

[29] 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)

[30] LinkedIn Learning: AWS Controlling Cost by Lynn Langit https://www.linkedin.com/learning/aws-controlling-cost/aws-service-types?autoAdvance=true&autoSkip=false&autoplay=true&resume=false&u=79182202 (Abgerufen am 29.11.2021)

[31] Plusserver: Kostenoptimierung in AWS https://get.plusserver.com/hubfs/Assets/aws/a/Whitepaper-Kostenoptimierung-in-AWS-DE.pdf?utm_campaign=IoT&utm_medium=email&_hsmi=188763947&_hsenc=p2ANqtz--pG4zb_6horYqX3d0QDpUAzNYdJL51HEBdAtK3IQRBKUfR226JxBly6n2ILDtAmkmPwlib5J7qYjL1Oc6Fs utm_content=188763947&utm_source=hs_automation (Abgerufen am 29.11.2021)

[32] TÜV Rheinland: Kurse zur Ausbildung von Cloud Architekten
https://akademie.tuv.com/weiterbildungen/architecting-on-aws-489176?
(Abgerufen am 29.11.2021)

[33] 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)

[34] 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)

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

[36] Statista: Wie schätzen Sie die Bedeutung Cloud-basierter Anwendungen in Ihrem Unternehmen ein?
https://de.statista.com/statistik/daten/studie/1221723/umfrage/
umfrage-zur-bedeutung-cloud-basierter-anwendungen-im-handel/ (Abgerufen am 25.11.2021)

[37] Statista: Corona-Krise: Anteile der Unternehmen mit geplanten Veränderungen im Arbeitsalltag nach Arbeitsbereichen in Deutschland im 2. Quartal 2020
https://de.statista.com/statistik/daten/studie/1140069/umfrage/
corona-krise-veraenderungen-im-arbeitsalltag/ (Abgerufen am 25.11.2021)

[38] Statista: Cloud infrastructure services vendor market share worldwide from 4th quarter 2017 to 3rd quarter 2021
https://www.statista.com/statistics/967365/
worldwide-cloud-infrastructure-services-market-share-vendor/ (Abgerufen am 25.11.2021)

[39] Statista: Wie viel planen Sie am Black Friday / Cyber Monday auszugeben?
https://de.statista.com/statistik/daten/studie/1074692/umfrage/
hoehe-der-geplanten-ausgaben-am-black-friday-und-cyber-monday-in-deutschlanc
(Abgerufen am 29.11.2021)

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.